

ESCUELA POLITÉCNICA NACIONAL

FACULTAD DE INGENIERIA ELECTRICA  
ESPECIALIZACION DE ELECTRONICA  
Y  
TELECOMUNICACIONES

DISEÑO DE CIRCUITOS INTEGRADOS DE APLICACION  
ESPECIFICA (ASICs) DIGITALES CON TECNOLOGIA CMOS

VOLUMEN I  
FUNDAMENTOS TEORICOS DE DISEÑO  
VLSI (VERY LARGE SCALE INTEGRATION)

IVAN BERNAL CARRILLO  
FREDY LEMUS CRIDLLO

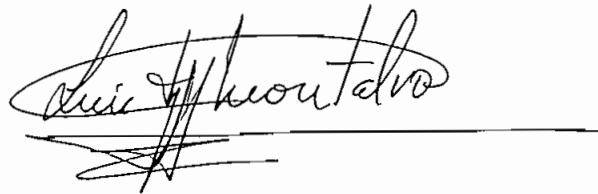
TESIS PREVIA A LA OBTENCION DEL TITULO DE  
INGENIERO EN LA ESPECIALIZACION DE INGENIERIA  
ELECTRONICA DE LA ESCUELA POLITÉCNICA NACIONAL

SEPTIEMBRE 1992

*Certifico que bajo mi dirección  
la presente tesis fue realizada  
en su totalidad por los  
señores:*

*Iván Bernal Carrillo*

*Fredy Lemus Criollo*

A handwritten signature in black ink, appearing to read "Luis Montalvo", is written over a horizontal line. Below the line, there are several scribbled-out strokes.

*Ing. Luis Montalvo Ramírez*

*DIRECTOR*

*DEDICATORIA*

*A MIS PADRES, QUIENES  
ENCAMINARON MI VIDA Y SEMBRARON  
EN ELLA LA CONSTANCIA Y EL  
DESEO DE SUPERACION.*

*IVAN BERNAL CARRILLO*

*DEDICATORIA*

*A MIS PADRES, CUYO EJEMPLO DE FE EN DIOS Y DE TRABAJO ME HA DADO EL ALIENTO Y LA CONSTANCIA PARA LOGRAR ESTA META. A MIS HERMANOS CON SU APOYO MORAL COADYUDAN PARA EL LOGRO DE NUESTRA SUPERACION.*

*FREDY LEMUS CRIOLLO*

## AGRADECIMIENTO

*Al Señor Ingeniero Luis Montalvo Ramírez cuyo valioso aporte fue determinante en la culminación de el presente trabajo de Tesis.*

*A los Señores Ingenieros Fernando Flores y Pablo Hidalgo y a todas las personas que contribuyeron para que el presente trabajo llegue a feliz término.*

## INTRODUCCION

Hasta hace pocos años la realización de circuitos integrados enfrentaba severas limitaciones debido al requerimiento de recursos humanos altamente calificados y elevados costos de fabricación. El conocimiento de los fundamentos teóricos y de las técnicas necesarias para el diseño de circuitos integrados eran privativos de una élite de especialistas y la infraestructura necesaria para la fabricación y evaluación de los prototipos obtenidos de los diseños era limitada debido a su alto costo.

El Proyecto Multiusuario (PMU) Iberoamericano, coordinado por el Centro Nacional de Microelectrónica (CNM) de Barcelona-España ha permitido a la Escuela Politécnica Nacional incursionar en el campo del diseño de circuitos integrados, superando las limitaciones económicas y de infraestructura indicadas gracias a la filosofía de cooperación internacional del proyecto. El Proyecto PMU Iberoamericano ha incentivado en la Escuela Politécnica Nacional la conformación y capacitación del "Grupo de Microelectrónica".

La presente Tesis reporta el trabajo de investigación del Grupo de Microelectrónica que ha debido afrontar múltiples dificultades y limitaciones propias de las investigaciones en campos inéditos en el país. Ello ha implicado la necesidad de un trabajo en varios frentes tanto a nivel de fundamentos físicos como metodológicos; ha sido necesaria además la adquisición y estudio de herramientas que permitan la realización del diseño de circuitos integrados.

El trabajo desplegado ha permitido a la Escuela Politécnica Nacional enfrentar con éxito el reto tecnológico que implicó el ingreso del Ecuador en un proyecto de esta trascendencia.

- En el Volumen I se introducen los conceptos fundamentales tanto a nivel de física de semiconductores como de tecnologías de fabricación y principios de caracterización de los circuitos integrados; se describe además la mecánica de operación del Proyecto Multiusuario tanto en los aspectos técnicos como administrativos.
- En el Volumen II se aborda el estudio de las herramientas para el diseño, simulación y descripción de los circuitos VLSI y se analizan en concreto dos paquetes con filosofías distintas, el uno consistente en el TENTOS desarrollado por la Universidad Federal de Río Grande del Sur (Brasil) y el otro denominado FPL desarrollado en la Universidad de UTAH (E.U.). Se describe

además a manera de tutorial el diseño de un circuito integrado sencillo "Medio Sumador" que fue fabricado en Diciembre de 1991.

- En el Volumen III se presenta la aplicación de los conceptos introducidos en los volúmenes anteriores en el diseño de un circuito integrado para el campo de las telecomunicaciones consistente en un "Codificador/Decodificador HDEn" y se resumen las conclusiones y recomendaciones finales.



## CONTENIDO GENERAL

### VOLUMEN I. FUNDAMENTOS TEORICOS DE DISEÑO VLSI (VERY LARGE SCALE INTEGRATION)

#### CAPITULO 1. GENERALIDADES DE LA METODOLOGIA DE DISEÑO VLSI.

- 1.1 Introducción.
- 1.2 Niveles de abstracción.
- 1.3 Estrategias para la reducción de la complejidad del diseño de un CI.
- 1.4 Alternativas de diseño de sistemas electrónicos.
- 1.5 Tecnologías de fabricación.
- 1.6 Aspectos económicos del desarrollo y producción de CIs.

#### CAPITULO 2. FUNDAMENTOS DE CIRCUITOS CMOS.

- 2.1 Generalidades.
- 2.2 El Transistor MOS como switch.
- 2.3 Funcionamiento y física del Transistor MOS.
- 2.4 El Inversor CMOS - Características DC.
- 2.5 La Compuerta de Transmisión - Características DC.
- 2.6 El problema del efecto parásito de LATCH-UP.
- 2.7 Comparación con otras tecnologías.

## CAPITULO 3. PROCESOS DE FABRICACION CMOS.

- 3.1 Técnicas básicas de fabricación.
- 3.2 Tecnologías de fabricación CMOS.
- 3.3 Interfaz Diseño-Fabricación.
- 3.4 Estructuras de Entrada/Salida: PADs.
- 3.5 Tipos de empaquetamiento.
- 3.6 Chips Multi-proyecto.

## CAPITULO 4. CARACTERIZACION Y ESTIMACION DEL FUNCIONAMIENTO DE CIRCUITOS INTEGRADOS.

- 4.1 Parámetros eléctricos: estimación de Resistencia y Capacitancia.
- 4.2 Características de conmutación y retardos.
- 4.3 Estimación de consumo de potencia.
- 4.4 Escalamiento de dimensiones del transistor MOS.
- 4.5 Estrategias de Temporización de los sistemas.

## VOLUMEN II. HERRAMIENTAS PARA DISEÑO DE ASICs

### CAPITULO 5. HERRAMIENTAS DE DISEÑO VLSI.

- 5.1 Tipos de herramientas de diseño.
- 5.2 Secuencias de diseño para la concepción de ASICs.
- 5.3 Principios de simulación eléctrica y lógica.
- 5.4 Formato CIF para descripción geométrica de layouts.

CAPITULO 6. DISEÑO DE ASICs BASADO EN CELDAS ESTANDAR.

- METODOLOGIA CONVENCIONAL.

- 6.1 Diseño basado en bandas y celdas estándar.
- 6.2 Herramientas del sistema "TENTOS" para diseño convencional.

CAPITULO 7. CASO DE ESTUDIO SENCILLO UTILIZANDO TENTOS:

DISEÑO DE UN "MEDIO SUMADOR".

- 7.1 Diseño a nivel funcional.
- 7.2 Diseño a nivel estructural.
- 7.3 Diseño a nivel físico.
- 7.4 Resultados del funcionamiento de los prototipos fabricados en la fundidora ES2(Francia).

CAPITULO 8. DISEÑO DE ASICs BASADO EN CELDAS ESTANDAR.

- METODOLOGIA ESTRUCTURADA PPL (*Physical Placement of Logic*).

- 8.1 Diseño convencional versus diseño PPL.
- 8.2 Metodología de diseño PPL.
- 8.3 Manipulación de las herramientas PPL. Caso de estudio sencillo: diseño de una compuerta "EXOR".

VOLUMEN III. CODIFICADOR/DEDOCIFICADOR  
DE LINEA PROGRAMABLE  
HDB<sub>n</sub> COMO CASO DE ESTUDIO.

CAPITULO 9. CONCEPCION DEL CIRCUITO CODIFICADOR/DECODIFI-  
CADOR.

- 9.1 Consideraciones teóricas del Código HDB<sub>n</sub>.
- 9.2 Diseño a nivel funcional del Codificador en base a Máquinas de Estados Finitos (MEF).
- 9.3 Diseño a nivel funcional del Decodificador en base a Máquinas de Estados Finitos (MEF).

CAPITULO 10. DISEÑO DEL CODIFICADOR/DECODIFICADOR HDB<sub>n</sub>  
BASADO EN CELDAS ESTANDAR - METODO CONVEN-  
CIONAL.

- 10.1 Diseño del Codificador Programable HDB<sub>n</sub>.
- 10.2 Diseño del Decodificador Programable HDB<sub>n</sub>.
- 10.3 Diseño de anillo de PADS e inserción del diseño global.

CAPITULO 11. DISEÑO DEL CODIFICADOR/DECODIFICADOR HDB<sub>n</sub>  
BASADO EN CELDAS ESTANDAR - METODO ESTRUCTURA-  
DO FFL.

- 11.1 Diseño del Codificador Programable HDB<sub>n</sub>.
- 11.2 Diseño del Decodificador Programable HDB<sub>n</sub>.
- 11.3 Inserción del diseño global en el PAD-FRAME.

## CAPITULO 12. CONCLUSIONES Y RECOMENDACIONES.

12.1 Conclusiones generales.

12.1 Conclusiones respecto al paquete TENTOS.

12.2 Conclusiones respecto al paquete FPL.

## BIBLIOGRAFIA

### ANEXOS

ANEXO A. DISPOSITIVOS Y COMANDOS SPICE EMPLEADOS EN SIMULACION DIGITAL.

ANEXO B. DISPOSITIVOS Y COMANDOS EMPLEADOS EN SIMULACION LOGICA NDL.

ANEXO C. DIRECTIVAS DE DESCRIPCION CIF.

ANEXO D. DETALLES DEL SISTEMA TENTOS.

ANEXO E. ARCHIVOS AUXILIARES DEL TENTOS.

ANEXO F. CELDAS ESTANDAR EMPLEADAS - METODO CONVENCIONAL.

ANEXO G. REGLAS DE DISEÑO .

ANEXO H. PLANOS DEL CIRCUITO CODIFICADOR/DECODIFICADOR HDBn.

## INDICE DEL VOLUMEN I

CAPITULO 1.	GENERALIDADES DE LA METODOLOGIA DE DISEÑO VLSI	
1.1	INTRODUCCION . . . . .	1
1.2	NIVELES DE ABSTRACCION . . . . .	3
1.2.1	El nivel físico . . . . .	4
1.2.2	Los niveles eléctrico y lógico (nivel estructural) . . . . .	7
1.2.3	El nivel funcional . . . . .	9
1.3	ESTRATEGIAS PARA LA REDUCCION DE LA COMPLEJIDAD DEL DISEÑO DE UN CI. . . . .	11
1.3.1	Abstracción . . . . .	11
1.3.2	Jerarquía y estructuración . . . . .	12
1.3.3	Modularidad . . . . .	12
1.3.4	Regularidad . . . . .	13
1.3.5	Localidad . . . . .	14
1.4	ALTERNATIVAS DE DISEÑO DE SISTEMAS ELECTRONICOS . .	14
1.4.1	PCBs (Printed Circuit Boards) . . . . .	15
1.4.2	ASICs (Aplication Specific Integrated Circuits) . . . . .	17
1.5	TECNOLOGIAS DE FABRICACION . . . . .	31
1.5.1	Bipolar . . . . .	32
1.5.2	Metal-Oxido-Semiconductor (MOS) . . . . .	32
1.5.3	Arseniuro de Galio (GaAs) . . . . .	33
1.5.4	CMOS-Bipolar (BiCMOS) . . . . .	34

1.6	ASPECTOS ECONOMICOS DEL DESARROLLO Y PRODUCCION DE	
	CIs . . . . .	34
1.6.1	Costes de producción . . . . .	36
1.6.2	Costes de desarrollo . . . . .	37
1.6.3	Conclusión. . . . .	38
CAPITULO 2.    FUNDAMENTOS DE CIRCUITOS CMOS . . . . .		40
2.1	GENERALIDADES . . . . .	40
2.1.1	Descripción básica de la estructura física de un Transistor MOS . . . . .	43
2.1.2	Breve descripción de los materiales utilizados en la fabricación del Transistor MOS . . . . .	44
2.2	EL TRANSISTOR MOS COMO SWITCH . . . . .	56
2.2.1	Redes combinatoriales CMOS . . . . .	59
2.3	FUNCIONAMIENTO Y FISICA DEL TRANSISTOR MOS . . . . .	80
2.3.1	Transistores de enriquecimiento nMOS . . . . .	80
2.3.2	Transistores de enriquecimiento y empobrecimiento . . . . .	92
2.3.3	Estudio analítico del transistor nMOS . . . . .	96
2.3.4	Características V-I . . . . .	110
2.3.5	Comparación entre transistores canal "n" y "p" . . . . .	112
2.3.6	Ajuste del voltaje umbral. . . . .	113
2.3.7	Efecto de cuerpo . . . . .	115

2.4	EL INVERSOR CMOS - CARACTERISTICAS DC . . . . .	117
2.4.1	Zona A . . . . .	122
2.4.2	Zona B. . . . .	123
2.4.3	Zona C . . . . .	127
2.4.4	Zona D . . . . .	129
2.4.5	Zona E . . . . .	131
2.4.6	Influencia de la relación $\beta_n/\beta_p$ en la curva de transferencia. . . . .	132
2.4.7	Influencia de la temperatura en la carac- terística de transferencia. . . . .	134
2.4.8	Margen de ruido. . . . .	135
2.5	LA COMPUERTA DE TRANSMISION - CARACTERISTICAS DC . .	139
2.5.1	Transistor de paso nMOS. . . . .	140
2.5.2	Transistor de paso pMOS . . . . .	142
2.5.3	Compuerta de transmisión CMOS . . . . .	143
2.6	EL PROBLEMA DEL EFECTO PARASITO DE <i>LATCH-UP</i> . . . .	145
2.6.1	Estructura física del Inversor CMOS . . .	145
2.6.2	El <i>latch-up</i> . . . . .	147
2.7	COMPARACION CON OTRAS TECNOLOGIAS . . . . .	152
2.7.1	MOSFETs vs. BJT . . . . .	153
2.7.2	CMOS vs. nMOS . . . . .	155
CAPITULO 3. PROCESOS DE FABRICACION CMOS . . . . .		157
3.1	TECNICAS BASICAS DE FABRICACION . . . . .	168
3.1.1	Fabricación de la oblea . . . . .	168
3.1.2	Oxidación . . . . .	171
3.1.3	Creación de patrones para difusión selec- tiva . . . . .	172



3.1.4	Deposición . . . . .	178
3.1.5	Proceso para la compuerta de Silicio (silicon gate) . . . . .	179
3.2	TECNOLOGIAS DE FABRICACION CMOS . . . . .	182
3.2.1	Proceso de pozo-p ( <i>p-well process</i> ) . . . . .	183
3.2.2	Proceso de pozo-n ( <i>n-well process</i> ) . . . . .	190
3.2.3	Proceso de doble pozo ( <i>twin tub process</i> ). . . . .	191
3.2.4	Silicio sobre aislante ( <i>silicon on insulator SOI</i> ). . . . .	192
3.2.5	Incremento de la capacidad de enrutamiento . . . . .	193
3.2.6	Protecciones contra el latch-up . . . . .	197
3.3	INTERFAZ DISEÑO - FABRICACION . . . . .	201
3.3.1	Reglas de diseño CMOS . . . . .	207
3.4	ESTRUCTURAS DE ENTRADA/SALIDA: PADS . . . . .	221
3.4.1	Estructuras de entrada/salida (I/O) . . . . .	224
3.4.2	Anillos de $V_{DD}$ y Gnd. . . . .	238
3.5	TIPOS DE EMPAQUETAMIENTO . . . . .	239
3.6	CHIPS MULTI-PROYECTO . . . . .	246
3.6.1	Proyecto multiusuario Iberoamericano . . . . .	251
CAPITULO 4.	CARACTERIZACION Y ESTIMACION DEL FUNCIONAMIENTO DE CIRCUITOS INTEGRADOS . . . . .	256
4.1	ESTIMACION DE RESISTENCIA Y CAPACITANCIA . . . . .	258
4.1.1	Estimación de la resistencia de regiones rectangulares. . . . .	258
4.1.2	Estimación de resistencia en transistores . . . . .	262
4.1.3	Características de un capacitor MOS . . . . .	264

4.1.4	Estimación de Capacitancias . . . . .	268
4.1.5	Efectos distribuidos RC . . . . .	280
4.1.6	Guía de diseño para la longitud de los caminos . . . . .	282
4.2	CARACTERISTICAS DE CONMUTACION Y RETARDOS . . . . .	286
4.2.1	Determinación del tiempo de bajada . . . . .	288
4.2.2	Determinación del tiempo de subida . . . . .	291
4.2.3	Tiempo de retardo . . . . .	292
4.3	ESTIMACION DEL CONSUMO DE POTENCIA . . . . .	293
4.3.1	Disipación estática (Ps). . . . .	294
4.3.2	Disipación dinámica (Pd). . . . .	296
4.3.3	ESCALAMIENTO DE DIMENSIONES DEL TRANSIS- TOR MOS . . . . .	299
4.3.1	Principios de escalamiento . . . . .	299
4.3.2	Escalamiento de los caminos de interco- nexión . . . . .	303
4.3.3	ESTRATEGIAS DE TEMPORIZACION DE LOS SIS- TEMAS . . . . .	304
4.3.4	Esquemas de temporización de circuitos síncronos . . . . .	306
4.3.5	Modelo lógico para la descripción de los esquemas de temporización . . . . .	311
4.3.6	Problemas con el esquema de temporización de una fase . . . . .	315
4.3.7	Esquema de temporización de una fase de reloj y su complemento . . . . .	318

# CAPITULO 1

## METODOLOGIA DE DISEÑO VLSI

### 1.1 INTRODUCCION

El término *Very Large Scale of Integration* VLSI refleja la capacidad de la industria de semiconductores de fabricar circuitos electrónicos complejos consistentes en millones de transistores en un solo sustrato de silicio.

Varios factores han contribuido al crecimiento acelerado de la densidad de integración en los últimos años, tales factores son en esencia:

- a) Los avances en las tecnologías de fabricación (técnicas litográficas de generación de máscaras y refinamiento de los métodos de procesamiento) que han permitido reducir los tamaños mínimos de los dispositivos básicos y de las líneas de interconexión. La mayor confiabilidad en el procesamiento ha permitido además un incremento en el tamaño de las obleas de silicio.

- b) La experiencia creciente en el diseño de layouts, que ha derivado en un mejor entendimiento de los problemas de diseño a diferentes niveles lo que ha permitido la generación de arquitecturas depuradas capaces de explotar la tecnología que se desarrolla.
- c) El surgimiento de un esquema claro de niveles de abstracción del problema de diseño de CIs resultado de los aportes de Mead y Conway (1980), que ha permitido esquematizar los conceptos de diseño en niveles bien definidos: funcional, circuital, y físico (de layout), y llegar a la etapa moderna del diseño VLSI caracterizada por una metodología jerárquica de concepción de CIs.
- d) El desarrollo de herramientas computacionales de diseño a diferentes niveles (CADs), que han liberado al diseñador de las complicadas tareas de descripción de diseños a bajo nivel, y de herramientas de simulación que han permitido la generación de diseños altamente depurados y confiables antes de su fundición.
- e) El desarrollo de los sistemas de fundición de *OBLEAS MULTIPROYECTO* que permiten enlazar directamente las fundidoras de silicio con los estudiantes permitiéndoles la adquisición de experiencia directa en el desarrollo de circuitos VLSI.

Estos factores han permitido el acceso al diseño VLSI a un grupo mayor de investigadores a nivel universitario, industrial y gubernamental, lo que se refleja en la difusión de publicaciones, conferencias, seminarios y fundiciones de investigación de vanguardia que garantizan el continuo desarrollo de las técnicas de diseño VLSI.

## 1.2 NIVELES DE ABSTRACCION

Los procesos de diseño VLSI pasan por un espectro diversificado de disciplinas involucradas con la física, la ingeniería química, ingeniería eléctrica, e informática.

Debido a la diversidad de tareas y problemas presentes en el proceso de diseño de CIs, es indispensable un refinamiento progresivo de dicho proceso, lo cual ha llevado a su división en varios niveles de abstracción que se podrían agrupar en:

- a) Nivel físico.
- b) Nivel estructural.
- c) Nivel funcional.

La Fig.1.1 resume la secuencia ideal en que se concatenan los niveles de diseño para la obtención de un circuito VLSI basado en la metodología Top-down.

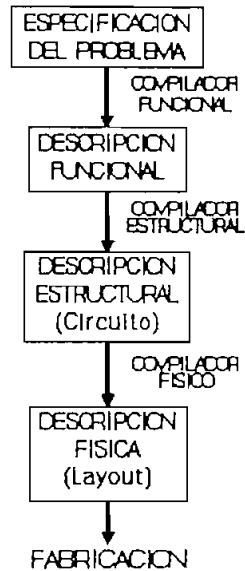


FIGURA 1.1 Niveles de abstracción.

A fin de facilitar la comprensión de cada nivel, cuyo fundamento reside en la abstracción y simplificación que se hace de elementos de los niveles inferiores, se ha preferido realizar su exposición comenzando por los niveles de menor abstracción.

### 1.2.1 El nivel físico

El nivel físico abarca las descripciones relativas a la realización física de cada uno de los componentes y su conexionado sobre el sustrato de silicio. Se lo puede considerar subdividido en dos niveles de mayor especialización:

a) Nivel físico propiamente dicho

Corresponde al nivel inferior en el que descansa el mundo físico de la conducción de semiconductores. Su objetivo consiste en controlar y explotar los procesos físicos a fin de traducirlos en dispositivos útiles.

Cuando estos dispositivos son agrupados para formar circuitos mayores, el flujo de corriente y los niveles de voltaje se convierten en los portadores de la información en el circuito. Los ingenieros eléctricos, para sus propósitos, usan una abstracción de este mundo físico basada en representaciones en forma de transistores, resistencias, capacitores y conectores. Los problemas de diseño inherentes a este nivel son:

- i) La velocidad de las señales determinada por los parámetros físicos del circuito.
- ii) La degradación de la señal en los conductores;
- iii) El consumo total de potencia del circuito, que determina su calentamiento; y,
- iv) La secuencia precisa de las formas de onda de voltaje y corriente que controlan el flujo de la información en el interior del circuito.

En un ambiente típicamente industrial, estos problemas ocupan la mayor atención de los ingenieros eléctricos que buscan diseñar circuitos *full-custom* óptimos.

#### b) Nivel de layout

Constituye el enlace entre el mundo físico y el eléctrico, es decir entre el circuito y su proceso de fabricación. Las descripciones se realizan en forma de *LAYOUTS*, consistentes en figuras geométricas agrupadas en capas yuxtapuestas que se corresponden con cada etapa del proceso de fundición.

Un ejemplo de layout se exhibe en la Fig.1.2.

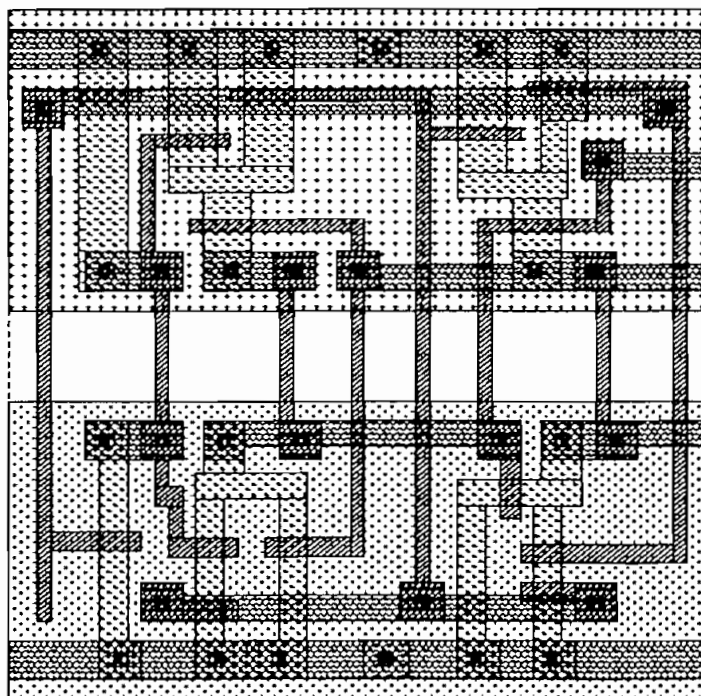


FIGURA 1.2 Ejemplo de un diseño al nivel de LAYOUT.



Problemas inherentes a ese nivel son:

- i) La optimización del área que ocupa el layout.
- ii) El cumplimiento de un conjunto de reglas de diseño dentro de las que el fabricante garantiza que aún considerando las variaciones inherentes al paso litográfico de generación de máscaras el circuito obtenido será en alto porcentaje funcionalmente correcto.
- iii) Problemas de interconexión a nivel eléctrico que en el layout implicarán una generación acumulada de resistencias y capacitancias en los conductores globales dependiendo del material de las capas involucradas en la conducción (metal, polysilicon o difusión).

### 1.2.2 Los niveles eléctrico y lógico (nivel estructural)

El nivel lógico en el proceso de diseño usa una abstracción de los circuitos eléctricos fundamentales en que las corrientes y voltajes son limitados a niveles discretos. En los circuitos digitales, los dos niveles de señales permisibles son aquellos que representan 0 y 1 lógico; el circuito es representado mediante diagramas tales como el de la Fig.1.3 (a) correspondiente al layout de la Fig.1.2.

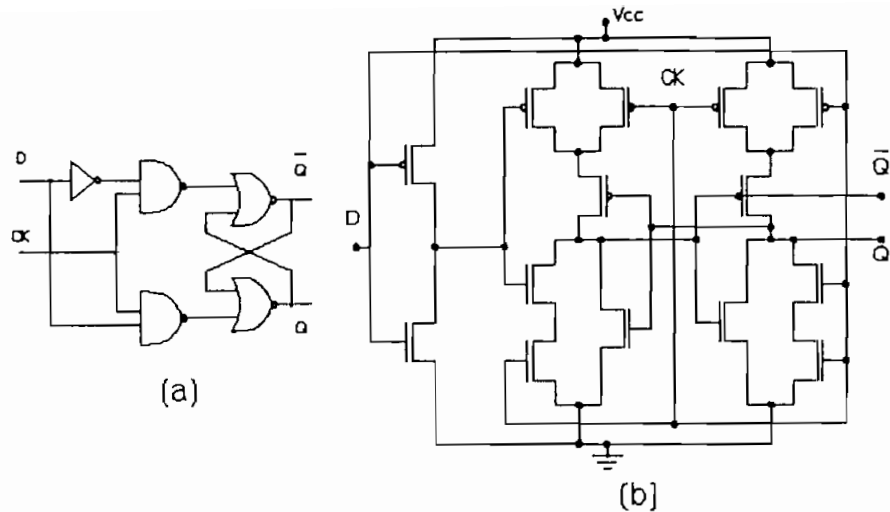


FIGURA 1.3 Esquemas lógico (a) y eléctrico (b) correspondientes al LAYOUT de la Fig.1.2.

Los símbolos usados a nivel lógico representan conmutadores, inversores y compuertas que manipulan señales lógicas.

Los esquemas a nivel circuital en cambio, son refinamientos del esquema a nivel lógico y se expresan en términos de transistores y elementos discretos (resistencias, capacitancias) con parámetros que representan las características físicas de los conductores.

### Sincronismo

El sincronismo consiste en el control del flujo de información entre las diferentes partes de la red de modo que ésta se procese en instantes de tiempo discretos determinados por la ocurrencia de "pulsos de reloj" que sirven como marcadores de tiempo dentro del circuito.

Las representaciones a nivel lógico esconden los detalles de retraso y consideraciones de temporización asumiendo que los circuitos trabajan sincrónicamente.

Son problemas inherentes a este nivel:

- a) El funcionamiento correcto de todas las operaciones lógicas del circuito en términos de operaciones elementales realizadas por los elementos lógicos en la secuencia de tiempo especificada.
- b) La simplicidad y elegancia del diseño, de modo que este se acople armónicamente con el proceso algorítmico a su nivel superior (nivel funcional) por un lado, y con el flujo de datos y la estructura de control en el nivel inferior (nivel de layout) por el otro lado.

### 1.2.3 El nivel funcional

La tarea del nivel funcional consiste en indicar las especificaciones para una solución conceptual del problema en términos de un proceso algorítmico, es decir la manera como un diseño particular debe responder a un conjunto dado de entradas.

Ejemplos: a)  $F = ((A+B+C) \cdot D)$

b) IF (LD==1) THEN Q = D;

Se trata, por tanto, de una descripción independiente de la tecnología, no existe ninguna noción sobre como se implementará esta función, ni sobre las características eléctricas implicadas.

A este nivel se analizan las diferentes opciones de algoritmos a implementarse a fin de evaluar de manera general la organización del flujo de datos, bloques funcionales, y necesidades de comunicación entre éstos.

Si la función a implementarse es demasiado compleja se estudiará la separación del proceso de diseño en subtarear más pequeñas que puedan ser ejecutadas independientemente. En la práctica, sin embargo, la independencia total de las subtarear no es factible debido a factores de diseño comunes a todas ellas y que forzan al diseño total a ir a través de varias iteraciones.

#### Plano de base.

No es considerado como un nivel de abstracción propiamente dicho, consiste en la proyección topológica de la arquitectura del circuito sobre el silicio.

En el plano de base se definen:

- a) La interconexión del sistema diseñado con subsistemas

mayores además de los interfaces de comunicación con el mundo externo;

- b) Las estrategias de conducción globales, selección de capas para conducción de las señales de control globales de datos, de polarización y potencia;
- c) La ubicación de puntos de entrada/salida, y estimaciones iniciales de área.

### 1.3 ESTRATEGIAS PARA LA REDUCCIÓN DE LA COMPLEJIDAD DEL DISEÑO DE UN CI.<sup>(1)</sup>

La complejidad de los sistemas VLSI y la cantidad de información que los sistemas actuales involucran es tal que independientemente del método a ser adoptado, se deben mantener algunos principios para la sistematización del diseño de un CI.

Tales principios son fundamentalmente:

#### 1.3.1 Abstracción

La delimitación de distintos niveles de abstracción para cada etapa permite al diseñador ignorar los detalles ajenos

---

<sup>(1)</sup> "Principles of CMOS VLSI Design A Systems Perspective", N. Weste - E. Karan, Págs 238 a 240.

al nivel en el que se halle trabajando, y concentrarse tan solo en los problemas propios de dicho nivel

### 1.3.2 Jerarquía y estructuración

Implica la división de un módulo en submódulos y la repetición iterativa de esta operación hasta que la complejidad de los submódulos sea tal que estos sean comprensibles al nivel deseado de detalle. Es similar al caso del desarrollo de software donde programas grandes son divididos en secciones cada vez más pequeñas hasta llegar a la definición de subrutinas con funciones e interfaces bien definidos (*Diseño top-down*).

### 1.3.3 Modularidad

Consiste en la división del sistema en un conjunto de submódulos funcionales, cuya interacción con otros módulos está bien caracterizada. El interfaz físico de cada módulo indica su posición, nombre, nivel, tamaño, y tipo de señal de interconexión externa, y es análogo en el caso del software a una lista de argumentos con tipos de variables.

La modularidad ayuda al ingeniero a clarificar y documentar la solución del problema, y aporta al diseño en equipo donde cada diseñador es responsable por una porción del diseño del CI total.

En el desarrollo de software estructurado se emplean tres construcciones básicas: *concatenación*, *iteración* y *selección condicional*. En el mundo del diseño de CIs la *concatenación* se realiza mediante el empalme de celdas básicas (en el nivel físico) que son interconectadas colocándolas en forma adyacente. La *iteración* es implantada mediante arreglos de una o dos dimensiones de celdas idénticas. Y la *selección condicional* se la realiza mediante los arreglos lógicos programables PLAs cuya función es determinada por la distribución de los transistores en el arreglo. Estas tres nociones de programación pueden ayudar al diseñador a modularizar sus diseños y a concebir de manera estructurada cualquier circuito de aplicación.

#### 1.3.4 Regularidad

Consiste en la implantación de estructuras priorizando su sistemática, uniformidad y ordenamiento, tendiendo de este modo a simplificar el diseño reduciendo además la probabilidad de error.

Su uso abarca todos los niveles de diseño, desde iteraciones para formar arreglos de celdas idénticas sencillas al nivel físico, hasta la implantación de conjuntos de compuertas iguales al nivel lógico y arquitecturas con estructuras de procesamiento similares a los niveles más altos.

### 1.3.5 Localidad

Mediante la definición de interfaces bien caracterizados para un módulo, se está estableciendo implícitamente que sus módulos internos no son de relevancia para el interfaz externo. De esta forma se realiza una especie de "ocultamiento de información" que reduce la complejidad aparente del módulo. En el mundo del software existe similitud con la minimización de variables globales. Ello permite, por ejemplo, modificar la estructura interna de un módulo sin alterar su interconexión física con el exterior.

*"Gracias a estos principios se puede afirmar que el diseño VLSI actual, es en concepción aun cuando no en forma similar al desarrollo de software".*

## 1.4 ALTERNATIVAS DE DISEÑO DE SISTEMAS ELECTRONICOS

La concepción y desarrollo de sistemas electrónicos se basan fundamentalmente en la realización de placas de circuito impreso PCBs, cuyos progresos han sido considerables tanto en la complejidad de su estructura como en sus aplicaciones. Este progreso se debe esencialmente a que el diseño electrónico se ha enriquecido en los últimos años con nuevas alternativas de circuitos integrados conocidas como (ASICs) que han ampliado el espectro opciones de los ingenieros en aplicaciones (Fig.1.4), puesto que además de permitírseles el uso



de componentes más complejos se les ha abierto la posibilidad de implantar sus propios circuitos integrados de aplicaciones específicas.

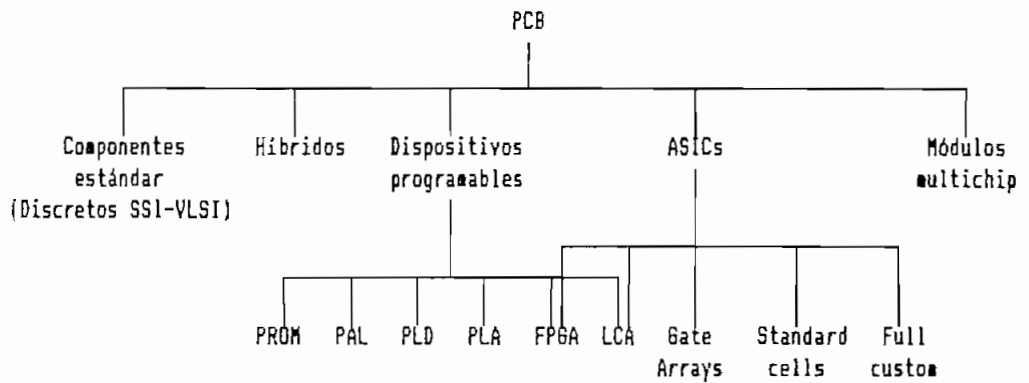


FIGURA 1.4 Diseño de sistemas electrónicos a base de PCB con las diversas modalidades ASICs.

A continuación se realiza una revisión panorámica de las distintas alternativas de diseño de sistemas electrónicos.

#### 1.4.1 PCBs (Printed Circuit Boards)

Las placas de circuito impreso (PCBs) albergan los distintos tipos de componentes y sus interconexiones.

La complejidad funcional de los PCBs ha crecido en forma exponencial a consecuencia de:

- a) El incremento considerable de la densidad de las placas debido a la aparición: de nuevas tecnologías de fabricación de PCBs, de técnicas más aficientes de montaje de

dispositivos y de poderosas herramientas de diseño asistido por computador.

- b) El incremento de la complejidad funcional de los componentes mismos debido a la evolución de las tecnologías de fabricación de los circuitos integrados.

Las siguientes son en resumen las técnicas de montaje utilizadas en los PCBs.

a) Trough-hole (TH)

Los componentes se colocan en la capa primaria insertándolos, de forma manual o automática, a través de los agujeros que para tal efecto se hacen en la placa, y a continuación se sueldan por el lado de la capa secundaria.

b) Surface Mounted Devices (SMD)

Los componentes se montan superficialmente sobre ambas caras de la placa incrementándose así su densidad, cabe anotar que no todos los componentes existen en SMD. El montaje se realiza mediante máquinas *pick & place* con las que primero se coloca la pasta de suelda en toda la placa y luego los componentes, finalmente se realiza la limpieza y extracción de la pasta sobrante.

c) **Tecnologías mixtas**

Se emplean componentes TH y SMD en la cara primaria y componentes pasivos SMD en la secundaria. Se colocan primero los componentes SMD y luego los TH en la cara primaria, finalmente se colocan los SMD en la cara secundaria.

d) **Módulos multi-chip**

Estos módulos son encapsulados que contienen sobre un mismo sustrato cerámico distintos circuitos integrados (ASIC o estándar en forma de dado sin encapsular) conectados entre sí por pistas de metal.

e) **Híbridos.**

En analogía a los módulos muti-chip, sobre un mismo sustrato cerámico que hace las veces de disipador, se colocan los distintos circuitos estándar conectados entre sí por líneas de metal; se pueden incluir además componentes pasivos propios de esta tecnología. Permite de forma bastante flexible mezclar circuitos de distintas tecnologías y reducir la superficie del PCB.

1.4.2 **ASICs (Application Specific Integrated Circuits)**

En función de la metodología de desarrollo, los ASICs

pueden agruparse en cuatro grandes categorías que posteriormente se analizarán con mayor detalle. y que son:

- a) totalmente a medida (*full-custom*),
- b) predifundidos (*semi-custom / gate arrays*)
- c) precaracterizados (*semi-custom / standard cells*), y
- d) lógica programable (*LCAs, PGAs, FPGAs*).

a) Full-Custom

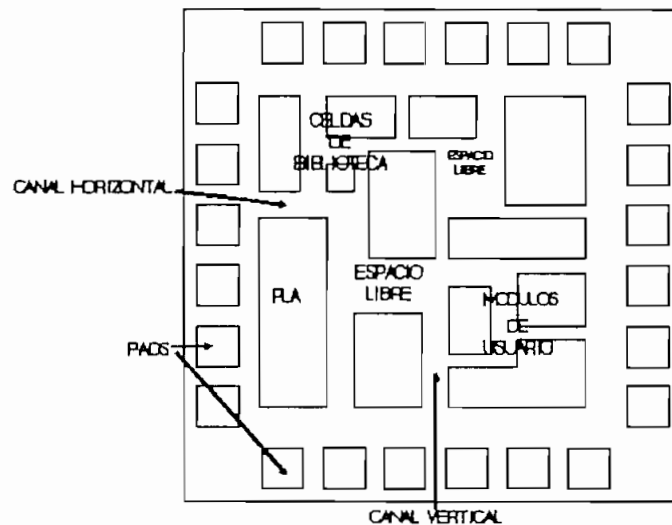


FIGURA 1.5 Ejemplo de un circuito integrado full-custom.

El diseñador de CIs *full-custom* se caracteriza principalmente porque debe disponer de la información, los conocimientos y las herramientas adecuadas para desarrollar sus propias celdas tanto a nivel eléctrico como geométrico, a partir de una descripción estructural del circuito, junto con un primer plano de base del mismo Fig.1.5.

En general, el proceso de diseño *full-custom* consta de tres fases:

- i) Descomposición jerárquica descendente (*diseño top-down*) de las distintas partes de la estructura a implantar en módulos y submódulos hasta llegar a determinar el conjunto de celdas a diseñar.
- ii) Diseño lógico, eléctrico y geométrico íntegro de cada celda, según su entorno de utilización. En algunos casos se puede disponer de bibliotecas de celdas que el diseñador podrá utilizar directamente, o modificarlas creando versiones que se adapten mejor a sus necesidades.
- iii) Composición jerárquica ascendente (*diseño bottom-up*) de las distintas celdas y módulos según el plano de base pre-establecido, hasta obtener el layout completo del circuito.

Se concluye que el diseñador deberá enfrentarse con las descripciones de nivel eléctrico y geométrico, en consecuencia deberá tener conocimientos profundos del proceso tecnológico, esto es:

- i) Comportamiento y características de los dispositivos básicos: transistores, capacidades y resistencias de las distintas capas.

ii) Modelos eléctricos y parámetros asociados para la simulación de tales dispositivos.

iii) Reglas de diseño geométrico para el dibujo de las caras.

Además de una buena dosis de experiencia/paciencia y un conjunto de herramientas CAD aptas para estas descripciones de bajo nivel de abstracción, que dentro de un plazo razonable le permitan finalizar el diseño con ciertas garantías de éxito.

#### Ventajas y desventajas.

Todas las características, ventajas y desventajas de esta metodología se derivan del elevado grado de libertad y control de que el diseñador dispone sobre las descripciones de menor nivel de abstracción de un CI y por ende de mayor complejidad.

Las ventajas podrían resumirse indicando que el diseño *full-custom* permite ajustar de forma muy precisa los compromisos entre superficie, velocidad, funcionalidad y consumo de un CI sin más restricciones que las de la tecnología a utilizar.

Las principales desventajas son:

- i) Se debe diseñar un conjunto completo de máscaras pasando por todos los niveles de fabricación.
- ii) Las herramientas de CAD consumen mucho tiempo de CPU en tareas específicas (simulación eléctrica de las partes más críticas) y en tareas de verificación.
- iii) La dedicación del diseñador debe ser elevada, y no obstante las herramientas de CAD, el riesgo de error y por tanto, la necesidad de rediseño son considerables.

#### **Aplicaciones:**

En general esta metodología trata de maximizar las prestaciones y minimizar el tamaño de los CIs; incidiendo de esta forma sobre el rendimiento del proceso tecnológico y sobre el precio unitario, con vistas a reducir los costes de producción, a costa de fuertes incrementos en los costes de desarrollo. Por ello, esta metodología solo se justifica para volúmenes de producción muy elevados.

Esta tecnología se usa además en el desarrollo de dispositivos de test, analógicos y bibliotecas de celdas, así como en prototipos de investigación tales como procesadores RISC y redes neurales artificiales.

## b) Arreglos de compuertas (Gate arrays)

Estos circuitos, también llamados predifundidos, son CIs semiprocesados que agrupan miles de celdas idénticas, capaces de configurar distintas funciones básicas según la definición del conexionado interno de sus transistores.

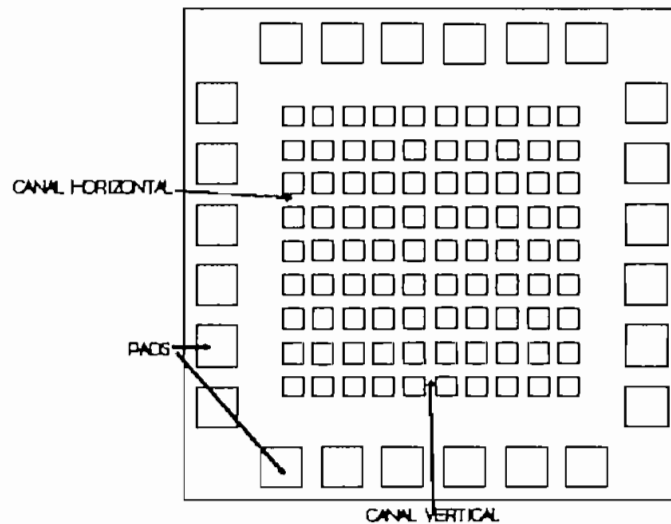


FIGURA 1.6 Ejemplo de Gate Array.

La distribución topológica de base se indica en la Fig. 1.6 y consiste en una disposición regular en forma de matriz en la que se van alternando filas o columnas de celdas, y canales de interconexión. El perímetro de esta matriz está ocupado por los módulos programables de entrada y/o salida. En algunos casos también existen bloques más complejos como memorias ROM o RAM.

El diseño se basa en el uso de una biblioteca de módulos precharacterizados que realizan distintas funciones lógicas, y



celdas de E/S obtenidas por personalización e interconexión de una o más celdas básicas. El enrutamiento entre celdas y entre celdas y módulos de E/S se realiza utilizando los canales de interconexión de tamaño prefijado que para tal efecto existen entre las filas de celdas, usualmente existe un control direccional estricto sobre el enrutamiento (p.ej. metal: vertical, polysilicon: horizontal).

La personalización de las celdas y su conexionado se realizan definiendo las máscaras de uno o más niveles de metal (según la tecnología) y los contactos entre ellos. Por lo tanto la representación del layout de los módulos de biblioteca, solo contendrá la geometría de estas máscaras, puesto que el resto ya está prefabricado sobre el silicio.

#### Ventajas y desventajas.

Las ventajas y desventajas de esta metodología se deben principalmente a que la distribución topológica de base (matriz de celdas rodeada por módulos de E/S) está fijada a priori.

Así pues las principales ventajas son:

- i) Disponibilidad de herramientas de CAD eficaces que simplifican el proceso de diseño.
- ii) Disponibilidad de bibliotecas de celdas standard que

facilitan el trabajo a aquellos diseñadores de sistemas habituados a las mismas.

iii) Menor costo y mayor rapidez de producción debido a que las máscaras, con excepción de las de personalización (1 ó 2 metales + contactos), son comunes a todos los usuarios del mismo tipo de Gate-Arrays.

Como desventajas podemos citar:

i) Las limitaciones de las bibliotecas a disposición del diseñador.

ii) La definición a priori del número de pines de entrada/salida.

iii) La utilización mediocre de la superficie de silicio. Las celdas básicas y los canales no utilizados no pueden compactarse ya que todos los transistores de un arreglo dado se hallan en su posición sean usados o no.

En general podemos decir que esta metodología trata de reducir notablemente los costos de desarrollo, a expensas de incrementar el tamaño del chip.

**Mar de puertas (Sea-of-Gates):**

En el afán de superar el desperdicio de área, los Gate-

Arrays han evolucionado hasta llegar a los llamados *Sea-of-Gates* consistentes en circuitos predifundidos en los que se han eliminado los canales de interconexión entre fila/columna de celdas y éstas ocupan toda la matriz interior.

La celda básica es diseñada para soportar un mayor grado de conectividad y puede ser usada para realizar conexiones o funciones de biblioteca, lo que permite un mejor ajuste de las zonas de conexión a las características de cada circuito mejorándose el aprovechamiento de la superficie de silicio.

### c) Celdas Estandar (*Standard Cells*)

Estos diseños se implantan sobre obleas vírgenes en las que no existe dispositivo preprocesado alguno por lo que cada nuevo circuito pasará por todas las etapas de fabricación.

El diseño se basa en un conjunto de celdas y macroceldas precharacterizadas que cubren un amplio rango de funciones digitales (puertas, registros, FLAs, memorias, procesadores, etc..), analógicas (convertidores A/D y D/A, amplificadores operacionales, osciladores, etc..) y circuitos de entrada-salida (amplificadores + pads). Estas celdas se diferencian de aquellas de los Gate-Arrays en que pueden ser libremente ubicadas en cualquier posición dentro del CI.

Las geometrías (layouts) de las celdas simples contienen todas las máscaras y respetan ciertas restricciones de uni-

formidad que permiten su ubicación en filas o columnas de longitud variable, separadas por canales de interconexión. Las macroceldas (bloques funcionales) digitales y analógicas son de tamaño variable y están separadas entre sí y de las filas de celdas simples por canales de interconexión Fig.1.7.

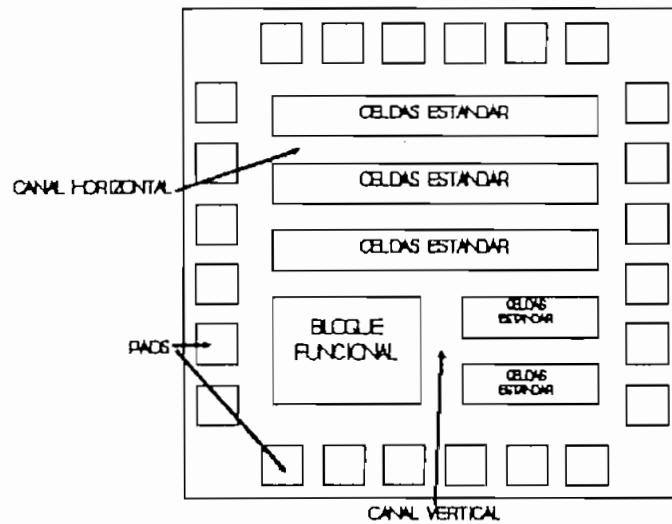


FIGURA 1.7 Estilo de layout en base a celdas estándar.

Los circuitos de E/S tienen todos la misma altura y pueden enlazarse lateralmente entre sí por simple adyacencia y se disponen formando un rectángulo alrededor del conjunto de celdas del núcleo.

#### Ventajas y desventajas.

Las ventajas y desventajas de esta metodología respecto a los Gate-Arrays, se derivan de los grados de libertad adicionales dados al diseñador y al sistema de CAD.

- i) Los algoritmos de ubicación y conexionado son más complejos, pero en general más eficientes.
- ii) Las bibliotecas contienen más variedad de celdas que en los Gate-Arrays y realizan funciones más complejas, adaptándose mejor a distintos tipos de aplicaciones.
- iii) Los circuitos resultantes son de mejores prestaciones y de menor área de silicio.

En contraste se tienen las siguientes desventajas:

- i) Se tiene que implantar un juego complejo de máscaras.
- ii) Los tiempos y costes de desarrollo (diseño + fabricación de prototipos + test) son mayores que en los Gate-Arrays pero mucho menores que en el caso de Full-Custom.

#### d) Lógica programable

Son dispositivos completamente fabricados y encapsulados, cuya funcionalidad se programa desde el exterior.

Estructura.

Los más comunes son los LCAs (Logic Cell Arrays), también conocidos como PGAs (Programmable Gate Arrays). Estos circuitos tienen una estructura regular (similar a un Gate-

Array) consistente en una matriz de bloques lógicos programables CLBs (*Configurable Logic Block*) separados por canales de interconexión lógica programable rodeada de celdas de entrada/salida configurables IOB (*Input/output Block*).

Se concluye que sus estructuras básicas son:

- i) **Bloques lógicos configurables (CLBs)** formados por una parte combinacional capaz de generar distintas funciones lógicas en base a multiplexores, cuyas entradas de datos son celdas de memoria RAM, donde se almacenan los distintos valores de la función; mientras que a las entradas de control se conectan las variables de la función. También dispone de uno o más biestables con distintas configuraciones posibles de *reloj*, *set*, *reset* y *data*.
- ii) **Bloques de entrada/salida configurables (IOBs)**, capaces de realizar funciones de entrada y/o salida en forma directa o a través de latches, pueden además definirse como CMOS ó TTL compatibles y su polaridad.
- iii) **Conexión programable**, formado por canales que separan las filas y columnas de CLBs, las conexiones se realizan en base a matrices de conmutación de señales programables. Existen además líneas específicas para cablear señales globales y otras para realizar interconexiones directas entre CLBs adyacentes.

Para su programación cada bloque dispone de celdas de memoria RAM, en las que se almacena su configuración. Debiendo por lo tanto existir una circuitería interna adicional para acceder a esta RAM distribuida.

#### Funcionamiento.

Existen dos fases durante el uso normal de un LCA dentro de una aplicación:

- i) **Inicialización:** al activarse el dispositivo, la RAM del LCA se carga con la configuración prevista desde un elemento externo (ROM, procesador, ordenador,...).
- ii) **Funcionamiento:** una vez configurado, el LCA cambia de estado y pasa a ejecutar la función programada en la RAM.

#### Diseño:

Respecto al proceso de diseño con LCAs se pueden distinguir tres fases fundamentales:

- i) **Captura de esquemas y simulación lógica,** basada en una biblioteca de celdas y/o descripciones lógicas.
- ii) **Personalización del diseño para configurar un LCA.** En primer lugar se hace la asignación y partición de los

distintos módulos de los circuitos CLBs e IOBs y a continuación se define la configuración de cada bloque y del conexionado necesario. Se trata pues de la fase de ubicación y conexionado de celdas, cuyo resultado final no es un *layout geométrico*, sino un *layout lógico* formado por el conjunto de *bits* para programar la función deseada sobre el LCA.

iii) **Implantación física del circuito y verificación en tiempo real.** Consistente en la transferencia sobre el LCA del programa obtenido en la fase anterior, y comprobación de su funcionalidad. Se asume en este caso que los dispositivos utilizados están libres de errores del proceso tecnológico.

#### **Ventajas, desventajas y aplicaciones.**

La principal ventaja se desprende de la reducción del tiempo de fabricación y test de prototipos (5-10 semanas), el cual consiste de un simple proceso de carga de una memoria RAM (unos pocos segundos) seguido de una comprobación funcional (unos cuantos minutos y horas, según el equipo disponible). Por lo tanto se reducen drásticamente los costes de desarrollo, y en especial los derivados del rediseño

Su mayor desventaja radica en el pobre aprovechamiento del área de silicio, y en la complejidad relativamente baja de los circuitos resultantes (<10.000 puertas).



En consecuencia, esta es una buena alternativa para el desarrollo económico y rápido de prototipos y series de circuitos de complejidad y volumen de producción bajos.

## 1.5 TECNOLOGIAS DE FABRICACION

Antes de la concepción de un circuito VLSI el diseñador deberá decidir la tecnología de fabricación, puesto que en función de ella se construirán las topologías circuitales y en consecuencia la estructura misma de las máscaras del LAYOUT que se obtengan.

La tecnología que se seleccione dependerá entre otros factores de:

- a) La aplicación para la cual se desarrolla el circuito VLSI propuesto.
- b) Las limitaciones tecnológicas que imponen las fundidoras de silicio, ya que son ellas quienes en última instancia definen los parámetros de los circuitos que están en capacidad de producir con características de funcionamiento óptimas.

De un modo general las tecnologías de fabricación que se han desarrollado hasta la actualidad son:

- a) Bipolar.
- b) Metal-Oxido-Semiconductor (MOS).
- c) Arseniuro de Galio (GaAs).
- d) CMOS-Bipolar (BiCMOS).

#### 1.5.1 Bipolar

La tecnología bipolar más difundida es la *Advanced Low Power Schottky* (ALS), debido a su capacidad de conmutación de muy alta velocidad, bajo ruido y reducción gradual del consumo de potencia consecuencia del mejoramiento de las tecnologías de procesamiento. Se emplea en aplicaciones tales como los circuitos integrados para "mainframes" de gran capacidad. La disipación de calor de los circuitos bipolares es bastante grande por lo que se requieren arreglos de enfriamiento especiales.

