

# UPC

UNIVERSITAT POLITÈCNICA DE CATALUNYA  
Departament d'Enginyeria Electrònica

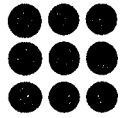
## TÉCNICAS DE REDUCCIÓN DEL RUIDO DE CONMUTACIÓN EN CIRCUITOS INTEGRADOS CMOS

Tesis doctoral presentada  
para la obtención del título de  
doctor Ing. Telecomunicación

*José Luis González Jiménez*

Director:  
*Antonio Rubio i Solà*





**UPC**

UNIVERSITAT POLITÈCNICA DE CATALUNYA  
Departament d'Enginyeria Electrònica

**TÉCNICAS DE REDUCCIÓN DEL RUIDO DE CONMUTACIÓN  
EN CIRCUITOS INTEGRADOS CMOS**

Tesis doctoral presentada  
para la obtención del título de  
doctor Ing. Telecomunicación

*José Luis González Jiménez*

Director:  
*Antonio Rubio i Solà*



*A Carmen, Jesús, Javier, Chechu, Jordi y Marta,  
y a todos mis compañeros del Departamento  
de Ingeniería Electrónica*



# Índice

---

<b>Agradecimientos</b>	<b>xi</b>
<b>Capítulo 1: Introducción</b>	<b>1</b>
1.1 Tecnología de fabricación de circuitos integrados CMOS y BiCMOS.....	1
1.1.1 Características específicas de los circuitos integrados .....	1
1.1.2 Estructura típica de un circuito integrado.....	2
1.2 El fenómeno del ruido en dispositivos semiconductores.....	4
1.2.1 Ruido en dispositivos semiconductores .....	4
1.2.2 Ruido debido a la conmutación de tensiones y corrientes.....	5
1.3 Cuantificación del ruido de conmutación.....	8
1.4 Esquema de la memoria de tesis.....	8
1.5 Convenciones usadas en esta memoria .....	10
Referencias.....	10
<b>Capítulo 2: Principios básicos del ruido de conmutación</b>	<b>13</b>
2.1 Generación del ruido de conmutación .....	13
2.1.1 Circuitos digitales integrados en tecnología CMOS.....	13
2.1.2 Circuitos integrados en tecnología BiCMOS .....	21
2.1.3 Comparación del ruido generado por puertas CMOS y BiCMOS .....	25
2.2 Modelos analíticos del ruido de conmutación para circuitos integrados CMOS.....	27
2.2.1 Modelos lineales del ruido de conmutación simultáneo (SSN) .....	27
2.2.2 Modelos sub-lineales e inclusión de efectos de canal corto.....	27
2.2.3 Parámetros influyentes en el ruido de conmutación .....	29
2.3 Mecanismos de transmisión del ruido de conmutación.....	30
2.3.1 Distribución de la alimentación en circuitos integrados.....	30
2.3.2 Polarización del sustrato .....	33
2.3.3 Comunicación entre dos circuitos integrados .....	34
2.4 Efectos del ruido de conmutación en circuitos digitales .....	35

2.4.1 Errores de retardo.....	37
2.4.2 Falsas conmutaciones.....	39
2.4.3 Almacenamientos erróneos en biestables.....	41
2.4.4 Sensibilidad al ruido de conmutación de los circuitos lógicos BiCMOS.....	41
2.5 Efectos del ruido de conmutación en circuitos analógicos.....	42
2.6 Conclusiones.....	43
Referencias .....	44

## Capítulo 3: Estrategias de diseño digital y ruido de conmutación 49

3.1 Introducción.....	49
3.2 Diseño y síntesis a alto nivel y ruido de conmutación .....	50
3.2.1 Proceso de síntesis a alto nivel .....	51
3.2.2 Ejemplo 1: controlador de un semáforo de tráfico .....	52
3.2.3 Ejemplo 2: contador/descontador binario de 8 bits.....	58
3.2.4 Ejemplo 3: resolución numérica de una ecuación diferencial .....	61
3.3 Arquitectura y ruido de conmutación.....	65
3.3.1 Circuitos combinatoriales y síntesis lógica.....	65
3.3.2 Máquinas de estados finitos.....	69
3.3.3 Unidad de control y <i>Datapath</i> .....	71
3.4 Alternativas lógicas y ruido de conmutación	
3.4.1 Lógicas estáticas.....	73
3.4.1.1 Lógica CMOS convencional.....	73
3.4.1.2 Lógica Pseudo-NMOS .....	73
3.4.1.3 Lógica de transistores de paso ( <i>pass transistor logic</i> ).....	74
3.4.2 Lógicas dinámicas sin estructura <i>pipeline</i> .....	74
3.4.2.1 Lógica Dominó y NP Dominó .....	75
3.4.2.2 Lógica CVSL ( <i>Cascode Voltag Switch Logic</i> ) .....	75
3.4.2.3 Lógica DSL ( <i>Differential Split-Level</i> ).....	75
3.4.3 Lógicas dinámicas con estructura <i>pipeline</i> .....	76
3.4.4 Lógicas diferenciales controladas por el reloj ( <i>clocked CMOS logics</i> ).....	76
3.4.5 Diseño síncrono <i>versus</i> diseño asíncrono .....	77
3.5 Técnicas de reducción del consumo.....	77
3.6 Conclusiones.....	80
Referencias .....	82

## Capítulo 4: Efectos del escalado en el ruido de conmutación para tecnologías CMOS 85

4.1 Introducción.....	85
4.1.1 Evolución de la tecnología de fabricación de circuitos integrados.....	87
4.1.2 Estado actual de la tecnología CMOS y perspectivas futuras .....	89
4.2 Modelos de escalado tecnológico CMOS.....	91
4.2.1 Modelos de escalado clásicos: voltaje constante y campo constante .....	92
4.2.2 Escalado y fiabilidad.....	95
4.2.2.1 Dimensiones físicas .....	96
4.2.2.2 Nivel de dopado del substrato .....	96



4.2.2.3 Tensión de alimentación.....	96
4.2.2.4 Óxido de puerta.....	97
4.2.3 Modelos para escenarios de bajo consumo y altas prestaciones .....	97
4.2.4 Escalado de las interconexiones.....	100
4.3 Efectos de canal corto en dispositivos submicrónicos.....	106
4.3.1 Reducción de la tensión umbral.....	106
4.3.2 Reducción de la movilidad debido al campo eléctrico.....	106
4.3.3 Saturación de la velocidad de los portadores.....	107
4.3.4 Características de conmutación de los dispositivos submicrónicos.....	108
4.4 Escalado y ruido de conmutación.....	114
4.4.1 Escalado del ruido de conmutación para escenarios clásicos .....	115
4.4.2 Escalado en escenarios realistas.....	117
4.4.3 Obtención por simulación HSPICE del ruido de conmutación para escenarios de escalado realistas.....	119
4.4.4 Implicaciones del escalado de las interconexiones y el aumento de la complejidad en el ruido de conmutación .....	123
4.5. Conclusiones .....	126
Referencias.....	128

## **Capítulo 5: Alternativas lógicas no convencionales de bajo nivel de ruido de conmutación** 133

5.1 Introducción.....	133
5.2 Lógicas con consumo de corriente constante .....	134
5.2.1 Lógica ESCL ( <i>Enhancement Source-Coupled Logic</i> ).....	134
5.2.2 Lógica FSCL ( <i>Folded Source-Coupled Logic</i> ).....	135
5.2.3 Otras alternativas lógicas CSL .....	136
5.3 Testabilidad de las lógicas de consumo de corriente constante .....	137
5.3.1 Análisis de la testabilidad de la lógica FSCL.....	137
5.3.2 Descripción del estudio.....	137
5.3.3 Simulaciones de los defectos tipo 'puente' .....	140
5.3.4 Simulaciones de los defectos tipo 'abierto'.....	143
5.3.5 Análisis de los resultados y discusión.....	145
5.4 Técnica de reducción del ruido de conmutación basada en lógicas diferenciales autotemporizadas.....	146
5.4.1 Metodología de diseño.....	147
5.4.2 Ejemplo de diseño: multiplicador de 4×4 bits .....	148
5.4.3 Resultados y discusión.....	151
5.5 Resultados experimentales del multiplicador 4×4 ECDL .....	153
5.5.1 Descripción del prototipo .....	153
5.5.2 Medidas sobre el prototipo .....	157
5.5.2.1 Límitadores de corriente.....	157
5.5.2.2 Multiplicador CMOS .....	159
5.5.2.3 Multiplicador ECDL.....	160
5.5.2.4 Análisis de los resultados .....	165
5.6 Conclusiones .....	166
Referencias.....	167

<b>Capítulo 6: Técnicas de alimentación de CIs CMOS de bajo nivel de ruido de conmutación</b>	<b>171</b>
6.1 Introducción.....	171
6.2 Distribución de la alimentación y condensadores de desacoplo .....	172
6.2.1 Modelo discreto de la distribución de la alimentación .....	173
6.2.2 Condensadores de desacoplo.....	173
6.3 Encapsulado .....	175
6.4 Técnicas de reducción del ruido de conmutación utilizando condensadores dentro del encapsulado .....	176
6.4.1 Modelo eléctrico simplificado de un CI encapsulado con condensador de desacoplo <i>on-chip</i> .....	177
6.4.2 Capacidades de desacoplo <i>on-chip</i> .....	180
6.4.3 Soluciones al problema de la resonancia de los condensadores de desacoplo <i>on-chip</i> .....	182
6.5 Técnica de alimentación con condensador tanque TCMOS .....	182
6.5.1 Diseño de un prototipo que implementa la técnica de alimentación de bajo nivel de ruido $dI/dt$ TCMOS.....	185
6.5.1.1 Especificaciones y parámetros de diseño del circuito TCMOS.....	185
6.5.1.2 Estructura del prototipo fabricado.....	187
6.5.1.3 Configuraciones de medida.....	191
6.5.2 Medidas realizadas sobre los prototipos.....	193
6.5.2.1 Funcionalidad del circuito TCMOS.....	193
6.5.2.2 Medida del ruido de conmutación.....	196
6.5.2.3 Medida del ruido de substrato.....	197
6.5.2.4 Medida de la variación en el retardo de propagación debido al TCMOS.....	198
6.5.2.5 Análisis del consumo .....	199
6.5.3 Discusión de los resultados y trabajo futuro .....	200
6.6 Tendencias futuras en la distribución de la alimentación a circuitos integrados.....	201
6.6.1 Viabilidad de integrar la fuente de alimentación.....	202
6.7 Conclusiones.....	204
Referencias .....	205
<b>Capítulo 7: Circuitos digitales modo corriente y comunicación entre CIs</b>	<b>209</b>
7.1 Introducción.....	209
7.2 Circuitos lógicos multivaluados en modo corriente (MVCML).....	212
7.2.1 Operaciones lógicas y aritméticas multivaluadas.....	213
7.2.2 Aritmética en modo corriente.....	214
7.2.3 Componentes básicos de los circuitos MVCML.....	214
7.2.4 Realización de un sumador total de números con signo (SDFA).....	218
7.2.5 Circuitos complejos multivaluados en modo corriente.....	220
7.3 Prestaciones frente al ruido de conmutación de los circuitos MVCML.....	221
7.3.1 Descripción de un prototipo de CI orientado al análisis y la experimentación.....	221
7.3.2 Generación del ruido de conmutación de los circuitos MVCML.....	223
7.3.3 Sensibilidad a las fluctuaciones estáticas y dinámicas en la tensión de alimentación.....	228

---

7.3.4 Acoplamientos capacitivos entre líneas que transportan señales digitales en modo corriente.....	230
7.4 Comunicación entre CIs en modo corriente .....	232
7.5 Conclusiones .....	237
Referencias.....	238
<b>Capítulo 8: Conclusiones</b>	<b>243</b>
<b>Apéndice A: Esquemáticos y <i>layouts</i> del circuito integrado ICECDL</b>	<b>249</b>
<b>Apéndice B: Esquemáticos y <i>layouts</i> del circuito integrado ICCML</b>	<b>255</b>



# Agradecimientos

---

Esta memoria de tesis resume el trabajo de investigación realizado durante los tres últimos años en el Departamento de Ingeniería Electrónica bajo la dirección del Dr. Antonio Rubio. Quisiera agradecer en primer lugar a Antonio por la forma en que se ha desarrollado el trabajo, no sólo en el ámbito académico e investigador, sino también en el plano personal, y hacer extensivo este agradecimiento al resto de miembros del grupo de investigación de diseño de circuitos VLSI de altas prestaciones, Patxi, Xavi, Diego y Pep, por su apoyo y colaboración durante todo este tiempo. Patxi y Antonio han revisado pacientemente cada capítulo de esta memoria y sus sugerencias han sido de gran utilidad para la redacción final del manuscrito.

Las discusiones con varios investigadores del departamento y de fuera de él han contribuido a profundizar en muchos de los aspectos tratados a lo largo de esta memoria de tesis. Pidiendo disculpas de antemano por si me dejo a alguien, quisiera expresar mi gratitud a Miquel, Jaume, Joan Figueras, Josep Rius, Toni Ferré, Alberto Poveda, Jordi Mádrenas, Juan Manuel Moreno, Jordi Cortadella, Javier Uceda, Robert H. Dennard y varios investigadores de la lista de correo de Integridad de Señales (SI LIST), así como a los revisores anónimos de las publicaciones realizadas que han contribuido a esclarecer algunos puntos poco claros de las propuestas realizadas o los resultados obtenidos. Quisiera agradecer especialmente a Vicente Jiménez su ayuda en todo lo relacionado con el análisis de datos de las simulaciones HSPICE.

Las implementaciones y medidas de alguno de los prototipos han sido realizadas junto con los estudiantes Jordi Bosch y Eloi Ramón, a los que quisiera agradecer su trabajo y dedicación y todo lo que hemos aprendido juntos.

El trabajo en el laboratorio de diseño VLSI y las medidas realizadas hubieran sido imposibles sin el soporte de Santi y Belén. También quisiera agradecer la colaboración de Dolors, Aurora, Elodia, Conxita y M<sup>a</sup>. Cruz y del resto de la administración del departamento en multitud de aspectos.

Durante estos tres años, y espero que durante muchos más, he tenido la suerte de convivir en el departamento con un grupo de personas excepcionales con las que he compartido, no solo el trabajo, sino también la amistad. Quisiera agradecer, además de a Antonio, Patxi, Xavi, Diego y Pep, a Toni, Dani, Vicente, Pau, Eva, Eduard y a Angel, desde la distancia, por la amistad brindada, y también pido disculpas si me dejo a alguien. El equipo de fútbol de los jueves ha contribuido a que pudiera seguir la máxima *mens sana in corpore sano* y quisiera agradecer a todos los compañeros los ratos de diversión vividos que han ayudado a aliviar la tensión que a veces conlleva este trabajo. También quisiera agradecer el apoyo del resto de amigos, a los que he dejado algo abandonados, a Angel, Susana, Nacho, Jordi, Pedro, Feli, Germán, Silvia, y un largo etc., y a la Comunidad Juvenil Tecla Sala por haber contribuido a hacerme mejor persona.

La peor parte de todo, como suele pasar, se la ha llevado mi familia. Quisiera agradecerles su comprensión y su apoyo durante todo este tiempo, teniendo en cuenta los momentos difíciles, que han sido superados. Especialmente agradezco a mi padre la supervisión ortográfica final del texto de la memoria. Quisiera agradecer también la paciencia de Marta, que ha tenido que soportar mis 'abducciones' por el trabajo y que me ha apoyado y animado en todo momento.

La realización de este trabajo ha sido posible en parte gracias a la financiación del proyecto de investigación TIC95/469 de la CICYT del Ministerio de Educación y Ciencia del Gobierno Español.

*José Luis González Jiménez*

*Campus Nord, Barcelona, Noviembre de 1997*

# Capítulo 1

---

## *Introducción*

*En este capítulo presentamos las motivaciones que nos han llevado a realizar el trabajo de investigación sobre el que versa esta memoria de tesis. En primer lugar tratamos de forma breve la evolución y el estado actual de la tecnología de fabricación de circuitos integrados CMOS y BiCMOS. A continuación se introduce el concepto de ruido de conmutación y la influencia de este fenómeno en las prestaciones de los sistemas fabricados utilizando dichas tecnologías. Finalmente se expone el esquema general en que esta estructurada esta memoria y las convenciones utilizadas en su redacción.*

### **1.1 Tecnología de fabricación de circuitos integrados CMOS y BiCMOS**

La implementación de sistemas enteros o parte de los mismos en pastillas semiconductoras de reducidísimas dimensiones ha revolucionado el mundo que nos rodea, sobre todo en los campos de las comunicaciones, la electrónica de consumo y la enseñanza. Las principales causas han sido la facilidad de integrar prestaciones en un área muy reducida y a un coste de fabricación realmente muy bajo. En la base de esta revolución tecnológica está la industria de fabricación de circuitos integrados (abreviados IC en inglés o CI en castellano).

#### **1.1.1 Características específicas de los circuitos integrados**

En las últimas décadas la tecnología de fabricación basada en dispositivos con estructura de Metal Óxido Semiconductor de tipo Complementario (que se abrevia con las siglas inglesas CMOS) ha jugado un papel cada vez más importante en la industria global de fabricación de circuitos integrados. Las principales ventajas respecto a otras alternativas (tecnologías basadas en dispositivos bipolares y tecnologías basadas en arseniuro de galio o compuestos III-IV) son el reducido consumo, la gran facilidad de integración y sus elevados márgenes de ruido [1][2]

Sin embargo los circuitos integrados CMOS tienen una capacidad limitada de *driving*<sup>1</sup>, lo que degrada sus prestaciones, sobre todo en chips de muy alto nivel de integración (VLSI), en los que las salidas deben conmutar grandes capacidades de carga con una gran disipación de potencia. Combinando la estructura de los circuitos CMOS con los bipolares se pretende obtener un compromiso entre las prestaciones de ambos tipos de dispositivos. Cuando se utiliza esta técnica a la tecnología se la denomina **BiCMOS** y el proceso de fabricación debe permitir implementar tanto dispositivos de efecto campo MOSFET como dispositivos bipolares. Éstos últimos ofrecen las siguientes ventajas: mayor corriente de salida, mejor apareamiento entre dispositivos teóricamente iguales, y menor sensibilidad a las variaciones del proceso. Los dispositivos MOSFET aportan como ventajas una menor área, debido a su característica de auto-aislamiento, menores capacidades parásitas, alta resistencia de entrada, menor consumo, mayor *yield* (término inglés que mide el rendimiento del proceso de fabricación), conmutación bidireccional de corriente y capacidad inherente de memorización. La tecnología BiCMOS ha surgido en los últimos años como una alternativa a la tecnología CMOS en algunas aplicaciones donde su mayor coste es compensado por el aumento de prestaciones que introduce.

Tradicionalmente los sistemas electrónicos se han subdividido entre **sistemas digitales**, que trabajan con dos o más valores discretos de señal, y **sistemas analógicos** que trabajan con señales continuas. El procesado digital de la señal ha permitido un gran desarrollo de los sistemas digitales que se han aprovechado de las prestaciones de las tecnologías de integración mediante las cuales es muy sencillo fabricar complejos sistemas electrónicos sobre pastillas semiconductoras. Durante un tiempo los circuitos integrados se clasificaban siguiendo esta misma división entre circuitos integrados analógicos y digitales. En los últimos años ha surgido la posibilidad de integrar sobre el mismo sustrato ambos tipos de sistemas dando lugar a los llamados **circuitos integrados mixtos**.

### 1.1.2 Estructura típica de un circuito integrado

En este punto debemos tener en cuenta otro de los aspectos específicos de la integración de circuitos: **el encapsulado**. Debido a las reducidas dimensiones de las pastillas semiconductoras es necesario encapsularlas en unos contenedores que permitan por un lado adaptar su reducido tamaño a las dimensiones mayores de las cartas de circuito impreso u otros sistemas con los que deban interactuar y por otro lado adaptar y servir de interfaz para las señales eléctricas que entran y salen del circuito integrado desde y hacia el mundo exterior y disipar el calor generado en su interior por los circuitos electrónicos. Debido a los elementos parásitos que introduce en los circuitos electrónicos que contiene y a los que sirve de soporte se ha de tener en cuenta sus efectos en las prestaciones finales de éstos [3].

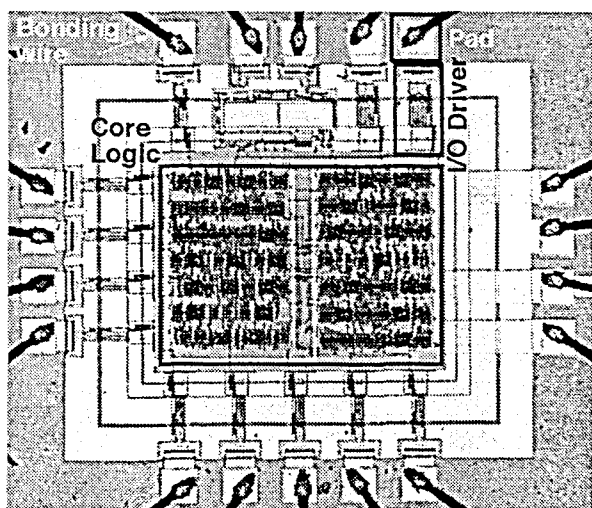
Aunque hoy en día existen multitud de alternativas de encapsulado podemos hablar de un conjunto común de elementos que forman la estructura pastilla semiconductor-encapsulado (Figura 1-1). Por una lado en el dado semiconductor (que en tecnologías CMOS y BiCMOS es

---

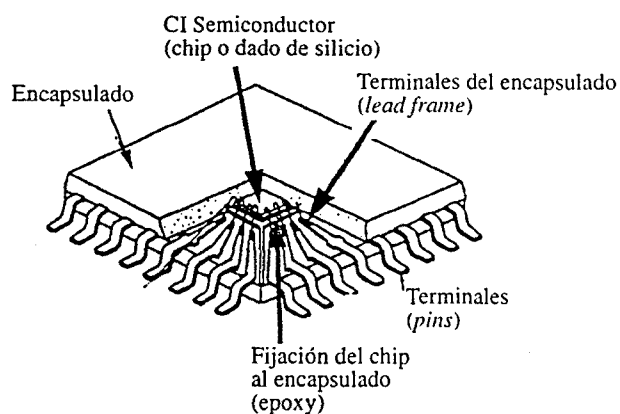
<sup>1</sup> El término proviene del sustantivo inglés *driver*, que se traduciría por 'actuador'. *Driving* es la capacidad o fuerza que tiene una puerta lógica para conmutar su salida.



de Silicio) deben reservarse unas áreas en las capas metálicas que permitan la conexión de las alimentaciones y las entradas y salidas del circuito integrado con el encapsulado. A estos elementos se les denomina *pads*. En el otro extremo, el encapsulado dispone de unos terminales a los que se conecta mediante hilos de oro (*bonding wires*) u otro sistema de conexión ( como por ejemplo las bolas de soldadura, en inglés *solder bumps*). Estos terminales pueden alargarse y ensancharse hasta formar los terminales de soldado del encapsulado (*pins*) formando lo que se conoce por *lead frame* o bien formar un complejo sistema de interconexión integrado en el encapsulado y que suele constar de diferentes capas superpuestas en los que se alternan planos de masa y alimentación y capas dedicadas a la interconexión de señales. Las interconexiones entre el dado de silicio y el mundo exterior se realiza mediante vías que conectan las diferentes capas entre sí, hasta formar un camino desde el *pad* del circuito integrado hasta el correspondiente *pin* del encapsulado. En el caso de las alimentaciones esta correspondencia unívoca no es tan sencilla. Si el encapsulado dispone de planos de masa y alimentación, se realizan múltiples conexiones de los *pads* a estos planos, y de forma independiente se conectan estos planos a múltiples *pins* del encapsulado.



(a)



(b)

### Figura 1-1

(a) Fotografía de una pastilla de silicio que contiene un circuito integrado digital. En el contorno del circuito puede observarse el anillo de pads y los hilos que conectan el circuito con el encapsulado.  
 (b) Dibujo representando un corte en un encapsulado de tipo MQFP o PQFP (siglas inglesas de Metric Quad Flat Pack y Plastic Quad Flat Pack, respectivamente). En él se indican las partes más importantes de un encapsulado del tipo 'lead frame', al que pertenece el aquí representado.

Finalmente hay que considerar el resto del circuito eléctrico formado por las pistas que conectan los diferentes chips entre si en la placa de circuito impreso y el sistema que distribuye la alimentación desde el rectificador hasta los terminales de cada integrado. Todos estos elementos no pueden ignorarse en el análisis de cualquier fenómeno relacionado con las prestaciones de los circuitos integrados.

## 1.2 El fenómeno del ruido en dispositivos semiconductores

Debido a las exigencias del mercado, la competitividad ha impulsado a la industria de fabricación de circuitos integrados a evolucionar a un ritmo vertiginoso [4][5]. Esto ha sido aún más patente en la tecnología CMOS que es con diferencia una de las más importantes alternativas hoy en día. El resultado ha sido una continua reducción de las dimensiones físicas de los dispositivos lo que ha permitido incrementar más y más el nivel de integración (el número de transistores que es posible integrar por unidad de área). Hablaremos de esta evolución y del estado actual y las previsiones de escalado futuras en un capítulo posterior. Esta tendencia a miniaturizar los dispositivos sumado a la tendencia a implementar sistemas cada vez más complejos en una misma pastilla semiconductor o la aparición de nuevas estructuras como los módulos multichip (referidos bajo las siglas MCM en inglés) ha provocado que determinados efectos de segundo orden tengan que ser tenidos en cuenta pues han llegado a ser las causas principales que podrían frenar la evolución de las tecnologías debido a las limitaciones en las prestaciones de los circuitos. Uno de los fenómenos más importantes que degrada el funcionamiento de los sistemas electrónicos es el ruido, es decir, la aparición de señales eléctricas espúreas no deseadas en los nodos y ramas del circuito. En el terreno de los circuitos integrados podemos diferenciar entre dos clases de ruido.

### 1.2.1 Ruido en dispositivos semiconductores

En primer lugar y partiendo de los dispositivos que componen los circuitos integrados aparecen una serie de fenómenos físicos que dan lugar a distintos tipos de ruido. Citaremos en este apartado los más importantes y remitimos al lector a otros textos si desea una información más detallada. Por ejemplo [6] es una referencia clásica sobre el ruido en sistemas electrónicos en general; [7] es un texto clásico sobre semiconductores en el que se abordan también los aspectos relacionados con el ruido y, por último, en el segundo capítulo de [8] se citan los tipos más comunes de ruido que aparecen en circuitos integrados semiconductores. De entre ellos destacamos los siguientes:

- El **ruido térmico** se debe a las variaciones producidas por la oscilación térmica de los portadores de electricidad en la conductividad de los materiales semiconductores o conductores. El valor cuadrático medio de este tipo de ruido se expresa como:

$$\langle v_n^2 \rangle = 4KTBR \quad (1-1)$$

dónde  $K$  es la constante de Boltzman,  $T$  la temperatura,  $B$  el ancho de banda y  $R$  la parte real del elemento semiconductor considerado. Este ruido tiene un espectro plano y por ello se conoce también como 'ruido blanco'.

- El **ruido impulsivo** (*shot noise* en inglés) es un ruido típico de uniones semiconductoras y se debe a la generación aleatoria de portadores que son capaces de atravesar una barrera de potencial (este tipo de ruido apareció por primera vez en las válvulas de vacío).

Este tipo de procesos aleatorios de generación provocan variaciones en el valor medio de la corriente que atraviesa la unión y su varianza puede representarse por:

$$\langle i_n^2 \rangle = 2qBI \quad (1-2)$$

donde  $q$  es la carga del electrón,  $B$  el ancho de banda e  $I$  el valor medio de la corriente que atraviesa la unión semiconductor.

- El ruido  $1/f$  o *flicker noise*, en inglés. Existe todavía todavía cierta controversia sobre su origen exacto. Está relacionado con la estructura de bandas de los semiconductores y los procesos de recombinación de portadores. Su característica más importante es la dependencia inversamente proporcional a la frecuencia, lo que hace que éste tipo de ruido sea muy importante para bajas frecuencias.

De estos tres tipos de ruido el más importante, sobre todo en lo reportado para dispositivos bipolares [7], es el ruido impulsivo. Ambos tipos de ruido son considerados en el diseño de circuitos integrados analógicos únicamente cuando se trata de trabajar con señales pequeñas, pues el nivel absoluto de todas estas clases de ruido debido a fenómenos físicos en el interior de los semiconductores está varios órdenes de magnitud por debajo de otros tipos de ruido [8] como los expuestos en el apartado siguiente.

### 1.2.2 Ruido debido a la conmutación de tensiones y corrientes

La conmutación de tensiones y corrientes es claramente la mayor fuente de ruido en los circuitos integrados actuales. La solución más simple para minimizar este tipo de ruidos, pero no siempre la más práctica, consiste pues en reducir tanto los niveles como la frecuencia de estas conmutaciones. Los circuitos lógicos CMOS, a pesar de ser la solución más popular, son de entre todas las alternativas para implementar sistemas lógicos (circuitos modo corriente, circuitos ECL balanceados y desbalanceados, circuitos TTL, ...) los que maximizan las excursiones (*swings*) de los niveles de corriente y tensión, y además concentran estas conmutaciones en intervalos muy pequeños de tiempo. En estas rápidas transiciones también puede producirse paso de corriente entre los dos terminales de alimentación a través de los circuitos lógicos en un fenómeno conocido como corriente de conmutación de cortocircuito. Como resultado de estas características específicas, los circuitos integrados CMOS, y su extensión BiCMOS, son la alternativa tecnológica para la integración de sistemas electrónicos más ruidosa. Es por ello que nuestro trabajo de investigación se ha centrado en analizar y proponer técnicas que permitan reducir este tipo de ruido en circuitos integrados CMOS y BiCMOS.

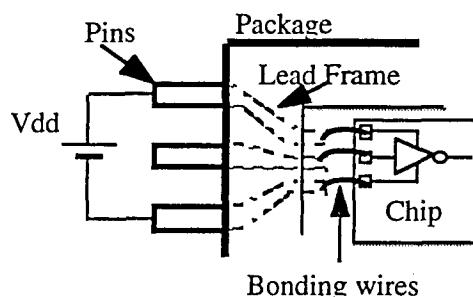
El ruido debido a conmutaciones de tensión o de corriente es un fenómeno de segundo orden que se da en los circuitos integrados puramente digitales. Es también especialmente importante en los circuitos integrados mixtos, debido a que puede transmitirse fácilmente a la parte analógica del integrado que es mucho más sensible a los fenómenos ruidosos.

Este tipo de ruido puede a su vez clasificarse en función del mecanismo de generación de la perturbación. Genéricamente se habla de **ruido  $dV/dt$**  para referirse al producido por conmutaciones en las tensiones de los nodos. El fenómeno más estudiado de este tipo es el *crosstalk* — que ya ha sido objeto de otros trabajos de investigación en nuestro grupo [9][10] — y del que no vamos a tratar más que algunos aspectos puntuales en capítulos posteriores.

El otro fenómeno de este tipo es el **ruido  $dI/dt$** , sobre el cual se centra el trabajo de investigación recogido en esta memoria de tesis. De ahora en adelante nos referiremos a éste como **ruido de conmutación**, tomando la traducción inglesa del término *switching noise*, y cuando nos queramos referir al otro tipo de ruido debido a las conmutaciones de tensión, usaremos explícitamente el término ' $dV/dt$ '. Cada vez que se realiza una conmutación en un circuito integrado digital el sistema de distribución de la alimentación debe proporcionar un pulso de corriente que el circuito que conmuta utiliza para cargar o descargar las capacidades presentes en sus nodos de salida. En este punto interviene uno de los elementos parásitos que introducen los encapsulados en los circuitos integrados: la inductancia de los terminales por los que se proporciona la alimentación al circuito electrónico (ver Figura 1-2). Según la teoría clásica la caída de tensión en bornes de un inductor responde a la siguiente expresión

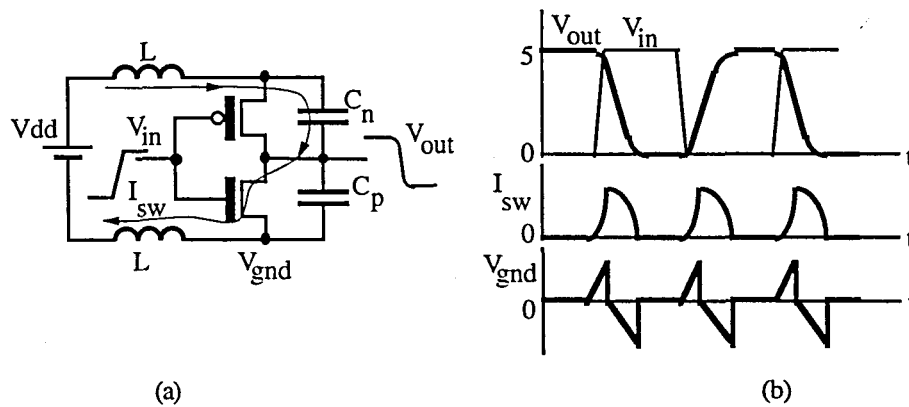
$$V(t) = L \cdot \frac{dI(t)}{dt} \quad (1-3)$$

dónde  $L$  es el valor de la inductancia serie parásita que introduce el encapsulado y  $V(t)$  e  $I(t)$  la corriente y la tensión, respectivamente, a través de la bobina. En la Figura 1-3 se muestra una representación esquemática de este fenómeno. Cada vez que se produce una conmutación en el circuito digital se genera un pulso de ruido en los terminales de alimentación del circuito integrado, cuya amplitud depende del valor de la inductancia parásita y de la forma de onda de la corriente de conmutación. Más adelante analizaremos en detalle cada uno de estos parámetros y su relación con las prestaciones del circuito. El ruido así generado puede afectar al resto de circuitos digitales (los que están quietos, es decir, no conmutan) y, en el caso de que se trate de un circuito integrado mixto, propagarse hasta la parte analógica y influir en su comportamiento. Este ruido puede también aparecer en las salidas finales del circuito integrado por lo que podría afectar a otros circuitos electrónicos conectados a éste.



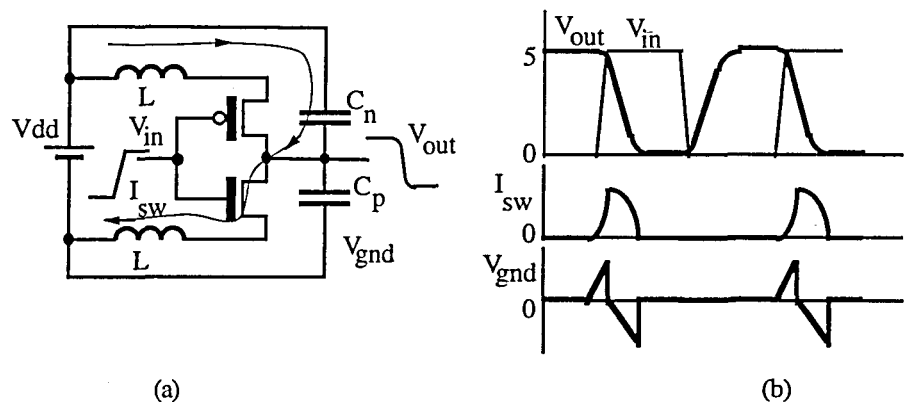
**Figura 1-2.**

*Representación esquemática de la estructura de un encapsulado de un circuito integrado.*

**Figura 1-3**

(a) Circuito digital mientras realiza una conmutación. La corriente de conmutación  $I_{sw}$  al circular a través de las inductancias parásitas  $L$  provoca pulsos de ruido de conmutación en ambos nodos de alimentación internos. En (b) se representa esquemáticamente las formas de onda para la entrada y la salida del circuito de la izquierda y la corriente de conmutación y la tensión en el nodo de tierra interno ( $V_{gnd}$ ).

Si las capacidades conmutadas se encuentran en el exterior del encapsulado, como es el caso de los *drivers* de salida de los CIs, el lazo de corriente se ha de cerrar por el exterior y solo tenemos ruido en uno de los dos terminales internos, o bien el de masa o bien el de alimentación, según sea el sentido de conmutación de la salida (ver Figura 1-4). Las capacidades externas (*off-chip*) son mucho mayores que las internas y por ello los *drivers* de salida son elementos del circuito con consumos dinámicos de corriente muy elevado. Sumando esto a que suelen estar agrupados en *buses* (grupos de señales binarias que representan conjuntamente una misma información) de anchura considerable (de 8 a 128 bits), y que todos las salidas de estos *buses* pueden realizar una conmutación simultánea en el mismo sentido, estos elementos del CI son los responsables de una gran cantidad de ruido de conmutación y por ello en la mayoría de aplicaciones se dispone de terminales de alimentación dedicados para ellos.

**Figura 1-4**

(a) Esquema de conmutación cuando las capacidades son externas (*off-chip*) como es el caso de los *drivers* de salida. (b) Formas de onda esquemáticas de la corriente y el ruido de conmutación en el caso del circuito anterior.

### 1.3 Cuantificación del ruido de conmutación

El ruido de conmutación  $dI/dt$  consiste en pulsos de tensión que aparecen en los terminales de alimentación del sistema. Las características a considerar de estos pulsos son: su sentido (positivo o negativo sobre la tensión nominal sobre la que se generan), su amplitud y su anchura. Estos tres parámetros tienen influencia en el tipo de efecto que produce el ruido de conmutación en los circuitos a los que afecta. La mayoría de autores cuando se refieren a los efectos perjudiciales del ruido de conmutación citan únicamente el nivel de pico de dicho ruido [11][12][13][14][15]. Nosotros adoptaremos el mismo criterio. Cuando el ruido se genera debido a conmutaciones internas normalmente los pulsos son simétricos en Vdd y Gnd, y una medida indicativa del ruido consiste en sumar la amplitud de ambos picos, es decir, la reducción del margen dinámico de la alimentación que se produce debido al ruido.

Existen otros criterios menos cuantitativos que permiten comparar diferentes alternativas respecto del ruido de conmutación que generan. No solo es importante que durante un instante se genere una gran pulso de ruido de conmutación sino también cada cuanto aparecen esos pulsos. En un sistema secuencial los pulsos de ruido de conmutación surgen en cada flanco de reloj, y dependiendo de la actividad del circuito en ese ciclo su amplitud será menor o menor. Pero puede haber sistemas que produzcan un solo pico de ruido en un instante dado y luego generen muy poco ruido de conmutación durante varios ciclos de reloj (presentaremos algunos casos de este tipo de circuitos en este trabajo de tesis). ¿Cual de los dos circuitos es más ruidoso? Evidentemente, todo depende de la concepción global del sistema. Si este puede sincronizarse de alguna manera con los pulsos de ruido, o estos aparecen en instantes conocidos, el circuito sensible puede habilitarse únicamente durante los intervalos ‘tranquilos’. Según este punto de vista es mejor la segunda alternativa que la primera.

La principal causa del ruido de conmutación es la imposibilidad de la alimentación de responder rápidamente a las demandas de consumo debido al carácter inductivo de los conductores que transportan la energía hacia el circuito integrado. Los pulsos de corriente son proporcionados a costa de un momentáneo transitorio en la tensión de la fuente. Otros elementos parásitos que son excitados por este transitorio contribuyen con sus frecuencias de resonancia a prolongar el transitorio con sobre niveles de tensión (*overshots*) y oscilaciones de la tensión de alimentación. La forma de proporcionar la energía al sistema tiene por tanto una influencia determinante en el ruido de conmutación. Una alternativa muy utilizada se basa en que elementos de almacenamiento (condensadores de desacoplo) colocados cerca de los circuitos que demandan los pulsos de corriente son los que proporcionan inicialmente la carga. Estos condensadores son devueltos después de una forma más lenta, y por lo tanto menos ‘ruidosa’ a su tensión original por la fuente de alimentación. Estas y otras ideas serán también aplicadas en soluciones originales propuestas a lo largo de este trabajo de tesis.

### 1.4 Esquema de la memoria de tesis

El primer objetivo de nuestro trabajo de investigación ha consistido en analizar las causas del ruido de conmutación, sobre todo en circuitos digitales CMOS. En segundo lugar se han

estudiado diferentes técnicas que podrían usarse en los diferentes niveles y/o fases del diseño de un sistema electrónico digital o mixto para reducir o eliminar los efectos nocivos de dicho ruido. Para ello se ha investigado el mecanismo por el que se genera el ruido de conmutación y cómo se transmite éste hasta nodos sensibles del propio circuito o de otros circuitos conectados a través de terminales comunes (como las alimentaciones o las entradas/salidas). Se ha investigado en diferentes tipos de circuitos digitales los módulos más susceptibles de generar ruido de conmutación. A partir de aquí se ha estudiado el estado del arte de las técnicas orientadas a reducir el ruido de conmutación y se han propuesto nuevas alternativas propias.

La presente memoria de tesis se ha estructurado siguiendo la consecución de estos objetivos. En el presente capítulo se hace un planteamiento breve del problema del ruido de conmutación y se enmarca dentro del ámbito tecnológico de la fabricación de CI CMOS y BiCMOS. En el segundo capítulo se analizan los principios básicos del ruido de conmutación: su generación, en circuitos CMOS y BiCMOS, su transmisión y los efectos que produce en las prestaciones de circuitos digitales y analógicos. En el mismo capítulo se presentan los diferentes modelos analíticos aparecidos en la literatura y se analizan sus principales parámetros de ajuste. En el tercer capítulo se analizan las diferentes técnicas utilizadas en el diseño y la síntesis a alto nivel de circuitos lógicos y se estudia su repercusión en el ruido generado por las implementaciones finales de los sistemas así diseñados. Se analizan también las estrategias de síntesis lógica y de sistemas secuenciales, sugiriéndose alternativas que permitan contemplar la minimización del ruido de conmutación como objetivo de optimización. También se estudian comparativamente las diferentes alternativas lógicas existentes para relizar las funciones digitales y se analiza el ruido de conmutación que generan. La primera parte que se centra en el análisis de este fenómeno se cierra con el cuarto capítulo en el que se profundiza en las tendencias actuales y las previsiones de escalado de CI CMOS y su repercusión en el ruido de conmutación.

La segunda parte del trabajo trata sobre las técnicas que permiten reducir el ruido de conmutación en circuitos integrados a diferentes niveles de diseño. En el quinto capítulo se analizarán técnicas de diseño de los circuitos digitales orientadas a generar bajos niveles de ruido de conmutación. Entre ellas las lógicas diferenciales del tipo FSCL, reportadas en la literatura, y otra de nuestras propuestas que se basa en el uso de estructuras regulares y lógicas diferenciales del tipo ECDL de la que se presentan resultados experimentales sobre prototipos fabricados. En el capítulo sexto se analizan los problemas de distribución de la alimentación y del encapsulado y los mecanismos que permiten minimizar el ruido de conmutación. Alguna de estas técnicas se ilustran con medidas realizadas sobre prototipos propios. Dentro de estas técnicas se presenta una aportación original en el campo de la alimentación de CI CMOS llamada TCMOS de la que se presentarán los resultados obtenidos por un prototipo fabricado con una tecnología comercial. En el séptimo capítulo se estudian y recopilan las diferentes técnicas aparecidas en la literatura para reducir el ruido de conmutación en los *drivers* de salida, uno de los elementos más ruidosos de los circuitos digitales. En el capítulo séptimo se presentan como alternativa a los circuitos digitales modo corriente que habitualmente integran los sistemas lógicos los circuitos lógicos multivaluados modo corriente, comparando sus prestaciones y su comportamiento frente a la generación de ruido de conmutación. En el mismo capítulo se presenta otra de las

propuestas de este trabajo de investigación consistente en el uso de *drivers* de entrada y salida en modo corriente. Finalmente, el capítulo octavo contiene las conclusiones y las líneas de investigación futuras que se proponen a partir del trabajo desarrollado.

## 1.5 Convenciones usadas en esta memoria

El idioma de esta memoria es el castellano, si bien dado que la totalidad de la literatura concerniente al tema objeto del trabajo se encuentra en inglés, se ha optado por dejar en el idioma original aquellos términos cuya traducción al castellano podría llevar a confusiones, términos que por lo general son sobradamente conocidos en el entorno académico en el que se inscribe la presente memoria de tesis. Estos términos han sido escritos en *cursiva*. Para otros términos anglosajones del argot tecnológico del ámbito de la ingeniería electrónica y las comunicaciones en que se ha considerado oportuno su transcripción al castellano se ha utilizado el manual “*Léxico básico de Telecomunicaciones*” publicado por el Servei de Llengües i Terminologia de la Universitat Politècnica de Catalunya [16] y diversos diccionarios de Inglés técnico.

En la presente memoria de tesis se hace un uso extensivo de las siglas y acrónimos. La mayoría corresponden a las iniciales de los términos en inglés y se han respetado debido al general uso de esta terminología en los ámbitos científicos y académicos. Por término general cada acrónimo se encuentra relacionado con la traducción del término inglés o castellano original al menos una vez en el texto. Las excepciones corresponden a abreviaturas comúnmente conocidas.

## Referencias

- [1] M.I. Elmasry (ed), *Digital MOS Integrated Circuits II*. New York: IEEE press, 1992.
- [2] J.P. Uyemura, *Circuit Design for CMOS VLSI*. Norwell, Massachusetts: Kluwer Academic Publishers, 1992
- [3] H.G. Bakoglu. *Circuits, Interconnections and Packaging for VLSI*. 1990. Massachusetts: Addison-Wesley Publishing Company.
- [4] *IEEE Spectrum : Technology 1997: Analysis and Forecast Issue*, January 1997.
- [5] Robert R. Schaller, ‘Moore’s Law: Past, Present and Future’, *IEEE Spectrum*, June 1997, pp. 53–59.
- [6] Aldert van der Ziel, *Noise: Sources, Characterization, Measurement*. Prentice-Hall, 1970.
- [7] S. M. Sze, *Semiconductors Devices, Physics and Technology*, New York: Wiley, 1985
- [8] Nishath K. Verghese, Timity J. Schmerbeck, David J. Allstot, *Simulation Techniques and Solutions for Mied-Signal Coupling in Integrated Cicuits*. Massachusetts: Kluwer Academic Publishers, 1995
- [9] Miquel J. Roca Adrover, *Estudio de Fallos Binodales en Circuitos Microelectrónicos*. Memoria para la obtención del grado de Doctor en Ciencias Físicas, Palma de Mallorca: Universitat de les Illes Balears, 1995



- [10] Francesc Moll Echeto, *Parasitic Effects Due to Interconnections in Microelectronic Design*. Dissertation for the obtention of the Ph. D. degree, Barcelona: Universitat Politècnica de Catalunya, 1996
- [11] W.D. Becker, B.D. McCredie, B. Shingh, and P. Lin, 'Performance Effects of Switching Noise on CMOS Microprocessors', *Proc. 4th Topical Meeting on Electrical Performance of the Electronic Packages*, 1995, pp. 24–26.
- [12] Keiko Makie-Fukeda, Takafumi Kikuchi, Tatsuji Matsuura, and Masao Hotta, 'Measurement of Digital Noise in Mixed-Signal Integrated Circuits', *IEEE J. Solid-State Circuits*, Vol. 30, No. 2, February 1995, pp. 87–92.
- [13] Bradley D. McCredie, and Wiren D. Becker, 'Modeling, Measurement, and Simulation of Simultaneous Switching Noise', *IEEE Tr. on Components, Packaging, and Manufacturing Techn.—Part B*, Vol. 19, No. 3, August 1996, pp. 461–472.
- [14] Johan Catrysse, 'Measured Distortion of the Output-Waveform of an Integrated OPAMP Due to Substrate Noise', *IEEE Tr. on Electromagnetic Compatiility*, Vol. 37, No. 2, May 1995, pp. 310–312.
- [15] Tallis Blalack, Bruce A. Wooley, 'The Effects of Switching Noise on an Oversampling A/D Converter', *Proc. of IEEE Int. Solid-State Circuits Conference*, 1995, pp. 200–201,367.
- [16] Servei de Llengües i Terminologia, *Lèxics Bàsics: Telecomunicacions*, Universitat Politècnica de Catalunya (UPC), 1996



# Capítulo 2

---

## *Principios básicos del ruido de conmutación*

*En este segundo capítulo se presentan los principios básicos del ruido de conmutación: su generación, su transmisión y el efecto que produce en circuitos digitales y analógicos. Se ven los diferentes modelos analíticos aparecidos en la literatura y se analizan los parámetros de ajuste y su relación con otros parámetros eléctricos de los circuitos integrados digitales.*

### 2.1 Generación del ruido de conmutación

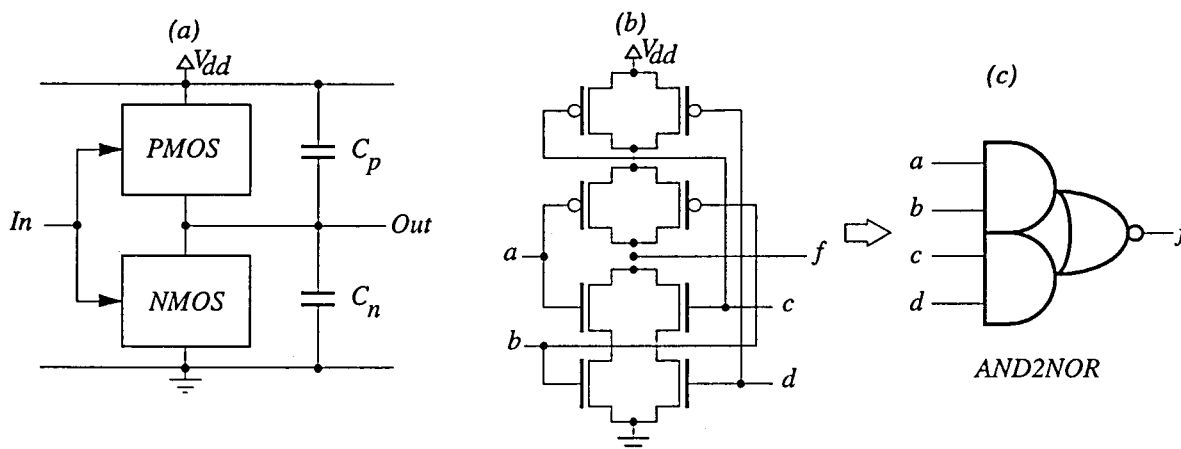
Como hemos visto en el capítulo anterior el **ruido de conmutación** (en inglés *switching noise*) es también conocido por **ruido  $dI/dt$**  (en inglés *delta-I noise* o  $\Delta I$  noise) haciendo referencia a su generación: la demanda de pulsos de corriente por parte de los circuitos digitales al conmutar, que al atravesar las inductancias de los conductores del sistema de distribución de la alimentación, se transforman en pulsos de ruido de tensión. El ruido de conmutación  $dI/dt$  es proporcional al valor de la inductancia multiplicado por la derivada de la forma de onda de la corriente de alimentación. Veremos en primer lugar el mecanismo de generación de este ruido para los circuitos fabricados con tecnología CMOS y en segundo lugar el ruido en los correspondientes circuitos BiCMOS.