#### 1.5.2 Metal-Oxido-Semiconductor (MOS)

Es la más apropiada para circuitos VLSI prototipos, debido a su alta capacidad de integración, pasos de procesamiento relativamente simples, y consumo de potencia bajo. Esta tecnología se está desarrollando rápidamente debido a su gran volumen de aplicaciones, las distancias mínimas son cada vez menores, lo que origina circuitos de mayor integración con confiabilidad probada. Así:

"Para 1985 la tecnología CMOS era de 1.5 micras llegándose a 0.5 micras en 1990. El número máximo de compuertas por CI en 1985 era aproximadamente de 65,000 llegando a 800,000 en 1990 (aproximadamente 4 transistores por compuerta). El retardo era de 2.5 ns por compuerta en 1985 reduciéndose a 1 ns en 1990. La tecnología de encapsulado de los CIs también ha mejorado significativamente de 80 pines por chip en 1985 a 1000 pines en 1990. La tecnología de definición de máscaras usaba comunmente haces de electrones y rayos-X hasta 1990 en que se usaron rayos-X exclusivamente".<sup>(2)</sup>

Las dos tecnologías MOS más difundidas son nMOS y CMOS, siendo esta última la más difundida en la actualidad, su mayor ventaja sobre la tecnología nMOS radica en su mínimo consumo de potencia estática. La tecnología nMOS por otro lado tiene ciertas ventajas en cuanto al área de silicio que el circuito requiere para obtener una funcionalidad equivalente.

### 1.5.3 Arseniuro de Galio (GaAs)

Los circuitos implantados en esta tecnología son altamente funcionales debido a su velocidad de conmutación extremadamente alta, baja disipación de potencia interna, alta temperatura de operación y propiedades de anti-radiación. Se espera que conforme los costes de fabricación de CIs de GaAs bajen, las aplicaciones comerciales vayan haciéndose factibles.

---

<sup>(2)</sup> "Introduction to nMOS and CMOS VLSI Systems Design", A. Mukherjee, pág. 7.

La tecnología de GaAs tiene grandes perspectivas en aplicaciones espaciales y militares debido a que los dispositivos construidos en base a ésta presentan excelentes características en ambientes con radiación y temperatura variables.

#### 1.5.4 CMOS-Bipolar (BiCMOS)

Los procesos bipolar y CMOS, han sido recientemente combinados con éxito en una misma oblea de silicio simple, obteniéndose como resultado la tecnología BiCMOS en la que se aprovechan las mejores características de los dos procesos; esto es la alta velocidad de conmutación del BJT y el bajo consumo de potencia del CMOS.

### 1.6 ASPECTOS ECONOMICOS DEL DESARROLLO Y PRODUCCION DE CIs

Conviene analizar someramente los aspectos económicos de la concepción y desarrollo de CIs tanto más cuanto la evolución del diseño en general y el de ASICs en particular se mueven guiados fundamentalmente por la economía.

Una expresión simplificada de los costes asociados al desarrollo y producción de un CI es la siguiente:

$$COSTE\ CI = f (C_{DESARROLLO} + C_{PRODUCCION})$$

[1.1]

$$C_{DESARROLLO} = C_{DISEÑO} + C_{PROTOTIPOS} + C_{TEST}$$

$$+ (C_{REDISEÑO} + C_{PROTOTIPOS} + C_{TEST}) * nrd$$

[1.2]

donde *nrd* se refiere al número de rediseños.

Por otra parte, los costes de producción serán:

$$C_{PRODUCCION} = Volumen * C_{UNITARIO}$$

[1.3]

Las diferentes alternativas de diseño y fabricación de CIs, tratan de reducir estos costes, incidiendo sobre el factor adecuado según el tipo, características y volumen de producción de cada circuito.

Siendo el objetivo global la disminución de costes asociados al desarrollo y producción del CI debe actuarse sobre cada uno de los parámetros involucrados: diseño, test, prototipos, costo unitario y volumen, y determinar en cada caso particular cual es la solución adecuada en función de las inter-relaciones existentes entre ellos.

### 1.6.1 Costes de producción

Los costes de producción están siempre dominados por el tamaño del *dato* (cuadrado de silicio que contiene el CI), que incide en el número de *datos* que entran en cada lote de obleas y por ende en el rendimiento del proceso tecnológico.

Es esencial tener en cuenta que:

"Las dimensiones mínimas que se ofrecen no significan ningún valor absoluto con significación en sí mismo, sino únicamente un compromiso obtenido estadísticamente entre la calidad de los procesos de producción, la infraestructura y el número de defectos de las máscaras (p.ej. 3 defectos/cm, 0.1 defecto/cm), frente a un rendimiento de fabricación aceptable económicamente por el mercado." (3)

Considerando que el coste de producción unitario del circuito integrado viene dado por la siguiente expresión:

$$C_{\text{UNITARIO}} = \frac{\text{Coste oblea}}{\text{CIs/oblea} * \text{Rendimiento}} + C_{\text{ENCAPSULADO}} + C_{\text{TEST}}$$

[1.4]

se tiene que los costes de producción se reducirán mediante:

- a) El incremento del rendimiento en la fabricación, que se consigue en base a mejorar la calidad de los procesos tecnológicos en general;

---

(3) "ASIC's Metodologías y Herramientas de Diseño", J. Llobet - L. Teres, pág. 9.

- b) El aumento del tamaño de las obleas procesadas lo que incrementa el número de circuitos por oblea, y por tanto, por lote de fabricación; y
- c) La disminución de las dimensiones mínimas de la tecnología lo que permite también tener un mayor número de CIs por oblea.

#### 1.6.2 Costes de desarrollo

Respecto al coste de desarrollo el factor más importante a minimizar es el riesgo de rediseño ya que al margen de la gravedad del error, los costes derivados siempre son considerables.

La disminución de los *costes de diseño* viene marcada por la eficacia de las herramientas de ayuda al diseño disponibles en la medida en que permitan al diseñador colocar más ítems/día en el CI objeto de diseño o que se aumente la complejidad de tales ítems, asegurando siempre su colocación correcta en el CI y, al mismo tiempo, facilitando su verificación, simulación y test previo a la fundición.

En cuanto a los *costes de prototipado*, las tres primeras categorías de CIs (full-custom, precharacterizados y predifundidos) corresponden a circuitos cuya realización física sobre silicio, requiere de todas o parte (predifundidos) de las

etapas del proceso tecnológico de fabricación de circuitos integrados. Por tanto, para estos casos hay que tener en cuenta que los costes de los prototipos serán elevados y que cualquier rediseño posterior exigirá una nueva fase de fabricación de prototipos.

En contraste, en los CIs de lógica programable sólo se requiere de la programación post-fundición del chip, para configurarlo según la función que se deba realizar. La rapidez y economía de las fases de desarrollo, hacen de esta alternativa ideal para implantar el prototipo y pequeñas series para circuitos de complejidad mediana (<10.000 puertas).

### 1.6.3 Conclusión.

La gran diferencia entre las distintas opciones de diseño radica en los aspectos de concepción y diseño, y en el tipo de proceso tecnológico final para la realización y test de los circuitos y las connotaciones en cuanto a costos y prestaciones que ello supone.

Existen distintas alternativas de realización de ASICs en función de la decisión sobre el factor a reducir de las expresiones de costos de la sección anterior. Cuando el volumen de producción es pequeño la parte dominante del costo corresponde al desarrollo ( $C_{desarrollo} > C_{producción}$ ); en este caso es necesaria una alternativa que permita reducir



los costes de desarrollo, aunque sea a expensas de incrementar el coste unitario durante la producción. Hacia este tipo de soluciones tienden los *gate-arrays*, *standard cells* y los *programmable gate-arrays*.

El extremo opuesto correspondería a aquellos casos en que el volumen de producción es muy elevado ( $C_{\text{producción}} \gg C_{\text{desarrollo}}$ ). Entonces se debe reducir el precio unitario (reducir las dimensiones del circuito) aunque ello imponga un incremento considerable en los costes de diseño. Este extremo corresponde al *full-custom*.

## CAPITULO 2

### FUNDAMENTOS DE CIRCUITOS CMOS

#### 2.1 GENERALIDADES

El diseñador de circuitos integrados (CIs) opera en un campo de complejidad siempre creciente, en el cual la sofisticación del diseño siempre tiende a ser mayor. Para trabajar de forma eficaz, el diseñador debe explotar totalmente el elemento primitivo utilizado en el diseño de sus circuitos; una comprensión básica de dicho elemento y su operación son esenciales para que el diseñador ejecute su tarea exitosamente y esté listo a absorber y a evaluar los avances provenientes de las investigaciones realizadas por los especialistas en el campo de diseño de circuitos VLSI.

En el diseño de CIs utilizando tecnología CMOS, el elemento primitivo de diseño es el transistor de efecto de campo (*Field Effect Transistor*) del tipo MOS.

Un transistor de efecto de campo es un dispositivo semiconductor cuya operación depende del control de su co-

riente realizado por medio de un campo eléctrico. Existen dos tipos de transistores de efecto de campo:

- a) El de juntura, denominado JFET (*junction field-effect transistor*), y
- b) El de compuerta aislada, denominado IGFET (*insulated-gate field-effect transistor*), más comunmente conocido como MOSFET (*metal-oxide-semiconductor field-effect transistor*) o simplemente transistor MOS.

Hay dos tipos básicos de transistores MOS: el transistor canal "n" (nMOS) y el transistor canal "p" (pMOS). La tecnología CMOS (*Complementary Metal-Oxide-Semiconductor*) involucra los dos tipos de transistores.

En los últimos años, la tecnología CMOS ha jugado un papel cada vez más importante en el mundo de la industria de los CIs. Al contrario de lo que puede pensarse, la tecnología CMOS no es nueva, en 1925 se planteó el principio básico del MOSFET y en 1935 ya se propuso una estructura muy similar a la de los transistores MOS modernos.

Fueron problemas con los materiales los cuales frustraron estas propuestas iniciales y más bien los experimentos con los primeros transistores de efecto de campo condujeron a la invención del transistor bipolar; dado el éxito de los transistores bipolares se perdió interés en los MOS.

Cuando se inventó el proceso planar de Silicio, al inicio de los 60, los transistores MOS tomaron actualidad. Sin embargo, nuevos problemas con los materiales y control de calidad hicieron que la introducción del dispositivo MOS, en usos comerciales, se retrase hasta 1967. Incluso entonces, se alentó únicamente el desarrollo de los transistores de una polaridad, "p" o "n".

El uso de los dos tipos de transistores MOS en un mismo CI (tecnología CMOS), utilizando el mismo sustrato o base, se reservó para aplicaciones de baja potencia (como en los relojes) y puesto que la tecnología requerida en la fabricación de estos circuitos era más compleja que la utilizada en los de simple polaridad, la tecnología CMOS fue generalmente evitada del diseño de sistemas.

Más tarde, debido al incremento de la complejidad de los procesos de fabricación nMOS, la complejidad adicional que implicaba el proceso básico CMOS disminuyó comparativamente en importancia. Además, los diseñadores se encontraron frente a CIs de tamaños muy grandes, con altos consumos de potencia.

Por las razones expuestas y otras adicionales, la tecnología CMOS ha crecido en nivel de importancia como una tecnología VLSI. Se pueden mencionar entre sus principales características las altas densidades de empaquetamiento, bajo consumo de potencia, excelente inmunidad al ruido, amplio rango de voltajes de polarización.

## 2.1.1 Descripción básica de la estructura física de un Transistor MOS

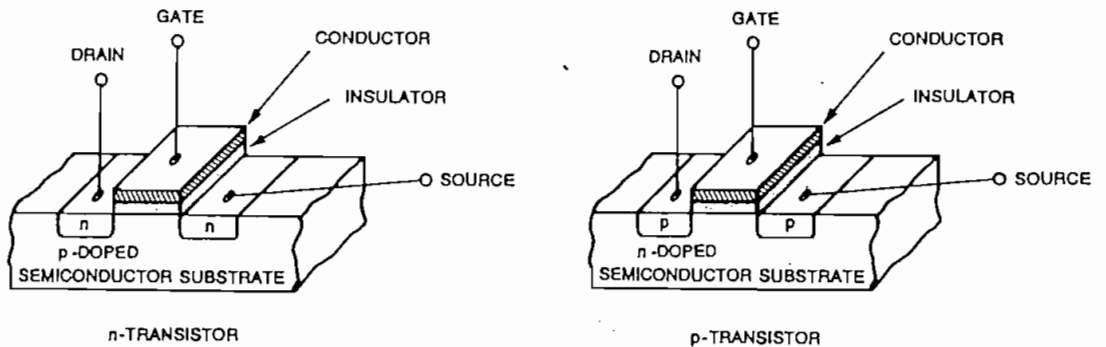


FIGURA 2.1 Estructuras físicas de los transistores "n" y "p".

En la Fig. 2.1 se presenta una vista de la sección transversal de la estructura típica para dos tipos de transistores MOS ya fabricados. Para el transistor "n", la estructura consiste de una sección de silicio tipo "p" (sustrato), parte de este material se encuentra separando dos áreas de silicio tipo "n". Estas dos áreas se forman por la inyección de impurezas que se difunden en el sustrato y reciben los nombres de sumidero o drenaje (*drain*) y fuente (*source*).

En el área que separa las regiones "n" se forma el canal, el cual junto con el terminal llamado compuerta (*gate*) forman las tapas de un "sánduche" (condensador) cuyo elemento central es una fina capa de aislante (*insulator*). El concepto de canal se aclara más adelante en este capítulo.

Para el caso del transistor tipo "p", la estructura consiste de: una sección de silicio tipo "n" (sustrato) que separa dos áreas de difusión tipo "p", y de todos los elementos necesarios para formar el sánduche de la compuerta.

La justificación del nombre asignado a cada tipo de transistor se presenta más adelante al explicar el funcionamiento del transistor MOS. Como se desprende de la Fig. 2.1, las estructuras de los transistores MOS se crean por la superposición de algunas capas de materiales conductores y aislantes.

### 2.1.2 Breve descripción de los materiales utilizados en la fabricación del Transistor MOS

A continuación se presenta una corta explicación de la estructura de los semiconductores y de la formación de nuevos materiales que se obtienen al agregar impurezas a base de semiconductores puros, se pretende también introducir cierta terminología utilizada en el desarrollo de este capítulo. Esta explicación introductoria no pretende ser un estudio detallado y mucho menos profundo de los semiconductores, el cual puede encontrarse en libros de Física del Estado sólido.

En las estructuras presentadas en la Fig. 2.1, los diferentes terminales (drenaje, fuente, compuerta y sustrato), para su interconexión con otros transistores requieren utili-

zar metal. Por lo tanto son partes constitutivas de los transistores, diversos materiales: conductores (como el Aluminio), aislantes (como el Dióxido de Silicio,  $\text{SiO}_2$ ) y semiconductores (como los materiales tipo "n", tipo "p", tipo "n+" o tipo "p+").

Un conductor es un término que se aplica a cualquier material que permita el flujo abundante de carga. Un semiconductor es un material que tiene un nivel de conductividad intermedio entre los extremos de un aislador, que presenta muy baja conductividad, y un conductor, que presenta un nivel alto de conductividad.

Los materiales semiconductores que han recibido el mayor grado de interés son el Germanio y el Silicio. Una de las razones que ha concentrado la atención en estos elementos es debido a la posibilidad de fabricarlos con un alto grado de pureza. Por ejemplo, pueden fabricarse materiales en el que se encuentra una impureza en 10000 millones de partes de material puro.

Los niveles de pureza son importantes ya que el agregar una parte de impureza, del tipo adecuado, por millón de material puro, por ejemplo, puede cambiarse la característica del material, de conductor pobre a uno muy bueno.

Al material semiconductor puro se le denomina también intrínseco. Para fines prácticos, materiales intrínsecos son

materiales cuidadosamente refinados para reducir las impurezas a un nivel muy bajo, al nivel que lo permita la tecnología.

La habilidad para cambiar significativamente las características del material semiconductor, relativamente puro, agregando impurezas, se denomina "doping". El material resultado del proceso de doping recibe el nombre de material extrínseco. En realidad el material usado en la fabricación de dispositivos debe tener una alta pureza inicial y posteriormente debe ser dopado, según se requiera.

#### a) Materiales tipo "p" y tipo "n"

Hay dos materiales extrínsecos de importancia para la fabricación de dispositivos semiconductores, el tipo "n" y el tipo "p". Los dos materiales se forman agregando un número predeterminado de átomos de impurezas a una base de Ge o Si.

Para dejar claro la formación de los materiales extrínsecos, es necesario revisar brevemente el concepto de valencia. En la Fig. 2.2 se presenta al átomo de Ge con sus 32 electrones y al de Si con sus 14, distribuidos en sus niveles internos y externos. En cada caso hay 4 electrones en la corteza más externa, este número de electrones, en la corteza más externa, determina la valencia. La valencia no es más que la capacidad de un átomo para agruparse con otros, tendiendo a formar una estructura que tenga ocho electrones



en la corteza más externa, que es energeticamente la estructura más favorable.

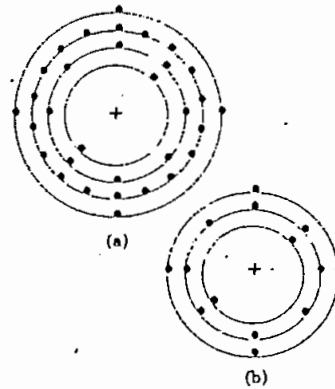


FIGURA 2.2 Estructura atómica: a) Germanio y b) Silicio.

Para el caso del Si y Ge puros, los cuatro electrones de valencia de cada átomo están unidos a cuatro átomos adyacentes; es decir, cada uno de los átomos comparte un par electrónico con 4 con cuatro átomos vecinos, formando un enlace denominado covalente. El enlace covalente se caracteriza entonces por la compartición de electrones por átomos adyacentes, lo que se traduce en un enlace más fuerte. En la Fig. 2.3 se esquematiza lo mencionado anteriormente, para el caso del Si.

A temperaturas muy bajas ( $0\text{ }^{\circ}\text{K}$ ), la estructura ideal de la Fig. 2.3 se aproxima a la realidad y los semiconductores se comportan como aislantes, puesto que no hay portadores de carga libres. Sin embargo, a temperatura ambiente, a pesar del enlace covalente formado, es posible todavía que los electrones de valencia absorban suficiente energía cinética,

a partir de causas naturales, para romper los enlaces covalentes y adquirir el estado "libre". Causas naturales pueden ser energía luminosa en forma de fotones o energía térmica de los alrededores entregadas al semiconductor.

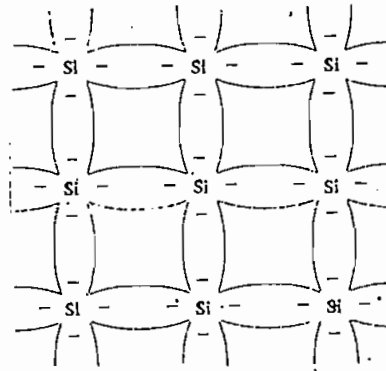


FIGURA 2.3 Enlace covalente en el átomo de Silicio.

Un cambio de temperatura en un material semiconductor puede alterar sustancialmente el número de electrones libres. Más aún, un aumento de temperatura presenta una reducción en la resistencia presentada por el semiconductor.

A temperatura ambiente hay aproximadamente  $1.5 \times 10^{10}$  portadores libres en un centímetro cúbico de material intrínseco de Si y  $2.5 \times 10^{10}$  para el Ge, son cifras altas pero en relación a metales como el cobre, estos materiales en estado intrínseco son considerados como conductores pobres.

i) Material tipo "n".

El material tipo "n" se forma añadiendo átomos llamados pentavalentes (valencia=5) al semiconductor intrínseco. Estos átomos de impureza tienen cinco electrones en su nivel de valencia y al formar la nueva estructura solamente cuatro de estos electrones forman parte de los enlaces covalentes, quedando el quinto disponible para la conducción. En el ejemplo de la Fig. 2.4 se muestra la estructura resultante de añadir Antimonio a una base de Si.

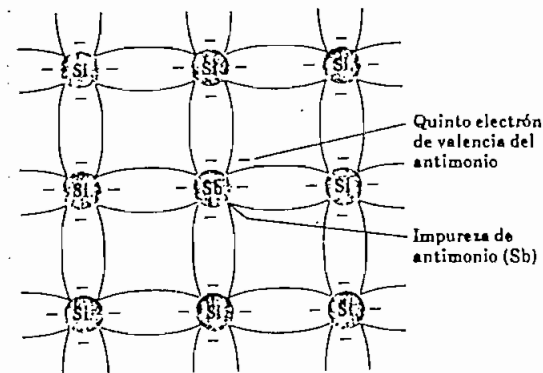


FIGURA 2.4 Impureza de Antimonio añadida a una base de Si para formar el material tipo "n".

En el material tipo "n", el quinto electrón está débilmente ligado a su átomo padre de impureza y es relativamente libre de moverse en el material formado. Sin embargo, se debe recordar que a pesar de los portadores libres añadidos, el material tipo "n" es aún eléctricamente neutro ya que los átomos de impureza tienen tantos protones cargados positivamente en su núcleo, como electrones cargados negativamente en sus órbitas.

Frente a la aplicación de un voltaje suficiente o por efecto de la temperatura, el electrón relativamente libre, abandona la impureza, y la impureza queda ionizada. Debido a que el átomo añadido dona el electrón para la conducción, a estos átomos de impureza se les denomina donadores. Elementos utilizados como donadores son por ejemplo: el Antimonio, Arsénico y Fósforo. En este tipo de material el portador de carga es el electrón, dado que su carga es negativa, a este material se lo denomina material "n".

#### ii) Material tipo "p"

El material tipo "p" se consigue dopando un semiconductor intrínseco con átomos que tienen un número insuficiente de electrones para completar el enlace covalente.

Los elementos utilizados tienen tres electrones de valencia, por lo que al agregarse al Si o Ge, en la estructura resultante se tiene un electrón faltante para formar uno de los enlaces covalentes. La vacante resultante se denomina hueco, esta vacante está lista a aceptar un electrón libre. Dado que son las impurezas agregadas las cuales aceptan los electrones, se denominan **aceptoras**.

Los átomos añadidos forman la estructura típica de la Fig. 2.5, en la que se esquematiza el caso particular en el que se añade Boro a una base de Si.

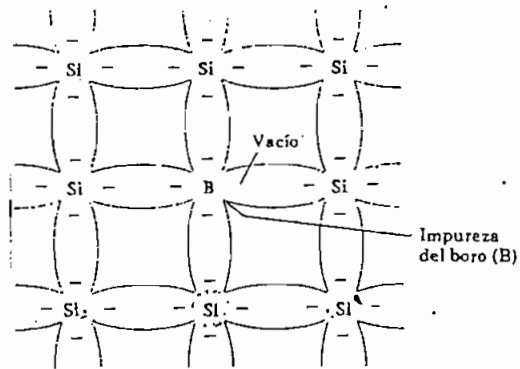


FIGURA 2.5 Impureza de Boro para formar el material tipo "p".

De igual manera que el material tipo "n", el tipo "p" continúa siendo eléctricamente neutro. Los elementos utilizados como impurezas aceptoras son por ejemplo: el Indio, Boro, Galio.

El efecto del hueco en la conducción se muestra en la Fig. 2.6. Cuando un enlace está incompleto, de tal manera que existe un hueco, es relativamente fácil para un electrón de valencia de un átomo vecino dejar su enlace covalente para llenar este hueco. Un electrón moviéndose de un enlace para llenar un hueco deja un hueco en su posición inicial, por lo que puede afirmarse que el hueco se mueve efectivamente en dirección opuesta a la del electrón. Aquí se tiene un mecanismo de conducción de electricidad que no involucra electrones libres. El portador en el material tipo "p", por lo tanto, se comporta como si tuviese una carga positiva de valor absoluto igual a la del electrón, y de ahí justamente el nombre "p" del material.

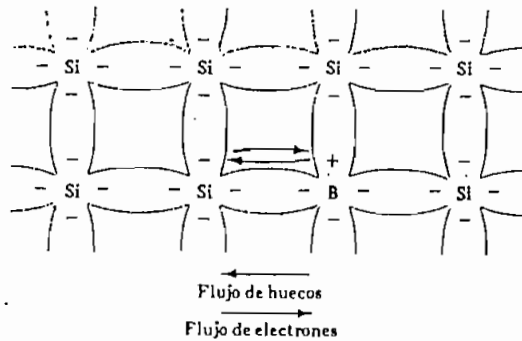


FIGURA 2.6 Flujo de electrones y huecos.

En el estado intrínseco, el número de electrones libres en el Ge y Si se deben solamente a aquellos pocos electrones de valencia que han adquirido suficiente energía de fuentes térmicas o luminosas para romper el enlace covalente, o las pocas impurezas que no pudieron retirarse. Las vacantes dejadas atrás por los electrones en la estructura del enlace covalente representan el limitado suministro de huecos.

En un material tipo "n" el número de huecos no ha cambiado significativamente respecto a los presentes en el material intrínseco. El resultado neto de las impurezas añadidas es que el número de electrones sobrepasa el número de huecos, por esta razón, al electrón se le denomina portador mayoritario y al hueco portador minoritario. En el material tipo "p" se tiene que el hueco es el portador mayoritario y el electrón el minoritario.

b) Materiales tipo "n+" y tipo "p+"

Como se ha explicado, puede disponerse de portadores de carga mayoritarios, electrones o huecos, dopando adecuadamente el material semiconductor. La concentración de los dopantes puede controlarse de manera exacta en muchos órdenes de magnitud, se tienen:

- i) Materiales dopados ligeramente (materiales que se notan "p" y "n"), aproximadamente  $10^{15}$  átomos de impureza por centímetro cúbico.
- ii) Materiales fuertemente dopados, con aproximadamente  $10^{19}$  átomos de impureza por centímetro cúbico. La densidad de átomos de impureza es siempre pequeña comparada a los aproximadamente  $5 \times 10^{22}$  átomos por centímetro cúbico del material intrínseco. Los materiales tipo "n" fuertemente dopados se notan como "n+" y para el caso de materiales fuertemente dopados del tipo "p", se utiliza la nomenclatura "p+".

Para la fabricación de transistores MOS, ya sea a nivel de elemento discreto o a nivel de CI, el elemento semiconductor utilizado generalmente es el Silicio. Para fabricar los transistores MOS, se utilizan: materiales tipo "p" y "n" para los sustratos y materiales tipo "p+" y "n+" para la fuente y drenaje. Las impurezas más utilizadas en tecnología MOS son el Boro y el Fósforo.

### c) Cristales

En algunos sólidos, los átomos o grupo de átomos están dispuestos de una manera irregular, más o menos aleatoria, como los átomos o moléculas en un líquido; tales sólidos son llamados no cristalinos o amorfos. En otros sólidos los átomos o grupos de átomos están dispuestos en un orden regular; tales sólidos se llaman cristalinos.

Si el orden regular se extiende sobre un trozo completo de material, dicho material se denomina monocristal; si la regularidad se extiende solo sobre una parte pequeña del material, que consiste entonces de una aglomeración de microcristales, se dice que el material es policristalino.

Los átomos del Ge y Si forman un sólido cristalino, cuyo rasgo fundamental es la periodicidad de su estructura, la cual consiste de un arreglo de átomos con un patrón regular y repetitivo. Este patrón se repite de manera regular e infinita en el espacio (tres dimensiones).

Como ya se mencionó, en el Ge y el Si un átomo puede compartir pares electrónicos con cuatro vecinos, los cuales están dispuestos en forma simétrica en torno a él; de esta manera, los cuatro vecinos forman un tetraedro regular, con el átomo mismo ubicado en el centro. Esto da origen a una disposición cúbica (tridimensional) conocida como estructura de diamante, que se indica en la Fig. 2.7. Las representacio-



El fabricar la compuerta de material semiconductor tiene algunas ventajas, pues el utilizar *polysilicon* permite una mayor integración y reduce la diferencia en el potencial de contacto, entre el electrodo de la compuerta y el dieléctrico ( $\text{SiO}_2$ ), lo que permite a la vez una reducción del voltaje umbral ( $V_t$ ), necesario para iniciar la conducción del dispositivo (este parámetro se detalla más adelante).

La reducción del valor de  $V_t$  permite que los circuitos MOS fabricados de esta manera requieran menores voltajes de polarización, lo que les permite ser directamente compatibles con CI bipolares, lo cual a su vez facilita el uso directo de CI bipolares y MOS dentro de un mismo sistema. La reducción del valor de las fuentes de polarización necesarias permite obtener menores tiempos de conmutación ya que se tiene una excursión de voltaje más pequeña en los instantes de conmutación (de ON a OFF y viceversa) de los transistores, lo que implica una operación más rápida que su equivalente utilizando compuerta metálica.

## 2.2 EL TRANSISTOR MOS COMO SWITCH

Para analizar la operación de los dispositivos MOS se ha concebido un modelo abstracto simplificado en el que se los considera como simples switches cerrados y abiertos (*on/off*). Esta representación se justifica entendiendo el proceso básico de operación del transistor MOS.

En un transistor MOS la compuerta es un terminal de control, que determina si existe o no flujo de portadores de la fuente al drenaje, lo que se traduce en la presencia o no de un camino entre estos dos terminales. Por lo tanto, drenaje y fuente pueden ser tratados como los terminales de un switch. Puede notarse también que el dispositivo presenta simetría bilateral, es decir que las regiones de fuente y drenaje son físicamente intercambiables y no están definidas hasta que el dispositivo está conectado a los nodos de un circuito. Esta característica ratifica la representación del transistor MOS como un simple switch.

En la Fig. 2.8 se muestra la representación de los dos switches con transistores MOS, *N-SWITCH* y *P-SWITCH*, junto con su representación circuital y el comportamiento de los mismos para permitir el paso del "0" y "1" lógicos. Para la figura se asignan las letras: "s" a la compuerta, "a" al drenaje y "b" a la fuente.

El *N-SWITCH* se cierra (ON) si en su compuerta se tiene un "1" y se abre (OFF) si se tiene un "0". El *N-SWITCH* es un switch casi perfecto (*good*) cuando debe permitir el paso de un "0" (desde uno de sus extremos al otro), pero es imperfecto (*poor*) al dejar pasar un "1" ya que produce una disminución del nivel de voltaje respecto al voltaje ingresado.

El *P-SWITCH* presenta características diferentes, éste se cierra (ON) cuando se tiene en la compuerta un "0" y se abre

(OFF) con un "1". Este switch es imperfecto (*poor*) para el paso de un "0" y casi perfecto (*good*) para el paso de un "1". Estas características se justifican en el numeral 2.5.

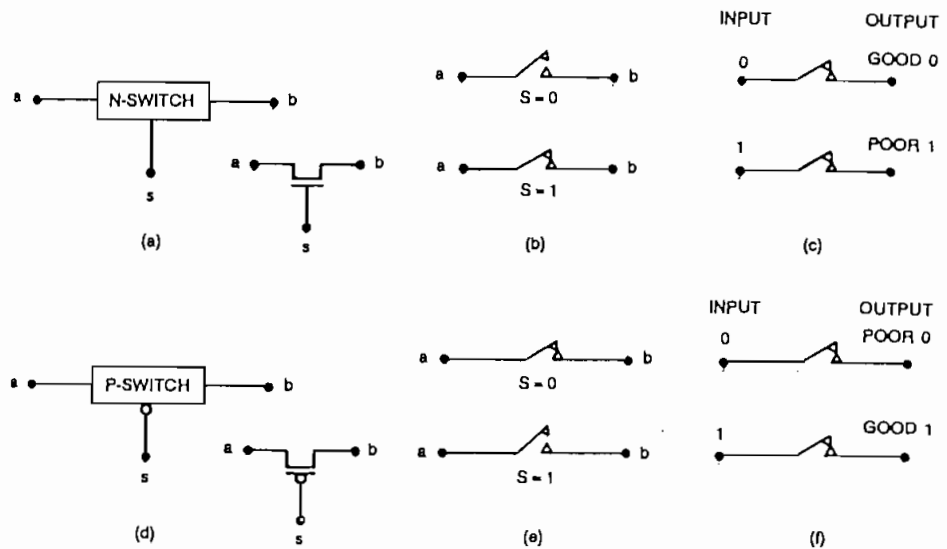


FIGURA 2.8 Transistores MOS vistos como switches.

Debe notarse que el N-SWITCH y el P-SWITCH están ON/OFF para valores complementarios de la señal de la compuerta. Esta diferencia se indica incluyendo una circunferencia en la compuerta de la representación de los transistores "p".

Combinando los dos tipos de switch en paralelo, como se indica en la Fig. 2.9, se obtiene un switch (C-SWITCH) que permite el paso en condiciones óptimas tanto del "0" como del "1". Para el funcionamiento adecuado del C-SWITCH, es necesario enrutar a las compuertas, la señal de control al transistor "n" y su complemento al transistor "p".

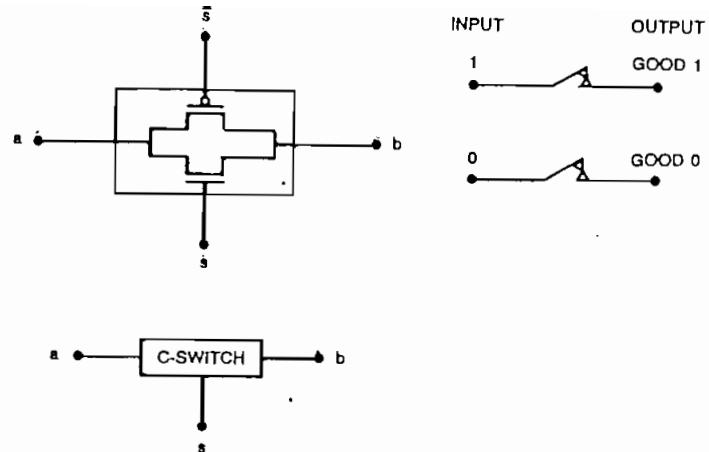


FIGURA 2.9 Switch complementario.

Se debe mencionar que en un circuito en el cual el switch se utiliza para pasar unicamente un "0" o un "1", debe eliminarse el sub-switch (N o P) del C-SWITCH, retornando al N-SWITCH o al P-SWITCH.

### 2.2.1 Redes combinatoriales CMOS

Una red digital puede construirse interconectando los elementos primitivos escogidos, en este caso los transistores MOS, siguiendo un conjunto de reglas de interconexión. En este numeral se determinan dichas reglas de interconexión.

#### a) El inversor

Cuando a la entrada del inversor se tiene un "0", a su salida se espera un "1"; esto sugiere el uso de un transistor

"p" que permita el paso de un "1", desde un punto que tenga este voltaje ( $V_{DD}$ ), hacia la salida del inversor.

Cuando a la entrada del inversor se tenga un "1", se espera a su salida un "0", esto sugiere un transistor "n" conectado a la salida y a un punto que entregue un "0" ( $V_{SS}$  o Gnd). El circuito completo que realiza las tareas descritas se presenta en la Fig. 2.10.

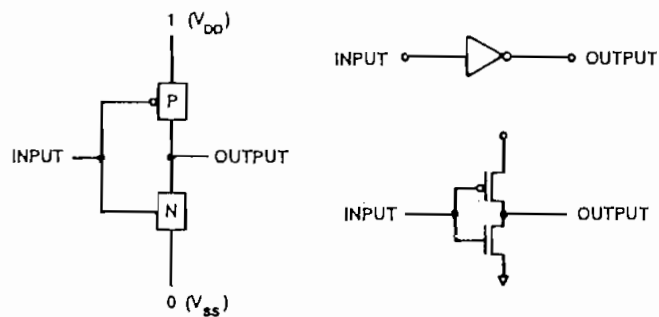


FIGURA 2.10 Construcción de un inversor CMOS.

Nótese que no se utilizan switches complementarios, sino únicamente un N-SWITCH y un P-SWITCH. Puesto que el switch inferior debe garantizar el paso adecuado de un "0", es necesario para este caso un N-SWITCH. Algo similar ocurre con el switch superior, en este caso se debe garantizar el paso adecuado de un "1" y es suficiente un P-SWITCH. La Fig. 2.10 incluye también la representación del inversor a nivel de transistor, con las conexiones a  $V_{DD}$  y Gnd. Con el arreglo presentado se garantiza que la salida nunca quede flotante.

En general, una compuerta CMOS complementaria tiene un arreglo de N-SWITCHES (*pull-down*) para conectar la salida a "0" (Gnd) y un arreglo de P-SWITCHES (*pull-up*) para conectar la salida a "1" ( $V_{DD}$ ).

## b) Lógica Combinacional

Si dos N-SWITCHES se colocan en serie, entonces se tiene un switch resultante que está en ON si los dos switches están ON (entradas en "1"), lo que se traduce en una función AND. Iguales resultados se tiene para el caso de los P-SWITCHES, todas las entradas en "0" para que el switch resultante se cierre. Estas dos configuraciones se muestran en la Fig. 2.11a y b.

Si dos N-SWITCHES se colocan en paralelo el switch resultante está en ON si cualquiera de los switches está en ON (un "1" en cualquiera de las compuertas), lo que se traduce en una función OR. En el caso de tener los P-SWITCHES en paralelo, se obtiene también con el arreglo la función OR, con la diferencia que el switch resultante está ON si cualquiera de sus entradas tiene un "0". Estas dos configuraciones se presentan en la Fig. 2.11c y d.

Utilizando combinaciones de estas estructuras, pueden construirse compuertas CMOS, que realicen cualquier función combinacional.

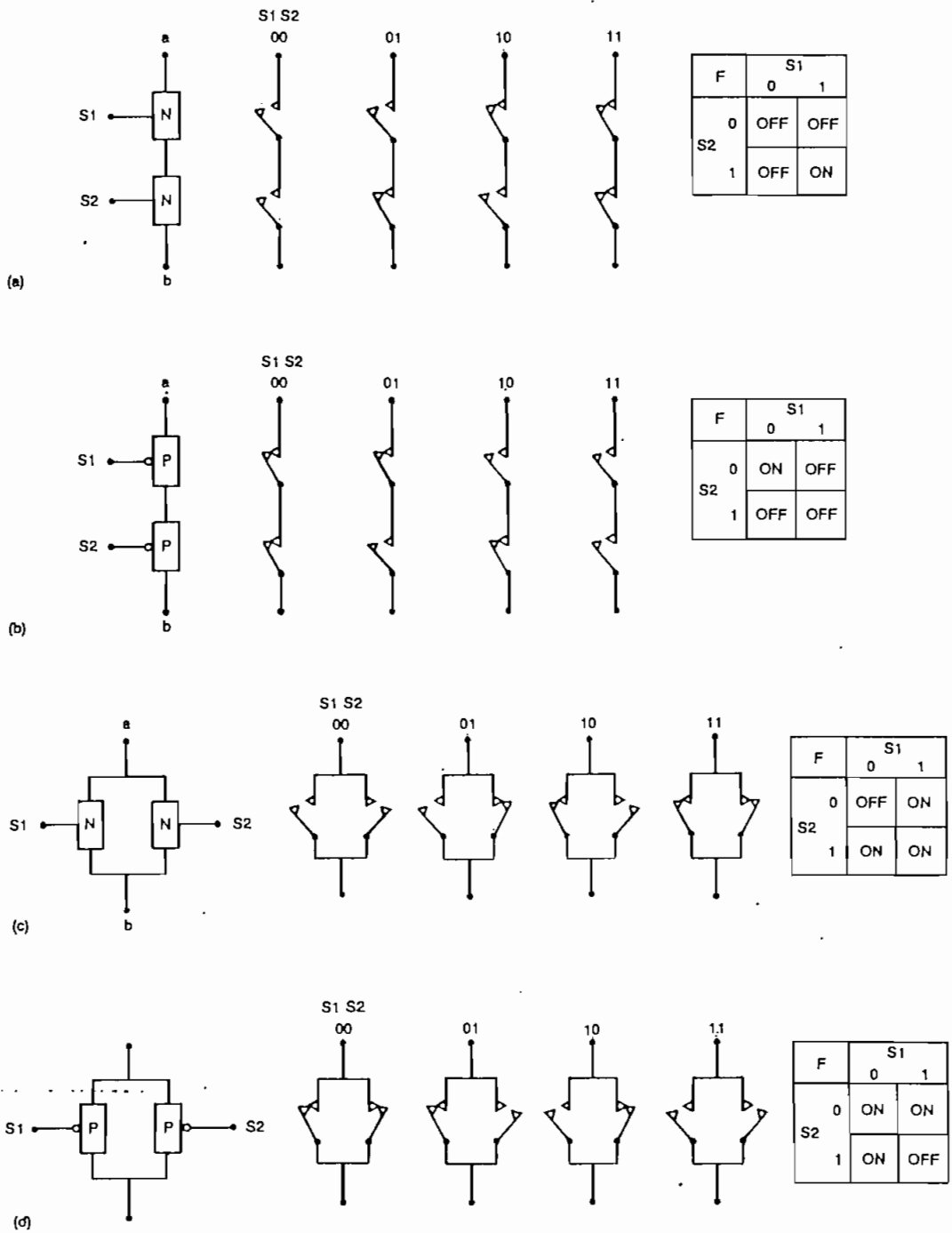


FIGURA 2.11 Combinaciones serie y paralelo de switches MOS.

En la explicación de los ejemplos que se exponen a continuación, se denomina término "1" al resultado de la agrupación en el mapa de Karnaugh de todos los "1s" y término "0" al resultado de la agrupación de los "0s".

c) La compuerta NAND

En la Fig. 2.12 se esquematiza el proceso mediante el cual se llega a determinar la estructura de una compuerta NAND, se utilizan para ello las construcciones de la Fig. 2.11a y d. La elección de cuales estructuras utilizar se deriva del Mapa de Karnaugh (Fig. 2.12a).

De los resultados obtenidos del mapa de Karnaugh se determina que:

- i) El término "0" (*pull-down*) determina una estructura AND ( $A \cdot B$ ), que se forma con un par de transistores "n" en serie.
- ii) El término "1" (*pull-up*) determina una estructura OR ( $\bar{A} + \bar{B}$ ), que se forma con un par de transistores "p" en paralelo.

Las señales complementadas necesarias en el término OR, no requieren de la inversión de las señales A y B, pues se obtienen directamente considerando la operación del dispositivo "p".



Si se analizan los dos términos,  $A \cdot B$  y  $\bar{A} + \bar{B}$ , se concluye que negando cualquiera de ellos se llega al otro, esto se debe a que al agrupar "1s" estamos considerando todas las combinaciones de las entradas que hacen que la salida tenga un voltaje alto ("1"), y cuando se agrupan "0s" se consideran los casos en que las combinaciones de la entrada determinan que la salida debe tener un voltaje bajo ("0").

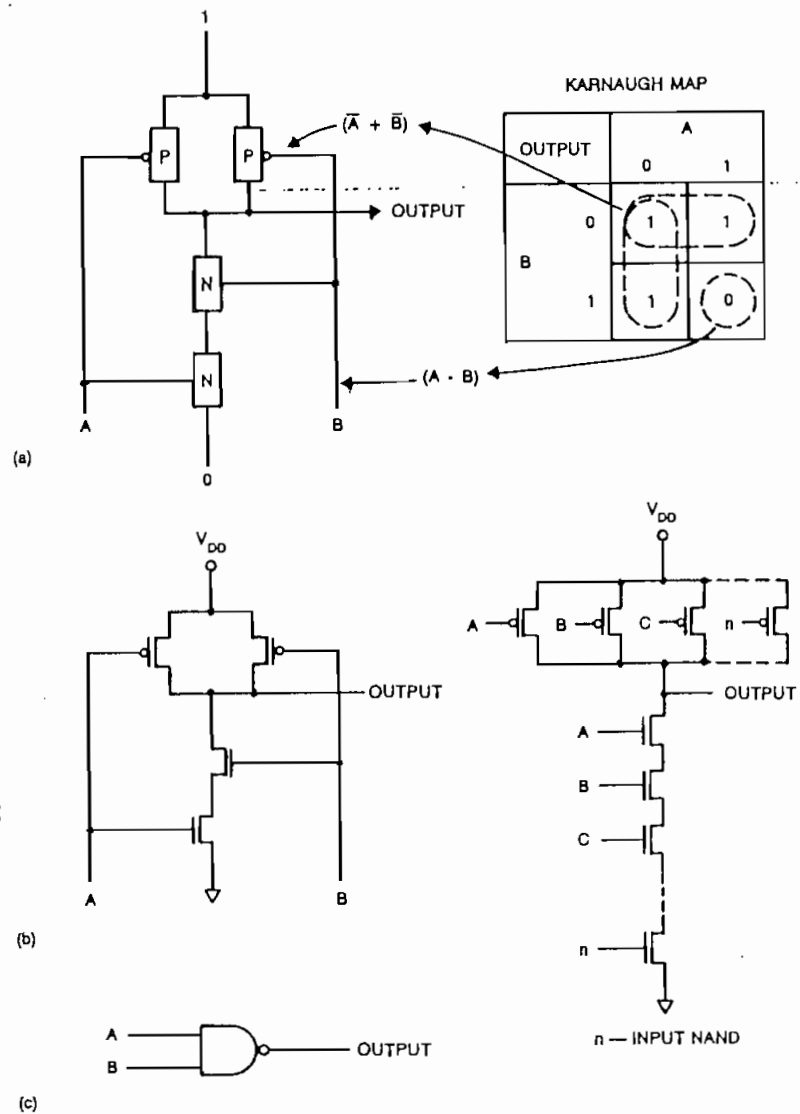


FIGURA 2.12 Compuerta NAND utilizando CMOS.

Observando los arreglos de transistores, se deduce que el arreglo "p" es dual del "n". Es decir, un arreglo serie y/o paralelo de transistores "n" del *pull-down*, será un arreglo paralelo y/o serie de transistores "p" en el *pull-up*, y viceversa.

De la Fig. 2.12 pueden obtenerse algunas características adicionales que son típicas de las compuertas lógicas CMOS:

- i) Para todas las entradas siempre hay un camino desde  $V_{DD}$  y Gnd ("1" y "0") a la salida, la disposición de los transistores nMOS y pMOS permite que los voltajes de las fuentes aparezcan a la salida sin degradación de los niveles de voltaje.
- ii) En CMOS nunca hay un camino de la fuente de un "1" ( $V_{DD}$ ) a la fuente del "0" (Gnd), para ninguna combinación de las entradas, como ocurre por ejemplo en los circuitos nMOS. Esta característica de CMOS es la razón para la baja disipación de potencia estática.

En la Fig. 2.12b puede apreciarse la generalización de la estructura de la compuerta NAND de dos entradas a una de "n" entradas; para ello, debe conectarse en el conjunto de transistores serie (*stack*) un transistor "n" y uno "p" en el conjunto de transistores en paralelo, por cada entrada adicional.

En la Tabla 2.1 se presenta el estado de cada uno de los switches y de la salida para las diferentes combinaciones de las dos entradas de la compuerta NAND de la Fig. 2.12a.

ENTRADAS		N-SWITCHES		P-SWITCHES		SALIDA
A	B	A	B	A	B	
0	0	OFF	OFF	ON	ON	1
0	1	OFF	ON	ON	OFF	1
1	0	ON	OFF	ON	OFF	1
1	1	ON	ON	OFF	OFF	0

TABLA 2.1 Tabla de verdad de la compuerta NAND.

d) La compuerta NOR

En la Fig. 2.13 se presenta una compuerta NOR de dos entradas. La estructura de la compuerta obtenida en base a las ecuaciones planteadas con ayuda del mapa de Karnaugh, se presenta utilizando switches en la Fig. 2.13a y transistores en la Fig. 2.13b.

i) El término "1" determina una función AND de los complementos de las señales de entrada, lo que se consigue con transistores "p" conectados en serie, los cuales no requieren las señales complementadas, debido a su funcionamiento.

ii) El término "0" determina la función  $(A+B)$ , que se estructura con transistores "n" en paralelo.

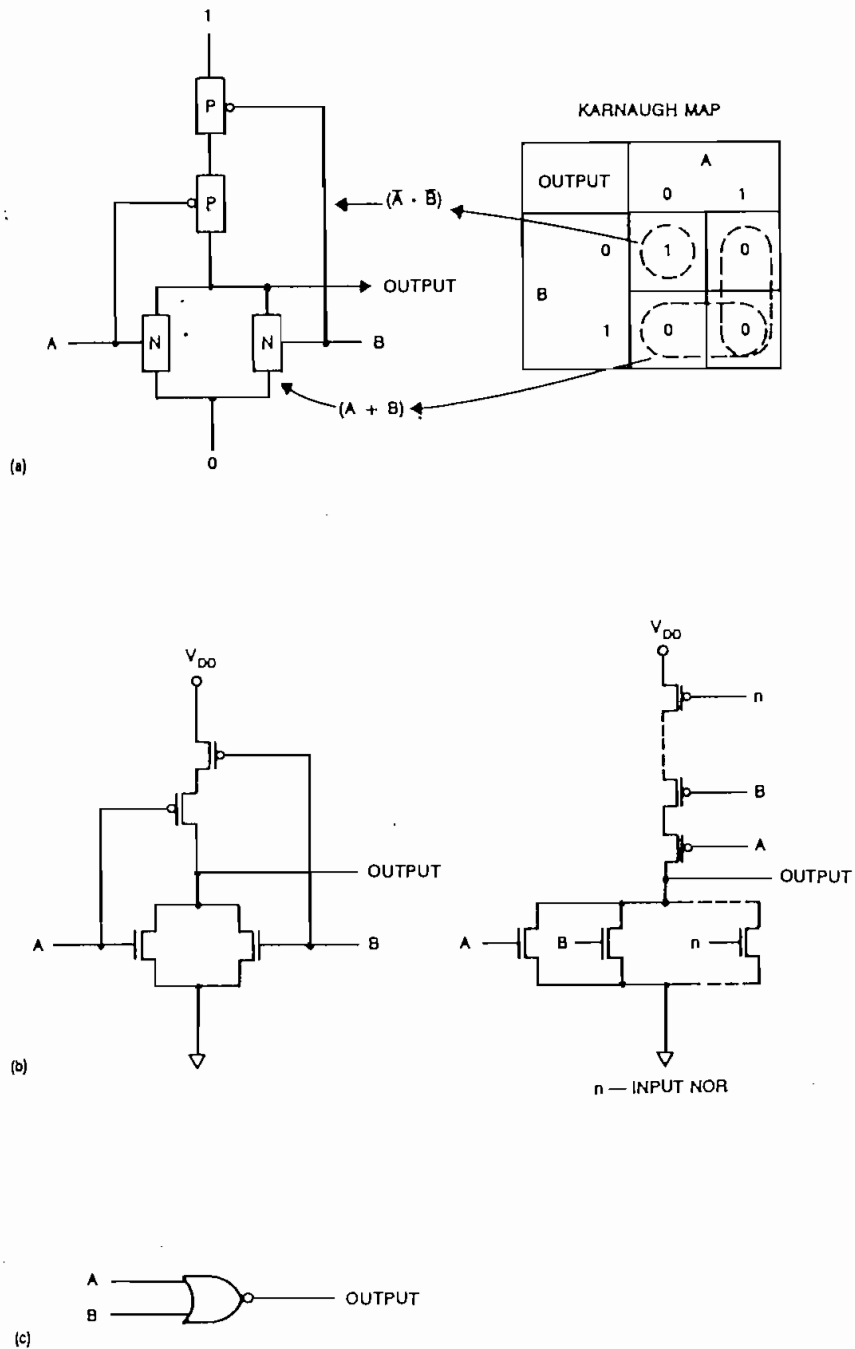


FIGURA 2.13 Compuerta NOR con tecnología CMOS.

La generalización de la estructura de la compuerta NOR para "n" entradas se presenta en la Fig. 2.13b. En este caso se agrega, por cada entrada, un transistor "p" en el conjunto de transistores en serie y uno "n" en el conjunto de transistores en paralelo.

Puede notarse que si se compara la disposición de los transistores de la compuerta NOR con la de la compuerta NAND, la configuración de transistores "p" y "n" de la compuerta NOR es la dual de la configuración obtenida para la compuerta NAND.

La Tabla 2.2 presenta el estado de cada uno de los switches y de la salida para las diferentes combinaciones de las dos entradas de la compuerta NOR.

ENTRADAS		N-SWITCHES		P-SWITCHES		SALIDA
A	B	A	B	A	B	
0	0	OFF	OFF	ON	ON	1
0	1	OFF	ON	ON	OFF	0
1	0	ON	OFF	OFF	ON	0
1	1	ON	ON	OFF	OFF	0

TABLA 2.2 Tabla de verdad de la compuerta NOR

e) Compuertas mixtas

Una compuerta mixta se forma por la combinación de estructuras de switches en serie y paralelo tanto en el *pull-up* como en el *pull-down*.

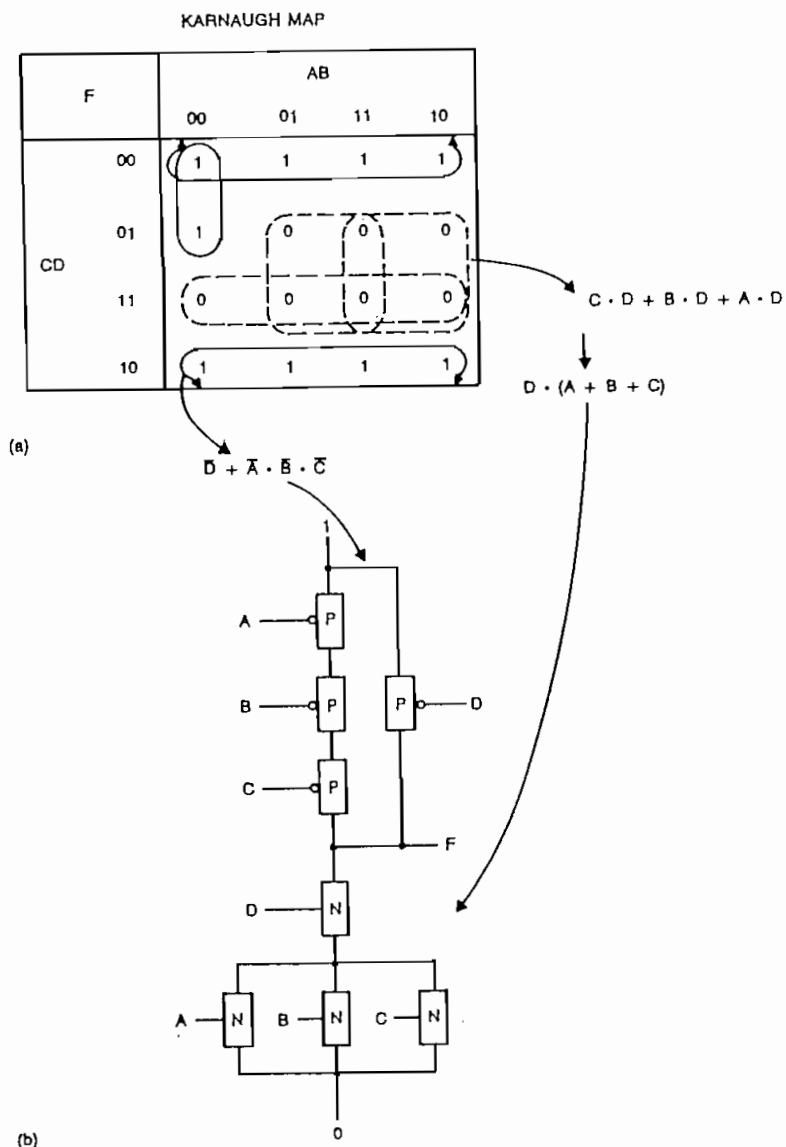


FIGURA 2.14 Construcción de la función  $F = \overline{((A+B+C) \cdot D)}$ .