#### 2.1.1 Circuitos digitales integrados en tecnología CMOS

Puede darse la confusión entre **circuito lógico CMOS** y **tecnología digital CMOS**. Para evitar malentendidos de ahora en adelante nos referiremos a un circuito lógico CMOS como aquel compuesto por puertas lógicas formadas por dos bloques complementarios PMOS y NMOS respectivamente. El término tecnología digital CMOS es mucho más amplio y abarca desde los circuitos lógicos CMOS hasta otro tipo de estructuras que contienen ambos tipos de transistores (PMOS y NMOS) como las lógicas dinámicas, las basadas en puertas de

transmisión o cualquier topología — además de la complementaria — que permita realizar funciones lógicas<sup>1</sup>.

La estructura básica de un circuito lógico CMOS estático se ilustra en la Figura 2-1. Las puertas basadas en esta tecnología consta de dos bloques complementarios: el superior formado por una red de transistores de tipo PMOS y que conectan el nodo de salida con el terminal positivo de la alimentación y el inferior formado por una red de transistores NMOS que conecta el nodo de salida con el terminal negativo de la alimentación. Ambos bloques funcionan de forma complementaria: cuando uno de ellos establece un camino de baja impedancia entre su terminal de alimentación y la salida el otro no, y viceversa. De esta forma la salida puede tomar el valor lógico '0' (*L* o *Low Level*) o '1' (*H* o *High Level*) dependiendo de la función lógica de la puerta y del valor de las entradas que harán que uno u otro de los bloques complementarios conduzca.



**Figura 2-1**

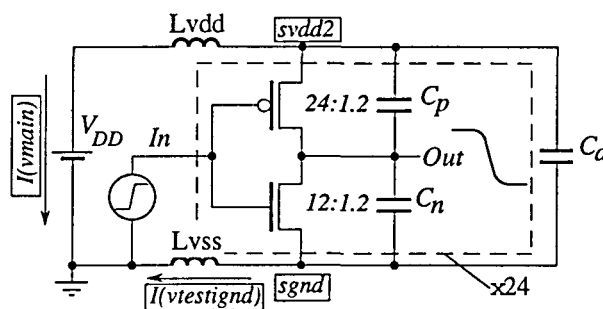
(a) Representación esquemática de un circuito lógico CMOS. (b) Ejemplo de circuito lógico CMOS: puerta AND2NOR y (c) símbolo lógico de esta puerta.

En los circuitos reales es necesario incluir en el circuito eléctrico correspondiente a cada puerta lógica las capacidades parásitas de entrada y salida además de los transistores. La capacidad de entrada se debe al solapamiento entre la región de puerta y las regiones de drenador y surtidor adyacentes al canal. Estas capacidades son pequeñas comparadas con la capacidad de puerta, que es la principal componente de la capacidad total de entrada de un MOSFET. La capacidad parásita de salida está formada por las uniones de las áreas de drenador y surtidor con el sustrato. Ambas capacidades parásitas tienen una componente proporcional al área de las regiones correspondientes a cada caso y otra componente debida al condensador que se forma en los laterales de las islas de difusión del drenador y el surtidor proporcional al perímetro. Todas estas capacidades son debidas a las zonas de depleción de las uniones puerta–drenador y puerta–surtidor y varían con las tensiones aplicadas a los tres terminales. Es necesario incluir también en la salida las capacidades parásitas debidas a las interconexiones formadas entre cada

<sup>1</sup> Para referirnos a circuitos integrados analógicos implementados utilizando tecnologías CMOS, es decir, que permiten fabricar transistores PMOS y NMOS utilizaremos explícitamente el término **tecnologías analógicas CMOS** o análogamente **circuitos analógicos CMOS**.

línea de metal y el substrato, que debe estar conectado a masa (Gnd) o a  $V_{DD}$ , en el caso del pozo tipo n, y también por las capacidades entre dos líneas debida a su proximidad o solapamiento. Considerar todas estas capacidades individualmente complicaría mucho el análisis que pretendemos llevar a cabo en este subapartado. Habitualmente se recurre, sin excesivo menoscabo de exactitud, a representar todas estas capacidades por sendos condensadores conectados entre el nodo de salida de la puerta lógica y los terminales de alimentación y masa del circuito ( $C_p$  y  $C_n$ , respectivamente, en la Figura 2-1).

El funcionamiento básico de cualquier puerta CMOS es el mismo: para unas determinadas combinaciones de las entradas el nodo de salida proporciona un nivel bajo (teóricamente 0 voltios) y para el resto un nivel alto (teóricamente  $V_{DD}$  voltios). Cuando las entradas cambian, de forma que la función lógica ha de hacerlo también, se produce una conmutación en el terminal de salida de nivel alto a bajo o viceversa. Esta conmutación supone siempre la carga o descarga de las capacidades de salida y la corriente necesaria para realizar esta operación debe circular por los terminales de alimentación produciéndose un pulso transitorio de corriente la duración y amplitud del cual determinará, como ya se vió en el capítulo introductorio, la cantidad de ruido de conmutación  $dI/dt$  generado. En la Figura 2-3 se muestra una simulación en la que se puede observar la forma de onda de la corriente de conmutación para la transición alto-bajo a la salida de un inversor CMOS (Figura 2-2). Podemos ver cómo la corriente circula por los dos terminales de alimentación y por tanto el pulso de ruido aparecerá en ambos, pero con polaridad inversa, negativo en el terminal interno de  $V_{DD}$  y positivo en el terminal interno de  $Gnd$ . El resultado es una disminución del margen dinámico de la alimentación, que tendrá sus efectos en otros circuitos “quietos” (es decir, que no conmuten su salida en ese instante) y en el propio circuito que conmuta, como se verá cuando se traten los modelos analíticos de ruido de conmutación.

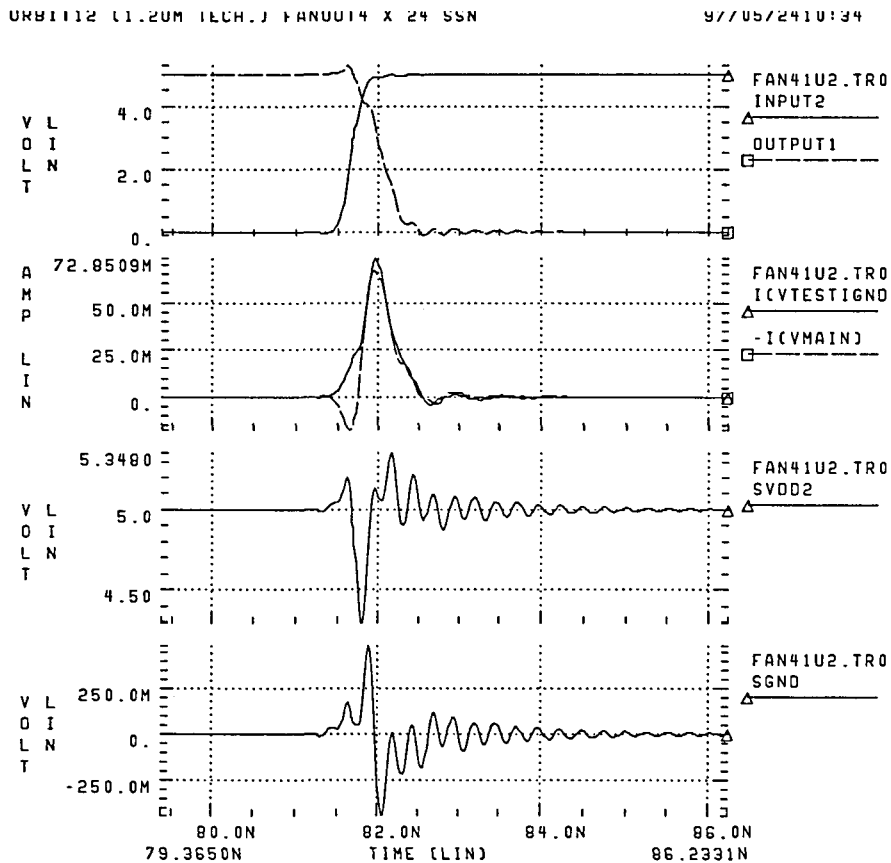


**Figura 2-2**

*Inversor CMOS con un fanout de 4 utilizado para simular la corriente de conmutación y el ruido generado en las bobinas que representan el efecto del encapsulado.  $L_{vss}=L_{vdd}=1nH$ . La capacidad de desacoplo on-chip  $C_d=1pF$ . El circuito consta de 24 inversores conectados a los mismos nodos de alimentación internos ( $svdd2$  y  $sgnd$ ) conmutando simultáneamente.*

La forma de los pulsos de corriente puede ser distinta en la transición alto-bajo que en la bajo-alto, debido a que la corriente circula por transistores distintos y carga o descarga capacidades distintas en ambos casos. Sin embargo si la puerta lógica se diseña para conseguir tiempos de conmutación simétricos, cabe esperar que las dos transiciones demanden pulsos de corriente muy similares. Otro aspecto que influye en la forma del pulso de corriente para puertas

lógicas con más de una entrada es la combinación de entradas que determina la transición. Tomemos como ejemplo la puerta NAND. La transición 0→1 puede producirse para las transiciones a la entrada 11→10, 11→01 o 11→00, mientras que la transición a la salida de 1→0 se da para las transiciones de entrada 00→11, 10→11 o 01→11. Vemos como en este segundo caso la descarga de la capacidad  $C_p$  se realiza en todos los casos por los dos transistores NMOS en serie (Figura 2-4). En el primer caso, la carga de la capacidad  $C_n$  se realiza en las dos primeras transiciones de entrada (11→10 y 11→01) por un único transistor PMOS, mientras que para el caso 11→00 intervienen los dos transistores PMOS en paralelo. Como la capacidad implicada es la misma en ambos casos, la forma de onda de la corriente depende de la transición a la entrada, como podemos ver en la Figura 2-5 y en la Figura 2-6.

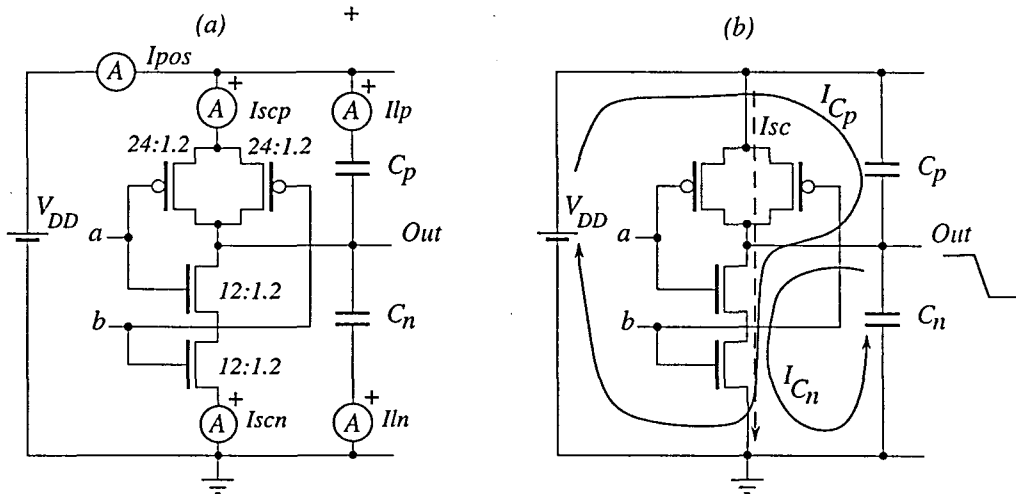


**Figura 2-3**

*Simulación de las corrientes de conmutación y el ruido generado en los nodos internos de alimentación (sgnd y svdd2) para 24 inversores implementados con una tecnología de 1 $\mu$ m CMOS conmutando simultáneamente. Los nombres de las formas de onda se corresponden con las etiquetas de la Figura 2-2.*

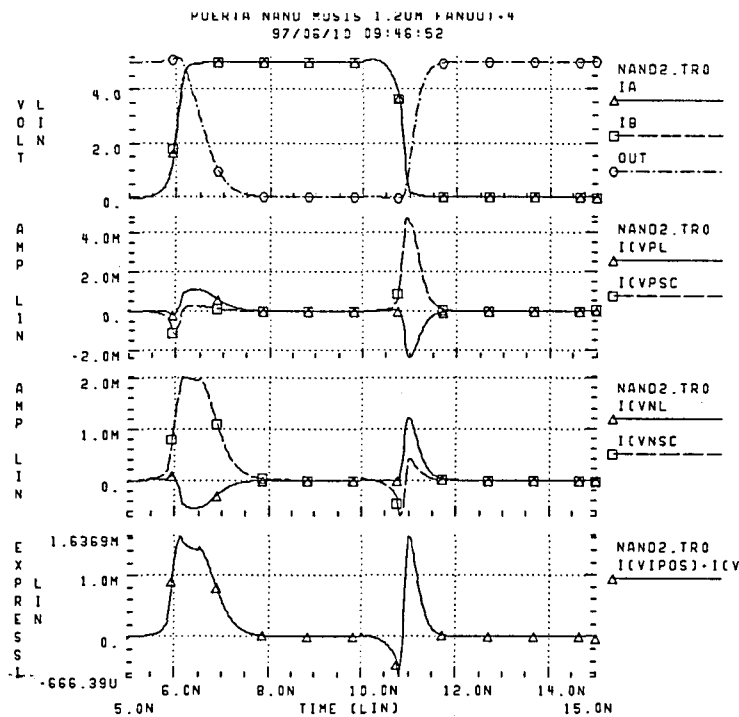
En la Figura 2-3 podemos observar varios aspectos que son importantes a la hora de analizar el ruido de conmutación  $dI/dt$ . En primer lugar, vemos como la corriente que circula por la bobina  $L_{vss}$  es ligeramente distinta a la que circula por  $L_{vdd}$ . Esto es debido a la corriente necesaria para cargar/descargar las capacidades de puerta de los inversores por el generador, que en esta simulación está referido a la masa externa del chip. Por ello existe un lazo de corriente generador→puertas transistores MOS→ $L_{vss}$  que no es simétrico en la bobina  $L_{vdd}$  (para la transición bajo-alto a la salida encontraremos el caso dual). Como veremos más adelante esto

también sucede con los *drivers* de salida, dado que hay un lazo externo en el que solo interviene uno de los dos terminales de alimentación. En estos casos el pulso de ruido en los nodos internos de alimentación y masa son distintos. Sin embargo, si todas las conmutaciones de entrada (generadores) y salida son internas, el lazo de corriente comprende las dos inductancias.



**Figura 2-4**

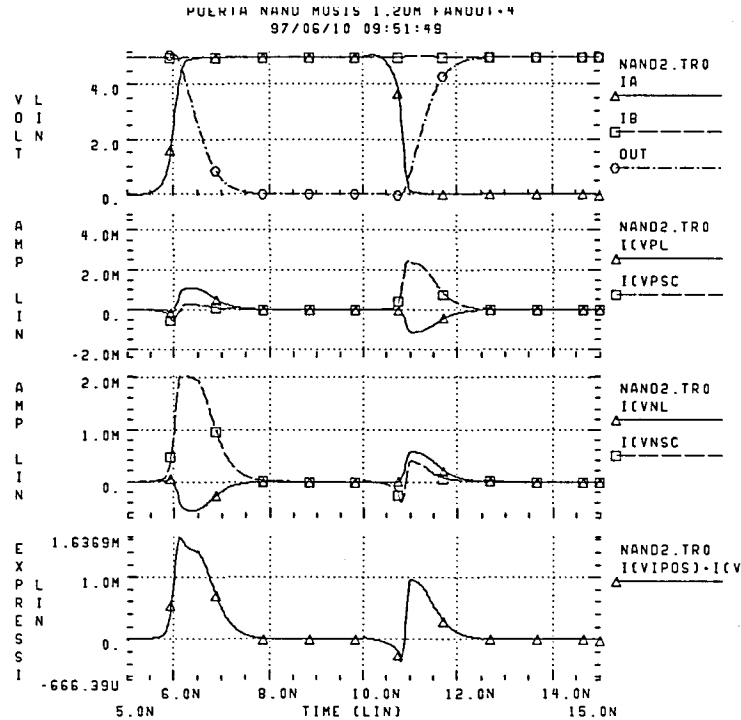
(a) Circuito eléctrico de una puerta NAND de dos entradas con un fanout de 4 ( $C_p$  y  $C_n$ ) para una tecnología de  $1.2\mu\text{m}$  (las dimensiones de los transistores están en  $\mu\text{m}$ ). (b) Esquema de circulación de las corrientes para una transición alto-bajo a la salida de la puerta. La línea discontinua representa la corriente de cortocircuito ( $I_{sc}$ ) y las líneas continuas las corriente de carga o descarga de las capacidades.



**Figura 2-5**

Formas de onda correspondientes a la simulación de una puerta NAND de dos entradas. El panel superior muestra la evolución de las dos entradas, que siguen la secuencia 11→00→11. Los otros paneles corresponden a las corrientes medidas por los amperímetros mostrados en la Figura 2-4.

En segundo lugar tenemos una serie de oscilaciones en la corriente y la tensión de alimentación. Como veremos más adelante, cuando analicemos la red de alimentación y los encapsulados, estas oscilaciones se deben a que el sistema de alimentación forma un circuito RLC resonante.



**Figura 2-6**

Formas de onda correspondientes a la simulación de una puerta NAND de dos entradas. El panel superior muestra la evolución de las dos entradas, que siguen la secuencia 11→10→11. Los otros paneles corresponden a las corrientes medidas por los amperímetros mostrados en la Figura 2-4.

Además de la combinación de entrada existe otro factor que determina la forma de onda de la corriente de conmutación y por lo tanto el ruido  $dI/dt$  que se genera debido a esta corriente. Este factor es la carga de las salidas de las puertas CMOS. Los transistores de salida de estas puertas conducen mientras conmutan en dos regiones: óhmica o lineal y saturación, en función de la relación entre su tensión de puerta a surtidor ( $V_{GS}$ ) y de drenador a surtidor ( $V_{DS}$ ). La primera tensión viene determinada por la evolución de las entradas y la segunda por la evolución de la salida. El tiempo que tarda en subir (o bajar) la salida depende de la corriente que esté proporcionando el transistor o transistores correspondientes a la carga o descarga de la salida, que es principalmente capacitiva, y al valor de esa capacidad.

En general una conmutación de la salida de una puerta CMOS puede dividirse en las etapas que se ilustran en la Figura 2-7 para un inversor CMOS. En la figura se muestra el caso de la descarga (transición alto-bajo). Mientras la tensión de entrada menos la tensión umbral sea superior a la tensión de salida el transistor NMOS está en saturación. Una vez  $V_{GS} - V_{TN}$  cae por debajo de  $V_{DS}$  el NMOS entra en la zona lineal. Para una parte de la transición la corriente de descarga depende cuadráticamente de la tensión de entrada. En el resto de la transición la corriente pasa a depender también de la tensión de salida y es siempre menor a la de saturación.



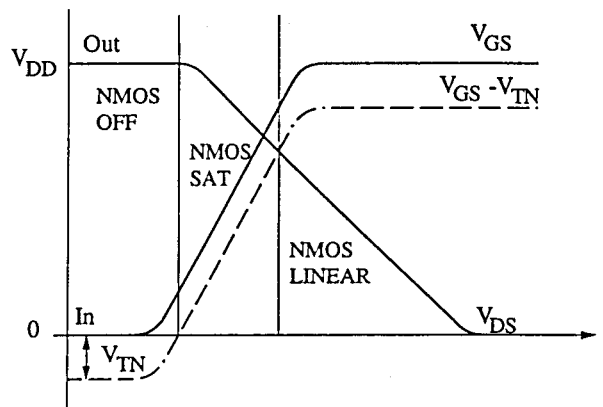
Por ello interesa que la mayor parte de la transición se realice con el NMOS en saturación. Cuanto mayor sea la capacidad de carga, mayor será el tiempo que tarde la salida en bajar. Si el tiempo que tarda en bajar la salida desde  $V_{DD}$  hasta  $V_{DD}-V_{TN}$  es mayor que el tiempo que tarda la entrada en subir de 0 a  $V_{DD}$  el transistor NMOS realiza toda la transición de salida en saturación. Este es el caso típico en circuitos VLSI donde se busca optimizar el *fanout*<sup>2</sup> de los circuitos lógicos. La influencia de la carga en la conmutación de circuitos CMOS ha sido estudiada en detalle en [1]. Es este trabajo se demuestra que por encima de una carga capacitiva crítica el ruido no aumenta con la capacidad de carga pues la corriente máxima queda limitada a la corriente de saturación para  $V_{GS}=V_{DD}-V_{TN}$ , y su evolución queda limitada al tiempo de transición de la entrada. El valor para esta capacidad crítica viene dado por:

$$C_{crit} = C_0 \left[ \frac{M}{n} - \sqrt{\frac{M^2}{n^2} + \frac{2M}{n}} \right]^2 \quad \text{Ec. (2-1)}$$

donde

$$M = \frac{t_r}{2V_{DD}\beta L} \quad C_0 = \frac{\beta t_r}{6V_{DD}V_T} [V_{DD} - V_T]^3 \quad \text{Ec. (2-2)}$$

En estas expresiones  $n$  es el número de inversores o *drivers* que conmutan simultáneamente,  $t_r$  el tiempo de subida (bajada) de la entrada,  $V_{DD}$  la tensión de alimentación,  $\beta$  la ganancia del NMOS (PMOS),  $L$  la longitud de canal y  $V_T$  la tensión umbral del NMOS (PMOS), donde los valores entre paréntesis corresponden al caso de una transición bajo-alto, y los otros a una transición alto-bajo a la salida del inversor.



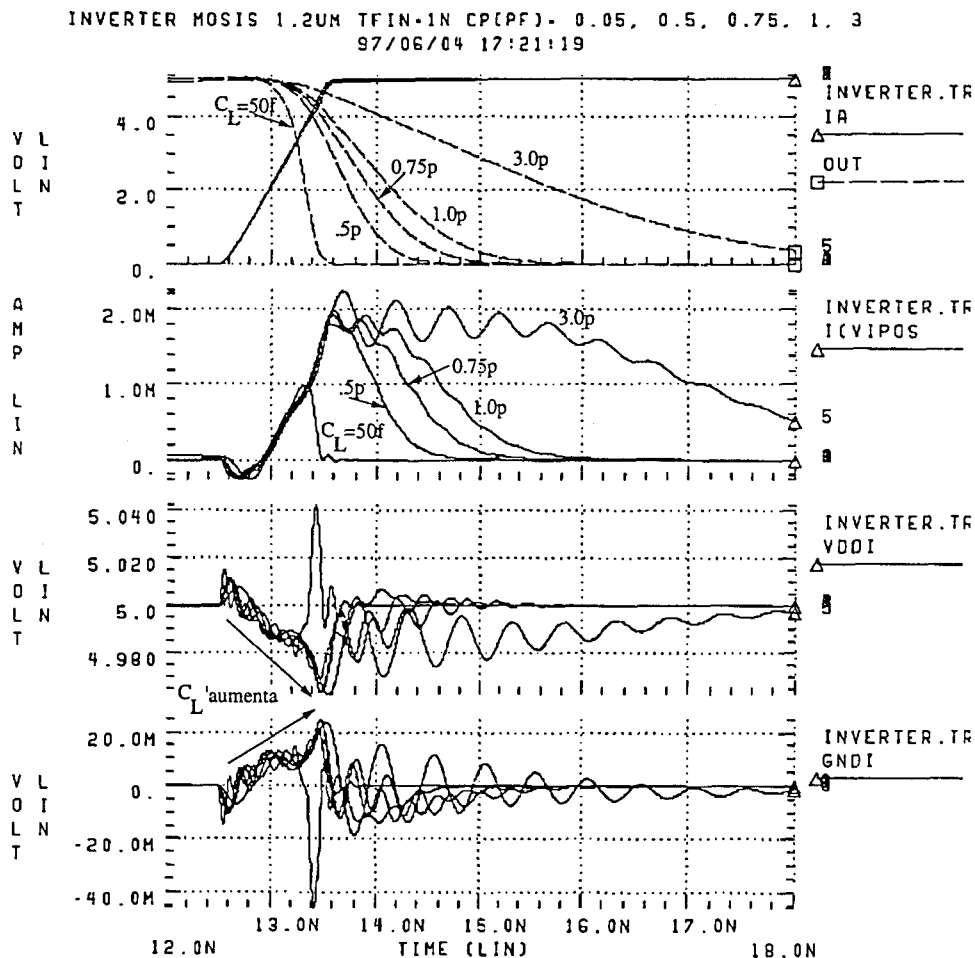
**Figura 2-7**

*Formas de onda de la salida y la entrada de un inversor CMOS y zonas de trabajo del transistor NMOS para una transición alto-bajo a la salida.*

En este trabajo se utilizan los modelos clásicos de canal largo para los MOS. Para dispositivos de canal corto o dominados por la velocidad de saturación de los portadores el

<sup>2</sup> El *fanout* de una puerta lógica se define como la relación entre la capacidad de carga de su salida y la capacidad de entrada. Un inversor cargado con cuatro inversores iguales a él tendrá un *fanout* de cuatro.

fenómeno básico es el mismo, si bien la expresión concreta para la capacidad límite por encima de la cual el ruido de conmutación no depende de la carga de la puerta es distinta. En nuestros análisis vamos a considerar siempre que las puertas CMOS están lo suficientemente cargadas como para verificar la condición anterior. Supondremos pues que la corriente con la que se cargan o descargan las capacidades de salida es siempre la corriente máxima de saturación y que la duración del flanco ascendente de corriente de conmutación coincide con el tiempo de subida o bajada de la entrada (el tiempo que tarda la entrada en realizar la excursión  $V_T \rightarrow V_{DD}$  o viceversa). En la Figura 2-8 podemos observar este efecto en una simulación realizada variando la capacidad de salida para el inversor CMOS de la Figura 2-2, dónde también podemos observar el ruido que se genera en la inductancia parásita del encapsulado debido a la forma de onda de la corriente. El ruido es proporcional a  $L_{eff} di(t)/dt.$ , dónde la inductancia efectiva ( $L_{eff}$ ) viene determinada por el encapsulado. En la simulación  $L_{eff}=8nH$  (4nH en cada terminal de alimentación).



**Figura 2-8**

Simulaciones del inversor de la Figura 2-2 con diferentes capacidades de carga. En el panel superior tenemos la tensión de entrada y la de salida. En el segundo la corriente de alimentación. En los dos inferiores el ruido en los terminales de  $V_{DD}$  y  $Gnd$  internos.

<sup>3</sup>  $L_{eff}$  en este entorno representa la inductancia efectiva del encapsulado. Más adelante, cuando hablemos del escalado tecnológico,  $L_{eff}$  se utilizará para referirse a la mínima longitud efectiva del canal de las tecnologías, y no debe confundirse con la significación anterior.

En la figura se observa como para capacidades por debajo de 0.5 pF, el pulso de corriente no alcanza su máximo posible. A partir de 0.5 pF el flanco ascendente es siempre el mismo y el descendente se va alargando a medida que aumenta la capacidad. Por ello la amplitud del pulso negativo de ruido en el nodo Vdd interno (y el positivo en el nodo Gnd interno) es prácticamente constante con la capacidad de carga por encima de 0.5 pF. El nivel y la frecuencia de la oscilación varían debido a que la capacidad de carga también influye en la resonancia RLC del circuito formado por la inductancia del encapsulado y la capacidad del circuito.

Si se cumple que la capacidad de carga en una puerta lógica es superior a la  $C_{crit}$  el ruido generado depende de la amplitud y el tiempo de subida de la corriente. La amplitud depende del tamaño de los transistores (PMOS para la transición bajo-alto a la salida y NMOS para la alto-bajo) y el tiempo de subida de la corriente es el tiempo de subida (de  $V_{TN}$  a  $V_{DD}$ ) de la entrada para transición bajo-alto y el tiempo de bajada (de  $V_{DD}-V_{TP}$  a 0) de la entrada para la transición alto-bajo a la salida. Si no se cumple esta condición la forma de onda de la corriente depende de la tensión de salida y además es necesario tener también en cuenta la corriente de cortocircuito ( $I_{sc}$ ). En este caso, si aumentamos la capacidad de carga, la transición se realiza más lentamente y el flanco de subida de la corriente se hace menos pendiente, por lo que el ruido generado disminuye [2]. Este efecto solo es apreciable para *fanouts* muy bajos para los que la mayor parte del tiempo el transistor encargado de la carga o la descarga de la salida se encuentra en zona óhmica.

### 2.1.2 Circuitos integrados en tecnología BiCMOS

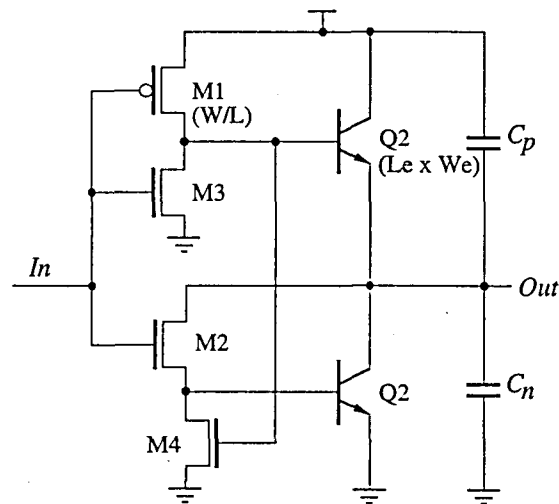
Los dispositivos CMOS presentan como principales ventajas una baja disipación de potencia, un alto nivel de integración y un elevado rendimiento de fabricación (*yield*). Los dispositivos bipolares conmutan más rápidamente, son capaces de conducir corrientes más elevadas y tienen unas propiedades de apareamiento (*matching*) mejores que los MOSFET. La tecnología BiCMOS que integra en el mismo sustrato ambos tipos de dispositivos aprovecha las ventajas de los dos tipos de dispositivos utilizándolos allí donde mejores prestaciones ofrecen. Se utilizan normalmente dispositivos CMOS (celdas de memoria, puertas lógicas, registros, multiplexores,...) y bipolares allí donde podemos sacar provecho de sus propiedades: *drivers* de la línea de palabra en las memorias, *sense amplifiers*<sup>4</sup>, *buffers*<sup>5</sup> para atacar a interconexiones globales, caminos críticos como la propagación del acarreo (*carry*) en circuitos aritméticos, *driver* de la señal de reloj, intercambio de señales de chip a chip,... En CI mixtos para la construcción de módulos analógicos se utilizan dispositivos CMOS siempre que se requiera una elevada impedancia de entrada y bipolares, por ejemplo, en los espejos de corriente y etapas diferenciales que requieren un buen apareamiento entre los dispositivos simétricos.

---

<sup>4</sup> Un *sense amplifier* es un circuito encargado de sensar valores bajos de tensión y detectar si corresponden a un valor lógico alto o '1' o bajo o '0'. Su uso más corrientes es en las memorias RAM donde los valores binarios se almacenan como pequeños valores de tensión sobre condensadores.

<sup>5</sup> Se conocen como *buffers* a las puertas lógicas diseñadas para conmutar rápidamente cargas de salida grandes.

Interesa analizar el ruido generado por las puertas lógicas BiCMOS, dado que estas se utilizan en los circuitos digitales precisamente para conmutar rápidamente señales de gran capacidad de carga (como el reloj o los buses internos). En la Figura 2-9 se presenta el circuito de un inversor BiCMOS estándar. Existen otras alternativas topológicas pero todas se basan en la estructura típica de la Figura 2-9 [3][4]. El circuito consta de un transistor PMOS (M1) y un NMOS (M2), cada uno seguido de un bipolar (BJTs Q1, Q2) en una configuración *totem pole*. Los transistores M3 y M4 sirven para mantener en *off* uno de los dos transistores BJT cuando el otro está cargando o descargando la carga. Cuando la entrada baja, M1 entra en conducción y proporciona corriente a Q1 que carga la capacidad de salida. Cuando la entrada sube, M2 entra en conducción y proporciona corriente a Q2 que descarga la salida. La superioridad de la puerta BiCMOS radica en la alta capacidad de conducción de los bipolares y el nulo consumo estático y la alta impedancia de entrada de los MOSFETs. Este inversor es fácilmente extensible para crear funciones lógicas de múltiples entradas del mismo modo que en circuitos lógicos CMOS. Por ejemplo, para la función NOR de tres entradas, M1 se substituye por 3 PMOS en serie y M2 por 3 NMOS en paralelo.



**Figura 2-9**

*Circuito de una puerta BiCMOS convencional.*

Vamos a analizar a continuación el comportamiento de la puerta BiCMOS convencional cuando conmuta su salida. Un análisis exacto queda fuera del objetivo de este trabajo y para ello referimos al lector a [5] y [6]. Llamaremos *pull-up* a la transición en la que Q1 carga la capacidad de salida de nivel bajo a alto y *pull-down* a la complementaria (realizada por Q2). Para la topología de la Figura 2-9 los niveles alto y bajo son  $V_{DD} - V_{BEON}$  y  $V_{BEON}$ , respectivamente, donde  $V_{BEON}$  es la tensión necesaria para que los BJT empiecen a conducir. Precisamente por eso, cuando estos transistores dejan de conducir, quedan unas diferencias de tensión entre la salida y los terminales de  $V_{DD}$  y  $Gnd$ . Si tomamos como referencia la transición *pull-up*, podemos analizar las diferentes fase del transitorio. En primer lugar cuando se produce una transición en la entrada (de alto a bajo) M1 tarda un cierto tiempo en entrar en conducción (el que necesita la entrada para bajar hasta  $V_{DD} - V_{TP}$ ). En ese momento M1 comienza a cargar las capacidades asociadas al nodo de base de Q1, hasta que transcurrido otro intervalo de tiempo el BJT empieza a conducir. A partir de entonces la corriente que circula por M1 es la corriente de

base de Q1 y éste comienza a cargar la capacidad de salida con su corriente de emisor. La capacidad de salida está compuesta, además de por la carga, por la capacidad parásita asociada a la salida, y que para esta transición principalmente corresponde a la capacidad entre el colector de Q2 y el substrato. De todas las corrientes que fluyen por el terminal de alimentación durante el transitorio la única que nos importa es la de emisor de Q1 que carga la capacidad de salida. El comportamiento de los BJTs depende de si entran en zona de alta inyección o no. Vamos a suponer para comenzar que Q1 se encuentra en baja inyección. La expresión que encontramos para la corriente de colector es:

$$I_c(t) = C_L^* \frac{dV_o(t)}{dt} = C_L^* (V_{DD} - V_{BEON}) \left[ \omega_b + \frac{1}{\omega_b \tau_b^2} \right] \sin(\omega_b t) e^{-\frac{t}{\tau_b}} \quad \text{Ec. (2-3)}$$

dónde  $C_L^* = C_L + C_s$ , es la capacidad efectiva de carga que incluye las capacidades parásitas del nodo de salida ( $C_s$ ). Los parámetros  $\omega_b$  y  $\tau_b$  son función de los parámetros del BJT y el PMOS M1:

$$\tau_b = 2\tau_F^* \beta^* \quad \omega_b = \frac{\sqrt{\frac{4\tau_F^*}{(R_1 + R_b)C_L^*} - (1/\beta^*)^2}}{2\tau_F^*} \quad \text{Ec. (2-4)}$$

dónde el tiempo de tránsito efectivo ( $\tau_F^*$ ) y la beta del BJT ( $\beta^*$ ) efectiva incluyen los efectos de las capacidades y resistencias parásitas del BJT (capacidad entre base y colector  $C_c$  y resistencia de colector  $R_c$ ) que modifican los valores normales del tiempo de tránsito ( $\tau_F$ ) y de la beta ( $\beta$ ):

$$\tau_F^* = \tau_F + R_c C_c \quad \frac{1}{\beta^*} = \frac{1}{\beta} + \frac{C_c}{C_L^*} \quad \text{Ec. (2-5)}$$

Las resistencias en (2-5) corresponden a la resistencia de base del bipolar ( $R_b$ ) y al modelo resistivo de M1:

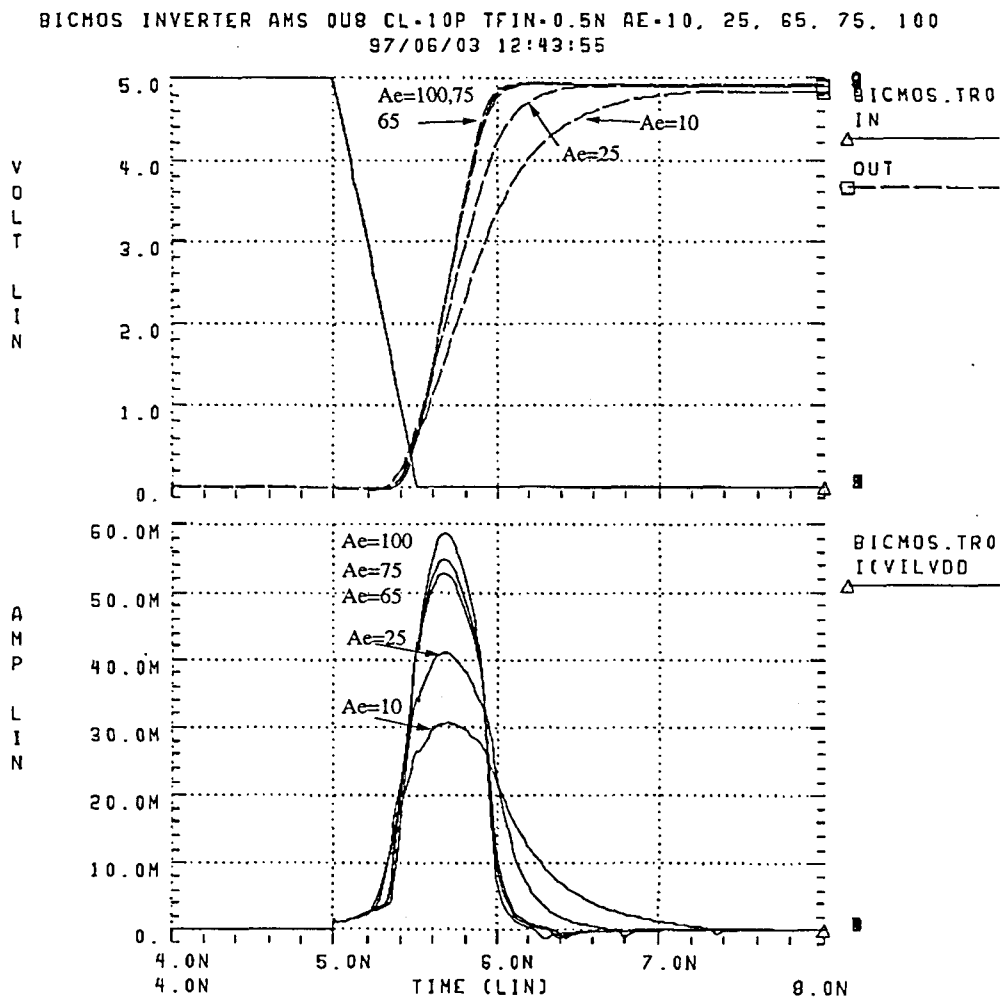
$$R_1 = \frac{2L_{eff}}{C_{ox}\mu_p W_1 (V_{DD} - |V_{TP}|)} \quad \text{Ec. (2-6)}$$

dónde  $L_{eff}$  es la longitud efectiva del canal de los MOS (que es la misma y la mínima para todos los dispositivos de este tipo de la puerta BiCMOS),  $C_{ox}$  la capacitancia del óxido de puerta,  $\mu_p$  la movilidad de los huecos,  $W_1$  la anchura de canal del PMOS M1 y el resto de parámetros tiene su significado habitual.

La forma de onda de la corriente de (2-3) puede aproximarse por una senoide si el término exponencial puede aproximarse por la unidad (es habitual que  $\tau_b \gg$  tiempo de conmutación de la salida). En la Figura 2-10 mostramos la simulación de una conmutación a la salida de una puerta BiCMOS para una tecnología de  $0.8\mu\text{m}$  y con una capacidad de carga de  $10\text{pF}$ . La puerta está diseñada de forma óptima para trabajar en el límite entre las zonas de baja y alta inyección,

lo que corresponde a un área de emisor para los bipolares (Q1 y Q2) de 65 (respecto al transistor BJT de área unitaria de la tecnología). En la gráfica se han representado diferentes casos para el área y se observa el efecto de la alta inyección cuando el área es menor de 65. Por encima de este valor de área no se mejora el retardo.

Hemos realizado también estudios sobre la influencia del tiempo de subida o bajada de la entrada. La disminución del tiempo de subida o bajada de la entrada, produce un ligero incremento en la pendiente del flanco ascendente de la corriente. Su efecto sólo es realmente importante si la entrada cambia muy lentamente [7].

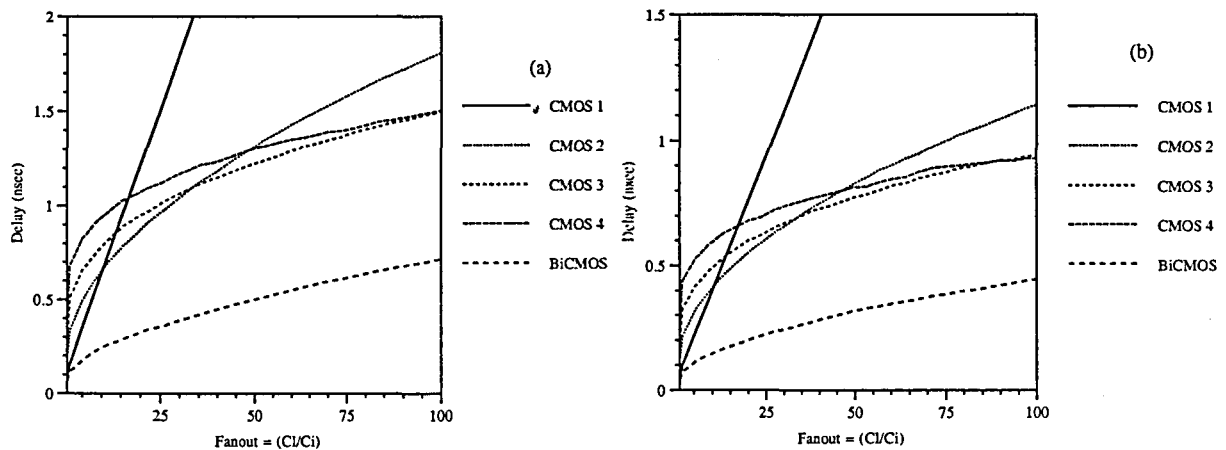


**Figura 2-10**

Simulación del inversor convencional BiCMOS para una tecnología de  $0.8\mu\text{m}$ . En la figura se presentan diferentes simulaciones de las tensiones de entrada y salida y la corriente de alimentación para diferentes áreas de emisor de los BJTs. El valor óptimo es  $Ae=65$ .

### 2.1.3 Comparación del ruido generado por puertas CMOS y BiCMOS

Trataremos ahora de comparar el ruido generado por dos inversores equivalentes CMOS y BiCMOS. Para que la comparación sea justa, sería necesario que ambas puertas tuvieran la misma capacidad de entrada y el mismo retardo para la misma capacidad de salida [3]. Sin embargo conseguir estas tres condiciones es imposible para las tecnologías BiCMOS que hemos escogido para el análisis. En la Figura 2-11 se han representado curvas de retardo en función de la capacidad de salida, para capacidad de entrada fija en dos tecnologías BiCMOS de AMS de  $1.2\mu\text{m}$  y  $0.8\mu\text{m}$  [8], respectivamente. Para cada punto de las gráficas se ha calculado el tamaño óptimo de los transistores MOS de una cadena de 1, 2, 3 y 4 inversores CMOS también de las tecnologías de  $1.2\mu\text{m}$  y  $0.8\mu\text{m}$ , y de los bipolares de un inversor convencional BiCMOS [9][10]. Como vemos ninguna de las curvas de puertas CMOS llega a cruzarse con la curva de BiCMOS por lo que el retardo BiCMOS es siempre inferior al CMOS.



**Figura 2-11**

*Retardos óptimos en función de la capacidad de carga para cadenas de inversores CMOS de 1, 2, 3 y 4 etapas y comparación con el retardo de un inversor BiCMOS con la misma capacidad de entrada, para dos tecnologías (a) de  $1.2\mu\text{m}$  y (b) de  $0.8\mu\text{m}$ , obtenidos por simulación.*

También se ha comprobado, aunque no se muestra aquí, que el aumento de la capacidad de salida provoca un aumento de la anchura del pulso de corriente y de su amplitud, pero la forma de su flanco ascendente permanece constante. Si se aumenta mucho la capacidad de salida puede llevarse a la puerta a trabajar en alta inyección y entonces sí se deforma el pulso de corriente (ver Figura 2-10). Como se ha comentado, un buen diseño de la puerta BiCMOS tiene en cuenta la máxima capacidad de salida y entonces se calcula el área de emisor para la cual los BJTs trabajan en el límite entre la zona de baja y alta inyección.

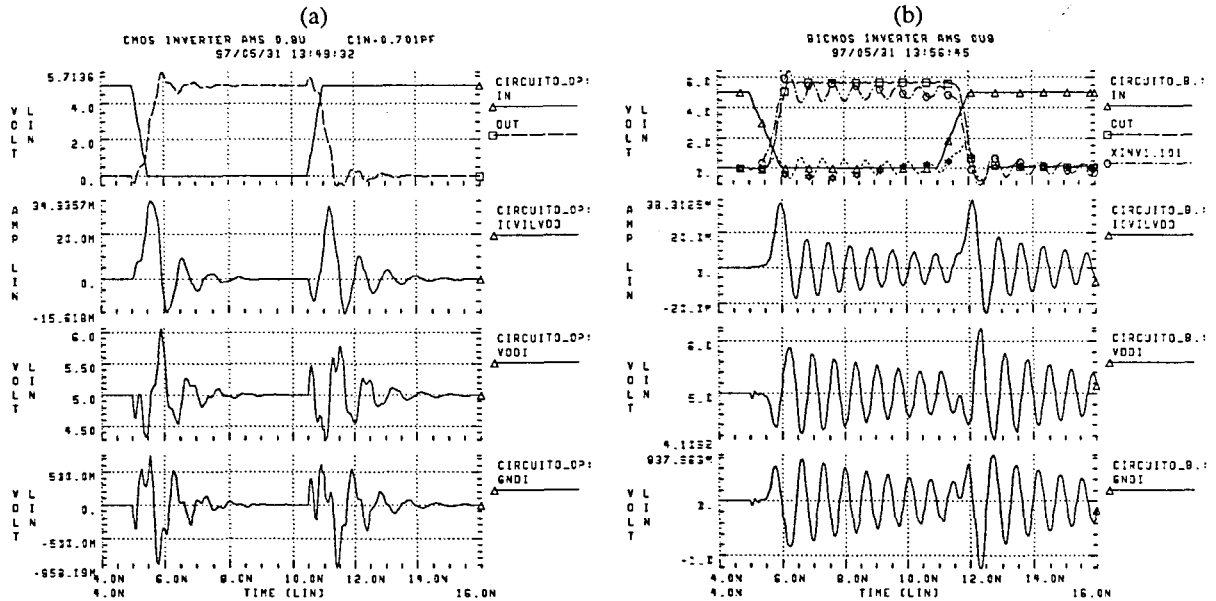


Figura 2-12

Formas de onda de tensión y corriente para una conmutación (a) de un inversor CMOS y (b) BiCMOS que permiten comparar el ruido de conmutación generado por ambos.

La comparación se ha realizado, ante la imposibilidad de satisfacer las tres condiciones anteriormente mencionadas, con un inversor CMOS optimizado para obtener el mismo retardo que un inversor BiCMOS con una capacidad de entrada de 191fF para la tecnología de 0.8 $\mu$ m. La carga ha sido en ambos casos de 3.82pF, lo que para el inversor BiCMOS representa un *fanout* de 20. La capacidad de entrada del inversor CMOS, con el mismo retardo para la misma carga de salida, es de 701fF. Es decir, el *fanout* para el inversor CMOS es tan solo de 3.67. El tamaño óptimo del área de emisor ( $A_e$ ) de los BJT's del inversor BiCMOS para trabajar en el límite de entre las zonas de baja y alta inyección es de 48, mientras que las anchuras de los transistores del inversor CMOS son de 102.7 $\mu$ m para el NMOS y de 289.84 $\mu$ m para el PMOS. Todas las longitudes son de 0.8 $\mu$ m. En la Figura 2-12.(b) tenemos el ruido generado por el inversor BiCMOS y en la Figura 2-12.(a) el correspondiente al inversor CMOS para un tiempo de subida de la entrada de 0.5n segundos. El circuito que simula el encapsulado y la alimentación es el mismo de la Figura 2-2, pero esta vez sólo conmuta un inversor (CMOS o BiCMOS) simultáneamente.

Como se puede apreciar en la simulación realizada, el ruido generado en ambos casos es muy similar. Aunque las formas de onda de la corriente no son exactamente iguales, si nos basamos en la hipótesis de que la mitad de la carga necesaria para realizar la conmutación se ha transferido ya en el momento en que la salida alcanza la mitad de la tensión de alimentación, y el tiempo que se tarda en realizar esta evolución es el mismo para las dos puertas (tienen el mismo retardo) el valor máximo de la corriente y la semi anchura del pulso de corriente deben ser aproximadamente el mismo, y por tanto la  $dI/dt$  será muy similar en ambos casos, como indica la simulación. Por tanto, cabe concluir que no puede decirse que un tipo de puerta sea más ruidosa que la otra, en igualdad de retardo y capacidad de carga, es decir, que ambas generan pulsos de ruido de conmutación de la misma amplitud.



## 2.2 Modelos analíticos del ruido de conmutación para circuitos integrados CMOS

Hasta ahora se ha tratado el problema de la generación, transmisión y efectos del ruido de conmutación de forma cualitativa, centrándonos sobre todo en los mecanismos físicos y electrónicos en los que se basa. En este apartado se resumen las diferentes tentativas aparecidas en la literatura para modelar de forma analítica el ruido de conmutación.