En la Fig. 2.14 se presenta el proceso para obtener el arreglo de switches para la función (F), se incluye el mapa de Karnaugh y los términos "1" y "0" resultantes. La función a obtenerse es  $F = \overline{(A+B+C)}.D$ .

- i) El término "0" determina la función  $D.(A+B+C)$ . La subfunción  $(A+B+C)$ , se implanta con tres N-SWITCHES en paralelo. Esta estructura se coloca luego en serie con un N-SWITCH que tiene a su entrada la señal D, formando la función AND con el arreglo en paralelo.
  
- ii) El término "1" determina la función  $\bar{D}+(\bar{A}.\bar{B}.\bar{C})$ , que requiere de tres transistores "p" en serie para las entradas A, B y C, y que luego se debe colocar en paralelo con el F-SWITCH que tiene a su entrada la señal D. El arreglo final se presenta en la Fig. 2.14b.

El proceso seguido para la construcción de cualquier función con lógica CMOS, para cuando se utiliza el mapa de Karnaugh, puede resumirse de la siguiente manera:

Se debe analizar el mapa de Karnaugh para obtener las ecuaciones lógicas resultantes de la agrupación de "0s", que se resuelven con transistores "n", y ecuaciones resultantes de la agrupación de "1s" que se, resuelven con transistores "p". A partir de las ecuaciones se plantean las combinaciones serie y paralelo de transistores necesarias, para finalmente integrar los dos conjuntos, de transistores "n" y "p".

Sin embargo, puede ser que se disponga ya de la ecuación de una función, siendo más cómodo en este caso plantear el arreglo de los transistores "n" y "p" sin tener que desarrollar el Mapa de Karnaugh. En la Fig. 2.15 se presenta la derivación de la interconexión de los switches para otra función. La función a implantar es:  $F = ((A.B) + (C.D))$ . Los pasos a seguirse son:

i) Para el extremo "n" se toma la función F complementada, para estructurar la parte del *pull-down*, lo que proporciona la función  $\bar{F} = ((A.B) + (C.D))$ .

i.1 Los términos AND: A.B y C.D, se forman con transistores en serie (Fig. 2.15a).

i.2 La función OR de los dos términos obtenidos se realiza colocando los arreglos anteriores en paralelo. (Fig. 2.15b).

ii) Para el lado "p" se expande la función F, utilizando las Leyes de DeMorgan, lo que proporciona la siguiente expresión:  $\overline{(A + B)}. \overline{(C + D)}$ .

ii.1 La expresión última sugiere dos estructuras OR, éstas son: (A + B) y (C + D). Como ya se ha mencionado, no es necesario tener las señales invertidas para controlar los transistores "p". (Fig. 2.15c).



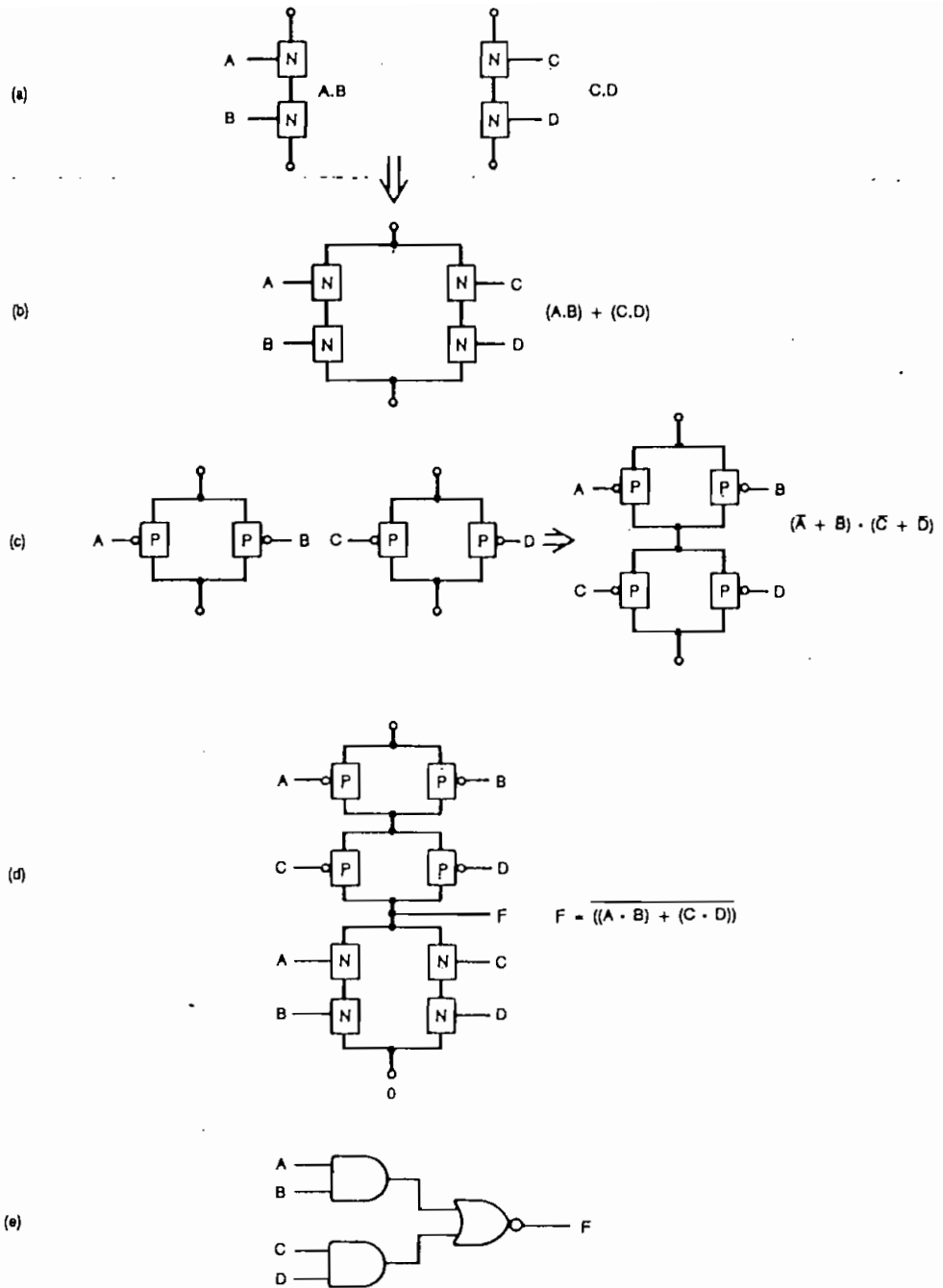


FIGURA 2.15 Construcción de la función  $F = \overline{((A \cdot B) + (C \cdot D))}$ .

ii.2 Los dos arreglos en paralelo obtenidos, se colocan en serie para estructurar la función del lado "p". La secuencia de los pasos necesarios para cumplir lo expuesto en 2.1 y 2.2 se presentan en la Fig. 2.15c.

iii) El último paso requiere conectar un extremo del arreglo "p" a  $V_{DD}$  y el otro a la salida y conectar un extremo del arreglo "n" a Gnd y el otro a la salida. Esto se presenta en la Fig. 2.15d. En la Fig. 2.15e se muestra la representación con compuertas de la función F.

Se presenta a continuación la implantación de una tercera función F que permite describir un procedimiento que evita realizar un proceso demasiado largo para el caso en el que se dispone de la ecuación de una función.

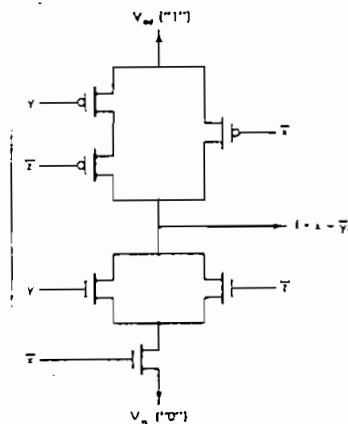


FIGURA 2.16 Construcción de la función  $F = X + \bar{Y}Z$ .

- i) Iniciar el proceso negando la función a implantar y aplicar las leyes de DeMorgan.

La función negada, implantada con transistores "n" proporciona los caminos necesarios para todas las combinaciones de las entradas (X,Y,Z) para los cuales  $F=0$ .

$$\bar{F} = \bar{X} \cdot (Y + \bar{Z})$$

- ii) Obtener la red serie-paralelo de transistores "n" para la función complementada.

Para este caso la red "n" consiste de dos transistores en paralelo, en serie con un tercero, manejados por las señales Y y  $\bar{X}$ ,  $\bar{Z}$  (Fig. 2.16).

- iii) Obtener la red dual de la estructura, utilizando unicamente transistores "p", cuyas transistores tienen las mismas señales de entradas que sus elementos duales correspondientes.

Para la función F, la estructura dual se obtiene colocando los transistores (Y,Z), que se encuentran en paralelo en la red "n", en serie, y estos dos transistores en paralelo con el (X), que en la red "n" se encuentra en serie con el arreglo paralelo de (Y,Z). La estructura completa para la función F se presenta en la Fig. 2.16.

Para este caso particular es necesario que las señales (X,Z) sean negadas como paso previo a ingresar a la compuerta.

#### f) Multiplexores

Los switches complementarios pueden utilizarse para realizar la función de un multiplexor. La Fig. 2.17 presenta el caso para un multiplexor de dos entradas.

Dado que los switches deben permitir el paso de "0s" y "1s" desde la entrada del multiplexor hasta la salida de manera íntegra, se utilizan switches complementarios.

La Tabla 2.3 permite resumir la operación de la estructura con switches complementarios de la Fig. 2.17a.

ENTRADAS		SELECCION		SALIDA
A	B	S	$\sim S$	
X	0	0	1	0(B)
X	1	0	1	1(B)
0	X	1	0	0(A)
1	X	1	0	1(A)

TABLA 2.3 Tabla de verdad del multiplexor de dos entradas

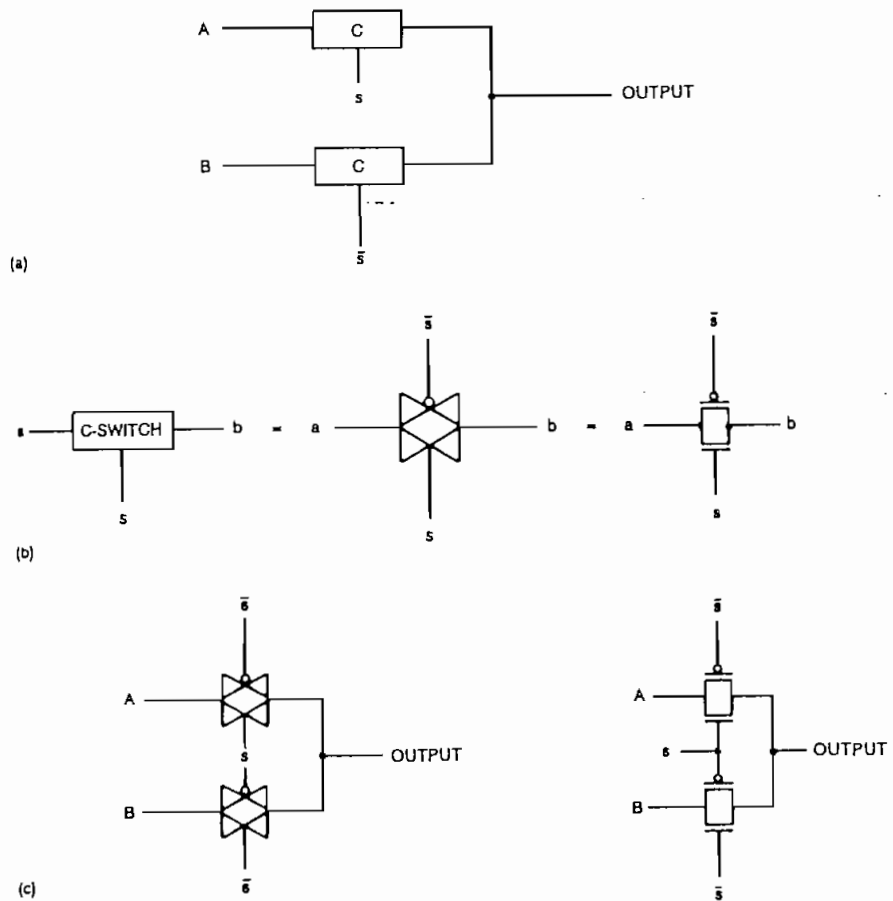


FIGURA 2.17 Multiplexor de dos entradas.

Para una clara comprensión de la estructura, se debe tener en cuenta también como está estructurado cada uno de estos switches y las señales de control que los comandan.

El análisis que se hace es simplificado, considerando a las combinaciones de los transistores (C-SWITCH) como simples switches, que están abiertos o cerrados.

Cuando la señal  $S=0$ , el switch de la parte superior (A), está abierto, lo que no permite el paso de ninguna señal que se encuentre a la entrada. La señal  $S$  complementada, por el contrario, tiene un valor de 1, por lo que el switch de la parte inferior (B) está cerrado, y permite el paso hacia la salida del valor que está presente a la entrada (ya sea un 0 o un 1). Lo contrario ocurre si  $S=1$ .

El switch complementario es también llamado **compuerta de transmisión** o **compuerta de paso** (*transmission gate* ó *pass gate*). Un par de representaciones circuitales para la compuerta de transmisión muy comunes, aunque no las únicas, se presentan en la Fig. 2.17b. En la Fig. 2.17c se muestra la estructura del multiplexor utilizando las representaciones indicadas para la compuerta de transmisión.

#### g) Elementos de memoria

Se ha presentado la estructura de un conjunto de estructuras CMOS suficiente para poder implantar un elemento de memoria. En la Fig. 2.18a se presenta un simple flip-flop, tipo "D" utilizando un multiplexor de dos entradas y dos inversores. Cuando  $LD = 1$ ,  $\bar{Q}$  tiene la información  $\bar{D}$  (Fig. 2.18b) y cuando  $LD$  cambia a 0, se crea un camino de realimentación a través del par de inversores (Fig. 2.18c), lo que permite que el estado actual de  $Q$  se conserve. Además cuando  $LD = 0$  se ignora la entrada  $D$ .

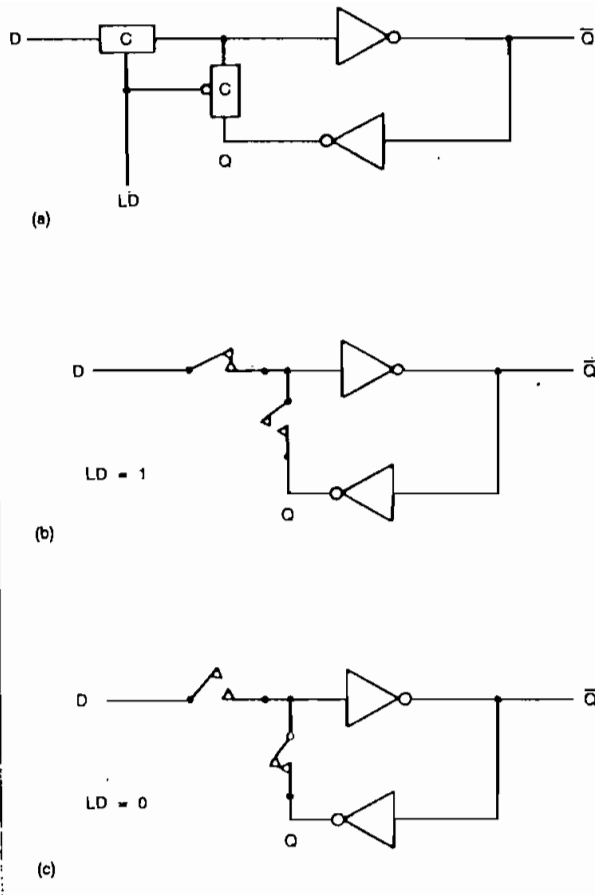


FIGURA 2.18 Estructura de un flip-flop básico.

Todos los ejemplos presentados hasta el momento de la estructura de elementos básicos de diseño lógico: switches, compuertas y circuitos combinatoriales, se introdujeron asumiendo que los transistores MOS actúan como simples switches. Como se analiza en este Capítulo, y posteriormente en otros, existen ciertas limitaciones que presentan los transistores MOS, que apartan la concepción idealizada de éstos como simples switches.

Las estructuras presentadas corresponden al tipo denominado "lógica totalmente complementaria" (*fully complementary logic*) o estática, que se caracteriza por tener siempre un transistor "p" por cada transistor "n". Sin embargo, existen configuraciones lógicas CMOS alternativas a la mencionada anteriormente.

Estas configuraciones alternativas son útiles en situaciones en las que: el área utilizada por la compuerta CMOS completamente estática es demasiado grande respecto a la que en realidad se requiere, la velocidad de operación es demasiado baja, o simplemente la compuerta no puede implantarse utilizando solo lógica totalmente complementaria, como ocurre en el caso de grandes PLAs.

Las configuraciones alternativas que se han planteado permiten disponer de compuertas más pequeñas y más rápidas, con desventajas como una posible disminución de la estabilidad e incremento de la complejidad de su operación.

Existe un buen número de estructuras lógicas alternativas CMOS que pueden ser utilizadas. Para enumerar las más difundidas:

1. Lógica Pseudo-nMOS.
2. Lógica CMOS Dinámica.
3. Lógica CMOS temporizada (C<sup>2</sup>MOS).
4. Lógica de "dominó" CMOS (disfraz).



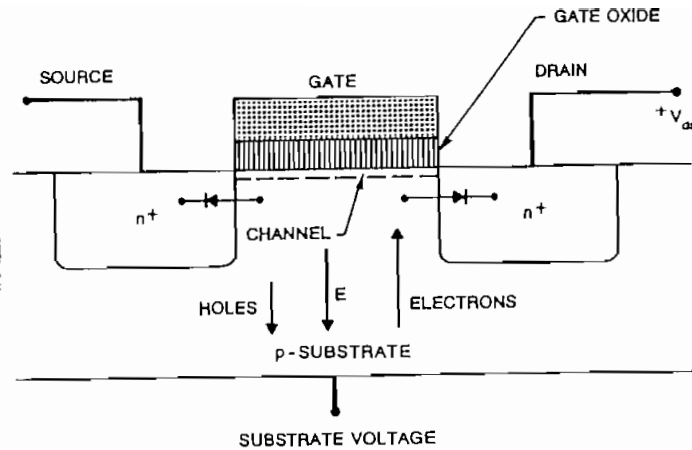


FIGURA 2.19 Estructura física de un transistor nMOS de enriquecimiento.

Son parte de este transistor:

- a) Un sustrato de silicio tipo "p" moderadamente dopado, que sirve de material de partida y sobre el cual se construyen las estructuras necesarias para formar el transistor, utilizando diferentes procesos tecnológicos.
- b) Dos regiones altamente dopadas  $n^+$ , el drenaje y la fuente, que se difunden en el sustrato. Se puede observar también la simetría bilateral de la estructura presentada.
- c) La región del "canal", la delgada capa de sustrato tipo "p" entre drenaje y fuente.

Sin voltaje de compuerta, drenaje y fuente están eléctricamente aisladas, con los voltajes adecuados se crea

por inducción un camino entre los dos terminales, que permitirá la circulación de corriente entre ellos. A este camino inducido se le denomina "canal". El canal se origina en la fuente y se extiende hacia el drenaje, ocurriendo en ocasiones, como se explica más adelante, que no alcanza la zona del drenaje. El canal tiene también una profundidad que depende del voltaje de compuerta, teniendo un límite máximo. Dado que el canal se forma entre drenaje y fuente se denomina a esta región del sustrato como "región del canal".

- d) Una delgada capa de aislante de dióxido de silicio  $\text{SiO}_2$ , denominada óxido de la compuerta que cubre la región del canal.
- e) El electrodo denominado compuerta, construido sobre el  $\text{SiO}_2$ . En la actualidad, la compuerta se fabrica de silicio policristalino altamente dopado (*polysilicon*).

Debido a la estructura del transistor, que incluye materiales tipo "p" y "n", se forman diferentes junturas p-n (diodos), dependiendo inclusive de las condiciones de operación del dispositivo. En los límites de contacto de estas junturas se produce la denominada barrera de potencial, que se explica de manera resumida, recordando los procesos físicos de difusión y desplazamiento.

La difusión consiste en el movimiento de los portadores desde un área de alta concentración a un área de baja concentración (gradiente de concentración). Debido a la difusión, los huecos del material tipo "p" se mueven a través de la juntura al material tipo "n", dejando atrás iones negativos, y los electrones del material tipo "n" se mueven, a través de la juntura, al material tipo "p", dejando atrás iones positivos.

La región a cada lado de la juntura desde la cual los electrones y huecos se han movido se denomina región de agotamiento o de transición. Los iones dejados atrás producen un campo eléctrico que se opone al proceso de difusión que lo causó. Este campo, denominado barrera de potencial o potencial de contacto causa un desplazamiento fuera de esta zona de huecos y electrones (corriente de desplazamiento). Un estado de equilibrio se alcanza si la corriente de difusión es igual y opuesta a la de desplazamiento, lo que resulta en una corriente neta de cero.

a) Regiones de operación del transistor .

i) Región de corte.

El transistor opera en la región de corte cuando en su compuerta se tiene un voltaje menor (incluso negativo) al voltaje umbral ( $V_t$ ), lo que impide la formación del canal de conducción.

En un inicio, con un voltaje aplicado a la compuerta de cero ( $V_{gs}=0$ ), las junturas formadas entre el sustrato y las regiones de difusión del drenaje y fuente (Fig. 2.19), aíslan efectivamente al drenaje y fuente por no estar directamente polarizadas .

Un voltaje negativo produciría un incremento de la barrera de potencial de las junturas p-n formadas entre el sustrato y drenaje y fuente, impidiendo aún más la formación del canal. Este efecto se representa en la Fig. 2.20a. con la acumulación de huecos bajo la compuerta.

Al aumentar el voltaje de compuerta ligeramente sobre el valor de cero ( $V_{gs}>0$ ) se produce un campo eléctrico  $E$  a través del sustrato que atrae a los electrones hacia la compuerta y repele a los huecos, dejando una zona de agotamiento de portadores mayoritarios móviles en el sustrato.

La carga inducida en estas condiciones se debe a la repulsión de los huecos de la superficie, lo que deja atrás iones dopantes y unos pocos portadores minoritarios (electrones en la superficie), produciendo una región de agotamiento. (Fig. 2.20b).

## ii) Región lineal

Si el voltaje  $V_{gs}$  se incrementa aún más, la profundidad de la región de agotamiento se incrementa y el potencial

electrostático en la superficie de la compuerta se incrementa lo que produce una disminución de la barrera de potencial de junturas p-n que aíslan drenaje y fuente.

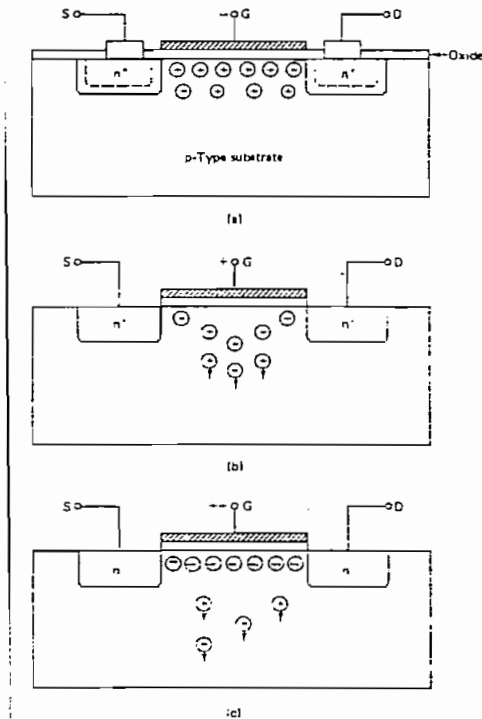


FIGURA 2.20 Condiciones de carga del canal de acuerdo a la magnitud de  $V_{GS}$ .

La disminución de la barrera de potencial permite "halar" electrones de la fuente y drenaje, lo que produce un alto incremento en la concentración de electrones en una delgada capa cercana a la superficie del sustrato, haciendo que esta capa cambie de "p" a "n", formando el canal. Este efecto se presenta en la Fig. 2.20c.

En conclusión, para el transistor nMOS, el canal no es sino una capa de sustrato tipo "p" que ha sido invertida. La Fig. 2.21a presenta la distribución inicial de los huecos en la capa de  $\text{SiO}_2$  al aplicar un pequeño voltaje positivo a la compuerta ( $V_t$ ). En estas condiciones se forma ya una pequeña capa invertida. Voltajes menores a  $V_t$ , no logran la formación del canal. Por lo tanto,  $V_t$  es el valor mínimo para formar el canal, es el voltaje a partir del cual el dispositivo MOS empieza a conducir o a "encenderse" y se denomina voltaje umbral. Aplicando un voltaje  $V_{gs} > V_t$ , se consigue una mayor orientación de los iones del aislante, y una mayor profundidad de la capa invertida, como se indica en la Fig. 2.21b.

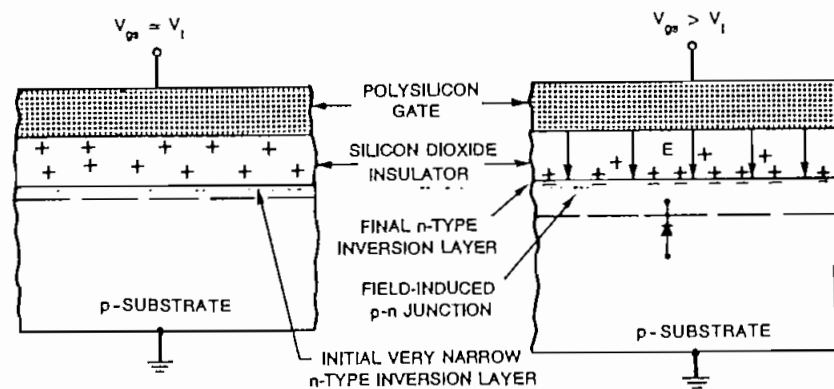


FIGURA 2.21 Creación de la capa de inversión en un transistor nMOS.

Una vez formado el canal, cualquier incremento de  $V_{gs}$  se traduce en un incremento de la concentración de electrones. Esta condición se denomina de "fuerte inversión", pues la concentración de los electrones en este punto excede la de

los iones dopantes. Bajo estas condiciones los electrones pueden ser desplazados por una diferencia de tensión aplicada entre drenaje-fuente, produciendo  $I_{d_s}$ .

El canal forma una nueva juntura con el sustrato. La diferencia entre una juntura p-n, presente en un transistor bipolar o en un diodo (o entre la fuente o drenaje y el sustrato) y la juntura canal-sustrato es que en la primera, el material tipo "n" es construido mediante procesos físicos, los electrones se introducen en el semiconductor, mediante impurezas donadoras; por el contrario, en una juntura canal-sustrato, la capa tipo "n" del canal es solamente inducida por un campo eléctrico (E), por lo que, esta juntura, en lugar de ser física, es una juntura inducida por campo.

De todo lo expuesto se puede decir que, electricamente, un transistor MOS actúa como un switch controlado por voltaje. Este switch inicia su conducción cuando el voltaje compuerta-fuente ( $V_{g_s}$ ) es igual al voltaje umbral ( $V_t$ ). Cuando se aplica un voltaje  $V_{d_s}$  entre la fuente y el drenaje, para  $V_{g_s} = V_t$ , las componentes horizontal y vertical del campo eléctrico, debidas al voltaje fuente-drenaje y voltaje compuerta a sustrato, interactúan causando la circulación de corriente a lo largo del canal. La componente horizontal del campo eléctrico asociado con el voltaje drenaje a fuente ( $V_{d_s} > 0$ ), es responsable de barrer los electrones del canal hacia el drenaje. La componente transversal del campo eléctrico en cambio se encarga de mantener el canal formado.

El óxido de la compuerta es muy delgado en comparación al espaciamiento fuente-drenaje. En un proceso típico 1988, el espaciamiento entre fuente-drenaje es 1.5 micras y el espesor del óxido es 400  $\text{\AA}$  ( $100 \text{\AA} = 0.01 \text{ micras}$ ). Por esta razón, el potencial en el canal, justo bajo la compuerta, es dominado por el voltaje de la compuerta.

En la Fig. 2.22a se aprecia la forma que tiene el canal cuando  $V_{gs} \geq V_t$  y  $V_{ds} = 0$ .

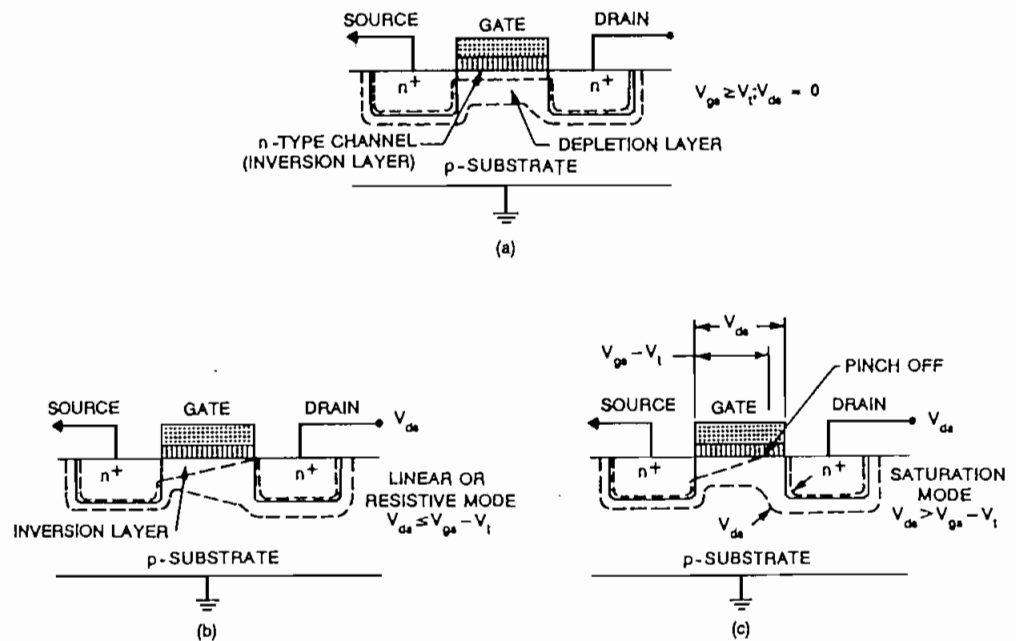


FIGURA 2.22 Transistor nMOS bajo la influencia de los diferentes voltajes terminales.

Puede observarse también en esta figura, la región de agotamiento formada en los límites de contacto de las junturas p-n formadas, tanto entre drenaje y fuente con el sustrato, así como en la junta canal-sustrato.



Si el voltaje drenaje-fuente se incrementa, la caída resistiva a lo largo del canal comienza a variar la forma característica del canal (Fig. 2.22b).

Para una comprensión clara del efecto que tiene lugar al incrementar  $V_{ds}$ , es necesario identificar los diferentes voltajes involucrados en la operación del transistor. En la Fig. 2.23 se presenta la estructura del transistor "n" en el que se señalan los voltajes  $V(z)$  y  $V_{chq}$ .

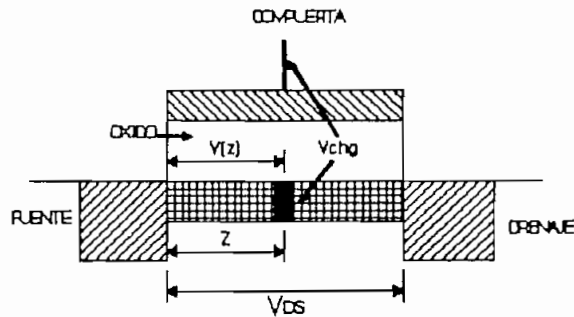


FIGURA 2.23 Sección transversal de un MOSFET canal "n".

El voltaje  $V(z)$  es la tensión del canal (en cada punto del mismo), expresado en función de  $z$ , que varía de 0 a  $L$ , siendo  $L$  la longitud total del canal.

La diferencia de tensión entre compuerta y canal se nota como  $V_{chq}$ , y es aquella aplicada al condensador de placas paralelas formado por el electrodo de la compuerta y el canal, teniendo al  $\text{SiO}_2$  como aislante. Este voltaje compuerta-canal está dado por:

$$V_{cbg} = V_{gs} - V(z)$$

[Ec.2.1]

El voltaje en el canal es 0 en el extremo de la fuente ( $V(0)=0$ ) y es igual al voltaje drenaje-fuente en el extremo del drenaje ( $V(L)=V_{ds}$ ). Por lo tanto, el valor del voltaje compuerta-canal está dado en el extremo de la fuente por: ( $V_{cbg} = V_{gs}$ ) y decrece linealmente hasta el extremo del drenaje a: ( $V_{cbg} = V_{gs} - V_{ds}$ ).

El voltaje efectivo que mantiene la formación del canal ( $V_{eff}$ ) está dado, en cada punto por:

$$V_{eff} = V_{cbg} - V_t$$

[Ec.2.2]

El voltaje  $V_{eff}$  debe ser diferente de cero en todos los puntos debajo de la compuerta para que exista el canal. Utilizando las relaciones presentadas, se encuentra que en el extremo de la fuente, el voltaje efectivo de la compuerta tiene un valor de: ( $V_{eff} = V_{gs} - V_t$ ) y cae en el extremo del drenaje al valor: ( $V_{eff} = V_{gs} - V_t - V_{ds}$ ).

Por lo tanto, en el extremo de la fuente, prácticamente todo el voltaje de la compuerta es efectivo para formar el canal; sin embargo, en la región próxima al drenaje, únicamente un valor menor a la diferencia entre los voltajes de la compuerta y el drenaje es efectivo.

Quando se cumple que  $(V_{gs} - V_t \geq V_{ds})$ , el canal se hace más profundo al aumentar  $V_{gs}$  y presenta la forma de la Fig. 2.22b. A esta región se le denomina como **lineal, resistiva o no-saturada**. La característica de esta región es la dependencia de la corriente del canal  $I_{ds}$  con los voltajes de compuerta y drenaje; además, se puede representar al dispositivo por una resistencia controlada por voltaje.

### iii) Región de saturación

Si  $V_{ds} \geq (V_{gs} - V_t)$ , entonces el voltaje  $V_{dr}$  en el extremo de la compuerta cercana al drenaje llega a ser cero o menor que cero y el canal desaparece. El dispositivo entra entonces en la región de saturación y el voltaje de drenaje en este punto se denomina voltaje de saturación o **estrangulamiento (pinch-off)**, cuyo valor es:  $(V_{dsat} = V_{gs} - V_t)$ . Esto se ilustra en la Fig. 2.22c.

La no formación del canal en el extremo del drenaje puede explicarse de la siguiente forma: la magnitud del voltaje  $V_{ds}$  de igual polaridad que el voltaje de la compuerta, se opone al proceso de inducción de electrones en el extremo del drenaje, y esta región ya no será invertida, simplemente es agotada. Posteriores incrementos del voltaje  $V_{ds}$  producen una mayor disminución de la longitud del canal.

El voltaje a través del canal estrangulado tiende a permanecer fijo en  $(V_{gs} - V_t)$ . Esta condición es el estado de

saturación en el cual la corriente del canal se controla con  $V_{gs}$  y es casi independiente de  $V_{ds}$ . En esta región, la corriente de drenaje se debe a un mecanismo de desplazamiento de electrones bajo la influencia del voltaje de drenaje positivo. Los electrones, que fluyen hacia la región de agotamiento próxima al drenaje, son acelerados a través de esta región por el alto valor del voltaje de drenaje.

### 2.3.2 Transistores de enriquecimiento y empobrecimiento

Los transistores de efecto de campo MOS pueden ser de enriquecimiento (o robustecimiento), tipo "n" y tipo "p", y de empobrecimiento (o agotamiento), también tipo "n" y "p". A estos transistores MOS también se les conoce con los nombres de transistores nMOS o transistores canal "n" y transistores pMOS o transistores canal "p", tanto de enriquecimiento como de empobrecimiento.

El funcionamiento de un transistor MOS, de cualquier tipo, se resume a la circulación de corriente por el canal de conducción. Esta corriente es modulada por el nivel de voltaje de la compuerta. Ya en operación se requiere también aplicar un voltaje drenaje-fuente de polaridad adecuada. Para un transistor nMOS un voltaje positivo de la compuerta induce un incremento o enriquecimiento (*enhancement*) del número de electrones en el canal. Los portadores mayoritarios presentes en el canal para este transistor son los electrones

inducidos. Para voltajes en la compuerta menores al voltaje umbral ( $V_t$ ), el canal no se forma.

Un transistor pMOS de enriquecimiento tiene un comportamiento similar al del nMOS, excepto que los portadores mayoritarios son los huecos y los voltajes de la compuerta son negativos, respecto al sustrato.

Se pueden fabricar transistores canal "n" que conducen para valores de voltaje de la compuerta igual al de la fuente y aún para valores inferiores a este voltaje. A estos transistores que conducen con una polarización de la compuerta de cero, se les denomina de **empobrecimiento** (*depletion*).

La diferencia establecida entre transistores de enriquecimiento y empobrecimiento, se debe a que en los últimos el canal es construido como una delgada capa durante la fabricación del dispositivo, y no debe su formación a la inducción de carga durante la operación, como ocurre en los dispositivos de enriquecimiento. El efecto neto del canal fabricado es cambiar el valor del voltaje umbral a un valor negativo.

Cada tipo de transistor nMOS tiene su dual en un transistor pMOS y viceversa, esto permite que entendiendo el funcionamiento de unos, para entender el correcto funcionamiento de los otros solamente se deben utilizar voltajes de polaridad opuesta.

En la tecnología CMOS, la misma pastilla (CI) incluye transistores nMOS y pMOS, actualmente la mayoría y de procesos de fabricación utilizan transistores de enriquecimiento. En la tecnología nMOS, se utilizan transistores nMOS de enriquecimiento para el *pull-down*, y transistores nMOS de empobrecimiento para el *pull-up*. Los transistores pMOS de empobrecimiento no han sido fabricados en CIs prácticos, aunque teóricamente su construcción es posible .

En el transistor nMOS, al establecerse la conducción, los portadores mayoritarios, los electrones, fluyen desde la fuente al drenaje. Si se asume la convención para la **representación** de la corriente, contraria al flujo de electrones, el sentido asignado es de drenaje a fuente ( $I_{d\downarrow}$ ); además, el sentido de circulación de la corriente determina la polaridad de los voltajes dentro de un circuito. Las curvas características que describen el comportamiento en las zonas de operación de los transistores MOS se realizan utilizando la corriente drenaje-fuente ( $I_{d\downarrow}$ ).

En el caso del transistor pMOS, los portadores mayoritarios, los huecos, fluyen de la fuente al drenaje, como en el caso anterior. En este caso la **representación** de la corriente coincide con el flujo de los portadores mayoritarios (fuente-drenaje). Sin embargo, para trazar las curvas características del transistor, se utiliza la corriente drenaje-fuente ( $I_{d\downarrow}$ ), por lo que la corriente tiene un signo negativo asociado a ella y determina un voltaje drenaje-fuente también negativo.

Puede graficarse la corriente fuente a drenaje ( $I_{ds}$ ) en función de la diferencia de voltaje compuerta-fuente ( $V_{gs}$ ). En la Fig. 2.24 se presentan 4 gráficos de ( $I_{ds}$ ) en función de ( $V_{gs}$ ) para un valor fijo del voltaje drenaje-fuente  $V_{ds}$ , para cada tipo de transistor "n" y "p" posibles.

En los gráficos se observa que el primer parámetro que caracteriza a un transistor MOS, es el voltaje umbral.

Como se indicó anteriormente, para tecnología CMOS se utilizan los transistores de enriquecimiento, cuyas curvas características y modos de operación se estudian a continuación.

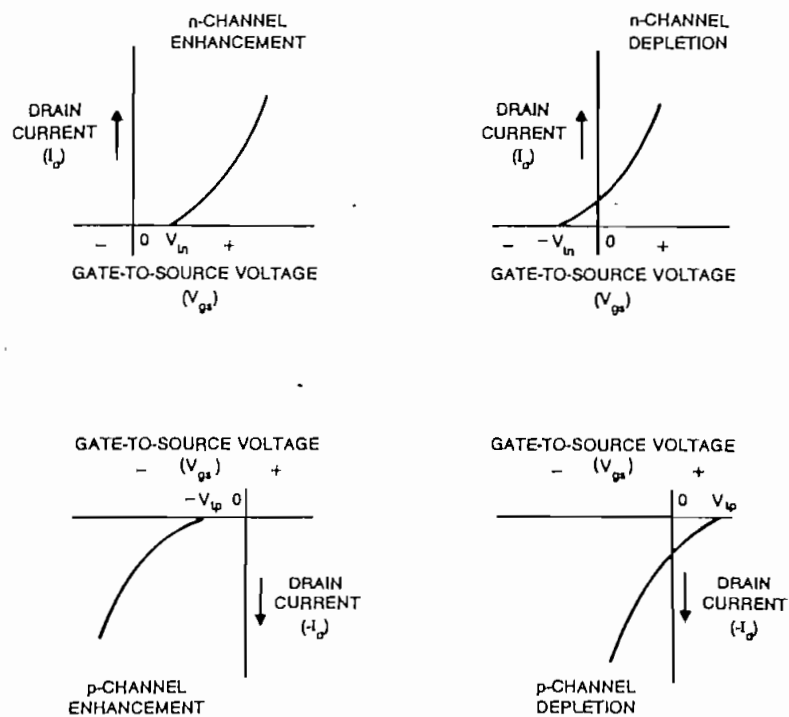


FIGURA 2.24 Características de conducción para transistores de enriquecimiento y empobrecimiento.

### 2.3.3 Estudio analítico del transistor nMOS

Las ecuaciones deducidas a continuación representan el comportamiento de los dispositivos de una manera aproximada, sin considerar efectos de segundo orden que pueden influir en las expresiones encontradas; sin embargo, cuando es necesario, se mencionan tales efectos y se presentan expresiones que los consideran.

#### a) Región de corte

La región de corte se denomina también región sub-umbral (subthreshold). En esta región la corriente  $I_{ds}$  crece exponencialmente con  $V_{ds}$  y  $V_{gs}$ , principalmente debido a efectos de difusión. Aunque el valor de la corriente es muy pequeño, aproximadamente cero, este valor finito puede llegar a influir en el comportamiento de los circuitos.

#### b) Región lineal

Para deducir las ecuaciones que determinan la característica fundamental corriente-voltaje de un MOSFET, se debe asumir que la capa de inversión está presente en el canal, que la fuente está conectada al sustrato, y que estamos operando el dispositivo en la región lineal, con corriente fluyendo entre drenaje y fuente.



Para la deducción se parte de dos relaciones fundamentales:

- i) La primera que establece una relación entre la densidad de corriente ( $J$ ) y el campo eléctrico ( $E$ ) establecido por el voltaje de la compuerta, a través de la conductividad ( $\sigma$ ).

$$J = \sigma E$$

[Ec. 2.3]

$J$  = es la corriente ( $I$ ) por unidad de área ( $A$ ) del medio conductor. Esta área es la transversal al desplazamiento de los portadores:  $J = I/A$  [ $A/m^2$ ]

$E$  = Campo eléctrico [ $V/m$ ].

$\sigma$  = Conductividad [ $1/\Omega.m$ ].

- ii) La conductividad ( $\sigma$ ) para un material semiconductor está determinada por la relación:

$$\sigma = (n\mu_n + p\mu_p) q$$

[Ec. 2.4]

En donde:

$n$  = magnitud de la concentración de electrones libres [ $electrones/m^3$ ].

De manera general, el valor de la concentración es el número de portadores (N) contenidos en una unidad de volumen. Para un medio conductor de área transversal A y longitud L, la concentración está dada por:  $N/(LA)$  [portadores/ $m^3$ ]. Para el caso que N representa el número de electrones, se tiene:  $n = N/(LA)$  [electrones/ $m^3$ ].

$p =$  Concentración de huecos. Está dada por la relación:  
 $p = N/(LA)$  [huecos/ $m^3$ ].

$q =$  Carga del electrón o del hueco [coulombs].

$\mu_n =$  movilidad de los electrones [ $m^2/Vs$ ].

$\mu_p =$  movilidad de los huecos [ $m^2/Vs$ ].

La movilidad se define como:

$$\mu = \frac{\text{velocidad-promedio-de-los-portadores-por-desplazamiento}}{\text{CampoEléctrico}}$$

[Ec. 2.5]

Para el caso del transistor nMOS se considera unicamente la concentración de los electrones que están presentes en la capa de inversión, pues la región del canal en la que se produce esta capa de inversión está agotada de huecos.

Con esta consideración la conductividad se expresa como:

$$\sigma = n\mu_n q$$

[Ec. 2.6]

$\mu_n$  es por lo tanto la movilidad efectiva de los electrones en el canal, es decir la facilidad con la cual los portadores se desplazan en el material del sustrato.

Reemplazando las expresiones equivalentes de la densidad de corriente  $J=I/A$ , de la conductividad  $\sigma=nq\mu$ , y de la concentración  $n=N/LA$ , se tiene:

$$\frac{I}{A} = \left(\frac{N}{LA}\right) q\mu E$$

[Ec. 2.7]

El valor  $Nq$  representa la carga total y agrupando  $Nq/L$ , se tiene la carga por unidad de longitud, que se representa con la letra  $Q$ . Para el caso de los transistores MOS,  $L$  representa la longitud del canal. Reemplazando el valor de  $Q$  en la ecuación anterior, y simplificando el área  $A$ , se tiene:

$$I = Q\mu E$$

[Ec. 2.8]

En la Fig. 2.25 se presenta la estructura de un transistor nMOS, con las dimensiones físicas y las variables necesarios para deducir la ecuación de la corriente del dispositi-

vo. La corriente  $I_{ds}$  a lo largo del canal tiene un valor constante, y su magnitud depende de los voltajes de compuerta y drenaje aplicados, como se demuestra a continuación:

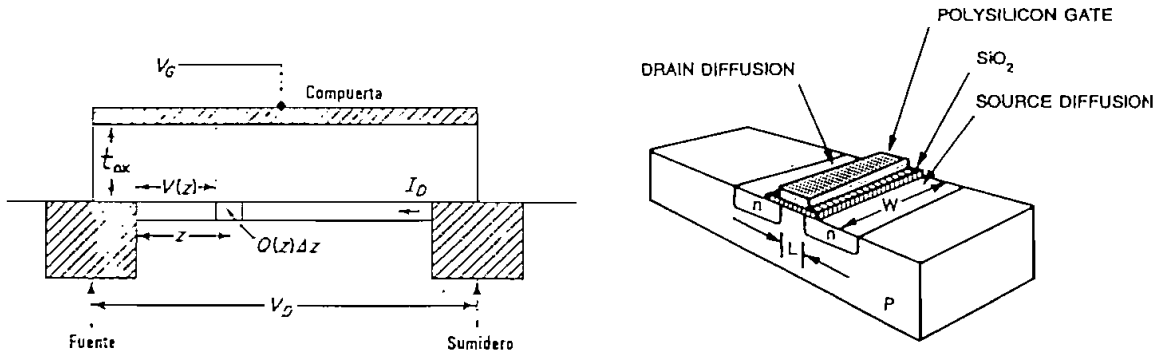


FIGURA 2.25 Estructura de un MOSFET cana "n".

En la Fig. 2.25  $t_{ox}$  es el espesor del óxido de Silicio,  $V(z)$  es el voltaje a lo largo del canal de longitud  $L$ . La variable  $z$  está comprendida en el rango  $(0 < z < L)$ .  $Q(z)$  es la carga por unidad de longitud a lo largo del canal.

Utilizando la Ec. 2.8, se tiene que la corriente que fluye por cada punto del canal comprendido entre drenaje y fuente está determinada por:

$$I_{ds} = -\mu Q(z) B(z)$$

[Ec. 2.9]

Partiendo de la relación que determina la capacitancia (C) para un condensador de placas paralelas  $C=Q/V$ , se tiene que  $Q = CV$ , siendo  $V$  la diferencia de potencial aplicado a las placas y  $Q$  la carga inducida en el sustrato por la diferencia de potencial aplicada y sabiendo que:

- i) La capacitancia de un condensador de placas paralelas, en términos de sus parámetros geométricos está dada por:

$$C = \frac{\epsilon A_p}{d}$$

[Ec. 2.10]

En donde  $A_p$  es el área de las placas,  $d$  es el espaciamiento entre las placas, y  $\epsilon$  es la permitividad relativa del dieléctrico.

Para el caso del transistor MOS los parámetros que definen la capacitancia del condensador de placas paralelas formado por la compuerta y sustrato, teniendo entre sus placas dióxido de Silicio como dieléctrico, están determinados por:

$\epsilon$ : la permitividad del aislante de la compuerta.

$t_{ox}$ : el espesor del aislante de la compuerta.

$W$ : el ancho del canal.

$L$ : la longitud del canal.

ii) El voltaje  $V_{cbg}$  en cada punto  $z$  aplicado a las placas del condensador formado, como ya se indicó anteriormente, está dado por:

$$V_{cbg} = V_{gs} - V(z)$$

[Ec. 2.11]

Podemos escribir la expresión para la carga por unidad de longitud, para cada punto del canal, como:

$$Q(z) = \frac{1}{L} \left[ -\frac{eN_L}{t_{ox}} (V_{gs} - V(z)) \right]$$

[Ec. 2.12]

Además, si la función del campo eléctrico  $E(z)$  a lo largo del canal se expresa como:

$$E(z) = \frac{d[V_{gs} - V(z)]}{dz} = -\frac{dV(z)}{dz}$$

[Ec. 2.13]

Finalmente podemos plantear la ecuación para la corriente como:

$$I_{ds} = \mu \left[ \frac{eN_L}{t_{ox}} (V_{gs} - V(z)) \right] \left[ \frac{dV(z)}{dz} \right]$$

[Ec. 2.14]

$$I_{ds} = -\frac{e\mu W}{2t_{ox}} \frac{d}{dz} [V_{gs} - V(z)]^2$$

[Ec. 2.15]

La ecuación planteada para  $Q(z)$  y la última, son válidas para cuando se cumple la condición:  $V(z) < V_0$ .

Multiplicando ambos lados de la Ec. 2.15 por  $dz$  e integrando en los límites del canal, en el extremo de la fuente y en el extremo del drenaje, se obtiene:

$$zI_{ds}|_0^L = -\frac{e\mu W}{2t_{ox}} [V_{gs} - V(z)]^2|_0^L$$

[Ec. 2.16]

$$I_{ds}L = \frac{e\mu W}{2t_{ox}} [V_{gs}^2 - (V_{gs} - V_{ds})^2]$$

[Ec. 2.17]

de modo que la expresión válida para  $V_{ds} < V_0$  es:

$$I_{ds} = \frac{e\mu W}{2t_{ox}L} [2V_{gs}V_{ds} - V_{ds}^2]$$

[Ec. 2.18]

Dado que la capa de inversión aparece solamente cuando  $V_0 > V_t$ , debemos reemplazar  $V_0$  por  $(V_0 - V_t)$ , en todo lugar que aparezca  $V_0$ . La Ec. 2.18 queda entonces como:

$$I_{ds} = \frac{e\mu W}{2t_{ox}L} [2(V_{gs} - V_t)V_{ds} - V_{ds}^2]$$

[Ec. 2.19]

De la Ec. 2.19, para voltajes fijos drenaje a fuente y de compuerta, los factores que tienen influencia en el nivel de corriente de drenaje  $I_{ds}$ , para una resistividad dada del sustrato, son aspectos que se derivan directamente de la geometría y proceso de fabricación:

- La distancia entre fuente y drenaje.
- El ancho del canal.
- El voltaje umbral  $V_t$ .
- El espesor de la capa de aislante óxido de la compuerta.
- La constante dieléctrica del aislante de la compuerta.
- La movilidad de los portadores (huecos o electrones).

Haciendo que:

$$\beta = \frac{\mu e}{t_{ox}} \left( \frac{W}{L} \right)$$

[Ec. 2.20]

la ecuación de la corriente puede escribirse como:

$$I_{ds} = \beta \left[ (V_{gs} - V_t)V_{ds} - \frac{V_{ds}^2}{2} \right]$$

[Ec. 2.21]



El parámetro  $\beta$  es el factor de ganancia del transistor MOS, dependiente de los parámetros del proceso de fabricación y de la geometría del dispositivo.

La dependencia del factor de ganancia  $\beta$  del proceso de fabricación está incluida en los términos  $\mu\epsilon/t_{ox}$ , que considera implícitamente el grado de dopaje y el espesor del  $\text{SiO}_2$ .

A la expresión  $\mu\epsilon/t_{ox}$ , se le denomina comúnmente *factor de ganancia del proceso* ( $K_p$ ). Para un proceso típico  $K_p$  está en el orden de 10 a 30  $\mu\text{A}/\text{V}^2$ . Para un proceso de fabricación dado, es común esperar una variación de  $K_p$  del 10 al 20 por ciento, principalmente debida a variaciones en las características de los materiales con los cuales se inicia la fabricación y variaciones en la construcción del  $\text{SiO}_2$ .

La dependencia de  $\beta$  de la forma geométrica está expresada en función de las dimensiones, largo y ancho. Los términos geométricos de la expresión para el cálculo de  $\beta$  están indicados en la estructura física del transistor nMOS de la Fig. 2.25.

Valores típicos de los términos de la Ec. 2.20 que determinan el valor de  $\beta$ , para un dispositivo nMOS son:

$\mu_n = 500 \text{ cm}^2/\text{V}\cdot\text{s}$ ,  $\epsilon = 4\epsilon_0 = 4 \times 8.85 \times 10^{-14} \text{ F/cm}$ , y  $t_{ox} = 500 \text{ \AA}$ . Con estos valores se tiene un valor de  $\beta$  de:

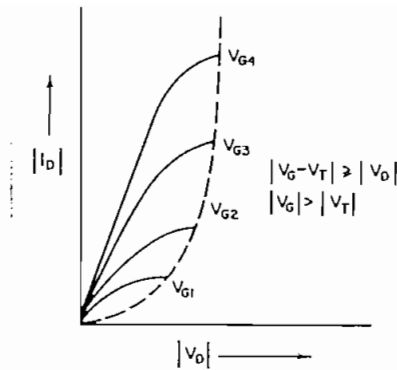


FIGURA 2.26  $I_{Dn}$  vs.  $V_{Dn}$  para cuando  $(V_{Gn} - V_{Tn}) \geq V_{Dn}$ .

$$\frac{500 \times 4 \times 8.85 \times 10^{-14}}{0.5 \times 10^{-5}} \left( \frac{W}{L} \right) = 35 \frac{W}{L} \mu A/V^2$$

[Ec. 2.22]

En la Fig. 2.26 se muestra una representación gráfica de la forma de la ecuación derivada para la característica  $I_{Dn}$  vs.  $V_{Dn}$ . En la parte presentada, la curva tiene una estructura lineal, justificando la designación de la región en estudio, como lineal.

c) Región de saturación

Como se explicó anteriormente, si  $V_{Dn}$  tiende a  $V_{Gn}$  o si llega a ser mayor, la capa de inversión no se forma en el extremo del drenaje e incrementos posteriores de  $V_{Dn}$  prácticamente no tienen influencia en el valor de la corriente.

Para determinar el valor del voltaje de estrangulamiento, para el cual se inicia la condición descrita, se debe determinar el valor para el cual la ecuación planteada para la región lineal llega a un máximo.

$$\frac{dI_{ds}}{dV_{ds}} \Big|_{V_{ds}=V_{sat}} = 0 = \beta [(V_{gs} - V_t) - V_{sat}]$$

[Ec. 2.23]

Y por lo tanto:

$$V_{sat} = V_{gs} - V_t$$

[Ec. 2.24]

La característica de voltaje-corriente para la región de saturación se obtiene sustituyendo el valor de  $V_{sat}$  en la ecuación planteada para la zona lineal, lo que nos proporciona el máximo valor de corriente del canal:

$$I_{ds} = I_{sat} = \frac{\beta}{2} (V_{gs} - V_t)^2$$

[Ec. 2.25]

En la Fig. 2.27a se muestra una representación gráfica de la condición para la cual se ingresa a la región de saturación ( $V_{sat}$ ), la cual obedece a una relación cuadrática; en la Fig. 2.27b se presenta el efecto de cualquier voltaje adicional aplicado entre drenaje-fuente más allá del voltaje

de estrangulamiento, este incremento se refleja en una disminución de la longitud de la capa de inversión, y no en un incremento de la corriente.

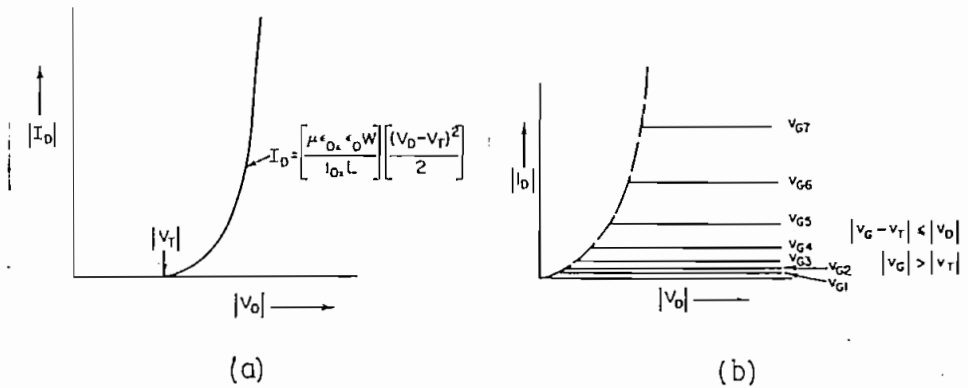


FIGURA 2.27  $I_{ds}$  vs.  $V_{ds}$  para cuando: a)  $V_{gs} - V_t = V_{ds}$ , b)  $V_{gs} - V_t < V_{ds}$ .

El resultado obtenido para  $I_{ds}$ , en la región de saturación, en realidad es una aproximación. En el dispositivo real, si la disminución de la longitud del canal es pequeña la corriente es esencialmente constante; sin embargo, si la longitud del canal es modificada o modulada apreciablemente, la relación  $W/L$  sufre un incremento efectivo, debido a la disminución de  $L$ , lo que se traduce en un incremento de la corriente, por la relación directa que existe entre este factor y la corriente. La longitud efectiva del canal está dada aproximadamente por la ecuación:

$$L_{efec} = L - \sqrt{2 \epsilon_0 \frac{e s l}{q n} (V_{ds} - [V_{gs} - V_t])}$$

[Ec. 2.26]

Una forma de la curva característica que considera el incremento de la corriente debida a la modulación de la longitud del canal se representa en la Fig. 2.28.

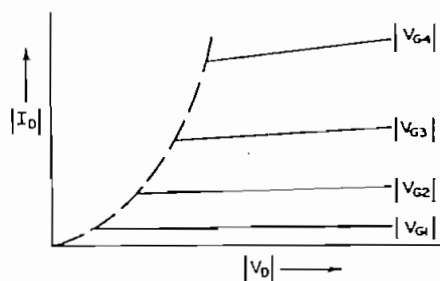


FIGURA 2.28  $I_{Dn}$  vs.  $V_{Dn}$  considerando la variación de la longitud del canal.

En resumen, las regiones de trabajo del transistor nMOS y sus principales características son:

- i) Región de corte (*cut-off*): la corriente fluye debido a la corriente parásita fuente-drenaje.
- ii) Región lineal: la capa de inversión y la corriente de drenaje se incrementan linealmente con el voltaje de compuerta y drenaje.
- iii) Región de saturación: la corriente de drenaje es prácticamente independiente del voltaje de drenaje.

El estudio realizado se ha hecho directamente orientado a los transistores nMOS. Sin embargo, las ecuaciones y el principio de funcionamiento explicados para el transistor nMOS son válidos también para el transistor pMOS; tan solo, es necesario recordar, como ya se mencionó anteriormente que en el transistor pMOS los portadores mayoritarios son los huecos y los voltajes utilizados son negativos respecto al sustrato.

#### 2.3.4 Características V-I

En la Fig. 2.29 se presenta la característica completa voltaje-corriente para los transistores MOS. Como se representa el valor absoluto de los voltajes, la forma de la curva es válida para transistores nMOS y pMOS.

En la figura se traza el límite entre las regiones lineal y de saturación, utilizando una línea entrecortada, para cuando  $|V_{ds}| = |V_{gs} - V_t|$ .

La resistencia de salida, es decir la resistencia del canal, en la región lineal puede obtenerse diferenciando la Ec. 2.21 con respecto a  $V_{ds}$ , lo que proporciona el valor de la conductancia de salida.

$$\lim_{V_{ds} \rightarrow 0} \frac{dI}{dV_{ds}} = \beta (V_{gs} - V_t)$$

[Ec. 2.27]

Lo que finalmente proporciona el valor aproximado de la resistencia del canal  $R_c$ :

$$R_{c(\text{lineal})} = \frac{1}{\beta (V_{gs} - V_t)}$$

[Ec. 2.28]

La última expresión indica que la resistencia del canal en la región lineal está controlada por el voltaje compuerta-fuente, y es válida para voltajes compuerta-fuente que mantienen constante la movilidad en el canal.

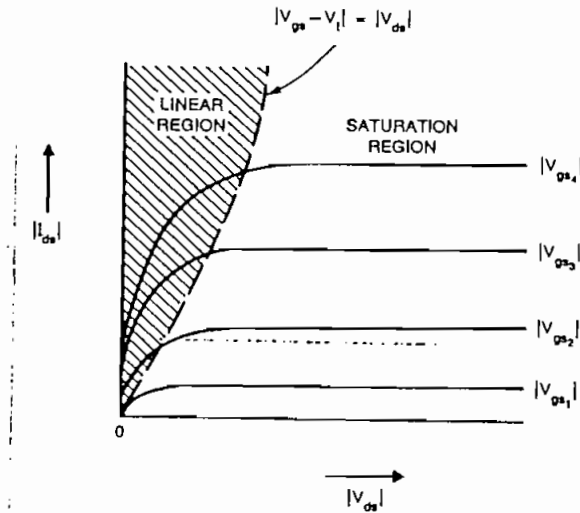


FIGURA 2.29 Característica de  $I_{ds}$  vs.  $V_{ds}$  para transistores nMOS y pMOS.

En la región de saturación, el transistor MOS se comporta como una fuente de corriente, siendo casi independiente de  $V_{ds}$ . Esto puede verificarse derivando la Ec. 2.25:

$$\frac{dI_{ds}}{dV_{ds}} = \frac{d\left(\frac{\beta}{2}(V_{gs}-V_t)^2\right)}{dV_{ds}} = 0$$

[Ec. 2.29]

### 2.3.5 Comparación entre transistores canal "n" y "p"

Tecnológicamente, la fabricación de los transistores "p" es más fácil que la de los dispositivos "n". Esto se debe a que la mayoría de los contaminantes en los procesos de fabricación MOS son iones móviles positivamente cargados, los cuales son atrapados en la capa de óxido aislante.

Ya en operación, en el transistor nMOS, la compuerta es positiva con respecto al sustrato, y los contaminantes que durante la fabricación se depositaron en el  $\text{SiO}_2$ , se agrupan ahora en el interfaz  $\text{SiO}_2$ -sustrato. La carga positiva de esta capa de iones atrae a los electrones libres del canal, lo cual tiende a encender el transistor prematuramente.

En el transistor pMOS, debido al voltaje negativo aplicado a la compuerta, los iones contaminantes positivos son atraídos, dentro del  $\text{SiO}_2$ , al lado opuesto del interfaz  $\text{SiO}_2$ -



sustrato, y en este lugar no afectan la operación del transistor. Por lo tanto, la fabricación de transistores nMOS requiere procesos tecnológicos más complejos que ayuden a evitar estos elementos contaminantes.

Las movilidades de los portadores mayoritarios, sean huecos o electrones son dependientes de la temperatura y de la concentración de impurezas en el material semiconductor. En el Silicio, bajo intensidades de campo normales, la movilidad de los electrones es mayor a dos veces la movilidad de los huecos. Esto se traduce en que el dispositivo pMOS tiene una resistencia de ON mayor al doble que su dispositivo nMOS equivalente, con la misma geometría y bajo las mismas condiciones de operación. En otras palabras, el dispositivo pMOS debe tener más allá del doble de área que el nMOS para presentar la misma resistencia. Por lo tanto, un circuito utilizando solo transistores nMOS es más pequeño que uno que utilice solo transistores pMOS, lo que implicaría mayores escalas de integración. Como se explica en el Cap. 4, una menor área del elemento le permite también ser más rápido.

#### 2.3.6 Ajuste del voltaje umbral.

Como se indicó anteriormente, el voltaje umbral  $V_t$  de un transistor MOS puede definirse como el voltaje, aplicado entre la compuerta y la fuente, bajo el cual la corriente de drenaje a fuente ( $I_{d_s}$ ), cae a cero.

El voltaje umbral es función de algunos parámetros, entre estos:

- a) El material de la compuerta.
- b) El material aislante de la compuerta.
- c) El espesor del aislante de la compuerta.
- d) Dopaje del canal.
- e) Impurezas del interfaz silicio-aislante.
- f) Voltaje entre fuente y sustrato  $V_{fb}$ .
- g) La temperatura. Un incremento de temperatura produce una disminución del valor absoluto del voltaje umbral. Esta variación es de aproximadamente  $-4\text{mV}/^\circ\text{C}$  para un sustrato con un alto nivel de dopaje, y  $-2\text{mV}/^\circ\text{C}$  para un bajo nivel de dopaje.

Existen dos técnicas utilizadas generalmente para ajustar el voltaje umbral a un valor deseado. Una es variar el nivel de dopaje en el interfaz silicio-aislante, por medio de la implantación de iones, y la otra, utilizar un material diferente para el aislante de la compuerta.

Para alterar las características del aislante se combinan el óxido de Silicio con un material de permitividad diferente. Se utiliza generalmente  $\text{Si}_3\text{N}_4$ , de permitividad relativa 7.5, que combinado con el Dióxido de Silicio, de permitividad relativa 3.9, proporciona una permitividad relativa efectiva de aproximadamente 6, sustancialmente mayor que la constante dieléctrica del Oxido de Silicio. Con este

proceso se logra disponer de una capa electricamente equivalente a una de menor espesor de  $\text{SiO}_2$ , conservando el mismo espesor de la capa original de  $\text{SiO}_2$ .

### 2.3.7 Efecto de cuerpo

Cualquier circuito que incluya dispositivos MOS, está construido sobre un sustrato común, lo que hace que todos sus elementos tengan el mismo voltaje de sustrato. Sin embargo, dentro del circuito no todos los transistores tienen iguales voltajes en sus terminales "fuente", respecto al sustrato.

Por ejemplo en la Fig. 2.30 se presentan dos transistores conectados en serie. En el transistor T1 se tiene un voltaje fuente-sustrato ( $V_{sb1}$ ) igual a cero; por el contrario, en el transistor T2 se tiene un voltaje fuente-sustrato:  $V_{sb2}$  diferente de cero, y así sucesivamente; es decir, un incremento del voltaje  $V_{sb}$  mientras se avanza en la cadena de transistores en serie. El efecto neto es un incremento del voltaje umbral ( $V_t$ ) ( $V_{t1} > V_{t2}$ ) de los transistores, efecto se denomina de "cuerpo" (*body effect*).

El voltaje umbral efectivo puede ser estimado, aproximadamente, con la siguiente expresión:

$$V_t = V_{t(0)} + \gamma (V_{sb})^{1/2}$$

[Ec. 2.30]

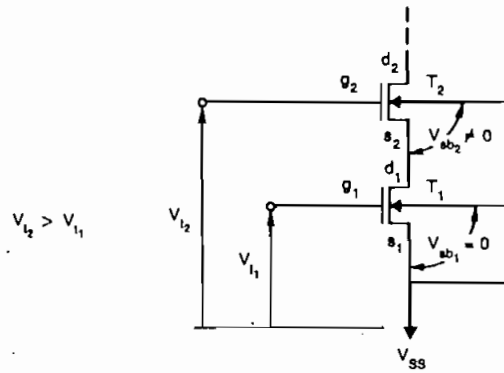


FIGURA 2.30 Efecto de cuerpo en transistores conectados en serie.

En la Ec. 2.30, el voltaje umbral resultante depende del voltaje umbral cuando  $V_{bs}=0$ , y de la constante  $\gamma$ , que describe el efecto de la polarización del sustrato. El signo negativo de la expresión se utiliza para el caso del transistor pmos.

El aumento resultante en el valor del voltaje umbral, lleva a tener corrientes menores, lo que a la vez conduce a tener circuitos más lentos.

El tipo de proceso utilizado para la fabricación de los transistores MOS tiene gran influencia en el valor de la constante  $\gamma$ , valores típicos en los que se encuentra esta constante están en el rango de 0.4 a 1.2. Una expresión más detallada para definir el voltaje umbral es:

$$V_c = V_{t(0)} + \gamma [\sqrt{V_{ds} + 2\phi_F} - \sqrt{2\phi_F}]$$

[Ec. 2.31]

Donde  $\phi_F$  es una constante y  $\gamma$  puede calcularse como:

$$\gamma = \left( \frac{t_{ox}}{\epsilon_{ox}} \right) \sqrt{2q\epsilon_{si}n}$$

[Ec. 2.32]

En la Ec. 2.32,  $q$  es la carga del electrón,  $\epsilon_{ox}$  es la constante dieléctrica del Dióxido de Silicio,  $\epsilon_{si}$  es la constante dieléctrica del sustrato de silicio, y  $n$  es la densidad de concentración del sustrato. Para  $V_{ds} \gg \phi_F$ , se obtiene la Ec. 2.30, planteada originalmente.

$K_p$ ,  $V_{t(0)}$  y  $\gamma$  son parámetros comunes, utilizados en la especificación del modelo de simulación de los transistores MOS, para el popular simulador eléctrico SPICE, u otros que acepten su formato.

## 2.4 EL INVERSOR CMOS - CARACTERÍSTICAS DC

En la Fig. 2.31 se presenta el esquema del inversor CMOS, incluida la nomenclatura de los terminales de sus transistores y los voltajes de salida  $V_o$  y de entrada  $V_i$  con los cuales se deriva la característica de transferencia DC.

Esta característica se deriva para determinar la variación del voltaje de salida debido a variaciones en el voltaje de entrada.

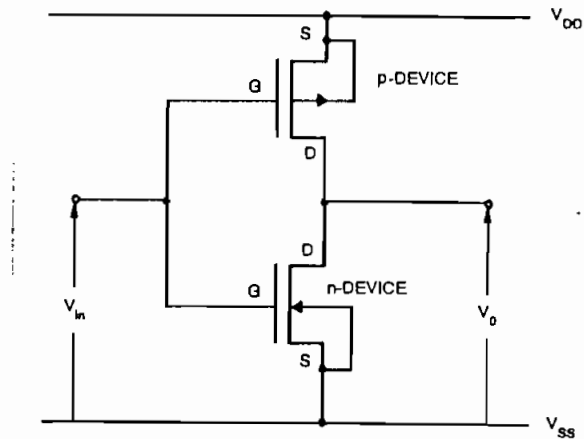


FIGURA 2.31 Inversor CMOS

De la Fig. 2.31 y del funcionamiento de cada dispositivo, se debe mencionar que:

- a) En el caso del transistor "n", los electrones fluyen de fuente a drenaje y la corriente se representa con un sentido de drenaje a fuente, determinando también un voltaje positivo de drenaje-fuente.
- b) En el caso del transistor "p", los huecos fluyen de fuente a drenaje, y la representación de la corriente es también de fuente a drenaje. Esta corriente determina por lo tanto un voltaje negativo drenaje-fuente.

c) La corriente del dispositivo "n" se representa en el mismo sentido que la del dispositivo "p". Sin embargo, considerando el flujo de los portadores mayoritarios, se tiene que:  $I_{d \rightarrow n} = -I_{d \rightarrow p}$ .

d) De acuerdo al esquema del inversor de la Fig. 2.31, se tiene que:  $V_{DD} = V_{\rightarrow dp} + V_{d \rightarrow n} = V_{\rightarrow dp} + V_o$ . Como lo común es referirse al voltaje drenaje-fuente, la última expresión se escribe como:  $V_{DD} = -V_{d \rightarrow p} + V_o$ , con  $V_{d \rightarrow p}$  negativo.

e) El voltaje de salida está determinado por la relación:

$$V_o = V_{DD} + V_{d \rightarrow p}$$

[Ec. 2.33]

f) Debe observarse que las dos compuertas de los dos transistores están unidas y conectadas al voltaje de entrada y para el transistor "n",  $V_i = V_{g \rightarrow n}$ .

g) Para el transistor "p" se puede deducir el voltaje  $V_{g \rightarrow p}$  de:  $V_{DD} = V_{\rightarrow gp} + V_i$ . Lo común es referirse al voltaje compuerta-fuente, reescribiendo la expresión anterior:  $V_{DD} = -V_{g \rightarrow p} + V_i$ , con  $V_{g \rightarrow p}$  negativo. Finalmente despejando  $V_{g \rightarrow p}$  obtenemos:

$$V_{g \rightarrow p} = V_i - V_{DD}$$

[Ec. 2.34]

Combinando las características  $I_{d_s}$  vs.  $V_{d_s}$  de los dos transistores puede determinarse la característica entrada-salida del inversor CMOS.

En la Fig. 2.32a se presenta, en un solo eje coordenado, las curvas de  $I_{d_s}$  vs.  $V_{d_s}$  para los dos transistores. En la Fig. 2.32b se toma el valor absoluto de  $I_{d_s}$  del transistor pMOS, lo que hace su característica simétrica, respecto a la característica del transistor nMOS, en el eje x.

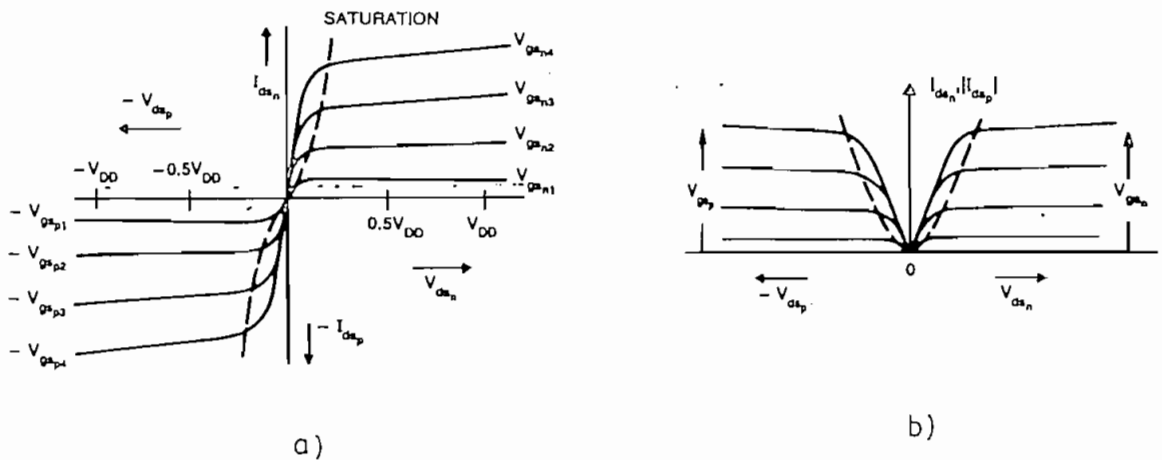


FIGURA 2.32 Derivación gráfica de la característica del inversor.

Finalmente se toma el valor absoluto de  $V_{d_s}$  del dispositivo pMOS, y se superponen las características de los dos dispositivos en el eje positivo, como se indica en la Fig. 2.33. La característica de transferencia queda determinada por los puntos en los cuales se tiene un mismo valor de  $V_{d_s}$ .



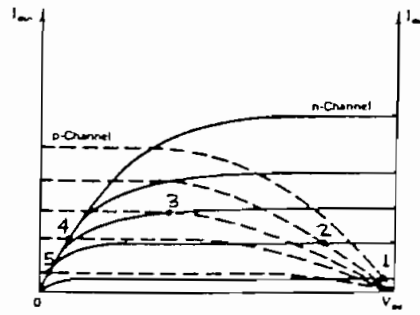


FIGURA 2.33 Corrientes  $I_{ds}$  de los transistores "n" y "p" superpuestas.

Resolviendo para  $I_{ds,n} = I_{ds,p}$  y  $V_{i1} = V_{i,p}$  se obtiene la curva de transferencia deseada (Fig. 2.34).

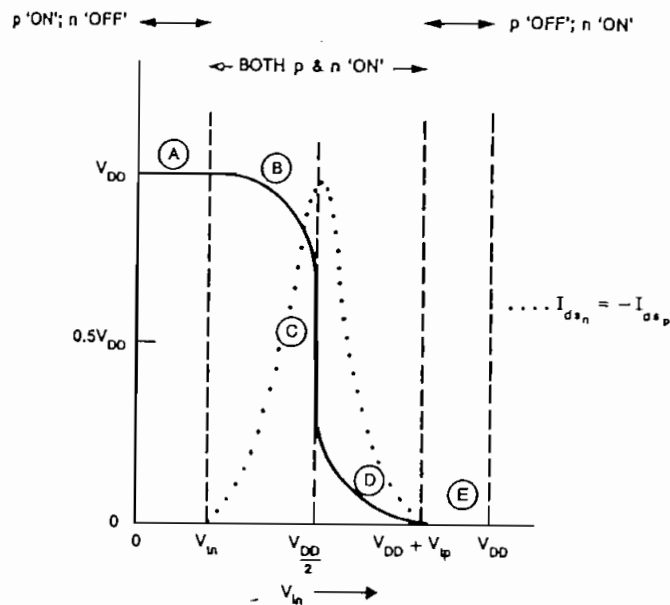


FIGURA 2.34 Curva de transferencia DC del inversor CMOS y zonas de operación.

En la Fig. 2.34 se presenta la curva del voltaje de salida en función del voltaje de entrada, y se señalan 5 zonas, comprendidas entre dos niveles de voltaje, que luego se analizan individualmente señalando el comportamiento de los dos transistores MOS en cada una de las zonas.

Como se explica más adelante, el punto de conmutación se diseña típicamente para que ocurra en el 50% de la magnitud de la fuente de polarización (aproximadamente  $V_{DD}/2$ ).

Durante la transición, tanto el transistor nMOS como el pMOS, están en el estado "ON", lo que produce un pulso de corriente asumido por la fuente de polarización y que se representa en la Fig. 2.34 con una línea de puntos.

#### 2.4.1 Zona A

En la zona A:  $0 \leq V_i \leq V_{tn}$ . En este rango de  $V_i$ , el transistor "n" opera en la región de corte y el "p" en la región lineal de sus curvas características respectivas.

El transistor "n" opera en la región de corte ( $I_{d-n}=0$ ), pues  $V_i=V_{g-n} \leq V_{tn}$ , lo que impide que haya un camino de circulación de corriente entre  $V_{DD}$  y Gnd. El análisis que se realiza es asumiendo que el inversor no tiene ninguna carga conectada a su salida.

El transistor "p", en esta zona, de acuerdo a la Ec. 2.34, tiene un voltaje  $V_{g-p}$  comprendido en el rango  $-V_{DD} \leq V_{g-p} \leq (-V_{DD} + V_{tn})$ , suficiente para establecer la capa de inversión en el canal. Considerando que no hay circulación de corriente, el voltaje de salida es prácticamente  $V_{DD}$ .

Si asumimos por un instante que se tiene una carga conectada al inversor, ésta proporcionaría un camino de corriente que produciría una pequeña caída de tensión en el transistor pMOS, con lo que se tendría una pequeña diferencia de tensión negativa entre drenaje-fuente que ubicaría el punto de trabajo del transistor "p" en la región lineal de operación. Un punto de operación típico (1) se representa en la Fig. 2.33. Este punto corresponde a un pequeño valor de  $V_{d\rightarrow p}$  para el transistor "p" y un voltaje  $V_{d\rightarrow n}=V_o$  de aproximadamente  $V_{DD}$ , para el transistor "n", lo que ubica al punto en el extremo derecho de la Fig. 2.33. Al incrementar el voltaje de entrada, el voltaje  $V_{g\rightarrow p}$  se hace menos negativo, provocando que el punto de operación se desplace hacia valores de corriente menores.

Como no existe carga conectada, en realidad no hay corriente que fluye por los dispositivos, hasta que  $V_i$  cruce el voltaje umbral del transistor nMOS. Como  $I_{d\rightarrow n}=-I_{d\rightarrow p}=0$ , el voltaje de salida de acuerdo a la Ec. 2.32, es por lo tanto:  $V_o = V_{DD} - 0V = V_{DD}$ .

#### 2.4.2 Zona B.

En esta zona el  $V_i$  está en el rango:  $V_{tn} \leq V_i < V_{DD}/2$ . El transistor "p" opera aún en la región lineal y el "n" en la región de saturación. Las condiciones iniciales en esta zona son:

- a) Un voltaje  $V_{o=p}$  suficiente para formar el canal.
- b) Un voltaje  $V_{d=p}$  aún suficiente también para mantener al dispositivo en la región lineal, permitiendo la circulación de una pequeña corriente.
- c) Un voltaje  $V_{o=n} \geq V_{tn}$  que permite formar la capa de inversión.
- d) Un voltaje  $V_{d=n}=V_o$ , aproximadamente igual a  $V_{DD}$  que lleva al transistor "n" a la región de saturación.

Bajo estas condiciones circula una corriente a través de los dos transistores. La corriente total que fluye por el circuito se incrementa al incrementar el voltaje de entrada y la salida tiende a disminuir abruptamente, como se indica en la Fig. 2.34. En la Fig. 2.32 se presenta un punto típico de esta zona de la curva característica del inversor (2).

La descripción dada puede demostrarse matemáticamente de la siguiente manera:

El inversor CMOS puede representarse en esta zona por una resistencia para el transistor pMOS (opera en la región lineal) y una fuente de corriente para el nMOS (opera en la región de saturación) (Fig. 2.35). La corriente de saturación del dispositivo nMOS ( $I_{d=n}$ ), utilizando la Ec. 2.25, puede escribirse reemplazando  $V_{o=n}=V_i$  como:

Reemplazando los valores anteriores en la Ec. 2.21:

$$I_{dnp} = -\beta_p [(V_1 - V_{DD} - V_{tp}) (V_o - V_{DD}) - \frac{1}{2} (V_o - V_{DD})^2] \quad [\text{Ec. 2.39}]$$

$$\beta_p = \frac{\mu_p \epsilon}{t_{ox}} \left( \frac{W}{L_p} \right) \quad [\text{Ec. 2.40}]$$

Igualando el valor de las corrientes de los dos dispositivos, según lo explicado anteriormente:  $I_{dnp} = -I_{dsn}$  y despejando el valor del voltaje de salida, se tiene:

$$V_o = (V_1 - V_{tp}) + [(V_1 - V_{tp})^2 - 2(V_1 - \frac{V_{DD}}{2} - V_{tp}) V_{DD} - \frac{\beta_n}{\beta_p} (V_1 - V_{tn})^2]^{1/2} \quad [\text{Ec. 2.41}]$$

La Ec. 2.41 puede escribirse, reemplazando todo el término elevado a " ½ " por X, como  $V_o = V_1 - V_{tp} + X$ , con X siempre positivo. Partiendo de esta relación podemos escribir que  $V_o > V_1 - V_{tp}$ , es decir que  $V_{dsn} = V_{dnp} = V_1 - V_o < V_{tp}$ , para el rango de voltaje  $V_{tn} \leq V_1 < V_{DD}/2$ .

La Ec. 2.41 puede graficarse asumiendo que:

- a) Que el valor de los voltajes de umbral de los dos dispositivos son iguales, recordando que tienen signo contrario:  $V_{tn} = -V_{tp}$ .

- b) Que los dos dispositivos tienen igual factor de ganancia  $\beta$ , considerando iguales dimensiones geométricas y que la movilidad de los electrones es igual a la de los huecos.
- c) Hablando ya de valores numéricos, puede utilizarse para  $V_{DD}$  un valor de 5V, y para  $V_{t_n}$ , alrededor de 1V.

Resolviendo la ecuación para  $V_{t_n} \leq V_i < V_{DD}/2$ , puede probarse que se obtiene la forma representada en la Fig. 2.34.

### 2.4.3 Zona C

Al incrementar el voltaje de entrada, existe un punto en el que el transistor "p" pasa de la región lineal a la región de saturación, debido principalmente a la disminución del voltaje de salida que provoca un voltaje  $V_{d_{n-p}}$  más negativo; mientras tanto, el transistor nMOS continúa en la región de saturación. En esta pequeña zona los dos dispositivos, trabajan en la zona de saturación.

Las expresiones correspondientes a las corrientes de saturación para los dos transistores se expresan como:

$$I_{d_{n-p}} = \frac{\beta_p}{2} (V_i - V_{t_n})^2$$

[Ec. 2.42]

$$I_{dnp} = \frac{\beta_p}{2} (V_i - V_{DD} - V_{tp})^2$$

[Ec. 2.43]

En esta zona la corriente alcanza su valor máximo y considerando nuevamente que las corrientes son iguales, podemos despejar el valor o valores de  $V_i$  para los cuales ocurre la condición de saturación para los dos transistores (zona C). Este valor, como se demuestra a continuación, es único. Un punto que representa la zona C se representa en la Fig. 2.33 (3).

$$V_i = \frac{V_{DD} + V_{tp} + V_{tn} \sqrt{\frac{\beta_n}{\beta_p}}}{1 + \sqrt{\frac{\beta_n}{\beta_p}}}$$

[Ec. 2.44]

La expresión de  $V_i$  proporciona la base para definir el voltaje umbral de la compuerta ( $V_{i_{nv}}$ ) para el cual  $V_o = V_i$ .

Considerando las condiciones  $\beta_n = \beta_p$  y  $V_{tn} = -V_{tp}$ , se obtiene que:

$$V_i = \frac{V_{DD}}{2}$$

[Ec. 2.45]

El circuito equivalente en esta zona es un par de fuentes de corriente en serie, lo que produce una condición de inestabilidad. Esta inestabilidad se refleja en que una pequeña variación del voltaje de entrada tiene un efecto muy significativo en la salida, lo que en la Fig. 2.34 se representa con una línea vertical (en la zona C).

Se debe considerar que para la deducción de este valor, se asume que las fuentes de corriente se mantienen constantes y son independientes de los voltajes drenaje-fuente. En los dispositivos reales, esta corriente aumenta ligeramente con los incrementos del voltaje drenaje-fuente, lo que se traduce en un pequeño grado de curvatura en la zona C, dejando de ser tan perpendicular como la representación de la Fig. 2.34.

#### 2.4.4 Zona D

En esta zona  $V_1$  está en el rango:  $V_{DD}/2 < V_1 \leq V_{DD}/2 + V_{tp}$ . El dispositivo "p" trabaja en la región de saturación y el "n" trabaja en la región lineal. Puede realizarse un análisis similar al realizado para la zona B, para confirmar las regiones de operación para cada dispositivo. El circuito equivalente en este caso es una resistencia, para el transistor nMOS, en serie con una fuente de corriente, que representa al pMOS. (Fig. 2.36). Las ecuaciones de corriente para cada uno de los dispositivos se escriben como:



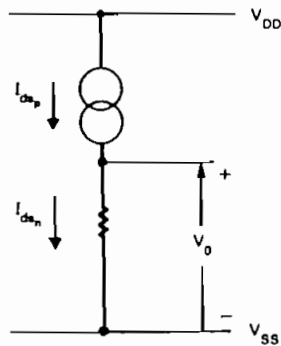


FIGURA 2.36 Circuito equivalente para la zona D de operación del inversor.

$$I_{dsp} = -\frac{\beta_p}{2} (V_i - V_{DD} - V_{tp})^2$$

[Ec. 2.46]

$$I_{dsn} = \beta_n \left[ (V_i - V_{tn}) V_o - \frac{1}{2} V_o^2 \right]$$

[Ec. 2.47]

Igualando las dos expresiones para las corrientes, considerando el signo para cada una, se puede despejar el voltaje de salida para la zona D.

$$V_o = (V_i - V_{tn}) - \left[ (V_i - V_{tn})^2 - \frac{\beta_p}{\beta_n} (V_i - V_{DD} - V_{tp})^2 \right]^{1/2}$$

[Ec. 2.48]

Un punto típico de operación (4) de esta zona se presenta en la Fig. 2.33.

La Ec. 2.48 puede escribirse, reemplazando todo el término elevado a la "  $\frac{1}{2}$  " por Y, como  $V_o = V_i - V_{tn} - Y$ , con Y siempre positivo. Partiendo de esta relación podemos escribir que  $V_o < V_i - V_{tn}$ , es decir que  $V_{odn} = V_{odp} = V_i - V_o > V_{tn}$ , para el rango de  $V_i$  de esta zona.

#### 2.4.5 Zona E

En esta zona  $V_i$  está en el rango:  $V_i \geq V_{DD} + V_{tp}$  y el voltaje compuerta-fuente, dado por  $V_{gsp} = V_i - V_{DD}$ , es más positivo que  $V_{tp}$ , lo que lleva al dispositivo a la región de corte; por el contrario, el dispositivo nMOS opera en la región lineal de su característica. Nuevamente, en esta zona la corriente que circula por los dos dispositivos es cero y el voltaje de salida es también cero, pues el transistor nMOS presenta una resistencia baja y no existe circulación de corriente. En la Fig. 2.33 se presenta un punto de trabajo para esta zona (5).

En la Tabla 2.4 se resumen las condiciones de voltaje que llevan a los dos transistores a su operación en sus diferentes regiones de trabajo. Los rangos de voltaje pueden establecerse con ayuda de las ecuaciones planteadas.

En la Tabla 2.5 se presenta un resumen de la operación del inversor CMOS en cada una de las zonas en que se dividió la curva de transferencia. Los valores presentados se calculan asumiendo que:  $V_{tn} = -V_{tp} = 1V$ ,  $V_{DD} = 5V$ ,  $\beta_n/\beta_p=1$ .

	CUTOFF	LINEAR	SATURATION
p-device	$V_{gs_p} > V_{tp};$	$V_{gs_p} < V_{tp};$ $V_{in} < V_{tp} + V_{DD}$	$V_{gs_p} < V_{tp};$ $V_{in} < V_{tp} + V_{DD}$
	$V_{in} > V_{tp} + V_{DD}$	$V_{gd_p} < V_{tp};$ $V_{in} - V_O < V_{tp}$	$V_{gd_p} > V_{tp};$ $V_{in} - V_O > V_{tp}$
n-device	$V_{gs_n} < V_{tn};$	$V_{gs_n} > V_{tn};$ $V_{in} > V_{tn}$	$V_{gs_n} > V_{tn};$ $V_{in} > V_{tn}$
	$V_{in} < V_{tn}$	$V_{gd_n} > V_{tn};$ $V_{in} - V_{tn} > V_{tn}$	$V_{gd_n} < V_{tn};$ $V_{in} - V_{tn} < V_{tn}$

TABLA 2.4 Relaciones de voltaje para la operación del inversor CMOS.

REGION	CONDITION	p-DEVICE	n-DEVICE	OUTPUT
A	$0 \leq V_{in} \leq V_{tp}$	linear	cut-off	$V_O = V_{DD}$
B	$V_{tp} \leq V_{in} < \frac{V_{DD}}{2}$	linear	saturated	$*V_O = (V_{in} + 1) + \sqrt{15 - 6V_{in}}$
C	$V_{in} = \frac{V_{DD}}{2}$	saturated	saturated	$V_O \neq f(V_{in})$
D	$\frac{V_{DD}}{2} < V_{in} \leq V_{DD} - V_{tp}$	saturated	linear	$*V_O = (V_{in} - 1) - \sqrt{6V_{in} - 15}$
E	$V_{in} \geq V_{DD} - V_{tp}$	cut-off	linear	$V_O = 0$

TABLA 2.5 Resumen de la operación del inversor CMOS

#### 2.4.6 Influencia de la relación $\beta_n/\beta_p$ en la curva de transferencia.

En la Fig. 2.37 se presentan graficadas las curvas de transferencia para distintos valores de la relación  $\beta_n/\beta_p$ . De la figura se desprenden las siguientes características:

- a) Al incrementarse la relación  $\beta_n/\beta_p$  el punto de transición de  $V_o$  se desplaza hacia la izquierda.
- b) A pesar del desplazamiento del punto de transición, dicha transición conserva aún su agudeza, es decir la forma practicamente vertical en la zona C de transición.
- c) Por lo mencionado en b), el comportamiento en la conmutación no es alterado.
- d) Puede deducirse que el voltaje umbral de la compuerta  $V_{INV}$ , definido para que  $V_i = V_o$ , es dependiente de la relación  $\beta_n/\beta_p$ .

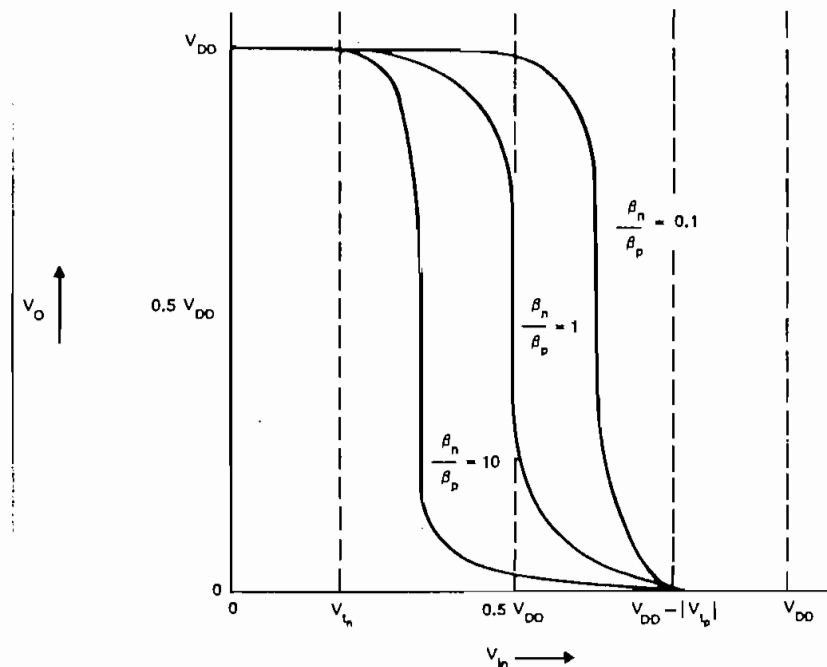


FIGURA 2.37 Influencia de la relación  $\beta_n/\beta_p$  en la característica DC del inversor CMOS.

Cuando se desea alterar la relación  $\beta_n/\beta_p$  es necesario alterar las dimensiones del canal: longitud L y ancho W.

Una relación de  $\beta_n/\beta_p=1$  es deseable desde el punto de vista de operación, ya que permite tiempos iguales de carga y descarga sobre una carga capacitiva, lo que provee iguales capacidades de drenar y entregar corrientes (Cap. 4).

#### 2.4.7 Influencia de la temperatura en la característica de transferencia.

En la Ec. 2.49 se reescribe la ecuación del factor de ganancia  $\beta$ . Este factor es directamente proporcional a la movilidad de los portadores mayoritarios de carga ( $\mu$ ).

$$\beta = \frac{\mu e}{t_{ox}} \left( \frac{W}{L} \right)$$

[Ec. 2.49]

Un incremento en la temperatura de un dispositivo MOS produce la disminución de la movilidad efectiva de los portadores ( $\mu$ ) en el canal. Esta disminución produce a su vez una disminución del factor  $\beta$ . El factor  $\beta$  está relacionado con la temperatura T de la siguiente manera:

$$\beta \propto T^{-3/2}$$

[Ec. 2.50]

y por lo tanto, también la corriente es dependiente de la temperatura en la misma proporción:

$$I_{ds} \propto T^{-3/2}$$

[Ec. 2.51]

La función de transferencia de voltaje depende de la relación  $\beta_n/\beta_p$ ; sin embargo, ya que tanto la movilidad de electrones y huecos es afectada por la temperatura en igual proporción, esta relación de factores de ganancia es prácticamente independiente de la temperatura, y por lo tanto, también lo es la función de transferencia.

#### 2.4.8 Margen de ruido.

El margen o inmunidad al ruido es un parámetro íntimamente ligado a la característica de voltaje entrada-salida. Este parámetro permite conocer el voltaje de ruido permisible a la entrada de una compuerta de tal manera que la salida no sea afectada y se lo especifica en términos del margen de ruido en bajo  $NM_L$  (LOW noise margin) y del margen de ruido en alto  $NM_H$  (HIGH noise margin). Estos dos valores están representados en la Fig. 2.38.

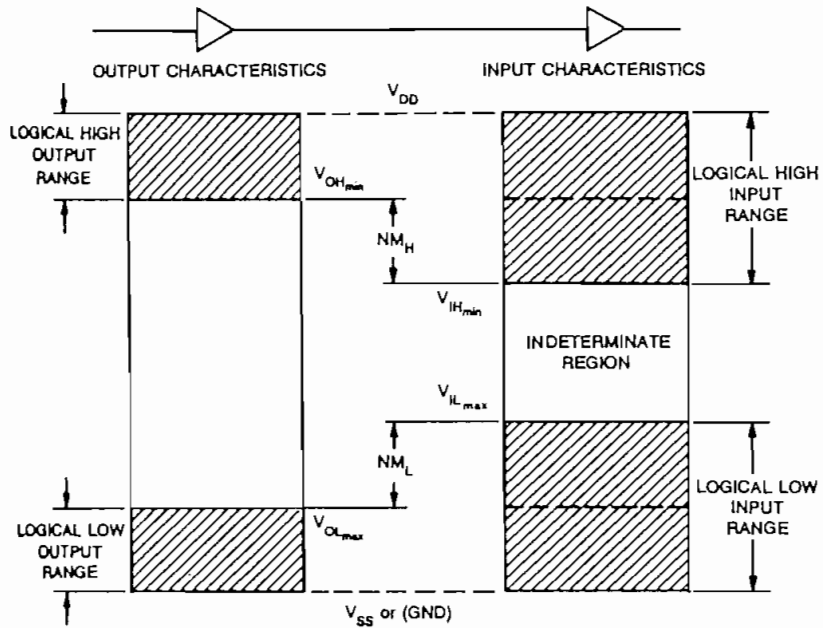


FIGURA 2.38 Definiciones del Margen de Ruido.

En la Fig. 2.38 se presentan dos compuertas en cascada de tal manera que la primera maneja a la segunda, bajo estas condiciones se tiene:

- a)  $NM_L$  se define como la diferencia en magnitud entre el voltaje máximo de salida en bajo ( $V_{OL_{max}}$ ) de la primera compuerta (que maneja a la segunda) y el voltaje máximo de entrada en bajo ( $V_{IL_{max}}$ ) reconocido por la segunda compuerta.

$$NM_L = |V_{IL_{max}} - V_{OL_{max}}|$$

[Ec. 2.52]

b)  $NM_H$  se define como la diferencia en magnitud entre el voltaje mínimo de salida en alto ( $V_{OHmin}$ ) de la primera compuerta y el voltaje mínimo de entrada en alto ( $V_{IHmin}$ ), reconocido por la segunda compuerta.

$$NM_H = |V_{OHmin} - V_{IHmin}|$$

[Ec. 2.53]

Generalmente es deseable tener  $V_{IH} = V_{IL}$  (para evitar la región de indeterminación) y que este valor de voltaje se encuentre en la mitad del rango comprendido entre  $V_{OL}$  y  $V_{OH}$ . El cumplir estos requerimientos implica que la característica de transferencia presente la zona de transición con una forma bastante abrupta.

En la Fig. 2.39 se presenta la curva de transferencia del inversor CMOS, incluidos los puntos  $V_{IL}$ ,  $V_{OL}$ ,  $V_{IH}$  y  $V_{OH}$ .

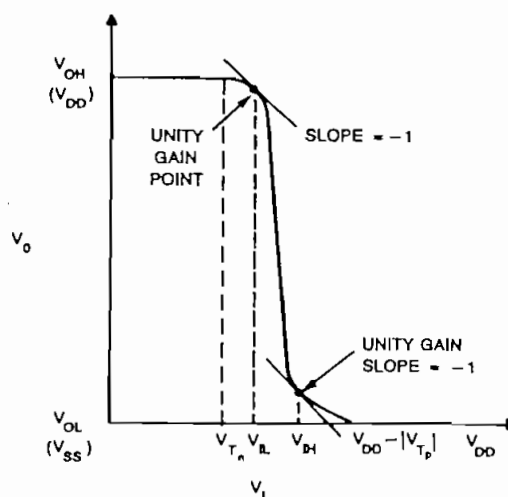


FIGURA 2.39 Márgenes de ruido en el inversor CMOS.



Considerando las características deseadas para los voltajes de entrada y salida, los márgenes de ruido se definen como:

$$NM_L = V_{IL} - V_{OL} \quad [\text{Ec. 2.54}]$$

$$NM_H = V_{OH} - V_{IH} \quad [\text{Ec. 2.55}]$$

Para determinar las expresiones para  $V_{IL}$  y  $V_{IH}$  se utilizan las ecuaciones de los dispositivos, aplicando criterios matemáticos para los puntos señalados en la Fig. 2.39.

Como  $V_{OL}=0$ ,  $NM_L = V_{IL}$ , determinando analíticamente el valor de  $V_{IL}$  se determina  $NM_L$ , si en la expresión que se obtenga se reemplazan la condición  $\beta_n/\beta_p=1$ , se obtiene:

$$NM_L = V_{IL} = \frac{3V_{dd} - 3|V_{tp}| + 5V_{tn}}{8} \quad [\text{Ec. 2.56}]$$

Realizando un proceso similar para  $V_{IH}$ , y como  $V_{OH}=5V$ , el margen de ruido en alto se reduce a:  $NM_H = V_{DD} - V_{IH}$ .

$$V_{IH} = \frac{5V_{dd} - 5|V_{tp}| + 3V_{tn}}{8} \quad [\text{Ec. 2.57}]$$

Realizando las operaciones algebraicas necesarias, finalmente se obtiene:

$$NM_H = \frac{3V_{DD} + 5|V_{tp}| - 3V_{tn}}{8}$$

[Ec. 2.58]

Reemplazando valores típicos:  $V_{tn} = -V_{tp} = 0.2V_{DD}$ , se obtiene que:  $NM_L = NM_H = 0.425V_{DD}$ .

Suponiendo que los valores de  $V_{tn}$  y  $V_{tp}$  se incrementan, también lo harán  $NM_L$  y  $NM_H$ . Si los valores de  $NM_L$  o  $NM_H$  se reducen, entonces una compuerta puede ser más susceptible a ruido de conmutación que puede estar presente en las entradas. Esta es la razón de preocuparse de los márgenes de ruido. A menudo, puede ser necesario comprometer los márgenes de ruido para mejorar la velocidad de un circuito.

## 2.5 LA COMPUERTA DE TRANSMISIÓN - CARACTERÍSTICAS DC

La estructura de la compuerta de transmisión o switch complementario se presentó anteriormente. Este switch está formado por un transistor "n" y un "p", con sus fuentes y drenajes respectivos conectados entre sí. La señal  $\phi$  de control se aplica a la compuerta del transistor "n" y su complemento a la compuerta del transistor "p", como se indica en la Fig. 2.40.

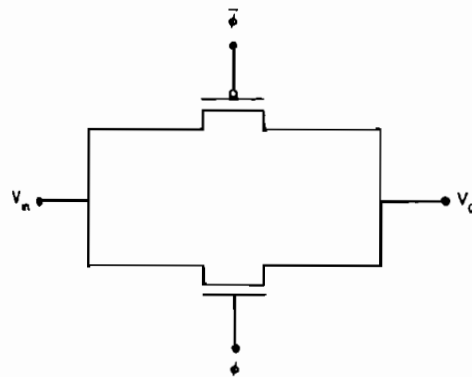


FIGURA 2.40 Compuerta de transmisión CMOS.

Se analiza el comportamiento de cada uno de los dispositivos MOS, para luego analizar la operación de la compuerta de transmisión. Al analizar el comportamiento individual de los dispositivos, generalmente, no se habla de compuertas de transmisión, sino de transistores de paso.

### 2.5.1 Transistor de paso nMOS.

En la Fig. 2.41 se presenta el transistor de paso nMOS, que tiene como carga el condensador  $C_L$ . Para el análisis se supone que  $C_L$  está inicialmente descargado ( $V_o=0V$ ). El comportamiento del transistor de paso nMOS, para las diferentes condiciones es:

- a) Cuando la señal de control  $\phi=0$  se tiene que  $V_{gs}=0$ ,  $I_{ds}=0$ , y por lo tanto independientemente del valor que se tenga en  $V_i$ , el voltaje de salida  $V_o=0$ .

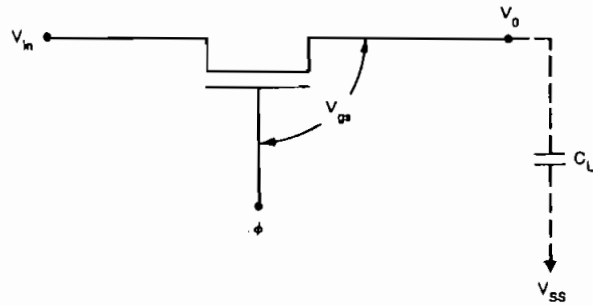


FIGURA 2.41 Transistor de paso nMOS.

- b) Cuando  $\phi=1$  ( $V_{gs}=V_{DD}$ ) y  $V_i=1$ , el transistor de paso empieza a conducir e inicia la carga del capacitor  $C_L$  hasta  $V_{DD}$ . Puesto que inicialmente  $V_i$  está inicialmente a un potencial superior a  $V_o$ , el sentido de la corriente es de izquierda a derecha. Cuando la salida alcanza el valor  $(V_{DD} - V_{tn})$ ,  $V_{gs} = \phi - V_o = V_{tn}$ , el transistor nMOS deja de conducir.

Cuando la señal  $\phi$  nuevamente regrese a 0, ( $V_{gs}$  es un valor negativo de voltaje) el capacitor  $C_L$  conserva la carga ya que no hay camino de circulación de corriente. Bajo estas condiciones el voltaje de salida permanece en el valor  $(V_{DD} - V_{tn})$ , lo que implica una degradación de un 1 lógico al pasar a través del transistor de paso nMOS.

- c) Cuando  $V_i=0$  y  $\phi=1$  el transistor de paso empieza a conducir y descarga al capacitor  $C_L$  hacia  $V_{ss}$ , es decir hasta tener que  $V_{gs}=V_{DD}$ . Puesto que al inicio  $V_i$  está a un potencial más bajo que  $V_o$ , la corriente fluye a través

del dispositivo de derecha a izquierda. Cuando el voltaje de salida se aproxima a 0V, como no existe capacitor de carga y  $V_1$  sigue siendo 0V, la corriente del dispositivo "n" disminuye, pues  $V_{ds}$  tiende a 0V. Así la transmisión de un 0 lógico no se degrada.

### 2.5.2 Transistor de paso pMOS

En la Fig. 2.42 se presenta el transistor de paso pMOS, que tiene como carga el condensador  $C_L$ . Se supone que  $C_L$  está inicialmente descargado ( $V_o=0V$ ) y se sigue un análisis similar al realizado para el caso del transistor de paso nMOS.

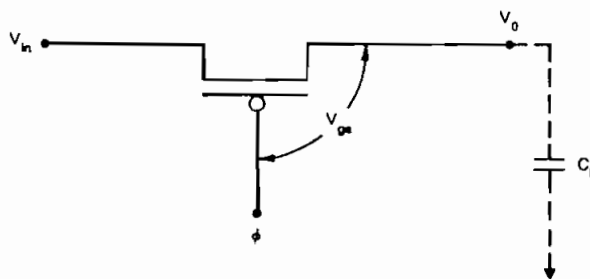


FIGURA 2.42 Transistor de paso pMOS.

- a) Con  $\phi=1$ ,  $V_o=0$  y  $V_1=1$ , el transistor pMOS no conduce y el capacitor  $C_L$  continúa descargado y la salida en 0V.
- b) Cuando  $\phi=0$ ,  $V_o=0$  y  $V_1=1$ , el voltaje  $V_{gs}=0$  y  $V_{ds}=V_{DD}$ . La corriente empieza a circular y a cargar el capacitor hasta  $V_{DD}$ .

c) Cuando en la entrada tengo 0V ( $V_i=0$ ) y  $V_o=1$ , se tiene un voltaje  $V_{gs}=-V_{DD}$ , nuevamente con un voltaje  $V_{gs}$  diferente de cero, el transistor conduce permitiendo que el capacitor  $C_L$  se descargue hasta que  $V_o=|V_{tp}|$ . En esta condición el transistor pMOS deja de conducir y la salida permanece en  $V_o=|V_{tp}|$ , lo que se traduce en una degradación del "0" al pasar por el transistor pMOS.

### 2.5.3 Compuerta de transmisión CMOS

En la Tabla 2.4 se resumen las características de operación de cada uno de los transistores de paso. El transistor de paso nMOS degrada el paso de un "1" y permite el paso óptimo de un "0"; por el contrario, el pMOS permite el paso óptimo del "1" y degrada el paso de un "0". Combinando las características de los dos dispositivos, se construye la compuerta de transmisión, que permite el paso óptimo tanto de un "0" como el de un "1".

Dispositivo	Transmisión de "1"	Transmisión de "0"
n	pobre	buena
p	buena	pobre

TABLA 2.6 Relaciones de voltaje para la operación del inversor MOS

Considerando que la compuerta de transmisión requiere la señal  $\phi$  y su complemento, el comportamiento de la compuerta de transmisión puede resumirse de la siguiente manera:

Para  $\phi=0$ :

Transistor "n" = OFF

Transistor "p" = OFF

$V_i=0$ ;  $V_o=Z$  ;  $Z =$  alta impedancia

$V_i=1$ ;  $V_o=Z$

Para  $\phi=1$ :

Transistor "n" = ON

Transistor "p" = ON

$V_i=0$ ;  $V_o=0$

$V_i=1$ ;  $V_o=1$

La característica de salida que presenta la resistencia de la compuerta de transmisión se presenta en la Fig. 2.43.