### 2.2.1 Modelos lineales del ruido de conmutación simultáneo (SSN)

Los primeros modelos de ruido de conmutación aparecen en el entorno de los circuitos de alta velocidad ECL y se centran en el estudio de los *drivers* de salida [16][25]. Se utiliza un pulso triangular para modelar el ruido presente en las alimentaciones y se estudia su propagación y su variación con los parámetros del proceso de fabricación. En estos trabajos se asume que si  $N$  *drivers* de salida conmutan simultáneamente, el ruido total es  $N$  veces superior. También se analiza el ruido que se genera cuando las transiciones de varios *drivers* no son exactamente simultáneas. Si la diferencia temporal entre dos pulsos de ruido separación es de la cuarta parte de la anchura en la base del pulso triangular utilizado en el modelo, el ruido resulta ser un 60% del que se obtendría si las transiciones fueran exactamente simultáneas. Esta reducción es del 20% si los pulsos están separados por la mitad de la anchura del pulso.

### 2.2.2 Modelos sub-lineales e inclusión de efectos de canal corto

El ruido presente en las alimentaciones modifica las tensiones de referencia para las entradas y las salidas de las puertas lógicas. De esta forma existe un efecto de retroalimentación que limita el ruido. Este efecto fue estudiado en primer lugar en [11]. Considerando que la forma de onda de la corriente era triangular y la semi anchura de la base era el tiempo de subida o bajada de la transición de entrada se llega a la siguiente expresión para el ruido de conmutación:

$$V_n = V_k + \frac{t_r}{L_{eff}NK} \left[ 1 - \sqrt{1 + 2V_k \frac{L_{eff}NK}{t_r}} \right] \quad Ec. (2-7)$$

donde  $V_k = V_{in} - V_t$  (la tensión de entrada menos la tensión umbral),  $t_r$  es el tiempo que tarda en subir la corriente desde cero hasta su valor máximo,  $L_{eff}$  es la inductancia efectiva que encontramos en el bucle desde la alimentación hasta el nodo que se descarga o se carga debido a la transición de la salida,  $K = \mu_n C_{ox}(W/L)$  para el transistor NMOS (y análogamente para el PMOS substituyendo adecuadamente la movilidad  $\mu_p$  y las dimensiones geométricas  $W$  y  $L$ ), y finalmente  $N$ , como ya hemos dicho, es el número de *drivers* que conmutan simultáneamente. En esta expresión no aparece ninguna dependencia con la capacidad de salida, ya que se ha realizado la suposición que se da la condición expuesta en el subapartado 2.1.1., Ec. (2-1).

Posteriormente en [1] se utiliza un pulso de corriente parabólico más realista basado en la asunción que el ruido aumenta linealmente durante el tiempo de transición de la entrada. De esta forma se llega a una expresión muy similar a la anterior:

$$V_n = V_k + \frac{V_k t_r}{2V_{DD} L_{eff} NK} \left[ 1 - \sqrt{1 + 4V_{DD} \frac{L_{eff} NK}{t_r}} \right] \quad Ec. (2-8)$$

Ahora  $V_k = V_{DD} - V_t$ , la tensión de alimentación menos la tensión umbral y el resto de parámetros tienen el mismo significado. Estos dos modelos incluyen el efecto de retroalimentación del propio ruido de conmutación. Este efecto supone una saturación del aumento del ruido de conmutación generado a medida que se incrementa el número de puertas conmutando simultáneamente. Otro modelo más complejo de este tipo podemos encontrarlo en [12].

Todos los modelos anteriores utilizan el modelo de Shichman-Hodges para los transistores MOS. Sin embargo en los dispositivos submicrónicos actuales caracterizados por las reducidas dimensiones del canal es necesario tener en cuenta los fenómenos de canal corto. Es necesario recurrir a modelos más realistas que contemplen la limitación de la movilidad debida a los campos verticales en el canal, la reducción de la longitud efectiva del canal, la saturación de la velocidad de los portadores debido al elevado campo longitudinal, etc... Los modelos del ruido de conmutación deberán adaptarse a estos nuevos fenómenos para predecir adecuadamente el comportamiento de los CI actuales. De los efectos de canal corto citados anteriormente, el que más afecta a la generación de ruido es sin duda el de la saturación de la velocidad de los portadores. Si en dispositivos de canal largo durante la transición los dispositivos podían trabajar en zona lineal o en saturación, en dispositivos de canal corto diseñados para conmutar rápidamente prácticamente siempre vamos a tener el transistor trabajando en saturación. Una saturación en la que la corriente no es proporcional de forma cuadrática con las tensiones, como sucedía para canal largo, sino que depende de las tensiones con un exponente menor que 2. En dispositivos fuertemente dominados por la velocidad de saturación la corriente de saturación puede llegar a ser linealmente proporcional a las tensiones, es decir, el exponente llega a tener un valor de 1. Si no se tienen en cuenta estas variaciones de la dependencia de la corriente de saturación con la tensión de entrada estaremos sobrestimando el nivel de ruido de conmutación. Un modelo de ruido de conmutación basado en estas consideraciones [13] utiliza una dependencia con un exponente variable de la corriente de saturación [14]. Este modelo propone la siguiente expresión para el ruido de conmutación  $dI/dt$ :

$$V_n = s_r NK L_{eff} f \left[ 1 - e^{-\frac{(t-t_n)}{NK L_{eff} f}} \right] \quad t_n \leq t \leq t_r \quad Ec. (2-9)$$

donde todo es conocido excepto  $s_r$  que representa la pendiente de la transición de la entrada y  $t_n$  que indica el tiempo en el que el transistor NMOS o PMOS, dependiendo del sentido de la transición de salida, entra en conducción ( $t_n = s_r / V_t$ ). El parámetro  $f$  debe determinarse

empíricamente e incluye la dependencia con el exponente  $\alpha_n$  de la expresión de la corriente de saturación en función de la tensión de entrada de los dispositivos. Este parámetro  $f$  es una constante si  $\alpha_n$  está próximo a la unidad. El producto  $Kf$  tiene unidades de admitancia [A/V].

Para dispositivos fuertemente controlados por la saturación de la velocidad de los portadores existe el modelo publicado en [15].

$$V_n = \frac{s_r N K L_{eff} V_C}{2} \left[ 1 - e^{-2 \frac{(t-t_n)}{N K L_{eff} V_C}} \right] \quad t_n \leq t \leq t_r \quad Ec. (2-10)$$

donde  $f$  ha sido substituida por  $V_C$  que es la tensión lateral umbral por encima de la cual se da la saturación de la velocidad de los portadores ( $V_C = Lv_{sat} / \mu_{eff}$ , con  $L$  la longitud efectiva del canal,  $v_{sat}$  la velocidad de saturación y  $\mu_{eff}$  es la movilidad efectiva que puede calcularse incluyendo los efectos del campo vertical en el canal).

### 2.2.3 Parámetros influyentes en el ruido de conmutación

Los parámetros de los que depende el ruido de conmutación pueden identificarse fácilmente a partir de las expresiones del subapartado anterior. En primer lugar la tensión de alimentación  $V_{DD}$  aparece directamente o bien a través de la tensión de entrada o de la pendiente de la tensión de entrada en todos los modelos. La dependencia es lineal en todos los casos. Sin embargo hay que tener en cuenta que la movilidad presente en la  $K$  y en el modelo de la Ec. (2-10) también depende de la tensión de alimentación. En este sentido una disminución de la tensión de alimentación por una parte hace disminuir directamente la amplitud del ruido, pero por otra parte hace aumentar las movilidades. Esto compensa, en cierto grado, la disminución del ruido producida por la reducción de la tensión.

Las dimensiones de los dispositivos afectan a la ganancia  $K$  y a medida que se reducen hacen aparecer los efectos de canal corto. Hemos dedicado más adelante un capítulo entero para analizar el escalado del ruido de conmutación donde presentamos los resultados obtenidos en función de las tendencias actuales y las perspectivas futuras de la evolución de la tecnología CMOS hasta dispositivos con longitud de canal por debajo de  $0.1\mu\text{m}$ . En [13] se comparan los resultados de una simulación SPICE con los valores que predicen los diferentes modelos presentados en el subapartado anterior. Para tecnologías de  $1.6\mu\text{m}$  y  $2.0\mu\text{m}$  las Ec. (2-7) y (2-8) se ajustan mucho mejor a los datos. Sin embargo para tecnologías menores de  $0.8\mu\text{m}$  y  $1.2\mu\text{m}$  es necesario recurrir al modelo de la Ec. (2-9).

Otro de los parámetros que afecta directamente al ruido de conmutación es el tiempo de transición de la entrada. Un tiempo menor implica un aumento del ruido generado en la misma proporción. Esos tiempos de transición vienen impuestos por la lógica anterior a los *drivers* o puertas que se consideran como generadores del ruido y su valor no es arbitrario sino que debe considerarse en el conjunto de parámetros de la tecnología y dependerán del diseño de las etapas anteriores. En general en el diseño VLSI se intenta maximizar la velocidad por lo que estos

tiempos de transición serán siempre los más rápidos posibles en cada caso, para evitar penalizar el retardo del circuito lógico.

Finalmente tenemos la dependencia con el número  $N$  de puertas lógicas que realizan la conmutación simultáneamente. Tanto en los modelos de canal largo como en los de canal corto vemos que existe un término de corrección que se resta de la expresión lineal y que depende de  $N$ . Este término es debido al efecto de retroacción que el propio ruido presente en las alimentaciones introduce en las tensiones de referencia de las entradas y las salidas de las puertas que conmutan, que son esas mismas alimentaciones contaminadas por el ruido. Cuanto mayor es el ruido generado menor es el margen entre  $V_{DD}$  y Gnd, por lo que las puertas se vuelven más lentas y se reduce el nivel de ruido generado. Este efecto no lineal es realmente importante si el nivel de ruido generado es alto.

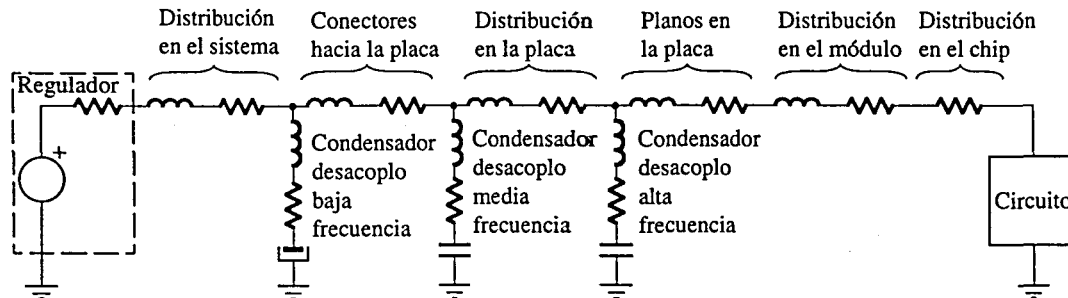
## 2.3 Mecanismos de transmisión del ruido de conmutación

El ruido de conmutación  $dI/dt$  generado por las partes digitales de un CI está presente, como se ha visto, en los terminales de alimentación. El ruido puede transmitirse a circuitos digitales o analógicos que comparten el mismo sistema de alimentación. Por tanto es necesario estudiar en primer lugar la influencia de la red de distribución de la tensión de alimentación en la transmisión del ruido de conmutación. Otro elemento que comparten todos los dispositivos de un CI es el sustrato semiconductor. Dado que dicho sustrato se polariza mediante contactos con las alimentaciones y que pueden estar contaminadas con ruido éste se introduce en el sustrato pudiendo afectar a otras partes del CI que incluso tengan diferentes terminales de alimentación que la parte generadora del ruido. El ruido generado en un CI encapsulado puede afectar también a otros CI que compartan con él señales de entrada/salida. Un emisor en un CI ruidoso refiere su salida a las tensiones de alimentación contaminadas y ese ruido está presente también en la salida. Un receptor de otro CI verá a su entrada esta señal contaminada por el ruido y por tanto puede funcionar incorrectamente o incluso transmitir ese ruido a su sistema de alimentación originalmente no ruidoso o “tranquilo”.

### 2.3.1 Distribución de la alimentación en circuitos integrados

El sistema de distribución de la alimentación comienza en el regulador de la fuente de alimentación. En un sistema electrónico general la alimentación debe distribuirse a través de conectores a las diferentes placas de circuito impreso y por estas a través de pistas o planos de alimentación a los diferentes CI encapsulados. Para CI complejos el propio encapsulado puede contar con planos de distribución de dicha alimentación hasta las diferentes partes del circuito o en módulos multi-chip (MCMs) el esquema de distribución de la placa base se reproduce a una escala más pequeña. Todo este sistema de distribución de la alimentación forma una compleja red RLC donde las pistas de metal aportan su resistencia y su inductancia. La capacidad, además de a las interconexiones, es debida a los condensadores de desacoplo que se introducen en diferentes puntos de la red de alimentación precisamente para disminuir en la medida de lo posible el ruido de conmutación (ver Figura 2-13). Se tratará con más detalle en un capítulo

específico esta visión global de la distribución de la alimentación en la que participan la placa de circuito impreso y el encapsulado, y nos centraremos aquí únicamente en la transmisión del ruido de conmutación dentro del propio CI debido a la estructura de su red de distribución de la alimentación (dentro del chip: *on-chip*), que es principalmente resistiva.

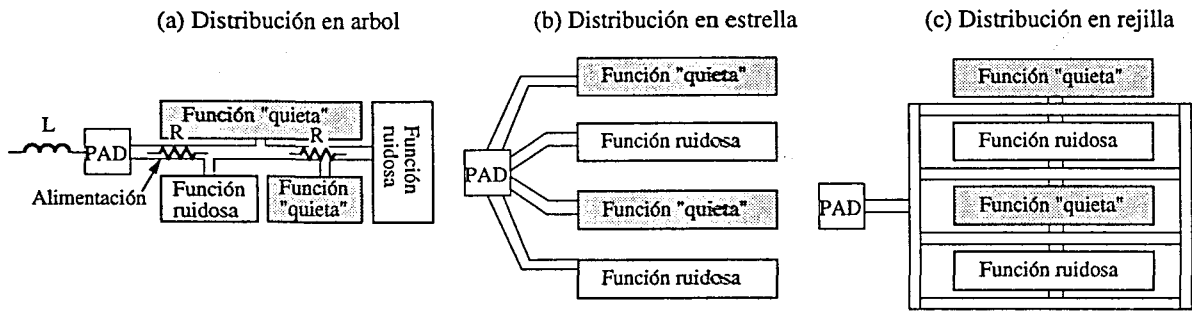


**Figura 2-13**

*Representación esquemática de la distribución de la alimentación en un sistema electrónico genérico. Los condensadores de desacoplo que se insertan en diferentes puntos se representan con su resistencia y su inductancia parásitas [16].*

En [17] se analizan las diferentes alternativas que podemos encontrar a la hora de distribuir la alimentación. Aunque uno de los terminales es común a todos los dispositivos (Gnd en tecnologías de sustrato tipo p y  $V_{DD}$  en las de sustrato tipo n) por compartir el mismo sustrato, la distribución se realiza de ambas alimentaciones. En el CI tenemos partes ruidosas y partes quietas (no ruidosas) que pueden ser o bien módulos digitales que en ese momento no están conmutando o partes analógicas. Una primera alternativa es alimentar todos los módulos funcionales con el mismo terminal interno (Figura 2-14). Podemos distribuir la alimentación en forma de árbol, de estrella o mediante una rejilla o *grid* regular. La segunda alternativa consiste en identificar aquellos módulos más ruidosos y alimentarlos por separado (con un terminal interno diferente) de los otros módulos no ruidosos (Figura 2-15). Cada grupo tiene su propia distribución de la alimentación que puede ser de cualquiera de los tipos anteriores. De esta forma se logra un aislamiento teórico entre las dos redes de distribución de la alimentación (o *buses* de alimentación) a costa de aumentar el área y el número de pads y pins dedicados a la alimentación. En la Tabla 2-1 se resumen las ventajas y desventajas de cada una de las técnicas expuestas. Se indica con un signo ‘↑’ o ‘↓’ la ventaja o desventaja, respectivamente, en un determinado aspecto como el aislamiento o la facilidad de automatizar el *routing*<sup>6</sup> de las pistas de alimentación. Se utiliza también los signos ‘+’ o ‘-’, indicando grande o pequeña respectivamente, para la impedancia total (la resistencia) de una determinada alternativa de distribución de la alimentación o para comparar el área ocupada en cada alternativa.

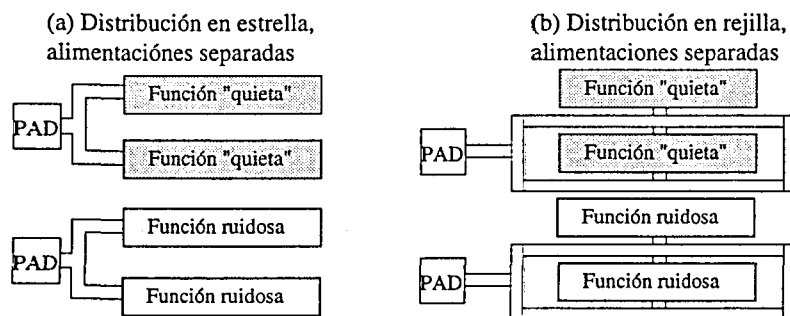
<sup>6</sup> El término inglés *routing* se utiliza aquí para referirse al trazado de las interconexiones en un CI, en el proceso de dibujo de las máscaras (*layout*) que permitirán la posterior fabricación del mismo.



**Figura 2-14**

*Esquemas de distribución de la alimentación entre módulos funcionales ruidosos y no ruidosos ("quietos") [17].*

De cara al ruido de transmisión nos interesa sobre todo conseguir un gran aislamiento. La estructura en árbol debido a la gran resistencia del bus de alimentación (*power bus*) produce un gran acoplamiento resistivo del ruido. Las otras dos estructuras reducen este acoplamiento resistivo. El valor de la resistencia del bus de alimentación es importante por tres aspectos. Uno, como ya hemos citado, es que una resistencia muy grande produce bajo aislamiento. Los otros dos están relacionados con aspectos que pueden reducir el ruido generado: una baja resistencia facilita la inserción de condensadores de desacoplo *on-chip* y, por contra, un valor alto ayuda a la atenuación de las oscilaciones producidas por la estructura resonante RLC que forma el sistema de distribución de la alimentación global (incluyendo los parásitos del encapsulado), lo que da lugar a un compromiso de diseño.



**Figura 2-15**

*Esquemas de distribución de la alimentación con separación de terminales entre las partes ruidosas y las quietas de un CI [17].*

Cuando se utiliza una distribución separada de la alimentación, que suele coincidir con la partición entre módulos digitales y analógicos en un CI mixto, nos enfrentamos al problema de la transmisión del ruido generado por un bus de alimentación ruidoso a otro no ruidoso en los puntos en los que bloques funcionales alimentados por uno de los dos buses se comunican con bloques funcionales alimentados por el otro bus. Dado que las salidas en el dominio ruidoso se refieren a tensiones de alimentación que están contaminadas por ruido, éste aparece también en las salidas y si estas señales van a parar a entradas de dispositivos del dominio no ruidoso pueden producir dos efectos: desplazamientos de fase (*time jitter*) de las conmutaciones a la salida de los módulos receptores [17][18] y acoplamiento del ruido hacia el bus no ruidoso. Un ejemplo del primer efecto es la recepción del reloj generado con un PLL en un dominio de

alimentación no ruidoso (analógico) por un *buffer* en el dominio ruidoso (digital). El reloj distribuido en la parte ruidosa tiene una variación en los tiempos de subida y bajada y los instantes de transmisión dado que el umbral de conmutación del *buffer* varía con el tiempo debido al ruido presente en sus terminales de alimentación. La solución a este problema es utilizar, siempre que sea posible, transmisión diferencial de las señales entre los dos dominios, aunque esto es complicado en circuitos digitales CMOS que son, por definición, monopolares. En cuanto al acoplamiento del ruido entre los dos dominios se produce por las capacidades entre las entradas y las alimentaciones y entre las entradas y la salida (si el receptor es una puerta CMOS estas capacidades son las de solapamiento entre la puerta y los surtidores y drenadores de los dispositivos PMOS y NMOS). Para reducir este acoplamiento es necesario minimizar estas capacidades. Sin embargo si hay varias señales que pasan del dominio ruidoso al no ruidoso el efecto del acoplamiento capacitivo es aditivo. Un caso crítico de este tipo de acoplamiento lo podemos encontrar en los circuitos CMOS que utilizan técnicas de capacidades conmutadas. La parte digital genera las señales de control de los interruptores que conmutan nodos analógicos muy sensibles al ruido. El ruido presente en la señal de control, a pesar de que esté fija a un valor lógico determinado, puede acoplarse fácilmente a través de las capacidades puerta-drenador y puerta-surtidor a los nodos que conecta el interruptor. La solución consiste en utilizar *buffers* alimentados por el bus no ruidoso entre la señal de control ruidoso y la puerta del interruptor MOS [17].

Aspecto	Alimentación Única			Alim. Separada	
	Árbol	Estrella	Rejilla	Estrella	Rejilla
Área	-	+	++	++	++
Impedancia	++	+	-	+	-
Routing	↓	↓	↑↑	↓	↑
Aislamiento	↓↓	↑	↓	↑↑	↑↑

Tabla 2-1

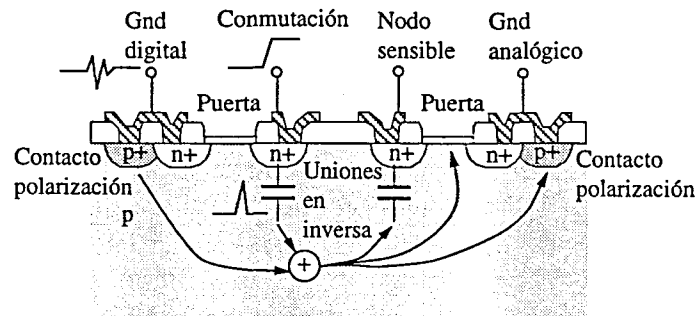
*Comparación de las distintas alternativas de distribución de la alimentación en función de varios aspectos como el área, la impedancia, la facilidad de routing o el aislamiento entre las partes ruidosas y las no ruidosas en un CI. Se ha sombreado la mejor alternativa de distribución de la alimentación para cada uno de los aspectos comparados.*

### 2.3.2 Polarización del sustrato

El sustrato sobre el que se construyen los dispositivos que forman el CI es una vía de transmisión de ruido. Existen numerosos trabajos que estudian este fenómeno entre los que podemos destacar [19][20][21][22][23]. El sustrato se comporta como una red resistiva y los puntos de inyección del ruido pueden ser por un lado las difusiones entre los drenadores y surtidores de los dispositivos (a través de su capacidad de unión polarizada en inversa) y por otro el contacto directo con las alimentaciones a través de los contactos de polarización. En tecnologías CMOS habitualmente disponemos de un sustrato dopado con un tipo de impurezas y pozos dopados con el tipo contrario. Los pozos pueden aislarse entre sí por lo que la transmisión del ruido entre ellos queda prácticamente eliminada, pero el sustrato es común a

todo el circuito. Existen tecnologías con dos pozos (*twin well*), uno de cada tipo, pero siempre uno de los dos es del mismo tipo que el sustrato. Si pozo y sustrato son del mismo tipo sus uniones quedan polarizadas en directa y no se resuelve el problema del acoplamiento del ruido. La única solución que elimina totalmente la transmisión de ruido por el sustrato es la tecnología SOI (*silicon on insulator*) que consiste en hacer crecer silicio semiconductor sobre un sustrato aislante y aislar todos los dispositivos semiconductores entre sí con barreras dieléctricas.

Los dos principales mecanismos de inyección de ruido en el sustrato, como se ilustra en la Figura 2-16, son la inyección capacitiva a través de las uniones PN que se forman entre los dispositivos y el sustrato o los pozos y el ruido presente en las líneas de alimentación que se utilizan para polarizar el sustrato o los pozos. De estos dos mecanismos el más importante es el debido al ruido de conmutación presente en las alimentaciones [24]. A los mecanismos de recepción del ruido presente en el sustrato por dispositivos remotos además de los dos ya citados hay que añadir el efecto sustrato (*body effect*) que provoca variaciones en la tensión umbral debido a la variación temporal de la tensión del sustrato producida por el ruido.



**Figura 2-16**

*Corte transversal de dos dispositivos NMOS pertenecientes a dominios de alimentación distinto (el de la izquierda pertenece a la parte digital ruidosa del circuito y el de la derecha a la parte analógica "quieta"). En el dibujo se muestran las dos principales vías de inyección de ruido en el sustrato y las tres mecanismos de recepción de este ruido por el dispositivo víctima.*

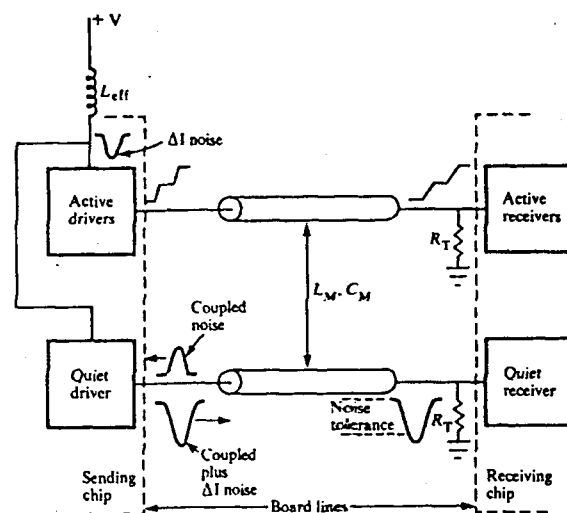
De los diferentes métodos utilizados para minimizar este tipo de transmisión del ruido cabe destacar la utilización de contactos, pads y pins dedicados exclusivamente a la polarización del sustrato. De esta forma se evita introducir ruido presente en las alimentaciones de la parte ruidosa del circuito en el sustrato común a todo el circuito. Del mismo modo es aconsejable utilizar sustratos poco conductivos (sin capa epitaxial fuertemente dopada) que atenúan más el ruido inyectado.

### 2.3.3 Comunicación entre dos circuitos integrados

Otro mecanismo por el que el ruido generado por la conmutación digital se propaga a otras partes del circuito y a otros circuitos del sistema es a través de las señales de chip a chip [25][26]. Los *drivers* de salida de un CI son cadenas de inversores diseñados para conmutar grandes cantidades de carga rápidamente. Todas las salidas que formen parte de un bus pueden llegar a conmutar simultáneamente y esto genera una gran cantidad de ruido  $dI/dt$  en las líneas de alimentación de estos *drivers* de salida. Por ello una práctica habitual consiste en separar las



alimentaciones de estos *drivers*, que suelen situarse en la periferia del CI, y las del resto del chip. Esto reduce la cantidad de ruido que puede propagarse hacia el interior del CI pero no hacia el exterior. El ruido produce que las alimentaciones fluctúen con el tiempo y las tensiones de umbral de los inversores que componen estos *drivers* de salida también lo hacen. Por ello los flancos de bajada y subida sufren errores de fase que hace disminuir o aumentar el retardo de la transmisión. Además el ruido generado por las salidas que conmutan se acopla directamente a las salidas estáticas a través del PMOS o del NMOS, según el nivel de la salida sea alto o bajo, respectivamente. Estos pulsos de ruido indeseados que aparecen en las salidas se propagan a través de las pistas de conexión entre el CI emisor y el receptor y pueden provocar fallos en el CI receptor o acoplarse a sus líneas de alimentación originalmente no ruidosas, ya que la alimentación de los *drivers* de las entradas de un CI suelen ser las mismas que la del resto de la circuitería interna. La mejor solución para evitar este tipos de problemas es utilizar señales diferenciales para la comunicación entre CIs [27], pero esto duplica la anchura de los buses. Otra posible solución, que es una de las aportaciones de esta tesis y que se expondrá más adelante, consiste en utilizar comunicación en modo corriente multivaluada entre CIs.



**Figura 2-17**

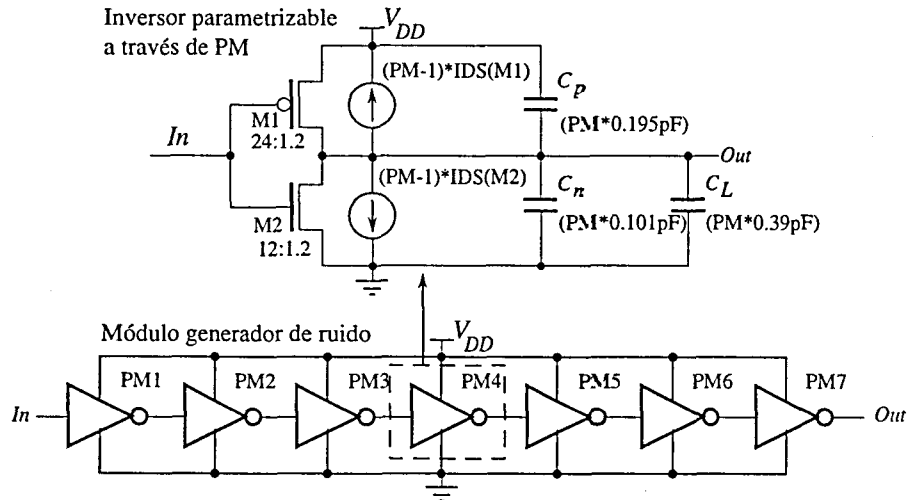
*Representación esquemática de los efectos del ruido de conmutación en las comunicaciones entre chips [16].*

En la figura se ilustra cómo el ruido presente en las alimentaciones del chip emisor se acopla al driver que conmuta produciendo error en el retardo. El ruido también se acopla a un driver que está quieto y se transmite al chip emisor. En el trayecto entre ambos y debido al crosstalk entre líneas existe un pulso de ruido de tipo  $dV/dt$  acoplado de la línea que conmuta a la que no lo hace que se suma al pulso de ruido de conmutación  $dI/dt$ .

## 2.4 Efectos del ruido de conmutación en circuitos digitales

Los primeros estudios sobre los efectos del ruido de conmutación se centraron en los circuitos puramente digitales [16][25][28]. En los circuitos digitales el ruido de conmutación afecta tanto a las transiciones que se producen en nodos de salida, causando errores de retardo,

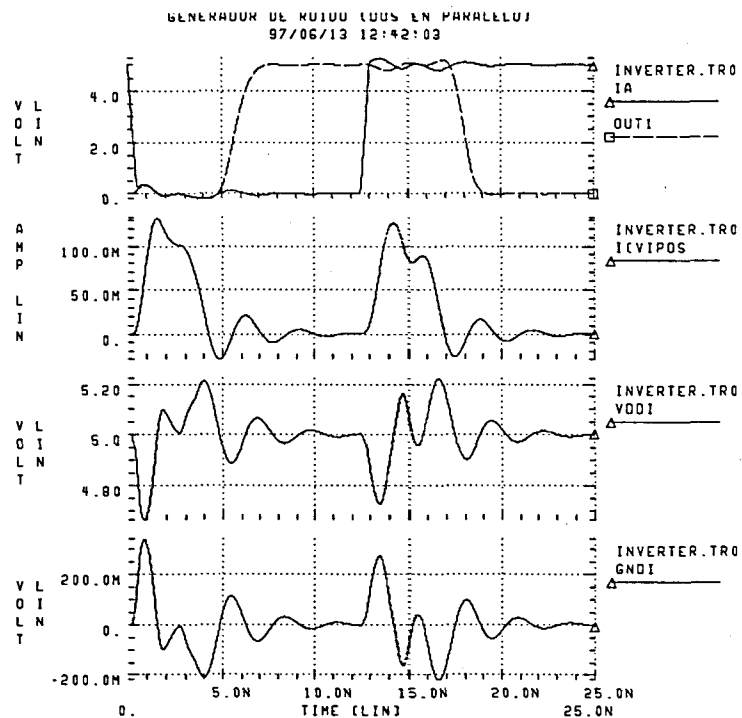
como a las señales estáticas, lo que puede provocar falsas conmutaciones y en el peor de los casos almacenamiento erróneo en biestables (*flip-flops*). Estos efectos son aún más importantes en lógicas dinámicas, como la dominó [29] u otras derivadas de ella, en la que existen nodos flotantes que pueden ser fácilmente descargados por el ruido en las alimentaciones [30]. Vamos a ver en este apartado los mecanismos por los que el ruido de conmutación produce estos efectos en el funcionamiento de los circuitos digitales, centrándonos en la lógica CMOS estática.



**Figura 2-18**

*Circuito generador de ruido de conmutación basado en una cadena de 7 inversores parametrizables.*

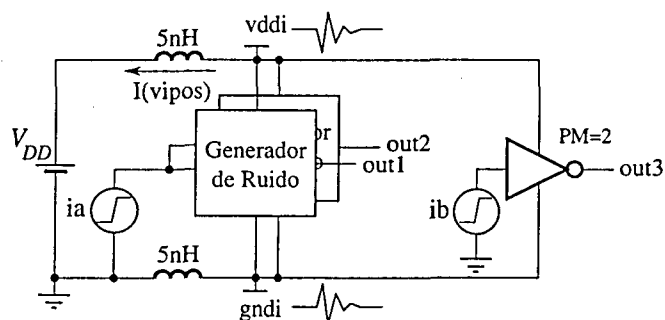
En los tres subapartados siguientes utilizaremos diferentes configuraciones en las que el ruido presente en las alimentaciones afecta a otras partes del circuito. Para generar el ruido de conmutación utilizaremos el circuito de la Figura 2-18, similar al utilizado en [17]. El bloque básico consta de siete inversores en serie parametrizables. El inversor está diseñado en una tecnología de  $1.2 \mu\text{m}$  y permite ajustar a través del parámetro  $PM$  su tamaño relativo. Las capacidades  $C_p$  y  $C_n$  se calculan para añadir las capacidades intrínsecas del PMOS y el NMOS, debidas sobre todo a las difusiones de drenador. La capacidad  $C_L$  se ha calculado para obtener un *fanout* aproximado de 3. Cada uno de los siete inversores de la cadena puede configurarse por separado. En la Figura 2-19 se muestra el ruido generado por uno de estos bloques generadores de ruido con parámetros  $PM1..7 = 22, 20, 20, 20, 1, 1, 1$  y alimentado a través de dos inductancias entre los extremos de la fuente de alimentación y los terminales internos ('vddi' y 'gndi') de  $5 \text{ nH}$  de valor.



**Figura 2-19**

Resultados de simulación para el circuito generador de ruido alimentado a través de dos inductancias de 5nH.

### 2.4.1 Errores de retardo



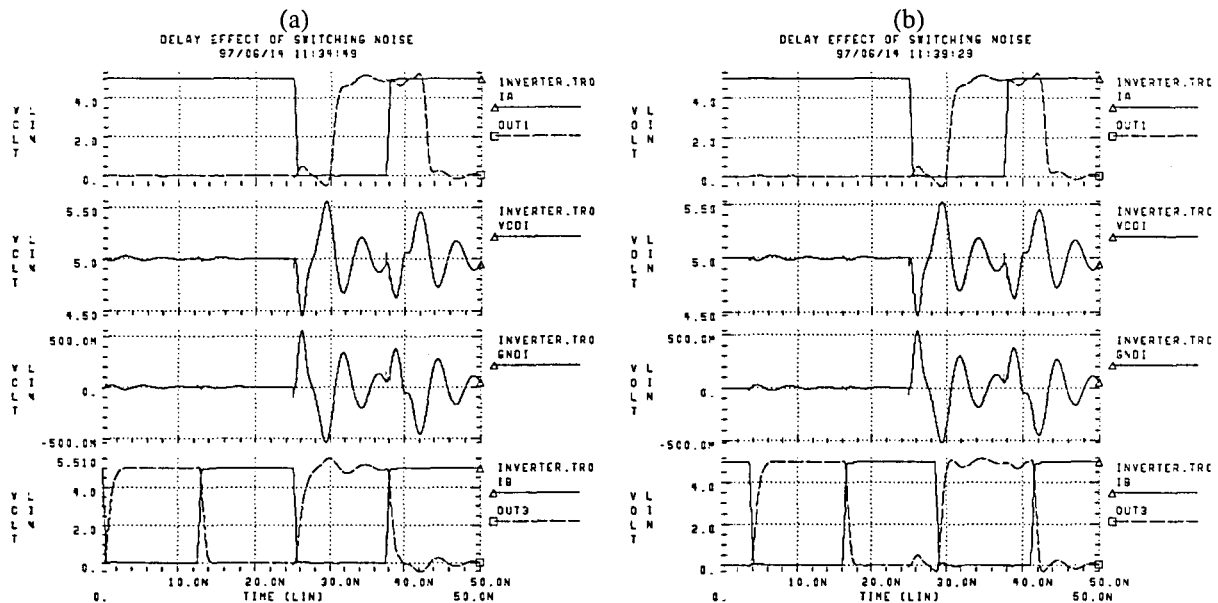
**Figura 2-20**

Circuito utilizado para medir por simulación el efecto del ruido de conmutación en el retardo de las conmutaciones del inversor.

Cuando un circuito lógico está realizando una conmutación y la tensión de alimentación varía, la evolución de esta transición se ve afectada. Si mientras se realiza un transición de alto a bajo o viceversa la tensión de alimentación positiva disminuye y/o la negativa (Gnd) aumenta la transición se realiza más lentamente de lo normal pues los transistores al disminuir su tensión entre puerta y surtidor disminuyen la corriente. En este caso el retardo aumenta. Si por el contrario mientras se realiza la transición la tensión positiva aumenta y/o la negativa (Gnd) disminuye dado que aumenta la tensión puerta surtidor de los dispositivos MOS, la corriente

aumenta y la conmutación se realiza más rápidamente de lo normal. Hemos realizado un experimento para comprobar este efecto ilustrado en la Figura 2-20, donde se han utilizado dos bloques generadores de ruido como el de la Figura 2-18 en paralelo y se ha medido el retardo en un inversor con  $PM=2$  para ocurrencias de pulsos negativos en 'vddi' y positivos en 'gndi' y viceversa ('vddi' y 'gndi' son los terminales de alimentación positivo y negativo, respectivamente, del inversor víctima contaminados por el ruido de conmutación). Se han realizado medidas con transiciones alto-bajo y bajo-alto definiendo los tiempos de propagación ( $t_{pHL}$  y  $t_{pLH}$ , respectivamente) como la diferencia temporal entre el cruce de la entrada y la salida por  $V_{DD}/2$ . En la Figura 2-21 se presentan dos resultados de simulación para casos en los que el retardo aumenta (a) y disminuye (b). En la Tabla 2-2 se resumen las medidas completas para todos los casos.

La transición en 'ia' puede realizarse en fase o en contrafase respecto a la de 'ib', lo que da dos posibles casos. Si retrasamos ligeramente 'ib' respecto de 'ia' la transición de la primera coincide con pulsos de ruido positivo en 'vddi' y negativo en 'gndi'. Si no la retrasamos la polaridad de los pulsos es la contraria. Esto supone un total de cuatro casos posibles, para este circuito de test en concreto. Aunque la amplitud de los pulsos de ruido no es muy elevada (500mV aproximadamente) las variaciones en el retardo llegan a ser en algún caso del orden de  $1\text{ns}$ .



**Figura 2-21**

*Resultados de simulación que muestran el efecto del ruido de conmutación en el retardo de un inversor CMOS. La gráfica (a) corresponde al caso en el que la transición se produce cuando tenemos un pico negativo en vddi y uno positivo en gndi, aumentando el retardo. En la gráfica (b) el caso es el complementario y el retardo disminuye.*

ia/ib	ruido en gndi	ruido en vddi	t <sub>PLH</sub> sin ruido	t <sub>PLH</sub> con ruido	Δt <sub>PLH</sub>	t <sub>PHL</sub> sin ruido	t <sub>PHL</sub> con ruido	Δt <sub>PHL</sub>
fase	+	-	6.7nseg	7.18nseg	0.48nseg	5.24nseg	5.64nseg	0.40nseg
fase	-	+	6.7nseg	6.31nseg	-0.39nseg	5.24nseg	4.66nseg	-0.42nseg
contrafase	+	-	6.7nseg	6.92nseg	0.22nseg	5.24nseg	6.25nseg	1.10nseg
contrafase	-	+	6.7nseg	6.37nseg	-0.27nseg	5.24nseg	4.60nseg	-0.64nseg

Tabla 2-2

*Modificaciones del retardo de propagación producidas por el ruido de conmutación presente en las alimentaciones. Los signos '+' y '-' se utilizan para indicar la polaridad del pulso de ruido en el terminal de alimentación correspondiente.*

### 2.4.2 Falsas conmutaciones

Las perturbaciones presentes en las alimentaciones debidas al ruido de conmutación se transmiten fácilmente a las salidas de las puertas lógicas que tienen un valor estático (nivel lógico alto o bajo) de tensión. Una vez el ruido está presente en las salidas puede propagarse hasta otras puertas lógicas y hacerlas conmutar, cuando no deberían de hacerlo, pues las señales de entrada originales están fijas a un nivel lógico. Existen dos procesos: el primero consiste en el acoplamiento del ruido de las alimentaciones a las salidas y el segundo la propagación de este ruido hacia otras puertas lógicas. El primer proceso es sencillo de explicar para tecnologías CMOS. Si la salida está, por ejemplo, fija a valor bajo ('0' lógico) al menos uno de los transistores NMOS conecta el nodo de salida con el nodo interno de tierra (Gnd) contaminado por el ruido. Este transistor se comporta como una resistencia (está en zona óhmica debido a la tensión en su puerta que deberá ser un '1' lógico), por lo que cualquier variación de la tensión en Gnd se transmite al otro extremo, la salida de la puerta [31]. En cuanto al segundo proceso, la transmisión de este pulso de ruido, ha sido objeto de diferentes estudios. Estos trabajos están relacionados con lo que se denomina el margen dinámico de ruido de una determinada lógica [32]. Una puerta lógica es insensible al ruido (es decir, no proporciona un valor erróneo en su salida en presencia de ruido en su entrada) si éste no supera los márgenes de ruido, que en general son distintos para el '0' que para el '1' lógico. Estos márgenes de ruido dependen de la anchura del pulso de ruido de la entrada. Un pulso muy estrecho puede alcanzar una amplitud elevada sin afectar a la puerta lógica, mientras que la misma amplitud provocaría una falsa conmutación en la salida si el pulso supera una determinada anchura. En [25] se presentan los calculos de los márgenes de ruido para lógicas ECL. En [30] se hace lo mismo para lógicas CMOS dinámicas. En estos dos casos los márgenes de ruido dinámicos son reducidos y el ruido acoplado a las salidas puede provocar falsas conmutaciones en otras puertas lógicas muy fácilmente. En [33] se realiza el mismo estudio para receptores CMOS (*drivers* de entrada de los CIs) con niveles lógicos compatibles con lógica TTL. En este caso el margen de ruido para el '0' es reducido dado que la tensión de conmutación para circuitos lógicos TTL está entorno a 1.2–1.6 V. Sin embargo la lógica estática CMOS convencional presenta una gran robustez frente a la propagación de pulsos de ruido, como ha estudiado detalladamente Francesc Moll [34][35]. En este trabajo se ha realizado un estudio por simulación con el circuito de test de la Figura 2-22. Un inversor parametrizable con PM=2 es alimentado con una tensión contaminada

por pulsos de ruido. La salida se conecta a una cadena de inversores alimentados con tensiones sin ruido. Mediante un proceso iterativo de simulación se encuentra para cada anchura de pulso ( $W_p$ ), la amplitud mínima del pulso de ruido ( $A_p$ ) en la alimentación, que provoca que el ruido se propague por la cadena, es decir, que la amplitud del pulso a la salida del primer (o2) y el tercer inversor (o4) de la cadena sea aproximadamente igual ( $ampl1 \approx ampl2$ ). Una simulación concreta para un pulso negativo en la tensión de alimentación positiva con dos anchuras distintas se muestran en la Figura 2-23. Puede observarse en la figura como a pesar de la gran amplitud del pulso (3.5 V) para una anchura de 0.75nseg (esta anchura se define al 50% de la amplitud del pulso, no en la base) el pulso a pesar de que aparece a la salida del inversor ruidoso no se propaga por la cadena hasta que supera una anchura de 3nseg. En la Figura 2-24 se presentan los resultados obtenidos por el proceso de iteración para la amplitud mínima del pulso de ruido que provoca la propagación de la falsa conmutación por la cadena de inversores para un '0' y un '1' a la salida del inversor.

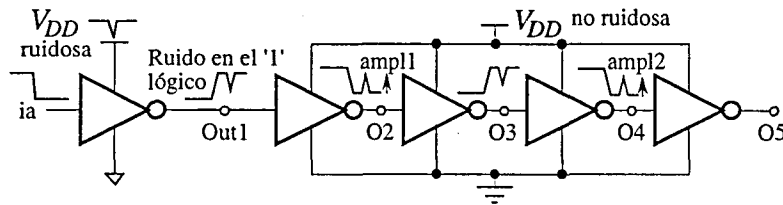


Figura 2-22

Circuito utilizado para simular el efecto de los pulsos de ruido de conmutación presente en la alimentación en las conmutaciones de otras puertas lógicas y la propagación de estas falsas conmutaciones.

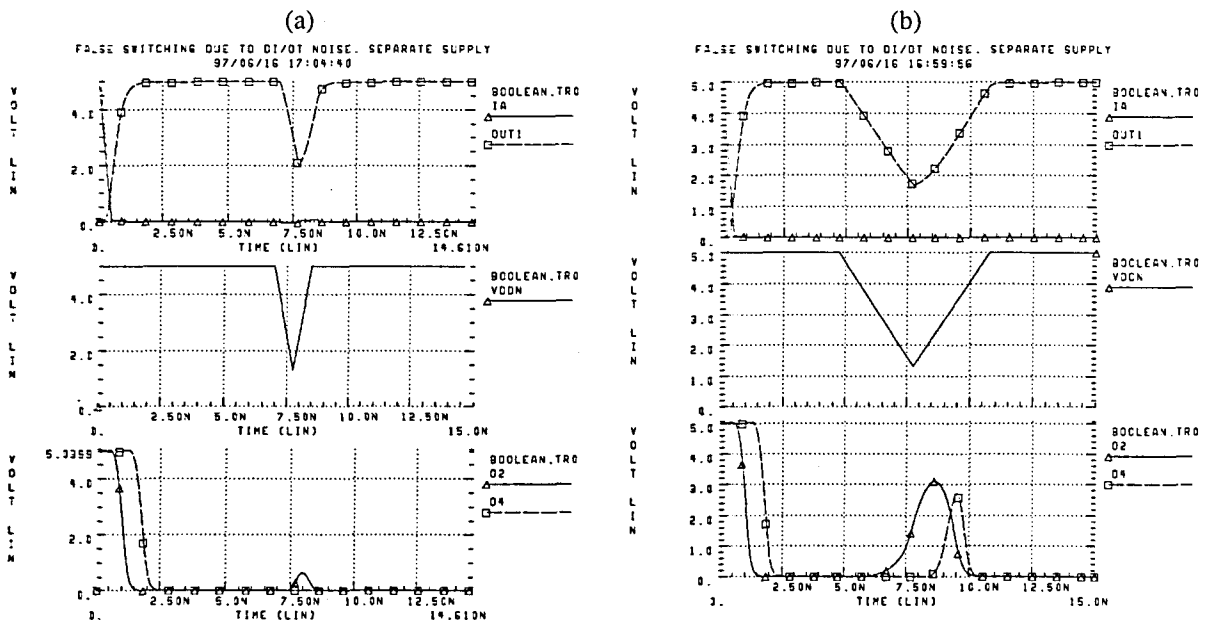
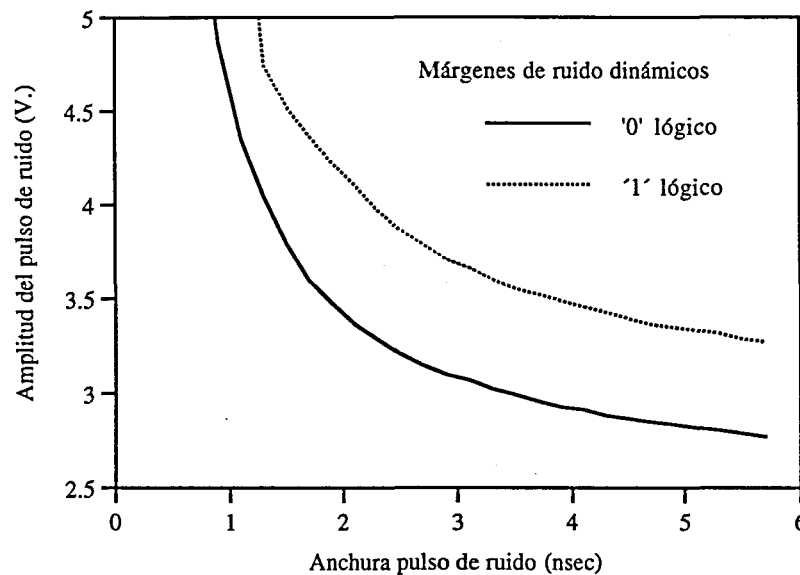


Figura 2-23

Resultados de simulación para dos anchuras de pulso diferentes (a) 0.75n y (b) 3n para la misma amplitud del pulso de ruido sobre  $V_{DD}$  de 3.5V.



**Figura 2-24**

*Márgenes de ruido dinámicos para el '1' y el '0' lógicos, en función de la anchura al 50% del pulso, definidos como la amplitud de pulso de ruido sobre los terminales de alimentación necesaria para que la falsa conmutación producida en una cadena de inversores se propague.*

### 2.4.3 Almacenamientos erróneos en biestables

Uno de los efectos más perjudiciales del ruido de conmutación en circuitos digitales es el almacenamiento de estados erróneos en biestables [36]. Los dos tipos de efectos vistos anteriormente afectan a los biestables: la variación en el retardo puede causar violaciones de los tiempos de *hold* y de *set-up*<sup>7</sup> [37] y las falsas conmutaciones en la señal de reloj o de datos del biestable, si el transitorio coincide con el flanco del reloj, puede provocar falsos almacenamientos. Pero también el ruido presente en las alimentaciones puede provocar los mismos errores de almacenamiento. En lógicas dinámicas CMOS este problema es especialmente grave pues basta un pulso de ruido en las alimentaciones del nivel de la tensión umbral de los MOSFETs para descargar los nodos dinámicos internos [30][38].