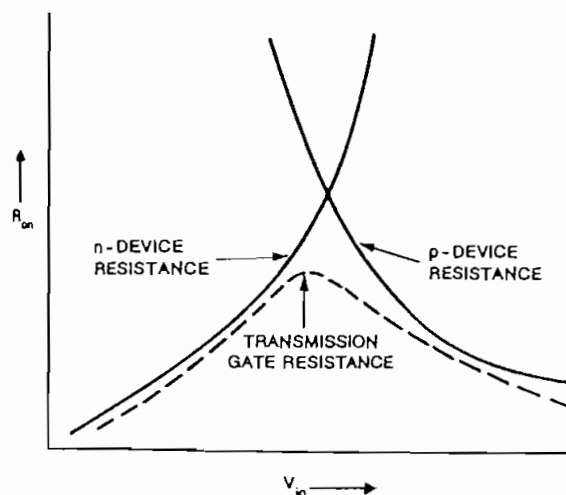


FIGURA 2.43 Característica de salida de la compuerta de transmisión.

## 2.6 EL PROBLEMA DEL EFECTO PARÁSITO DE *LATCH-UP*

Para entender el efecto parásito de *latch-up* es necesario conocer la estructura de un circuito CMOS, y establecer el origen de dicho efecto.

### 2.6.1 Estructura física del Inversor CMOS

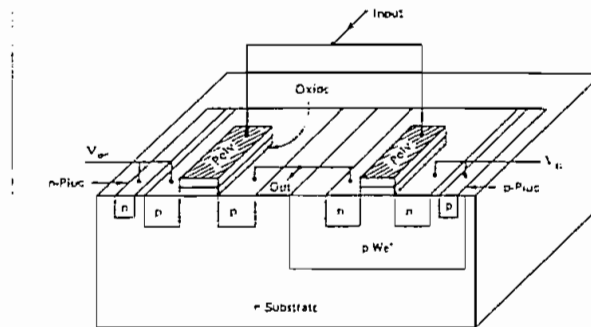


FIGURA 2.44 Estructura de un inversor CMOS.

Para la tecnología CMOS, se requieren construir sobre una misma base (sustrato) los dos tipos de transistores. Se toma como ejemplo la estructura física del Inversor CMOS, cuya representación física se presenta en la Fig. 2.44.

Para el caso presentado el sustrato de Silicio es de tipo "n". Para construir el transistor pMOS, solo se deben difundir dos regiones altamente dopadas de tipo "p" y construir los demás elementos de la compuerta.



Para estructurar el inversor, se deben realizar las conexiones tanto del sustrato (tipo "n") como de la fuente a  $V_{DD}$  y conectar el drenaje al terminal de salida. La señal de entrada está aplicada a la compuerta.

Para construir el transistor nMOS, se debe disponer de una estructura que realice las funciones del sustrato, para el ejemplo se crea por difusión, un pozo de material tipo "p", como se indica en la Fig. 2.44. El transistor nMOS, con todos sus elementos, se construye luego sobre la región del pozo-p. Para completar las conexiones del inversor se conectan el pozo-p (que hace de sustrato del transistor nMOS) y la fuente a Gnd. Se conecta el terminal de drenaje a la salida, y finalmente se conecta el terminal de la compuerta al de la compuerta del transistor pMOS.

Puede optarse también por construir los circuitos CMOS utilizando un sustrato original de tipo "p", en cuyo caso el pozo difundido sería de tipo "n".

En la Fig. 2.44 se aprecia que para las conexiones del sustrato tipo "n" y del pozo-p se utilizan plugs especiales "n" y "p" respectivamente. Esta estrategia de conexión se utiliza para disminuir la influencia de efectos parásitos que se hacen presentes en la tecnología CMOS. El efecto parásito que se tiende a eliminar con los plugs especiales es el denominado de *latch-up*, que se describe a continuación.

### 2.6.2 El latch-up

El *latch-up* (traba) es un efecto parásito que ha afectado a la tecnología CMOS desde sus orígenes, y más aún fue un factor fundamental para la no aceptación de los procesos de fabricación CMOS iniciales; sin embargo, en la actualidad, este efecto parásito se controla con innovadores procesos y técnicas resultantes de la investigación y adecuada comprensión del problema.

El resultado del efecto de *latch-up* es el de proporcionar un camino de corriente entre las líneas de  $V_{DD}$  y Gnd ( $V_{SS}$ ), lo que puede ocasionar la auto-destrucción del chip, o al menos la falla del sistema, provocando el retiro de la alimentación.

En la Fig. 2.45 se presenta la estructura de un inversor CMOS con sustrato tipo "n" (pozo-p). Se incluye también la representación de los transistores bipolares parásitos que se forman entre los distintos tipos de materiales. Se puede entender claramente la formación de los transistores bipolares "nnp" y "pnp", si en la representación de los mismos se sigue cada uno de los terminales: colector, emisor y base, y se observa el tipo de material semiconductor asociado a ellos. Se ha designado al transistor "pnp" como T1 y al "nnp" como T2; además, se puede observar la representación de dos resistencias  $R_s$  y  $R_w$ .

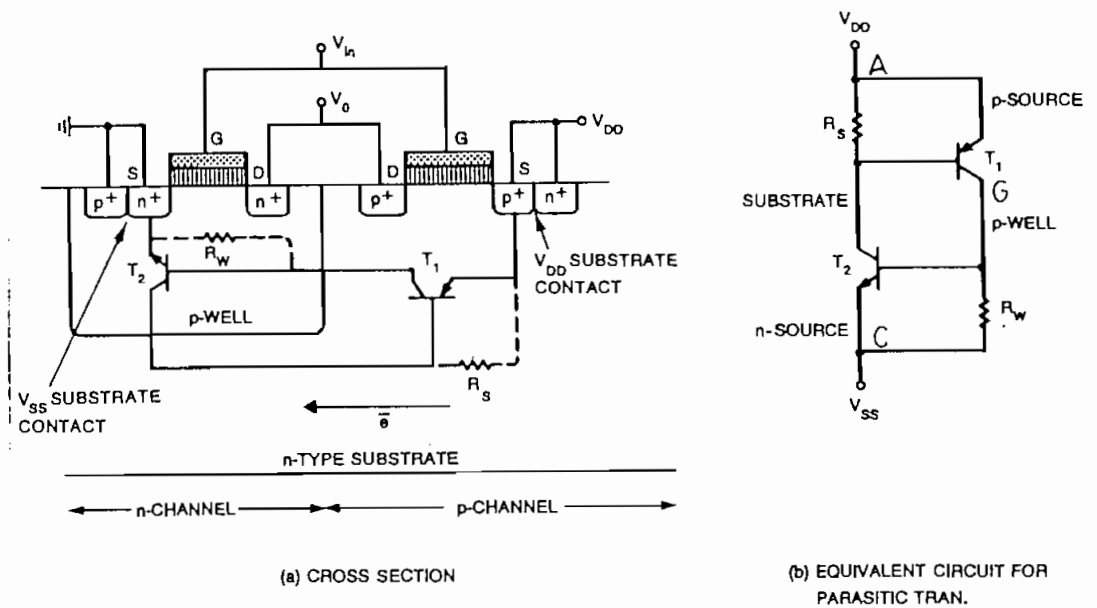


FIGURA 2.45 Origen del latch-up en CMOS.

La resistencia  $R_s$  es debida al sustrato (tipo "n" en este caso) y la resistencia  $R_w$  debida al pozo (tipo "p" en este caso). El valor de estas resistencias son determinantes en la susceptibilidad del circuito al efecto de latch-up, mientras más grandes son estas resistencias, más probable es que la estructura sea susceptible al latch-up. En la Fig. 2.45b se presenta el circuito equivalente de los transistores parásitos, incluidas las resistencias  $R_s$  y  $R_w$ .

Para estudiar el latch-up, se debe entender la operación básica de un SCR (*silicon-controlled rectifier*), dispositivo pnpn, con tres terminales (ánodo (A), cátodo (C) y compuerta (G)), como se indica en la Fig. 2.46. El circuito equivalente del SCR, utilizando transistores se muestra en la Fig. 2.47.

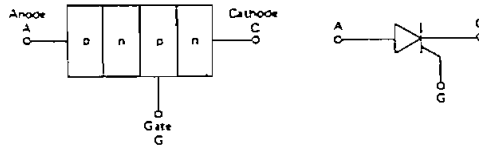


FIGURA 2.46 Representación de un SCR.

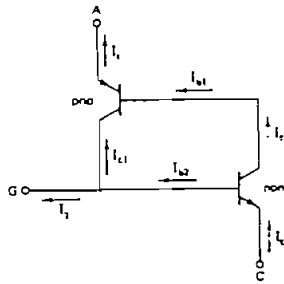


FIGURA 2.47 Circuito equivalente con transistores de un SCR.

Como puede observarse, el arreglo de los dos transistores coincide con el circuito equivalente presentado en la Fig. 2.45b. Por esta razón puede utilizarse el principio de funcionamiento del SCR para entender el efecto de *latch-up*.

En la Fig. 2.47 se incluye la representación de las corrientes de los dos transistores, siguiendo el sentido de los electrones. Si la corriente de la compuerta  $I_g$  crece, lo hacen también la corriente de base  $I_{b2}$  del transistor npn, y por lo tanto la corriente  $I_{c2}$ . Como  $I_{c2} = I_{b1}$ , la corriente  $I_{c1}$ , del transistor pnp, también aumenta, provocando un mayor incremento en  $I_{b2}$  del transistor npn ( $I_{b2} = I_{c1} + I_g$ ).

Si la ganancia de los dos transistores  $\alpha_1$  y  $\alpha_2$  son de magnitud tal que  $\alpha_1\alpha_2 \geq 1$ , la acción de la realimentación positiva hará que el SCR se "encienda" permanentemente, y la corriente incrementada resultante auto destruya el dispositivo. Bajo estas condiciones, el SCR está "disparado".

Un disparo del SCR puede ocurrir por un impulso proporcionado a la compuerta (del SCR), el voltaje en la resistencia del pozo "p",  $R_w$ , induce una corriente de realimentación positiva para el SCR. Esta condición puede originarse, por ejemplo en el instante de encendido, pues puede producirse una condición de "racing" entre las entradas y el voltaje de polarización, recibiendo un pulso el terminal G.

Algo similar ocurre si el terminal A sube a un voltaje superior a  $V_{DD}$ , ahora se induce una corriente por  $R_s$ , también de realimentación. Esta condición puede ocurrir cuando las entradas del circuito conmutan, o en la operación de conmutación dinámica del circuito CMOS, la cual demanda grandes corrientes en cortos periodos de tiempo, lo que puede producir caídas del voltaje de polarización, induciendo de esta manera un *latch-up*.

En cualquiera de las situaciones mencionadas, si la corriente inducida es  $I$  y el voltaje  $IR_s$  o  $IR_w$  es suficientemente alto para polarizar directamente la juntura base-emisor de alguno de los transistores parásitos, se produce el efecto de *latch-up*.

Para reducir las oportunidades en que se produzca este efecto se han planteado diversas soluciones. Una es reducir los valores de  $R_S$  y  $R_w$  en cuanto sea posible, reduciendo de esta manera la magnitud del posible voltaje de polarización de las juntas base-emisor de los transistores, para ello se utilizan los denominados **anillos de guarda** (*guard rings*), que se explicarán en el Cap. 3. Estos anillos de guarda son conexiones de baja resistividad hacia los voltajes de polarización, contruídos alrededor de los transistores canal "n" y canal "p". La colocación de estos elementos implican mayor espaciamiento entre transistores; es decir, mayor área y por ende menor densidad de integración.

La colocación de los plugs "p" y "n", en puntos cercanos a los puntos de polarización, tiene el mismo efecto que reducir  $R_w$  y  $R_S$ , respectivamente.

Otra alternativa es controlar el valor de las ganancias  $a_1$  y  $a_2$ .  $a_2$  está determinado por el proceso de fabricación, mientras que  $a_1$  puede reducirse incrementando la distancia entre el pozo-p y la difusión "p", incrementando de manera efectiva el ancho de la región de la base de los transistores pnp. En este caso, nuevamente se reducen los efectos del *latch-up*, sacrificando área y reduciendo por lo tanto la densidad de integración.

En la mayoría de los procesos actuales la posibilidad de que el *latch-up* se presente en la circuitería interna se ha

reducido al grado que el diseñador no necesita preocuparse de este efecto, simplemente colocando "convenientemente" los plugs de los sustratos. El término "convenientemente" se deriva más bien de la experiencia del diseñador o de quienes preparan las herramientas de diseño. Aunque los parámetros parásitos pueden ser evaluados, es difícil establecer el número de plugs necesarios.

Como se explica en el Cap. 3, el circuito integrado puede considerarse constituido por una circuitería interna y por las estructuras (PADs) que permiten realizar el interfaz entre la circuitería interna y el mundo exterior. Y justamente, el lugar más favorable para que ocurra el *latch-up* son las estructuras de entrada/salida, en donde fluyen corrientes altas y pueden ocurrir condiciones anormales de voltaje. Las posibles estrategias de prevención del *latch-up* y algunas reglas típicas que se siguen, se explican en el Cap. 3.

## 2.7 COMPARACION CON OTRAS TECNOLOGIAS

Se presenta a continuación un resumen comparativo de las principales características de la tecnología nMOS, de la tecnología bipolar y de la tecnología CMOS, de la cual ha tratado el desarrollo de este capítulo; sin embargo, para cumplir el objetivo propuesto, se utiliza una visión muy general de las características de las tecnologías nMOS y bipolar, y no se pretende profundizar en sus características.

### 2.7.1 MOSFETs vs. BJT

Haciendo una comparación de los MOSFETs, utilizados en tecnología nMOS y CMOS, con los transistores bipolares de juntura (*Bipolar Junction Transistor*, abreviado BJT), se puede señalar:

- a) La operación del MOSFET depende del flujo de portadores mayoritarios solamente, y es por lo tanto un dispositivo unipolar.
- b) El MOSFET es más simple de fabricar y ocupa menor área que un BJT. Un CI fabricado con tecnología bipolar utiliza un área considerablemente mayor que su correspondiente fabricado con tecnología MOS.
- c) A diferencia de un BJT, un transistor MOSFET es un dispositivo controlado por voltaje, y no por corriente como el BJT.
- d) En el MOSFET, el  $\text{SiO}_2$  es un material aislante y la corriente entre el canal y la compuerta es básicamente cero, lo que le permite al MOSFET presentar una alta impedancia de entrada, superior a la presentada por los BJTs.



- e) La alta impedancia de entrada y debido a que no se consume ninguna corriente de entrada, permite a un dispositivo MOSFET manejar a muchos dispositivos similares.
- f) En un MOSFET se pueden aplicar voltajes relativamente altos en la compuerta, en el orden de los 15V; por el contrario, en los CIs bipolares los voltajes están en el orden de los 5V.
- g) Debido a que la capa de  $\text{SiO}_2$  es extremadamente delgada, la aplicación de voltajes excesivos (superiores a los especificados para un determinado CI) pueden facilmente perforar esta capa de aislante.
- h) Uno de los problemas de los dispositivos MOS es su susceptibilidad a la carga estática. El desarrollo de carga estática en la manipulación de los elementos puede llegar a perforar el óxido de la compuerta, causando el daño permanente del dispositivo, por lo que para prevenir daños se incluye circuitería de protección.
- i) La simetría bilateral del MOSFET, permite que sea utilizado para cargar o descargar un nodo de un circuito, reduciendo en muchos casos el número de dispositivos necesarios para implementar determinadas funciones.
- j) Con tecnología MOS se obtiene menores consumos de potencia.

- k) Un MOSFET requiere de un número mucho menor de pasos de fabricación que los que requiere un proceso bipolar. Sin embargo, se debe señalar que en CMOS no es tan cierto.
- l) El MOSFET puede trabajar como dispositivo activo, o como una resistencia pasiva. Las resistencias de carga para componentes activos pueden ser por lo tanto implementados con dispositivos MOS, con un ahorro sustancial de área.
- m) En tecnología MOS, los resistores y capacitores pueden fabricarse fácilmente siguiendo la estructura metal-óxido-semiconductor.

### 2.7.2 CMOS vs. nMOS

Algunas de las características de la tecnología CMOS que se mencionan han sido ya presentadas en este capítulo, otras se presentan en el Cap. 3 y Cap. 4.

- a) En CMOS las salidas de sus circuitos entregan los niveles de voltaje correspondientes a  $V_{DD}$  y Gnd, sin degradación. En nMOS los niveles lógicos de voltaje se degradan.
- b) En CMOS pueden conseguirse tiempos de transición (tiempo de subida y tiempo de bajada) similares. En nMOS los tiempos de subida son mayores que los de bajada.

## CAPITULO 3

### PROCESOS DE FABRICACION CMOS

Un diseñador no necesariamente tiene que involucrarse con los procesos de fabricación; sin embargo, un conocimiento de dichos procesos le ayuda a comprender las bases físicas del interfaz entre el diseño y el proceso de fabricación. Además, para un diseñador involucrado en el campo de los CIs es indispensable absorber las innovaciones técnicas de los procesos de fabricación para comprender como afectan el comportamiento de los dispositivos en el CI, evitando convertirse en un simple autómeta que pasa sus diseños para que sean fabricados.

La industria Microelectrónica moderna es una de las disciplinas de ingeniería más sofisticadas y exactas que haya sido desarrollada.

La implantación de un CI involucra tres grandes pasos: el diseño, la preparación de máscaras y la fabricación. Para cada paso, se han desarrollado ambientes completos de herramientas.

Como se expuso en el Cap. 1, el nivel de *layout* constituye el enlace entre el mundo físico y el eléctrico; en otras palabras entre el circuito diseñado y su proceso de fabricación.

En el nivel de *layout* se describen de manera detallada y exacta todas las formas geométricas que deben ser construídas sobre el Silicio base (sustrato).

En la Fig. 3.1 se presenta una vista transversal de la estructura física de un transistor MOS.

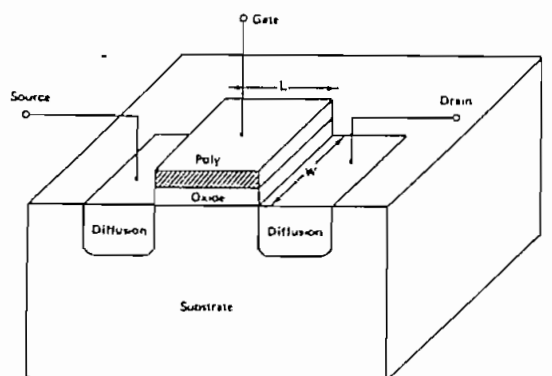


FIGURA 3.1 Estructura física de un Transistor MOS.

En la Fig. 3.2 se presenta la vista superior de la estructura del transistor de la Fig. 3.1, incluidas las dimensiones L y W. La vista superior incluye la representación de la capa de *polysilicon* y de la capa de difusión que determina los terminales de drenaje y fuente.

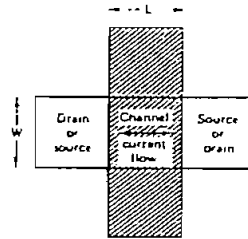


FIGURA 3.2 Vista superior de la estructura de un transistor MOS.

La región del canal y la capa de óxido de la compuerta están definidos implícitamente en el área de intersección de las dos capas. Son representaciones de este tipo las que el diseñador debe manipular.

Para tener una idea más clara, en la Fig. 3.3 se presenta un inversor CMOS, con su representación a nivel circuital, una vista transversal de su estructura física, y la forma de su layout.

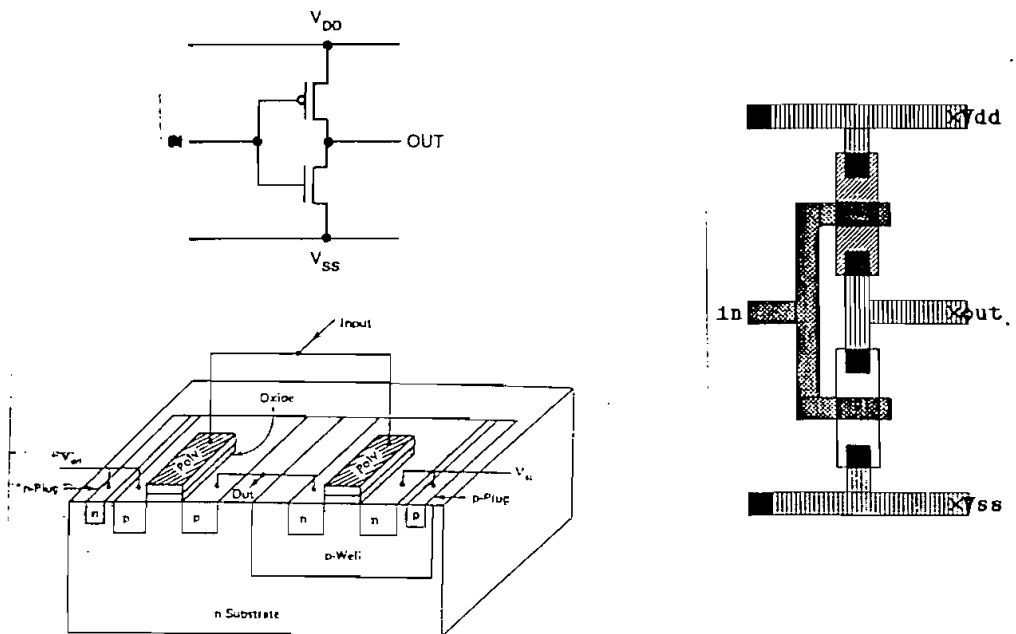


FIGURA 3.3 El inversor CMOS.

Como puede apreciarse en la Fig. 3.3c, los materiales utilizados se representan de diferente manera; así, los caminos de metal se representan rayados con líneas verticales, los contactos entre difusión y metal con pequeños rectángulos negros, el camino de difusión para el transistor "p" presenta un rayado denso de forma oblicua, el camino de difusión para el transistor "n" no tiene ningún rayado y finalmente para los caminos de *polysilicon* se utilizan puntos densos de color oscuro.

En el layout se distingue claramente la estructura de los dos transistores MOS, en los lugares en donde existe cruce de *polysilicon* y difusión. Se puede observar que su representación coincide con la de la Fig. 3.2, pero rotada 90 grados. El transistor-"p" está representado en la parte superior y el "n" en la parte inferior.

Para el caso del transistor "p" se conecta la fuente a  $V_{DD}$ , que corre en la parte superior de la Fig. 3.3c en metal, y el drenaje se conecta a un nuevo camino de metal que permite obtener la salida (Xout).

El transistor "n" tiene conectada su fuente a  $V_{SS}$  (Gnd), que corre en el camino de metal de la parte inferior, y su drenaje al camino de metal de la salida (Xout).

Para las 4 conexiones mencionadas se utiliza siempre un "contacto" para unir eléctricamente las difusiones con metal.

La señal de entrada, común a los dos transistores, es aplicada a las compuertas por medio de un único camino de *polysilicon*, por lo que no se requiere de ningún contacto.

En la Fig. 3.4 se presenta una forma alternativa del layout para el inversor CMOS, de las numerosas que un diseñador puede concebir para cumplir el mismo objetivo.

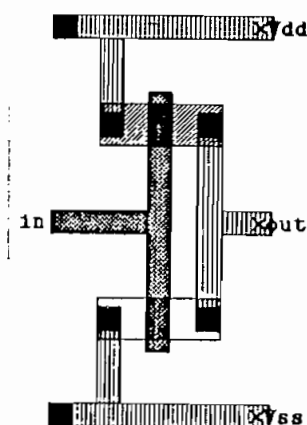


FIGURA 3.4 Forma alternativa del layout del inversor CMOS.

Se puede analizar el layout de la Fig. 3.4 y concluir que la estructura de los transistores y sus interconexiones son las mismas que para el layout de la Fig. 3.3c.

A pesar de la sencillez de los layouts de las Fig. 3.3 y 3.4, no debe pensarse que son las únicas capas de las cuales debe preocuparse el diseñador, como se demuestra en la sección 3.3, el proceso de fabricación exige un mayor número de capas que deben entregarse para la fabricación de su CI.

Para dibujar las formas geométricas de las distintas capas, el diseñador dispone de herramientas computacionales denominadas "editores de layout" (o llamados también editores de máscaras), que permiten el diseño de las capas a nivel individual y conjunto.

En la Fig. 3.5 se presentan separadamente las capas representadas en el layout del inversor de la Fig. 3.3c.

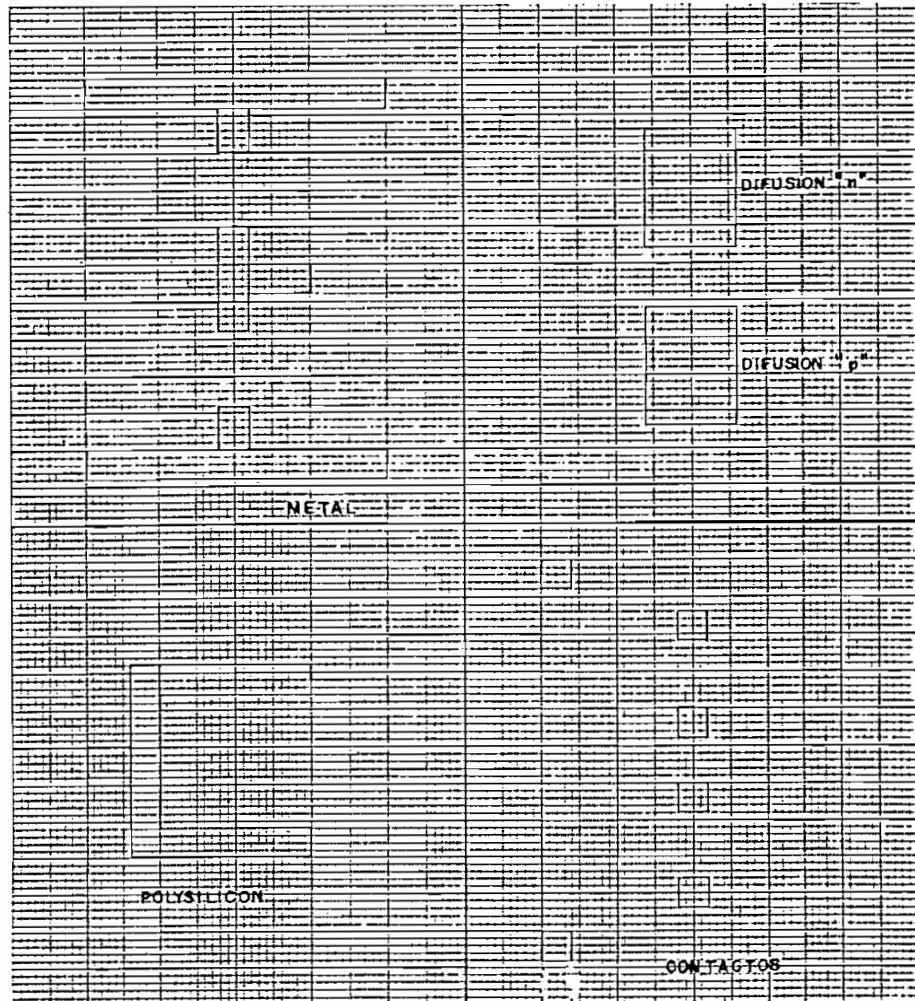


FIGURA 3.5; Capas separadas del layout del inversor CMOS.



Intuitivamente se pueden enumerar las posibles capas cuyos layouts serían necesarios para la fabricación de un CI con tecnología CMOS, conforme se avance en la descripción de los procesos de fabricación se comprueba si estos son todos. Las formas geométricas necesarias deberían incluir la definición para el material conductor (metal para caminos y contactos), materiales semiconductores (*polysilicon*, difusiones, pozos).

La descripción de las formas geométricas que se obtengan se lo hace utilizando lenguajes con el denominado formato intermedio (intermedio entre el diseñador y el proceso de fabricación). El formato utilizado generalmente es el lenguaje CIF (*Caltech Intermediate Form*), descrito en detalle en el Cap. 5. Este lenguaje es capaz de describir estructuras geométricas planas utilizando notaciones simples y claras para cada una de las capas. Archivos que contienen la descripción de los layouts, son el objetivo final del trabajo realizado por el diseñador.

En la Fig. 3.6 se esquematiza de manera general los pasos involucrados en el proceso de fabricación desde que se dispone de los archivos con formato intermedio, que contienen la información de un diseño y que se asume cumple todas las especificaciones requeridas para realizar su fabricación. La tarea básica del proceso de fabricación es llevar cada una de las formas geométricas de los layouts al Si, en los materiales especificados.

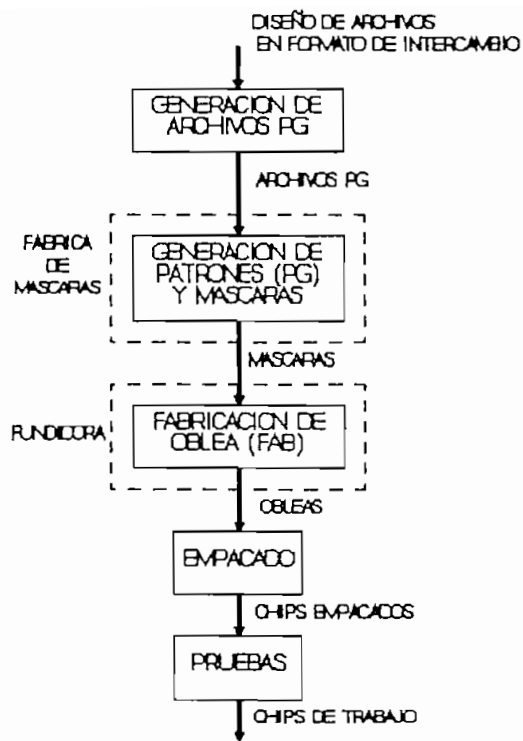


FIGURA 3.6 Proceso de fabricación de un CI.

El primer paso en este proceso es la construcción de un conjunto de máscaras, que contienen imágenes exactas de las formas de los layouts, en materiales transparentes u opacos.

La primera etapa de fabricación de las máscaras consiste en generar, a partir de los archivos de diseño, nuevos archivos que se entregarán a la fábrica de máscaras (*mask houses*) para controlar el mecanismo óptico utilizado para la fabricación de los patrones. Estos archivos se conocen con el nombre de "generadores de patrones" (*Pattern Generator, PG*).

En las fábricas de máscaras, mediante una secuencia de pasos de litografía, se construyen las máscaras, que luego son utilizadas para la fabricación misma en las fundidoras (*fab houses*).

En la Fig. 3.7 se presenta un diagrama simplificado de un mecanismo óptico que construye las máscaras. El equipo consta de una fuente de luz, de una sección móvil en los ejes (x,y) que permite desplazar los materiales utilizados y exponerlos a la fuente de luz. Todo el proceso es controlado por un sistema computacional.

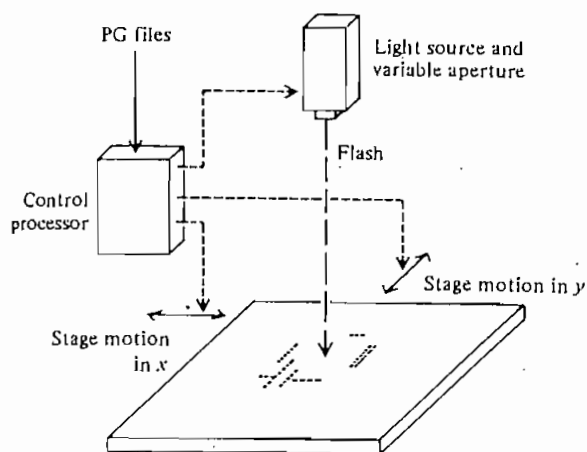


FIGURA 3.7 Generador de máscaras.

Generalmente, una máscara terminada (Fig. 3.8b) contiene al mismo sistema diseñado repetido cierto número de veces, número determinado por el tamaño del sistema diseñado y del Silicio disponible. La Fig. 3.8a es el elemento base de repetición (*reticle*) de la máscara terminada de la Fig. 3.8b.

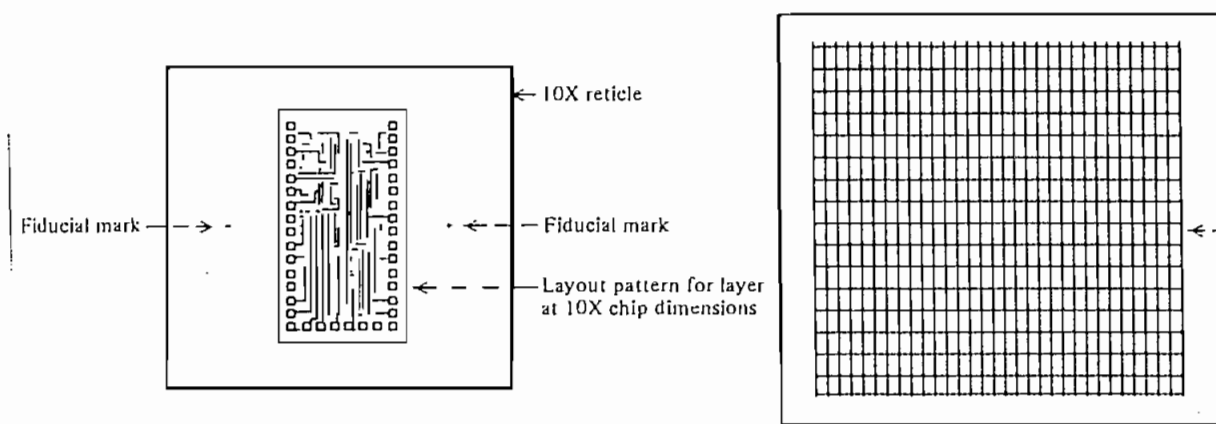


FIGURA 3.8 a) Elemento de repetición, b) Máscara concluida.

Una característica importante en la construcción de las máscaras es la inclusión en el "reticle" de un par de marcas (Fig. 3.8a), también fabricadas en el Si, y que permiten durante el proceso de fabricación el alineamiento de las máscaras (*fiducial mark*).

Las fundidoras son modernas y muy sofisticadas fábricas químicas que realizan las "impresiones" de las máscaras en el Silicio. A pesar de los controles precisos realizados durante la fabricación, siempre se debe estimar una tolerancia de desalineamiento.

El desalineamiento de las máscaras conduce a establecer un conjunto de reglas de diseño, determinado por el proceso de fabricación, para garantizar que los layouts se creen en el Si en los lugares esperados y de las dimensiones adecuadas. La naturaleza de estas reglas se explica en la sección 3.3.

Se procede ahora a dividir la estructura construída en fragmentos (chips), por las líneas de división (*scribe lines*). Finalmente se realiza el empaquetamiento de los chips, los cuales son sometidos a pruebas funcionales. Estas últimas dos etapas requieren tecnologías complejas. Los pasos descritos se esquematizan en la (Fig. 3.9).

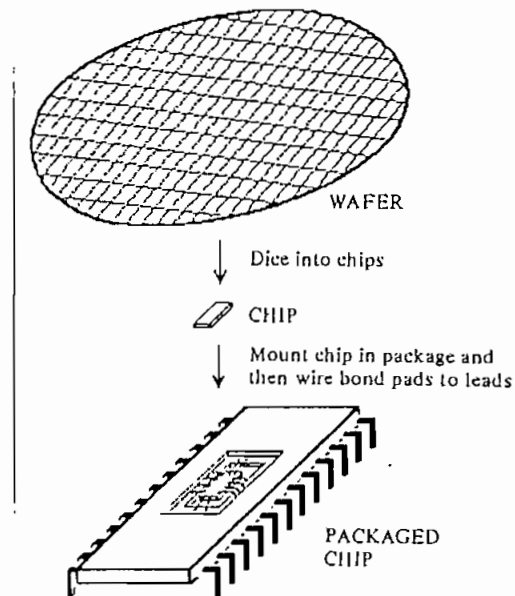


FIGURA 3.9 Secuencia del empaquetamiento.

Se debe señalar, de acuerdo a lo descrito, que las fabricas de máscaras y las fundidoras no requieren información detallada de los CIs que están fabricando; sin embargo, es necesario proporcionar información de carácter general, como por ejemplo el número de pines.

A continuación se describen técnicas de fabricación básicas que son comunes en muchos procesos, sean estos nMOS o CMOS.

### 3.1 TECNICAS BASICAS DE FABRICACION

#### 3.1.1 Fabricación de la oblea

La materia prima básica del proceso de fabricación es el Dióxido de Silicio, del cual se tienen reservas abundantes en el planeta. El  $\text{SiO}_2$  es purificado al grado de silicio policristalino, haciéndolo reaccionar con Carbono (C) y luego es cristalizado mediante procesos especiales.

La implantación del diseño en el Silicio se lo hace partiendo de una oblea o disco de Silicio (Fig. 3.10). Cada oblea tiene un diámetro comprendido entre 75 y 150 mm y de espesor inferior a 1 mm.

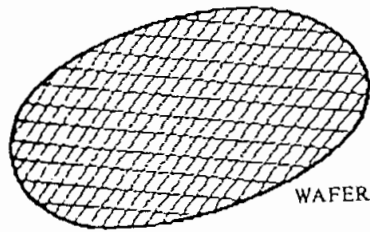


FIGURA 3.10 Oblea de Silicio.

Las obleas se cortan de lingotes de Silicio monocristalino que se obtiene a partir de Silicio policristalino puro fundido. El proceso empleado se denomina "Método de Czochralski" y actualmente es el método más utilizado para la producción de Silicio monocristalino.

La Fig. 3.11a y 3.11b esquematizan los elementos y procesos involucrados en la fabricación de los lingotes. La Fig. 3.11a es una buena aproximación al dispositivo real mientras que la Fig. 3.11b da una idea más clara del proceso.

El Silicio está contenido en un crisol (*crucible*) de cuarzo rodeado de un radiador de grafito, el cual es calentado mediante inducción de radio frecuencia, utilizando un arreglo de bobinas adecuadamente distribuido, que funde el Si. La temperatura se mantiene en un valor ligeramente superior a la de fusión del Si ( $1500^{\circ}\text{C}$ ). La atmósfera sobre el material fundido es Helio o Argón, o se hace al vacío para reducir la posibilidad de contaminación.

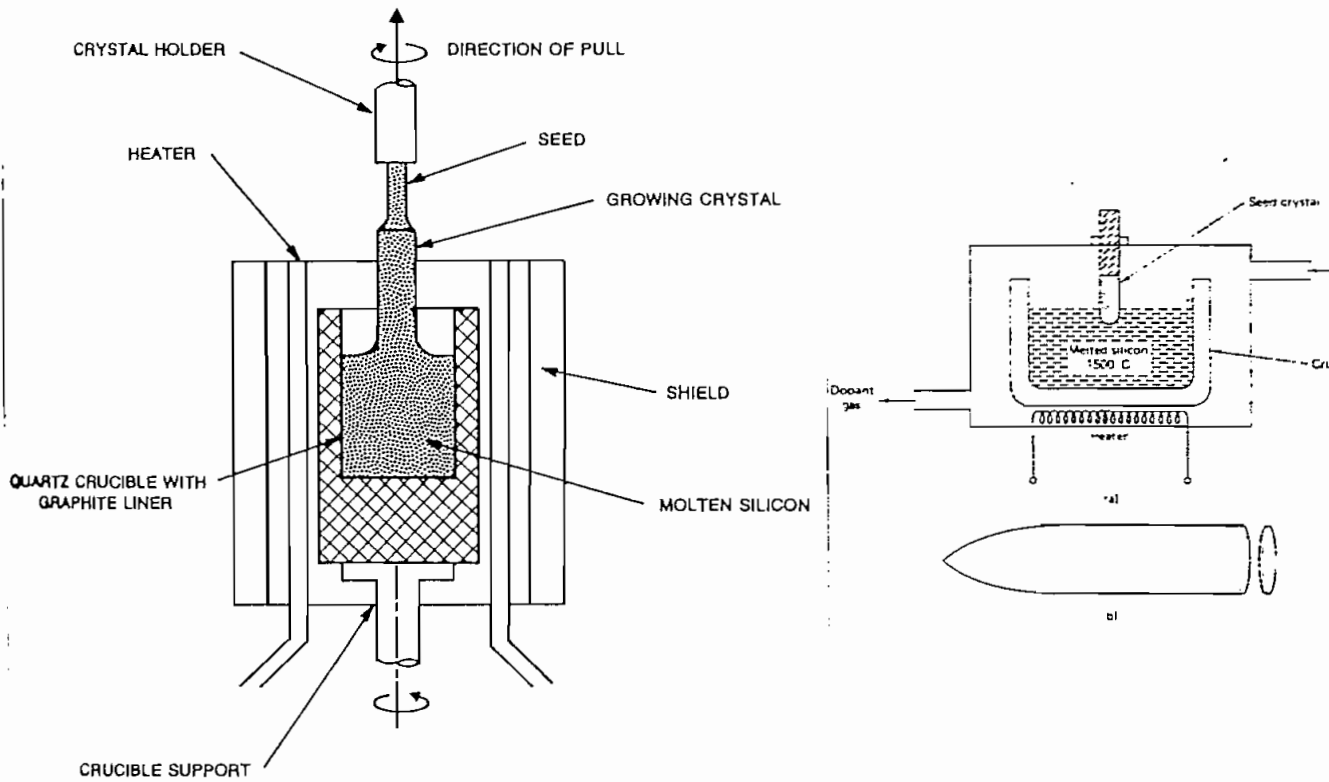


FIGURA 3.11 Proceso de fabricación de lingotes de Si.

Una "semilla" o guía de material monocristalino se introduce en el material fundido, para luego ser gradualmente extraída de manera vertical, mientras que al mismo tiempo se la hace girar, lo que permite que los átomos se adieran a la guía. Al ir extrayendo la guía, el material se enfría, y asumen la estructura monocristalina de la guía. El proceso continúa hasta que se termina el material fundido.

Como parte del proceso pueden agregarse cantidades controladas de material dopante para producir una concentración uniforme y así obtener materiales tipo "p" y "n".



En la Fig. 3.11b puede apreciarse la forma del lingote resultante. El diámetro del lingote está determinado por la velocidad con que se extrae la guía y la velocidad de rotación de la misma. Velocidades de crecimiento están en el rango de 30 a 180 mm/hora. Una vez fabricado el lingote, éste debe ser cortado para obtener las obleas finales; para ello se utilizan dispositivos con puntas de diamante, lo que finalmente entrega obleas con al menos un lado totalmente pulido, una superficie de espejo libre de rayaduras.

### 3.1.2 Oxidación

La oxidación consiste en obtener una fina capa de Dióxido de Silicio para posteriormente cumplir diferentes objetivos en el proceso de fabricación. Muchas de las estructuras y técnicas de fabricación de CIs utilizando Si aprovechan la propiedad del  $\text{SiO}_2$  para actuar como barrera frente a las impurezas durante la creación de los diferentes materiales de Si, por lo que la creación de éste es sumamente importante.

La oxidación del Silicio (Fig. 3.12) se consigue mediante calentamiento de las obleas en una atmósfera oxidante (oxígeno o vapor de agua). Los dos métodos más comunes son:

- a) Oxidación húmeda: si la atmósfera oxidante contiene vapor de agua. La temperatura está entre 900 y 1000 °C. Este proceso es sumamente rápido.

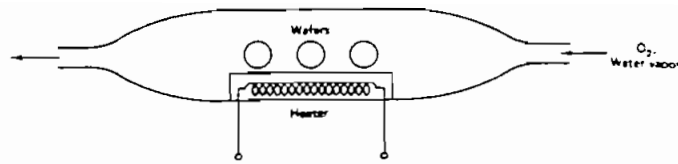


FIGURA 3.12 Proceso de oxidación.

- b) Oxidación seca: si la atmósfera oxidante es oxígeno puro. Para obtener resultados aceptables se debe mantener la temperatura alrededor de 1200°C. Se utiliza para la producción de capas de óxido más delgadas que las obtenidas con el proceso anterior. Las capas de óxido resultantes son delgadas pero robustas; sin embargo, requiere de más tiempo que el método húmedo.

### 3.1.3 Creación de patrones para difusión selectiva

Dado que un proceso de fabricación se resume a la creación de patrones para una secuencia particular de capas, los pasos para la creación de patrones, mediante los cuales las distintas formas geométricas especificadas por el diseñador se transfieren a las obleas de Si, son similares para cada una de las capas.

Un ejemplo claro para la creación de un patrón se presenta frente a la necesidad de construir Silicio de diferente tipo (para pozos, difusiones, etc), conteniendo proporciones

diferentes de impurezas donadoras o aceptoras, haciendo uso de las propiedades del  $\text{SiO}_2$  para actuar como una barrera frente a las impurezas, el proceso resultante se denomina "difusión selectiva".

El proceso consiste en trasladar con ayuda del  $\text{SiO}_2$  el patrón de una máscara al sustrato base. En otras palabras, los lugares en donde no esté presente el  $\text{SiO}_2$  permitirán el paso de las impurezas hacia la oblea, cambiando de esta manera las propiedades del Si base, y las áreas en donde se disponga de  $\text{SiO}_2$  estarán protegidas del efecto de las impurezas pues el  $\text{SiO}_2$  actuará como barrera frente a los átomos dopantes.

#### a) Creación del patrón.

El proceso se inicia con una oblea limpia (Fig. 3.13a). La oblea se somete a un proceso de oxidación, lo que da como resultado la forma de la Fig. 3.13b. Una vez enfriada la oblea, ésta es cubierta por una delgada capa de resina orgánica (Fig. 3.13c). La resina se somete a un proceso de "horneado" para asegurar su integridad.

En el instante de la fabricación, el patrón que se desea transferir a la oblea existe como una **máscara**. Una máscara, como ya se dijo anteriormente, no es sino una estructura transparente, que sirve de soporte de una fina capa de material opaco (Fig. 3.13d).

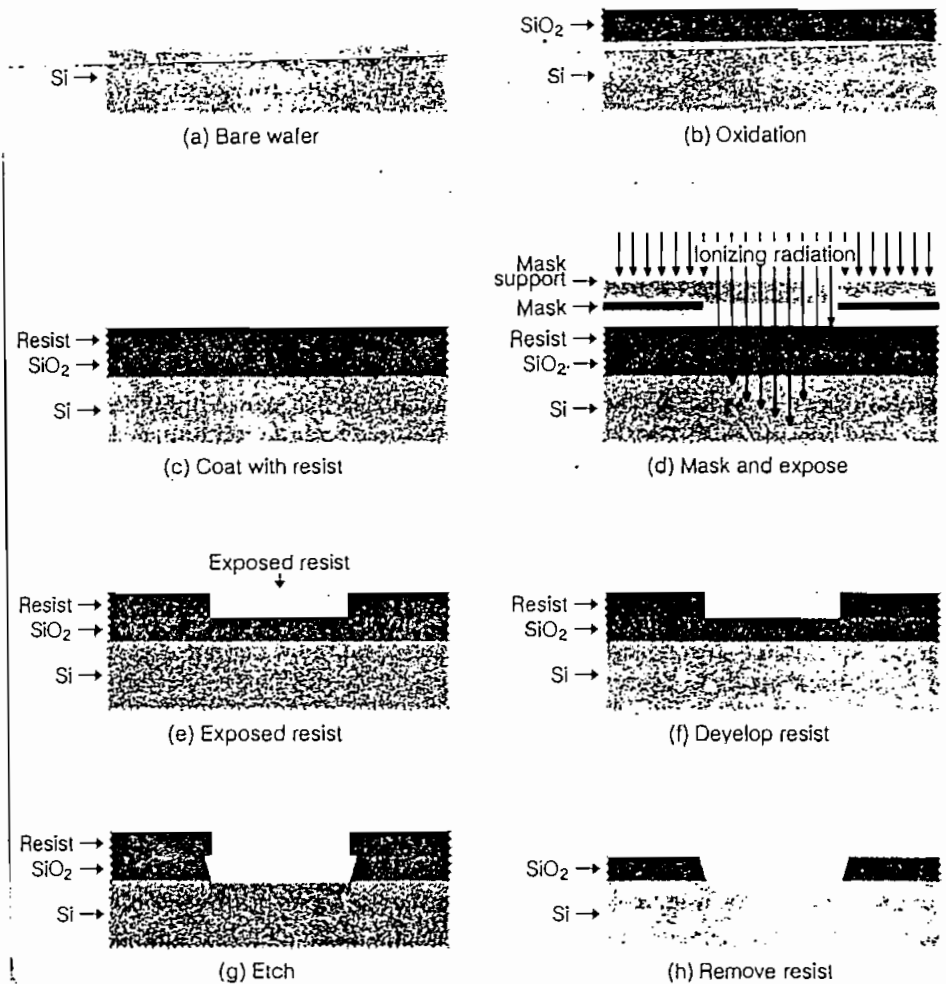


FIGURA 3.13 Creación de un patrón para difusión selectiva.

En la Fig. 3.13d se aprecia que ciertas secciones del material opaco de la máscara son removidas, dejando en la máscara el patrón preciso a ser transferido a la oblea. Las áreas oscuras del material opaco de la máscara coinciden con las áreas en las cuales se desea dejar al  $\text{SiO}_2$ , mientras que las áreas en las cuales se desea remover el  $\text{SiO}_2$  corresponden a las áreas en las cuales se desea remover el  $\text{SiO}_2$ .

Una imagen de la máscara se proyecta a la superficie de la oblea con una fuente intensa de radiación ionizante, que puede ser luz ultravioleta o rayos-X de baja energía, como se ilustra en la Fig. 3.13d. La radiación es detenida en los lugares en que existe material opaco, pero atraviesa la máscara en donde no está presente. La radiación pasa a través de la resina, el  $\text{SiO}_2$  y el Si de la oblea; sin embargo, tiene poco efecto en el  $\text{SiO}_2$  y el Si, pero rompe la estructura molecular de la resina en moléculas considerablemente más pequeñas.

Luego de haber sido expuesta a la radiación, la oblea se encuentra en el estado de la Fig. 3.13e. Las áreas de la resina expuestas a la radiación pueden disolverse sumergiendo la oblea en solventes orgánicos, que respetan a la resina no expuesta, dejándola intacta. (Fig. 3.13f).

Como resultado de todos los pasos realizados hasta el momento, se ha logrado transferir el patrón deseado a la capa de resina que se encuentra sobre la superficie del  $\text{SiO}_2$ . Ahora se debe transferir el patrón al  $\text{SiO}_2$ , para lo cual se expone la oblea a un "grabador" (*etchant*), que no es más que un gas reactivo ionizado que actuará sobre el  $\text{SiO}_2$ , pero no sobre la resina orgánica o el Silicio de la oblea. Este proceso generalmente se hace con Acido Fluorhídrico gaseoso. El resultado de este proceso de grabado se presenta en la Fig. 3.13g.

El paso final en la creación del patrón sobre el  $\text{SiO}_2$  es retirar la resina, para lo cual puede utilizarse diferentes sustancias: fuertes disolventes orgánicos o fuertes ácidos que atacan la materia orgánica. Una vez retirada la resina, el patrón deseado ya transferido a la oblea se presenta en la Fig. 3.13h.

En los procesos que utilizan Fotolitografía (el proceso que acaba de ser descrito utilizando la resina y luz ultravioleta), ocurre difracción cerca de los extremos de las máscaras y problemas de alineamiento de las mismas han limitado el ancho de las caminos a aproximadamente  $1.5$  o  $2 \mu\text{m}$ , sin poder disminuir más este valor. Sin embargo, durante los últimos años, (tomado de esragian 1988) la Litografía por haz de electrones (*Electron Beam Litography* EBL) ha surgido como una alternativa para la generación de patrones para conseguir buenas definiciones con anchos de hasta  $0.5 \mu\text{m}$ .

Las principales características de la generación de patrones con EBL son:

- i) Los patrones se derivan directamente de información digital.
- ii) Ya no existe la necesidad de preparar máscaras intermedias y todo el proceso involucrado para trasladar los patrones a las obleas.

iii) Cambios en los patrones pueden implementarse muy rápidamente.

iv) La desventaja de esta técnica es el alto costo de los equipos y el tiempo excesivo necesario para acceder a todos los puntos de la oblea.

#### b) Difusión.

El proceso de difusión permite la creación de las regiones "p" y "n" en los lugares no protegidos por el  $\text{SiO}_2$ . El primer paso de este proceso se denomina *pre-deposición*, el cual consiste en calentar a  $1000^\circ\text{C}$  la oblea en conjunto con átomos dopantes, como el Boro o Fósforo, mezclados con gas inerte, como el Nitrógeno. Los átomos se difunden en una delgada capa en la superficie del Silicio, formando una solución de gas y sólido. La concentración de impurezas de la capa se incrementa al incrementar la temperatura. La profundidad de penetración depende del tiempo que se lleva a cabo esta operación.

El siguiente paso consiste en calentar la oblea con gas Nitrógeno puro, lo que distribuye la difusión de la superficie hacia el interior del sustrato. La profundidad de penetración depende de la temperatura y del tiempo empleado. En las "fab houses" modernas la difusión se realiza con técnicas de *implantación de iones*, lo que provee un mayor control del proceso de difusión.

### c) Implantación de iones.

En este proceso, el gas dopante atraviesa un ionizador, Los iones resultantes son acelerados por un par de electrodos, con un potencial de 150 KV. Posteriormente los iones pasan por un fuerte campo magnético que permite agrupar los iones dopantes en un haz, con el cual se impacta a la oblea. Los iones chocan con la superficie del Si a alta velocidad y son colocados en el Si a una profundidad determinada por el campo acelerador y la concentración de los iones dopantes. Finalmente se realiza un proceso para una redistribución de los iones e incrementa la profundidad de penetración. Como barreras de los iones pueden utilizarse resinas, capas gruesas de óxido o metal. Si la capa de óxido de la barrera es delgada, puede hacerse la implantación a través de la capa.

La implantación de iones puede también utilizarse para construir los pozos "n" y "p" y para ajustes del voltaje umbral.

#### 3.1.4 Deposición

Una técnica muy utilizada dentro de los procesos de fabricación es la "deposición", que puede ser de  $\text{SiO}_2$ , *poly-silicon* o metal, sobre la oblea. El proceso de deposición se realiza a altas temperaturas dentro de una cámara, utilizando técnicas químicas de deposición de vapor.



Para depositar  $\text{SiO}_2$ , una mezcla de Nitrógeno, Oxígeno y Silano (compuesto de Si e Hidrógeno) se introducen en la cámara a temperaturas entre los 300 y 500 °C. El Silano reacciona con el Oxígeno para producir el  $\text{SiO}_2$ , el cual se deposita sobre la oblea. Debe notarse la diferencia con el proceso de oxidación en el cual el Si del sustrato reacciona con el oxígeno y a temperaturas mucho más altas.

El *polysilicon* se deposita simplemente por calentamiento del Silano a 1000 °C, lo que permite que se deposite el Silicio, y se recoja el Hidrógeno como gas.

El Aluminio se deposita vaporizando Aluminio de un filamento calentado dentro de un alto vacío.

### 3.1.5 Proceso para la compuerta de Silicio (silicon gate)

Como se mencionó en el Cap. 2, actualmente puede construirse la compuerta de *polysilicon*, pudiéndose además utilizar este material para realizar interconexiones dentro del CI (como en el inversor de la Fig. 3.3).

Una de las principales ventajas de utilizar *polysilicon* para construir la compuerta de los MOSFETs, es la posibilidad de utilizarlo como una máscara, la cual permite una definición bastante precisa de las zonas de la fuente y drenaje. En otras palabras, se consigue tener la zona de la compuerta

con un mínimo solapamiento con las difusiones del drenaje y fuente, lo que permite mejorar las condiciones de operación de los dispositivos fabricados.

Para la fabricación de las compuertas de los transistores MOS, se deposita *polysilicon* puro sobre el óxido de la compuerta. Posteriormente el *polysilicon* y las regiones de la fuente/drenaje son dopadas simultáneamente. El dopaje del *polysilicon* se hace ya que el *polysilicon* puro tiene una resistividad muy alta.

Los pasos necesarios en un proceso típico de compuerta de Si incluye el proceso para la creación de un patrón sobre el  $\text{SiO}_2$ , cuyo resultado se presenta en la Fig. 3.14a. La gruesa capa de  $\text{SiO}_2$  se denomina **óxido de campo** y no está presente en el lugar destinado para construir un transistor.

Luego, toda la superficie de la estructura de la Fig. 3.14a se cubre con una capa de  $\text{SiO}_2$ , de espesor cuidadosamente controlado (Fig. 3.14b). El área de Silicio expuesta, entra en contacto con esta fina capa de  $\text{SiO}_2$ , denominada **óxido de la compuerta** (o *thin oxide* o *thinox*).

El siguiente paso es depositar *polysilicon* sobre la superficie de la oblea, para luego dejar luego solamente el patrón con las interconexiones y compuertas necesarias. El resultado del proceso para el caso de la compuerta del transistor nMOS se presenta en la Fig. 3.14c.

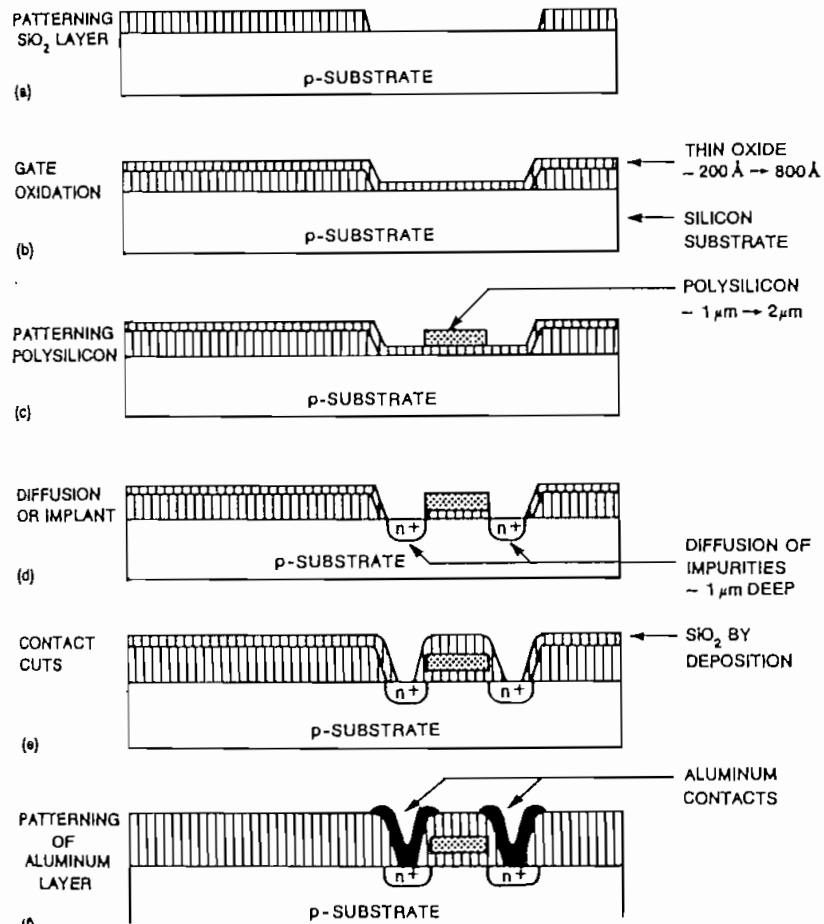


FIGURA 3.14 Pasos para la fabricación de un transistor nMOS con compuerta de Si.

Las áreas de thinox no cubiertas por el *polysilicon* son luego retiradas (Fig. 3.14d).

A continuación la oblea completa debe ser sometida a un proceso de dopaje. Con esto se consigue formar las zonas del drenaje y fuente y dopar al *polysilicon*. (Fig. 3.14d).

Finalmente toda la estructura se cubre con  $\text{SiO}_2$  y se repite el proceso de grabado para dejar el patrón de los

contactos que permiten la conexión con las capas inferiores de difusión (Fig. 3.14e). Para las interconexiones se utiliza Aluminio, mediante deposición (Fig. 3.14f).

Como puede observarse en la Fig. 3.14d, el proceso permite que las zonas de difusión se formen solamente en las regiones en las cuales el *polysilicon* no cubre el sustrato, razón por la cual se denomina al proceso auto-alineado (*self-aligned*). En los procesos de fabricación iniciales, muchas veces las zonas de difusión se extendía por debajo de la compuerta, causando sobrelapamiento que altera la estructura y características del transistor.

### 3.2 TECNOLOGIAS DE FABRICACION CMOS

El haber descrito detalladamente la secuencia de pasos necesarios para crear un patrón permite tener una visión global del proceso completo de fabricación. Ahora se puede analizar una secuencia particular de patrones para determinado proceso, evitando la repetición de los detalles de la creación de cada patrón para cada capa, el principio básico se repite varias veces cambiando unicamente los materiales involucrados.

Para la fabricación de CIs CMOS existen actualmente cuatro tecnologías dominantes, que son:

1. Proceso de pozo-p (p-well process).
2. Proceso pozo-n (n-well process).
3. Proceso de doble pozo (twin tub process).
4. Silicio sobre aislante (silicon on insulator).

Seguidamente se describe la secuencia de pasos requeridos en el proceso de pozo-p. Los procesos de pozo-n y de doble pozo son muy similares al de pozo-p. Se da una rápida explicación del proceso de Silicio sobre aislante.

### 3.2.1 Proceso de pozo-p (p-well process)

En este proceso se utiliza un sustrato tipo "n" moderadamente dopado (oblea), luego se crea el pozo-p para construir los transistores canal "n", los transistores canal "p" se construyen en el sustrato tipo "n".

De este proceso existen variaciones que vuelven un tanto complejos a los procesos. La descripción que se expone a continuación es una simplificación de todos los pasos requeridos en un proceso CMOS típico de pozo-p.

En la Fig. 3.15 se ilustra la secuencia de pasos necesarios en un proceso típico pozo-p. La figura incluyen para cada paso la vista transversal de la oblea y la máscara correspondiente con el patrón deseado a ser trasladado. Se numeran las máscaras para saber el número total necesario.

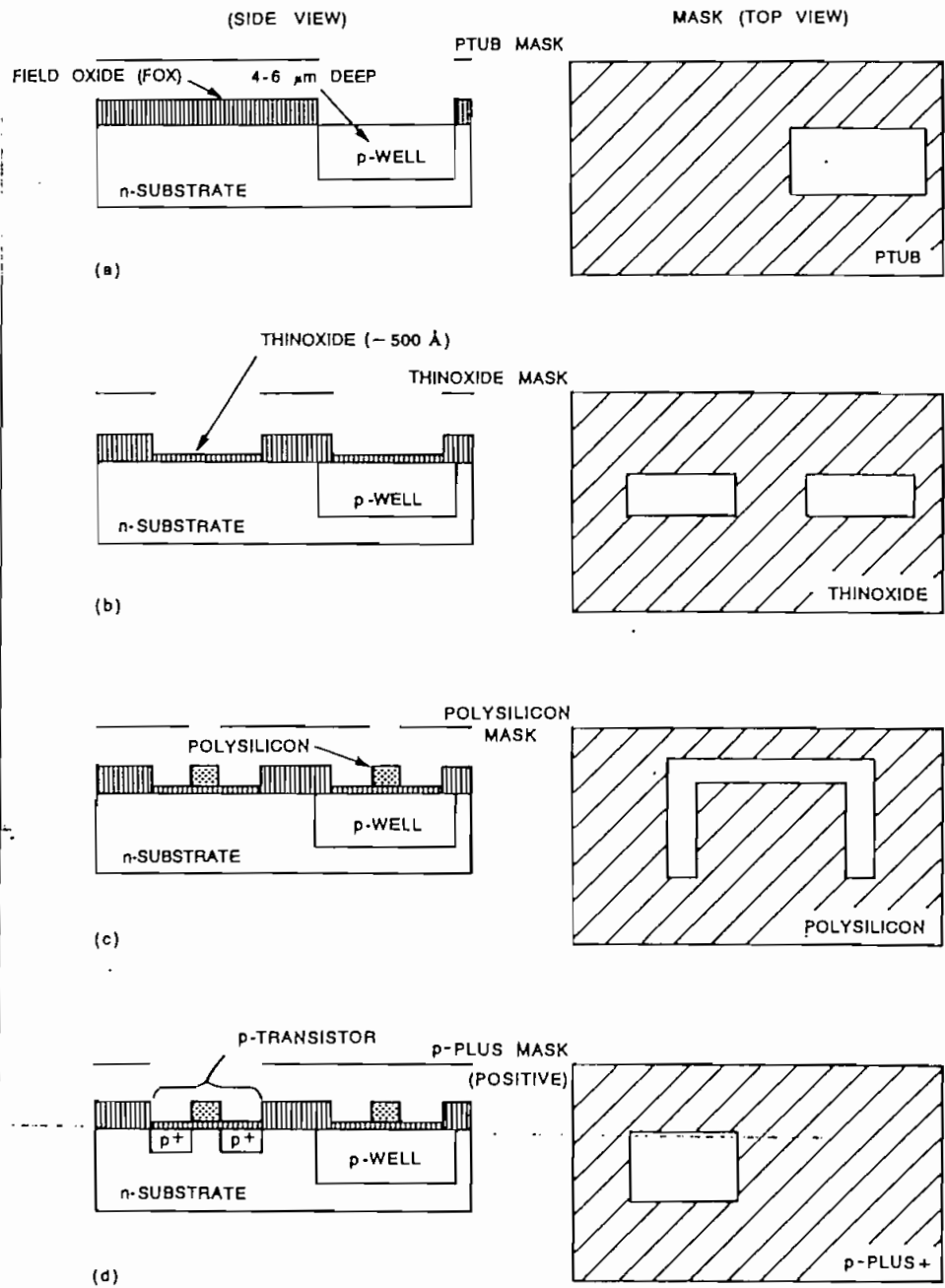


FIGURA 3.15 Pasos típicos de un proceso CMOS de pozo-p.

La primera máscara utilizada define el área correspondiente al pozo-p (*p-well* o *p-tub* (1)), sobre el se fabrica un transistor canal "n" (Fig. 3.15a).

La siguiente máscara se denomina *thinoxide* o *thinox* y comprende dos áreas, una para el transistor "n" (área de menor dimensión que la del pozo) y otra para el "p". Esta máscara define las áreas en donde se requiere el  $\text{SiO}_2$  para formar las compuertas de los transistores (Fig. 3.15b). Es común referirse a esta máscara con el nombre de área activa (2).

Se debe ahora, de acuerdo al proceso de compuerta de Si, cubrir la superficie con *polysilicon* (3) y dejar grabado únicamente el patrón deseado, (una U invertida Fig. 3.15c).

La siguiente máscara a utilizar es la denominada p-plus (*p+*), que indica aquellas áreas bajo el *thinox* (sumamente delgado) que van a ser implantadas para formar el material *p+* (Fig. 3.15d). También el área de *polysilicon* de la compuerta es sometida a la implantación. Si el patrón de la máscara permite que la implantación se haga también sobre el pozo-p, entonces el área *p+* construida permite disponer de un contacto óhmico. En este tipo de contacto no se forma la juntura rectificadora, a pesar de que los materiales "p" y "n" están en contacto. La corriente puede circular en ambos sentidos. El contacto óhmico se utiliza como plug de conexión del pozo y es una medida para evitar el efecto parásito de *latch-up*. A la máscara *p+* se le denomina también "implante *p+* (4)".

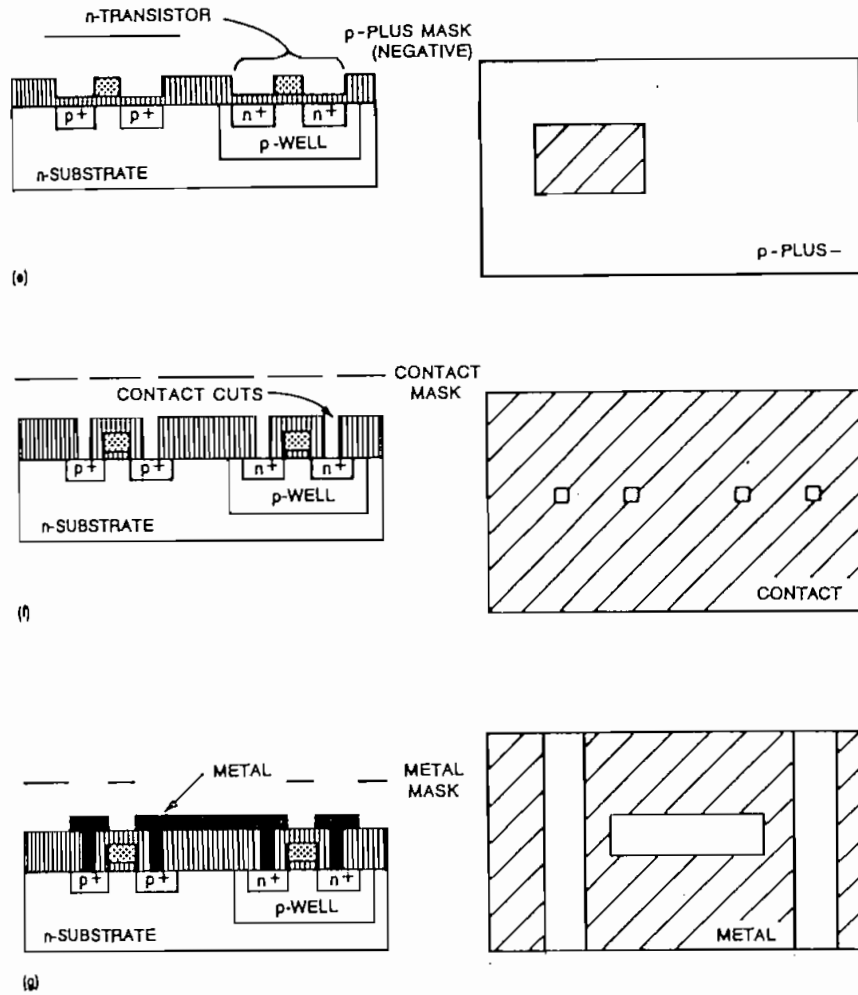


FIGURA 3.15 CONTINUACION Pasos típicos de un proceso CMOS de pozo-p.

Utilizando el complemento de la máscara anterior se definen las regiones para difusión del tipo n+. La ausencia de una región p-plus sobre una región de thinox indica que esa área va a ser una difusión n+. A esta máscara se le conoce también con el nombre de "implante n+ (5)".

Como en el caso de la máscara anterior, se puede incluir contactos óhmicos n+ para utilizarse como plugs del sustrato.



Algunos pasos adicionales pueden requerirse si se desean ajustes de los valores del voltaje umbral  $V_t$ . Influyen en el valor de  $V_t$  de los transistores n la concentración del pozo-p y la profundidad de penetración. Para conseguir valores bajos (0,6-1V) se requiere ya sea gran penetración de la difusión del pozo o gran resistividad del mismo. Sin embargo, una resistividad demasiado alta del pozo puede acentuar el problema del *latch-up*.

Valores precisos de  $V_t$  se consiguen en los procesos pozo-p, haciendo que la concentración del pozo sea aproximadamente un orden de magnitud más alta que la del sustrato, con la contraparte que se acentúa el efecto de cuerpo (*body effect*) para los transistores "n". Además, debido a esta alta concentración los transistores nMOS tienen una excesiva capacitancia de drenaje/fuente a pozo-p, haciéndolos más lentos.

Por las características mencionadas, los transistores nMOS son de inferiores características que aquellos que se construyen directamente sobre un sustrato tipo "p" (sin pozo) y los circuitos que los utilicen serán de menor velocidad que circuitos fabricados con tecnología nMOS.

La convención para la representación gráfica de los *layouts*, que se utiliza en algunos de los gráficos que se presentan a continuación, se resume en la Fig. 3.16.

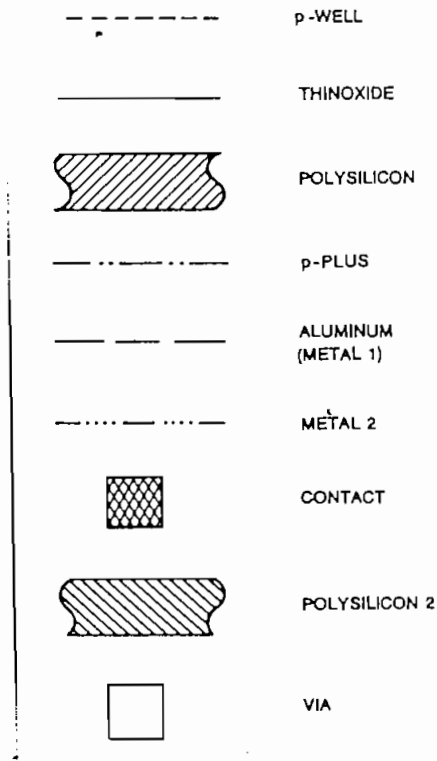


FIGURA 3.16 Convención para la representación gráfica de los layouts.

La estructura de la Fig. 3.15g, resultado del proceso descrito, se presenta en la Fig. 3.17c y corresponde a la fabricación de un inversor (Fig. 3.17a).

En la Fig. 3.17b se presenta el layout completo de la estructura, con la indicación del lugar en donde se realiza el corte que proporciona la vista de la Fig. 3.17c y utilizando la convención de la Fig. 3.16. En la Fig. 3.17d se presenta la forma de la estructura final fabricada, ajustada un poco más a la realidad, pues en este caso las formas no están definidas por líneas rectas.

Los pasos típicos de la fabricación para el proceso de pozo-n son similares a los ya explicados para el de pozo-p, excepto que en este proceso se utiliza pozo-n. El proceso puede resumirse de la siguiente manera:

El proceso se inicia sobre un sustrato tipo "p". Las primeras máscaras definen las regiones para el pozo "n". A continuación se procede al dopaje de éstas regiones con la concentración adecuada, formando de esta manera el pozo "n". El siguiente paso es definir las compuertas de los dispositivos y realizar las difusiones necesarias. Finalmente se crean los contactos y se procede a la metalización. Las máscaras utilizadas tendrán similar estructura a las descritas para el proceso pozo-p.

En el caso del proceso en estudio, problemas como el efecto de cuerpo y grandes valores de capacitancia hacen que los transistores "p" no sean óptimos; algo similar a lo que ocurre en el proceso pozo-p, en el cual fueron los transistores canal "n" los de relativamente baja calidad.

### 3.2.3 Proceso de doble pozo (twin tub process).

Este proceso permite optimizar independientemente los transistores "n" y "p", tomando acciones totalmente independientes para ajustar el voltaje umbral, el problema del efecto de cuerpo, etc.

Generalmente, el material inicial es un sustrato tipo p+ o n+, cubierto por una capa de Silicio monocristalino, de espesor controlado y de muy alta pureza, esta delgada capa es luego dopada uniformemente. El objetivo de esta capa es el de constituirse un medio de protección contra el *latch-up*. La secuencia del proceso de fabricación es muy similar al proceso pozo-p, con la diferencia que se deben formar dos pozos.

#### 3.2.4 Silicio sobre aislante (silicon on insulator SOI).

El proceso se inicia sobre una base de aislante, como por ejemplo el safiro, y luego se crea una delgada capa de Silicio monocristalino. Utilizando diferentes procesos con máscaras y técnicas de dopaje, se crean los transistores. A diferencia de los procesos convencionales CMOS, esta tecnología no incluye los pasos requeridos para la formación de los pozos, pues estos se crean sobre la estructura de safiro.

El proceso SOI tiene algunas ventajas determinantes sobre los procesos tradicionales de fabricación CMOS, entre ellas:

- a) Permiten una mayor densidad de integración.
- b) No presentan el problema de *latch-up*.
- c) Bajas capacidades parásitas.
- d) No se presenta el efecto de cuerpo.
- e) Mayor resistencia a la radiación.

Sin embargo, presenta también algunas desventajas, entre ellas:

- a) El CI es más difícil de proteger.
- b) Las estructuras de entrada/salida son de mayor tamaño.
- c) Los sustratos de safiro son considerablemente más caros que los de Silicio.
- d) Los estudios de técnicas de fabricación están orientados hacia los de sustrato de Silicio.
- e) De las técnicas de fabricación actuales, la SOI es la más costosa.

### 3.2.5 Incremento de la capacidad de enrutamiento

Uno de los principales problemas del diseñador es disponer de la libertad de trazar los caminos, sin incremento exagerado del área utilizada, para realizar el enrutamiento de las señales dentro del CI. Se han desarrollado algunas técnicas con la finalidad principal de incrementar la capacidad de enrutamiento dentro de los CIs. Las soluciones planteadas requieren añadir una o algunas capas de materiales, de los ya utilizados, a los procesos descritos anteriormente. Para cumplir el objetivo planteado se utilizan:

- a) Doble nivel de metal.
- b) Doble nivel de *polysilicon*.
- c) Combinaciones de las dos anteriores.

La primera solución para incrementar los caminos de enrutamiento es utilizar otra capa de metal, lo que implica en el proceso de fabricación la utilización de una nueva máscara.

Como regla general, la segunda capa de metal tiene un patrón más grueso que el de la primera capa (Fig. 3.18). La interconexión entre el primer nivel de metal (metal 1 (7)) y el segundo (metal 2 (9)) se realiza por medio de lo que se denomina "vía", como se indica en la Fig. 3.18.

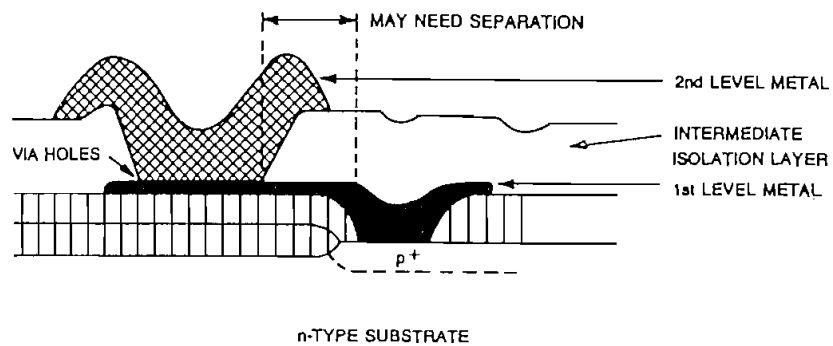


FIGURA 3.18 Vista transversal de un proceso que incluye metal 2.

Una vía servirá entonces para crear la máscara que determine las áreas en las cuales la metalización pone en contacto metal 1 y metal 2. El nombre "vía (10)" se asigna también como nombre a la máscara y se lo utiliza para diferenciarlo de la interconexión entre metal 1 y *polysilicon* o difusión que se denominó "contacto".

En el caso de que simultáneamente se requiera el uso de una vía y un contacto, puede ser necesario establecer un

espaciamiento entre la vía y el contacto (Fig. 3.18) o incluso se los puede sobreponer, dependiendo de los requerimientos de la casa fundidora.

En la Fig. 3.19 se presentan tres posibles configuraciones geométricas para una vía.

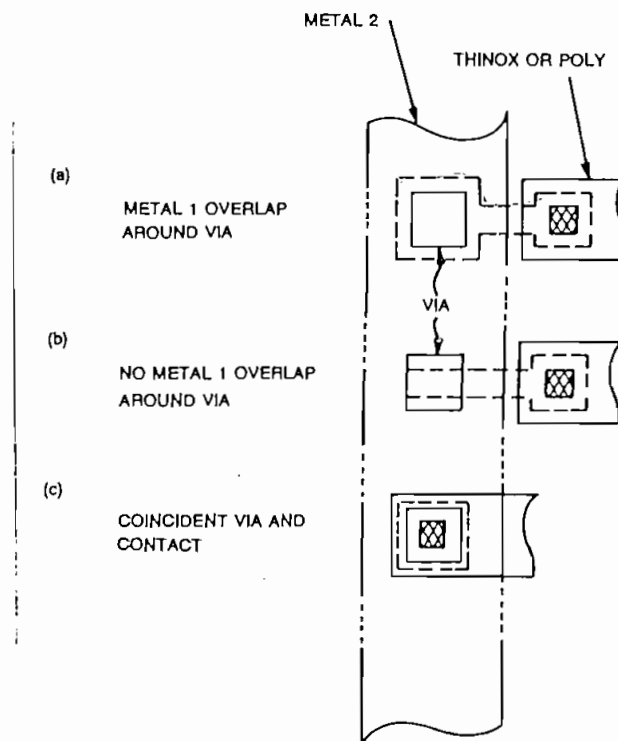


FIGURA 3.19 Configuraciones posibles de una vía.

En la Fig. 3.19a se observa que el metal 1 debe cubrir la vía y aún más debe extenderse más allá de ella (sobrelapándola). En la Fig. 3.19b la vía debe sobrepasar los límites del metal 1. En los dos casos (a y b), el metal 2 sobrelapa a la vía.

Finalmente en la Fig. 3.19c se presenta el caso en el cual coinciden una vía y un contacto, y no se requiere espaciado entre ellos. En este caso el metal 1 y metal 2 sobrelapan a la vía y al contacto. Las relaciones de tamaño entre vía y contacto son evidentes, como lo son también en la Fig. 3.18.

Las ideas generales expresadas en la descripción dada, son también válidas para cuando el proceso incluye un segundo nivel de *polysilicon*.

Un proceso completo de fabricación con doble nivel de metal requiere las máscaras para:

1. Fozo-n ó pozo-p
2. Area activa (*thinox*)
3. *polysilicon* (*poly*)
4. p-plus+ (implante p+)
5. p-plus- (implante n+)
6. Contactos
7. Metal 1
8. Overglass (pasivación)
9. Metal 2
10. Vía



### 3.2.6 Protecciones contra el latch-up

#### a) Contactos del pozo y sustrato a $V_{DD}$ y $V_{SS}$

La colocación de los plugs "p+" y "n+" para sujetar el sustrato y el pozo a los puntos de polarización, tiene el mismo efecto que reducir  $R_w$  y  $R_s$ , como se indicó en el Cap.2, reduciendo de esta manera la posible ocurrencia del efecto parásito de *latch-up*. En el proceso pozo-p, el sustrato "n" debe conectarse a  $V_{DD}$  por medio de un contacto óhmico. Por el contrario, el pozo-p debe conectarse a Gnd, también con un contacto óhmico. En la Fig. 3.20 se presenta la estructura física y el layout correspondiente, para cuando se incluyen las estructuras descritas.

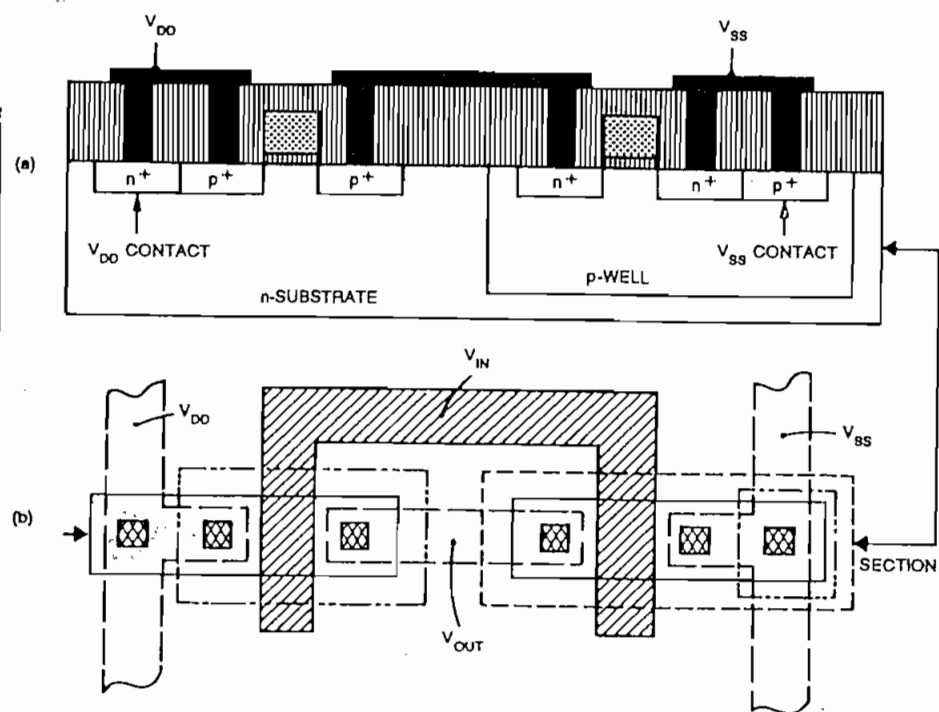
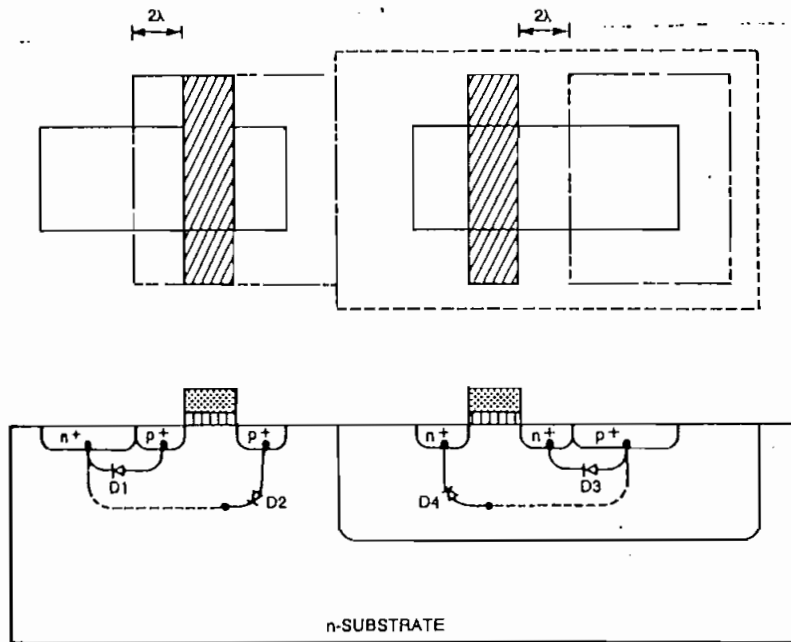


FIGURA 3.20 Contactos óhmicos formados con las difusiones  $n^+$  (sustrato) y  $p^+$  (pozo).

La Fig. 3.21 presenta la utilización de las máscaras para la construcción de los contactos del sustrato y pozo. Como se observa en la figura, los dos tipos de transistores requieren de regiones de *thinox* y *polysilicon*, para el transistor "n" las mencionadas regiones están rodeadas por la máscara del pozo "p", y para el transistor "p" están rodeadas por la máscara implante p+.

Las regiones de *thinox* que no están rodeadas por la máscara implante p+, son del tipo n+ y son las difusiones que formarán los transistores "n", dentro del pozo "p". Por lo tanto, un transistor será tipo "p" si está dentro de la máscara implante p+, de lo contrario será tipo "n".



DIODE BREAKDOWN VOLTAGES

$BV_{D1}$	20 - 40 VOLTS	$BV_{D3}$	20 - 40 VOLTS
$BV_{D2}$	50 - 70 VOLTS	$BV_{D4}$	50 - 70 VOLTS

FIGURA 3.21 Construcción de los contactos óhmicos n+ (sustrato) y p+ (pozo).

Observando la Fig. 3.21 se desprende que hay dos tipos de combinación de máscaras implante/difusión utilizadas para formar los transistores "n" y "p", que se obtienen de la siguiente manera:

- i) Las difusiones  $p^+$  se obtienen realizando el AND lógico de las máscaras implante  $p^+$  y  $thinox$ , que a la vez permite la formación de las regiones  $p^+$  dentro del pozo.
  - ii) Las difusiones  $n^+$  se derivan del AND lógico entre las máscaras  $thinox$  y el complemento o negación de la máscara implante  $p^+$ , que a la vez permite la formación de las regiones  $n^+$  dentro del sustrato.
- b) Anillos de guarda (*guard rings*)

Los anillos de guarda son conexiones de baja resistividad hacia los voltajes de polarización, contruidos alrededor de los transistores canal "n" y canal "p", que permiten reducir la resistencia del pozo y sustrato y además polarizar inversamente las junturas  $p-n$  que forman el SCR (Capítulo 2), con que se presenta el efecto de *latch-up*.

Para el caso de un proceso de fabricación de pozo-p, los anillos de guarda son difusiones " $n^+$ " en el sustrato "n", y difusiones " $p^+$ " en el pozo "p", y se encargan de recoger los portadores minoritarios.

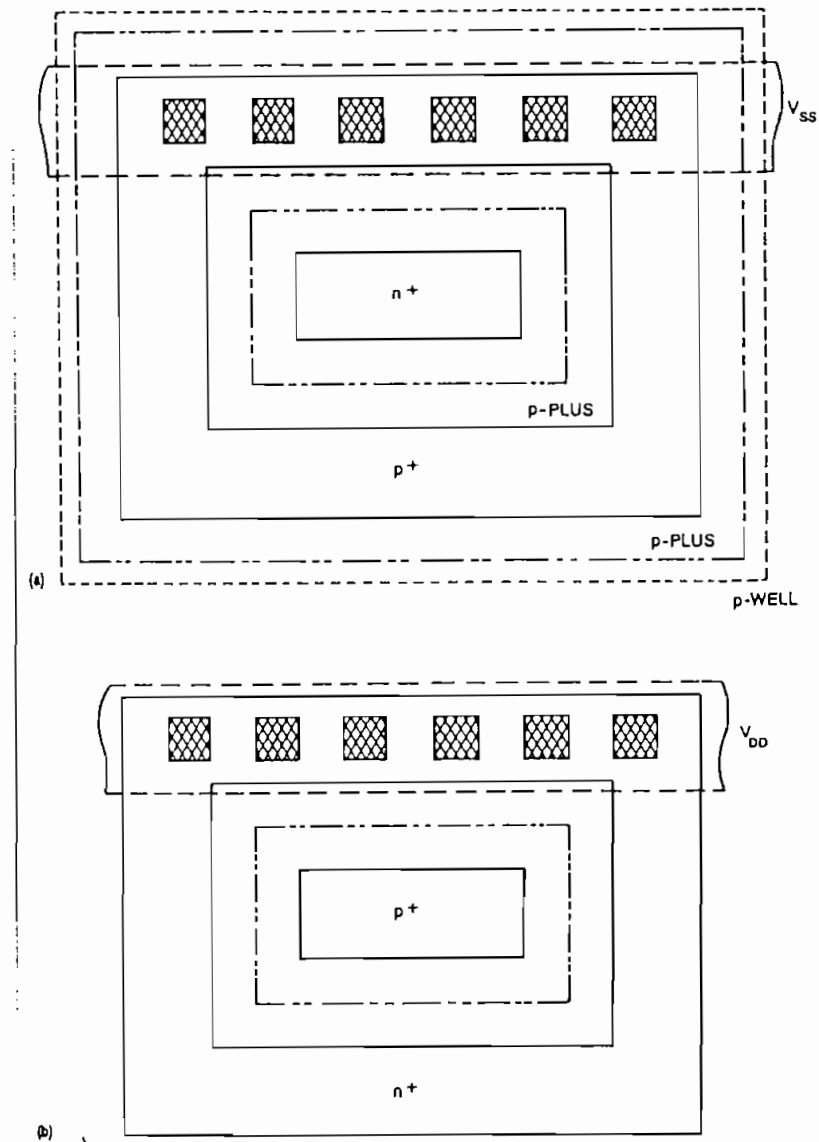


FIGURA 3.22 Anillos de guarda a) dentro del pozo "p" b) en el sustrato.

En la Fig. 3.22 se presentan las máscaras necesarias para la construcción de las difusiones de los anillos de guarda para un transistor "n" y un transistor "p". Para la construcción del anillo de guarda dentro del pozo "p" se requiere de una máscara implante p+, que definirá la forma y áreas donde se construye la región de difusión p+ del anillo.

Entre los límites de la máscara implante  $p^+$  se construyen las zonas de difusión  $p^+$  que rodean completamente a la difusión  $n^+$ , tomando la forma de un verdadero "anillo".

Existe entre la difusión  $n^+$ , que formará un transistor "n", y el anillo de guarda " $p^+$ " una región del material original del pozo (material "p"), que separa estas dos regiones. El anillo de guarda construido debe ser sujetado a Gnd, de esta manera el diodo que se forma entre las regiones " $n^+$ " y " $p$ " tiene una amplia región para polarizarlo inversamente. Por las dimensiones del anillo se realiza la unión al metal utilizando varios contactos.

Para el transistor " $p$ ", el anillo de guarda se construye en el sustrato, con material " $n^+$ ", también rodeando completamente a la difusión " $p^+$ ". Se utilizan también varios contactos para sujetar el anillo a  $V_{DD}$ , la juntura formada entre los materiales " $p^+$ " y " $n$ " es polarizada inversamente a través del anillo.

### 3.3 INTERFAZ DISEÑO - FABRICACION

Durante la fabricación existen factores que ocasionan que las formas transferidas a la oblea no sean exactamente las mismas que tenían los layouts preparados por el diseñador. Estos factores pueden ser: el desalineamiento de las máscaras, variaciones en los extremos de los patrones en la

resina, variaciones en las dimensiones de las áreas requeridas durante el proceso de grabado, esparcimiento de las zonas de difusión causando sobrelapamientos con las compuertas de los transistores, pequeñas variaciones en las dimensiones de los patrones realizados con  $\text{SiO}_2$ , distorsión de la oblea debido a procesos a altas temperaturas o tiempos de exposición de la resina inferiores o superiores a los necesarios.

Existe también la probabilidad que durante la operación del circuito se produzcan alteraciones de las dimensiones de ciertos materiales. Por ejemplo, puede ocurrir el fenómeno denominado "migración del metal" si la densidad de corriente excede ciertos límites en un camino de metal. Este fenómeno provoca que los átomos del metal se muevan en la dirección del flujo de corriente, dejando atrás caminos más delgados en ciertas áreas, que inclusive pueden llegar a "volar" como fusibles, debido a la excesiva densidad de corriente. Para prevenir este fenómeno, se debe garantizar un ancho mínimo del camino de metal, de acuerdo a los requerimientos de la densidad de corriente.

Todos los problemas mencionados pueden acumularse, produciendo fallas catastróficas de los circuitos. Como ya se ha mencionado, actualmente el diseñador opera a un nivel cuya meta final es entregar las descripciones de sus layouts y no se preocupa de los detalles de parámetros muchas veces complejos dependientes de los procesos de fabricación. Para conseguir este objetivo, se han desarrollado un conjunto de

reglas que el diseñador debe cumplir durante la preparación de sus layouts. Estas reglas, reducidas a su forma más simple, se conocen como "reglas de diseño o de layout".

Las reglas de diseño no son sino un conjunto de limitantes en el diseño de los layouts de tal manera que los patrones resultantes sobre la oblea conserven la topología y geometría de los diseños. Las reglas de diseño no son un límite estricto entre una correcta o incorrecta fabricación, más bien proporcionan cierto margen de tolerancia y ocurrencia de ciertas variaciones en el proceso de fabricación, garantizando que subsecuentemente el circuito funcionará adecuadamente. Puede haber ocasiones que un layout viole las reglas de diseño y aún funcione correctamente y viceversa. Sin embargo, cualquier violación de las reglas puede llevar con gran probabilidad a errores y al fracaso de un diseño.

Un conjunto de reglas de diseño debe también garantizar que los parámetros eléctricos considerados en el diseño no sean alterados por variaciones de los procesos de fabricación. Los valores de resistencias y capacitancias de los caminos son determinados por sus dimensiones físicas (como largo y ancho) y alteraciones en sus valores podrían llegar a causar degradación en la operación del circuito.

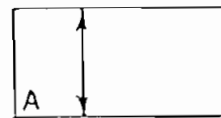
En conclusión, las reglas de diseño proveen un medio de comunicación directo entre el diseñador del circuito y el ingeniero involucrado en el proceso de fabricación.

ellos, en cada capa. Ejemplos claros de los problemas que contempla este conjunto de reglas son: casos en los que si el ancho de los caminos se hace demasiado pequeño en los layouts, en la oblea fabricada los caminos pueden resultar discontinuos o pueden ocurrir efectos como el de la migración del metal. Si se especifican dos caminos demasiado cercanos, es posible que resulten contruados uno sobre otro, provocando cortocircuitos.

- b) Un conjunto de reglas que prevee posibles interacciones entre diferentes capas. En este caso se consideran restricciones en cuanto a sobrelapamientos y extensiones.

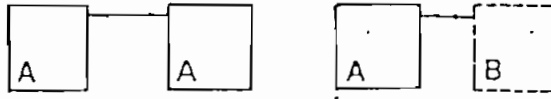
A continuación se resumen las representaciones gráficas generalmente utilizadas para exponer las reglas de diseño. Se introducen también algunos términos y su significado.

- a) Ancho: distancia mínima entre los extremos interiores de una figura.

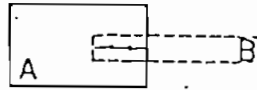


- b) Separación: distancia mínima entre los extremos de dos figuras (en la misma capa o en diferentes capas), para preveer interacción entre ellas.

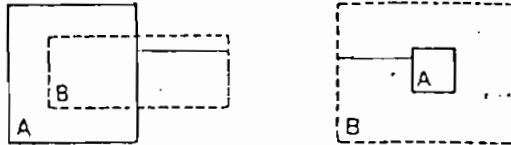




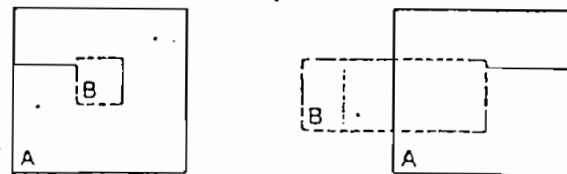
- c) Coincidencia: distancia mínima común a dos figuras en diferentes capas.



- d) Sobrelapamiento: distancia mínima que la figura de una capa (B) debe extenderse por fuera de la figura en otra capa (A).



- e) Margen: distancia mínima que la figura en una capa (B) debe estar interna en la figura de otra capa (A).



Hay varias alternativas a considerar para describir las reglas de diseño.

- a) Las reglas "micron", que son el estilo normal en la industria, se presentan como una lista de dimensiones mínimas y espaciamientos para todas las máscaras requeridas en cierto proceso. Así por ejemplo, el ancho mínimo del *thinox* puede expresarse como 4  $\mu\text{m}$ .
- b) Las reglas  $\alpha$  y  $\beta$ . En estas reglas, la dimensión básica se define en términos de  $\beta$ , mientras que el tamaño mínimo de la grilla que se requiere se describe con  $\alpha$ .  $\alpha$  y  $\beta$  se relacionan con un factor constante.
- c) Las reglas basadas en el parámetro *lambda*. Estas reglas fueron popularizadas por Mead y Conway.

### 3.3.1 Reglas de diseño CMOS

Se presentan a continuación un conjunto representativo de reglas de diseño para tecnología CMOS de pozo-p. Estas reglas son el resultado de haber realizado un promedio de un gran número de procesos <sup>(1)</sup>. Las reglas se presentan utilizando el parámetro *lambda*.

---

<sup>1</sup> "Principles of CMOS VLSI Design", Weste R. - Eshragian K., págs. 103 a 106.

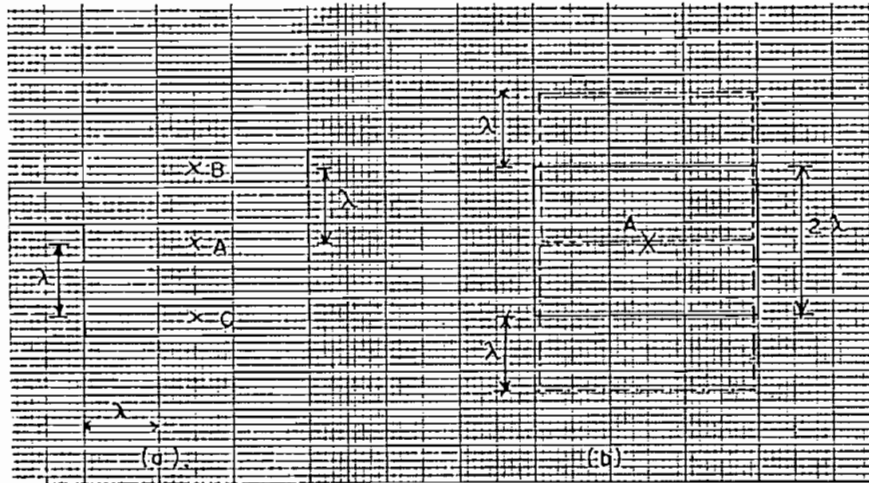


FIGURA 3.23 Ancho mínimo de un camino.

Si el camino fuese de un ancho mínimo de  $2\lambda$ , se puede asegurar que siempre por el punto  $A$  pasará el camino deseado (Fig. 3.23b). De lo expuesto se desprende que geométricamente el ancho mínimo de cualquier camino debe ser  $2\lambda$ .

En la Fig. 3.24 se presenta el caso en que se desea tener dos figuras rectangulares, ya sea en una misma capa o en diferentes. Si el proceso de fabricación determina que la figura de la izquierda se desplace un valor  $\lambda$  a la derecha, y lo mismo ocurra para la otra, pero desplazándose hacia la izquierda, se forma un único rectángulo, si los dos caminos están en la misma capa, y se sobrelapan causando **posiblemente** efectos que comprometan el funcionamiento del circuito o incluso catastróficos, si las figuras están en capas diferentes.

Si se establece como valor de separación  $2\lambda$  seguramente no se producirá ninguno de los efectos mencionados y el circuito funcionará adecuadamente. Sin embargo, este valor es considerado en ocasiones, de acuerdo a las capas involucradas y a los posibles efectos de los solapamientos, como demasiado conservador.

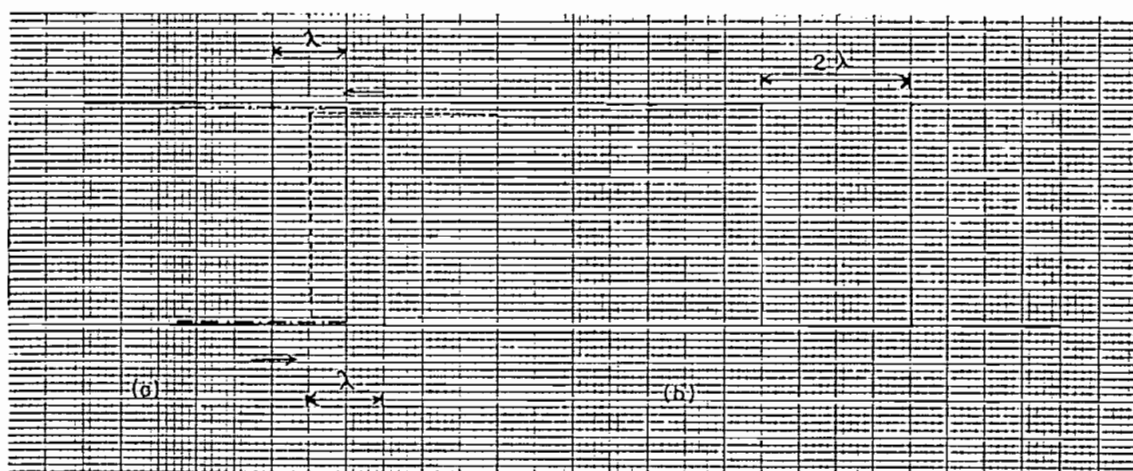


FIGURA 3.24 Separación mínima entre dos regiones.

En las Figs. 3.25 y 3.27 se presentan el conjunto de reglas para el proceso de fabricación de pozo-p. Cuando sea sencillo, se presenta la justificación de la necesidad de determinada regla. En las figuras se señalan el nombre de la máscara utilizada para definir las reglas, se asignan nombres a las reglas y se define el significado de cada una. Para entender el significado de cada regla se deben revisar la nomenclatura asumida para la representación de las capas de la Fig. 3.16.

a) Máscara 1

Para la máscara *thinox*, que define las regiones de difusión ( $n+$  o  $p+$ ) de los transistores, se señalan las reglas:

- A1. Un valor típico para el ancho mínimo de un camino en difusión es  $2 \lambda$ .
- A2. El espaciamiento requerido entre dos regiones de difusión, no es función solamente de la resolución geométrica del proceso, sino también del funcionamiento de los dispositivos formados.

Si dos regiones de difusión corren demasiado próximas, ya en operación las regiones de agotamiento asociadas a cada una de las junturas formadas, pueden sobrelaparse y producir una circulación de corriente no prevista entre las dos regiones. En procesos típicos, se utiliza como separación entre dos regiones de difusión  $3 \lambda$ , un valor menos conservador es  $2 \lambda$ .

- A3. La separación mínima entre una difusión  $p+$  ubicada en el sustrato y una difusión  $n+$  ubicada en el pozo es  $8 \lambda$ .

b) Máscara 2

Para la máscara del pozo-p, se señalan las siguientes reglas:

- B1. El ancho mínimo del pozo es 4  $\lambda$ .
- B2. La separación mínima entre dos pozos al mismo potencial es de 2  $\lambda$  y está garantizada en la superficie por el óxido de campo del límite del pozo.
- B3. La separación mínima para dos pozos a diferente potencial es 6  $\lambda$ .
- B4. El solapamiento mínimo del pozo sobre las difusiones internas es 3  $\lambda$ .
- B5. El pozo-p es una difusión profunda y por lo tanto las dimensiones exteriores deben garantizar la separación entre los límites del pozo y áreas adyacentes de difusión p+ fuera del pozo, para que no se produzcan caminos imprevistos entre las dos regiones. En procesos actuales con 5  $\lambda$  se garantiza lo mencionado.

c) Máscara 3

Para la máscara de polysilicon se señalan las siguientes reglas:

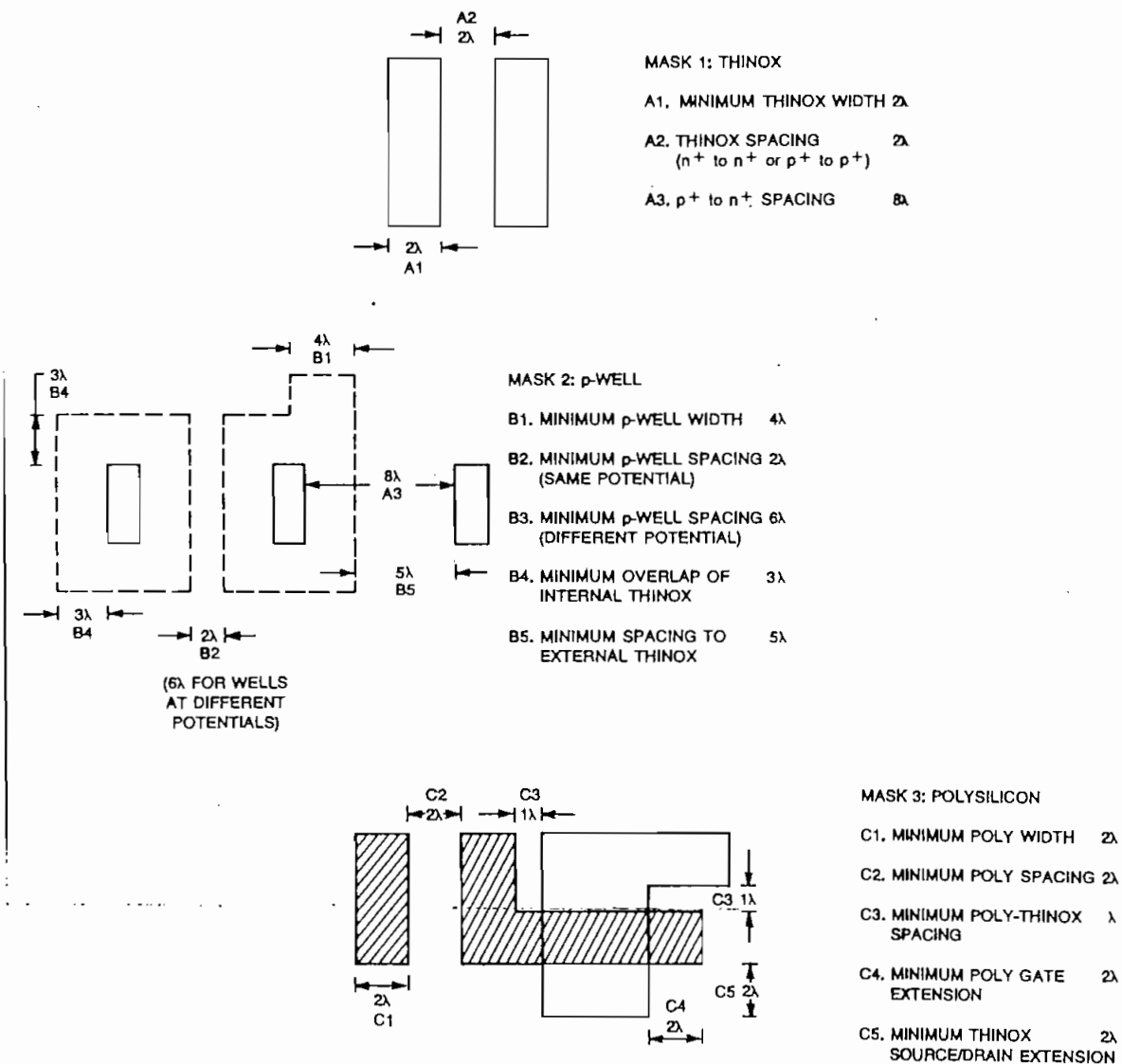


FIGURA 3.25 Reglas de diseño  $\lambda$  para un proceso CMOS representativo.

C1. El ancho mínimo para los caminos en *polysilicon* es  $2\lambda$ .

C2. A esta capa no están asociadas regiones de agotamiento, la separación entre dos caminos de *polysilicon* puede ser  $2\lambda$ .

C3. Esta regla se aplica al caso en que un camino de difusión corre paralelo a otro en *polysilicon*. El único requerimiento en este caso es que los dos caminos no se sobrelapen, si esto ocurre se forma un capacitor no deseado (*polysilicon-SiO<sub>2</sub>-difusión*).

Para preveer el sobrelapamiento, se requiere una separación *polysilicon* - difusión de 1 *lambda*. Si las capas llegan a sobrelaparse se tiene incrementos en los valores de las capacitancias parásitas, lo que haría mas lento al circuito, pero se considera como un efecto no catastrófico.

C4. Esta regla se aplica para la formación de transistores. Una región de difusión cruzada por una de *polysilicon*, forma un transistor. Para asegurar que la región de difusión en el extremo (derecho de la Fig) de la compuerta no se forme un camino entre las regiones de drenaje y fuente, se extiende el *polysilicon* 2 *lambda*. Para dejar clara esta regla debe recordarse que a pesar que en el *layout* se representan las zonas de difusión de drenaje y fuente con un solo rectángulo (Fig. 3.26a), durante la fabricación el *polysilicon* cubre la difusión y sirve como barrera durante el dopaje y las difusiones se construyen como dos áreas efectivamente separadas y alineadas (Fig. 3.26b). Al extender el *polysilicon* más allá del límite de las difusiones, se prevee que en el extremo de las mismas no se cree un camino.