### 2.4.4 Sensibilidad al ruido de conmutación de los circuitos lógicos BiCMOS

Las puertas lógicas BiCMOS son menos sensibles al ruido de conmutación presente en las alimentaciones que las CMOS debido a las uniones entre base y emisor y base y colector de los BJT de *pull-up* y *pull-down*, que cuando están en inversa, bloquean la transmisión del ruido de las alimentaciones a otras partes de la puerta. Además los transistores bipolares son unidireccionales lo que reduce el número de casos en los que el ruido puede propagarse a través

<sup>7</sup> Estos dos parámetros se derivan de las restricciones temporales debidas a los tiempos de transición finitos del reloj de los biestables. El tiempo de *set-up* es el tiempo mínimo que debe separar una transición de la entrada de datos de la entrada de reloj para que ésta quede memorizada correctamente. De igual forma el tiempo de *hold* es el tiempo que debe permanecer la señal de datos estática desde el flanco de reloj para que ésta sea correctamente reconocida por el biestable.

de ellos. El ruido que se acopla desde la alimentación de una puerta BiCMOS que no está conmutado a su salida depende del estado de la puerta y de si este ruido está presente en el terminal de alimentación positivo o en el de Gnd ya que la estructura de las puertas no es simétrica. El peor caso se da cuando la salida está a nivel bajo y tenemos un pulso positivo en el nodo de Gnd. Como el transistor de *pull-down* está en conducción la salida sigue al nodo de Gnd con una diferencia de tensión de la caída de un diodo en directa [39]. Por tanto, además de ser menos sensibles al ruido generan pulsos en las salidas de menor amplitud, por lo que se reduce el riesgo de falsas conmutaciones booleanas y otros efectos que ya hemos explicado anteriormente.

## 2.5 Efectos del ruido de conmutación en circuitos analógicos

Uno de los fenómenos que más limitan las prestaciones de los sistemas analógicos es el ruido. El ruido de conmutación  $dI/dt$  es la fuente de ruido más intensa en los CI actuales y, como se verá más adelante, en el futuro el problema se agravará. En los CI mixtos que integran en el mismo sustrato partes digitales y partes analógicas el ruido generado en las primeras puede fácilmente afectar a las segundas. Existen varios elementos comunes: las alimentaciones, el sustrato y las señales que comunican ambas partes del circuito. A pesar de que se consiguiera separar totalmente las alimentaciones de la parte digital de la analógica y emplear señales diferenciales para la comunicación entre el mundo digital y el analógico que eliminase el efecto del ruido en las entradas, nos quedaría siempre el sustrato común. Por ello, el mayor esfuerzo en el campo del aislamiento del ruido en CI mixtos se ha realizado en el campo de la propagación del ruido por sustrato [40][41][42].

Los principales efectos del ruido en las partes analógicas de los CI mixtos están relacionados con la degradación de la relación entre la señal y el ruido. En [43], por ejemplo, se presenta el proceso de diseño de un sistema de comunicaciones en el que el nivel de la señal de entrada en la parte analógica es de 70nV y el ruido presente en las alimentaciones digitales medido llega a los 40mV (no se informa del ruido final presente en la parte analógica, pero estos números nos pueden dar una idea de la magnitud del problema). La reducción en la relación señal a ruido (SNR) obliga a trabajar con técnicas más refinadas de procesado que aumentan el coste y el consumo, dos parámetros cruciales en el diseño de equipos de comunicación portátiles. Otros ejemplo de este tipo de restricciones son [44][45]

Dos circuitos analógicos especialmente importantes, los conversores analógico digitales (ADCs) y los amplificadores operacionales (OPAMPs) se ven afectados por el ruido de conmutación presente en CI mixtos en diferentes formas. En [46] se presentan medidas sobre los efectos del ruido en OPAMPs. El ruido produce un aumento de la distorsión en la forma de onda de la salida y errores de fase. Respecto de los ADCs el efecto más importante es una limitación en el nivel mínimo de señal distinguible del ruido, límite que fija la resolución máxima posible del conversor [47][48][49][50]. Por tanto el ruido de conmutación que se transmite hasta los circuitos analógicos dificulta el trabajo con niveles reducidos de tensión, limitando el margen dinámico, distorsiona las señales amplificadas por los OPAMPs e introduce errores de fase que podrían afectar, por ejemplo, a filtros lineales o a circuitos de demodulación. En



resumen, no se produce en general un fallo funcional, como en el caso digital, sino que las prestaciones de la parte analógica de estos CI mixtos se ven limitadas por la presencia del ruido de conmutación.

## 2.6 Conclusiones

En este capítulo hemos recopilado el estado del arte del ruido de conmutación, completando con trabajos y aportaciones propias allí donde ha sido necesario. Nos hemos centrado en el ruido de conmutación presente dentro del encapsulado, *on-chip*, representado éste simplemente por la inductancia equivalente de los terminales de alimentación. Del mismo modo hemos enfocado el análisis en la conmutación de la circuitería lógica interna de los CI. La mayoría de lo dicho es aplicable también a los *drivers* de salida si bien las características específicas del ruido de conmutación para estos circuitos se analizarán en un capítulo aparte.

En primer lugar se ha tratado el mecanismo de la generación de el ruido de conmutación cuyo origen está en las transiciones de los circuitos lógicos y las demandas de corriente que al circular a través de las inductancias presentes en la red de distribución de la alimentación generan ruido inductivo proporcional a la derivada de estos pulsos de corriente. Se ha estudiado el ruido generado por circuitos CMOS y también por circuitos BiCMOS observándose que no existen diferencias importantes entre ambas alternativas por lo que al nivel de ruido generado se refiere. Se han analizado los parámetros que determinan el nivel de ruido de conmutación tanto a nivel circuital (combinación de las entradas, capacidad del nodo de salida, número de puertas que conmutan simultáneamente) como tecnológico (tensión de alimentación, tamaño de los transistores, velocidad de las transiciones de las entradas, ...) Respecto al primer tipo de parámetros, en circuitos reales VLSI se ha asumido que para optimizar el área y los tiempos de transición de las entradas el diseño de las puertas mayoritariamente culpables del ruido de conmutación en el interior de los CI (no nos referimos aquí a los *drivers* de salida de los que nos ocuparemos en un capítulo aparte) se realiza de forma que el ruido de conmutación acaba siendo independiente de la capacidad de carga de la puerta y debido únicamente a la corriente máxima de saturación de los dispositivos que cargan o descargan el nodo de salida. En este sentido se ha hecho un resumen de los diferentes modelos tanto para dispositivos de canal largo como de canal corto aparecidos en la literatura. Se ha indicado también como la dependencia del nivel de ruido generado con el número de puertas lógicas que realizan la conmutación de forma simultánea no es lineal, como originariamente se planteó, si no que existe cierta saturación del incremento del ruido debido a un mecanismo de retroalimentación y que el nivel de ruido se reduce si desplazamos temporalmente unas transiciones de otras. En cuanto al segundo tipo de dependencia del ruido de conmutación generado, la tecnológica, nos ocuparemos en detalle en un capítulo posterior.

En segundo lugar nos hemos ocupado de los mecanismos de transmisión del ruido de conmutación generado por los circuitos digitales pensando tanto en un entorno puramente digital como en un CI mixto con partes digitales y analógicas. La transmisión del ruido se produce por los elementos que comparten las diferentes partes del circuito integrado: las alimentaciones, el substrato y las señales que comunican entre sí diferentes partes como las que intercambian la

parte digital y la analógica. De entre estos medios de transmisión del ruido cabe destacar el substrato, pues los demás pueden diseñarse adecuadamente para minimizar el ruido transmitido (separación de alimentaciones, comunicación diferencial entre partes ruidosas y no ruidosas, ...). Aunque el análisis se ha centrado en el ruido de conmutación  $dI/dt$  mucho de lo dicho es también aplicable a la transmisión del ruido que tenga como origen los acoplamientos capacitivos o el *crosstalk* entre líneas y que es proporcional a la  $dV/dt$ .

Se ha estudiado a continuación los efectos que produce el ruido de conmutación acoplado de diferentes formas en circuitos digitales y analógicos. Se ha demostrado como la lógica CMOS tradicional presenta una mejor inmunidad en comparación con otras alternativas como las lógicas dinámicas, las ECL o las basadas en niveles TTL. En los circuitos digitales el ruido de conmutación produce errores de funcionamiento como el aumento del retardo, las falsas conmutaciones o los almacenamientos erróneos en biestables. Esto obliga a tener en cuenta en el proceso de diseño técnicas que permitan detectar estos fallos en lo que podríamos llamar estrategias de test para errores debidos al ruido de conmutación [51][52]. En el campo analógico los efectos son muy importantes llegando a ser el ruido una de las principales causas que limitan las prestaciones de elementos clave como los conversores ADC o los amplificadores operacionales.

Hemos dejado otros temas relacionados con la generación del ruido de conmutación como el encapsulado y los *drivers* de salida para capítulos específicos que veremos más adelante debido a que requieren un análisis más detallado.

## Referencias

- [1] Arun Vaidyanath, Birgir Thoroddsen, and J.L. Prince, 'Effect of CMOS Driver Loading Conditions on Simultaneous Switching Noise', *IEEE Tr. on Components, Packagind, and Manufacturing Technology-Part B*, Vol. 17, No. 4, pp. 450–485, 1994.
- [2] Yi-Shing Chang, Sandeep K. Gupta, and Melvin A. Breuer, ' Analysis of Ground Bounce in Deep Sub-Micron Circuits', private communication.
- [3] Prasad Raje, 'Digital Design' in *BiCMOS Technology and Applications*, A.R. Álvarez (ed), Kluwer Academic Publishers, Massachussets, 1993.
- [4] M.I. Elmasry (ed), *BiCMOS Integrated Circuit Design*, IEEE Press, Piscataway, 1994.
- [5] P.Raje, K. Saraswat, and K. Cham, 'Accurate Delay Models of Digital BiCMOS', *IEEE Trans. Electron Devices*, Vol. 39, No. 6, 1992.
- [6] Vicente Jiménez, and Ramón Alcubilla, 'An Accurate Transient Model for BiCMOS Gate Optimization', *Int. J. Electronics*, Vol. 82, No. 5, pp. 503–514, 1997.
- [7] Shayan Zhang, T.S. Kalkur, 'Analysis of BiCMOS Buffer for Input Voltages with Finite Rise Time', *IEEE J. of Solid-State Circuits*, Vol. 29, No. 7, pp. 797–807, 1994.
- [8] 0.8  $\mu\text{m}$  and 1.2  $\mu\text{m}$  BiCMOS Porcess Parameters. Austria Mikro Systeme International. Schloß Premstätten, A-8141 Unterpremstätten, Austria.
- [9] Wen Fang, Arthur Brunnschwiler, and Peter Ashburn, 'An Accurate Analytical BiCMOS Delay Expression and its Application to Optimizing High-Seed BiCMOS Circuits', *IEEE J. of Solid-State Circuits*, Vol. 27, No. 2, pp. 191–202, 1992.

- [10] Prasad A. Raje, Krishna C. Saraswat, and Kit M. Cham, 'A New Method for Design of BiCMOS Gates and Comparison with CMOS', *IEEE Tran. Electron Devices*, Vol. 39, No. 2, pp. 339–347, 1992.
- [11] R. Senthinathan, and J. L. Prince, 'Simultaneous Switching Ground Noise Calculation for Packaged CMOS Devices', *IEEE J. of Solid-State Circuits*, Vol. 26, No. 11, pp. 1724–1728, November 1991.
- [12] Yaochao Yang, and John R. Brews, 'Design Trade-Offs for the Last Stage of an Unregulated, Long-Channel CMOS Off-Chip Driver with Simultaneous Switching Noise and Switching Time Considerations', *IEEE Tran. on Components, Packaging, and Manufacturing Technology—Part B*, Vol. 19, No. 3, pp. 481–486, August 1996.
- [13] Srinivasa R. Vemuru, 'Accurate Simultaneous Switching Noise Estimation Including Velocity-Saturation Effects', *IEEE Transactions on Computers, Packaging, and Manufacturing Technlonogy—Part B*, Vol. 19, No. 2, pp. 344–349, May 1996.
- [14] Takayasu Sakurai, and A. Richard Newton, 'Alpha-Power Las MOSFET Model and its Applications to CMOS Inverter Delay and Other Formulas', *IEEE J. of Solid-State Circuits*, Vol. 25, No. 2, pp. 584–593, April 1990.
- [15] Yaochao Yang, and John R. Brews, 'Design for Velocity Saturated, Short-Channel CMOS Drivers with Simultaneous Switching Noise and Switching Time Considerations', *IEEE J. of Solid-State Circuits*, Vol. 31, No. 9, pp. 1357–1360, September 1996.
- [16] E.E. Davidson, 'Electrical Design of a High Speed Computer Package', *IBM J. of Research and Developement*, Vol. 26, No. 3, pp. 349–361, 1982.
- [17] Nishath K. Verghese, Timothy J. Schmerbeck, David J. Allstot, 'Chapter 8: Modeling Chip/Package Power Distribution' in 'Simulation Techniques and Solutions for Mixed-Signal Coupling in Integrated Circuits', Kluwer Academic Publishers, Boston, 1995, pp. 149–182.
- [18] Perng-Shyong Lin, and A. Zukowski, 'Analysis and Control of Timing Jitter in Digital Logic Arising from Noise Voltage Sources', *IEEE Int. Conf. Computer Design*, pp. 352–356, 1993.
- [19] D.K. Su, M.J. Loinaz, S. Masui, B.A. Wooley, 'Experimental Results and Modeling Techniques for Substrate Noise in Mixed-Signal Integrated Circuits', *IEEE J. Solid-State Circuits*, vol. 28, no.4, pp. 420-430, April 1993.
- [20] N.K. Verghese, D.J. Allstot, S. Masui, 'Rapid Simulation of Substrate Coupling Effects in Mixed-Mode ICs', *Proc. IEEE Custom Integrated Circuits Conference*, pp. 18.3.1-18.3.4, 1993.
- [21] B.R. Stanistic, N.K. Verghese, R.A. Rutenbar, L.R. Carley, D.J. Allstot, 'Addressing Substrate Coupling in Mixed-Mode IC's: Simulation and Power Distribution Synthesis' *IEEE J. Solid-State Circuits*, vol. 29, no. 3, pp. 226-237, March 1994.
- [22] K. Joardar, 'A Simple Approach to Modeling Cross-Talk in Integrated Circuits', *IEEE J. Solid-State Circuits*, vol. 29, no. 10, pp. 1212-1219, October 1994.
- [23] N.K. Verghese, D.J. Allstot, M.A. Wolfe, 'Verification Techniques for Substrate Coupling and Their Application to Mixed-Signal IC Design', *IEEE J. Solid-State Circuits*, vol. 31, no. 3, pp. 354-365, March 1996.
- [24] Xavier Aragones, 'Caracterització de l'acoblament per substrat' en *Contribució a l'Estudi de l'Acoblament per Substrat en Circuits Integrats Mixtes*, Tesis Doctoral, Dept. d'Enginyeria Electrònica, Univ. Politècnica de Catalunya, 1997.
- [25] George A. Katopis, 'Delta-I Noise Specification for a High-Performance Computing Machine', *Proc. of the IEEE*, Vol. 73, No. 9, pp. 1405–1415, September 1985.

- [26] Michael Dolle, 'Analysis of Simultaneous Switching Noise', *IEEE Symp. Circuits and Systems*, pp. 904–907, 1995.
- [27] A. J. Rainal, 'Eliminating Inductive Noise of External Chip Interconnections', *IEEE J. Solid-State Circuits*, Vol. 29, No. 2, February 1994, pp. 126–129.
- [28] E.E. Davidson, P.W. Hardin, G.A. Katopis, M.G. Nealon, L.L. Wu, 'Physical and Electrical Design Features of the IBM Enterprise System/9000 Circuit Module', *IBM J. of Research and Development*, Vol. 36, No. 5, pp. 877–888, September 1992.
- [29] R.H. Krambeck, Charles M. Lee, and Huang-Fai Stephen Law, 'High-Speed Compact Circuits with CMOS', *IEEE J. of Solid-State Circuits* Vol. SC-17, No. 3, pp. 614–619, June 1982.
- [30] Patrik Larsson, and Christer Svensson, 'Noise in Digital Dynamic CMOS Circuits', *IEEE J. of Solid-State Circuits*, Vol. 29, No. 6, pp. 655–662, June 1994.
- [31] Mary Sue Haydt, Robert Owens, and Samiha Mourad, 'Modeling the Effect of Ground Bounce on Noise Margin', *Proc. International Test Conference 1994*, pp. 279–285.
- [32] Jacek M. Zurada, Yu Sang Joo, Samuel V. Bell, 'Dynamic Noise Margins of MOS Logic Gates', *Proc. of ISCAS' 89*, pp. 1153–1156.
- [33] R. Senthinathan, J.L. Prince and S. Nimmagadda, 'Noise Immunity Characteristics of CMOS Receivers and Effects of Skewing/Damping CMOS Output Driver Switching Waveform on the Simultaneous Switching Noise', *Microelectronics Journal*, Vol. 23, 1992, p. 29–36.
- [34] Francesc Moll, and Antonio Rubio, 'Analysis of Spurious Signal Propagation in Digital CMOS Circuits', *Int. J. Electronics*, Vol. 73, No. 3, pp. 621–625, 1992.
- [35] Francesc Moll, and Antonio Rubio, 'Spurious Signals in Digital CMOS VLSI Circuits: A Propagation Analysis', *IEEE Tran. on Circuits and Systems-II*, Vol. 39, No. 10, pp. 749–752, October 1992.
- [36] Francesc Moll Echeto, 'Parasitic Effects Due to Interconnections in Microelectronic Design', Dissertation for the obtention of the Ph. D. degree, Electronic Engineering Department, Universitat Politècnica de Catalunya, 1995.
- [37] S. H. Unger, and C.-J. Tan, 'Clocking Schemes for High-Speed Digital Systems', *IEEE Tran. on Computers*, Vol. C-35, No. 10, October 1986.
- [38] Daniel W. Dobberpuhl, 'A 200-Mhz 64-b Dual-Issue CMOS Microprocessor', *IEEE J. of Solid-State Circuits*, Vol. 27, No. 11, pp. 1555–1564, November 1992.
- [39] David A. Secker, and John L. Prince, 'Effects and Modeling of Simultaneous Switching Noise for BiCMOS Off-Chip Drivers', *IEEE Tran. on Components, Packaging, and Manufacturing Technology—Part B*, Vol. 19, No. 3, pp. 474–480, August 1996.
- [40] Balsha R. Stanasic, Rob A. Rutenbar, and L. Richard Carley, 'Addressing Noise Decoupling in Mixed-Signal IC's: Power Distribution Design and Cell Customization', *IEEE J. of Solid-State Circuits*, Vol. 30, No. 3, pp. 321–326, March 1995.
- [41] Sujoy Mitra, Rob. A. Rutenbar, L. Richard Carley, and David J. Allstot, 'Substrate-Aware Mixed-Signal Macrocell Placement in WRIGHT', *IEEE J. of Solid-State Circuits*, Vol. 30, No. 3, March 1995.
- [42] Martin Pfost, Hans-Martin Rein, and Thomas Holzwarth, 'Modeling Substrate Effects in the Design of High-Speed Si-Bipolar IC's', *IEEE J. of Solid-State Circuits*, Vol. 31, No. 10, pp. 1493–1501, October 1996.
- [43] Patrik J. Zabinski, Mark E. Cickberg, Barry K. Gilbert, Philip J. Zucarelli, Darren V. Weninger, Theodore W. Keller, and David M. Yee, 'Case Study: The Design of a Mixed-Signal Global Positioning System Receiver Using Multichip Module Packaging',

- IEEE Tran. on Components, Packaging, and Manufacturing Technology—Part B, Vol. 19, No. 1, pp. 215–224, February 1996.
- [44] A BiCMOS Time Interval Digitizer Based on Fully-Differential, Current-Steering Circuits', *IEEE J. of Solid-State Circuits*, Vol. 29, No. 6, June 1994, pp. 707–714.
- [45] K.J. Schultz, R.G. Gibbins, J.S. Fujimoto, R.S. Philips, G.F.R. Gibson, A.L. Silburt, 'Low-Supply-Noise Low-Power Embedded Modular SRAM', *IEE Proc.-Circuits, Devices and Systems*, Vol. 143, No. 2, April 1996, pp. 73–82.
- [46] Johan Catrysse, 'Measured Distortion of the Output-Waveform of an Integrated OPAMP Due to Substrate Noise', *IEEE Tran. on Electromagnetic Compatibility*, Vol. 37, No. 2, pp. 310–312, May 1995.
- [47] B.M.J. Kup, E.C. Dijksmans, P.J.A. Naus, J. Sneep, 'A Bit Stream Digital-to-Analog Converter with 18-b Resolution', *IEEE J. of Solid-State Circuits*, vol. 26, no. 12, pp. 1757–1763, December 1991
- [48] Brian P. Brandt, and Bruce A. Wooley, 'A 50-MHz Multibit Sigma-Delta Modulator for 12-b 2-MHz a/D Conversion', *IEEE J. of Solid-State Circuits*, Vol. 26, No. 12, pp. 1746–1756, December 1991.
- [49] J.W. Fattaruso, S. Kiriaki, M. de Wit, C. Warmar, 'Self-Calibrating Techniques for a Second-Order Multibit Sigma-Delta Modulator', *IEEE J. of Solid-State Circuits*, Vol. 28, No. 12, pp. 1216–1223, Decembre 1993.
- [50] Tallis Blalack, and Bruce A. Wooley, 'The Effects of Switching Noise on an Oversampling A/D Converter', *1995 IEEE International Solid-State Circuits Conference*, pp. 200–201, 367.
- [51] Alan Rogers, 'Reliability Issues Due to Switching Currents in CMOS ASICs', *Proc. NORTHCON '89 Conference*, pp. 304–309.
- [52] Melvin A. Breuer, and sandeep K. Gupta, 'Process Aggravated Noise (PAN): New Validation and Test Problems', *Proc. 1996 International Test Conference*, pp. 914–923.



# Capítulo 3

---

## *Estrategias de diseño digital y ruido de conmutación*

*En este capítulo se repasan las diferentes estrategias de diseño digital desde las descripciones utilizando lenguajes de alto nivel y la síntesis automática a la elección del tipo de lógica (dinámica, estática) pasando por las diferentes arquitecturas (paralelización, pipeline, ...). En cada caso se analiza las diferentes alternativas y se estudia la mejor opción de cara a la minimización del ruido de conmutación generado por el circuito digital.*

### 3.1 Introducción

Como se ha visto en los capítulos anteriores el origen del ruido de conmutación está en la actividad de la circuitería digital de los CIs. El proceso de diseño de un circuito digital comienza con una especificación abstracta de la funcionalidad (comportamiento) del circuito. Hoy en día existen lenguajes de alto nivel como el VHDL [1] o el Verilog [2] que permiten especificar formalmente los circuitos digitales en esta fase del proceso de diseño. Las herramientas CAD de diseño de CIs son capaces de generar la estructura final de los circuitos digitales descritos a nivel abstracto. El proceso de síntesis a alto nivel de circuitos digitales puede dar resultados distintos para una misma especificación dependiendo de los criterios utilizados, que generalmente son o bien minimizar el área o bien maximizar la velocidad. Las herramientas de síntesis utilizan una biblioteca básica de bloques digitales para construir los circuitos digitales. Las libertades del diseñador una vez proporcionada la especificación funcional o de comportamiento quedan limitadas a los parámetros de control del proceso de síntesis. En un primer apartado se analiza los resultados obtenidos con los dos criterios de síntesis (área o velocidad) respecto del ruido de conmutación generado por los circuitos digitales obtenidos a partir de una misma especificación.

Si el diseñador requiere un mayor control sobre la estructura final del circuito digital puede realizarse una especificación también utilizando los lenguajes de alto nivel que contenga

descripciones estructurales de partes del circuito o bien del circuito completo. En este caso el resultado del proceso de síntesis es más predecible. En un segundo apartado se analizarán las diferentes alternativas estructurales que nos podemos encontrar en un circuito digital genérico y se estudia comparativamente el ruido de conmutación generado por cada una de ellas.

En una fase posterior, el diseñador tiene la libertad de implementar los diferentes bloques de la estructura mediante varias alternativas lógicas. Incluimos en esta fase de diseño las alternativas de sincronización de los circuitos digitales: sistemas puramente combinacionales, sistemas síncronos con una o varias fases de reloj, sistemas asíncronos o autotemporizados, etc. Todas estas alternativas tienen diferentes repercusiones en el ruido de conmutación generado que se estudian en otro apartado de este capítulo.

Finalmente el esquema eléctrico o lógico del circuito digital debe plasmarse en una serie de máscaras que determinan la estructura física final sobre un chip de silicio del circuito digital microelectrónico. Esta fase se realiza también de forma automática mediante herramientas CAD de diseño que sintetizan el *layout* (el conjunto de máscaras del circuito) a partir del esquemático utilizando compiladores de silicio o técnicas de diseño con celdas estándar. En todo caso el diseñador ha de decidir la tecnología que utilizará de entre las disponibles proporcionadas por los fabricantes. Esta última decisión afecta también al ruido de conmutación, y será tratada en un capítulo aparte en el que se analizará la repercusión en el ruido de conmutación de los parámetros tecnológicos y su evolución para CIs CMOS.

Este capítulo se cierra con un breve apartado en el que se analizan las técnicas de diseño utilizadas para minimizar el consumo de potencia, que se basan en la minimización del número total de transiciones y en la selección de alternativas lógicas y estructurales que minimicen la disipación de potencia, y se estudia su conveniencia para la reducción del ruido de conmutación, ya que la reducción de la actividad de los circuitos digitales es una de las mejores técnicas de minimizar el ruido producido por la conmutación simultánea de varias puertas lógicas.

## 3.2 Diseño y síntesis a alto nivel y ruido de conmutación

A pesar de que el ruido de conmutación está directamente relacionado con la estructura del circuito, ciertos aspectos de las especificaciones a alto nivel de los circuitos digitales influyen en el nivel de ruido generado por la implementación final. A este nivel hablar de ruido de conmutación se reduce a hablar de transiciones simultáneas. Partiendo de la base de que los pulsos de corriente demandados en las conmutaciones por las celdas básicas de la librería de diseño no difieren mucho entre sí, el factor determinante es el número máximo de estos elementos básicos de diseño que pueden llegar a conmutar a la vez en el circuito. El nivel de ruido de conmutación generado por un circuito digital está relacionado con el máximo de actividad del circuito. Sin embargo hay que resaltar que lo importante es la actividad instantánea o, dicho de otra forma, simultánea<sup>1</sup> independientemente que la actividad promedio sea alta o

---

<sup>1</sup> El concepto de simultaneidad debe completarse con las consideraciones realizadas en el apartado 2.2.1 de esta memoria de tesis sobre la superposición de pulsos de corriente ligeramente desplazados entre si.



baja. Las especificaciones que impliquen un alto grado de paralelismo son susceptibles de producir implementaciones potencialmente más ruidosas. Por lo general en especificaciones completamente comportamentales es difícil estimar el grado de paralelismo y mucho menos el número máximo de elementos de la estructura final del circuito digital susceptibles de conmutar simultáneamente.

Es en el proceso de síntesis donde podemos encontrar las primeras alternativas que lleven a generar circuitos más o menos ruidosos. Por ello hemos realizado un estudio con dos ejemplos de síntesis de sistemas a alto nivel con diferente grado de complejidad que nos permitirán establecer la relación entre el proceso de síntesis y el ruido de conmutación generado por el circuito final. La posibilidad de contar con reglas o indicaciones sobre el ruido de conmutación en estas primeras fases del diseño, aunque sean a nivel cualitativo, será de gran utilidad para los diseñadores.

### 3.2.1 Proceso de síntesis a alto nivel

El proceso de síntesis a alto nivel puede dividirse en tres etapas. El punto de partida es una descripción en un lenguaje de alto nivel (VHDL, Verilog, HardwareC, ...) del circuito. La primera etapa consiste en un análisis sintáctico y en una traducción de la especificación a una representación interna que facilite el posterior proceso del circuito. Normalmente el resultado de esta fase del proceso de síntesis es lo que se denomina un grafo de secuencia.

La segunda etapa, denominada síntesis de la arquitectura, genera la representación estructural de la especificación de partida en base a una serie de restricciones y objetivos como el área, el retardo o la latencia. En esta fase podemos encontrar varios procesos que pueden realizarse de forma consecutiva o bien simultánea. La planificación consiste en una organización temporal de las operaciones que debe realizar el sistema. La asignación de recursos es una organización espacial de los elementos necesarios para realizar esas operaciones. Generalmente un sistema puede dividirse entre recursos operativos y unidades de control. Los recursos operativos realizan las operaciones con los datos y la secuencia de estas operaciones es gestionada por la unidad de control. El proceso de síntesis de la arquitectura requiere dividir el sistema en estos dos tipos de elementos y después sintetizar cada uno de ellos. El resultado de esta segunda etapa es una descripción estructural del sistema.

Finalmente es necesario traducir esa estructura en un circuito lógico implementable mediante una tecnología dada. Este proceso denominado síntesis y optimización lógica consta también de varias subetapas. Primero se realiza una optimización lógica tanto para los recursos operativos como para las unidades de control cuyo objetivo es encontrar la realización óptima mediante puertas y elementos lógicos de la arquitectura sintetizada en la etapa anterior. Posteriormente se ha de encontrar la estructura final de cada uno de los bloques del sistema construida con el conjunto de puertas disponibles en la librería de la tecnología escogida. A este proceso se le denomina mapeo tecnológico.

La relación entre el proceso de síntesis automática y el ruido de conmutación del circuito final puede establecerse en diferentes etapas. Los procesos de planificación y asignación de recursos determinan de forma directa la estructura final del sistema. Los algoritmos utilizados en las herramientas comerciales de síntesis utilizan criterios para minimizar el área, lo que se consigue con una asignación de recursos limitada o maximizar la velocidad, lo que se consigue con una adecuada planificación reduciendo el número de operaciones consecutivas y aumentando el grado de paralelismo. En esta fase del proceso de síntesis podrían introducirse criterios relacionados con el ruido de conmutación. Por ejemplo, dada una planificación que minimice el tiempo de ejecución de una determinada operación, las operaciones concurrentes deberían distribuirse de forma uniforme por todo el intervalo temporal minimizando el número de operaciones realizadas simultáneamente. Del mismo modo la realización de las unidades de control podrían incluir criterios que tendiesen a minimizar el ruido de conmutación, reduciendo el número de bits que cambian en las transiciones entre estados (utilizando codificaciones adyacentes u otras técnicas). Otras estrategias de síntesis utilizadas para reducir los tiempos de operación de los circuitos como el uso de estructuras *pipelines*<sup>2</sup> también tienen influencia en el ruido de conmutación generado como veremos más adelante. La etapa de síntesis y optimización lógica tendrá repercusiones en la reducción del ruido de conmutación en la medida en que reduzca el número de puertas en paralelo para realizar una determinada operación o implementar un bloque dado de la estructura del circuito. Aquí también tendrá repercusiones el objetivo de síntesis. Una realización en la que prime reducir el retardo intentará minimizar el *fanout* de las salidas aumentando el paralelismo y reduciendo la longitud de las cadenas combinatorias. Sin embargo una optimización basada en el área producirá circuitos lógicos con menos puertas, aunque dependiendo del algoritmo, no necesariamente con menos puertas susceptibles de conmutar simultáneamente. Por último la fase de mapeo tecnológico nos permitirá escoger entre librerías de puertas y elementos lógicos diferentes. Aquí también cabe la posibilidad de escoger entre lógicas más o menos ruidosas, como ya veremos también más adelante.

### 3.2.2 Ejemplo 1: Controlador de un semáforo de tráfico

Con este ejemplo sencillo realizaremos todo el proceso de síntesis y simulación lógica y eléctrica, lo que nos permitirá relacionar el ruido de conmutación observado en las implementaciones finales del circuito con las alternativas ofrecidas por el proceso de síntesis.

El circuito digital implementa el sistema de control de un cruce entre dos carreteras mediante semáforos. Está especificado por una máquina de estados finitos. El resultado de la síntesis será un sistema secuencial síncrono con la estructura canónica formada por un sistema combinatorial y un banco de biestables que almacenan el estado del sistema y sincronizan las salidas. El circuito está descrito a nivel de comportamiento en lenguaje VHDL y es uno de los

---

<sup>2</sup> Una estructura *pipeline* consta de varias etapas combinatorias separadas entre sí por elementos de memoria como registros o biestables. Varias unidades de información van avanzando por la *pipeline* y se van procesando progresivamente a la vez que nueva información entra de forma continua.

*benchmark*<sup>3</sup> de síntesis existentes para VHDL desde 1992 [3]. El fichero original ha sido retocado para adecuarlo a la versión del lenguaje compatible con el software Synergy de Cadence [4] que ha sido el utilizado en este capítulo como herramienta de síntesis. La descripción del circuito controlador del tráfico mediante semáforos (TLC) se encuentra en la Figura 3-1. El primer paso realizado ha sido la síntesis mediante Synergy del circuito TLC con dos criterios distintos: minimización del área y maximización de la velocidad. Las dos implementaciones obtenidas tras el proceso de síntesis se han llamado TLCcost y TLCtime, respectivamente. La Tabla 3-1 muestra los principales parámetros de cada uno de las dos soluciones sintetizadas.

```

-----
--
-- Traffic Light Controller (TLC)
--
-- Source: Hardware C version written by David Ku on June 8, 1988 at Stanford
--
-- VHDL Benchmark author Champaka Ramachandran
--                University Of California, Irvine, CA 92717
--                champaka@balboa.eng.uci.edu
--
-- Developed on Aug 11, 1992
--
-- Verification Information:
--
--                Verified      By whom?      Date      Simulator
--
-- Syntax          yes          Champaka Ramachandran  Aug 11, 92  ZYCAD
-- Functionality   yes          Champaka Ramachandran  Aug 11, 92  ZYCAD
-----

entity TLC is
  port (
    Cars : in BIT;
    TimeoutL : in BIT;
    TimeoutS : in BIT;
    StartTimer : out BIT;
    HiWay : out BIT_VECTOR(1 downto 0);
    FarmL : out BIT_VECTOR(1 downto 0);
    state : out BIT_VECTOR(1 downto 0);
    clk: in bit
  );
end TLC;

architecture TLC of TLC is
  signal newstate, state_aux, current_state : BIT_VECTOR(1 downto 0);
  signal newHL, newFL : BIT_VECTOR(1 downto 0 );
  signal newST : BIT;

begin
  -----
  traffic:process(current_state, Cars, TimeoutS, TimeoutL)
  begin
    -- current_state := newstate;

    -- combinational logic to determine nextstate
    case current_state is
      when "00" => newHL <= "10"; newFL <= "11";
                   if (Cars = '1') and (TimeoutL = '1') then
                     newstate <= "10"; newST <= '1';
                   else

```

<sup>3</sup> El concepto de circuito o sistema *benchmark* que podríamos traducir por circuito patrón o de referencia se aplica a ciertos circuitos, a veces sin utilidad concreta, que son de dominio público y han sido aceptados como referencias comunes en el entorno científico e industrial para comparar resultados.

```

        newstate <= "00"; newST <= '0';
    end if;

    when "10" => newHL <= "01"; newFL <= "11";
        if (TimeoutS = '1') then
            newstate <= "01"; newST <= '1';
        else
            newstate <= "11"; newST <= '0';
        end if;

    when "01" => newHL <= "11"; newFL <= "10";
        if (Cars = '0') or (TimeoutL = '1') then
            newstate <="11"; newST <= '1';
        else
            newstate <= "01"; newST <= '0';
        end if;

    when "11" => newHL <= "11"; newFL <= "01";
        if (TimeoutS = '1') then
            newstate <= "00"; newST <= '1';
        else
            newstate <= "11"; newST <= '0';
        end if;

    when others =>
        newstate <= "00";
        newHL <= "00";
        newFL <= "00";
        newST <= '0';

    end case;

    -- wait for 10 ns;

end process traffic ;
-----

current_state <= state_aux;
state <= state_aux;

fsm:process
begin
    wait until (clk'event and clk='1');
    state_aux <= newstate;
    HiWay <= newHL;
    FarmL <= newFL;
    StartTimer <= newST;

end process;

end TLC;

```

**Figura 3-1**

*Descripción en lenguaje VHDL del circuito controlador del tráfico mediante semáforos del ejemplo.*

Parámetro	TLCcost	TLCtime
Número de celdas	22	25
Área ocupada (µm <sup>2</sup> )	8553.60	9375.68
Frec. máxima	223.52MHz	311.95MHz
Retardo camino más largo	4.47nsec	3.21nsec

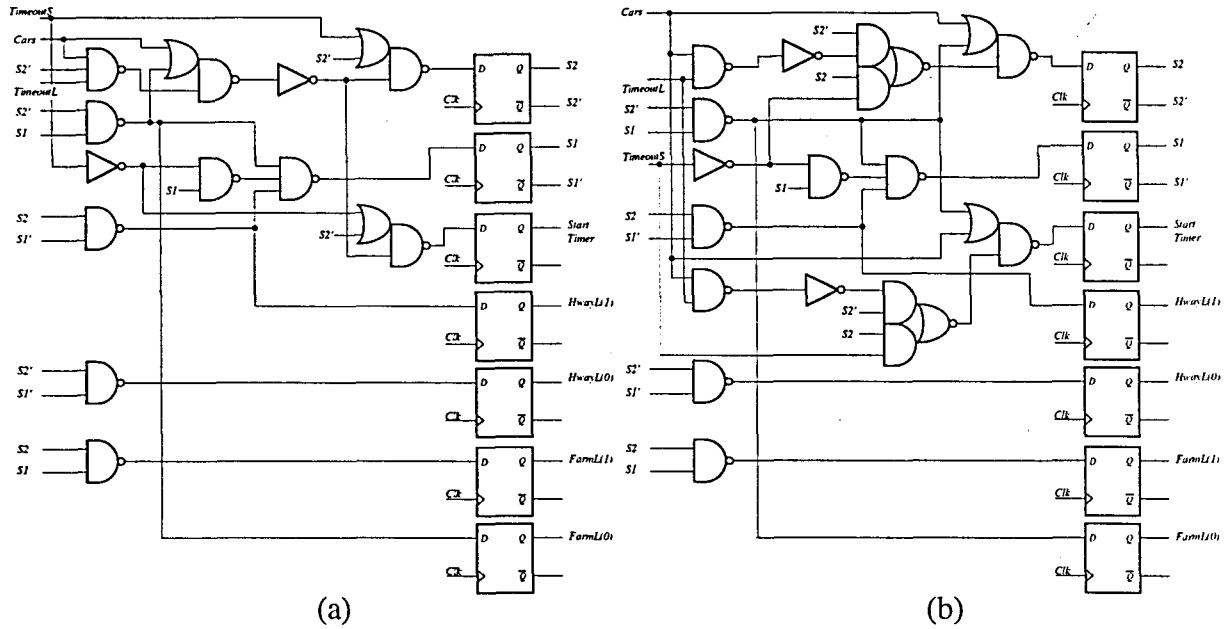
**Tabla 3-1**

*Resultados del proceso de síntesis. TLCcost es la solución obtenida con el criterio de mínima área y TLCtime la obtenida con la optimización de la velocidad.*

La estructura de las dos soluciones es muy similar (ver Figura 3-2). Las diferencias están en la lógica utilizada para generar las entradas de los biestables que almacenan las variables S2 y StartTimer. La implementación de mínimo área tiene menos puertas en paralelo que la de mínimo retardo, como puede apreciarse en la figura. El siguiente paso a realizar ha sido la simulación mediante Verilog [5] de las dos implementaciones obtenidas del circuito TLC. La

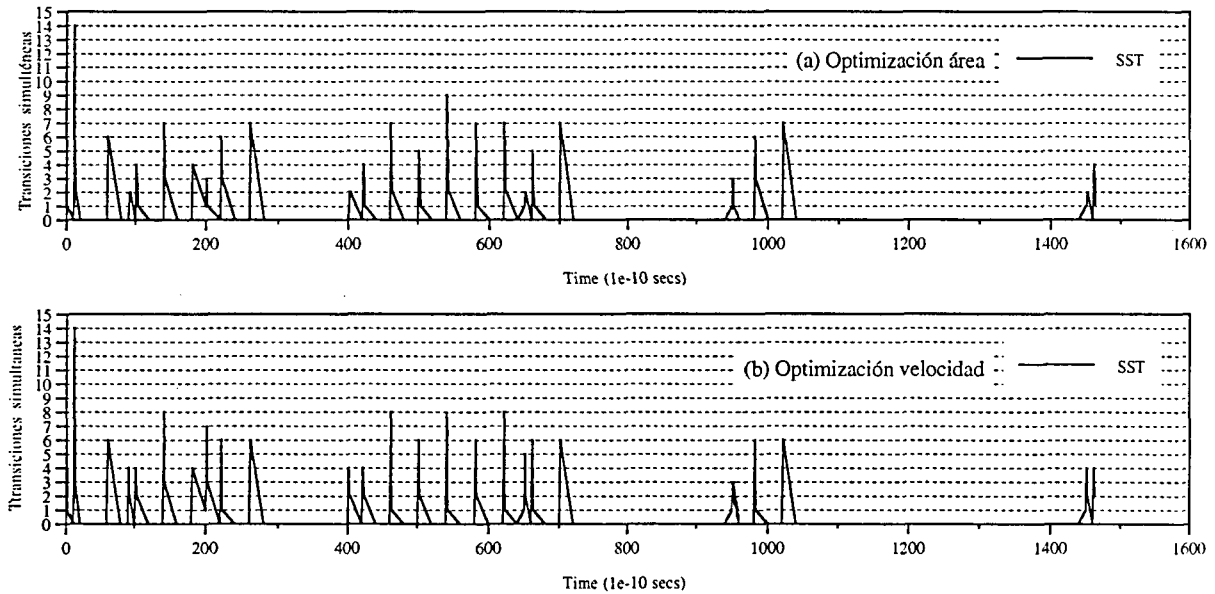
instante de tiempo los valores de todos los nodos internos. Mediante un programa escrito en C recorreremos ese fichero de resultados y vamos contando el número de nodos que cambian de valor de un instante de tiempo al siguiente. De esta forma es posible representar gráficamente el número de transiciones simultáneas que realiza el circuito en cada instante de tiempo. A este nivel lógico no tenemos información sobre la amplitud de los pulsos de corriente de las diferentes celdas de la librería, pero vamos a suponer que son todos equivalentes. El ruido de conmutación generado en un momento dado será considerado directamente proporcional al número de transiciones simultáneas. La Figura 3-3 muestra dos gráficos, cada uno de ellos correspondiente a una de las dos implementaciones TLCcost o TLctime, que presentan los resultados obtenidos mediante este proceso de simulación y proceso de los resultados. Como puede observarse, tras el transitorio inicial, el circuito TLCcost produce en general una actividad menor que el circuito TLctime.

Para comprobar la bondad de la aproximación realizada en la simulación digital se ha generado el esquemático correspondiente a cada implementación y se ha realizado una simulación eléctrica a nivel de transistor mediante el simulador mixto del software de Cadence DFVII llamado Analog Artist. Este simulador permite utilizar partes digitales y analógicas y simularlas conjuntamente. Las primeras mediante Verilog y las segundas mediante un simulador eléctrico muy similar al SPICE llamado Spectre. En nuestro caso los estímulos son digitales y son los mismos utilizados para la simulación anterior y el resto del circuito es analógico. Los modelos para los dispositivos son los proporcionados por el fabricante AMS para su tecnología CMOS de  $0.8\mu\text{m}$ . La Figura 3-4 muestra los resultados de simulación del circuito TLCcost y la Figura 3-5 los del circuito TLctime. Estas simulaciones corresponden a una fracción de la simulación digital anterior, de 0 a 1000 ns. Se han trazado las entradas y alguna de las salidas y la corriente proporcionada por la fuente de alimentación, que es lo que más interesa conocer. Como se puede observar la concordancia de estas curvas con las gráficas de actividad extraídas de una simulación puramente digital es buena y la comparación entre las dos soluciones arroja los mismos resultado. El circuito TLCcost presenta varios picos máximos de aproximadamente 8mA mientras que el circuito TLctime alcanza los 11mA. Esta diferencia es poco significativa debido a que la estructura de ambas soluciones se diferencia en pocas puertas lógicas. En el siguiente subapartado realizaremos el mismo proceso de síntesis y simulación con un circuito más complejo para corroborar la relación entre el criterio de optimización utilizado y el ruido de conmutación generado por las diversas soluciones.



**Figura 3-2**

Circuitos generados a partir de la descripción VHDL del circuito TLC con dos criterios de síntesis: (a) mínima área y (b) mínimo retardo. Los elementos utilizados son celdas estándar de la librería HARDLIB de la tecnología de 0.8µm de AMS.



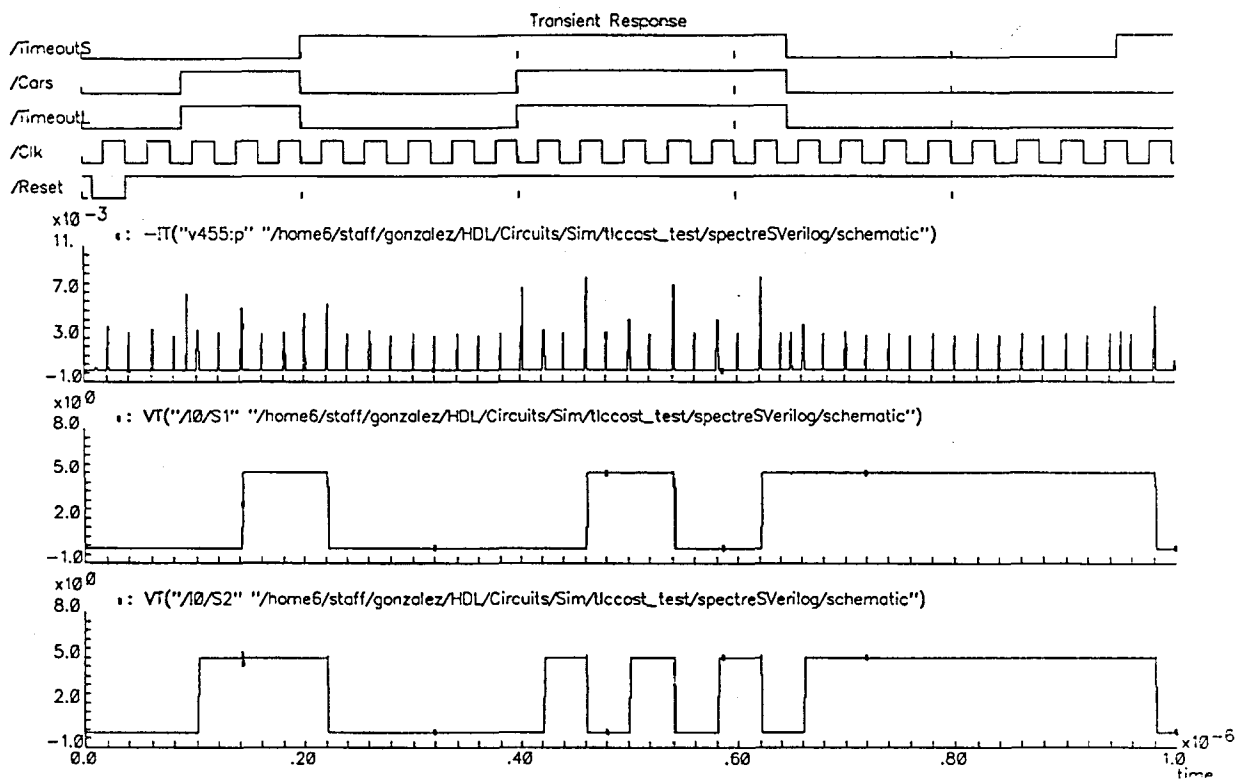
**Figura 3-3**

Representación gráfica del número de transiciones simultáneas extraídas de la simulación mediante Verilog de las dos soluciones sintetizadas del circuito TLC: (a) solución de área mínima, (b) solución de retardo mínimo.

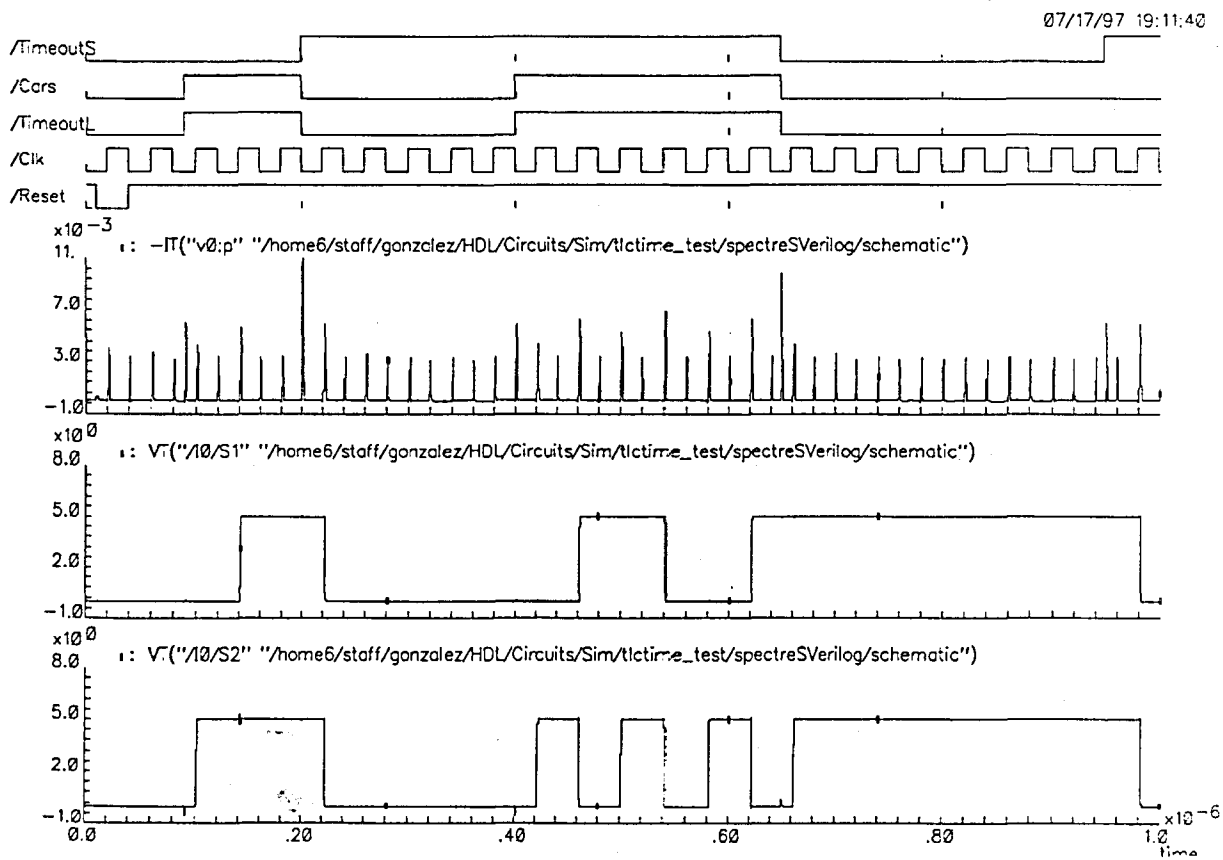
(página siguiente →)

**Figura 3-4 y Figura 3-5**

(a) Simulación eléctrica del circuito TLCcost. (b) Simulación TLCtime Las formas de onda superiores corresponden a los estímulos de entrada. La gráfica etiquetada IT corresponde a la corriente de conmutación del circuito y el resto son alguna de las salidas.



(a)



(b)

### 3.2.3 Ejemplo 2: contador/decontador binario de 8 bits

Con este ejemplo se pretende demostrar como, a pesar de proporcionar una descripción estructural del circuito, el sintetizador produce circuitos finales muy diferentes en base al criterio de optimización (área o retardo) que se utilice. El circuito descrito a alto nivel mediante VHDL consiste en un contador/decontador de 8 bits con reset síncrono activo a nivel bajo y entrada de activación. Mediante una entrada de control llamada *updown* se selecciona la dirección del conteo. El contador se describe estructuralmente interconectando entre si ocho celdas básicas llamadas *tffud* consistentes en un biestable tipo T más una lógica combinacional que calcula la salida para la siguiente celda en función de la dirección de conteo. La Figura 3-6 muestra el listado del contador (llamado *count\_8*) y la Figura 3-7 la celda básica *tffud*.

El proceso seguido ha sido el mismo que en el ejemplo anterior. La descripción del circuito *count\_8* en VHDL se ha sintetizado mediante Synergy con los dos criterios de área (circuito COUNTcost) y retardo (circuito COUNTtime). Ambas implementaciones se han realizado con las celdas de la librería de celdas estándar de la tecnología CMOS de 0.8 $\mu$ m de AMS. El resultado del proceso de síntesis se resume en la Tabla 3-2. Posteriormente se ha simulado mediante Verilog cada una de las implementaciones y se ha obtenido el número de transiciones simultáneas para cada instante. La simulación consta, por este orden, de una inicialización, una activación del contador y al cabo de unos ciclos una cambio de sentido (Figura 3-8).

```

library ieee;
use ieee.std_logic_1164.all;

library work;

entity count_8 is
port(
  t: in std_logic; -- entrada de activacion del contador
  clearn: in std_logic; -- reset del comptador
  clk: in std_logic; -- reloj
  q: out std_logic_vector(7 downto 0); -- salida
  qn: out std_logic_vector(7 downto 0); -- salida negada
  updown: in std_logic; -- control de la direccion
  carry: out std_logic -- fin de conteo
);
end count_8;

architecture estructural of count_8 is

  component tffud -- celda basica del contador
  port(
    t: in std_logic; -- entrada T
    clearn: in std_logic; -- reset activo nivel bajo
    clk: in std_logic; -- reloj
    v: out std_logic; -- para el siguiente tffud
    q: out std_logic; -- salida
    qn: out std_logic; -- salida negada
    updown: in std_logic -- control direccion
  );
  end component;

  signal carry_aux: std_logic_vector(6 downto 0);

begin
  compta_bloc: for i in 7 downto 0 generate
    tff_0: if i = 0 generate
      gen_0: block
        for all: tffud use entity work.tffud(behavioral);
        begin
          tff_biest_0: tffud
            port map
            (

```



```

        t => t,
        clearn => clearn,
        clk => clk,
        v => carry_aux(0),
        q => q(0),
        qn => qn(0),
        updown => updown
    );
    end block;
    end generate;

tff_7: if i = 7 generate
gen_7: block
    for all: tffud use entity work.tffud(behavioral);
    begin
        tff_biest_7: tffud
        port map
        (
            t => carry_aux(6),
            clearn => clearn,
            clk => clk,
            v => carry,
            q => q(7),
            qn => qn(7),
            updown => updown
        );
        end block;
    end generate;

tff_1_6: if i /= 0 and i /= 7 generate
gen_1_6: block
    for all: tffud use entity work.tffud(behavioral);
    begin
        tff_biest_1_6: tffud
        port map
        (
            t => carry_aux(i-1),
            clearn => clearn,
            clk => clk,
            v => carry_aux(i),
            q => q(i),
            qn => qn(i),
            updown => updown
        );
        end block;
    end generate;
end generate;

end estructural;

```

**Figura 3-6**

*Descripción estructural en VHDL del contador/decontador de 8 bits.*

```

library ieee;
use ieee.std_logic_1164.all;

entity tffud is
    port(
        t: in std_logic; -- entrada
        clearn: in std_logic; -- reset activo nivel bajo
        clk: in std_logic; -- reloj
        v: out std_logic; -- para el siguiente tffud
        q: out std_logic; -- salida
        qn: out std_logic; -- salida negada
        updown: in std_logic -- control direccion
    );
end tffud;

architecture behavioral of tffud is
    signal q_aux: std_logic; -- senal auxiliar de salida
    signal o_xor: std_logic; -- conexior xor_and
begin
    t1: process
        -- espera el flanco del reloj y comprueba el reset (sincrono)

```

```

begin
  wait until (clk'event and clk='1');
  if (clearn = '0') then
    q_aux <= '0';
  elsif (t = '1') then
    q_aux <= not q_aux;
  end if;
end process t1;

-- asignacion de las salidas

q <= q_aux;
qn <= not q_aux;

-- generacion de la salida para el siguiente biestable

o_oxor <= q_aux xor updown;
v <= t and o_oxor;

end behavioral;

```

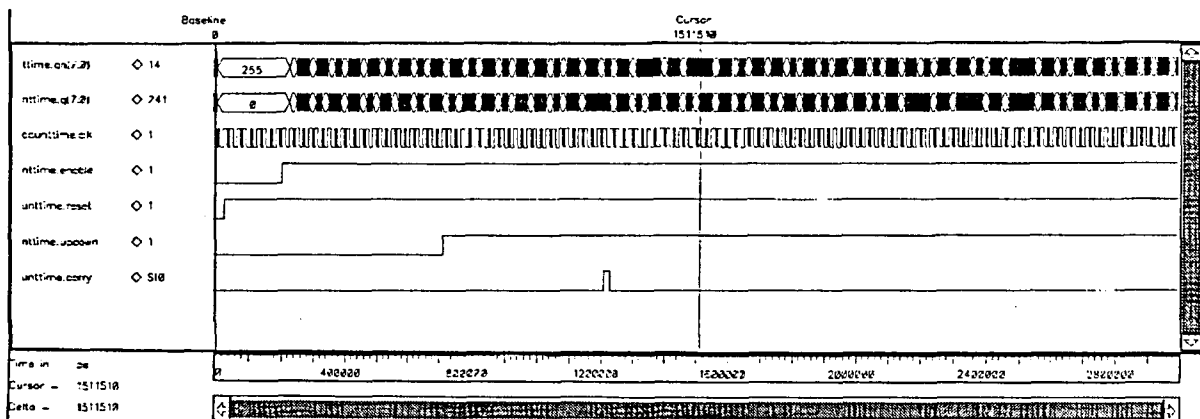
**Figura 3-7**

Descripción comportamental en VHDL de la celda básica (tffud) del contador/decontador.

Parámetro	COUNTcost	COUNTtime
Número de celdas	62	166
Área ocupada (µm <sup>2</sup> )	40601	130125
Frec. máxima	122.257MHz	201.355MHz
Retardo camino más largo	8.179nsec	4.966nsec

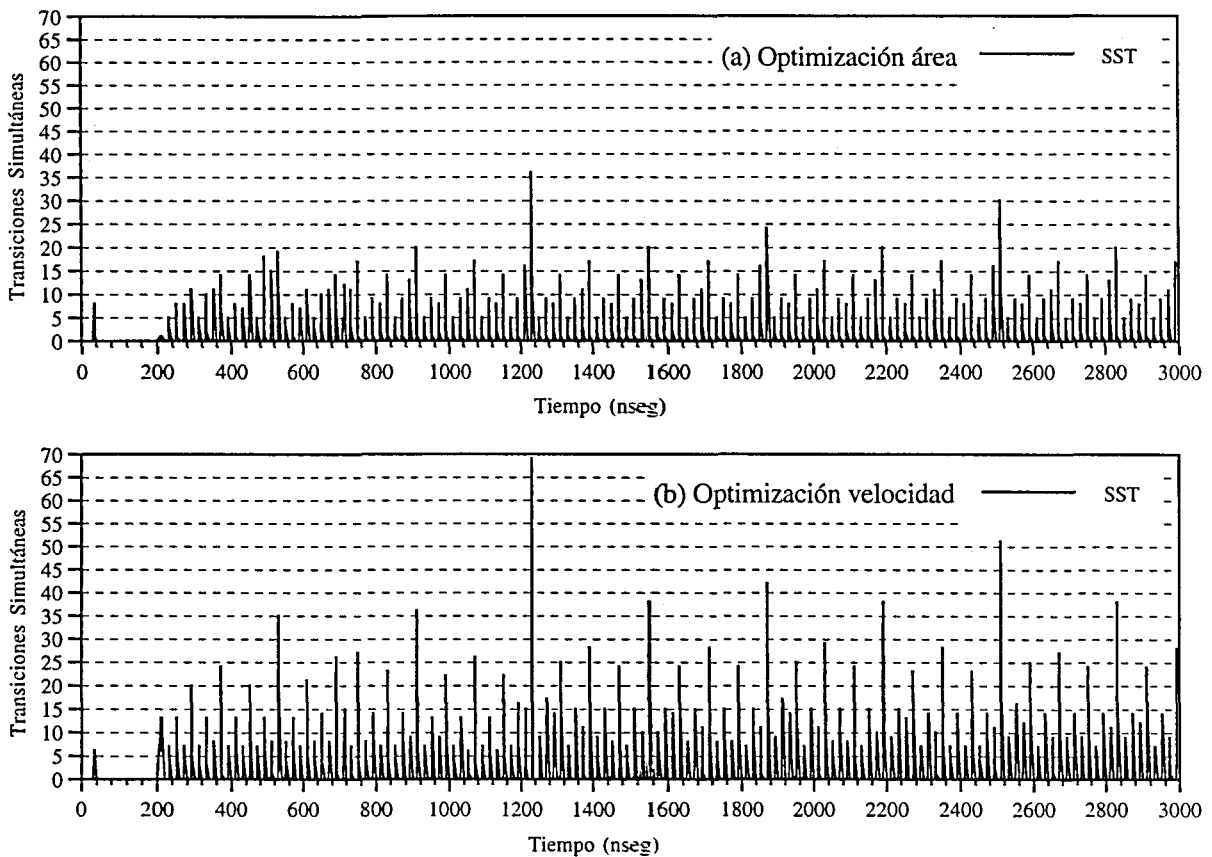
**Tabla 3-2**

Resultados del proceso de síntesis. COUNTcost es la solución obtenida con el criterio de mínima área y COUNTtime la obtenida con la optimización de la velocidad.



**Figura 3-8**

Formas de onda de los estímulos (clk, reset, enable y updown) y de las salidas (q, qn y carry) del contador/decontador correspondientes a una simulación Verilog.



**Figura 3-9**

*Transiciones simultáneas (SST) para las soluciones sintetizadas del circuito count\_8 (a) con optimización de área y (b) con optimización de retardo o velocidad.*

La Figura 3-9 muestra las transiciones simultáneas que se realizan en cada uno de los dos circuitos obtenidos tras el proceso de síntesis. Puede apreciarse mucho más claramente que en el ejemplo anterior como la solución de mínimo área (COUNTcost) presenta en general muchas menos transiciones simultáneas que la solución de mínimo retardo (COUNTtime). El máximo se produce al llegar la cuenta a 255 y volver a 000, un poco más tarde del instante 1200nseg, como puede apreciarse en la simulación digital de la Figura 3-8 cuando la salida *carry* que indica el final de la cuenta vale '1' lógico durante un ciclo de reloj (20nseg).

### 3.2.4 Ejemplo 3: resolución numérica de una ecuación diferencial

Este sistema digital es ampliamente utilizado en toda la bibliografía sobre síntesis. Se trata de un pequeño lazo de cálculo en coma fija que resuelve la siguiente ecuación diferencial [6]:

$$\frac{d^2y}{dx^2} + 3x \frac{dy}{dx} + 3y = 0$$

Se utiliza la variable  $u$  para representar la  $dy/dx$ . La  $dx$  se aproxima por  $x_i - x_{i-1}$ . Del mismo modo  $dy$  se aproxima por  $y_i - y_{i-1}$  y  $du$  por  $u_i - u_{i-1}$ . El valor  $a$  controla el número de veces que el bucle es ejecutado comparando el valor de  $x_i$  con  $a$ . El cálculo que debe realizar el bucle en cada pasada es el siguiente:

$$\begin{aligned}
 x_1 &= x + dx \\
 y_1 &= y + udx \\
 u &= u - 3xudx - 3ydx \\
 x &= x_1 \quad y = y_1
 \end{aligned}$$

Además de estas operaciones se ha de realizar una comparación para determinar si el bucle ha de seguir o detenerse. La Figura 3-10 muestra el listado de la especificación en VHDL de este circuito al que llamaremos DIFFEQ. El fichero es otro de los *benchmarks* de VHDL y también ha sido retocado para que su formato sea sintetizable utilizando Synergy.

```

-----
--
-- Differential Equation Benchmark
--
-- Source: Adapted from example in paper
--         "HAL: A Multi-Paradigm Approach to Automatic Data Path Synthesis"
--         by P. Paulin, J. Knight and E. Girczyc
--         23rd DAC, June 1986, pp. 263-270
--
-- Benchmark author: Joe Lis
--
-- Copyright (c) 1989 by Joe Lis
--
-- Modified by Champaka Ramachandran on Aug 17th 1992
--
-- Verification Information:
--
--           Verified      By whom?           Date           Simulator
--           -----      -
-- Syntax           yes     Champaka Ramachandran  17th Aug 92     ZYCAD
-- Functionality    yes     Champaka Ramachandran  17th Aug 92     ZYCAD
-----
-
library ieee;
use ieee.std_logic_1164.all;
use ieee.std_logic_arith.all;
library work;
use work.all;

entity diffeq is
  port (Xinport: in std_logic_vector(15 downto 0);
        Xoutport: out std_logic_vector(15 downto 0);
        DXport: in std_logic_vector(15 downto 0);
        Aport: in std_logic_vector(15 downto 0);
        Yinport: in std_logic_vector(15 downto 0);
        Youtport: out std_logic_vector(15 downto 0);
        Uinport: in std_logic_vector(15 downto 0);
        Uoutport: out std_logic_vector(15 downto 0);
        clk: in std_logic);
end diffeq;

--VSS: design_style BEHAVIORAL
architecture diffeq of diffeq is

begin

P1 : process (Aport, DXport, Xinport, Yinport, Uinport,clk)

  variable x_var,y_var,a_var,dx_var,u_var,x1,y1,t1,t2,t3,t4,t5,t6,t7:
std_logic_vector(15 downto 0);

begin

if (clk'event and clk='1') then
  x_var := Xinport;
  a_var := Aport;
  dx_var:= DXport;
  y_var := Yinport;
  u_var := Uinport;

  if (a_var > x1) then
    x1 := x_var + dx_var;
    t1 := u_var * dx_var;
    y1 := y_var + t1;
    t2 := "0000000000000011" * x_var;
    t3 := "0000000000000011" * y_var;

```

```

    t4 := t1 * t2;
    t5 := dx_var * t3;
    t6 := u_var - t4;
    u_var := t6 - t5;
    x_var := x1;
    y_var := y1;
end if;
end if;

Xoutport <= x_var;
Youtport <= y_var;
Uoutport <= u_var;

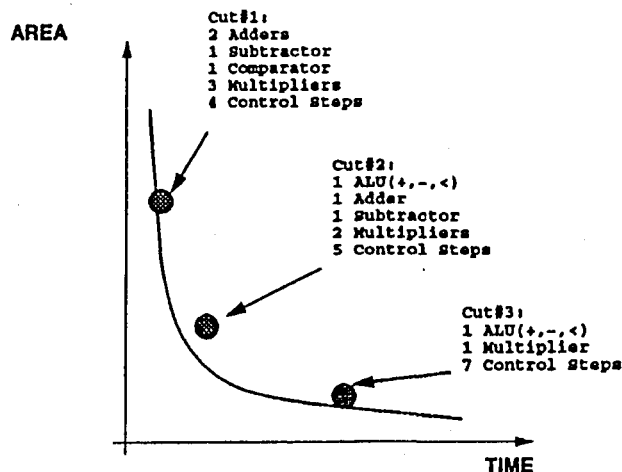
end process P1;
end diffeq;

```

**Figura 3-10**

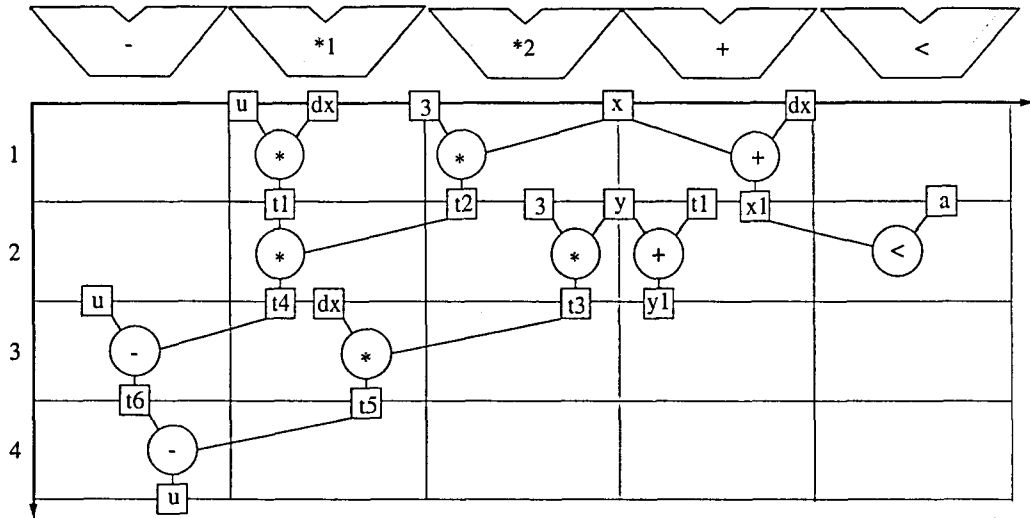
*Descripción a nivel de comportamiento del circuito DIFFEQ escrita en VHDL.*

El proceso de síntesis implica siempre encontrar un compromiso entre el área y el retardo de los circuitos sintetizados. Este compromiso se suele representar en forma de curva donde se indican los diferentes puntos sintetizables (como las soluciones del proceso de síntesis son soluciones enteras no todos los puntos de la curva son posibles). La Figura 3-11 muestra esta curva para el circuito DIFFEQ. Lógicamente las soluciones concretas dependen del tipo de restricciones que se impongan al diseño. Los resultados serán diferentes dependiendo si el retardo de la multiplicación es el mismo que el de la suma/resta o no, etc. A pesar de estas diferencias la estructura final sintetizada respecto al criterio utilizado (minimizar el retardo o minimizar el área) siempre tiene las mismas características. En el caso de la curva de la Figura 3-11 La solución de mínimo retardo está compuesta por más elementos operativos (7 recursos en total) que las soluciones de área mínima (2 recursos). Sin embargo los diferentes operadores en la solución de mínimo retardo trabajan en paralelo, aumentando el número de transiciones simultáneas que se realizan y, por tanto, es la peor solución en cuanto al ruido de conmutación. Un compromiso entre los dos extremos lo representan el punto central. La implementación de ésta solución se muestra en la Figura 3-12 y la solución de mínimo área en la Figura 3-13.



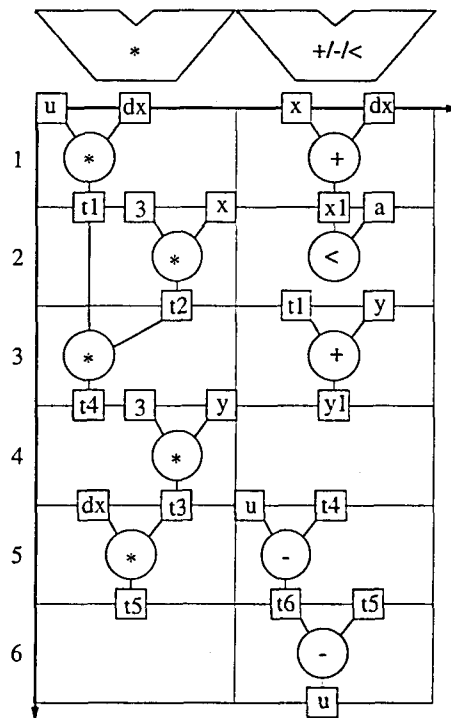
**Figura 3-11**

*Curva de síntesis para el circuito DIFFEQ con los puntos óptimos de diseño [7].*



**Figura 3-12**

Representación temporal de una posible implementación del circuito DIFFEQ correspondiente al punto central de la curva de la Figura 3-11.



**Figura 3-13**

Implementación del circuito DIFFEQ con restricciones de área mínima.

En concreto para la solución de menor retardo de la Figura 3-12 tenemos en uno de los pasos de la operación, el segundo, cuatro operadores trabajando simultáneamente. Sin embargo en la solución de área mínima de la Figura 3-13, es decir, una solución generada con el número de recursos limitado, como máximo tendremos dos operadores trabajando simultáneamente. Podríamos decir, en resumen, que siempre que el criterio sea minimizar el área las soluciones obtenidas serán circuitos digitales que generarán menos ruido de conmutación que las soluciones obtenidas si las restricciones temporales hacen aumentar el número de operadores trabajando simultáneamente.

### 3.3 Arquitectura y ruido de conmutación

Las alternativas estructuras del diseño lógico pueden clasificarse en función del nivel a que se apliquen. En primer lugar se analiza el proceso de síntesis lógica combinacional y las estructuras habituales que se encuentran en este tipo de circuitos. A continuación se estudia la estructura de las máquinas de estados finitos (FSMs) y finalmente las estructuras más complejas compuestas por una unidad operativa (*datapath*) y una unidad de control.

#### 3.3.1 Circuitos combinacionales y síntesis lógica

Si consideramos los circuitos digitales partiendo del nivel más bajo de complejidad tenemos los circuitos combinacionales, para los que existen varias alternativas de arquitectura: las matrices lógicas programables (PLA), la estructura tipo memoria ROM (Figura 3-14), la estructura basada en selectores y la implementación multinivel mediante puertas lógicas sencillas de pocas entradas (Figura 3-15). Las tres primeras son implementaciones multinivel [8]. La ventaja de la alternativa PLA respecto a la ROM y la basada en selectores es su menor área. La que presenta mejores prestaciones en cuanto a área y retardo es la lógica multinivel. Sin embargo es mucho más sencilla de diseñar la implementación mediante PLA o ROM.

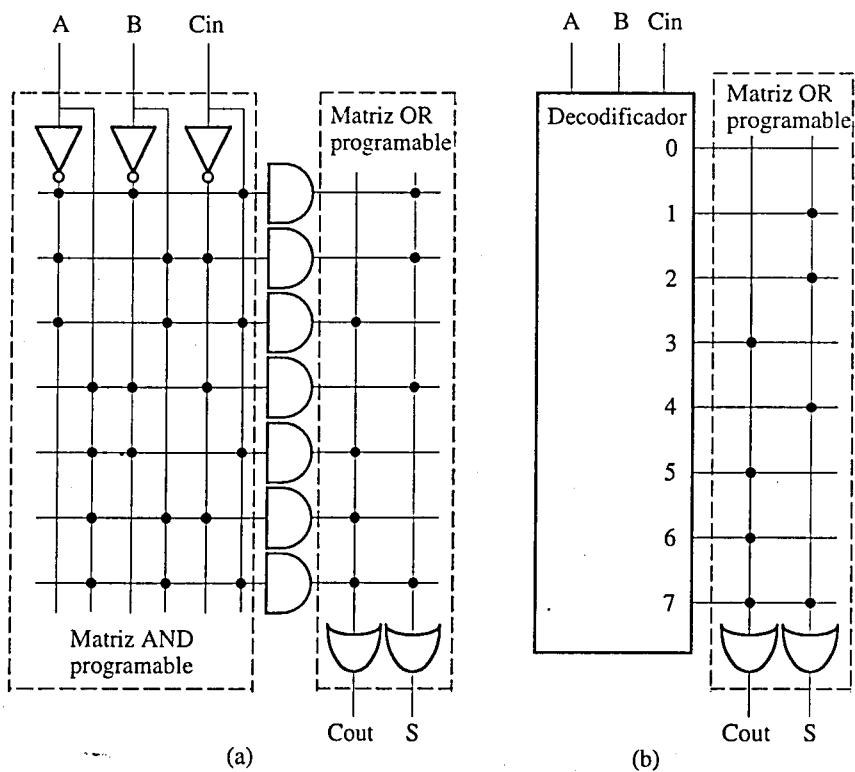
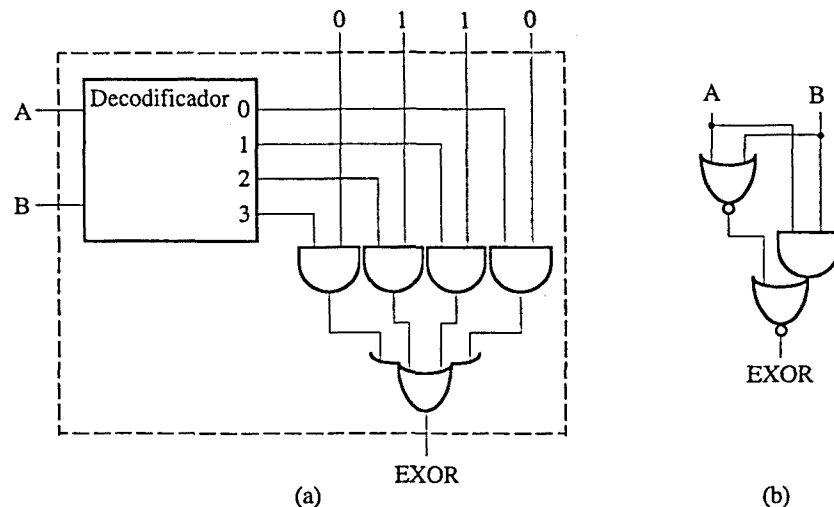


Figura 3-14

Implementación de la función sumador total (Full Adder) mediante (a) una PLA y (b) una ROM.



**Figura 3-15**

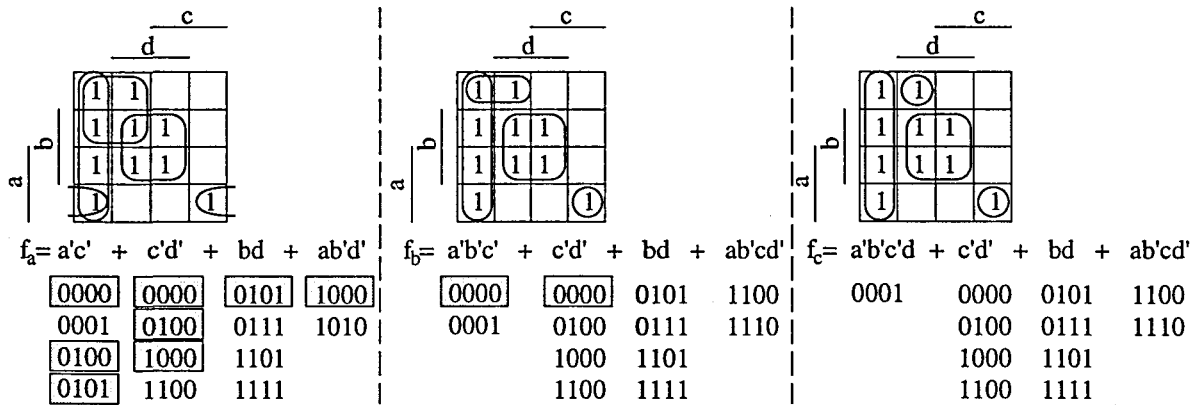
*Implementación de la función OR exclusiva (EXOR) mediante (a) un selector y (b) puertas lógicas.*

Respecto del ruido de conmutación las soluciones PLA y ROM son estructuras con dos niveles lógicos y por tanto las funciones combinatorias complejas requerirán muchas puertas en cualquiera de los dos niveles que pueden llegar a conmutar simultáneamente. La estructura multinivel optimiza el número de puertas que deben trabajar en un mismo intervalo temporal y al ser una estructura mucho más irregular las conmutaciones en los diferentes niveles de la lógica tienden a desincronizarse. Sin embargo no todas las puertas en un mismo nivel de una PLA o estructura a dos niveles tienen que conmutar simultáneamente. En el segundo nivel, si las salidas son independientes y todos los valores posibles están especificados si puede darse la situación en la que todas las salidas cambien de valor a la vez. En el primer nivel donde se calculan los factores de la función, también denominados cubos, generalmente nos encontramos con un número de puertas mayor. Cada puerta representa un cubo de la función. Cada cubo cubre una serie de unos de la función (en la estructura disyuntiva). Si existe un '1' de la función común a varios cubos cuando se presente la combinación de entrada correspondiente a ese '1' todas las puertas correspondientes a los cubos que cubren ese '1' se activarán y cuando esa combinación de entrada cambie todas esas puertas conmutarán si la nueva combinación de entrada no corresponde a otro '1' de la función cubierto por alguno de los cubos anteriores o si la corresponde a un '0'. En la Figura 3-16 se muestran varias coberturas de una misma función completamente especificada. La cobertura óptima en cuanto a número de puertas y de literales (número de entradas de las puertas) es la  $f_a$ .

La cobertura  $f_a$  está formada por cuatro cubos. Existen varios '1s' de la función que son cubiertos por varios de estos cubos. Cuando se produce la transición de entradas  $0101 \rightarrow 1000$  dos puertas (las correspondientes a los cubos  $a'c'$  y  $bd$ ) realizan una transición de nivel alto a bajo y otras dos (las correspondientes a los cubos  $c'd'$  y  $ab'd'$ ) realizan la transición contraria: en total cuatro transiciones, es decir, todas las puertas del primer nivel de la PLA que implementase esta función combinatorial conmutarían simultáneamente. Las coberturas  $f_b$  y  $f_c$  de la misma función solucionan esta situación a costa de aumentar el número de literales en los cubos de la función buscando siempre que cualquier '1' de la función esté cubierto por un único cubo. Si la implementación final del circuito se va a realizar con una PLA esto no supone un



incremento del área pues la matriz AND de la PLA ya está preparada para soportar un número determinado de entradas en cada puerta. Sin embargo si el objetivo final es una implementación a dos niveles con puertas de una librería estándar el aumento de los literales de los cubos representa incrementar el número de entradas de las puertas aumentando el área final de la implementación y también su retardo. Si existe un conjunto de entradas no especificadas (el denominado *don't care set*: DC) tenemos más libertad a la hora de formar los cubos y el criterio de no superposición de cubos podría realizarse, dependiendo del caso, sin incremento del número de literales.



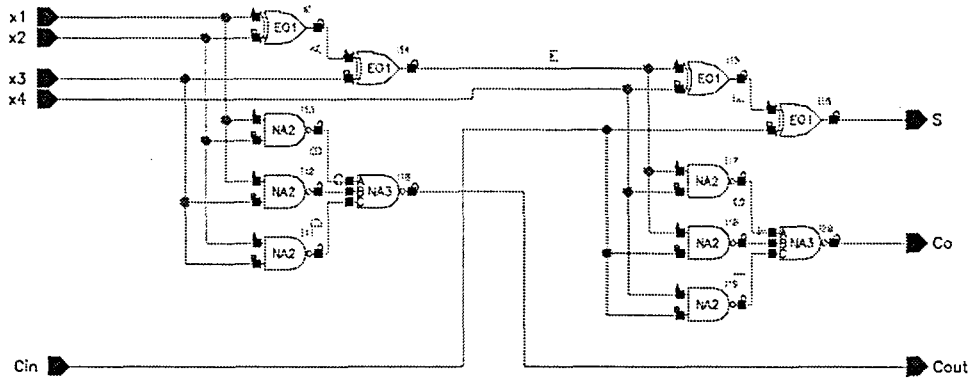
**Figura 3-16**

Varias posibles coberturas de una misma función. Bajo cada cubo se han escrito los unos de la función que cubre y se han resaltado aquellas combinaciones de entrada cubiertas por dos o más cubos.

Esta misma cuestión se encuentra si se analiza el proceso de síntesis lógica. Si se realiza una síntesis a dos niveles (mediante el algoritmo EXPRESSO [9][10], por ejemplo) tras la fase de expansión de todos los factores de la función, se pasa a la fase de selección de los cubos que formarán la cobertura óptima de dicha función. Utilizando los DCs se escogen aquellos cubos que minimizan una función de coste dada. Normalmente el coste está relacionado con el número de literales de cada cubo. El orden en el que el algoritmo encuentra los cubos es importante si el coste, en cuanto al número de literales, es igual en dos cubos consecutivos. Por tanto, una técnica sencilla de optimizar el número de transiciones simultáneas consistiría en calcular otra función de coste alternativa que tuviera en cuenta la superposición de los cubos entre sí. Si esta superposición se produce en '1s' de la función, minimizar el número de transiciones representaría escoger cubos con mayor número de literales. Es posible que utilizando adecuadamente el conjunto de DCs puedan encontrarse cubos que cubran todos los '1s' con el mínimo número de superposiciones posibles, minimizando de esta forma el máximo número de puertas en el primer nivel de la estructura que conmutarán simultáneamente.

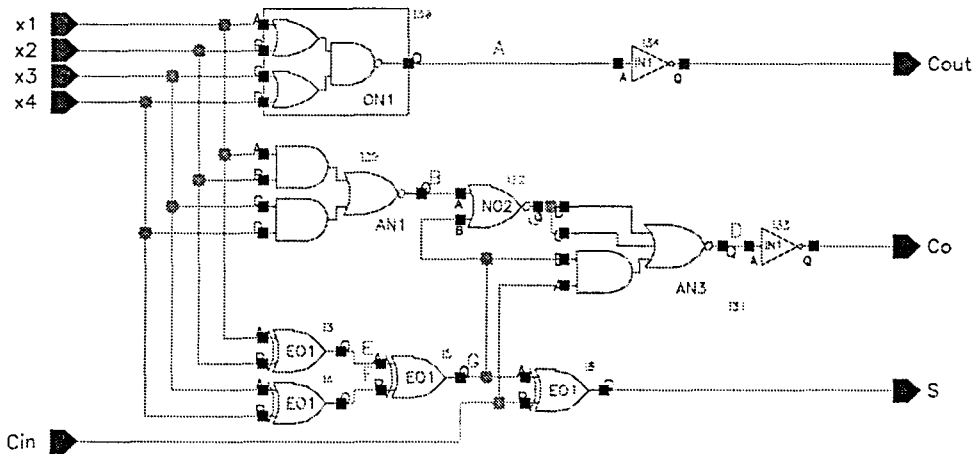
En cuanto a la síntesis lógica multinivel, no existen criterios tan claros como anteriormente de cara a cuantificar el número de transiciones simultáneas y las técnicas que podrían utilizarse para sintetizar redes de puertas menos ruidosas. Al nivel que trabajan los algoritmos de síntesis lógica sería necesario contar con información precisa de los retardos de cada puerta, incluyendo una estimación de las interconexiones. De esta forma se podrían establecer diferentes tramos temporales en los que un número de puertas determinadas puede conmutar. El número máximo de conmutaciones en esos intervalos de simultaneidad dependerá de la función lógica y de la

estadística de las entradas. Una pequeña modificación del circuito afectará a un gran número de caminos de propagación y por tanto a los intervalos de simultaneidad. Del mismo modo las transiciones espúreas (*glitches*) deben tenerse en cuenta pues contribuyen de forma importante al número de transiciones totales que se realizan en un determinado instante de tiempo. Esto queda ilustrado con el siguiente ejemplo. La Figura 3-17 corresponde a un encadenamiento de dos sumadores totales (*full adders*) al que llamaremos *2xfa* realizados con lógica a dos niveles y puertas XOR.



**Figura 3-17**

*Circuito 2xfa correspondiente a dos sumadores totales encadenados.*

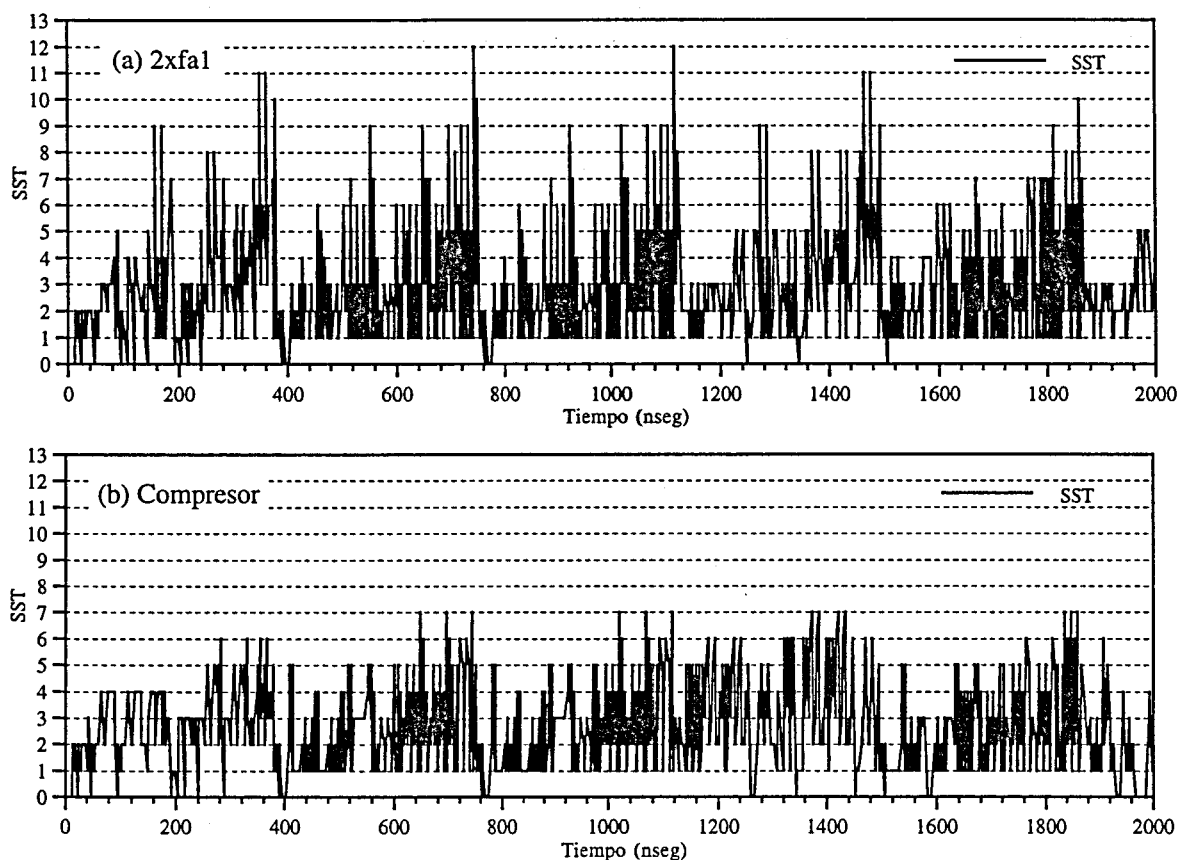


**Figura 3-18**

*Circuito compresor que realiza la misma función lógica que el circuito 2xfa pero con un menor retardo y número de puertas.*

La Figura 3-18 corresponde a un circuito alternativo denominado *compresor* que reduce el retardo y el número de puertas. El retardo del camino más largo del circuito *2xfa* es de 5.60nseg desde cualquiera de las entradas *x1* o *x2* hasta la salida *S*. Para el circuito *compresor* este retardo es de 4.20nseg entre cualquiera de las cuatro entradas *x1*, *x2*, *x3* o *x4* y la salida *S*. Estos retardos corresponden a una implementación con puertas de las librerías de celdas estándar de la tecnología de 0.8 $\mu$ m de AMS sin incluir los retardos debidos a las interconexiones. La simulación digital realizada con Verilog revela un mayor número máximo de transiciones simultáneas en el circuito *2xfa* que en el *compresor*. Estas simulaciones se han realizado con una secuencia de entradas exhaustiva, es decir, se ha probado todos los posibles

pares consecutivos de vectores de entrada. En las gráficas se muestran las zonas de la simulación donde se ha detectado máxima actividad en ambos circuitos.



**Figura 3-19**

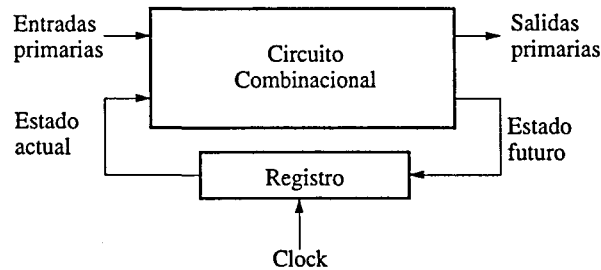
*Resultados de la simulación digital mediante Verilog de los circuitos (a) 2xfa y (b) compresor. Las gráficas indican el número de conmutaciones simultáneas. La secuencia de estímulos es común a los dos circuitos.*

A partir del ejemplo se puede ver también como un buen mapeo tecnológico ayuda a reducir la complejidad del circuito lógico y por tanto el número de conmutaciones simultáneas. El mismo circuito *compresor* de la Figura 3-18 realizado únicamente con puertas XOR, NOR y NAND y al que llamaremos *compresor2*, a pesar de que presenta el mismo retardo que el *compresor* (debido a que el retardo viene limitado por las 3 puertas XOR encadenadas) produce un número mayor de conmutaciones simultáneas (numerosos máximos de 9 y 10 conmutaciones simultáneas) en igualdad de condiciones, como se ha comprobado con una nueva simulación Verilog.

### 3.3.2 Máquinas de estados finitos

Los sistemas secuenciales más sencillos, las máquinas de estados finitos (FSM) tienen una representación canónica fijada con un registro que almacena el estado del sistema y un sistema combinacional que calcula las funciones de estado futuro y de salida (Figura 3-20). La cantidad de ruido de conmutación generada por la parte propiamente secuencial del FSM dependerá del número de bits de las variables que codifican el estado que cambien simultáneamente. Una codificación siguiendo los criterios de asignar códigos adyacentes a

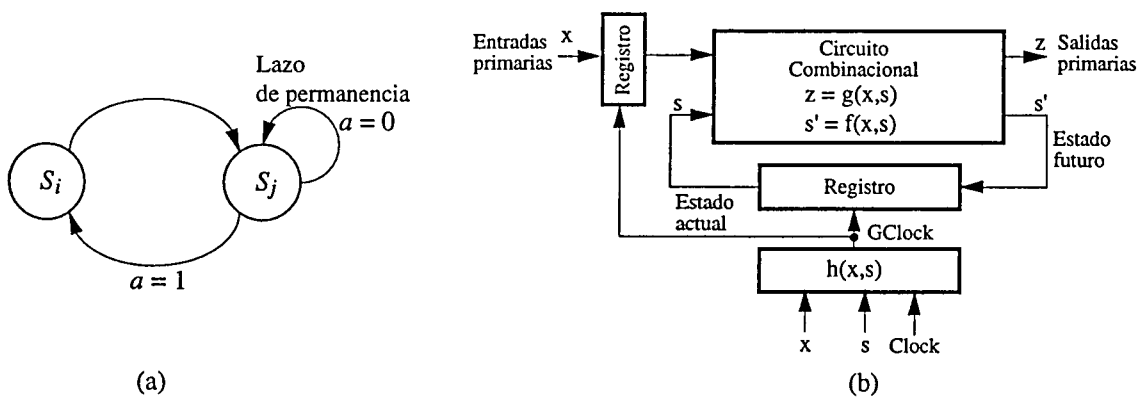
estados entre los que se produzcan muchas transiciones será beneficiosa en este sentido [11]. En todo caso, la mejor solución respecto al ruido de conmutación sería una codificación redundante tal que en cualquier transición de estado posible sólo cambiaran unos pocos bits.



**Figura 3-20**

*Estructura canónica de una FSM. El registro almacena el estado actual y el sistema combinacional calcula el estado futuro y las salidas en función del estado actual y las entradas.*

Otra alternativa que se utiliza para reducir el consumo en las FSM consiste en detener la señal de reloj cuando se produce una situación de permanencia en un mismo estado (lazo de permanencia) como sucedería en el estado  $S_j$  mientras la variable  $a$  valga '0' en la Figura 3-21.a. Esta arquitectura, que se conoce en inglés como *clock gating* [12][13], reduciría también el ruido de conmutación en esos estados mientras el sistema está detenido. En todo caso esta técnica reduciría el ruido de conmutación sólo en las situaciones en las que el sistema secuencial estuviese en esos estados, por lo tanto sólo durante unos intervalos de tiempo. Esta técnica puede utilizarse sincronizando los elementos sensibles al ruido para que funcionen precisamente durante estos lazos de permanencia en los que no se genera ruido de conmutación. Para ello podemos utilizar como señal de activación en estas partes sensibles del circuito la misma función que detiene el reloj,  $h(x,s)$  ilustrada en la Figura 3-21.b. Una técnica complementaria a ésta consistiría en introducir estos estados de permanencia de forma intencionada en el diagrama de la FSM de forma que asegurásemos cada cierto tiempo, y para cualquier secuencia de operaciones, unos intervalos de "tranquilidad" en el sistema en los que pudiesen trabajar las partes más sensibles al ruido de conmutación, sin peligro de verse afectadas por éste.

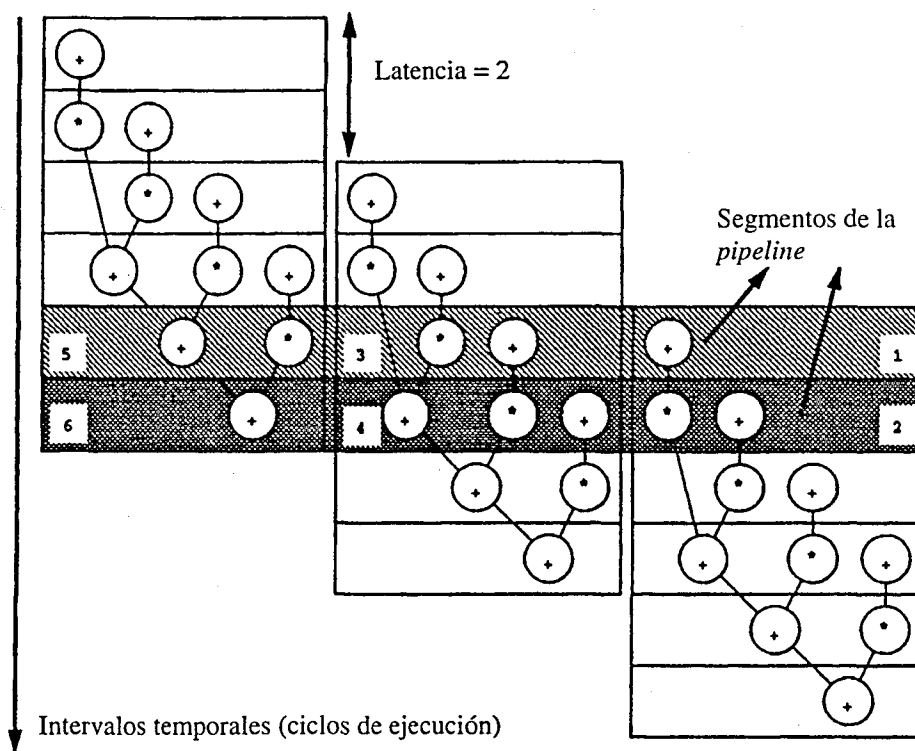


**Figura 3-21**

*(a) Diagrama de estados ilustrando un lazo de permanencia en el estado  $S_j$ , (b) Estructura modificada de una FSM incluyendo la función  $h(x,s)$  que detiene el reloj cuando se detecta un lazo de permanencia.*

### 3.3.3 Unidad de control y *Datapath*

El caso más general de un sistema digital es el que, debido al gran número de estados que se necesitarían para implementarlo mediante una FSM, se divide el sistema en una unidad *datapath* (o varias) y una unidad de control implementada mediante una FSM mucho más sencilla. La unidad *datapath* comprende registros para almacenar palabras de varios bits de información, operadores que procesan esos datos y multiplexores para seleccionar las entradas de los registros y los operadores. Normalmente el *datapath* está conectado a los buses de datos de entrada y salida y por tanto sus elementos trabajan con palabras de varios bits de anchura. La unidad de control, por el contrario, trabaja con señales sencillas, de un bit en la mayoría de los casos. La unidad de control proporciona las señales de habilitación y inicialización (*reset*) de los registros y de selección de los multiplexores y las unidades operativas (ALU's, multiplicadores, etc...) y recibe como entrada señales de estado que proporciona el *datapath* y que modifican la evolución del sistema digital.



**Figura 3-22**

*Representación esquemática mediante un grafo de secuencia de operaciones de la implementación del circuito DIFFEQ mediante una pipeline.*

Las alternativas en este tipo de arquitecturas se centran en la estructura del *datapath*. Como hemos visto en el apartado anterior, un mismo proceso puede realizarse de diferentes formas en función del número de operadores disponibles y de como se organice su utilización a lo largo del tiempo (planificación). El principal objetivo de los diseñadores, movidos por las demandas del mercado, es minimizar el retardo de los circuitos para aumentar su velocidad. La estructura de un *datapath* puede modificarse de dos formas diferentes para aumentar el *throughput*, es decir, la velocidad a la que la información puede ser procesada. Una de ellas es

umentar el grado de paralelismo y otra segmentar el *datapath* mediante una estructura *pipeline*. Un ejemplo de la primera alternativa lo hemos visto en el apartado anterior con el circuito DIFFEQ. La estructura con más operadores concurrentes generará más ruido de conmutación al realizarse más transiciones simultáneamente. La segunda alternativa, aplicada al mismo ejemplo, queda ilustrada en la Figura 3-22. En una *pipeline* cada segmento funciona de forma simultánea con los demás. En la figura, los segmentos 1, 3 y 5 funcionan concurrentemente durante un ciclo de reloj y posteriormente lo hacen los segmentos 2, 4 y 6 durante otro ciclo del reloj. La operación necesita seis ciclos de reloj para completarse pero pueden introducirse los datos a un ritmo del inverso de dos ciclos de reloj, por lo que la velocidad de proceso, el *throughput* se ve incrementado. El número de transiciones simultáneas aumentará proporcionalmente con el número de segmentos en los que se divida la *pipeline*. Una vez más se observa como la disminución del ruido de conmutación debería hacerse a costa de una penalización en la velocidad de los circuitos.