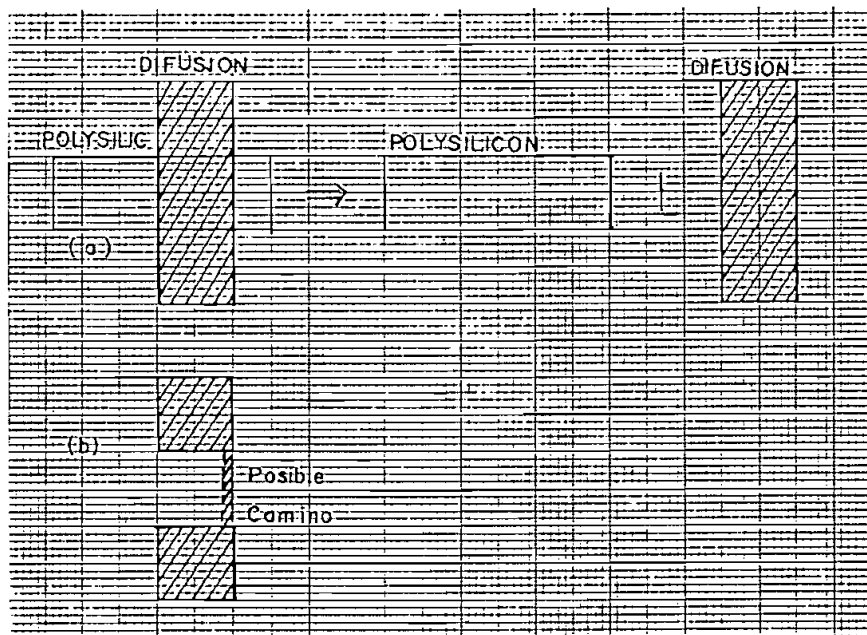


FIGURA 3.26 Formación de un Transistor MOS.

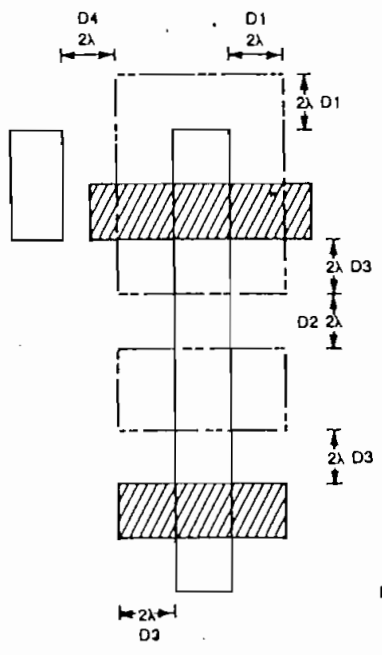
C5. De la Fig. 3.26b se desprende también que el ancho mínimo de cada área de difusión del transistor resultante de acuerdo a la primera regla establecida debe ser  $2\lambda$ .

Los gráficos correspondientes a la descripción de las reglas de diseño con las máscaras p-plus, contactos y metal se presentan en la Fig. 3.27

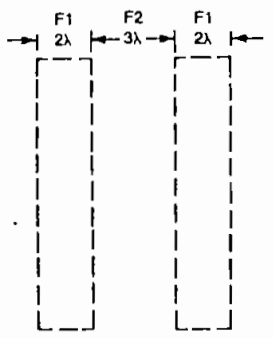
#### d) Máscara 4

Para la máscara p-plus se indican las siguientes reglas:

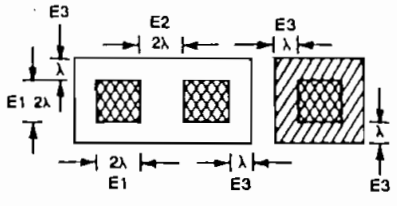
D1. El mínimo solapamiento de la máscara implante p+ sobre una difusión es  $2\lambda$ .



- MASK 4: p-PLUS**
- D1. MINIMUM OVERLAP OF THINOX  $2\lambda$
  - D2. MINIMUM p-PLUS SPACING  $2\lambda$
  - D3. MINIMUM GATE OVERLAP OR DISTANCE TO GATE EDGE  $2\lambda$
  - D4. MINIMUM SPACING TO UNRELATED THINOX  $2\lambda$



- MASK 8: METAL**
- F1. MINIMUM METAL WIDTH  $2\lambda$
  - F2. MINIMUM METAL SPACING  $3\lambda$
  - F3. MINIMUM METAL OVERLAP  $\lambda$  OF CONTACT



- MASK 5: CONTACT**
- E1. MINIMUM CONTACT AREA  $2\lambda \times 2\lambda$
  - E2. MINIMUM CONTACT SPACING  $2\lambda$
  - E3. MINIMUM OVERLAP OF POLY, OR THINOX OVER CONTACT  $\lambda$
  - E4. MINIMUM SPACING TO GATE POLY  $2\lambda$
  - E5. n+ SOURCE/DRAIN CONTACT
  - E6. p+ SOURCE/DRAIN CONTACT
  - E7. V<sub>SS</sub> CONTACT
  - E8. V<sub>DD</sub> CONTACT
  - E9. V<sub>SS</sub> SPLIT (OR MERGED) CONTACT (ELONGATED CONTACT SHOWN)
  - E10. V<sub>DD</sub> SPLIT CONTACT ( $2\lambda \times 2\lambda$  CONTACTS SHOWN)

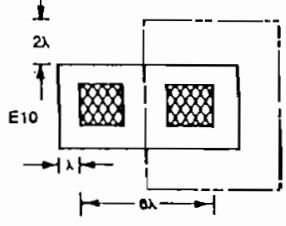
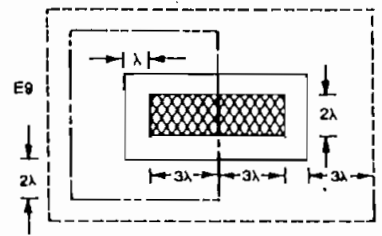
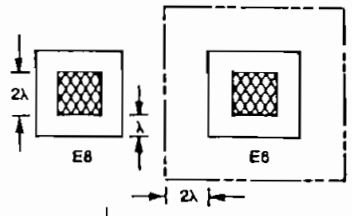
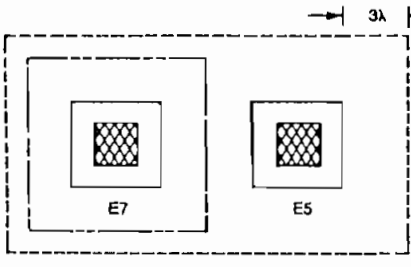
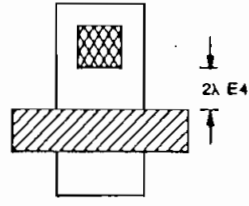


FIGURA 3.27 CONTINUACION Reglas de diseño  $\lambda$  para un proceso CMOS representativo.

- D2. El mínimo espaciamiento entre dos implantes p+ es  $2 \lambda$ .
- D3. El mínimo solapamiento de un implante p+ sobre una compuerta de *polysilicon* o la mínima distancia al límite de una compuerta es de  $2 \lambda$ .
- D4. El mínimo espaciamiento de un implante p+ a una difusión es  $2 \lambda$ .

e) Máscara 5

Para la máscara de contacto se indican las siguientes reglas:

- E1. El área mínima de un contacto es de  $2 \lambda \times 2 \lambda$ .
- E2. Cuando se desea hacer el contacto entre una gran región de metal y una gran región de difusión, puede requerirse no solo un contacto sino varios, espaciados  $2 \lambda$ .
- E3. Un contacto se forma generalmente entre la capa de metal y *polysilicon* o difusión. Para estos casos los materiales de cada capa deben proveer un área para rodear al contacto, de esta manera en la parte interior no se producen conexiones no deseadas hacia otros puntos. La regla señala que cualquier capa debe extenderse un  $\lambda$  de todos los límites del contacto.

E4. La distancia mínima de un contacto a la compuerta de un transistor es  $2 \lambda$ .

Las reglas desde la E5. hasta la E10. presentan las reglas de diseño para varios casos de contactos. Los contactos que pueden requerirse son:

- Metal a difusión "n".
- Metal a difusión "p".
- Metal a *polysilicon*.
- Contactos del pozo y sustrato a  $V_{DD}$  y Gnd.
- El contacto "*split*", equivalente a dos contactos metal-difusión.
- Existe también el denominado "*buried contact*" que se utiliza cuando se requiere una conexión directa entre *polysilicon* y difusión.

f) Máscara 6

Para el metal se indican las siguientes reglas:

F1. El ancho mínimo de un camino de metal debe ser  $2 \lambda$ . Sin embargo, por consideraciones de fabricación, ya que el metal se deposita al final sobre la superficie de la oblea, cuando ésta tiene una forma irregular, se sugiere generalmente que el ancho mínimo sea  $3 \lambda$ . El ancho del camino debe incrementarse de acuerdo a la corriente necesaria, para evitar el fenómeno de migración.

F2. La mínima separación entre dos caminos de metal es  $3\lambda$ .

F3. El mínimo solapamiento del metal sobre un contacto, por todos los límites del mismo es  $\lambda$ .

En la Tabla 3.1 se presenta la derivación de algunas reglas  $\lambda$  de un conjunto representativo de reglas "micron".

MASK	FEATURE	DIMENSIONS	
		Micron rule	$\lambda$ rule
1: Thinox	Minimum thinox width	4 $\mu\text{m}$	2 $\lambda$
	Minimum thinox spacing	4 $\mu\text{m}$	2 $\lambda$
	Minimum p-thinox to n-thinox spacing	8 $\mu\text{m}$	4 $\lambda$
3: Polysilicon	Minimum poly width	3.75 $\mu\text{m}$	2 $\lambda$
	Minimum poly spacing	3.75 $\mu\text{m}$	2 $\lambda$
	Minimum gate poly width (p)	4.5 $\mu\text{m}$	3 $\lambda$
	Minimum gate poly width (n)	4.0 $\mu\text{m}$	2 $\lambda$
	Minimum gate poly extension	3.5 $\mu\text{m}$	2 $\lambda$
6: Aluminum	Minimum Al width	4.5 $\mu\text{m}$	3 $\lambda$
	Minimum Al spacing	4.5 $\mu\text{m}$	3 $\lambda$

TABLA 3.1 Derivación de reglas  $\lambda$  a partir de reglas micron.

Debido a los valores resultantes para las reglas  $\lambda$ , puede ocurrir que el comportamiento de un circuito no sea tan óptimo, ni el área tan pequeña, comparado al circuito que se fabricaría utilizando las reglas micron. Estos factores hacen los diseños con reglas  $\lambda$  inadecuados para circuitos comerciales y aún para circuitos de carácter experimental, principalmente por los costos adicionales que implica el

incremento de área. Lo más común, especialmente para usos comerciales, es por lo tanto expresar las reglas de diseño con las dimensiones en  $\mu\text{m}$  (reglas *micron*).

Desde que fueron formuladas las reglas de diseño de Mead y Conway para nMOS, se han hechos algunos intentos para desarrollar un conjunto de reglas para la tecnología CMOS, que sean tan manejables como aquellas.

El proceso de fabricación CMOS tiene un mayor número de capas que las requeridas en nMOS, lo que implica también un mayor número de reglas de diseño, esto asociado a la complejidad y variaciones de su proceso de fabricación es uno de los mayores problemas en formular un buen conjunto de reglas para CMOS. Además, cada fundidora elige o incluso desarrolla un determinado proceso de fabricación y asociado al mismo se dispone del conjunto de reglas de diseño.

Como las herramientas automatizadas para diseño de CIs son cada día más accesibles, la necesidad de que un diseñador conozca detalladamente las reglas de diseño es cada vez menos importante.

En los ambientes de desarrollo de CI, se incluyen los denominados "chequeadores de reglas de diseño", que son programas computacionales que determinan si las reglas de diseño han sido violadas, al final de su ejecución entrega reportes al diseñador.

Un programa "chequeador de reglas" utiliza el conjunto de reglas para un determinado proceso de fabricación. Así por ejemplo, puede utilizar las reglas para un proceso de doble pozo de  $2.0\ \mu\text{m}$  o las reglas para un proceso de pozo-p también de  $2.0\ \mu\text{m}$ . Podría también utilizarse las reglas para un proceso de doble pozo ya sea de  $1.0\ \mu\text{m}$  o de  $2\ \mu\text{m}$ . En este último caso, las reglas pueden ser escaladas, por un factor determinado, y utilizarse para los dos procesos.

En el ANEXO G se presenta el conjunto de reglas utilizadas por la fundidora ES2 (European Silicon Structures) e incluidas en el Paquete de diseño TENTOS. Estas reglas son para un proceso CMOS de pozo-n, con dos niveles de metal y de  $1.5\ \mu\text{m}$ .

En el ANEXO G se presentan también las reglas de MOSIS, utilizadas por el paquete FPL. Estas reglas son para un proceso de doble pozo, con dos niveles de metal y de  $2.0\ \mu\text{m}$ .

### **3.4 ESTRUCTURAS DE ENTRADA/SALIDA: PADs**

En la Fig 3.28 se presenta el esquema general de un CI fabricado. El CI puede considerarse constituido por :

- a) La estructura de Silicio resultante del proceso de fabricación, que está conformada por:

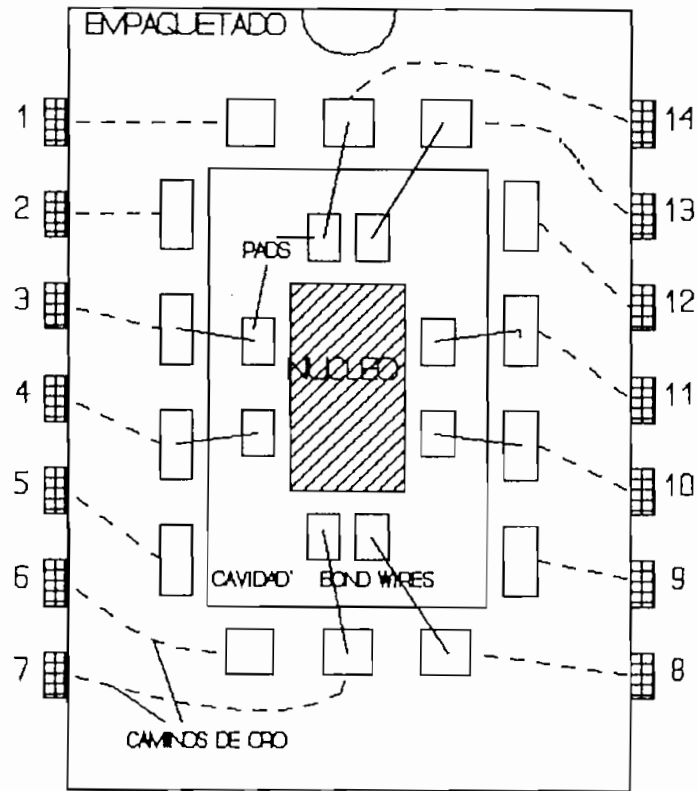


FIGURA 3.28 Estructura general de un CI.

- i. El nucleo, que es toda la circuitería interna planificada por el diseñador.
- ii. Los PADS que son las estructuras que permiten realizar el interfaz entre el núcleo y el mundo externo.



Se utilizan tantos PADs, cuantas señales de entrada y salida tenga el circuito, más los PADs necesarios para proveer la polarización al circuito. En la metodología de diseño *semicustom*, el diseñador no se preocupa de la estructura interna y diseño de los PADs, simplemente hace uso de las bibliotecas de caracter estándar que se disponen para el efecto. Debe existir el enrutamiento necesario entre el núcleo y los PADs.

- b) El empaquetamiento, que es la forma rectangular, de plástico o cerámica, que contiene:
  - i. Una cavidad (*die cavity*), en la cual se deposita la estructura de Silicio.
  - ii. Los "*bonding PADs*", que no son sino pequeñas formas rectangulares (almohadillas), que mediante caminos de oro se conectan a los pines externos del empaquetado. Estos elementos se encuentran rodeando la cavidad.
  - iii. Los pines externos y los caminos de oro que unen los "*bonding PADs*" y los pines externos.
- c) Mediante alambres conductores se unen electricamente los PADs de la estructura de Silicio con los "*bonding PADs*". A estos alambres se les denomina "*bond wires*".

### 3.4.1 Estructuras de entrada/salida (I/O)

Los PADS sirven para proporcionar al CI las señales de entrada (*IN*) y obtener de él, las señales de salida (*OUT*). Es también a través de los PADS que se proporciona la alimentación necesaria para la operación del circuito.

De todas las estructuras CMOS que se diseñan, las estructuras de entrada/salida (*input/output, I/O*) requieren del mayor grado de conocimiento y experiencia del diseñador, así como de un profundo conocimiento de los procesos de fabricación. Por esta razón se considera inapropiado que el diseñador se preocupe por la concepción de estas estructuras. Es más bien generalizado utilizar bibliotecas preparadas que se escojen de acuerdo al proceso de fabricación y requerimientos eléctricos.

De acuerdo a la función que realice la estructura, se tienen PADS de entrada, salida, de tres estados, bi-direccionales y los PADS para  $V_{DD}$  y Gnd.

#### a) PADS para $V_{DD}$ y Gnd

Estos PADS se diseñan fácilmente y consisten de una almohadilla (*pad*) de metal, de forma rectangular o cuadrada, que sirve de área de contacto para realizar la suelda con el "*bond wire*" y conectarlo al potencial adecuado.

Como se verá más adelante, los PADS de salida pueden requerir de un *driver* inversor y posiblemente de algunas compuertas adicionales, para conseguir las corrientes necesarias y así poder manejar otros CI. Estos elementos necesitan también ser polarizados. Por esta razón, alrededor del núcleo, en las áreas en donde se encuentran los PADS, corren buses de  $V_{DD}$  y Gnd.

En la Fig. 3.29 se representa la estructura de un PAD que proporciona  $V_{DD}$  para los PADS y para el núcleo. Este PAD está entonces conectado al bus de  $V_{DD}$ . Bajo el bus de  $V_{DD}$  corre el bus para Gnd ( $V_{SS}$ ), que debe ser interrumpido para atravesar el camino de metal que lleva  $V_{DD}$  hacia el núcleo. El camino interrumpido puede completarse con *polysilicon*, como se indica en la Fig. 3.29. En un proceso con dos niveles de metal no es necesaria la interrupción de los caminos en las áreas de cruce.

Es importante mencionar que las dimensiones de las almohadillas y demás estructuras que se utilizan en los PADS, son considerablemente mucho más grandes que las que se utilizan en el núcleo. Por lo tanto, cuando se requiere realizar contactos y/o vías, por tratarse de grandes áreas, se utilizan múltiples contactos y/o vías. Así por ejemplo, en la Fig. 3.29 en cada unión de metal con *polysilicon* se utilizan seis contactos. La estructura de un PAD para Gnd es muy similar a la presentada en la Fig. 3.29, excepto que está conectado al bus para Gnd.

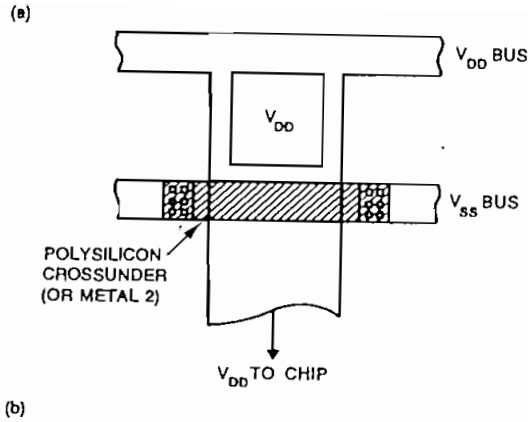


FIGURA 3.29 Estructura de un PAD para  $V_{DD}$ .

Las almohadillas de los PADs se diseñan generalmente de una altura y ancho constantes, con los puntos de conexión en ubicaciones previamente especificadas. El tamaño de la almohadilla está determinado generalmente por el mínimo tamaño requerido para realizar la conexión mediante un alambre a los "bonding PADs". Esta área está en el orden de  $150\mu\text{m} \times 150\mu\text{m}$ . Se ha generalizado también la ubicación de las estructuras necesarias para  $V_{DD}$  y Gnd, generalmente en las esquinas y puntos centrales.

El ancho de los caminos de  $V_{DD}$  y Gnd se estiman para las peores condiciones de consumo de potencia. La circulación de corrientes de valores altos pueden causar caídas de voltaje considerables en los caminos que distribuyen la polarización, haciendo a los circuitos inoperables. Estas variaciones de voltaje se deben a las caídas  $IR$  y  $(L \cdot di/dt)$ . Siendo  $R$  y  $L$ , la resistencia e inductancia asociadas a los caminos.

Una fuente de grandes valores de corriente se presenta cuando los grandes *drivers* manejan grandes capacitancias asociadas con el mundo externo.

Por ejemplo si la inductancia equivalente asociada a un camino de polarización para 20 FADs es 10 nH y la rapidez a la que cambia la corriente es 10 mA/ns, asumiendo que todos los FADs conmutan simultáneamente, la caída debido únicamente al término  $L \cdot di/dt$  es:

$$[10 \times 10^{-9} \times (10 \times 10^{-3} / 10^{-9})] \times 20 \text{ FADs} = 2V,$$

que es un valor que debe ser considerado en el momento del diseño.

Un método para reducir estas caídas de voltaje es distribuir separadamente la polarización para los FADs y el núcleo. Generalmente en estos casos se recomienda utilizar dos FADs para  $V_{DD}$  y dos FADs para Gnd. De ser necesario el número de FADs puede ir incrementándose para la distribución de la polarización en el núcleo.

#### b) FADs de salida

La principal característica de un FAD de salida es que debe tener la suficiente capacidad de manejo de corriente para proporcionar adecuados tiempos de subida y bajada sobre una carga capacitiva.

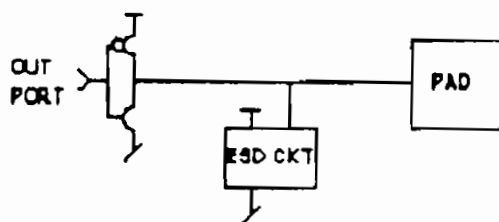


FIGURA 3.30 Estructura de un PAD de salida.

Para cumplir con los requerimientos de manejo de carga y velocidad se utilizan grandes transistores como parte de buffers inversores, a más de la almohadilla necesaria (Fig. 3.30). Los transistores deben ser adecuadamente dimensionados. El *buffer* permite presentar una carga menor a la circuitería interna.

Dada la carga y los tiempos de subida y bajada se calculan los tamaños de los transistores de salida. Generalmente en un PAD se utiliza un par de inversores en cascada para obtener la señal original y no la invertida como en el caso de utilizar un solo inversor.

Por los grandes transistores de las estructuras de I/O circulan corrientes también altas y debido a su conexión con el mundo externo, si se producen transientes de voltaje que vayan sobre  $V_{DD}$  o bajo Gnd, es en los PADs donde la susceptibilidad al *latch-up* es máxima.

Para combatir el *latch-up*, se deben por lo tanto tomar las medidas adecuadas utilizando los anillos de guarda en los transistores de las estructuras de salida y conectándolos a los voltajes de polarización adecuados.

Una vez determinados los tamaños de los transistores se inicia el diseño del layout. Si el pad maneja cargas que no sean CMOS, se deben también cumplir las características DC requeridas. Cuando las compuertas CMOS manejan cargas TTL, se deben considerar los niveles umbrales necesarios para conmutar la compuerta TTL.

El voltaje de entrada en bajo máximo ( $V_{ILmax}$ ) de una compuerta TTL es 0.8V, el voltaje de salida en bajo ( $V_{OL}$ ) de una compuerta CMOS es 0V, por lo que no hay ningún problema. El voltaje de entrada en alto máximo ( $V_{IHmax}$ ) para una compuerta TTL es 2.4V. El voltaje de salida en alto ( $V_{OH}$ ) de una compuerta CMOS es 5V (para una polarización de 5V) y por lo tanto no hay problema.

Para el estado en bajo, el *buffer* CMOS debe absorber 1.6 mA para una carga TTL con un  $V_{IL}$  menor a 0.8V, que no es problema para los grandes transistores *drivers* típicos.

### c) PADS de entrada

El desarrollo de los PADS de entrada es similar al de los de salida en lo que respecta al dimensionamiento de los

transistores, a menudo simplemente se cambia de sentido los transistores diseñados para los FADs de salida. Sin embargo, se deben tener precauciones de otra naturaleza. La impedancia de entrada de un transistor MOS es sumamente alta ( $10^{12}$  a  $10^{13} \Omega$ ). Y el voltaje al cual el óxido de la compuerta se perfora está en el rango de (40-100)V. El voltaje que puede tenerse en la compuerta se puede determinar de la relación:  $C=Q/V$  y  $I=Q/t$ , entonces:

$$V = \frac{I \Delta t}{C_g}$$

[Ec. 3.1]

En donde:

$V$  = voltaje de la compuerta.

$I$  = corriente de carga.

$\Delta t$  = tiempo requerido para cargar la compuerta.

$C_g$  = capacitancia de la compuerta.

Si se tienen los siguientes valores:  $I=10\mu A$ ,  $C_g=0.03pF$  y  $\Delta t = 1 \mu seg$ , el voltaje que aparece en la compuerta es de aproximadamente 330V.

También los momentos en que se manipulan los dispositivos, pueden generarse descargas electrostáticas (*electrostatic discharges, EDC*) a través de los FADs, que si llegan a las compuertas internas producen el daño permanente del óxido de la compuerta.



Existen varias soluciones para limitar los sobrepicos de voltaje. Típicamente consisten de diodos polarizados inversamente y resistores de *polysilicon* o resistores activos para disipar energía. En la Fig. 3.31 se representa la estructura de un PAD de entrada.

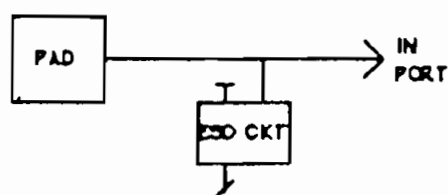


FIGURA 3.31 Estructura de un PAD de entrada.

En la Fig. 3.32a se representa el caso en que la circuitería de protección se realiza con un par de diodos y una resistencia para limitar los potenciales destructivos. La Fig. 3.32, incluye los layouts aproximados de los elementos de protección.

Los diodos D1 y D2 conducen si el voltaje en el nodo X sobrepasa  $V_{DD}$  o baja del valor de voltaje de Gnd. La resistencia R se utiliza para limitar el pico de corriente que puede fluir en cualquiera de los diodos en caso de un pico de voltaje inusual.

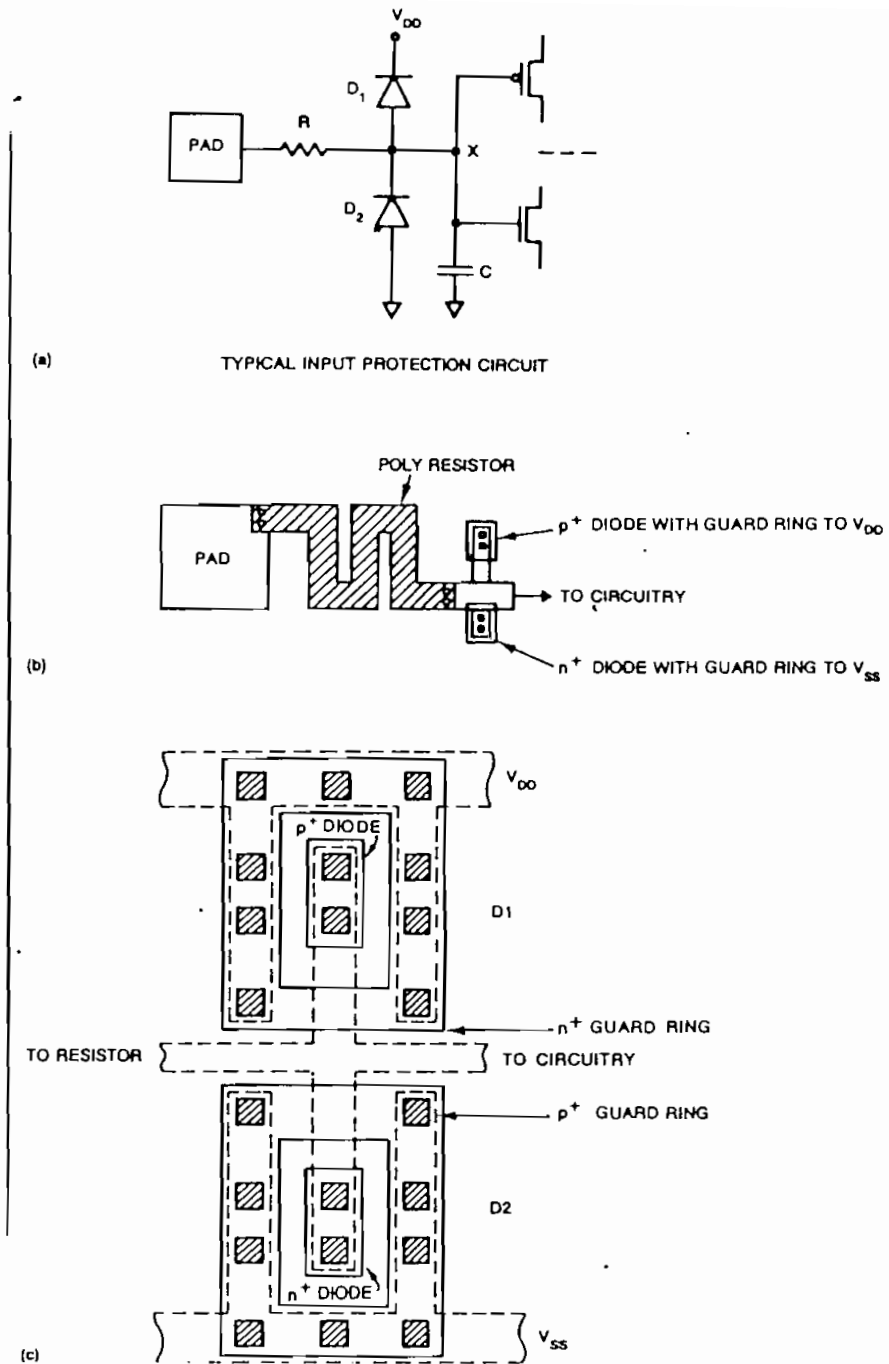


FIGURA 3.32 Estructura de un PAD de entrada y layout de los elementos de protección.

El valor del resistor está en el rango de  $200\Omega$  a  $3K\Omega$ . Esta resistencia en conjunto con cualquier capacitancia de entrada  $C$ , proporciona una constante  $RC$ , que debe considerarse en circuitos de alta velocidad. En un proceso pozo-p, para evitar posibles inyecciones extras de carga al sustrato, facilitando el latch-up, la resistencia  $R$  se construye con *polysilicon* en lugar de difusión. En un proceso pozo-n, toda la circuitería de I/O puede diseñarse con dispositivos "n".

En la Fig. 3.32b se presenta el layout de la resistencia de protección en *polysilicon* y en la Fig. 3.32c, el layout de los diodos  $D1$  y  $D2$ . Para un proceso pozo-p, el diodo  $D1$  se estructura construyendo una difusión "p+" en el sustrato "n". El diodo  $D2$  se forma con una difusión "n+" en el pozo "p". Las dos difusiones "p+" y "n+" (ánodo de  $D1$  y cátodo de  $D2$ , respectivamente) se unen utilizando metal y se enrutan a la resistencia  $R$  y a los transistores de entrada. La figura incluye también los anillos de guarda de cada una de las difusiones construídas y los contactos necesarios para sujetarse a los buses de polarización.

**d) PADs de salida tres estados.**

Una alternativa para un PAD de salida tres estados se presenta en la Fig. 3.33a, junto con la tabla de funcionamiento que incluye la señal de control ( $C$ ), la señal de datos ( $D$ ) y los valores  $P$  y  $N$ , que determinan el estado de los transistores.

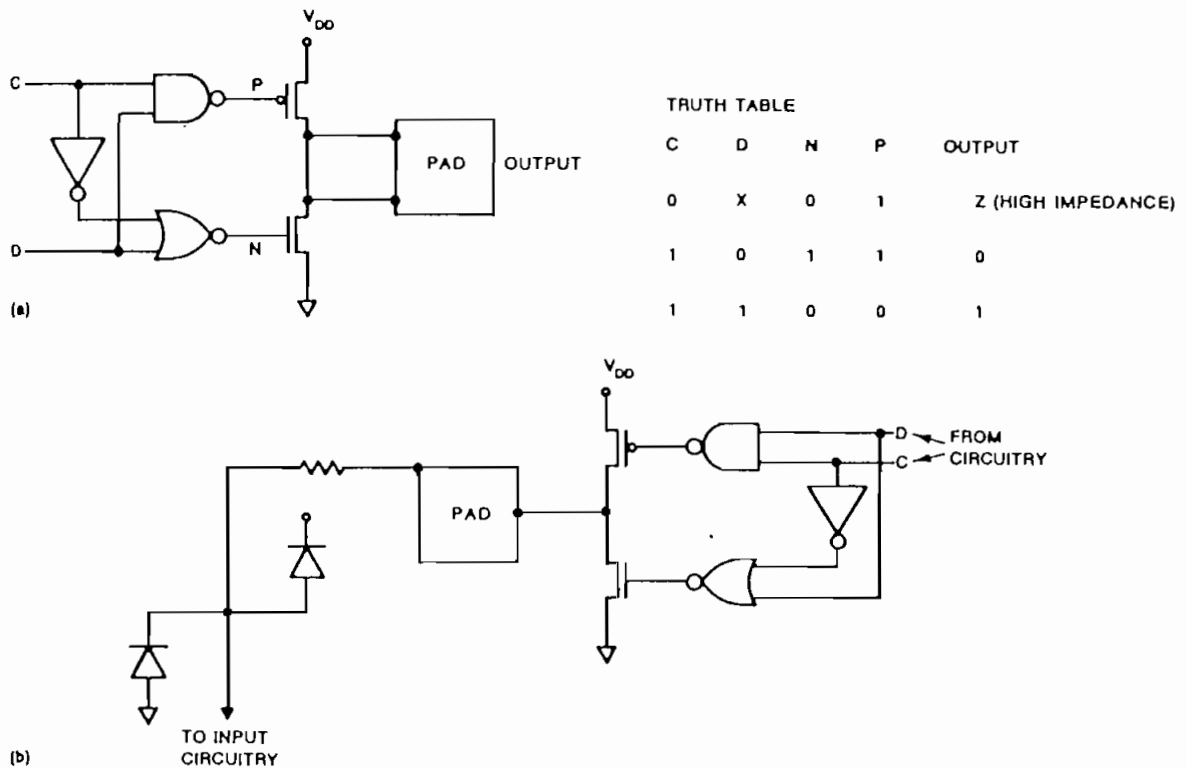


FIGURA 3.33 Estructura de un PAD de a) tres estados b) bidireccional.

### e) PADs bidireccionales.

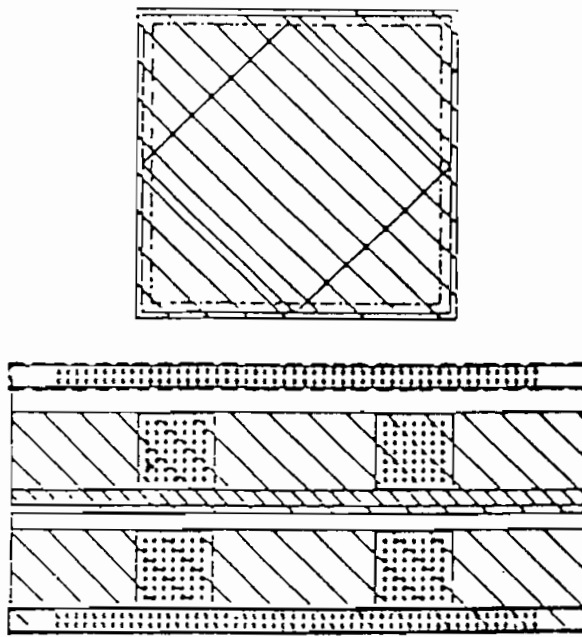
Un PAD bidireccional se consigue combinando un PAD de entrada y un PAD tres estados (Fig. 3.33b). La almohadilla es el medio de ingreso de la señal desde el exterior, la cual pasa por la circuitería de protección y luego sigue al núcleo. En estas condiciones la salida del *driver* debe estar en alta impedancia. Utilizando la estructura del PAD de tres estados, la señal de salida se proporciona a la almohadilla, por medio del *driver*, y de ahí hacia el pin externo. Se supone que circuitería adicional aísla la señal del *driver* de la circuitería de entrada hacia el núcleo.

f) PADS "blancos" y de "esquinas".

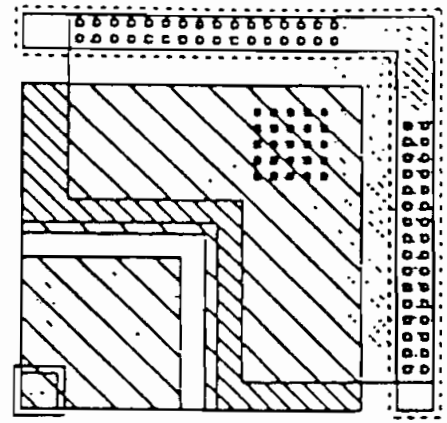
Muchas veces las fundidoras proporcionan estructuras denominadas "PAD-FRAMES", que tienen la ubicación de los PADS y en cuyo interior el usuario puede ubicar el diseño de su CI. La fundidora determina la ubicación obligatoria de los PADS para polarización y el número máximo y ubicación de PADS adicionales a utilizarse, pero no determina su tipo.

Al inicio, la estructura tiene en la ubicación de los PADS, un tipo denominado "blanco" (*blank*), de estructura similar a los PADS de  $V_{DD}$  y Gnd, que permiten satisfacer los requerimientos de ubicación de los PADS. Un PAD "blanco" tiene la almohadilla y los caminos que permiten que los buses de polarización de los PADS corran hacia los dos lados, pero los dos elementos están completamente aislados. Por lo tanto, la estructura no permite el paso de ninguna señal hacia los "bonding PADS".

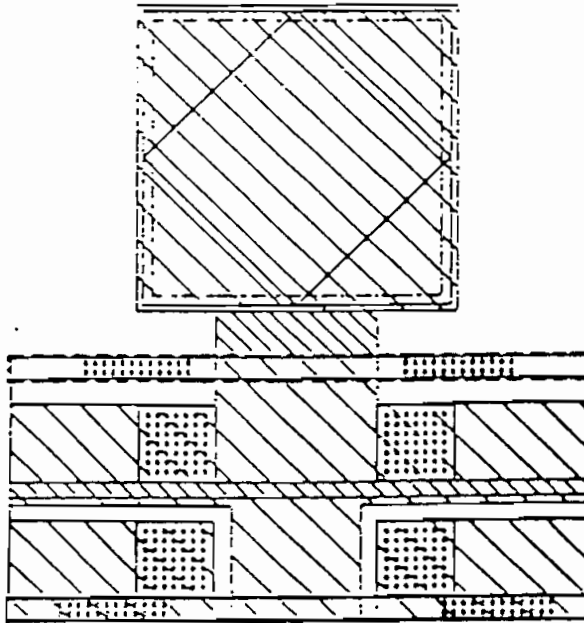
En la Figs. 3.34 y 3.35 se presentan layouts completos de los PADS para  $V_{DD}$  y Gnd y la diferencia existente con el PAD blanco es evidente. También puede apreciarse el gran número de contactos requeridos cuando se necesitan realizar uniones. Para las esquinas se utilizan también una configuración especial para permitir que el paso de los buses de polarización de los PADS no sean interrumpidos. En la Fig. 3.35 se presentan layouts para los PADS de entrada, salida y tres estados.



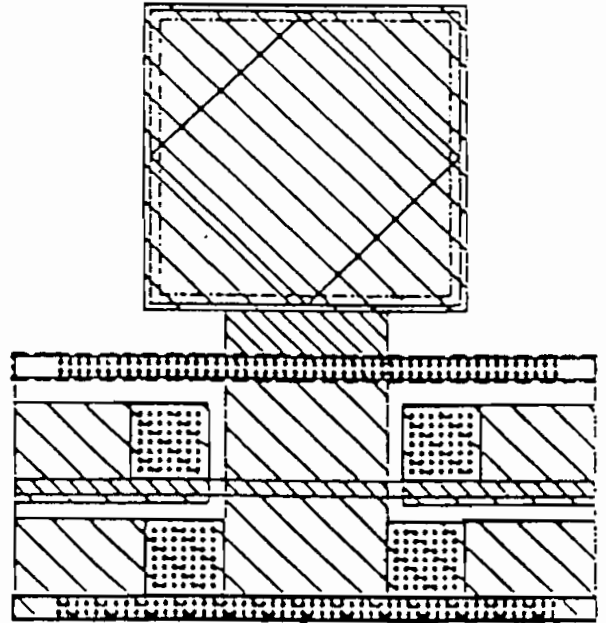
BLANK



CORNER



VDD

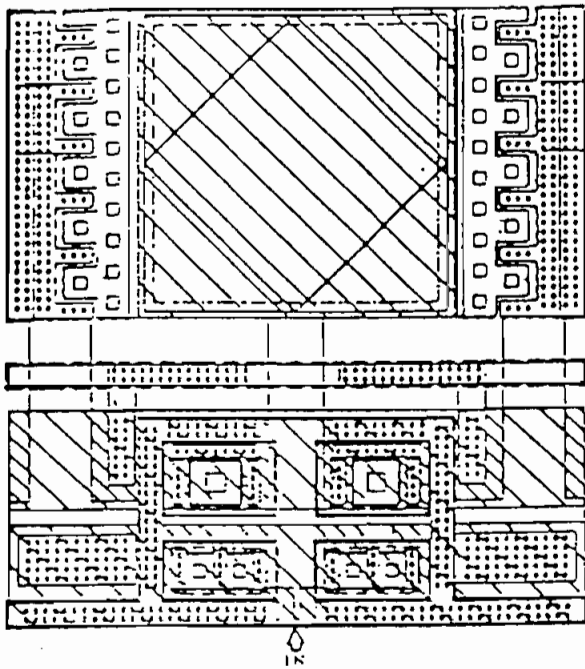


GND

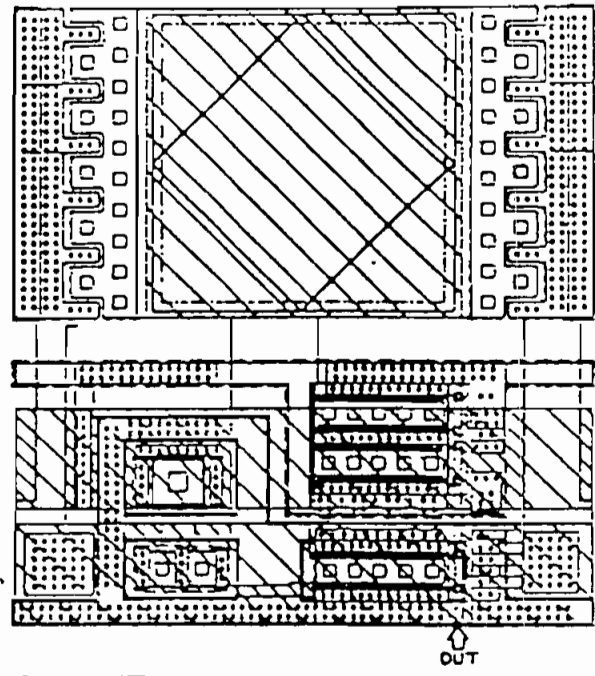
VDD

GROUND

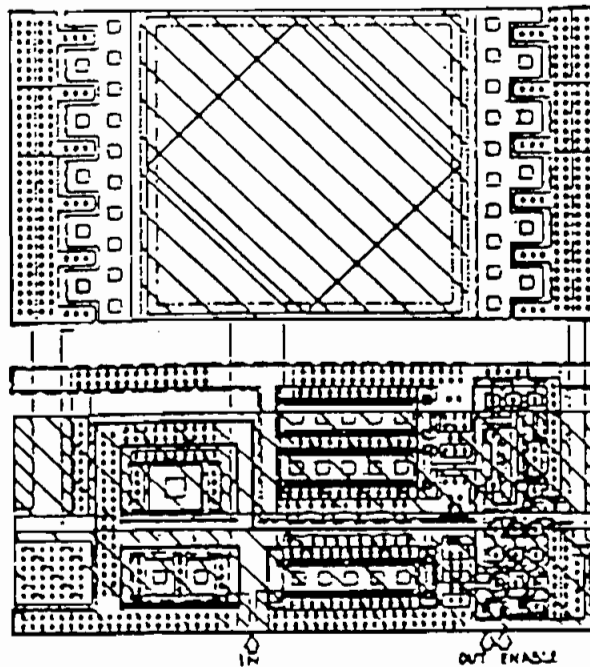
FIGURA 3.34 Estructura de los PADs a) blank b) esquina c) VDD d) Gnd.



INPUT



OUTPUT



TRI-STATE

FIGURA 3.35 Estructura de los PADs a) entrada b) salida c) tres estados.

### 3.4.2 Anillos de $V_{DD}$ y Gnd.

Para evitar la interferencia electromagnética, se acostumbra a rodear a la estructura completa, núcleo y PADS, con una anillo ya sea de  $V_{DD}$  o Gnd, dependiendo del proceso de fabricación (Fig. 3.36).

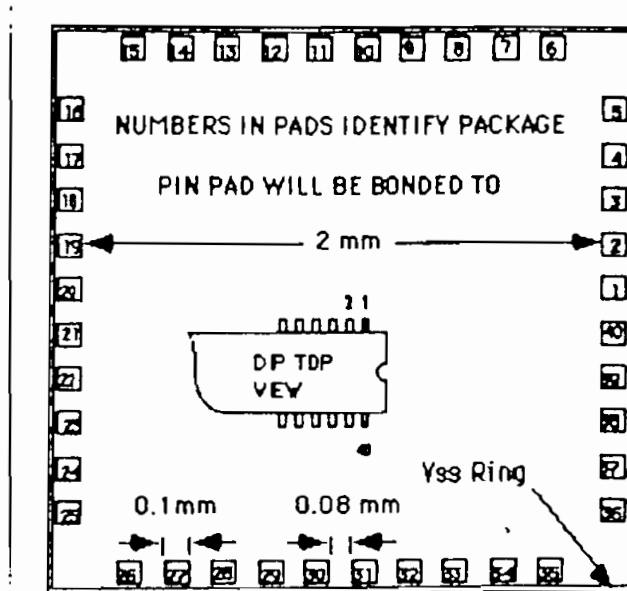


FIGURA 3.36 Anillo de  $V_{DD}$  o Gnd.

Para un proceso pozo-n se utiliza un anillo de Gnd conectado al sustrato "p". En la Fig. 3.37 se presenta la estructura del anillo de Gnd para un proceso de dos niveles de metal.

Primeramente se construye un anillo de metal 2, que se conecta por medio de "vías" a un anillo de metal 1. El anillo de metal 1 a la vez se conecta por medio de contactos a un anillo p+ construido en el sustrato "p".



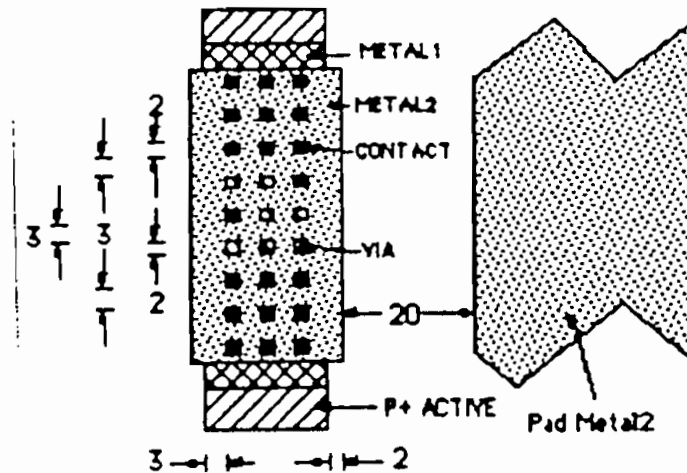


FIGURA 3.37 Estructura de un anillo de Gnd para un proceso pozo-n.

Al conectar el anillo de Gnd a tierra se protege a la estructura de interferencia electromagnética, del latch-up y además obtenemos una mejor conexión del sustrato a tierra, mejorando el funcionamiento del circuito. Para el caso de un proceso pozo-p, se debería construir un anillo de  $V_{DD}$ .

### 3.5 TIPOS DE EMPAQUETAMIENTO

El proceso de fabricación de los empaquetados requiere procesos tecnológicos complejos, deben ser el resultado de muchas pruebas mecánicas, térmicas y eléctricas, para asegurar el buen funcionamiento y hermeticidad del CI. Los materiales utilizados pueden ser plástico o cerámica. El empaquetado contiene a los pines externos, a los bonding PADS y los caminos de oro que interconectan a los dos anteriores.

El proceso de empaquetamiento incluye dos fases:

- a) La fijación del chip a la cavidad del empaquetamiento, lo que se realiza con la ayuda de resina epóxica u otras sustancias.
  
- b) La unión de los PADS del chip con los "banding PADS", mediante un alambre conductor que puede ser de oro o aluminio. Se utiliza para esta fase diversos métodos de soldadura, por ejemplo:
  - i) Suelta ultrasónica
  - ii) Suelta de termo-compresión
  - iii) Suelta termosónica
  
- c) Sellado de la estructura.

De acuerdo a las dos tecnologías básicas de montaje, se tiene diversos tipos de empaquetamiento. Las dos tecnologías de montaje son:

- a) La tecnología "through board" (Fig. 3.38), cuyas principales características son:
  - i) Los componentes son grandes y los montajes seguros.
  - ii) El número de pines es limitado.
  - iii) Atraviesan las placas en las que se los monta por lo que las placas se utilizadan de un solo lado.

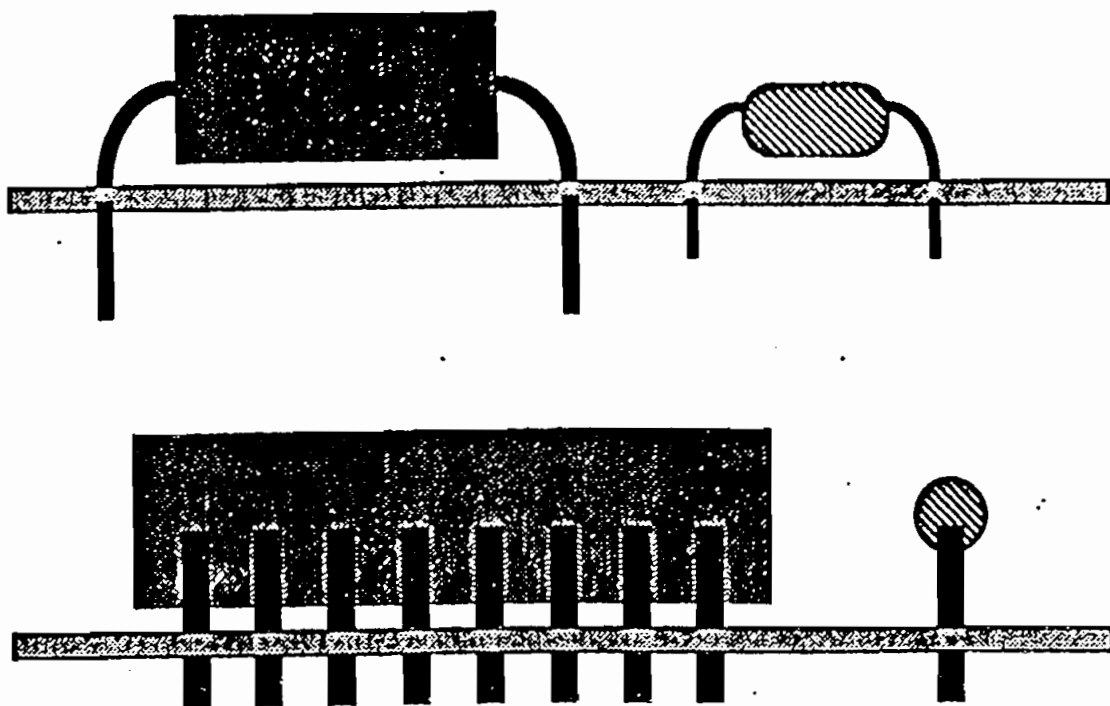


FIGURA 3.38 Tecnología "trough board".

- b) La tecnología SMD (*surface mounted devices*), cuyas principales características, respecto a la anterior son (Fig. 3.39):
- i) Los montajes no son tan reforzados.
  - ii) Se puede utilizar los dos lados de la placa para montar los dispositivos.
  - iii) Requiere mayor soporte técnico y de ingeniería.
  - iv) Los equipamientos necesarios son generalmente caros.
  - v) Mayor potencia y densidad de información concentradas en el mismo espacio.

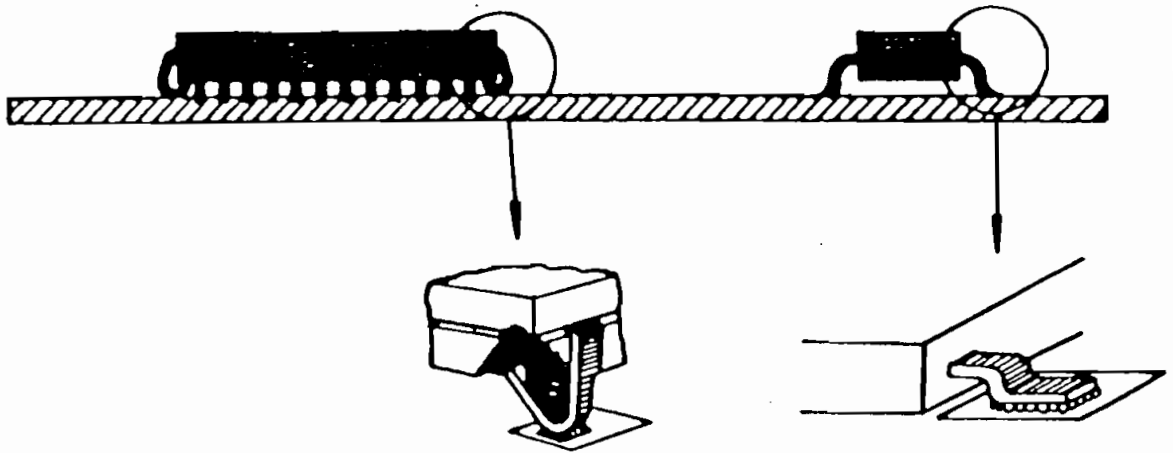


FIGURA 3.39 Tecnología SMD.

Los tipos de empaquetamiento para la tecnología "through board", fabricados con cerámica o plástico, son:

- a) La convencional DIP (*dual in line package*), que tiene sus pines distribuidos en dos lados de su estructura, en dos hileras. Número de pines típicos son: 8, 14, 16, 18, 20, 24, 28, 40, 48 (Fig. 3.40).

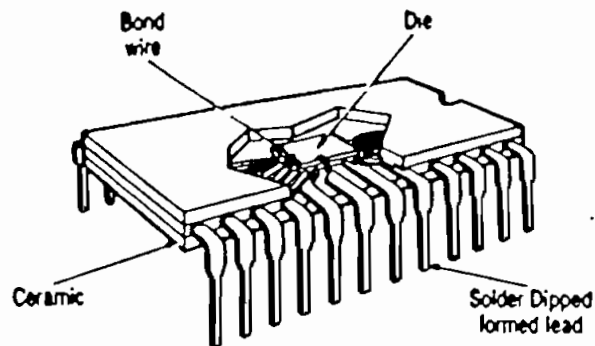


FIGURA 3.40 Empaquetamiento DIP.