### 3.4 Alternativas lógicas y ruido de conmutación

La alternativa lógica utilizada para las puertas digitales básicamente influencia la velocidad, el tamaño, la disipación de potencia y la complejidad del cableado de un circuito. El retardo del circuito depende del número de niveles de inversión, del número de transistores en serie, de los tamaños de los transistores (i.e., la anchura del canal) y de las capacidades dentro de la celda y entre celdas. El tamaño del circuito depende del número de transistores y de sus tamaños y de la complejidad del cableado. La disipación de potencia o consumo depende sobre todo de la actividad de conmutación del circuito y de las capacidades de los nodos (compuesta de la capacidad de puerta, de las difusiones y de las interconexiones). La complejidad del cableado depende del número de conexiones y de sus longitudes y de si la lógica utilizada es unipolar o diferencial<sup>4</sup>. Estos cuatro aspectos son los que se utilizan a la hora de comparar las diferentes alternativas lógicas entre sí. En este apartado haremos un análisis de las principales alternativas lógicas introduciendo como elemento de comparación que completa a los anteriores el ruido de conmutación que generará un circuito compuesto de puertas según las diferentes lógicas analizadas.

El criterio que se utiliza para determinar el nivel de ruido de conmutación generado por cada alternativa lógica se basa en dos aspectos. Por un lado está el ruido generado por cada puerta individualmente que depende de la capacidad de conducir corriente de la puerta (*driving*) y de los tiempos de transición. Por otro lado está la simultaneidad implícita en un determinado estilo lógico que determina el número de puertas que conmutan en un mismo instante para realizar una función lógica en un circuito. Dado que hay alternativas lógicas que permiten trabajar con retardos menores que otras, y por tanto los tiempos de transición serán distintos dependiendo de la alternativa, se ha establecido como criterio de comparación que los tiempos

---

<sup>4</sup> Entendemos por *lógica unipolar* aquella que utiliza señales que representan un solo valor para cada variable en contraposición a las *lógicas diferenciales* que utilizan para una misma variable dos señales que corresponden al valor de la variable y a su complemento. Las puertas unipolares solo tienen una salida para cualquier función lógica mientras que las diferenciales tienen dos que proporcionan el valor de la función afirmada y complementada.

de transición vengan prefijados como especificación a cualquier lógica considerada. De esta forma, las lógicas que presenten menores capacidades en los nodos del circuito necesitan para realizar las transiciones en los tiempos dados pulsos de corriente de menor amplitud que las que tienen mayor capacidad en los nodos del circuito. Por tanto éstas últimas generan pulsos de ruido de mayor amplitud en igualdad de tiempos de transición.

La clasificación realizada de las diferentes alternativas lógicas se ha basado en [14] y [15] y se ha completado y actualizado donde se ha creído oportuno. Las asignaciones de una determinada lógica a clasificaciones como “estática”, “dinámica” o “*clocked*”<sup>5</sup> a veces son discutibles y se utilizan aquí como un recurso meramente organizativo de la exposición.

### 3.4.1 Lógicas estáticas

La principal característica de este tipo de lógicas es que realizan una evaluación continua de la función lógica. Cualquier cambio en las entradas es inmediatamente evaluado y propagado a las salidas. Un encadenamiento de este tipo de puertas propaga los cambios en las entradas por los diferentes caminos formados por las puertas. Existe el problema de que dos entradas a una puerta dada tengan retardos de propagación a lo largo de la cadena distintos, con lo que antes de evaluarse el valor definitivo pueden llegar a evaluarse valores intermedios debido a los diferentes tiempos de llegada de cada entrada. Esto produce conmutaciones espúreas o *glitches* que se añaden a las transiciones útiles realizadas en cada instante por el circuito lógico.

#### 3.4.1.1 Lógica CMOS convencional

Las puertas lógicas CMOS convencionales están compuestas por dos redes complementarias formadas por transistores PMOS y NMOS, respectivamente. Presentan unos márgenes de ruido excelentes. No tienen consumo estático, trabajan a velocidades relativamente altas y son bastante insensibles a variaciones del proceso de fabricación. Su principal desventaja es un consumo grande de área de silicio debido al uso de PMOS, que suelen ser 2 o 3 veces mayores que sus equivalentes NMOS. Tomaremos esta alternativa lógica como referencia para comparar las demás. Las características de generación de ruido de este tipo de lógica han sido analizadas extensamente en los capítulos precedentes.

#### 3.4.1.2 Lógica Pseudo-NMOS

Esta lógica es idéntica a la de acumulación/depleción NMOS excepto en que se utiliza un transistor PMOS como carga, en vez de un NMOS de depleción. El número de transistores, la carga capacitiva y el área son menores respecto a la lógica CMOS convencional. Sin embargo no es tan robusta frente a variaciones del proceso y consume potencia estática. La baja carga capacitiva de salida compensa la reducción impuesta por la carga activa PMOS (*PMOSpull-up*)

---

<sup>5</sup> Dejamos el término original en inglés *clocked* para evitar traducciones ad hoc que podrían llevar a confusión. Las lógicas *clocked* tienen un funcionamiento controlado por una señal de reloj que normalmente establece dos fases, una de ellas en la que la puerta memoriza la salida (actúa de *latch*). Se utiliza este término para diferenciarlas de las lógicas dinámicas, también controladas por una señal de reloj, pero en las que no se produce una memorización de la salida sino una precarga a un valor lógico fijado (uno o cero) y una posterior evaluación que hará cambiar este valor o no según la función lógica y las entradas.

al bloque NMOS en su capacidad de conmutar. La velocidad de las lógicas pseudo-NMOS es algo superior a la lógica CMOS convencional. Sus márgenes de ruido son algo más reducidos que los de la lógica CMOS convencional. Debido a su menor capacidad de carga el ruido generado por este tipo de puertas es menor, a igualdad de tiempos de transición, que en las puertas CMOS convencionales.

### 3.4.1.3 Lógica de transistores de paso (*pass transistor logic*)

Este tipo de lógicas ha sido propuesta en los últimos años como alternativa para bajo consumo y alta velocidad (i.e. la lógica de transistores de paso complementaria CPL [16]). La principal diferencia con la lógica CMOS convencional es que el surtidor de las redes de transistores se conecta a algunas entradas en vez de a las líneas de alimentación. La ventaja es que una sola red de transistores de paso (ya sea NMOS o PMOS) es suficiente para realizar la función lógica, lo que resulta en un número reducido de transistores y menor capacidad de carga. Por ello el ruido de conmutación generado por este tipo de puertas será menor que en sus equivalentes CMOS convencionales. La principal desventaja se debe a la caída de tensión en los NMOS que impiden tener salidas de excursión completa. El nivel alto en este tipo de lógica es de  $V_{DD}-V_{tm}$ . Los *buffers* que se han de situar a las salidas para restaurar los niveles lógicos consumen potencia estática. Debido a esta caída de tensión los márgenes de ruido son menores que en la lógica CMOS y si se reduce la tensión de alimentación se degradan más que en las puertas CMOS clásicas o pseudo-NMOS, así como también lo hace la velocidad, si la tensión umbral no se escala tan agresivamente como la tensión de alimentación, cosa usual como ya se verá en el capítulo siguiente. Este tipo de lógicas permite realizar determinadas funciones lógicas, como la multiplexación o la XOR con muy pocos transistores. Las entradas deben de estar disponibles afirmadas y negadas por lo que podríamos clasificar esta lógica también como diferencial. Una alternativa introducida para eliminar la caída de tensión debida a los NMOS consiste en colocar en paralelo un transistor PMOS dando lugar a la que se conoce como lógica de transistores de paso doble (DPL) [17]. Las ventajas debido a la baja capacidad de los nodos de salida de la lógica CPL se reducen en la DPL debido al uso de transistores PMOS, incluida la reducción en el ruido de conmutación generado.

### 3.4.2 Lógicas dinámicas sin estructura *pipeline*

Los circuitos CMOS dinámicos se utilizan ampliamente en el diseño VLSI cuando el objetivo es conseguir elevadas velocidades. Los circuitos dinámicos CMOS tienen la particularidad de realizar las operaciones lógicas con muy pocos transistores, de forma similar a la lógica pseudo NMOS, y por tanto, tienen bajas capacidades de carga. Sin embargo la operación de los circuitos dinámicos requiere del proceso de precarga en cada ciclo de reloj. La alta velocidad de operación se consigue a costa de un elevado consumo, debido a la conmutación en la precarga. Estas lógicas funcionan de forma síncrona con una señal de reloj. En uno de los semiciclos del reloj todas las puertas de la cadena realizan la precarga de forma simultánea. En el siguiente semiciclo la operación se va evaluando a lo largo de la cadena de forma progresiva. De cara al ruido de conmutación, una puerta dinámica aislada generará menos ruido que su equivalente CMOS debido a la baja capacidad de carga. Sin embargo el circuito



global realiza una conmutación simultánea por puerta en la fase de precarga, por lo que el ruido generado es, en total, mayor [18]. Por otro lado, los márgenes de ruido de este tipo de lógicas son los más reducidos de todos, del orden de las tensiones umbrales de los transistores.

#### 3.4.2.1 *Lógica Dominó y NP Dominó*

La lógica dominó consta únicamente de bloques N, es decir, a base de transistores NMOS, que realizan la evaluación de la función lógica [19]. Es muy rápida pero necesita intercalar un inversor estático entre puerta y puerta para poder encadenarlas. La lógica NP Dominó utiliza bloques N y P (árbol de evaluación con PMOS) alternados, ahorrándose el inversor entre las puertas [20]. Es más lenta debido al uso de transistores PMOS en los bloques P y además necesita dos fases de reloj (el reloj afirmado y su complemento). De cara al ruido de conmutación esto duplica la distribución del reloj y por tanto multiplica por dos el número de *buffers* que conmutarán simultáneamente en los flancos del reloj. Debido a la mayor capacidad de entrada de los bloques P, los bloques N deberán diseñarse para conmutar corrientes mayores si se quiere conservar la misma velocidad y el ruido generado por éstos bloques será mayor que en el caso de la lógica Dominó de tipo N solamente. La lógica Zipper Dominó [21], alternativa propuesta para solucionar el problema del reparto de carga en la fase de evaluación de las lógicas anteriores, utiliza cuatro fases de reloj, por lo que su uso es desaconsejable desde el punto de vista del ruido de conmutación.

#### 3.4.2.2 *Lógica CVSL (Cascode Voltage Switch Logic)*

Estas lógicas se basan en el mismo principio de funcionamiento que la lógica Dominó pero duplican el árbol de evaluación para obtener salidas complementarias. Se trata por tanto de una lógica diferencial y también se la conoce por DCVSL (*Differential Cascode Voltage Switch Logic*) [22]. Dos transistores PMOS entrelazados realizan la función de *latch* y aceleran la operación de conmutación de las salidas. La principal ventaja de este tipo de lógicas es la sencillez de trasladar una función lógica a un árbol de transistores NMOS que implemente la función y su complemento. Estas puertas duplican el número de nodos a conmutar respecto de las lógicas unipolares ya que una conmutación de la salida supone dos transiciones, una de '0' a '1' y otra de '1' a '0'. Debido al *latch* formado por los transistores PMOS entrelazados el comportamiento de estas puertas es estático.

#### 3.4.2.3 *Lógica DSL (Differential Split-Level)*

Esta alternativa lógica, derivada de la anterior, es aproximadamente dos veces más rápida que la DCVSL. Su estructura es la misma pero se intercalan entre el *latch* y los nodos de salida dos transistores NMOS controlados por una tensión de referencia que limitan la excursión de las tensiones de salida [23]. Es posible encadenar varias puertas DSL entre las que se pasan señales con solo  $V_{DD}/2$  de excursión. Las salidas finales son restauradas a los niveles convencionales. Dada la reducción en los niveles de tensión internos, la carga total que ha de conmutarse es mucho menor y por tanto el ruido de conmutación generado se reducirá también. Sin embargo estas lógicas presentan consumo estático y unos reducidos márgenes de ruido.

### 3.4.3 Lógicas dinámicas con estructura *pipeline*

La velocidad de los sistemas CMOS aún puede aumentarse más mediante el uso de estructuras de tipo *pipeline* como ya se vio en el apartado 3.3.3. Los retardos de los diferentes bloques encadenados no se acumulan. La máxima frecuencia de operación de una *pipeline* viene determinada por el retardo de uno solo bloque, el más lento de todos. Las lógicas dinámicas CMOS con *latches* a las salidas para evitar las conmutaciones espúreas debido a diferentes caminos de propagación en la cadena (*signal races*<sup>6</sup>) conducen a estructuras de tipo *pipeline*. De cara al ruido de conmutación esta técnica supone, como ya vimos también en el apartado anterior, que todos los bloques que funcionan en el mismo segmento temporal conmutan simultáneamente. En el caso de lógicas dinámicas el problema es más acusado pues mientras unos bloques realizan la precarga, los otros realizan la evaluación en el mismo semiciclo de reloj. En el semiciclo siguiente las operaciones se invierten. Cuanto más se fragmente el circuito total, mayor será el número de bloques que conmuten simultáneamente y mayor el ruido de conmutación generado.

Entre las lógicas dinámicas con estructura de *pipeline* tenemos, en primer lugar, las lógicas C<sup>2</sup>MOS (*Clocked CMOS*) [24] y la lógica NORA [25] que necesitan dos fases de reloj (afirmado y negado) con la consiguiente duplicidad de la distribución de dicha señal y de los *buffers*. Un poco menos problemáticas, las lógicas TSPC (*True Single Phase Clock*) [26] y ANL (*All N Logic*) [15] utilizan una sola fase del reloj para toda la *pipeline*.

### 3.4.4 Lógicas diferenciales controladas por el reloj (*clocked CMOS logics*)

Este tipo de alternativas lógicas que se conocen en inglés por el nombre de *clocked CMOS logics* incorporan en la estructura de la puerta un circuito que realiza las funciones de *latch*. Su origen está en la lógica DCVSL que se ha ido modificando para solucionar alguno de sus problemas como el consumo estático, la no simetría en la conmutación de las dos salidas complementarias y el excesivo tiempo de bajada producido por árboles N de evaluación con varios transistores NMOS en serie [23]. La lógica SSDL (*Sample-Set Differential Logic*) [27] añade un amplificador de sensado (*sense amplifier*) para acelerar el proceso de conmutación de la puerta al ser éste circuito el que descarga rápidamente una de las dos salidas tras la fase de muestreo (en la que las dos salidas se precargan a '1' lógico). Esta lógica presenta consumo estático. La lógica ECDL (*Enable/disable CMOS Differential Logic*) resuelve el problema del consumo estático de la lógica SSDL introduciendo dos inversores entrelazados en vez del amplificador de sensado. Todas estas lógicas presentan la ventaja frente a la lógica estática convencional CMOS de necesitar menos transistores para realizar la misma función lógica ya

---

<sup>6</sup> En las cadenas de puertas de lógicas dinámicas el problema de los diferentes caminos de propagación es mucho más importante que en las lógicas estáticas. En una lógica estática una señal que llegue antes que otra simplemente provoca un valor erróneo temporal a la salida de la puerta. Sin embargo en cadenas dinámicas si una de las entradas de una puerta descarga erróneamente la salida (al llegar antes que las otras entradas), el valor no puede restaurarse debido a que los nodos son dinámicos. Para solucionar este problema se insertan *latches* que mantienen cada cierto número de puertas el valor de las salidas hasta que son validados por la señal de reloj.

que con un único árbol NMOS es posible implementar fácilmente cualquier función booleana [28][29]. Por ello tienen menor capacidad de carga y son más rápidas que la lógica CMOS convencional [30]. Sin embargo el hecho de ser lógicas diferenciales compensa en cierta manera la reducción en el número de transistores.

Respecto al ruido de conmutación, todas ellas trabajan en dos fases controladas por la señal de reloj. En una primera fase (primer semiciclo de la señal de reloj) las dos salidas (afirmada y negada) se precargan a un determinado valor. Durante esta fase las entradas pueden cambiar, cosa que no era posible en las lógicas dinámicas, lo que permite una mayor velocidad de operación en *pipelines* realizadas con este tipo de lógicas. En la siguiente fase (segundo semiciclo de la señal de reloj) el *latch* interno se habilita y el árbol N realiza la evolución de la función. Una de las dos salidas estará conectada a tierra a través del árbol y la otra no, por tanto el *latch* bascula y hace que una de las dos salidas valga '1' y la otra '0'. Debido a la estructura del *latch* estas lógicas son bastante robustas frente al ruido que sólo puede afectar en determinados momentos, como durante la fase de basculación del *latch*. Pero debido a la fase de precarga (estática, no dinámica) nos encontramos con el mismo problema de la simultaneidad de las conmutaciones de las lógicas dinámicas.

Existe, no obstante, una posibilidad muy interesante con este tipo de lógicas que consiste en implementar encadenamientos de puertas autotemporizadas. La señal de reloj no se distribuye a todas las puertas de la cadena sino sólo a la primera puerta. Cada puerta genera la señal de reloj para la siguiente puerta una vez ha evaluado sus salidas. De esta forma todas las conmutaciones son consecutivas y se reduce al máximo la simultaneidad. Pueden realizarse también *pipelines* de esta forma, pero en vez de funcionar todas las celdas a la vez, los datos se van propagando progresivamente por la cadena en forma de 'olas'. A estas estructuras se las denomina *wave pipelines* [31][32] y, a pesar de la dificultad de su diseño debido a las diferencias en los retardos de diferentes caminos, representan una atractiva alternativa para la minimización del ruido simultáneo de conmutación.

### 3.4.5 Diseño síncrono *versus* diseño asíncrono

Los sistemas secuenciales asíncronos han sido constantemente propuestos como una alternativa mucho más veloz que los sistemas síncronos. Sin embargo su complejidad de diseño y su sensibilidad a las variaciones del proceso que influyen en la dispersión de los retardos de las puertas han impedido que su uso se generalice. Los sistemas asíncronos no necesitan una señal de reloj general para todo el sistema por lo que las transiciones se realizan en cualquier instante de tiempo. Este funcionamiento es óptimo para reducir al máximo el ruido de conmutación pues se minimiza la simultaneidad [33].

## 3.5 Técnicas de reducción del consumo

El análisis detallado de las técnicas de reducción del consumo en CI CMOS requeriría trabajo extenso, que queda fuera de los objetivos de esta tesis. Sin embargo, dada la relación que puede establecerse entre algunas técnicas de reducción del consumo y la reducción del ruido

de conmutación se ha realizado un breve resumen de dichas técnicas. La reducción del consumo puede aplicarse a todos los niveles del proceso de diseño y será éste el orden que seguiremos.

A nivel de sistema están las técnicas de deshabilitación del consumo (*power down*) y los modos de funcionamiento en consumo reducido de los modernos microprocesadores y CIs similares. El principal problema de estas técnicas es que en los cambios de modo de activo a *power down* y viceversa el sistema demanda grandes pulsos de corriente en tiempos reducidos, lo que supone grandes  $dI/dt$  y ruido resistivo e inductivo. Esta característica ha modificado los criterios de diseño del sistema de distribución de la alimentación en los sistemas de altas prestaciones con varios modos de consumo [34][35][36].

Cuando ya se tiene una visión de la estructura del diseño se ha de identificar cuales son las principales fuentes de consumo de potencia. Estas fuentes, en un CI CMOS, son tres: el consumo debido a las conmutaciones de las salidas, el consumo debido a la corriente de cortocircuito de las puertas CMOS y el consumo debido a las pérdidas, principalmente al consumo sub-umbral de los MOSFET. En un circuito “bien diseñado” el término dominante es el debido a las conmutaciones y es este término el que más influencia tiene en la reducción del ruido de conmutación. La reducción de la corriente de cortocircuito siempre tendrá un efecto beneficioso en la reducción del ruido de conmutación pues tendrá como resultado la generación de pulsos de corriente de amplitud ligeramente menor. El consumo debido a la actividad del circuito suele representarse por la expresión  $P_{\sigma} = \alpha \cdot C_L \cdot V \cdot V_{DD} \cdot f_{CLK}$ . Donde  $\alpha$  es un factor que representa la actividad del sistema,  $C_L$  la capacidad de carga,  $V$  la excursión de salida, que puede o no coincidir con la tensión de alimentación ( $V_{DD}$ ) y  $f_{CLK}$  la frecuencia del reloj. La técnicas de reducción del consumo se centran en reducir cada uno de los factores de la expresión anterior.

Reducir la tensión de alimentación tiene un efecto cuadrático en la reducción del consumo. Pero también provoca una reducción del retardo. Disminuir la tensión de alimentación también provoca una reducción del ruido de conmutación, pero también se reducen los márgenes de ruido, por lo que la importancia relativa del ruido sigue siendo la misma, como se verá en el capítulo siguiente. Para compensar el aumento de retardo algunos autores del entorno del bajo consumo proponen el uso de paralelismo o estructuras *pipeline* [37]. Este aumento de la concurrencia tiene efectos perjudiciales al aumentar el número de elementos susceptibles de conmutar simultáneamente por los que estas alternativas deben estudiarse con detalle estableciendo el compromiso entre la reducción del ruido producida por la reducción de la tensión y el aumento debido al de la actividad concurrente.

A nivel algorítmico y de arquitectura pueden aplicarse varias técnicas. Algunas de ellas se basan en la estrategia anterior y buscan minimizar el incremento de retardo producido por la reducción de la tensión de alimentación con técnicas que aceleran la velocidad de proceso. Otras técnicas buscan minimizar directamente la actividad a través de optimizar las operaciones y la complejidad de los algoritmos [38]. Todas estas técnicas, *a priori*, tienen un efecto beneficioso si logran minimizar el número máximo de elementos simultáneos necesarios para realizar esas operaciones, al reducir la complejidad general de todo el sistema. En el proceso de síntesis de la

arquitectura, como se indica en [39], determinadas propiedades pueden ser explotadas para reducir el consumo. La *localidad espacial* permite identificar operaciones que están fuertemente conectadas entre sí. La compartición de recursos puede hacerse aprovechando esta propiedad en la fase de asignación de recursos a las operaciones. De esta forma la partición se realiza de acuerdo con agrupaciones de operaciones inherentes en el algoritmo y se reduce el número de transferencias altamente capacitivas. La reducción de este tipo de conexiones beneficia a la reducción de ruido de conmutación pues como ya se ha visto los *buffers* y etapas de salida que deberán conmutar estas grandes capacidades generan elevados niveles de ruido de conmutación. Otras propiedades de los algoritmos como la *regularidad* si se explotan en el diseño de bajo consumo producen una reducción de las interconexiones (*buffers*, multiplexores y *buses*) lo que beneficia también a la reducción del ruido de conmutación. Otra de las posibles alternativas, la selección de la representación numérica (complemento a 2, modulo y signo, etc.) tiene también repercusiones en el consumo. Por ejemplo, una comparación de la actividad de conmutación del formato complemento a 2 (CA2) con el formato modulo y signo presentada en [40] muestra como esta última produce menos transiciones dado que al atravesar el cero solo cambia el bit más significativo, mientras que en los números expresados en CA2 conmutan todos los bits de signo.

A nivel lógico las técnicas de reducción del consumo se centran en minimizar la capacidad conmutada a través de la actividad o bien a través de la estructura de la lógica. La reducción de la capacidad de los nodos de salida tiene un efecto directo en el ruido de conmutación. El tamaño de los transistores para realizar una transición en un tiempo dado será menor si la capacidad de carga se reduce y los pulsos de corriente tendrán menor amplitud. No obstante si esto se aprovecha para aumentar la velocidad (dado que ahora las capacidades de entrada de las puertas son también menores) los tiempos de transición a la entrada se reducirán y se deberá realizar un estudio detallado para comprobar si realmente la  $dI/dt$  disminuye o no. Hay que recordar también que por encima de una capacidad crítica, como se vio en el capítulo dos, el aumento de la capacidad no supone un aumento del ruido de conmutación. Si la reducción de la capacidad no sobrepasa este umbral, no se reducirá el ruido de conmutación, aunque si se reduzca el consumo.

Otro campo muy activo se centra en la reducción de la actividad de los circuitos lógicos y existen propuestas de algoritmos de síntesis lógica con objetivos de reducción del consumo [41][42][43][44]. Se ha de realizar un estudio mucho más profundo de la relación de estas técnicas con el ruido de conmutación, pues la reducción de la actividad total de un circuito no necesariamente implica que la actividad simultánea máxima se reduzca en la misma proporción. Esta es una de las futuras líneas de investigación que se proponen. La reducción del consumo en los sistemas secuenciales, mediante técnicas de síntesis y codificación de estados adecuadas, tiene un efecto directo beneficioso en la reducción del ruido de conmutación, como ya se vio en el sub-apartado 3.3.2.

En resumen, no puede decirse que todas las técnicas que reducen el consumo provocan una reducción del ruido de conmutación. Existen alternativas que lo incrementan, alternativas

que lo disminuyen y otras en las que no está claro y sobre las que es necesario un estudio más detallado.

### 3.6 Conclusiones

En este capítulo se han analizado las alternativas que se le presentan al diseñador de circuitos integrados digitales CMOS en las diferentes fases del diseño.

A nivel abstracto se ha visto como puede influir el proceso de síntesis de una descripción a alto nivel en la estructura final del circuito y por tanto en el ruido de conmutación generado. Normalmente los sintetizadores comerciales permiten establecer dos criterios de síntesis: mínima área o mínimos retardos. La primera alternativa produce soluciones que generan menos ruido de conmutación que la segunda, como se ha demostrado con varios ejemplos. La causa está en que para reducir los retardos se ha de recurrir a la concurrencia, aumentando el número de bloques lógicos que trabajan simultáneamente. Se ha demostrado también cómo este resultado no depende de si la especificación del sistema se ha realizado de forma comportamental (ejemplos 1 y 3 del apartado 3.2) o estructural (ejemplo 2). Queda propuesto como posible línea futura de investigación analizar en detalle los algoritmos de síntesis usualmente utilizados e introducir las modificaciones oportunas que permitieran, dentro de las restricciones de área o retardo dadas, minimizar el grado de paralelismo. Una posible estrategia sería: dado un grafo de secuencia ya planificado temporalmente, en el que queda fijado un número máximo de operaciones encadenadas, distribuir el resto de operaciones que se realizan concurrentemente por los diferentes intervalos temporales en los que se subdivide la secuencia de operaciones.

En el siguiente nivel de abstracción, el nivel estructural, también se han analizado las diferentes alternativas en cuanto a la arquitectura de los diferentes módulos lógicos. Para sistemas combinatoriales se han considerado sobre todo las implementaciones mediante PLA o lógica a dos niveles y las implementaciones con lógica multinivel, haciendo referencia al proceso de síntesis lógica. Las estructuras a dos niveles, dada su regularidad, se prestan a realizar con mayor simultaneidad las transiciones lógicas. Sin embargo se ha demostrado que utilizando determinados criterios en el proceso de minimización lógica a dos niveles como la no superposición de los cubos de la función reduce el número de conmutaciones simultáneas, a costa de incrementar en algunos casos el número de literales de los factores de la función y por tanto el área y el retardo. Estas estrategias de minimización orientadas a reducir el número de transiciones simultáneas en el proceso de optimización lógica podrían incorporarse a los algoritmos utilizados en los procesos de síntesis lógica (de forma similar a las técnicas utilizadas para reducir la actividad en el campo del bajo consumo), y queda propuesto como línea de investigación futura. En cuanto a la síntesis multinivel no se han encontrado criterios tan claros que permitan relacionar las técnicas de minimización con el ruido de conmutación, si bien se ha demostrado con un ejemplo que diferentes implementaciones de una misma función pueden dar lugar a circuitos con un número máximo de transiciones simultáneas bastante diferente. No necesariamente las estructuras más irregulares dan lugar a menores cotas de actividad simultánea debido, entre otras cosas, a las conmutaciones espúreas o *glitches*.

Para los sistemas secuenciales formados por una máquina de estados finitos, basándonos en su estructura canónica, pueden aplicarse los criterios comentados anteriormente para el sistema combinacional que calcula el estado futuro y las salidas. Para el registro que almacena el estado actual pueden aplicarse técnicas de codificación de los estados en el proceso de síntesis del sistema secuencial que minimicen el número de bits que cambian en las transiciones entre estados. También se han propuesto otras alternativas que se utilizan en el diseño de sistemas secuenciales de bajo consumo como la parada del reloj en los lazos de permanencia de los diagramas de estados. La inserción artificial de estos estados de permanencia en una sistema secuencial permitiría disponer de intervalos de funcionamiento ‘tranquilos’ (con la citada técnica del *clock gating*) en los que podrían operar las partes sensibles al ruido del mismo circuito integrado.

En la estructura más general de un sistema basado en un *datapath* y una unidad de control deben aplicarse los criterios deducidos en los apartados anteriores sobre síntesis a alto nivel y síntesis de FSMs. La paralelización o las estructuras tipo *pipeline* aumentan las cotas de actividad simultánea produciendo una mayor generación de ruido de conmutación. En este caso existe un compromiso entre la velocidad de proceso del sistema y el ruido de conmutación generado que debe resolverse en cada aplicación concreta limitando el número de unidades en paralelo o el número de segmentos de la *pipeline*.

Por último se ha descendido al nivel lógico analizándose las alternativas lógicas CMOS más usuales y comparándose cualitativamente respecto del ruido de conmutación. Tomando como referencia la lógica CMOS estática convencional existen algunas alternativas interesantes que, mostrando peores prestaciones en otros parámetros como el consumo o la velocidad, son mejores en cuanto a la generación de ruido. Entre ellas podríamos destacar las lógicas Pseudo NMOS que debido a su menor capacidad de carga permiten conmutar con pulsos de corriente de menor amplitud en el mismo tiempo la salida, generando por tanto menor ruido de conmutación  $dI/dt$ . Lo mismo sucede con la lógica CPL, si bien las mejoras introducidas por sus versiones posteriores como la DPL eliminan esta ventaja por lo que respecta al ruido de conmutación. Las lógicas dinámicas son en general desaconsejables si el ruido de conmutación es una restricción de diseño debido a que realizan una precarga simultánea todas las puertas del circuito. Si además se estructuran en forma de *pipeline* para evitar los problemas de *signal races* la cota de actividad simultánea del circuito aún aumenta más. Una alternativa interesante, no obstante, es la lógica DSL que trabaja internamente con la mitad de la tensión de alimentación. A su baja capacidad de carga une la reducida excursión de tensión en las salidas por lo que los pulsos de corriente que se necesitan para realizar las transiciones en el mismo tiempo que su equivalente CMOS convencional son mucho más pequeños y por tanto también lo es el ruido de conmutación generado. No obstante los márgenes de ruido son también muy reducidos. La alternativa más interesante la representan las lógicas *clocked CMOS* (DCVSL, SSDL, ECDL) dada la sencillez con la que pueden encaenarse en forma autotemporizada. De esta forma la actividad simultánea se reduce al máximo. Extrapolando esta idea al sistema global se ha de considerar como una opción muy interesante de cara a minimizar el ruido de conmutación el diseño secuencial asíncrono, a pesar de su complejidad de diseño, pues puede ser una alternativa idónea cuando el ruido de conmutación y la velocidad son dos limitaciones críticas del sistema.

## Referencias

- [1] Peter J. Ashenden. (1996). *The Designer Guide to VHDL*. 1st edition. San Francisco: Morgan Kaufmann Publishers.
- [2] D.E. Thomas, P.R. Moorby. (1991). *The Verilog Hardware Description Language*. Boston: Kluwer Academic Publishers.
- [3] *Hamburg VHDL Archive*, Página consultable desde el Web, URL <<http://tech-www.informatik.uni-hamburg.de/vhdl/vhdl.html>>, Universidad de Hamburgo.
- [4] Referencia del software synergy
- [5] D.E. Thomas, P.R. Moorby (1991) *The Verilog Hardware Description Language*. Massachusetts: Kluwer Academic Publishers.
- [6] Pierre G. Paulin, Hohn P. Knight, 'Force-Directed Scheduling for the Behavioral Synthesis of ASIC's', *IEEE Tr. on Computer-Aided Design*, Vol. 8, No. 6, June 1989, pp. 661–679.
- [7] Sabine März. 'High-Level Synthesis' en *The Synthesis Approach to Digital System Design*, Petra Michel, Ulich Lauter, and Peter Duzy, Ed. (1992). Massachusetts: Kluwer Academic Publishers.
- [8] Daniel Gajski, Nikil Dutt, Allen Wu, and Steve Lin. (1992). *High-Level Synthesis: Introduction to Chip and System Design*. Massachusetts: Kluwer Academic Publishers.
- [9] R.K. Brayton, G.D. Hatchel, C.T. McMullen, and A. Sangiovani-Vicentelli, *Logic Minimization Algorithms for VLSI Synthesis*. Boston, Massachusetts: Kluwer Academic Publishers, 1984.
- [10] R. Rudell and A. Sangiovani-Vicentelli, 'Multiple-Valued Minimization for PLA Optimization', *IEEE Tr. on Computer-Aided Design of Integrated Circuits and Systems*, Vol. CAD-6, Sept. 1987, pp. 727–750.
- [11] Luca Benini, and Giovanni De Micheli, 'State Assignment for Low Power Dissipation', *IEEE J. Solid-State Circuits*, Vol. 30, No. 3, March 1995, pp. 258–267.
- [12] Luca Benini, and Giovanni De Micheli, 'Automatic Synthesis of Low-Power Gated-Clock Finite-State Machines', *IEEE Tr. on Computer-Aided Design of Integrated Circuits and Systems*, Vol. 15, No. 6, June 1996, pp. 630–643.
- [13] Manfred Koegst, Günter Franke, Steffen Rülke, and Hlaus Feske, 'A Strategy for Low Power FSM-Design by Reducing Switching Activity', *Proc. of 7th Intl. Workshop PATMOS'97*, Lauvain-la Neuve, Belgium, 8–10 September 1997, pp. 119–128.
- [14] Neil H.E. Weste, Kamran Eshraghian, *Principles of CMOS VLSI Design. A systems perspective*. 2d. Edition (1994). Massachusetts: Addison-Wesley Publishing Company.
- [15] Richard X. Gu, Khaled M. Sharaf, Mohamed I. Elmasry. *High-Performance Digital VLSI Circuit Design*. 1996. Massachusetts: Kluwer Academic Publishers.
- [16] K. Yano, et al., '3.8-ns CMOS 16x16-b Multiplier Using Complementary Pass-Transistor Logic', *IEEE J. Solid-State Circuits*, Vol. SC-25, April 1990, pp. 388–395.
- [17] M. Suzuki, et al., 'A 1.5-ns 32-b CMOS ALU Double Pass-Transistor Logic', *IEEE J. Solid-State Circuits*, Vol. SC-28, Nov. 1993, pp. 1145–1151.
- [18] Patrik Larsson, and Christer Svensson, 'Noise in Digital Dynamic CMOS Circuits', *IEEE J. Solid-State Circuits*, Vol. 29, No. 6, June 1994, pp. 655–662.
- [19] R.H. Krambeck, Charles M. Lee, and Hung-Fai Stephen Law, 'High-Speed Compact Circuits with CMOS', *IEEE J. Solid-State Circuits*, Vol. SC-17, No. 3, June 1981, pp. 614–619.



- [20] V. Friedman and S. Liu, 'Dynamic Logic CMOS Circuits', *IEEE J. Solid-State Circuits*, Vol. SC-19, April 1984, pp. 263–266.
- [21] C.M. Lee, and E. W. Szeto, 'Zipper CMOS', *IEEE Circuits & Devices Mag.*, May 1986, pp. 10–16.
- [22] Kan M. Chu, and David L. Pulfrey, 'A comparison of CMOS Circuit Techniques: Differential Cascode Voltage Switch Logic Versus Conventional Logic', *IEEE J. Solid-State Circuits*, Vol. SC-22, No. 4, August 1987, pp. 528–532.
- [23] Leo C.M.G. Pfenning, Wim G.J. Mol, Joseph J.J. Bastiaens, and Jan M.F. Van Dijk, 'Differential Split-Level CMOS Logic for Subnanosecond Speeds', *IEEE J. Solid-State Circuits*, Vol. SC-20, No. 5, October 1985, pp. 1050–1055.
- [24] Yasoji Susuki, Kaicho Odagawa, and Toshio Abe, 'Clocked CMOS Calculator Circuitry', *IEEE J. Solid-State Circuits*, Vol. SC-8, No. 6, Dec. 1973, pp. 462–469.
- [25] Nelson F. Goncalves, Hugo J. De Man, 'NORA: A Recefree Dynamic CMOS Technique for Pipelined Logic Structures', *IEEE J. Solid-State Circuits*, Vol. SC-18, No.3, June 1983, pp. 261–266.
- [26] Jiren Yuan and Christer Svensson, 'High-Speed CMOS Circuit Technique', *IEEE J. Solid-State Circuits*, Vol. 24, No. 1, February 1989, pp. 62–70.
- [27] Timothy A. Grotjohn, and Bernd Hoeflinger, 'Sample-Set Differential Logic (SSDL) for Complex High-Speed VLSI', *IEEE J. Solid-State Circuits*, Vol. SC-21, No. 2, April 1986, pp. 367–369.
- [28] Kan M. Chu, and David L. Pulfrey, 'Design Procedures for Differential Cascode Voltage Switch Circuits', *IEEE J. Solid-State Circuits*, Vol. SC-21, No. 6, December 1986, pp. 1082–1087.
- [29] Jordi Cortadella, 'Mapping BDD's Into DCVSL Gates (or Switch-Level Synthesis)', Research report of the Computer Architecture Dept, Universitat Politècnica de Catalunya, August 1992.
- [30] Shih-Lien L. Lu, and Milos D. Ercegovac, 'Evaluation of Two-Summand Adders Implemented in ECDL CMOS Differential Logic', *IEEE J. Solid-State Circuits*, Vol. 26, No. 8, August 1991, pp. 1152–1160.
- [31] Wei-han Lien, and Wayne P. Burleson, 'Wave-Domino Logic: Theory and Application', *IEEE Tr. on Circuits and Systems—II*, Vol. 42, No. 2, February 1995, pp. 78–91.
- [32] C.H. Lau, D. Renshaw, and J. Mavor, 'A Self-Timed Wavefront Array Multiplier', *Proc. of 1989 Intl. Symp. on Circuits and Systems (ISCAS'89)*, pp. 138–141.
- [33] Vitit Kantaburata, and Andreas G. Andreou, 'A State Assignment Approach to Asynchronous CMOS Circuit Design', *IEEE Tr. on Computers*, Vol. 43, No. 4, April 1994, pp. 460–469.
- [34] Bob Mammano, 'Fueling the Megaprocessors - Empowering Dynamic Energy Management', —.
- [35] Rich Evans, and Michael Tsuk, 'Modeling and Measurement of High-Performance Computer Power Distribution System', *IEEE Tr. on Components, Packaging, and Manufacturing Techn.—Part B*, Vol. 17, No. 4, November 1994, pp. 467–471.
- [36] *VRM 9.1 DC-DC Converter Design Guidelines*. Intel Application Notes, Order Number 243408-001. May 1997.
- [37] Anatah P. Chandrakasan, Samuel Sheng, and Robert W. Brodersen, 'Low-Power CMOS Digital Design', *IEEE J. Solid-State Circuits*, Vol. 27, No.4, April 1992, pp. 473–484.
- [38] Anantha P. Chandrakasan, Miodrag Potkonjak, Renu Mehra, Jan Rabaey, and Robert W. Brodersen, 'Optimizing Power Using Transformations', *IEEE Tr. on Computer-*

*Aided Design of Integrated Circuits and Systems*, Vol. 14, No. 1, January 1995, pp. 12–31.

- [39] Jan M. Rabaey, and Massoud Pedram, editors. *Low Power Design Methodologies*, 1996, Massachusetts: Kluwer Academic Publishers.
- [40] Vasily G. Moshnyaga and Keikichi Tamaru, 'A Comparative Study of Switching Activity Reduction Techniques for Design of Low-Power Multipliers', *Proc. of 1995 IEEE Int. Symp. Circuits and Systems: ISCAS'95*, pp. 1560–1563.
- [41] Anantha P. Chandrakasan, and Robert Brodersen. *Low Power Digital CMOS Design*. 1995. Massachusetts: Kluwer Academic Publishers.
- [42] P. Girard, C. Landrault, S. Pravossoudovitch, and D. Severac, 'Technique for Reducing Power Consumption in CMOS Circuits', *Electronics Letters*, Vol. 33, No. 6, March 1997, pp. 485–486.
- [43] R. Iris Bahar, Fabio Somenzi, 'Boolean Techniques for Low Power Driven Re-Synthesis', *Proc. of IEEE ICAD Conf.*, 1995, pp. 428–432.
- [44] G. Theodoridis, S. Theoharis, D. Soudris, O. Koufopavlou, and C. Goutis, 'A Novel Approach for Reducing the Switching Activity in Two-Level Logic Circuits', *Proc. of ICECS*, 1996, pp. 840–843.

# Capítulo 4

---

## *Efectos del escalado en el ruido de conmutación para tecnologías CMOS*

*La tecnología de fabricación de semiconductores, y en especial la tecnología basada en dispositivos de efectos de campo de tipo CMOS es una de las industrias que más rápidamente evoluciona. El creciente requerimiento de mejores prestaciones que se exige a los circuitos integrados hace de motor de la evolución de estas tecnologías. Esta evolución tiene como objetivo principal conseguir reducir progresivamente las dimensiones de los dispositivos y sus interconexiones. Otros parámetros que evolucionan junto con las dimensiones de los dispositivos son la tensión de alimentación, la velocidad y el nivel de complejidad de los circuitos integrados. En este capítulo se analizan todos estos parámetros y sus tendencias futuras y se investiga el efecto de la evolución tecnológica en el ruido de conmutación para CI CMOS.*

### 4.1 Introducción

La industria microelectrónica ha experimentado un rápido crecimiento en las últimas décadas movida por el gran impulso económico que los avances en este tipo de tecnologías permiten en un gran número de áreas como las comunicaciones, la electrónica de consumo, la medicina, los transportes, la instrumentación de todo tipo, etc. La demanda de mejores prestaciones de los CI se centra sobre todo en tres aspectos:

- Aumento de la velocidad.
- Reducción de las dimensiones mínimas de los dispositivos, lo que permite integrar en el mismo área cada vez mayor número de transistores aumentando el número de prestaciones integradas y la complejidad del CI.
- Reducción del consumo, debido a las necesidades de ahorro energético y reducción de la generación de calor, y a la aparición de equipos electrónicos portátiles, sobre todo en el campo de las comunicaciones, la instrumentación y los ordenadores personales.

La consecución de estos tres objetivos a la vez normalmente no es posible y por ello surgen diferentes alternativas en la evolución tecnológica de las que hablaremos más adelante.

La progresiva miniaturización de las dimensiones de los dispositivos ha llevado a éstos a regiones de funcionamiento donde los modelos convencionales ya no son válidos debido a la aparición de efectos de segundo orden. Ha sido necesario introducir nuevos modelos que contemplen los fenómenos físicos que suceden en estos dispositivos actuales, como los efectos de alta inyección (BJT), de canal corto, conducciones subumbral (MOSFET), etc. Uno de los frenos a la evolución tecnológica proviene de estos efectos que afectan a la fiabilidad<sup>1</sup> de los dispositivos [1][2][3] y para los que se han ido introduciendo soluciones a medida que ha evolucionado la técnica de fabricación.

El aumento de la complejidad y de las prestaciones de los circuitos integrados ha provocado también que fenómenos que originalmente no eran considerados hoy en día deban ser tenidos en cuenta desde las primeras fases del diseño. Entre ellos cabe destacar los aspectos relacionados con la fiabilidad: fenómenos como la electromigración [4][5][6][7] y los efectos de portadores calientes (*hot carrier effects*) [8][9][10][11]. En el campo del diseño de circuitos integrados digitales además de una correcta funcionalidad y estabilidad ha sido necesario incluir en el diseño aspectos como el consumo de potencia [12][13][14] y los efectos de reducir la tensión de alimentación [15]. Entre estas consideraciones de diseño situamos el ruido de conmutación  $dI/dt$  y  $dV/dt$  [16][17]. El diseño de las interconexiones y del encapsulado ha ido aumentando su importancia debido a los efectos parásitos que éstos introducen en los circuitos integrados actuales [16].

En este capítulo se trata, en primer lugar, cuál ha sido la evolución, cuál es el estado actual de la tecnología de CI CMOS y cuál se prevee que puede ser su evolución en los próximos años. Se estudia también como repercute la evolución tecnológica en el ruido de conmutación  $dI/dt$ , incluyendo tanto los aspectos relacionados con el escalado de los dispositivos como los relacionados con el escalado de los sistemas de interconexión.

#### 4.1.1 Evolución de la tecnología de fabricación de circuitos integrados

Históricamente existen varios hitos que han marcado la evolución de lo que se denomina tecnología de muy alto nivel de integración (del inglés *Very Large Scale of Integration: VLSI*). Los primeros estudios teóricos provienen de los años 30' [18][19] cuando Lilienfeld y Heil pusieron las bases de lo que más tarde se denominaría transistor de efecto de campo (FET). En los años 1947 y 1948 Brattin, Bardeen y Shockley desarrollaron en los laboratorios Bell el

<sup>1</sup> El concepto de **fiabilidad** (*reliability*) se aplica a la fabricación de CI para referirse al estudio del funcionamiento de los posibles errores que puedan producirse en campo, durante el tiempo de vida del sistema electrónico. Es importante asegurar unos niveles elevados de fiabilidad respecto de los diferentes mecanismos físicos que tienden a degradar los dispositivos y las interconexiones que forman estos CI que aseguren una vida media aceptable y unas probabilidades de fallo en campo reducidas. Existen varios estándares de calidad que regulan estos aspectos como por ejemplo las normas MIL-HDBK-217F, publicadas por el Departamento de Defensa de EUA, Wanshington DC 20301, 1990.

transistor de unión bipolar (BJT). La aparición de este tipo de dispositivo marcó el inicio de la industria microelectrónica. En los 15 años posteriores se produjeron una gran variedad de este tipo de dispositivos para numerosas aplicaciones, sustituyendo a las válvulas de vacío y permitiendo la construcción de nuevos equipos electrónicos.

En 1958 Jack Kilby de Texas Instruments inventó el primer circuito integrado. El año siguiente Robert Noyce de Fairchild presentó una técnica más depurada y cercana a lo que hoy en día son los circuitos integrados. El trabajo de estos dos pioneros puso las bases de lo que hoy es el diseño de circuitos VLSI.

El germanio era el semiconductor más utilizado en los primeros dispositivos semiconductores. El silicio es hoy en día el semiconductor dominante, si bien otros materiales como el arseniuro de galio son utilizados en algunos nichos tecnológicos como los circuitos digitales de muy alta velocidad.

Tres factores han permitido el desarrollo que ha llevado a obtener las tecnologías de las que disponemos hoy en día: el avance en el campo de los materiales, las técnicas fotolitográficas y la aparición y perfeccionamiento de herramientas de ayuda al diseño CAD. En un inicio, el mercado de circuitos integrados estaba dominado por las tecnologías bipolares (TTL, ECL) si bien a partir de los años 80 el aumento del mercado de la tecnología CMOS ha sido constante siendo hoy en día y con diferencia la tecnología dominante para circuitos VLSI (ver Figura 4-1). La principal causa es la gran capacidad de integración que proporciona esta tecnología, como podemos observar en la Tabla 4-1, extraída de [20]. Las otras conservan nichos especiales de aplicación como la alta velocidad, etc.

Tecnología	Complejidad (Nº de puertas /CI)
CMOS	500K
BiCMOS	400K
GaAs	300K
ECL	60K

**Tabla 4-1**

*Niveles de complejidad típicos, expresados como número de puertas por circuito integrado, para diferentes tecnologías.*

A partir de ahora el análisis se centrará en la tecnología CMOS. En la Figura 4-2.(a) puede verse la evolución en la reducción de las dimensiones mínimas de los dispositivos, según [21] y en la Figura 4-2.(b) se presenta la evolución del nivel de integración para memorias y microprocesadores [20].

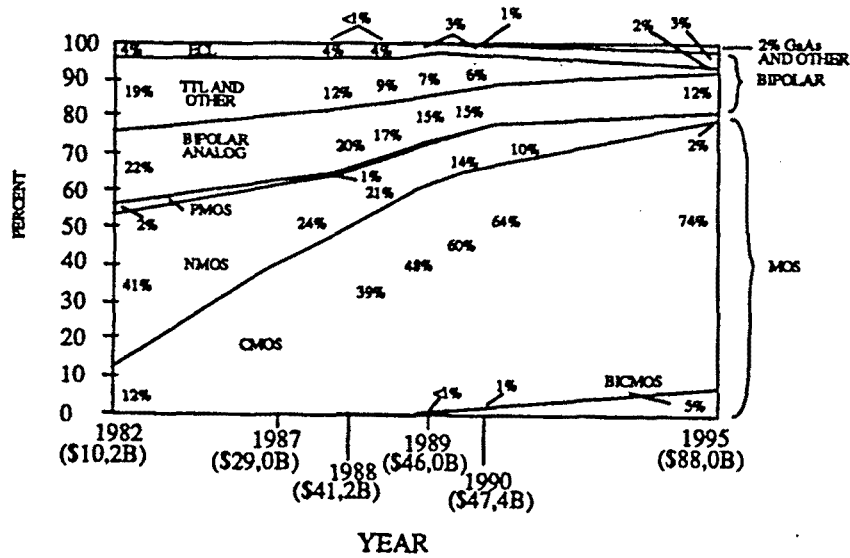


Figura 4-1 Evolución del mercado de circuitos integrados en función de las tecnologías [20]

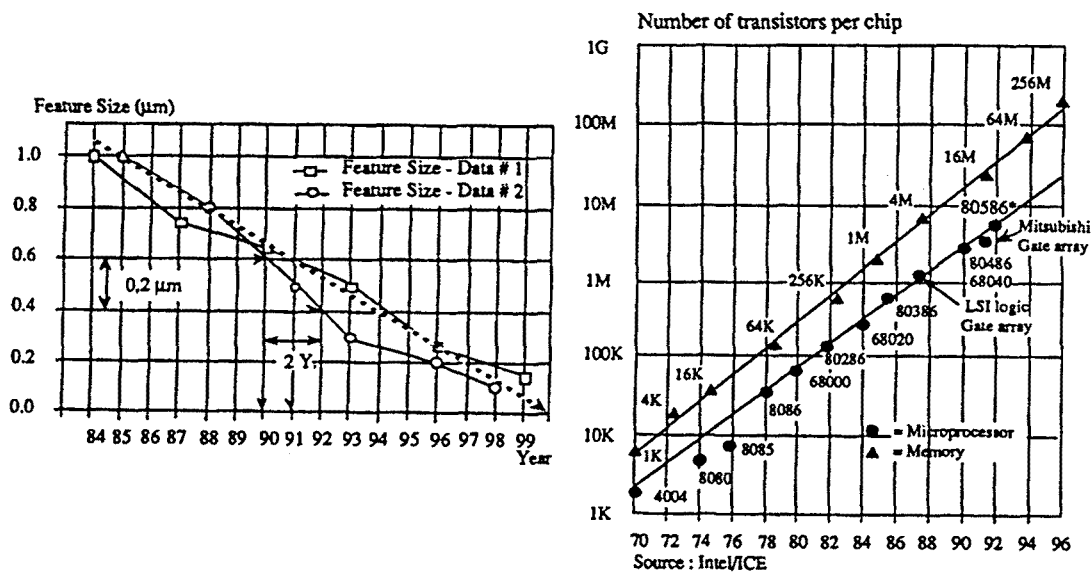


Figura 4-2 (a) A la izquierda se muestra una gráfica con la evolución de las dimensiones mínimas de la tecnología de los CI en función del año de aparición, incluyendo una prospectiva hasta el año 1999 [21]. (b) En la parte derecha se muestra una gráfica con la evolución del nivel de integración para dos productos representativos de la industria de fabricación de circuitos integrados: los microprocesadores y las memorias.

### 4.1.2 Estado actual de la tecnología CMOS y perspectivas futuras

En la Tabla 4-2 se presentan varios parámetros de tecnologías disponibles hoy en día para la fabricación de circuitos integrados de propósito específico (ASICs). Los datos se han extraído en parte de Europractice [22].

Los datos presentados en la Tabla 4-2 corresponden a tecnologías disponibles para la fabricación de ASICs. Existen ya, o están a punto de aparecer, versiones comerciales de

tecnologías CMOS de  $0.35\mu\text{m}$ , por ejemplo la de MOSIS [23], disponible a partir del cuarto trimestre de 1997 o la ya existente de Toshiba de  $0.3\mu\text{m}$  y con dos extensiones para  $3/5\text{ V}$  y  $2/2.5\text{ V}$  [24], lista para producción a partir del segundo trimestre de 1997. En julio de 1997 NEC ha presentado una tecnología de  $0.25\mu\text{m}$  para la realización de ASICs [25].

Fabricante	Longitud mínima	Tensión de alimentación	$V_T$ (NMOS/PMOS)	Niveles de Metal	$T_{ox}$
Alcatel-Mietec	$0.5\mu\text{m}$	3.3 V	0.6 V/0.59 V	3	$100\text{ \AA}$
AMS	$0.6\mu\text{m}$	5 V y 3.3 V	(no disponibles)	2	
ATMEL-ES2	$0.5\mu\text{m}$	3.3 V	0.55 V/0.75 V	2	$120\text{ \AA}$

**Tabla 4-2**

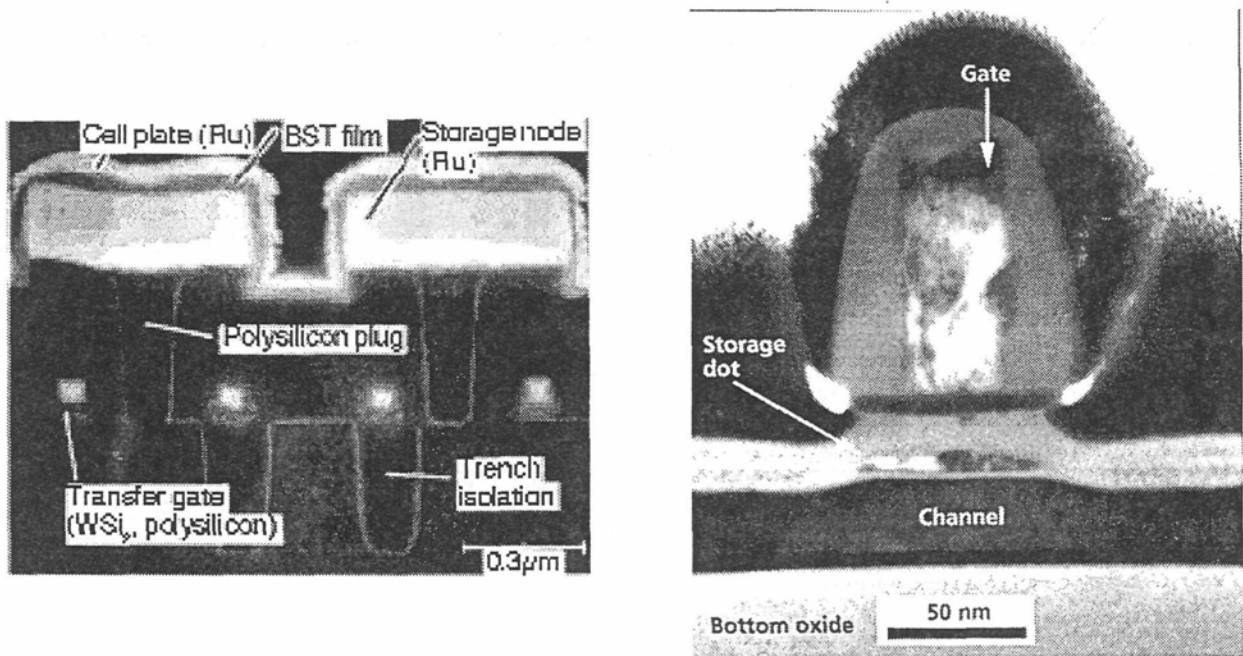
*Valores típicos de los principales parámetros para las últimas generaciones de tecnologías CMOS de tres fabricantes europeos.  $V_T$  es la tensión umbral de los transistores tipo N y P, respectivamente.  $T_{ox}$  es el grosor del óxido de puerta de los transistores.*

Para la fabricación de memorias el estado de la técnica está siempre un paso por delante y las innovaciones introducidas en este campo son aplicadas después a la fabricación de ASICs y otros CI. En la Tabla 4-3 se muestran datos sobre memorias DRAM (memorias RAM dinámicas), que son las más avanzadas, extraídos de la información que proporcionan sus fabricantes.

	Mitsubishi [26]	Toshiba [27]	Hitachi [28]
Capacidad	256 Mb	64 Mb	64 Mb
Tecnología	$0.25\mu\text{m}$	$0.35\mu\text{m}$	$0.35\mu\text{m}$
Tiempo de acceso	34 ns	40 ns/50 ns	60 ns/70 ns
Tensión alimentación	3.3 V externos, 2.5 V circuitos periférico y 2.0 V para la matriz de memoria	3.3 V+/-0.3 V	3.3 V+/-0.3V
Consumo (dinámico)	62 mA	100 mA	152 mA
Área (dado de silicio)	$304\text{ mm}^2$	$167.7\text{ mm}^2$	(no disponible)

**Tabla 4-3**

*Memorias DRAM de última generación disponibles en el mercado para tres de los más importantes fabricantes.*



**Figura 4-3**

(a) A la izquierda se presenta la microfotografía SEM de una celda de memoria fabricada para una DRAM de 1Gb de capacidad [31]. (b) En la derecha tenemos otra microfotografía de una celda de memoria de tipo flash, fabricada por J.J. Welser, Sandip Tiwari y sus colegas del centro de IBM Thomas J. Watson. La imagen está tomada de [29].

En el terreno de la investigación se están desarrollando técnicas que permitan implementar transistores con una longitud de canal menor de  $0.1 \mu\text{m}$ . El objetivo de estos trabajos se centra en obtener la memoria RAM dinámica (DRAM) de 1 Gbyte en los próximos años [29]. En el momento actual existen ya procesos experimentales que fabrican transistores con dimensiones de  $0.25 \mu\text{m}$ . Para implementar las máscaras en estos avanzados procesos se utilizan *steppers*<sup>2</sup> que usan longitudes de onda en el ultravioleta profundo de  $245 \text{ nm}$  de longitud de onda<sup>3</sup>. Las próximas generaciones se prevee que serán de  $0.18 \mu\text{m}$  y  $0.13 \mu\text{m}$  [29]. Para obtener dimensiones por debajo de  $0.1 \mu\text{m}$  será necesario pasar al espectro de rayos X [30][31], lo que introduce grandes dificultades técnicas pues no existe una óptica que pueda focalizar la luz en estas longitudes de onda. En la Figura 4-3 tenemos dos ejemplos de dispositivos experimentales con dimensiones por debajo de  $0.1 \mu\text{m}$ .

<sup>2</sup> Un *stepper* es la maquinaria que se utiliza para impresionar la oblea de silicio con el dibujo de las máscaras. El término proviene del proceso de impresión: se va repitiendo y desplazando el mismo patrón de máscaras para rellenar toda la oblea con un mosaico de circuitos iguales. La oblea se divide después de forma que cada uno de estos patrones repetidos forma un *chip*: un circuito integrado en un dado de silicio.

<sup>3</sup> La longitud de onda de la fuente de luz que se utiliza en el proceso de fotolitografía es un parámetro crucial de la técnica de fabricación de CI. La máxima resolución, y por tanto las dimensiones mínimas implementables, está estrechamente relacionada con la longitud de onda de fuente de luz que se utiliza para trasladar los patrones de las máscaras al silicio.



Las tendencias actuales de evolución pueden resumirse en varios aspectos fundamentales:

- Disminución progresiva de las dimensiones mínimas, aunque a un ritmo menor que el observado hasta ahora debido entre otras cosas a las limitaciones inherentes al proceso de fotolitografía. Actualmente existen tecnologías comerciales de hasta  $0.5\mu\text{m}$  para el diseño de ASIC y  $0.35\mu\text{m}$  en las memorias DRAM. Las siguientes generaciones serían  $0.25\mu\text{m}$ ,  $0.18\mu\text{m}$  y  $0.13\mu\text{m}$ . El objetivo futuro, para comienzos del siglo que viene, es bajar de  $0.1\mu\text{m}$ .
- Aumento progresivo del número de niveles de metalización, para soportar la creciente complejidad de los circuitos integrados en un solo chip. Volveremos más adelante sobre este tema cuando veamos el escalado de las interconexiones.
- Disminución de la tensión de alimentación, forzada no sólo por los requerimientos de bajo consumo de algunas aplicaciones sino por aspectos relacionados con la fiabilidad. Actualmente la reducción ha sido de 5 V a 3.3 V y 2.5 V, y esta tendencia seguirá en el futuro.
- Utilización de tecnologías del tipo SOI (del inglés *Silicon-on-Insulator*), que permiten eliminar algunos de los problemas que surgen en los nuevos dispositivos submicrónicos como las corrientes de fuga, el *latch-up* y los acoplamientos por el sustrato [32].
- Tendencia a integrar todo el sistema en un solo chip (*system-on-chip*), mediante técnicas MCM que incluyan memoria, microprocesadores, puertos de entrada/salida, módulos multimedia, etc., en un solo circuito integrado[33]. Inclusión de partes analógicas y digitales en un mismo circuito integrado (*mixed-mode* o *mixed-signal IC*) [20].

## 4.2 Modelos de escalado tecnológico CMOS

Los modelos de escalado tecnológicos pretenden representar de una forma cercana a la realidad la evolución de la tecnología. Estos modelos son muy útiles a la hora de observar posibles causas que limiten la evolución futura las prestaciones de los CIs y anticiparse a su aparición. Del mismo modo permiten predecir las diferentes características de los CIs en el futuro y analizar el efecto de las modificaciones introducidas en el diseño y la fabricación en las tendencias de escalado tecnológico.

La evolución de la tecnología puede describirse mediante la ley Moore [34], que predecía que el número de transistores por chip se doblaría cada año. Según [2] históricamente las tendencias en el escalado tecnológico se pueden resumir de la siguiente forma:

- Cada tres años se desarrolla una nueva generación tecnológica.
- Con cada nueva generación la densidad de memoria (por chip) se multiplica por cuatro y la densidad de los circuitos lógicos por un factor que va de dos a tres.
- En cada dos generaciones (seis años) las dimensiones mínimas de los dispositivos disminuyen un factor de dos, la densidad de corriente, la velocidad del circuito (ciclo

de reloj), el área del chip, la disipación de corriente del chip, y el máximo número de terminales de entrada/salida se multiplican todos por un factor de dos.

En la Tabla 4-4 se presentan los parámetros más importantes de la tecnología MOSFET y su evolución a lo largo de las décadas 70'-80'.

	1977	1979	1982	1985	1988	1991
Dimensiones mínimas ( $\mu\text{m}$ )	3	2	1.5	1	0.7	0.5
Longitud de la puerta ( $\mu\text{m}$ )	3	2	1.5	1.1	0.9	0.6
Longitud del canal ( $\mu\text{m}$ )	3	2	1.2	0.9	0.7	0.4
Grosor óxido de puerta: $t_{ox}$ (Å)	700	400	250	250	200	135
Profundidad de las uniones ( $\mu\text{m}$ )	0.6	0.4	0.3	0.25	0.2	0.15
Tensión de alimentación: $V_{dd}$ (V)	5	5	5	5	5	5
Corriente de saturación NMOS para tensión de puerta-surtidor $V_{gs} = 5\text{V}$ ( $\text{mA}/\mu\text{m}$ )	0.1	0.14	0.23	0.27	0.36	0.64
Corriente de saturación PMOS para tensión de puerta-surtidor $V_{gs} = 5\text{V}$ ( $\text{mA}/\mu\text{m}$ )		0.06	0.11	0.14	0.19	0.31
Retardo con un <i>Fan-Out</i> = 1 (ps)	800	350	250	200	160	90

**Tabla 4-4**

*Evolución del escaldado para la tecnología CMOS [2].*

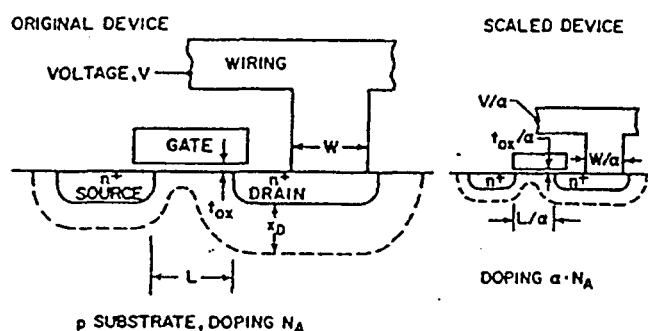
Los modelos clásicos de escalado se basan en esta tendencia observada por la tecnología, si bien se introducen modificaciones como el escalado de campo constante previendo dificultades en mantener los valores de algunos parámetros eléctricos como la tensión de alimentación. A partir de los 90' surgieron una serie de fenómenos asociados a la necesidad de superar la barrera submicrónica que han hecho que las predicciones de Moore y de los modelos clásicos derivados de las reglas citadas anteriormente no sean correctas. Han surgido límites impuestos por la fiabilidad que han modificado los criterios utilizados por las compañías de fabricación de CI para escalar algunos parámetros físicos y eléctricos, lo que ha hecho necesaria la aparición de modelos de escalado más cercanos a la nueva tendencia.

#### 4.2.1 Modelos de escalado clásicos: voltaje constante y campo constante

El parámetro de partida que utilizan todos los modelos es la reducción en las dimensiones mínimas (llamaremos  $\lambda$  al valor de este parámetro). El factor por el que se reducen estas dimensiones de una generación a otra se suele expresar mediante la letra griega  $\alpha$ , siendo  $\alpha$  siempre mayor que la unidad. Los dos modelos clásicos que se han venido utilizando se denominan escalado de voltaje constante (siglas inglesas CV) y campo constante (CF). En el primero se reducen todas las dimensiones geométricas por el mismo factor  $\alpha$ . Esto provoca que

el campo en el óxido aumente por un factor  $\alpha$ , aumento que tiende a degradar la fiabilidad de los dispositivos facilitando la ruptura del óxido de puerta. Para solventar este problema se utiliza como alternativa el segundo tipo de escalado, manteniendo el campo constante. Para ello es necesario escalar la tensión el mismo factor  $\alpha$  que las dimensiones físicas. Para evitar que las zonas de depleción de las uniones entre las regiones de drenador y de surtidor con el sustrato lleguen a tocarse es necesario reducir también, en ambos casos, el nivel de dopado del sustrato ( $N_s$ ), como queda ilustrado en la Figura 4-4. En la Tabla 4-5 se presentan los dos tipos de escalado clásicos. Llamamos parámetros primarios a los que dependen del diseño y la fabricación (dimensiones, tensiones, niveles de dopado, etc.) y parámetros secundarios a aquellos que resultan de los primarios y que determinan las prestaciones del circuito: las capacidades, los retardos, las corrientes máximas y el consumo, etc.

En el caso del escalado de campo constante el retardo de puerta, para el cálculo del cual se tienen únicamente en cuenta las capacidades de los dispositivos, no las de las interconexiones, se reduce sólo en un factor  $\alpha$ , mientras que en el escalado CV esta reducción es en un factor cuadrático  $\alpha^2$ . Esto explica la resistencia de la industria a reducir las tensiones de alimentación, como se observa en la Tabla 4-4, pues una de las razones que mueven la evolución tecnológica es la búsqueda de velocidades de cálculo cada vez mayores. Sin embargo, si el criterio es la reducción del consumo y el aumento de la fiabilidad, se puede observar como el escalado CF es claramente mejor que el CV. Para el cómputo del consumo se ha tenido en cuenta que, a pesar de la reducción de las dimensiones de los dispositivos, el nivel de integración (Nº de transistores/chip) va aumentando debido a que el área de los CIs queda igual o aumenta. Para representar este aumento de la complejidad de los CIs CMOS se ha utilizado el parámetro  $\alpha_C$ , que mide el aumento del tamaño del chip de una generación respecto de la anterior, y es un número siempre mayor que la unidad.



**Figura 4-4**

*Principios del escalado de campo constante (CF) para transistores MOS y circuitos integrados [35].*

La tendencia seguida por los fabricantes en la última década no ha sido ni un escalado CV ni un escalado CF, sino un caso intermedio entre ambos denominada por Robert H. Dennard escalado generalizado [35], y en el que la resistencia a cambiar las tensiones de alimentación introducen un nuevo parámetro de escalado llamado  $\epsilon$  ( $\epsilon > 1$ ) que expresa la reducción en la tensión de alimentación en un factor menos agresivo que las dimensiones geométricas. Otra de

las correcciones consiste en contemplar que la anchura de los transistores no se escale por igual que su longitud, debido a que el escalado de las interconexiones por razones de fiabilidad y del aumento de las corrientes que han de soportar tampoco se escalan tanto como la longitud del canal [32]. Se utiliza un parámetro  $\alpha_w$  ( $\alpha_w > 1$ ) que expresa el factor por el que se escalan las interconexiones y la anchura del canal de los transistores. Se ha de contemplar también que el óxido de puerta se escale de forma distinta a la longitud de canal. Expresamos este factor de escalado para el óxido de puerta como  $\alpha_g$ . Los diferentes parámetros asociados con este tipo de escalado se muestran en la Tabla 4-6.

Parámetros primarios	Escalado CF	Escalado CV
Anchura y longitud del canal ( $W$ y $L$ , respectivamente)	$1/\alpha$	$1/\alpha$
Grosor del óxido de campo y de puerta ( $F_{ox}$ y $t_{ox}$ , respect.)	$1/\alpha$	$1/\alpha$
Tensión de alimentación y tensión umbral ( $V$ y $V_p$ , respect.)	$1/\alpha$	1
Lado del dado de silicio (incremento de tamaño del chip)	$\alpha_c$	$\alpha_c$
Nivel de dopado del sustrato ( $N_s$ )	$\alpha$	$\alpha$
Parámetros secundarios		
Campo eléctrico ( $E_F$ )	1	$\alpha$
Capacidad de puerta ( $C_g$ )	$1/\alpha$	$1/\alpha$
Retardo de puerta, suponiendo un Fan-Out=1	$1/\alpha$	$1/\alpha^2$
Transconductancia ( $g_m$ )	1	$\alpha$
Consumo de potencia ( $P_{av}$ )	$\alpha_c^2$	$\alpha^3 \alpha_c^2$
Nº de transistores MOS/chip	$\alpha^2 \alpha_c^2$	$\alpha^2 \alpha_c^2$

**Tabla 4-5**

*Modelos clásicos de escalado de campo constante (CF) y voltaje constante (CV) y su efecto en las prestaciones de los circuitos integrados.*

Parámetros primarios	Factores de escalado
$L$	$1/\alpha$
$W$	$1/\alpha_w$
$t_{ox}$	$1/\alpha_g$
$F_{ox}, t_m$	$1/\alpha_t$
$V, V_t$	$\epsilon/\alpha$
$NS$	$\epsilon \cdot \alpha$
Lado del dado	$\alpha_c$
Parámetros secundarios	
$E_F$	$\epsilon$
$C_g$	$\alpha_g/(\alpha_w \cdot \alpha)$
Retardo de puerta	$1/(\epsilon \cdot \alpha)$
$g_m$	$\epsilon$
$P_{av}$	$(\epsilon^3 \cdot \alpha_c^2 \cdot \alpha_{g,t})/\alpha$
Nº de transistores MOS/chip	$\alpha \cdot \alpha_w \cdot \alpha_c^2$

Tabla 4-6

Escalado generalizado incluyendo diferentes factores de escalado para la anchura del canal, el grosor de los óxidos, y las tensiones.

### 4.2.2 Escalado y fiabilidad

A partir de los primeros 90' los aspectos de fiabilidad empezaron a dominar las tendencias en escalado de los CI CMOS. Además de los objetivos de aumentar la corriente de los dispositivos, para poder conmutar más rápidamente sus salidas, y reducir las dimensiones para incrementar la densidad de integración, los fabricantes se han visto obligados a tener en cuenta una serie de restricciones centradas en la disipación de potencia y la fiabilidad [2][36]:

1) Niveles aceptables de corriente de fugas cuando los transistores están en "off".

- Disminución de la tensión umbral ( $V_T$ ) debido a la corta longitud del canal, provocando el paso de corriente en la región subumbral del dispositivo aún cuando la tensión de puerta ( $V_g$ ) sea nula.
- Fugas en el drenador inducidas por la puerta (GIDL: *gate-induced drain leakage*)
- Ruptura del óxido por ionización de impacto debido a "portadores calientes" (*hot-carriers*).

2) Tiempos de vida debidos a fiabilidad y tasas de fallo aceptables

- Fiabilidad frente a efectos de “portadores calientes”. Puede mejorarse con estructuras de tipo LDD, bajando la tensión de alimentación, óxidos de puerta de alta calidad, etc.[37]
- Fiabilidad del óxido: ruptura del óxido dependiente del tiempo. Para asegurarla es necesario reducir el campo a través del óxido de puerta y disminuir la densidad de defectos del mismo [38].
- Fiabilidad de las metalizaciones. Puede mejorarse con metales de alto punto de fusión y niveles adicionales de metalización.

Todos estos fenómenos afectan al diseño de los dispositivos de las nuevas generaciones y hacen que las tendencias de escalado se aparten de los modelos clásicos citados en el apartado anterior. De entre estos parámetros de diseño destacamos los siguientes, comentando su efecto en los diferentes aspectos relacionados con la fiabilidad:

#### 4.2.2.1 Dimensiones físicas:

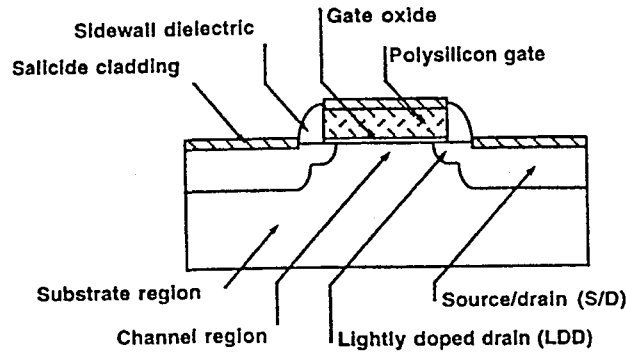
La reducción en las dimensiones físicas produce los llamados efectos de canal corto. Debido a la reducción de la distancia entre el drenador y el surtidor es necesario utilizar óxidos más delgados para conseguir uniones más estrechas de forma que el dispositivo siga estando controlado por la puerta y el campo del drenador no se extienda dentro del canal. Este efecto provoca la reducción de la tensión umbral y se conoce por las siglas inglesas DIBL (*drain induced barrier lowering*). La reducción de la tensión  $V_t$  aumenta las pérdidas cuando el transistor está en “off” incrementado el consumo estático. Otro fenómeno de canal corto consiste en el paso por efecto túnel de banda a banda (GIDL) inducido por el campo vertical en el solapamiento entre la puerta y el drenador. Ambos fenómenos se combaten reduciendo la tensión de alimentación y utilizando estructuras LDD [1](ver Figura 4-5), esto último a costa de incrementar la resistencia intrínseca de drenador y surtidor.

#### 4.2.2.2 Nivel de dopado del substrato:

El fenómeno de DIBL fija un límite inferior para el nivel de dopado del substrato que permita asegurar una separación adecuada entre las zonas de depleción de las uniones de drenador y surtidor.

#### 4.2.2.3 Tensión de alimentación:

La reducción de la tensión de alimentación tiene numerosos efectos beneficiosos en la fiabilidad, pero tiene también sus inconvenientes. Si se quieren mantener los mismos niveles de prestaciones en la velocidad y los márgenes de ruido es necesario escalar la tensión umbral  $V_t$  a la vez. Sin embargo esto aumenta las corrientes de pérdidas por conducción sub-umbral, como se ha visto antes. Además, como se ha demostrado en [35][39], el incremento de la corriente sub-umbral aumenta de forma exponencial con la reducción de  $V_t$ . Estas pérdidas fijan un límite para la reducción de la tensión umbral entorno a 0.3V, si no se recurre al uso de polarización del substrato o de tecnologías SOI [3][35]. Si aceptamos un valor de la tensión de alimentación cuatro veces superior a  $V_t$ , el límite inferior para  $V_{dd}$  es de 1.2V.



**Figura 4-5**

*Estructura típica de un dispositivo MOSFET con estructuras LDD (Light Doped Drain) en el drenador/surtidor. Estas regiones permiten que los campos eléctricos en los extremos del canal sean menos intensos, reduciendo los efectos de DIBL que afectan al control de la puerta sobre el canal.*

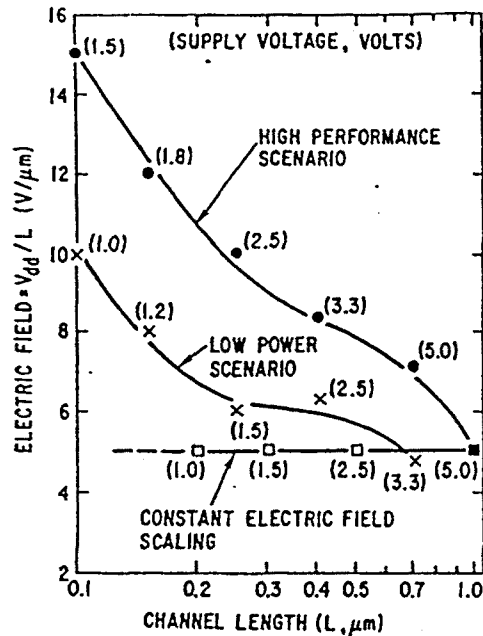
#### 4.2.2.4 Óxido de puerta:

La necesaria reducción en el óxido de puerta al escalar las dimensiones lo hace más vulnerable a los fenómenos de desgaste temporal, como los efectos de “portadores calientes”, y a los esporádicos, como las descargas electrostáticas (ESD) [40]. La introducción de trampas en la superficie del óxido o en su interior [41] degrada su calidad aumentando las pérdidas y pudiendo llegar a provocar su perforación, lo que daría lugar a fallos de tipo GOS (*gate oxide short*) muy difíciles de detectar [42][43]. La relación entre las tensiones aplicadas y el grosor del óxido de puerta fija un límite al campo en el óxido ( $E_{ox}$ ) que para un tiempo de vida de 10 años a 125°C deberá ser inferior a 4.24MV/cm [3]. Para sobrepasar este umbral será necesario utilizar otro tipo de dieléctricos y perfeccionar el proceso de crecimiento del óxido para asegurar el mínimo nivel posible de defectos en el mismo.

Además de los valores típicos para los que se ajustan los procesos de fabricación, es importante conocer cual puede ser la variación de estos parámetros con el proceso de fabricación. Según [44], los efectos de las fluctuaciones típicas del proceso de fabricación pueden ser hasta 600 veces más pronunciados en dispositivos profundamente submicrónicos, como los actuales y las siguientes generaciones, que en sus antecesores micrónicos. Los dos parámetros que más se ven afectados por estas fluctuaciones son el grosor del óxido de puerta y la longitud del canal, precisamente los que más influencia tienen en los aspectos de fiabilidad.

#### 4.2.3 Modelos para escenarios de bajo consumo y altas prestaciones

Como se demuestra en [35], se puede encontrar un valor óptimo para la tensión de alimentación limitada por los efectos de portadores calientes en un escenario donde se pretende obtener la máxima velocidad para unas dimensiones mínimas determinadas. Este valor es de 2.5V para una tecnología de 0.25 $\mu$ m, y va disminuyendo a medida que se reducen las dimensiones del canal. Sin embargo también es posible utilizar tensiones de alimentación más bajas sin que el retardo se vea excesivamente afectado.



**Figura 4-6**

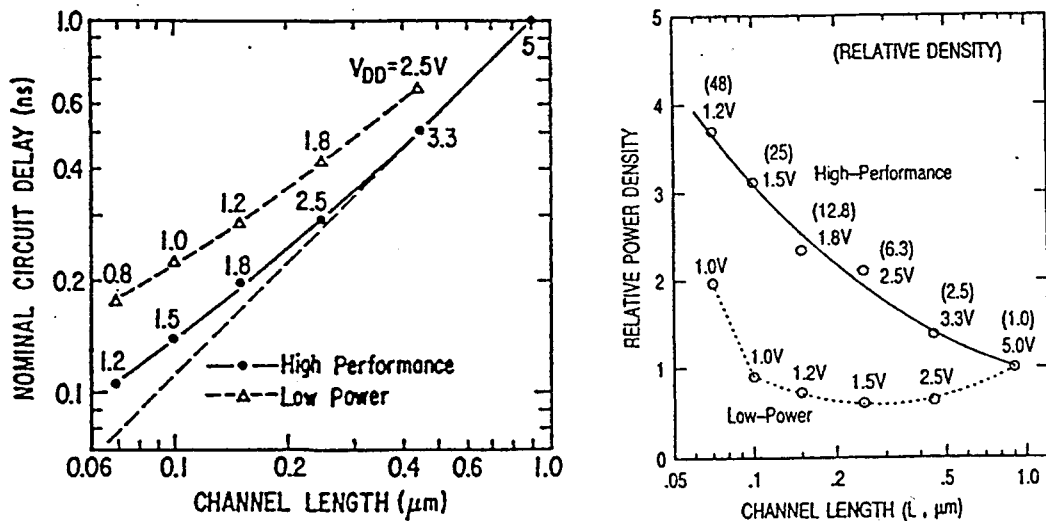
Representación gráfica del campo eléctrico para los dos escenarios de altas prestaciones (*high performance*) y bajo consumo (*low power*) en comparación con un escalado de campo constante. Se indican las dimensiones del canal y las tensiones de alimentación para cada tecnología [35].

Esta alternativa se traduce en dos estrategias distintas según prime la velocidad o el consumo. Llamaremos a estas dos tendencias escenarios de altas prestaciones o alta velocidad (cuando se busca obtener la máxima velocidad, y, por tanto, los mínimos retardos) y escenarios de bajo consumo (Figura 4-6). Basándose en estos requerimientos y los límites establecidos por los aspectos de fiabilidad diferentes autores han propuesto unas previsiones sobre el escalado futuro de los dispositivos [2][3][35][45]. Los valores para los distintos parámetros presentados en estos trabajos no difieren mucho entre sí. En la Tabla 4-7 se presentan los diferentes parámetros para los dos escenarios basados en [35]. El parámetro de partida es la resolución litográfica general ( $\lambda$ ). A nivel de puerta la resolución puede llegar a ser algo inferior. Los datos de la tabla corresponden a la dimensión más pequeña que podría dibujarse en el layout de los circuitos para la longitud del canal. La longitud efectiva del canal es siempre menor que la dibujada, y sus valores corresponden a la fila etiquetada "longitud del canal". Todos los valores de velocidad, consumo por unidad de área y densidad de integración (Nº de transistores por unidad de área) se expresan de forma relativa respecto la tecnología de  $1.25\mu\text{m}$  de los últimos 80's (primera columna de datos). En la Figura 4-7 se presentan dos gráficas con los retardos y la densidad de potencia para los dos escenarios de los que hemos hablado.



	Finales 80'	1992	1995	1998	2001	2004
Resolución litográfica ( $\mu\text{m}$ )						
General	1.25	0.8	0.5	0.35	0.25	0.18
A nivel de puerta : $L$ 's cortas	—	0.6	0.35	0.25	0.18	0.13
Longitud de canal ( $\mu\text{m}$ )	0.9	0.6/0.45	0.35/0.25	0.2/0.15	0.1	0.07
Grosor óxido puerta ( $t_{ox}$ ) (nm)	23	15/12	9/7	6/5	3.5	2.5
Densidad relativa	1.0	1.4/2.0	2.7/3.4	4.25/5.1	7.2	9.6
Escenario Altas Prestaciones						
Tensión de Alimentación	5	5/3.3	3.3/2.5	2.5/1.8	1.5	1.2
Velocidad relativa	1.0	1.4/2.0	2.7/3.4	4.2/5.1	7.2	9.6
Consumo/área relativo	1.0	2.25/1.38	3.0/2.1	3.7/2.34	3.12	3.7
Escenario Bajo Consumo						
Tensión de Alimentación	—	3.3/2.5	2.5/1.5	1.5/1.2	1.0	1.0
Velocidad relativa	—	1.0/1.6	2.0/2.4	3.2/3.5	4.5	7.2
Consumo/área relativo	—	0.7/0.63	1.25/0.6	1.02/0.72	0.92	1.97

**Tabla 4-7**  
*Guía del escalado para tecnologías CMOS según [35].*



**Figura 4-7**  
 (a) La figura de la izquierda corresponde al retardo nominal para una puerta NAND con un Fan-Out=1. La línea a trazos en la parte inferior de la gráfica representa el objetivo ideal de reducir el retardo linealmente con la longitud del canal. (b) La figura de la derecha muestra la densidad de potencia en función de la longitud del canal. Los números entre paréntesis representan la densidad de integración para cada generación tecnológica [35].

Como vemos en la Figura 4-7.(a) no es posible seguir la tendencia clásica de reducción del retardo linealmente con la reducción de la longitud del canal (indicada con una línea discontinua) debido a la necesidad de escalar las tensiones de alimentación por debajo de unos límites que aseguren una fiabilidad aceptable. En la Figura 4-7.(b) observamos como en el escenario de bajo consumo se tiende a mantener constante la densidad de potencia disipada por

el CI a pesar del aumento del nivel de integración para las diferentes longitudes de canal. Se observa también como para el escenario de altas prestaciones el incremento de consumo es proporcional a la reducción de la longitud del canal y al aumento de la densidad de integración.

#### 4.2.4 Escalado de las interconexiones

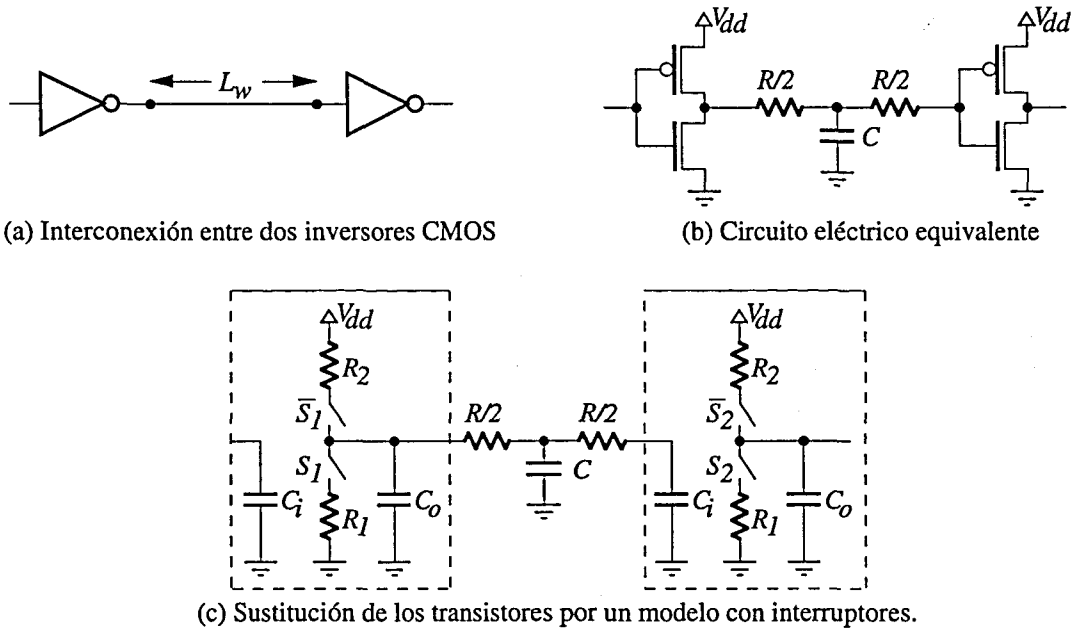
Las tendencias en el escalado de interconexiones están fijadas por dos aspectos bien diferenciados: el aumento de la complejidad de los CI y las necesidades de asegurar unos tiempos de vida y tasas de fallo en campo aceptables, es decir, asegurar unos niveles de fiabilidad adecuados. El primer aspecto implica aumentar la complejidad de la red de interconexiones, incrementando la longitud de las mismas y el número de niveles de metalización. La necesidad de integrar el circuito en un área reducida induce a disminuir la anchura y, por tanto, sección transversal de las interconexiones. Por otro lado la demanda en el incremento de la velocidad de conmutación implica un incremento de las corrientes de los nodos del circuito, corrientes que han de circular por unas interconexiones con una sección cada vez más pequeña. Esto implica que todos los fenómenos que degradan la calidad de las interconexiones, como la electromigración, que son proporcionales a la densidad de corriente ( $J$ ) ven aumentados sus efectos, reduciendo la fiabilidad. En la Tabla 4-8 se muestra la evolución de la densidad de corriente para los dos escenarios clásicos de escalado CF y CV y para el escalado más generalista de la Tabla 4-6. Podemos ver como para el caso del escalado generalizado el aumento es considerable, y mucho más para el caso CV. El aumento es algo menor en el escalado CF, escalado que, como se vio, no se sigue realmente debido al reducido beneficio que supone respecto al incremento de la velocidad de los CI:

	Escalado CF	Escalado CV	Escalado generalizado
Long. canal ( $L$ )	$1/\alpha$	$1/\alpha$	$1/\alpha$
Anchura canal e interconexiones ( $W$ )	$1/\alpha$	$1/\alpha$	$1/\alpha_w$
Grosor interconexiones ( $t_m$ )	$1/\alpha$	$1/\alpha$	$1/\alpha_m$
Tensiones ( $V$ )	$1/\alpha$	1	$\epsilon/\alpha$
Densidad de corriente ( $J$ )	$\alpha$	$\alpha^3$	$\alpha \cdot \alpha_m \cdot \epsilon$

**Tabla 4-8**

*Efectos del escalado en la densidad de corriente que han de soportar las interconexiones.*

El segundo factor a tener en cuenta en la evolución de las interconexiones es el retardo introducido por las mismas. En la Figura 4-8 se presenta una representación esquemática de dos inversores CMOS conectados con una línea de metal de longitud  $L_w$ . El modelo eléctrico de los transistores NMOS y PMOS para este caso será el de una capacidad de entrada debido a las puertas de ambos dispositivos ( $C_i$ ), una resistencia y un interruptor que modela los transistores ( $R_1$  para el NMOS y  $R_2$ , para el PMOS) y una capacidad de salida debida a las capacidades de las difusiones de drenador de ambos dispositivos ( $C_o$ ). El modelo para la línea es de elementos concentrados donde  $R$  y  $C$  son la capacidad y la resistencia por unidad de longitud. En el



**Figura 4-8**

*Circuito equivalente de una interconexión entre dos inversores CMOS.*

circuito de la Figura 4-8 se ha optado por representar la interconexión con un circuito en T con elementos concentrados. El retardo total incluyendo los elementos de los dispositivos y los de la interconexión, para una transición de nivel lógico alto a bajo del primer inversor, viene dado según la expresión de la Ecuación (4-1) calculada siguiendo el método de [46] y agrupando términos.

$$t_d = R_1(C_o + C_i) + (R_1C + RC_i)L_w + \frac{1}{2}RCL_w^2 \quad \text{Ec. (4-1)}$$

dónde  $t_d$  es el retardo total, y el resto de parámetros ya los definimos en el párrafo anterior.

El valor de la resistencia equivalente de los transistores NMOS,  $R_1$ , se calcula a partir de los parámetros del dispositivo y la tensión de alimentación [47]:

$$R_1 = \frac{2V_{DD}L_1}{K'W_1(V_{DD} - V_{iN})^2} \quad \text{Ec. (4-2)}$$

donde  $L_1, W_1$  son las dimensiones del transistor NMOS,  $K'$  la ganancia,  $V_{iN}$  la tensión umbral y  $V_{DD}$  la tensión de alimentación.

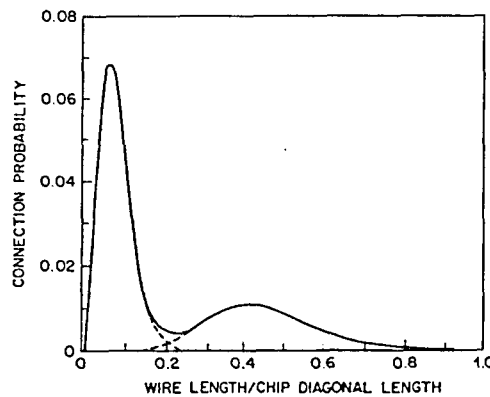
En la Ecuación (4-1) pueden observarse tres términos. El primero corresponde al retardo intrínseco debido a los dispositivos, el segundo es el retardo de carga y el tercero el retardo RC debido a la interconexión. En función de como se escalen las dimensiones de los dispositivos y de las interconexiones cada uno de los términos se escala de forma distinta. En la Tabla 4-9 se muestran las expresiones para el escalado de los tres términos del retardo para cada uno de los tres escenarios clásicos. En este caso se supone que el grosor de las interconexiones ( $t_m$ ) y el óxido de campo e intermedio ( $t_{Fox}$ ) se escalan siguiendo un factor distinto ( $\alpha_m$ ) de las

dimensiones de los dispositivos ( $L$  y  $W$ ). Por otro lado, la longitud de las interconexiones se escala según un factor llamado  $\alpha_{LW}$ .

Parámetro	Escalado CF	Escalado CV	Escalado Generalizado
$L, W, W_{interconexiones}$	$1/\alpha$	$1/\alpha$	$1/\alpha$
$L_{interconexiones}$	$\alpha_{LW}$	$\alpha_{LW}$	$\alpha_{LW}$
$V_{DD}, V_t$	$1/\alpha$	$1$	$\epsilon/\alpha$
$t_{ox}$	$1/\alpha$	$1/\alpha$	$1/\alpha$
$t_m$	$1/\alpha_m$	$1/\alpha_m$	$1/\alpha_m$
$t_{Fox}$ (grosor óxido de campo)	$1/\alpha_m$	$1/\alpha_m$	$1/\alpha_m$
$R_{NMOS, PMOS}$	$1$	$1/\alpha$	$1/\epsilon$
$C_i, C_o$	$1/\alpha$	$1/\alpha$	$1/\alpha$
$C$	$\alpha_m/\alpha$	$\alpha_m/\alpha$	$\alpha_m/\alpha$
$R$	$\alpha_m \alpha$	$\alpha_m \alpha$	$\alpha_m \alpha$
$t_D$ intrínseco	$1/\alpha$	$1/\alpha^2$	$1/\epsilon \alpha$
$t_D$ carga	$(\alpha_m/\alpha) \alpha_{LW}$	$(\alpha_m/\alpha^2) \alpha_{LW}$	$(\alpha_m/\epsilon \alpha) \alpha_{LW}$
$t_D$ línea	$\alpha_m^2 \alpha_{LW}^2$	$\alpha_m^2 \alpha_{LW}^2$	$\alpha_m^2 \alpha_{LW}^2$

**Tabla 4-9**

Escalado de los tres términos del retardo de la Ecuación 4-1 para los escenarios clásicos CF, CV y Generalizado.



**Figura 4-9**

Distribución bimodal de la longitud de las interconexiones para un CI CMOS típico. El eje de las abscisas representa la longitud de la interconexión relativa a la diagonal del dado de silicio [16].

Un aspecto importante es la evolución que ha seguido la longitud media de las interconexiones en los CI [48]. Podemos distinguir entre dos tipos distintos de señales dentro del circuito. Por un lado tenemos señales que intercambian puertas o bloques cercanos entre sí. Llamaremos **interconexiones locales** a las líneas de metal que transportan estas señales. Por

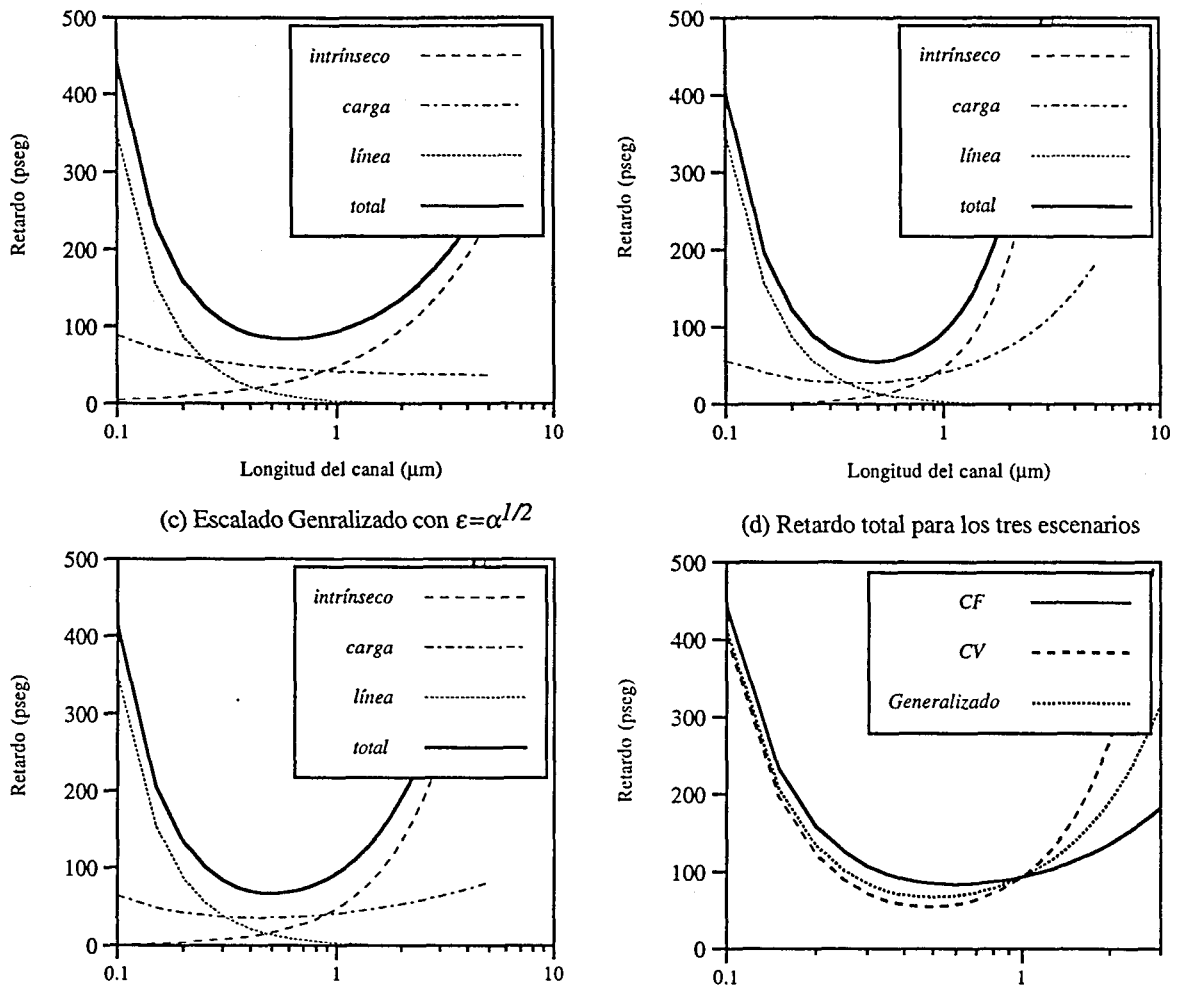
otro lado están las señales globales a todo el circuito, como el reloj, o que deben distribuirse por todo el integrado como las alimentaciones o señales de carácter global. Las interconexiones utilizadas para distribuir estas señales se conocen como **interconexiones globales**. En la Figura 4-9 se presenta un histograma con la distribución de longitudes en un CI típico extraído de [16], donde podemos ver el carácter bimodal de dicha distribución. El escalado de estos dos tipos de interconexiones es radicalmente distinto. Las interconexiones locales ven escalada su longitud de la misma forma que se reducen las dimensiones geométricas de los dispositivos, ya que el tamaño de las puertas y los bloques y las distancias que los separan son más pequeñas. Sin embargo, las interconexiones globales se escalan siguiendo el aumento del nivel de integración, pues el tamaño total del CI crece a pesar de que las dimensiones geométricas de los dispositivos y la anchura de las interconexiones se reducen con el escalado[1][48]. Podemos encontrar varios modelos que permiten expresar la longitud media de las interconexiones globales en función de parámetros como el área del chip, el número de entradas y salidas y el número de puertas del circuito [16][49], y en resumen vienen a establecer que la longitud media es proporcional a la raíz cuadrada del área del chip.

$$L_{max} = \frac{\sqrt{A_{chip}}}{2} \quad Ec. (4-3)$$

En los circuitos actuales y en las futuras generaciones el retardo total está dominado por el retardo debido a las interconexiones [1][3][16], como puede observarse en la Figura 4-10 donde se ha representado la Ecuación (4-1) y cada uno de sus términos para los tres escenarios clásicos presentados anteriormente. En el escenario generalizado se ha escogido un valor de  $\varepsilon = \alpha^{1/2}$ . En este estudio se ha mantenido la longitud de la interconexión constante, con una longitud de 2000 $\mu\text{m}$ . Se ha utilizado un circuito como el de la Figura 4-8, escalando las dimensiones y las tensiones a partir de los valores de una tecnología comercial de 1 $\mu\text{m}$ , que se listan en la Tabla 4-10. Las dimensiones del grosor ( $t_m$ ) de la línea de interconexión y del óxido entre la línea y el sustrato ( $t_{Fox}$ ) se han escalado igual que las dimensiones de los dispositivos y la anchura de la interconexión:  $\alpha_w = \alpha_m = \alpha$ .

Como puede observarse en la Figura 4-10, reducir el grosor de las interconexiones tiene efectos negativos en el retardo y también, como vimos anteriormente, en la fiabilidad de las metalizaciones debido al incremento de la densidad de corriente. La solución extrema sería mantener constante la sección transversal anchura×grosor ( $\alpha_w = \alpha_m = 1$  para las interconexiones). Pero esto limita el nivel de integración, pues el espaciado entre interconexiones (*interconnect pitch*) no podría reducirse. Otra alternativa es reducir la anchura pero mantener el grosor. Con esta tendencia se iría incrementando el área lateral de las líneas de metal y reduciéndose la separación, lo que aumentaría el acoplamiento entre líneas paralelas y se reduciría el acoplamiento con el sustrato. Esta estrategia hace aumentar el nivel de acoplamiento de ruido  $dV/dt$  entre interconexiones (*crosstalk* en el mismo nivel de metalización) [50] y el término dominante de la capacidad al sustrato corresponde a los campos de tipo *fringing*. Esta capacidad no se escala al reducir la anchura de las interconexiones pues depende del área lateral, y por tanto del grosor de las líneas de metal. Por ejemplo, en [51] se reporta que el incremento

del acoplamiento entre interconexiones (*cross coupling*) aumenta un 40% en un escalado de una tecnología de  $0.6\mu\text{m}$  a otra de  $0.35\mu\text{m}$ .



**Figura 4-10**

Escalado del retardo entre dos inversores conectados mediante una línea de metal de  $2000\mu\text{m}$  de longitud. Todas las dimensiones geométricas se han escalado linealmente tomando como partida los parámetros para una longitud de canal  $L=1\mu\text{m}$ . Las gráficas (a),(b) y (c) presentan los resultados para los tres escenarios CF, Cv y generalizado de cada término del retardo (intrínseco, carga y línea) y de la suma de los tres (total). La cuarta gráfica (d) presenta una comparación entre los valores del retardo total de las otras tres gráficas.

El incremento de la longitud de las interconexiones globales, como se muestra en [52] y en la Tabla 4-11, hace aumentar el retardo cuadráticamente con  $\alpha$  y cuadráticamente con  $\alpha_c$  (que representan la disminución de las dimensiones y el aumento del nivel de integración, respectivamente). Esto es cierto suponiendo que la anchura y el grosor de estas interconexiones se escalan linealmente con el resto de dimensiones de la tecnología

Parámetro	Valor
Longitud canal NMOS y PMOS ( $L$ )	$1\mu\text{m}$
Anchura canal NMOS/PMOS ( $W_n/W_p$ )	$100\mu\text{m}/300\mu\text{m}$
Anchura interconexión ( $W_{intercon}$ )	$10\mu\text{m}$
Grosos de óxidos y de metal ( $t_{ox}/t_{Fox}/t_m$ )	$200\text{Å}/1.16\mu\text{m}/0.44\mu\text{m}$
Capacidad de entrada y salida ( $C_i/C_o$ )	$0.44\text{pF}/0.33\text{pF}$
Tensiones ( $V_{DD}/V_t$ )	$5\text{V}/0.7\text{V}$
Longitud interconexión ( $L_{intercon}$ ): constante	$2000\mu\text{m}$

**Tabla 4-10**

Parámetros de partida para el escalado del retardo correspondientes a una tecnología comercial de  $1\mu\text{m}$ . Todos los parámetros se escalan siguiendo los modelos clásicos excepto la longitud de la interconexión que se mantiene constante.

Parámetros de las interconexiones	Factor de escala
Dimensiones de la interconexión global: $t_{Fox}$ , $t_m$ , $W_{intercon}$	$1/\alpha$
Longitud: $L_{intercon\_global}$	$\alpha_c$
Resistencia de la línea ( $R = \rho L_{intercon\_global}/t_m W_{intercon}$ )	$\alpha^2 \alpha_c$
Capacidad de la línea ( $C = \epsilon_{SiO_2} L_{intercon\_global} W_{intercon}/t_m$ )	$\alpha_c$
Tiempo de respuesta de la línea: $0.89RC$	$\alpha^2 \alpha_c^2$

**Tabla 4-11**

Escalado del tiempo de respuesta para las interconexiones globales.

Ante todas estas perspectivas (restricciones debido a la fiabilidad de la máxima densidad de corriente y aumento del retardo con el escalado, especialmente en las interconexiones globales) la industria de semiconductores ha seguido diferentes estrategias [35][48][50], que podemos resumir en:

- Utilizar varios niveles de metalización con relaciones anchura/grosor distintos en cada nivel. Introducir planos de masa entre niveles de metalización utilizados para líneas de señal para facilitar un camino de retorno de las corrientes de conmutación, aumentando la capacidad a tierra y por tanto reduciendo el acoplamiento lateral.
- Utilizar los niveles inferiores con una separación mínima entre líneas para interconexiones locales lo que permite un elevado nivel de integración, y los niveles superiores con una anchura, un grosor y una separación mínima mayores para distribuir señales globales como el reloj o las alimentaciones. De esta forma se cumplen los objetivos de aumento de integración y disminución del retardo para señales críticas globales, como el reloj, simultáneamente.
- Utilizar varias capas en forma de sandwich con diferentes metales de elevado punto de fusión para evitar fenómenos como la electromigración. Tendencia a utilizar metales con menor resistividad como el cobre o el oro en vez del aluminio. Planarizar el óxido antes de depositar

cada capa de metalización para evitar escalones que degradan la fiabilidad de las interconexiones. Utilizar vías de material refractario (como el wolframio) para interconectar los diferentes niveles entre sí.

### 4.3 Efectos de canal corto en dispositivos submicrónicos

Para analizar los efectos del escalado en el ruido de conmutación es necesario conocer el comportamiento real de los dispositivos fabricados con las tecnologías actuales. En estos dispositivos el comportamiento de la tensión umbral, la corriente de saturación o de la movilidad, por poner tres ejemplos, se apartan del modelo clásico de Schichman-Hodges, exhibiendo comportamientos de segundo orden. En este apartado veremos cual es la influencia de estos efectos debidos a las reducidas dimensiones del canal y a los elevados campos que se generan en los dispositivos en los parámetros eléctricos necesarios para comprender las características de conmutación de los circuitos de transistores CMOS.

#### 4.3.1 Reducción de la tensión umbral

Como ya explicamos anteriormente, la reducción en las dimensiones del canal de los dispositivos provoca que las regiones de deplexión creadas en las uniones del drenador y el surtidor con el sustrato se aproximen corriendo el peligro de entrar dentro de la región de canal. De esta forma el control sobre el canal deja de ser ejercido por la tensión de puerta y pasa a tener influencia de la tensión del drenador (efecto DIBL). El resultado eléctrico de este fenómeno es una reducción de la tensión umbral en función de la tensión del drenador. Este efecto es más importante cuanto menor es la longitud del canal [32][53].

Sin embargo el efecto más importante de la reducción del canal es el incremento de la corriente de fugas cuando el transistor esta en "off". Por ello, no tendremos este efecto en cuenta a lo hora de analizar el ruido de conmutación en circuitos CMOS.

#### 4.3.2 Reducción de la movilidad debido al campo eléctrico

La movilidad de los portadores es un parámetro que expresa la facilidad con la que éstos se mueven impulsados por un campo eléctrico entre la red cristalina del semiconductor. En el modelo clásico se considera la movilidad una constante del material que sólo depende de la temperatura. Sin embargo, en los dispositivos reales se ha de tener en cuenta que la movilidad depende del campo transversal en el canal debido a la puerta. Uno de los modelos empíricos más ampliamente usados [54] relaciona la movilidad efectiva con la tensión de puerta a través de una constante  $\theta$  que depende del proceso y de la polarización del sustrato:

$$\mu_{eff} = \frac{\mu_o}{1 + \theta(V_g - V_t)} \quad \text{Ec. (4-4)}$$



dónde  $\mu_o$  es la movilidad para campo nulo, y  $V_g$  y  $V_t$  son las tensiones de puerta y umbral, respectivamente. La movilidad efectiva ( $\mu_{eff}$ ) es menor cuanto mayores son los campos en la puerta. Esta variación en la movilidad afecta a las corrientes y por tanto a la velocidad de los circuitos CMOS. Existen modelos algo más precisos y que permiten introducir otros parámetros de ajuste como el grosor del óxido de puerta que podemos encontrar en [41].

### 4.3.3 Saturación de la velocidad de los portadores

La velocidad de arrastre (*drift velocity*) de los portadores, aquella que es proporcional al campo eléctrico longitudinal que se aplica en el canal, se satura para valores elevados de dicho campo. Existen dos clases de modelos. El primero considera la velocidad proporcional al campo eléctrico hasta que éste alcanza un valor por encima del cual la velocidad se considera constante:

$$v = \begin{cases} \frac{\mu_{eff} E}{1 + (E/E_{sat})} & E < E_{sat} \\ v_{sat} & E \geq E_{sat} \end{cases} \quad Ec. (4-5)$$

donde  $\mu_{eff}$  es la movilidad efectiva,  $E$  el campo eléctrico longitudinal,  $E_{sat}$  el valor del campo de saturación y  $v_{sat}$  el valor de la velocidad de saturación. El otro modelo evita la discontinuidad, pero complica mucho la ecuación de transporte que se ha de utilizar después para calcular la corriente de drenador ( $I_D$ ) de los transistores MOS:

$$v = \frac{\mu_{eff} E}{[1 + (E/E_{sat})^{n_v}]^{1/n_v}} \quad Ec. (4-6)$$

donde  $n_v$  es un exponente que vale 2 para los electrones y es menor que 2 para los huecos. Este último modelo empírico se ajusta muy bien a los datos experimentales[55]. Podemos utilizar también una versión simplificada de este modelo fijando  $n_v=1$ , con lo que reducimos la complejidad a costa de subestimar la velocidad para campos moderados y altos.

El valor del campo para el que la velocidad de los electrones se satura debido a la dispersión en el canal por el aumento del campo eléctrico está entorno a los  $10^4$  V/cm [56]. La velocidad de saturación para los electrones es, por encima de este campo eléctrico, de  $8 \cdot 10^6$  cm/seg. La velocidad de saturación para los huecos es ligeramente inferior:  $6.5 \cdot 10^6$  cm/seg. Utilizando que el campo de saturación puede expresarse en función de la velocidad de saturación como  $E_{sat} = v_{sat} / \mu_{eff}$  y a partir del modelo simplificado con  $n_v=1$ , encontramos una expresión para la velocidad de saturación más adecuada para los posteriores cálculos que realizaremos:

$$v = \frac{\mu_{eff} E}{1 + \frac{\mu_{eff}}{v_{sat}} E} \quad Ec. (4-7)$$

en la que la velocidad depende de la movilidad efectiva, una constante (la velocidad de saturación) y del campo eléctrico longitudinal en el canal.

Una vez la velocidad de los portadores no depende del campo eléctrico longitudinal se pierde la influencia de la reducción de la longitud del canal en el aumento de la corriente de los transistores MOS. Este efecto de saturación de la velocidad implica que la saturación de la corriente ya no se produce por el estrangulamiento del canal si no por superar el campo eléctrico de saturación a una tensión drenador-surtidor ( $V_{DS}$ ) inferior y por tanto el valor de la corriente de saturación ( $I_{DSAT}$ ) es menor del esperado. Debido al escalado los campos en el canal son cada vez mayores; los efectos de reducción de la movilidad, vistos en el apartado anterior, y de saturación de la velocidad de los portadores limitan el aumento de la corriente de saturación máxima, por lo que el incremento de velocidad real es inferior al esperado si no se consideran estos efectos de canal corto. Veremos en el apartado siguiente como afectan estos fenómenos a las corrientes y los tiempos de conmutación de los transistores MOS submicrónicos.

#### 4.3.4 Características de conmutación de los dispositivos submicrónicos

El fenómeno del ruido de conmutación está estrechamente relacionado con la característica de conmutación de los dispositivos. Determinar las expresiones de los modelos para dispositivos submicrónicos permite realizar una aproximación analítica al ruido de conmutación, como se verá más adelante

En la teoría estándar las expresiones que se encuentran para la corriente de drenador en las dos regiones de conducción (zona óhmica y zona de saturación) son, respectivamente:

$$\begin{aligned}
 I_D &= \frac{W}{L} C_{ox} \mu \left( V_G - V_t - \frac{V_D}{2} \right) V_D & 0 < V_D < V_G - V_t (= V_{Dmax}) \\
 I_D &= \frac{W}{2L} C_{ox} \mu (V_G - V_t)^2 & V_D > V_G - V_t
 \end{aligned}
 \tag{Ec. (4-8)}$$

donde  $I_D$  es la corriente de drenador,  $W$  y  $L$  las dimensiones del canal,  $C_{ox}$  la capacitancia por unidad de área del óxido de puerta,  $\mu$  la movilidad,  $V_G$  y  $V_D$  las tensiones de puerta y drenador referidas al surtidor, respectivamente, y  $V_t$  la tensión umbral. En estas expresiones el límite entre las zonas óhmica y de saturación lo marca  $V_{Dsat} = V_G - V_t$ . La dependencia de la corriente máxima de saturación con la tensión, cuando  $V_G = V_{DD}$ , es cuadrática.

Si incluimos los efectos de canal corto las expresiones para la corriente de drenador pasan a ser [57]:

$$I_D = \left( \frac{\mu_{eff} C_{ox} W}{L} \right) \frac{(V_G - V_t) V_D - (1/2) V_D^2}{1 + \frac{\mu_{eff}}{L v_{sat}} V_D} \quad 0 < V_D < V_{Dmax} \quad Ec. (4-9)$$

$$I_D = I_{Dsat} = \frac{1}{2} \left( \frac{\mu_{eff} C_{ox} W}{L} \right) \cdot V_{Dmax}^2 \quad V_D > V_{Dmax}$$

donde la movilidad efectiva ( $\mu_{eff}$ ) se calcula según se explicó en el apartado 4.3.2 y la tensión máxima de drenador se expresa como:

$$V_{Dmax} = \left( \frac{L v_{sat}}{\mu_{eff}} \right) \left[ \sqrt{1 + 2 \frac{\mu_{eff}}{L v_{sat}} (V_G - V_t)} - 1 \right] \quad Ec. (4-10)$$

que depende, como ya vimos anteriormente, de la velocidad de saturación  $v_{sat}$ .

Los efectos de canal corto se tienen en cuenta al calcular la movilidad efectiva y en la relación de ésta con el producto  $L v_{sat}$ . Los dos extremos son: si el cociente  $\mu_{eff}/L v_{sat} \ll 1$  la saturación está controlada por el estrangulamiento del canal como en los dispositivos de canal largo y la expresión para  $I_{Dsat}$  es la de la ecuación (4-8). Si  $\mu_{eff}/L v_{sat} \gg 1$ , el efecto dominante es la saturación de la velocidad de los portadores, y la expresión para la corriente de saturación queda:

$$I_{Dsat} = W C_{ox} v_{sat} (V_G - V_t) \quad Ec. (4-11)$$

donde vemos que, al contrario que en el otro extremo, la dependencia con la tensión de control es lineal. Además la dependencia con la longitud del canal ha desaparecido con lo que el posterior escalado de ésta no incrementaría la corriente de saturación. En dispositivos submicrónicos y profundamente submicrónicos la dependencia de la corriente de saturación con la tensión se encuentra entre estos dos extremos con un exponente entre 1 y 2. A medida que los efectos de canal corto dominan el funcionamiento del dispositivo el exponente se aproxima a 1. Esto tiene una gran influencia en el retardo intrínseco y en los tiempos de subida y bajada de las puertas CMOS como veremos a continuación.

Supongamos que a la salida de un inversor CMOS tenemos una capacidad de valor  $C_L$ , supongamos también que la transición a la entrada es lo suficientemente rápida como para que la tensión de puerta sea  $V_{DD}$  durante toda la transición de salida (ver en [58] las condiciones necesarias para que se cumpla esta suposición). En estas circunstancias si definimos el retardo como el tiempo que la salida tarda en alcanzar el valor  $V_{DD}/2$ , podemos encontrar una expresión para el retardo sin considerar los efectos de canal corto, es decir, a partir de la ecuación (4.8). Consideremos el caso de una transición alto-bajo a la salida en la que el NMOS descarga la capacidad  $C_L$ . Mientras  $V_D(t)$  sea mayor que  $V_{Dsat} = V_{DD} - V_{tN}$  tendremos la descarga de un

condensador a través de una fuente de corriente de valor la corriente de saturación. Cuando rebasemos el límite de la zona de saturación la corriente depende de la tensión de drenador (la tensión del condensador) y hemos de resolver una ecuación diferencial. Igualando la solución para el caso de que  $V_D(t)=V_{DD}/2$  encontramos la expresión del retardo para dispositivos de canal largo:

$$T_D = \underbrace{\left( \frac{C_L}{\beta_N (V_{DD} - V_{iN})} \right)}_{t_o} \left\{ 2 \frac{v_{iN}}{1 - v_{iN}} + \ln(3 - 4v_{iN}) \right\} \quad \text{Ec. (4-12)}$$

donde  $\beta_N = \mu(W/L)C_{ox}$  es la ganancia del NMOS,  $V_{iN}$  su tensión umbral y  $v_{iN} = V_{iN}/V_{DD}$ . Llamamos al primer factor de la expresión  $t_o$ .

El cálculo para el caso de dispositivos de canal corto es algo más complejo. En primer lugar hay que considerar dos casos, según sea o no  $V_{Dmax} > V_{DD}/2$ . En vez de trabajar directamente con la tensión de drenador máxima lo haremos con otro parámetro al que llamaremos  $V_o = Lv_{sat}/\mu_{eff}$ . Cuando  $V_{Dmax} = V_{DD}/2$  nos encontramos en el límite entre los dos casos y el valor de  $V_o = V_{oC}$  se toma como referencia. Despejando de la ecuación (4-10) con  $V_{Dmax} = V_{DD}/2$  y  $V_G = V_{DD}$  encontramos que:

$$V_{oC} = \frac{1}{1 - 2v_{iN}} \left( \frac{V_{DD}}{4} \right) \quad \text{Ec. (4-13)}$$

donde todas las variables son ya conocidas. En función del valor de  $V_o$  correspondiente a otros valores de  $V_{Dmax} \neq V_{DD}/2$  tenemos dos casos:

1) Si  $V_o < V_{oC}$  (dispositivo fuertemente dominado por la velocidad de saturación), la tensión de drenador  $V_D(t)$  viene dada por:

$$V_D(t) = V_{DD} - \frac{I_{Dsat}}{C_L} t \quad \text{Ec. (4-14)}$$

y el retardo se encuentra despejando  $t$  para  $V_D(t) = V_{DD}/2$  y sustituyendo  $I_{Dsat}$  por la ecuación (4-9) correspondiente a la condición de saturación:

$$T_D = \frac{1}{2} \left( \frac{C_L}{\beta_N V_{DD}} \right) \frac{1}{v_o^2} \frac{1}{\left[ \sqrt{1 + 2 \left( \frac{1 - v_{iN}}{v_o} \right) - 1} \right]^2} \quad \text{Ec. (4-15)}$$

donde  $v_o = V_o/V_{DD}$ .

2) Si  $V_o > V_{oC}$  (dispositivo débilmente dominado por la velocidad de saturación), tenemos dos regiones ya que el NMOS está en saturación hasta que se alcanza el instante temporal  $t_S$ , donde se satisface que  $V_D(t_S) = V_{Dmax}$ . El tiempo  $t_S$  se encuentra como:

$$t_S = \left( \frac{C_L}{\beta_N V_{DD}} \right) \frac{2(1 - v_{Dmax})}{v_{Dmax}^2} \quad \text{Ec. (4-16)}$$

donde  $v_{Dmax} = V_{Dmax}/V_{DD}$ . Cuando  $V_o \rightarrow \infty$ ,  $V_{Dmax} \rightarrow V_{DD} - V_{iN}$ , es decir, se tiende al modelo de canal largo. Después del tiempo  $t_S$ ,  $V_D$  satisface la ecuación diferencial:

$$C_L \frac{dV_D}{dt} = - \frac{\beta_N}{1 + \frac{V_D}{V_o}} \left( V_{DD} - V_{iN} - \frac{1}{2} V_D \right) V_D \quad \text{Ec. (4-17)}$$

Esta ecuación se resuelve por separación de variables. La condición inicial es que  $V_D = V_{Dmax}$  para  $t = t_S$  y la solución es:

$$t - t_S = \left( \frac{C_L}{\beta_N V_{DD}} \right) \Phi_1(v_D, v_o, v_{Dmax}, v_{iN}) \quad \text{Ec. (4-18)}$$

dónde todo es conocido y la función  $\Phi_1$  es:

$$\begin{aligned} & \Phi_1(v_D, v_o, v_{Dmax}, v_{iN}) \\ &= \frac{1}{1 - v_{iN}} \ln \left[ \frac{(2(1 - v_{iN}) - v_D) v_{Dmax}}{(2(1 - v_{iN}) - v_{Dmax}) v_D} \right] + \left( \frac{2}{v_o} \right) \ln \left[ \frac{(2(1 - v_{iN}) - v_D)}{(2(1 - v_{iN}) - v_{Dmax})} \right] \end{aligned} \quad \text{Ec. (4-19)}$$

El retardo se obtiene igualando  $v_D = V_D/V_{DD} = 1/2$  en la ecuación (4-18). Cuando  $v_o \rightarrow \infty$ , la ecuación (4-18) tiende a la expresión del retardo para dispositivos de canal largo, como la ecuación (4-12). Por el contrario si  $v_o \rightarrow 0$ , el retardo es proporcional a  $1/v_o$ .

El retardo para el caso de una transición de nivel Bajo a Alto a la salida se calcula de forma análoga substituyendo los parámetros del transistor NMOS por los del PMOS.

El cálculo de los tiempos de subida y bajada de la salida se realiza de forma similar, cambiando únicamente los límites de la integración. Éstos se definen como el tiempo que tarda la salida en subir del 10% de su valor final al 90% o de bajar del 90% al 10%. La expresión para dispositivos de canal largo del tiempo de bajada es:

$$t_f = t_o \left( \frac{2v_{iN} - 0.2}{1 - v_{iN}} + \ln(19 - 20v_{iN}) \right) \quad \text{Ec. (4-20)}$$

donde  $t_o = C_L/\beta_N(V_{DD} - V_{iN})$  como vimos en la ecuación (4-12). Para el dispositivos de canal corto volvemos a tener dos casos, según la tensión  $V_{Dmax}$  sea o no menor que un 10% de  $V_{DD}$ .

Si  $V_{Dmax} < 0.1V_{DD}$  durante todo el tiempo de bajada el transistor NMOS se comporta como una fuente de corriente cuyo valor viene dado por la ecuación (4-9) para  $I_{Dsat}$ . El tiempo de bajada será:

$$t_f = \left( \frac{C_L}{\beta_N V_{DD}} \right) \frac{1.6}{v_o^2 \left[ \sqrt{1 + 2 \frac{1 - v_{iN}}{v_o}} - 1 \right]^2} \quad \text{Ec. (4-21)}$$

En caso de que  $0.1V_{DD} < V_{Dmax} < 0.9V_{DD}$  durante un tramo del tiempo de bajada el NMOS trabajará en saturación y luego en zona lineal. En este caso el tiempo de bajada se encuentra de forma similar al tiempo de retardo pero cambiando los límites. La expresión es muy compleja y muestra las mismas dependencias con los parámetros que los del tiempo de retardo.

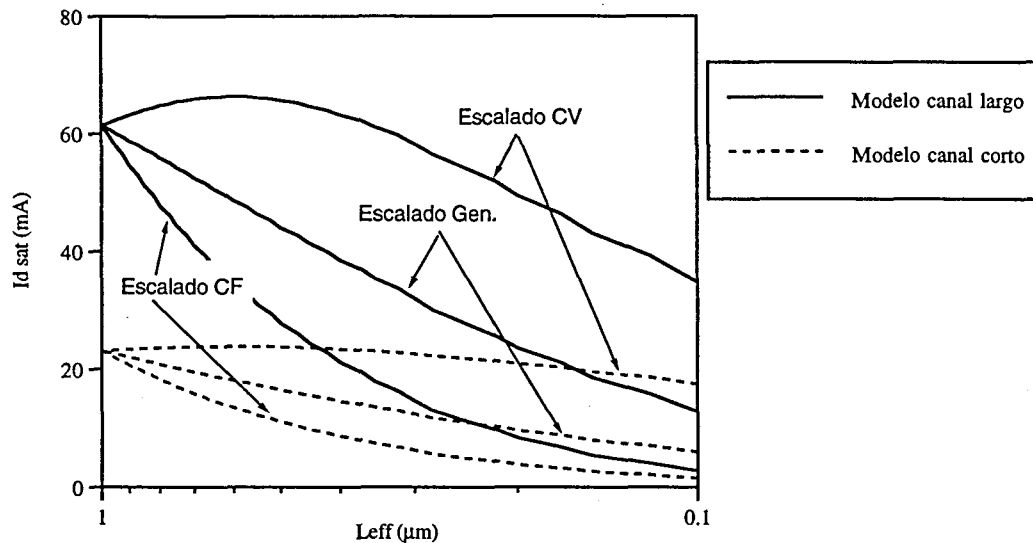
$$t_f = 2 \left( \frac{C_L}{\beta_N V_{DD}} \right) \left[ \frac{1 - v_{Dmax}}{v_{Dmax}^2} + \frac{\ln \left( \frac{(19 - 20v_{iN})v_{Dmax}}{2(1 - v_{iN}) - v_{Dmax}} \right)}{2(1 - v_{iN})} + \frac{\ln \left( \frac{(1.9 - 2v_{iN})}{2(1 - v_{iN}) - v_{Dmax}} \right)}{v_o} - \frac{0.1}{v_o^2 \left[ \sqrt{1 + 2 \frac{1 - v_{iN}}{v_o}} - 1 \right]^2} \right] \quad \text{Ec. (4-22)}$$

El cálculo para el tiempo de subida se realiza de forma análoga en todos los casos pero substituyendo los parámetros por los del transistor PMOS.

El interés de este análisis se centra en conocer cual será el comportamiento de la corriente máxima de conmutación y de los tiempos de subida y bajada con el escalado pues de esta forma podremos prever como se escalará el ruido de conmutación  $dI/dt$ . En la Figura 4-11 hemos representado el valor de la corriente máxima de saturación para los tres escalados clásicos utilizando el modelo de canal largo (ec. (4-8)) y el de canal corto (ec. (4-9)). En esta figura se ha tomado como punto de partida los valores de los parámetros para la tecnología de  $1\mu\text{m}$  de la Tabla 4-10 y a partir de éstos se han escalado los correspondientes al resto de tecnologías según los diferentes criterios de cada escenario. El valor de la movilidad efectiva ( $\mu_{eff}$ ) se ha calculado para cada caso con la ec. (1.81) de [41].

En la Figura 4-11 puede observarse como el modelo de canal largo predice una mayor sensibilidad de la corriente máxima de saturación respecto del escalado que el modelo de canal corto. Esto es debido a que en el segundo caso se tienen en cuenta los efectos de saturación de la velocidad de los portadores. La dependencia con la tensión en el modelo de canal largo es cuadrática (si no tenemos en cuenta la dependencia de la movilidad con el campo vertical en el canal) mientras que la dependencia para los dispositivos de canal corto con la tensión oscila entre un exponente de valor 1 (en el caso extremo para dispositivos fuertemente dominados por la velocidad de saturación, ec. (4-11)) y el valor de 2 para dispositivos de canal largo. Según algunos autores este exponente está en torno a los valores 1.3 y 1.4 para tecnologías actuales submicrónicas [59][60]. El principal efecto de este cambio de exponente es que el aumento de la

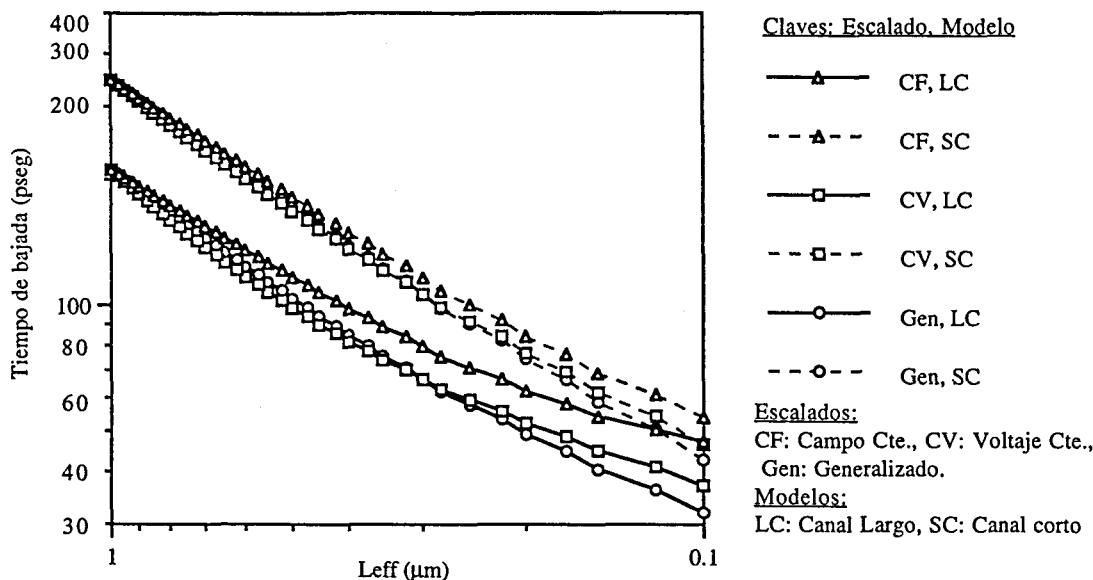
corriente de saturación por unidad de longitud del canal se satura, hasta hacerse constante a medida que nos adentramos más en la región profundamente sub-micrónica.



**Figura 4-11**

*Representación gráfica de la corriente máxima de saturación para un inversor con relaciones de aspecto 300:1 para el PMOS y 100:1 para el NMOS, en función del escalado para los tres escenarios clásicos (campo constante: CF, voltaje constante: CV y Generalizado con  $\epsilon=\alpha^{1/2}$ ). Las líneas continuas corresponden al modelo para dispositivos de canal largo y las líneas a trazos al modelo para canal corto.*

Por otro lado tenemos los tiempos de transición, el otro parámetro que determina el ruido  $dI/dt$  que generan los dispositivos lógicos al conmutar. En la Figura 4-12 se han representado los modelos de tiempo de bajada para una transición de nivel alto a bajo a la salida de un inversor cargado con un Fan-Out de cuatro inversores iguales a él. En esta figura puede observarse como la reducción de los tiempos de transición se frena para dimensiones menores de  $0.5\mu\text{m}$ , incluso en el escenario de voltaje constante (CV) y para el modelo de canal largo, debido sobre todo a la disminución de la movilidad producida por el campo vertical que se genera en el canal. Vemos que el escenario generalista es el que mejores prestaciones ofrece (se consideren los efectos de canal corto o no) debido al compromiso entre reducción del voltaje y mantenimiento de la movilidad efectiva para dispositivos por debajo de  $0.3\mu\text{m}$ . Si consideramos los efectos de canal corto los valores obtenidos para los tiempos de transición son siempre menores, como cabe esperar debido a la saturación de la velocidad de los portadores debido al campo transversal en el canal.



**Figura 4-12**

Representación gráfica de los modelos del tiempo de bajada para una transición alto-bajo a la salida de un inversor con un Fan-out de 4. Las relaciones de aspecto de los transistores son de 300:1 para el PMOS y 100:1 para el NMOS. Las líneas a trazos corresponden al modelo de canal corto mientras que las líneas continuas corresponden al modelo de canal largo, para los tres escenarios clásicos CV, CF y Generalizado. Los tiempos de bajada se definen como el tiempo que tarda la salida en evolucionar del 90% al 10% de su valor inicial.

En resumen, podemos decir que es necesario utilizar modelos que contemplen los efectos de canal corto si queremos evitar ser demasiado pesimistas al predecir el ruido de conmutación, pues los modelos clásicos proporcionan niveles de corriente máximos mayores y tiempos de transición menores. En el apartado siguiente vamos a ver cual es la predicción que proporcionan estos modelos del ruido de conmutación  $dI/dt$ .

## 4.4 Escalado y ruido de conmutación

En [61] se presenta un estudio de cómo evolucionaría el ruido de conmutación en un escenario de escalado de voltaje constante. La conclusión es que éste aumenta exponencialmente al reducir las dimensiones físicas de los dispositivos de los CI. Sin embargo es necesario realizar un estudio más amplio que contemple los escenarios reales, que como hemos visto, no pueden clasificarse dentro de los dos extremos de campo constante o voltaje constante. En este apartado se presenta, en primer lugar, una estimación del ruido de conmutación basándose en los modelos del apartado anterior para los tres escenarios clásicos. Después se utilizan estos modelos para calcular el ruido de conmutación para datos más realistas teniendo en cuenta dos escenarios: de altas prestaciones y de bajo consumo, como vimos en el subapartado 4.2.3. Finalmente se comparan estos cálculos con resultados obtenidos por simulación para estos mismos escenarios en los que se utilizan datos de tecnologías actuales y modelos HSPICE escalables para las tecnologías futuras.

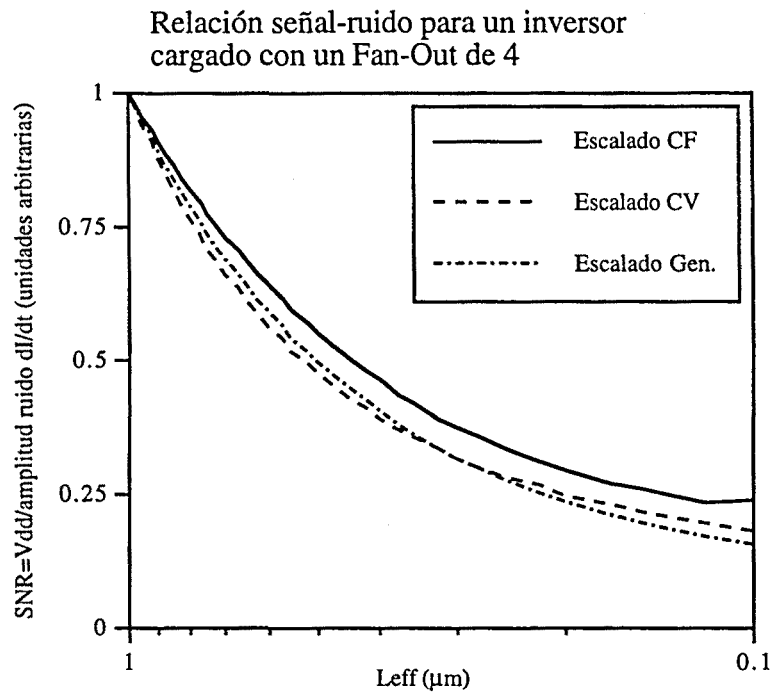


#### 4.4.1 Escalado del ruido de conmutación para escenarios clásicos.

Nos interesa en este apartado analizar el resultado de las tendencias de escalado en el ruido de conmutación  $dI/dt$  suponiendo escenarios de campo constante, voltaje constante y generalizados. Para ello basta con calcular el cociente entre el valor máximo de la corriente de saturación para cada longitud de canal y correspondiente tiempo de transición, que en nuestro estudio aproximaremos por el tiempo de bajada. Lo que nos interesa no son los valores absolutos de ruido sino estimar su evolución a medida que reducimos las dimensiones físicas. Partiendo de los datos para una tecnología de  $1\mu\text{m}$  (Tabla 4-10) y utilizando el modelo de canal corto para estimar  $I_{Dsat\ max}$  y  $t_f$ , según el procedimiento explicado en el apartado anterior, se han representado gráficamente los valores del cociente  $I_{Dsat\ max}/t_f$ , expresándolos de forma relativa a los valores obtenidos para la tecnología de  $1\mu\text{m}$ . Los resultados se presentan en la Figura 4-14 donde se ha utilizado como circuito de prueba un inversor cargado con cuatro inversores idénticos de relaciones de aspecto 300:1 para el PMOS y 100:1 para el NMOS. En el estudio se ha mantenido el valor de la inductancia del encapsulado constante durante el escalado y su valor no es importante pues los resultados presentados para las diferentes longitudes efectivas de canal son relativos a los obtenidos para la tecnología de  $1\mu\text{m}$ .

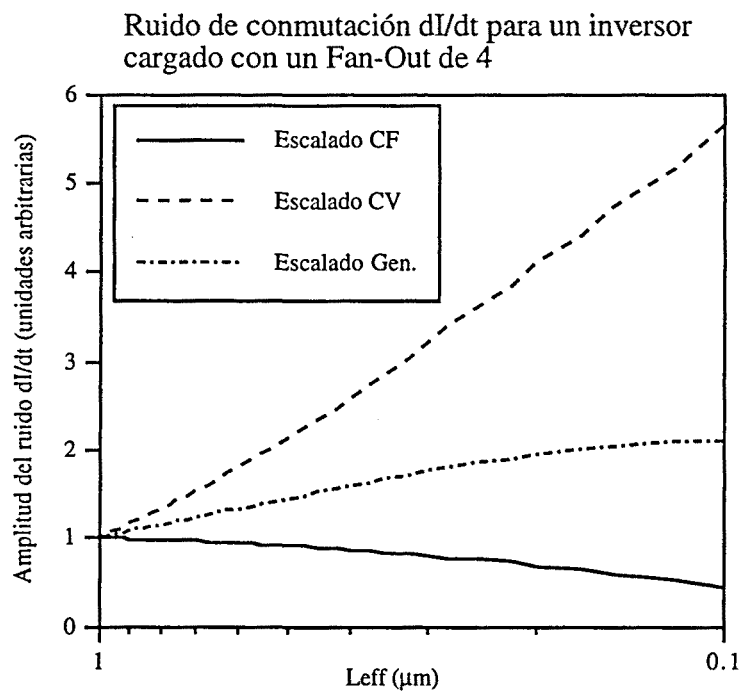
En la Figura 4-14 se ha representado el ruido de conmutación calculado como el cociente entre la corriente máxima de saturación y el tiempo de bajada para un inversor cargado con un Fan-Out de 4 inversores iguales utilizando el modelo de canal corto del apartado anterior. Como podemos observar, los resultados del escenario de escalado de voltaje constante (CV) coinciden con los de [61]. Lógicamente el escenario de campo constante (CF) en el que la tensión se reduce, se produce una cierta disminución del ruido de conmutación pero a costa de escalar muy lentamente los tiempos de transición. El compromiso entre el crecimiento exponencial del ruido para el escalado CV y la pérdida de prestaciones del escalado CF lo tenemos en el escalado generalizado. Para este escenario se ha utilizado un factor de escalado para las tensiones igual a la raíz cuadrada del que se utiliza para escalar el resto de parámetros.

Sin embargo el valor de la amplitud del ruido de conmutación es necesario compararlo con el valor de las tensiones de alimentación, es decir obtener el valor de ruido relativo a las tensiones de alimentación para cada tecnología. Hemos llamado a este valor la relación señal ruido (SNR) entre la tensión de alimentación y el ruido de conmutación y lo calculamos como el cociente entre ambas magnitudes. En la Figura 4-13 se ha representado gráficamente la relación SNR así definida para los valores de ruido de la Figura 4-14 y las tensiones de alimentación que se derivan de cada escalado, partiendo de que la tensión para la tecnología de  $1\mu\text{m}$  es de 5V. Como puede observarse en la Figura 4-13, a pesar de que el comportamiento de la amplitud del ruido de conmutación es bastante diferente para los tres escenarios (CF, CV y Generalizado), la relación de éste con la tensión de alimentación sigue las mismas pautas en los tres escenarios. La tendencia general es una disminución muy rápida hasta dimensiones entorno a las  $0.5\mu\text{m}$  de la SNR, frenándose después debido a la influencia de la reducción de la movilidad y la saturación de la velocidad que hacen que la reducción en los tiempos de transición con el escalado sea menor en la región profundamente submicrónica.



**Figura 4-13**

Representación gráfica de la relación señal-ruido entre la tensión de alimentación y el ruido de conmutación  $dI/dt$  para los tres escenarios clásicos de escalado.



**Figura 4-14**

Representación gráfica del ruido de conmutación calculado como el cociente entre la corriente máxima de saturación y el tiempo de bajada para un inversor cargado con un Fan-Out de 4 inversores iguales utilizando un modelo de canal corto.

### 4.4.2 Escalado en escenarios realistas

La tendencia real de la industria de fabricación de circuitos integrados, como ya vimos en el apartado 4.2.3, no sigue exactamente ninguno de los tres escenarios clásicos estudiados en el apartado anterior. Es necesario utilizar valores para los parámetros tecnológicos que reflejen con más fidelidad la evolución real de la tecnología. Existen numerosos trabajos que dan indicaciones de esta evolución, como ya vimos anteriormente, y en los que se apunta una doble tendencia hacia un escalado de altas prestaciones, que busca maximizar la velocidad de los CI, y un escalado de bajo consumo, que intenta mantener la densidad de potencia disipada lo más constante posible. Utilizaremos los datos para los parámetros de las tecnologías futuras de uno de los trabajos de la literatura [35]. Para las tecnologías actuales hemos usado los parámetros que proporciona MOSIS para sus tecnologías de  $2\mu\text{m}$ ,  $1.8\mu\text{m}$ ,  $0.8\mu\text{m}$  y  $0.6\mu\text{m}$ . Los diferentes valores se listan en la Tabla 4-12.

La tensión umbral se ha calculado a partir de la tensión umbral para los dispositivos de 5V de tensión de alimentación que es de 0.7V. Esta tensión umbral se ha escalado proporcionalmente a la tensión de alimentación para cada tecnología pero limitándola a un valor mínimo de 0.3V, límite inferior fijado, como ya vimos en el subapartado 4.2.2.3, por aspectos de fiabilidad.

Utilizando un inversor cargado con un Fan-Out de 4 inversores iguales como circuito de test hemos calculado, utilizando los modelos de canal largo y canal corto, la corriente máxima de saturación ( $I_{Dsat\ max}$ ) y el tiempo de bajada ( $t_f$ ) para una transición alto-bajo a la salida del inversor (Figura 4-15). A partir de los resultados obtenidos con el modelo de canal corto se ha estimado el ruido de conmutación  $dI/dt$  como el cociente entre  $I_{Dsat\ max}$  y  $t_f$  que encontramos representado en la Figura 4-16.(a). Se ha calculado también la relación entre la tensión de alimentación y el ruido de conmutación para cada tecnología, representándose en la Figura 4-16.(b).

Longitud puerta ( $\lambda$ )	$2\mu\text{m}$	$1.2\mu\text{m}$	$0.8\mu\text{m}$	$0.6\mu\text{m}$	$0.35\mu\text{m}$	$0.25\mu\text{m}$	$0.18\mu\text{m}$	$0.13\mu\text{m}$
Longitud canal ( $L_{eff}$ )	$1.8\mu\text{m}$	$1.2\mu\text{m}$	$0.7\mu\text{m}$	$0.5\mu\text{m}$	$0.25\mu\text{m}$	$0.15\mu\text{m}$	$0.1\mu\text{m}$	$0.07\mu\text{m}$
$V_{DD}$ (Altas prestaciones)	5	5	5	3.3	2.5	1.8	1.5	1.2
$V_{DD}$ (Bajo consumo)	5	5	5	2.5	1.8	1.2	1	1
Óxido de puerta ( $T_{ox}$ )	$356\text{\AA}$	$224\text{\AA}$	$171\text{\AA}$	$120\text{\AA}/96\text{\AA}$	$70\text{\AA}$	$50\text{\AA}$	$35\text{\AA}$	$25\text{\AA}$

**Tabla 4-12**

*Valores de los parámetros tecnológicos principales para dos escenarios de escalado realistas basados en objetivos de altas prestaciones y bajo consumo, respectivamente.*

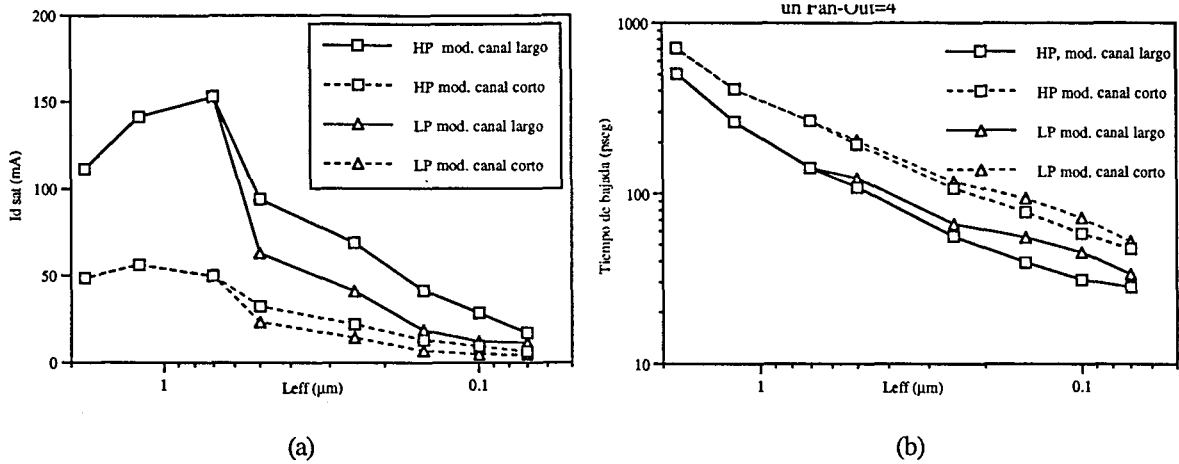


Figura 4-15

Corriente máxima de saturación (a) y tiempo de bajada (b) para tecnologías actuales y futuras siguiendo dos escenarios realistas de escalado: HP que corresponde a un escenario de altas prestaciones y LP para bajo consumo. Las líneas continuas representan los valores obtenidos con un modelo de canal largo mientras que las discontinuas enlazan los valores obtenidos con un modelo de canal corto.

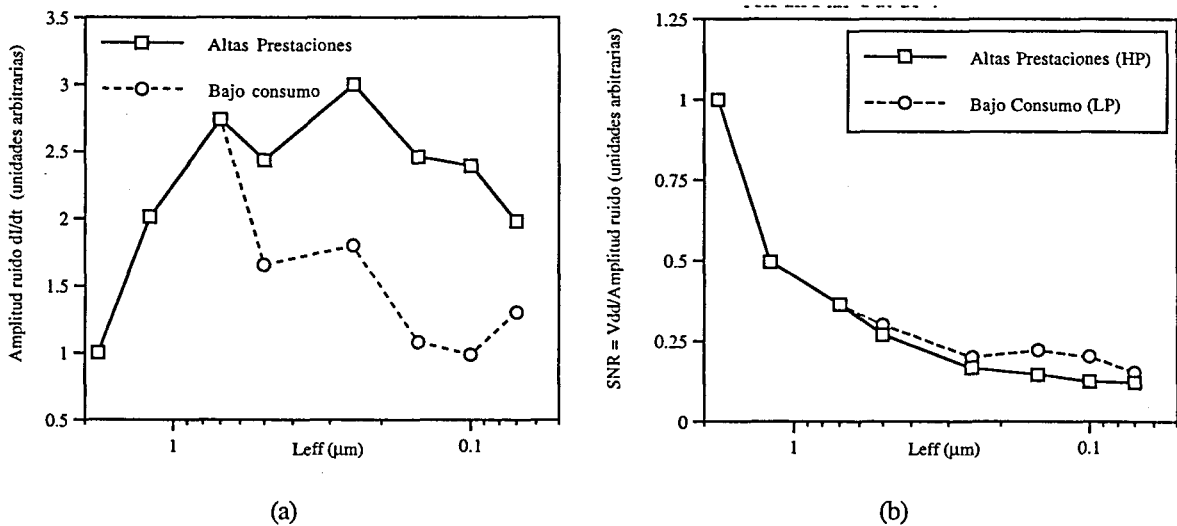


Figura 4-16

Amplitud del ruido de conmutación (a) y relación entre la tensión de alimentación y éste ruido (b) para los dos escenarios realistas de altas prestaciones (HP) y bajo consumo (LP). Para calcular los valores se ha utilizado un modelo de canal corto.

Como podemos observar en la Figura 4-16 mientras mantenemos constante la tensión de alimentación (generaciones tecnológicas de  $L_{eff}$  de  $1.8 \mu m$ ,  $1.2 \mu m$  y  $0.8 \mu m$ ) el ruido de conmutación aumenta rápidamente. Para generaciones posteriores y a medida que se va reduciendo la tensión de alimentación este aumento se ve frenado y finalmente disminuye en la región profundamente sub-micrónica (sobre todo en el escenario LP). Sin embargo, al igual que en los escenarios clásicos, la relación entre el ruido de conmutación generado y la tensión de alimentación va disminuyendo progresivamente hasta la generación de  $0.5 \mu m$ , manteniéndose luego bastante constante. No existe tampoco gran diferencia entre el comportamiento de los dos escenarios, como puede apreciarse en la Figura 4-16.(b).