- b) FGA (*Pin grid array*), con sus pines distribuidos en la superficie inferior del empaquetado, formando una verdadera grilla. Números de pines típicos son: 84, 120, 144 (Fig. 3.41).

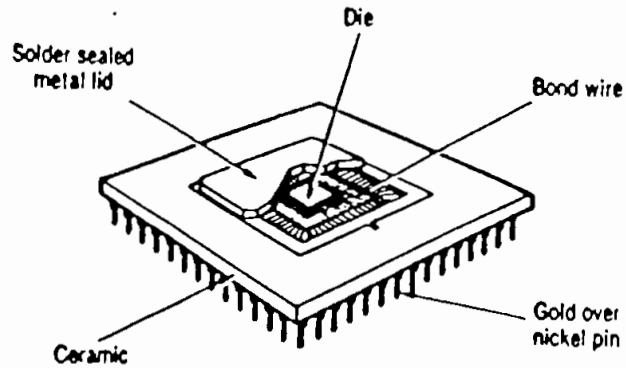


FIGURA 3.41 Empaquetamiento PGA.

Los tipos de empaquetamiento para la tecnología SMD, fabricados con cerámica o plástico, son:

- a) SOIC (*small outline ICs*), con sus pines distribuidos en dos hileras. Números típicos de pines son: 8, 16, 20, 24, 28 (Fig. 3.42)

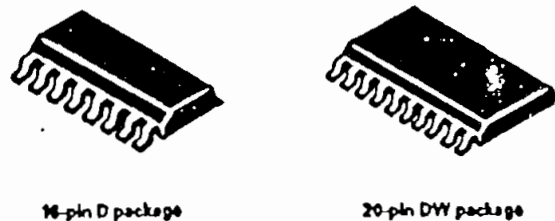


FIGURA 3.42 Empaquetamiento SOIC.

2. *Quad flat pack*, de forma planar con sus pines distribuídos en los cuatro lados de su estructura. Números de pines típicos son: 100, 208 (Fig. 3.43).

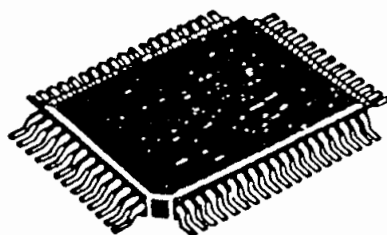


FIGURA 3.43 Empaquetamiento *Quad flat pack*.

- c) *Chip carrier- J leaded*, este tipo de empaquetamiento se utiliza para insertar los chips en zócalos construídos para el efecto. Para realizar el contacto con el zócalo se tiene los pines formando una especie de "J" alrededor del encapsulado (Fig. 3.44). Números de pines típicos son: 28, 44, 68, 84.

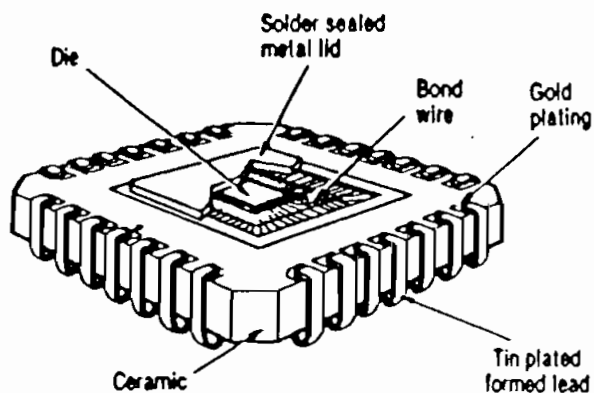


FIGURA 3.44 Empaquetamiento *chip carrier - J leaded*

d) *Chip carrier - leadless*, de características similares al anterior, con la diferencia que para el contacto con el zócalo se tienen ahora láminas conductoras en hendiduras del empaquetado. El zócalo tiene un elemento saliente que ejerce presión para lograr el contacto (Fig. 3.45).  
 Números de pines típicos son: 28, 44, 68, 84, 100.

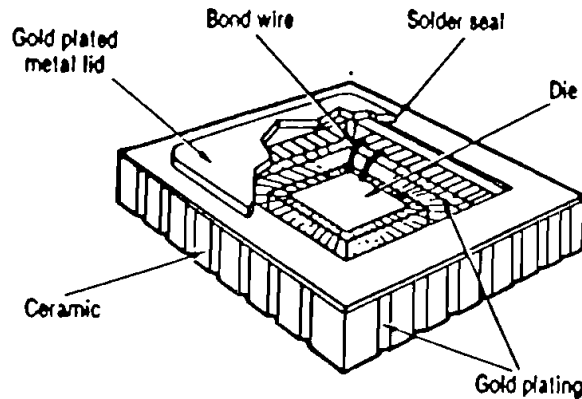


FIGURA 3.45 Empaquetamiento chip carrier - leadless.

En la Tabla 3.2 se presentan estadísticas y proyecciones del consumo de circuitos integrados (en billones de unidades) de acuerdo al tipo de empaquetamiento (Agosto de 1991).

<b>Table 1. Forecast of Worldwide IC Package Consumption (Billions of Units)</b>							
	1989	1990	1991	1992	1993	1994	1989-94 CAQR
Total ICs	28.62	41.90	49.93	57.87	63.96	69.82	12.6%
Plastic DIP	24.00	23.70	25.09	25.17	22.50	18.53	-5.0%
Ceramic DIP	2.94	2.90	3.11	3.22	3.34	3.38	2.9%
Small outline	6.48	8.48	11.88	15.68	19.42	22.27	28.5%
Chip carrier	2.04	2.64	3.69	5.01	6.70	9.04	34.7%
Quad flat pack	1.40	2.28	3.91	6.22	9.24	13.15	56.6%
Pin grid array	0.47	0.52	0.64	0.74	0.83	0.91	14.4%
Other	1.32	1.39	1.62	1.82	1.95	2.03	9.1%

Source: Advanced IC Packaging, Electronic Trend Publications

TABLA 3.2 Tendencia en el consumo de componentes en el mundo.

En la Fig. 3.46 se presenta un diagrama de una cavidad, con la distribución de los "bonding PADS" alrededor de la misma. El número asociado a cada "bonding PAD" corresponde al pin externo del empaquetado, como se indica en el gráfico de la parte inferior. El ejemplo presentado corresponde a un empaquetado de 40 pines del tipo DIP.

En la Fig. 3.47 se presenta un esquema similar al de la Fig. 3.45. En este caso se utiliza un empaquetado del tipo PGA de 84 pines. En los "bonding PADS" se incluye un nombre asignado a cada pin, que permite realizar la correspondencia con los pines del empaquetado. La vista superior indica la distribución de los pines en toda la superficie inferior, a excepción del área bajo la cavidad.

### 3.6 CHIPS MULTI-PROYECTO

En el ambiente de diseño de CIs, ha ganado popularidad la técnica de diseñar sistemas grandes y complejos en módulos. Estos módulos o subsistemas se fabrican y prueban hasta tenerlos completamente depurados. Al final se integran todos los subsistemas y se genera el layout de todo el sistema. Utilizar esta técnica permite verificar si los diseños lógicos, layout e implementación de un subsistema son correctos, con la ventaja adicional de medir las características de retardo y consumo de potencia, obteniendo de esta manera más información que la que se obtendría con la fabricación del sistema completo.



Sin embargo, esta técnica es práctica solamente si existe un esquema que permita disponer de los módulos fabricados en un tiempo relativamente corto. Los esquemas denominados chips multiproyecto o proyectos multiusuario (PMU) permiten distribuir y organizar los layouts de muchos pequeños proyectos sobre una oblea para su fabricación. Una vez fabricados, la oblea se divide en chips individuales y varias muestras de cada diseño son distribuidas entre los participantes.

Previa a la distribución de los CIs se deben soldar los "bonding PADS" necesarios unicamente a los PADS del circuito correspondiente a cada uno de los participantes. Los chips multiproyecto permiten aún la confidencialidad de los diseños, de tal manera que un participante no conozca la función de los diseños de los otros, pues no tiene acceso a los PADS de los demás circuitos; sin embargo, no se puede garantizar confidencialidad absoluta. La organización de un chip multiproyecto involucra:

- a) Crear el layout de una estructura inicial, en la cual se introducen todos los proyectos.
- b) Coleccionar, distribuir el espacio disponible, redistribuir y finalmente unir los layouts de todos los proyectos, en la estructura inicial creada. Para obtener una distribución óptima y minimizar el tiempo necesario se utilizan diferentes algoritmos de ubicación.

c) Generar el archivo final de descripción de los layouts y documentar los parámetros y especificaciones requeridos por las fundidoras.

El tiempo necesario para disponer de los prototipos fabricados puede reducirse si se utilizan estructuras iniciales estándar. El tiempo requerido para todo el proceso en 1978 <sup>(3)</sup> era de 3 a 5 semanas para la elaboración de las máscaras y de 3 a 4 semanas para la fabricación; un mayor tiempo se requería si habían colas en las fábricas de máscaras. Actualmente todo el proceso se realiza en unas 5 semanas.

El uso de los chips multiproyecto permite una drástica disminución de los costos en el desarrollo de CIs de bajo volumen de producción y de ASICs. Generalmente, dado que en los chips multiproyecto se fabrican solamente prototipos de prueba, que no serán comercializados, no se realiza la "pasivación" ni se prepara la máscara de "overglass". Esto permite ahorro en la preparación de una máscara y un menor número de pasos en la fabricación, que contribuyen a una reducción del costo total de los CIs.

Algunos Centros de Microelectrónica ofrecen servicios de fabricación de prototipos para diseño VLSI, recogiendo diseños de universidades, politécnicas, centros de investigación

---

<sup>3</sup> "Introduction to VLSI systems", Mead C. - Conway L., pág. 128

y desarrollo y firmas comerciales. Esto permite compartir los costos relacionados con la fundición. En este caso los diseños entregados por las instituciones mencionadas se someten al chequeo de las reglas de diseño (DRC) y al chequeo de reglas eléctricas (ERC). Los archivos de descripción de los layouts pueden estar en diferentes formatos, que luego serán convertidos a los utilizados por la fundidora escogida. Todos estos aspectos permiten diseñar CIs a un gran número de diseñadores con una mínima inversión de herramientas.

### 3.6.1 Proyecto multiusuario Iberoamericano

El presente trabajo de Tesis se desarrolla dentro del proyecto "Capacitación en Concepción y Diseño de Circuitos Integrados" enmarcado en el Programa Iberoamericano de Ciencia y Tecnología para el Desarrollo (CYTED-D V CENTENARIO) dentro del Subprograma IX: Microelectrónica y reconocido como Proyecto Multi Usuario (PMU), en el que participa la Escuela Politécnica Nacional. El Centro Nacional de Microelectrónica (CNM) de Barcelona - España es el coordinador del proyecto.

El Programa PMU iberoamericano tiene como finalidad facilitar la formación e investigación en el campo del diseño de CIs de grupos investigadores de carácter universitario, haciendo posible el desarrollo y fabricación de prototipos de circuitos integrados de aplicación específica (ASICs) a bajo coste, al compartir entre los distintos participantes en el

programa los gastos de fabricación. La fundición de los circuitos integrados ha sido convenida con la compañía "European Silicon Structures ES2".

Con la finalidad de recorrer todo el proceso necesario desde la concepción hasta la fabricación y verificación de prototipos de ASICs, se realizó el diseño de un circuito sencillo de prueba, un medio sumador. El archivo CIF de este diseño se envió al CNM.

La Fig. 3.48 describe el esquema operativo del Programa PMU. Las fases de concepción, diseño y pruebas finales de los CIs están a cargo de cada uno de los países participantes, la coordinación del chequeo final de reglas, ensamblado del proyecto global para su posterior fabricación y la distribución de los prototipos, son responsabilidad del Centro Nacional de Microelectrónica de Barcelona - España. La fabricación es realizada por la compañía ES2 (Francia).

A finales de noviembre de 1991, se enviaron el archivo CIF e información adicional del diseño solicitados por el CNM para su fundición en la "corrida" (run) de diciembre de 1991. El diseño enviado cumplió los requerimientos establecidos por el CNM, pasó el DRC de ES2 y fue fundido en diciembre de 1991.

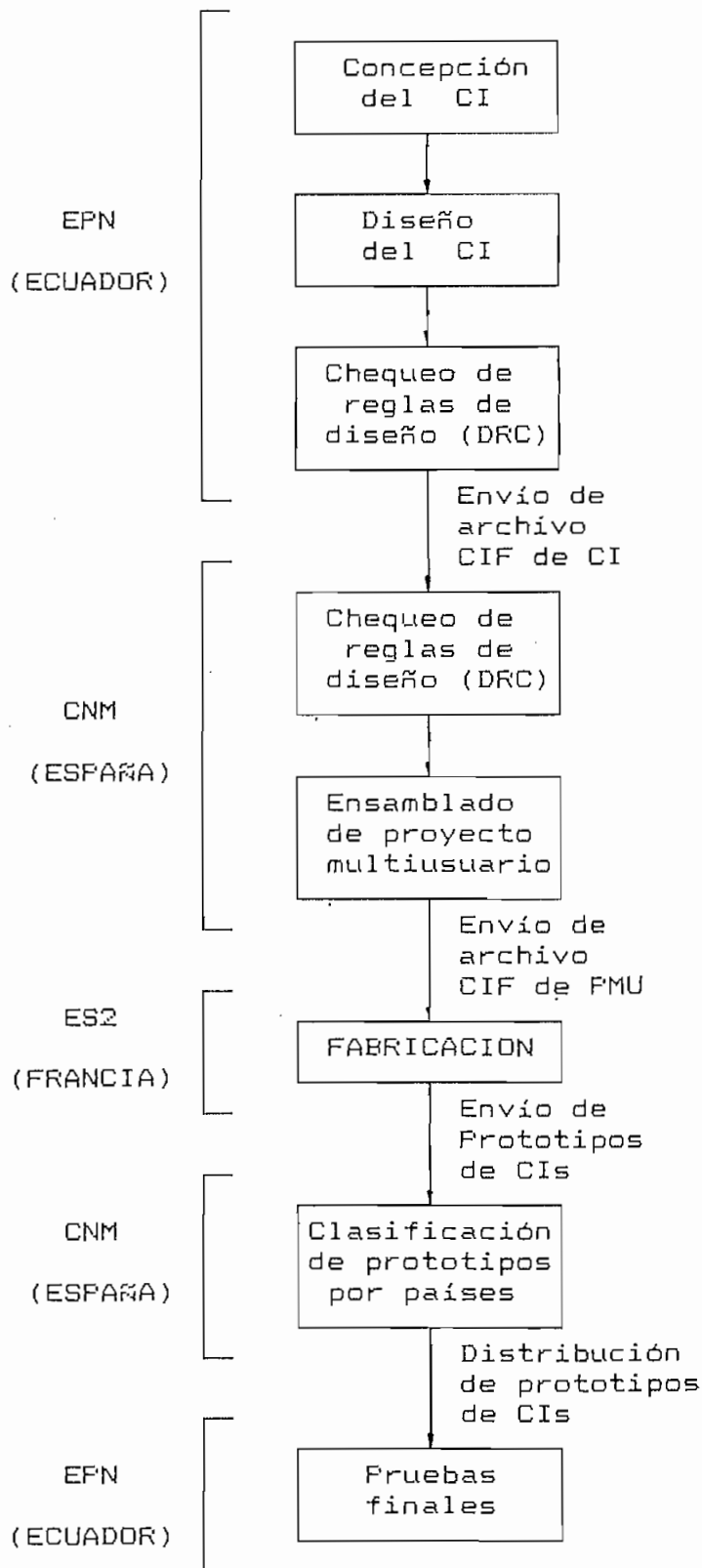


FIGURA 3.48 Esquema operativo del Programa PMU Iberoamericano

En mayo de 1992 se recibieron cinco prototipos del Medio Sumador y fueron probados, encontrándose que eran funcionalmente correctos. La realización del diseño del medio sumador, y del resultado de las pruebas funcionales obtenidos se exponen en detalle en el Cap. 7.

En la Fig. 3.49 se presenta la distribución del área disponible para los diferentes participantes en el proyecto. Este gráfico fue enviado por el CNM junto con los prototipos. La estructura de la esquina superior derecha (Fig.3.49) es e área destinada para la fabricación del medio sumador. En la figura puede observarse fácilmente el núcleo o circuitería interna y los PADs.

La Fig. 3.48 también presenta la conexión de los 6 PADs utilizados hacia los "*bonding PADs*", que permiten el interfaz hacia los pines externos del CI. El empaquetamiento utilizado es del tipo DIP.

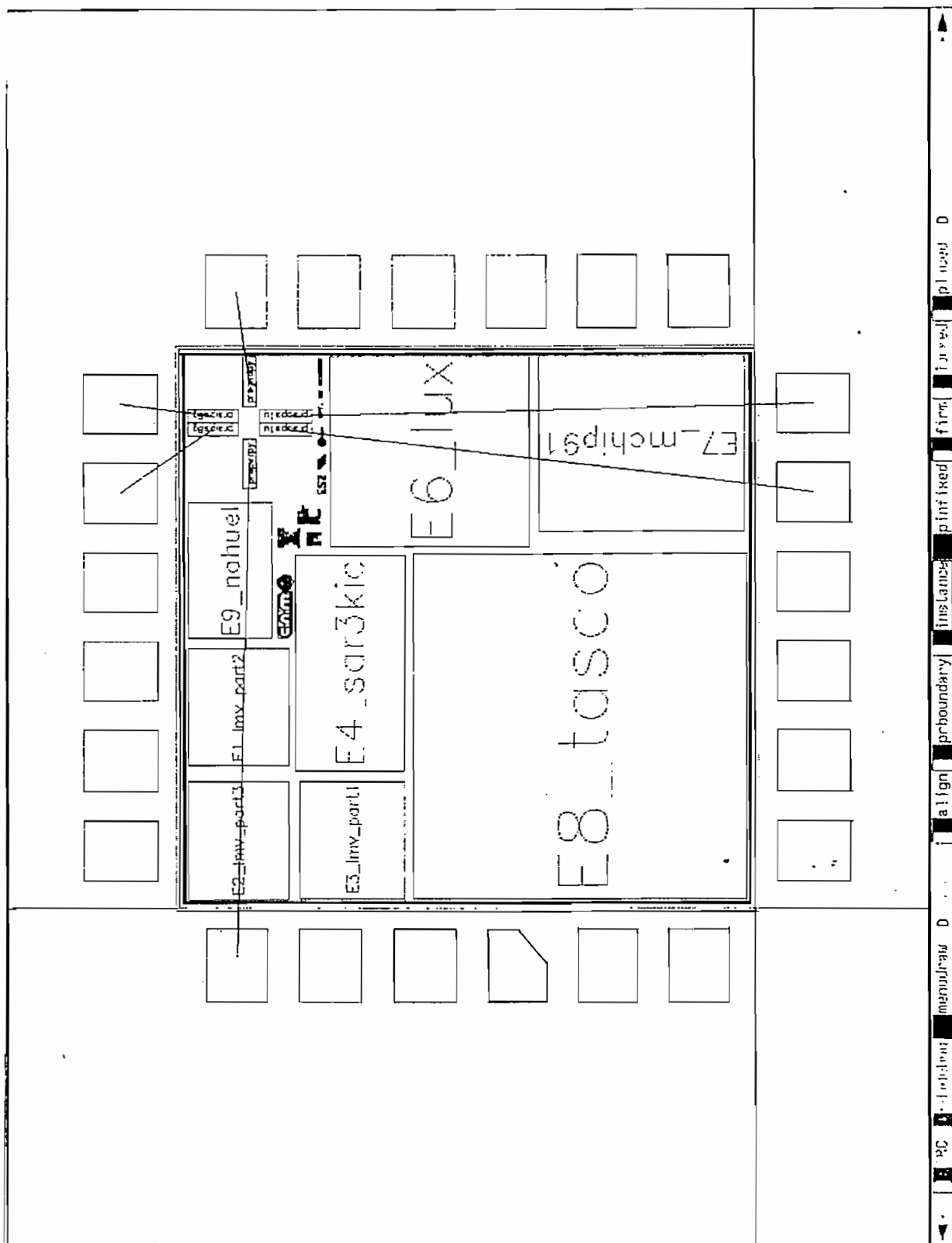


FIGURA 3.49 Distribución del área de la corrida de diciembre de 1991.

## CAPITULO 4

### CARACTERIZACION Y ESTIMACION DEL FUNCIONAMIENTO DE CIRCUITOS INTEGRADOS

El sujetarse a las reglas de diseño garantiza que las formas geométricas resultantes del proceso de fabricación serán muy aproximadas a las originalmente especificadas en los *layouts*, si los parámetros del proceso varían con una tolerancia máxima de  $\lambda$ . Para completar un diseño es necesario realizar una estimación aproximada del comportamiento del circuito (velocidad, potencia, corriente, etc), para lo cual es necesario conocer los parámetros eléctricos de los transistores y de los diferentes materiales que conforman el CI.

Cada capa de material del *chip* tiene una resistencia y capacitancia asociadas, que son componentes fundamentales al estimar el comportamiento del circuito diseñado. También se tiene una inductancia asociada, que por su valor y efectos mínimos, puede ser despreciada. La definición de modelos para determinar estos parámetros eléctricos es fundamental para la estimación de los retardos de las señales y del consumo de potencia.



Las estimaciones que se realizan, a pesar que se utilizan modelos simplificados, intentan proveer al diseñador del suficiente conocimiento de los parámetros fundamentales que afectan el comportamiento del circuito. Este análisis simplificado puede presentar aproximaciones gruesas; sin embargo, su utilidad radica en permitir la comprensión de dichos parámetros, que luego pueden ser incorporados a los modelos de los programas de simulación escritos para manipular circuitos grandes y complejos.

En este capítulo se presentan algunas ideas fundamentales de los parámetros eléctricos y de la manera de calcularlos, para luego utilizarlos en la estimación del comportamiento de los circuitos.

Los ambientes actuales de trabajo para el diseño de CIs incluyen programas que extraen los valores de los parámetros eléctricos a partir de los *layouts* y otros programas permiten simular los circuitos utilizando los parámetros extraídos, facilitando la predicción del comportamiento eléctrico de los circuitos VLSI.

A pesar de que las herramientas computacionales evitan al diseñador el cálculo detallado de los parámetros eléctricos, es importante tener una idea clara de los efectos físicos que representan y de los valores típicos que tendrán dichos parámetros.

## 4.1 ESTIMACION DE RESISTENCIA Y CAPACITANCIA

### 4.1.1 Estimación de la resistencia de regiones rectangulares.

La resistencia eléctrica de una muestra uniforme de material conductor (Fig. 4.1), puede calcularse como:

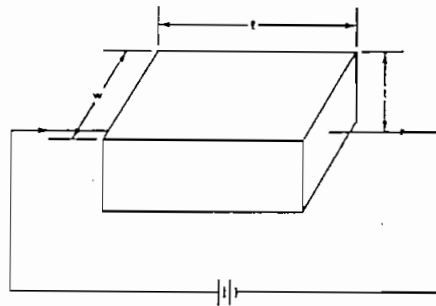


FIGURA 4.1 Resistor simple de cualquier material.

$$R = \left(\frac{\rho}{t}\right) \left(\frac{l}{w}\right)$$

[Ec. 4.1]

en donde:

- $\rho$  = resistividad del material [ $\Omega \cdot \text{cm}$ ]
- $t$  = espesor de la lámina
- $l$  = longitud de la muestra
- $w$  = ancho de la muestra

Si  $l=w$ , se tiene la resistencia de una muestra de forma cuadrada. La Ec. 4.1 se reduce a:  $R = \rho/t$ , valor denominado resistividad de la lámina (*sheet resistivity*) y se denota como  $R_{\square}$  o  $R_{\square}$ . En la última expresión, el valor  $t$  es fijo para una muestra dada,  $\rho$  se expresa en  $[\Omega \cdot \text{cm}]$  y  $R_{\square}$  se expresa en  $[\Omega]$ , aunque generalmente se lo hace en  $[\Omega/\square]$ . La Ec. 4.1 puede escribirse en función de  $R_{\square}$  como:

$$R = R_{\square} \left( \frac{l}{w} \right)$$

[Ec. 4.2]

La definición de  $R_{\square}$  permite obtener de manera aproximada la resistencia de muestras de cualquier material que presente formas rectangulares, simplemente multiplicando el valor de  $R_{\square}$  por la relación de la longitud al ancho de la muestra.

Es importante notar que en una muestra dada, "l" es la dimensión que corre paralela con el flujo de corriente, mientras que "w" es transversal a dicho flujo.

La resistencia de regiones no rectangulares, pero limitadas por líneas horizontales y verticales puede calcularse dividiendo la muestra en rectángulos de longitud  $l_1$  y ancho  $w_1$ , conectados en serie. Lo que puede expresarse como:

$$R = R_0 \cdot \sum \left( \frac{l_i}{w_i} \right)$$

[Ec. 4.3]

En la última ecuación, cada sumando puede ser el resultado de la combinación de elementos en paralelo. Así por ejemplo en la Fig. 4.2a se presenta una región de  $l=1$  y  $w=3$ , la resistencia puede calcularse directamente multiplicando  $R_0$  por la relación  $l/w = 1/3$ .

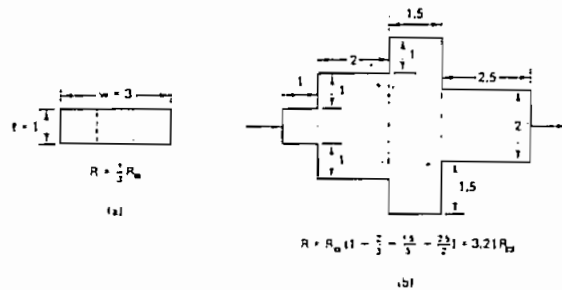


FIGURA 4.2 Cálculo de resistencia: a) Simple rectángulo b) región arbitraria.

Aplicando la Ec. 4.3, se consideran tres regiones de dimensiones  $(l,w)$ , por cada una de las cuales circula la corriente en el sentido de "1" y están por lo tanto conectadas en paralelo. El resultado en este caso es también  $R_0(l/w)$ .

En la Fig. 4.2b se presenta una forma más irregular, en ella la corriente circula en sentido horizontal y la resistencia total es igual a la suma de las resistencias de cuatro

rectángulos (indicados por las líneas entrecortadas), conectados en serie. El valor resultante es  $R = 3.21 R_s$ , como se indica en la Fig. 4.2b.

Las Ecs. 4.2 y 4.3 no son exactas pues no toman en cuenta posibles variaciones de  $\rho$  con la profundidad y cambios en los valores de "l" y "w" debido al proceso de fabricación, especialmente en las regiones creadas por difusión.

La Tabla 4.1 presenta valores típicos de  $R_s$  que se encuentran en procesos MOS de 3 a 5  $\mu\text{m}$ . Los valores de esta tabla son muy aproximados y pueden utilizarse como valores promedios para el cálculo de diferentes parámetros que ayuden a evaluar el comportamiento de un circuito.

Material	$R_s$ [ $\Omega/\square$ ]		
	Mínimo	Típico	Máximo
Metal (Al)	0.03	0.05	0.08
Difusión(n+, p+)	10	25	50
Polysilicon	15	50	100

TABLA 4.1 Valores típicos de  $R_s$  para materiales conductores

Mientras que la resistividad del metal es conocida para un valor "t" dado, la resistividad del *polysilicon* y difusión están determinados también por la concentración de impurezas. En otras palabras, para los semiconductores se deben conocer

los parámetros del proceso para estimar las características eléctricas de manera precisa.

Hasta el momento se han considerado solamente regiones rectangulares y que pueden calcularse tomando como base regiones rectangulares; para capas que tengan regiones de formas diferentes, la estimación de los valores de resistencia requiere cálculos más elaborados <sup>(1)</sup>

#### 4.1.2 Estimación de resistencia en transistores

La característica voltaje-corriente de un transistor MOS es no-lineal, a excepción de una pequeña región de operación. Un transistor utilizado como transistor de paso trabaja en la región lineal de su característica, pero un transistor utilizado en el *pull-up* o *pull-down* de una compuerta, varía su resistencia significativamente, ya que conmuta rápidamente entre las regiones de corte y saturación.

A pesar de las variaciones indicadas, a menudo es útil aproximar el comportamiento de un transistor en términos de la "resistencia del canal", una resistencia efectiva que permite estimar el comportamiento del transistor.

---

<sup>1</sup> "Introduction to nMOS and CMOS VLSI Systems Design", Mukherjee Amar, págs. 157 a 159.

Como se demostró en el Cap. 2, para la región lineal la resistencia del canal está dada por:

$$R_{c(\text{lineal})} = \frac{t_{ox}}{\mu c} \left( \frac{L}{W} \right) \frac{1}{(V_{gs} - V_t)}$$

[Ec. 4.4]

La Ec. 4.4 puede escribirse como:

$$R_{c(\text{lineal})} = k \left( \frac{L}{W} \right)$$

[Ec. 4.5]

En donde  $k$  representa ahora la resistividad de lámina del canal. Para transistores canal "n" y "p",  $k$  puede tomar un valor comprendido entre 5000 y 30000  $\Omega/\square$ .

Modelos más complejos y exactos definen varias resistencias efectivas para el transistor. Estas son: la resistencia estática (resistencia efectiva del canal) y resistencias dinámicas que corresponden a las corrientes transitorias que circulan durante las transiciones de la entrada de alto a bajo y viceversa <sup>(2)</sup>

---

<sup>2</sup> "Simulation tools for Digital LSI Design", Terman C.

### 4.1.3 Características de un capacitor MOS

Primeramente se explica el funcionamiento de un capacitor MOS, lo que permitirá entender fácilmente las capacitancias presentes en la estructura de un transistor MOS.

La característica capacitancia-voltaje de una estructura MOS depende de las condiciones de carga de la superficie del sustrato: acumulación, agotamiento e inversión, determinadas por el voltaje de la compuerta.

En la Fig. 4.3a se presenta una estructura MOS de sustrato "p" y en la Fig. 4.3b su característica capacitancia-voltaje, su comportamiento se explica para cada una de las condiciones de la superficie del sustrato.

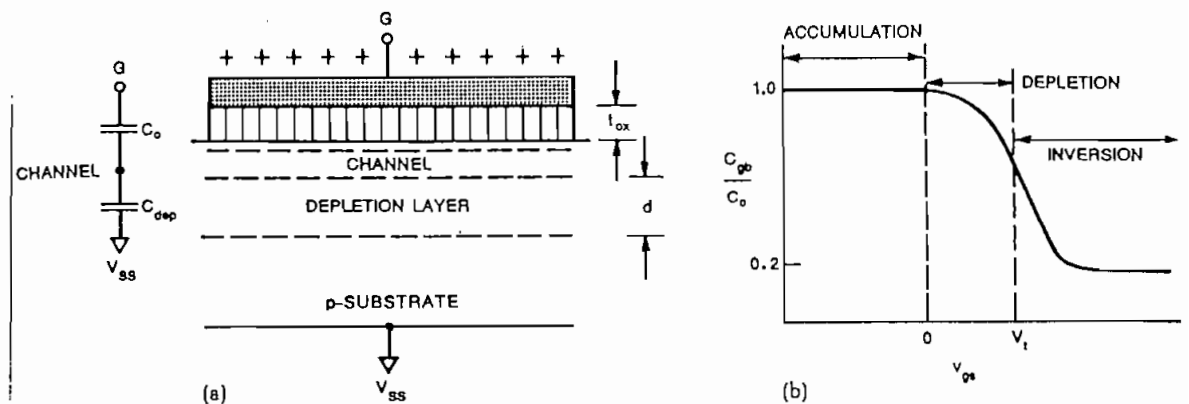


FIGURA 4.3 Estructura de un capacitor MOS.



a) Condición de acumulación.

Con un voltaje de compuerta negativo, la carga negativa de la compuerta atrae a los huecos del sustrato hacia la superficie límite con el  $\text{SiO}_2$ , formando una capa de acumulación de huecos. La estructura resultante se comporta como un capacitor de placas paralelas. La compuerta de *polysilicon* constituye la una placa y la capa de acumulación del sustrato la otra, el aislante es el Oxido de Silicio. Bajo estas condiciones la capacitancia compuerta-sustrato ( $C_{gs}$ ) está dada por el valor de la capacitancia del Oxido de Silicio ( $C_o$ ), cuyo valor es:

$$C_o = \left( \frac{\epsilon_{\text{SiO}_2} \epsilon_o}{t_{\text{ox}}} \right) \cdot A$$

[Ec. 4.6]

En donde:

A = área de la compuerta.

$\epsilon_{\text{SiO}_2}$  = permitividad relativa del  $\text{SiO}_2$   
(=3.9).

$t_{\text{ox}}$  = espesor del  $\text{SiO}_2$ .

$\epsilon_o$  = permitividad del vacío.

En la Fig. 4.3b se grafica la relación  $C_{gs}/C_o$  en función del voltaje de compuerta. Para la condición de acumulación, puede verse que la capacitancia de la compuerta está dada por el valor de  $C_o$ .

b) Condición de agotamiento.

Con un pequeño voltaje positivo en la compuerta (menor a  $V_t$ ), se repelen los huecos del sustrato dejando una región de iones cargada negativamente y libre de portadores. Con el pequeño voltaje de compuerta se tiene ya un canal de espesor muy pequeño, pero que junto con la capa de agotamiento (*depletion*) formada y el sustrato bajo la región de agotamiento dan origen a un nuevo capacitor de placas paralelas ( $C_{d-p}$ ), que está en serie con el valor de  $C_o$ , como se indica en la Fig. 4.3a. El valor de esta capacitancia está dado por:

$$C_{dep} = \left( \frac{\epsilon_{si} \epsilon_0}{d} \right) \cdot A$$

[Ec. 4.7]

En donde:

$\epsilon_{si}$  = permitividad relativa del Si, (= 12)  
 $d$  = profundidad de la capa de agotamiento.

La combinación serie de  $C_o$  con  $C_{d-p}$ , proporciona una capacidad resultante ( $C_{ob}$ ) menor a  $C_o$ , como se esquematiza en la Fig. 4.3b.

Los incrementos de voltaje de cero a  $V_t$ , producen una mayor profundidad de la capa de agotamiento, disminuyendo el valor de  $C_{d-p}$  y por lo tanto el de  $C_{ob}$ .

c) Condición de inversión.

Con voltajes de compuerta mayores a  $V_t$ , se crea una capa de fuerte inversión (canal), de alta conductividad, que retorna la capacitancia  $C_{gb}$  al valor  $C_o$ .

Se debe mencionar que en la estructura MOS en estudio, los electrones que forman el canal se originan por procesos de generación-recombinación en la superficie de la capa de agotamiento. Este mecanismo de generación no provee instantáneamente los electrones para la capa de inversión; por lo que si se aplican a la compuerta voltajes de frecuencia alta, el canal puede no formarse y desaparecer con la misma rapidez y en promedio el dispositivo parece estar en la condición de agotamiento, lo que produce la aparición de  $C_{d-p}$  y la reducción de  $C_{gb}$ .

Se ha determinado que si la frecuencia del voltaje de compuerta utilizado es mayor a 100 Hz, la capacitancia permanece en el valor correspondiente a la máxima condición de agotamiento alcanzada (Fig. 4.3b).

Si la frecuencia es menor a los 100 Hz, los electrones se obtienen de los procesos de generación-recombinación en cantidad suficiente para seguir a la señal aplicada a la compuerta. El valor de capacitancia  $C_{gb}$ , en estas condiciones, se aproxima al valor de  $C_o$ , como se indica en la Fig. 4.4.

- b) La capacitancia de las regiones de difusión de los drenajes conectados a su salida.
- c) La capacitancia de los caminos de interconexión entre su salida y las entradas de las otras compuertas.

A continuación se encuentran modelos y se presentan valores típicos de cada una de las componentes de capacitancia, para finalmente presentar un ejemplo en el que se utilizan sus valores típicos.

a) Capacitancia del transistor MOS

En la Fig. 4.5 se presenta un transistor MOS con sus capacitancias parásitas asociadas. El transistor MOS tiene en su compuerta la estructura de un capacitor MOS, pero se forman también capacitores similares, de placas paralelas, entre los diferentes materiales del transistor.

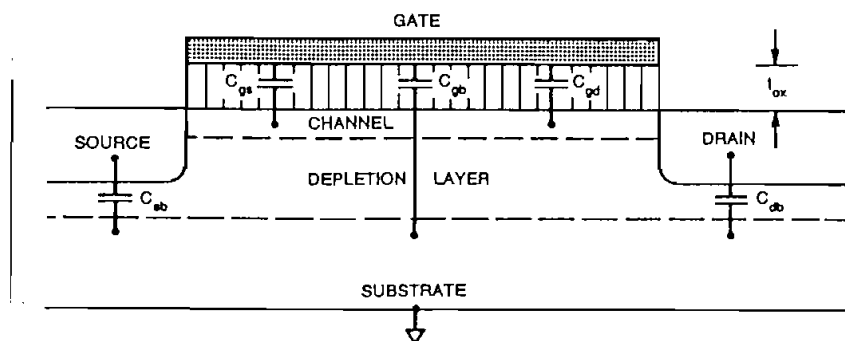


FIGURA 4.5 Representación de las capacidades parásitas de un transistor MOS.

El mismo análisis aplicado al capacitor MOS puede aplicarse a un transistor MOS para determinar el valor de  $C_{gb}$ . Para la condición de inversión, se debe considerar que en el transistor están presentes las regiones de drenaje y fuente, éstas proporcionan fácilmente los electrones para el canal ya que tienen altas concentraciones de electrones. Por lo tanto, para altas y bajas frecuencias, la capacitancia  $C_{gb}$  del transistor es también la representada en la Fig. 4.4.

$C_{gs}$  y  $C_{gd}$  se forman si existen regiones de sobrelapamiento entre compuerta-fuente y compuerta-drenaje, respectivamente. Aún si el proceso es "auto alineado" (Cap. 3), existen siempre pequeños valores de estas capacitancias. Además, cuando se forma el canal, se lo puede mirar como una extensión física de las regiones de drenaje y fuente. En este caso,  $C_{gs}$  y  $C_{gd}$  representan las capacitancias concentradas en las regiones del canal próximas a la fuente y drenaje, respectivamente; sin embargo, en la región de saturación debido al estrangulamiento del canal  $C_{gd} = 0$ .

En la Fig. 4.5 se representan también  $C_{sb}$ ,  $C_{db}$ , que son las capacidades de las difusiones de fuente y drenaje al sustrato. El cálculo de estas capacitancias se presenta en la parte b) de este numeral.

La representación de las capacidades parásitas a nivel circuital se presenta en la Fig. 4.6.

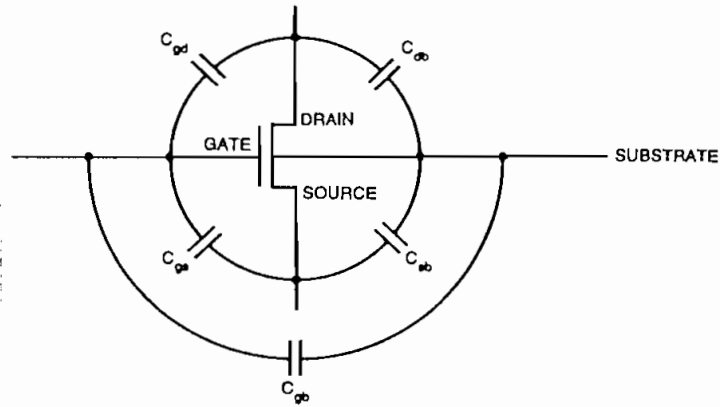


FIGURA 4.6 Representación circuital de las capacidades parásitas de un transistor MOS.

Para realizar estimaciones de la capacitancia total de la compuerta, generalmente se asume que  $C_{db} = C_{sb}$ , aproximación que estaría despreciando la región en donde  $0 < V_{gs} < V_t$  (Fig. 4.4); pero debido a que los transistores en los circuitos digitales pasan rápidamente por esta región, esta aproximación proporciona buenos resultados.

Con esta condición, se tienen en paralelo  $C_{gd}$ ,  $C_{gs}$  y  $C_{gb}$ . Por lo tanto, la capacitancia de entrada o capacitancia total de la compuerta ( $C_g$ ) de un transistor MOS está dada por:

$$C_g = C_{gs} + C_{gd} + C_{gb}$$

[Ec. 4.8]

Los valores de  $C_{gd}$  y  $C_{gs}$  son pequeños en comparación con  $C_{gb}$ , y también suelen despreciarse. En conclusión, se puede asumir que la capacitancia de un transistor es constante e igual a la del óxido de la compuerta ( $C_{ox}$ ).

El valor de  $C_g$  puede expresarse solamente en función de las dimensiones de los transistores, para esto:

$$C_g = C_{ox} \cdot A$$

[Ec. 4.9]

en donde  $C_{ox}$  es la capacitancia del óxido de la compuerta por unidad de área, dado por:

$$C_{ox} = \frac{\epsilon_o \epsilon_{SiO_2}}{t_{ox}}$$

[Ec. 4.10]

Con un espesor del óxido de la compuerta en el rango de 500 a 1000 Å, y una permitividad relativa aproximada de 4, el valor de  $C_{ox}$  es:

$$C_{ox} = \frac{4 * 8.854 * 10^{-14}}{(500 - a - 1000) * 10^{-9}}$$

[Ec. 4.11]

$$C_{ox} \approx (0.8 - a - 0.4) \text{ fF}/\mu\text{m}^2$$

[Ec. 4.12]

Como ejemplo, para el transistor de la Fig. 4.7, con  $\lambda = 2\mu\text{m}$  y  $t_{ox} = 1000\text{Å}$ , el valor de la capacitancia de la compuerta o capacitancia de entrada es:

$$C_g = 3.5 * 10^{-4} * 8 \lambda^2 \text{ pF} = 11.2 \text{ fF}$$

[Ec. 4.13]

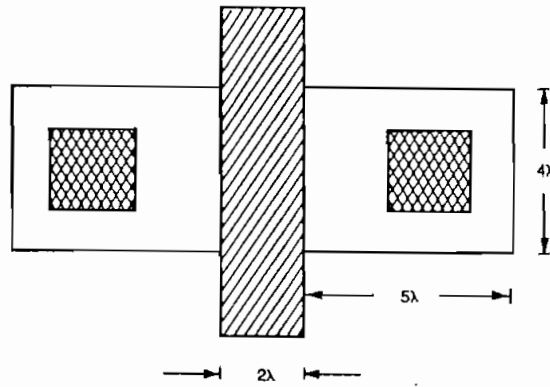


FIGURA 4.7 layout de un transistor para determinar  $(C_d)$ .

#### b) Capacitancia de difusión

Las regiones de difusión "p+" y "n+" que forman los transistores o caminos de interconexión, tienen una capacitancia asociada respecto al sustrato (o pozo). El valor de la capacitancia depende de la diferencia de potencial entre las regiones de difusión y el sustrato (o pozo), así como del área de la región de agotamiento que separa las regiones de difusión y el sustrato (o pozo).

La capacitancia de difusión ( $C_d$ ) es proporcional al área total de la juntura difusión-sustrato; es decir, al área de la juntura en la base de la difusión ( $C_{d,b}$ ) y al área de la juntura de toda la periferia de las paredes de la difusión ( $C_{d,p}$ ), debida a la profundidad finita de la difusión. En la Fig. 4.8 se presentan las regiones de difusión de un transistor y las dos componentes de la capacitancia  $C_d$ , las dimensiones de las regiones de difusión son "a" y "b".



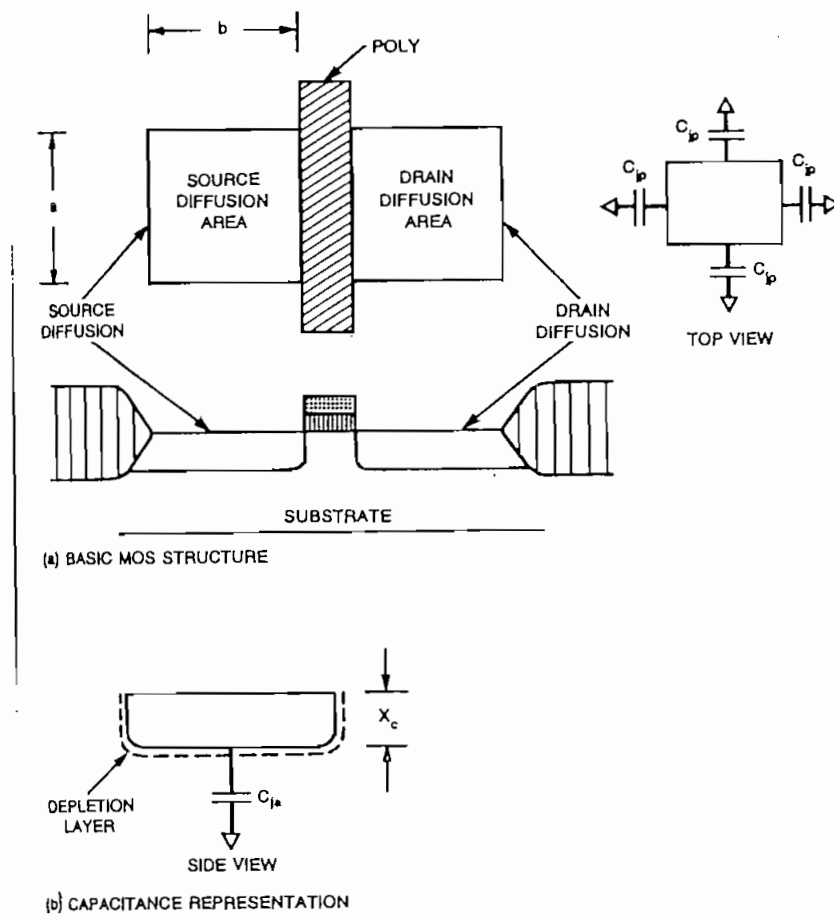


FIGURA 4.8 Componentes de la capacitancia de difusión debidas al área y a la periferia.

La capacitancia de las paredes puede caracterizarse por una capacitancia por unidad de longitud de la periferia.

La capacitancia total  $C_d$  puede expresarse como:

$$C_d = C_{j_a} * (ab) + C_{j_p} * (2a + 2b)$$

[Ec. 4.14]

En donde:

- $C_{j_n}$  = capacitancia de la juntura /  $\mu\text{m}^2$   
 $C_{j_p}$  = capacitancia de la periferia /  $\mu\text{m}$   
 $a$  = ancho de la región de difusión  
 $b$  = extensión de la región de difusión

Valores típicos para  $C_{j_n}$  y  $C_{j_p}$  se presentan en la Tabla 4.2, para transistores "p" y "n".

	Dispositivo/camino "n"	Dispositivo/camino "p"
$C_{j_n}$	$1 \cdot 10^{-4}$ pF/ $\mu\text{m}^2$	$1 \cdot 10^{-4}$ pF/ $\mu\text{m}^2$
$C_{j_p}$	$9 \cdot 10^{-4}$ pF/ $\mu\text{m}$	$8 \cdot 10^{-4}$ pF/ $\mu\text{m}$

TABLA 4.2 Valores típicos de las capacitancias de difusión

Por ejemplo para el transistor de la Fig. 4.7, de dimensiones  $a = 8 \mu\text{m}$  y  $b = 10 \mu\text{m}$ , la capacitancia del área de difusión del drenaje, que constituye la capacitancia de salida del transistor, es:

$$C_d = 1 \cdot 10^{-4} \left[ \frac{\text{pF}}{\mu\text{m}^2} \right] (10 \cdot 8) [\mu\text{m}^2] + 9 \cdot 10^{-4} \left[ \frac{\text{pF}}{\mu\text{m}} \right] (20 + 16) [\mu\text{m}] = 40 \text{ fF}$$

[Ec.4.15]

c) Capacitancias de enrutamiento

Como se justifica en el numeral 4.1.6 los materiales utilizados para el enrutamiento son el metal y el polysilici-

con. Estos caminos de enrutamiento forman también capacitores con el sustrato, que pueden representarse con el modelo de placas paralelas, una placa está formada por un camino de metal o *polysilicon* y la otra por el sustrato, teniendo como aislante al óxido de campo (FOX).

Sin embargo, el modelo de placas paralelas ignora el efecto de los campos periféricos (Fig. 4.9), conocidos como "*fringing fields*". El resultado neto de este efecto es incrementar el área efectiva de las placas y con ello los caminos de metal y *polysilicon* tienen en realidad una capacitancia más alta, que puede ser de hasta dos veces el valor previamente estimado.

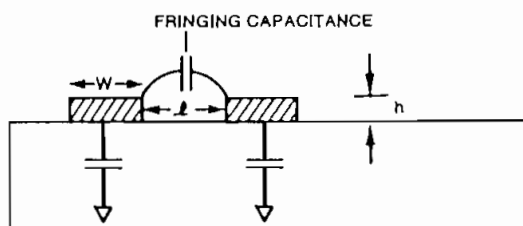


FIGURA 4.9 Efecto de los "*fringing fields*" en la capacitancia.

#### d) Guía para el cálculo de las capacitancias

En la Tabla 4.3 se presentan valores típicos para un proceso CMOS de compuerta de Silicio de  $4 \mu\text{m}$  ( $\lambda=2 \mu\text{m}$ ). En esta tabla se resumen todos los valores de capacitancia estudiados anteriormente, estos son:

- i) La capacitancia de entrada de una compuerta ( $C_{in}$ );
- ii) La capacitancia de las regiones de difusión ( $C_{jn}$ ,  $C_{jp}$ );
- iii) Las capacitancias de los caminos de enrutamiento:

- Polysilicon sobre FOX
- Metal sobre polysilicon
- Metal sobre FOX
- Metal sobre difusión

La tabla puede utilizarse como guía para elegir el material de los caminos, pero debe señalarse que no incluye el efecto de los "fringing fields".

PARAMETER	MIN.	MAX.	COMMENTS
$C_g(\text{pF}/\mu\text{m}^2)$	$4.0 * 10^{-4}$	$5.0 * 10^{-4}$	Gate
$C_p(\text{pF}/\mu\text{m}^2)$	$0.4 * 10^{-4}$	$0.6 * 10^{-4}$	Polysilicon over field
$C_{mp}(\text{pF}/\mu\text{m}^2)$	$0.4 * 10^{-4}$	$0.6 * 10^{-4}$	Metal over poly
$C_m(\text{pF}/\mu\text{m}^2)$	$0.15 * 10^{-4}$	$0.3 * 10^{-4}$	Metal over field
$C_{md}(\text{pF}/\mu\text{m}^2)$	$0.8 * 10^{-4}$	$1.0 * 10^{-4}$	Metal over diffusion; ( $p^+$ AND $n^+$ )
$C_{\mu n}(\text{pF}/\mu\text{m}^2)$	$0.8 * 10^{-4}$	$1.0 * 10^{-4}$	n-diffusion
$C_{\mu p}(\text{pF}/\mu\text{m}^2)$	$0.8 * 10^{-4}$	$1.0 * 10^{-4}$	p-diffusion
$C_{jn}(\text{pF}/\mu\text{m})$	$7.0 * 10^{-4}$	$9.0 * 10^{-4}$	n-channel device
$C_{jp}(\text{pF}/\mu\text{m})$	$6.0 * 10^{-4}$	$8.0 * 10^{-4}$	p-channel device

TABLA 4.3 Valores típicos de capacitancias de un proceso de 4  $\mu\text{m}$  de compuerta de Si

En la Tabla 4.4 se presentan valores típicos de las capacitancias para un proceso de doble metal. Se incluyen los valores de capacitancia de:

- Metal 2 a sustrato
- Metal 2 a *polysilicon*
- Metal 2 a metal 1

PARAMETER	MIN.	MAX.	COMMENTS
$C_{m2}(pF/\mu m^2)$	$0.1 * 10^{-4}$	$0.15 * 10^{-4}$	Metal 2 to substrate
$C_{m2p}(pF/\mu m^2)$	$0.2 * 10^{-4}$	$0.3 * 10^{-4}$	Metal 2 to poly
$C_{m21}(pF/\mu m^2)$	$0.3 * 10^{-4}$	$0.5 * 10^{-4}$	Metal 2 to metal 1

TABLA 4.4 Valores típicos de capacitancias de un proceso de  $4 \mu m$  con metal 2

En base a la Fig. 4.10, se calculan las capacitancias parásitas de un camino de metal que se interconecta con la compuerta de un transistor. Se utiliza el valor de  $\lambda = 2 \mu m$  y los valores máximos de la Tabla 4.3.

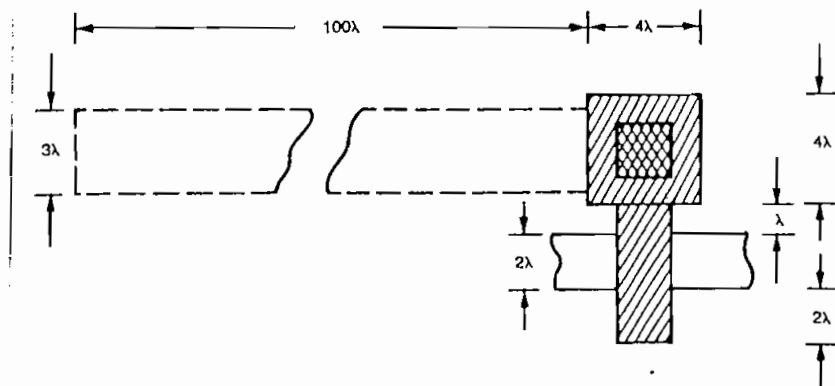


FIGURA 4.10 Ejemplo de cálculo de capacitancias parásitas.

Para el metal:

$$C_{m2p} = (3\lambda * 100\lambda) 0.3 * 10^{-4} = 0.036 pF$$

[Ec. 4.16]

Para el *polysilicon*:

$$C_p = [(4\lambda * 4\lambda) + (\lambda + 2\lambda) * 2\lambda] 0.6 * 10^{-4} = 0.0053 pF$$

[Ec. 4.17]

Para el área de la compuerta:

$$C_g = [2\lambda * 2\lambda] 5.0 * 10^{-4} = 0.008 pF$$

[Ec. 4.18]

La capacitancia total es por lo tanto:

$$C_T = C_{in} + C_p + C_g = 0.049 pF$$

[Ec. 4.19]

De esta manera se estiman los valores de capacitancia sin recurrir a análisis profundos.

#### 4.1.5 Efectos distribuidos RC

La propagación de una señal a lo largo de un camino depende de la resistencia y capacitancia distribuidas del camino, la impedancia de la fuente de la señal y la impedancia de carga. Para caminos largos predominan los retardos de propagación debidos a la resistencia y capacitancia (RC) distribuidas. Este efecto de transmisión es particularmente severo en caminos de *polysilicon* debido a su resistencia relativamente alta.

Un camino largo puede representarse por varias secciones RC, como se indica en la Fig. 4.11. La respuesta temporal del nodo  $V_j$  está dada por:

$$C \frac{dV_j}{dt} = (I_{j-1} - I_j)$$

[Ec. 4.20]

$$C \frac{dV_j}{dt} = \frac{(V_{j-1} - V_j)}{R} - \frac{(V_j - V_{j+1})}{R}$$

[Ec. 4.21]

Un análisis discreto del circuito de la Fig. 4.11 conduce a un retardo aproximado de la señal de:

$$t_n = \frac{RCn(n+1)}{2}$$

[Ec. 4.22]

siendo "n" el número de secciones.

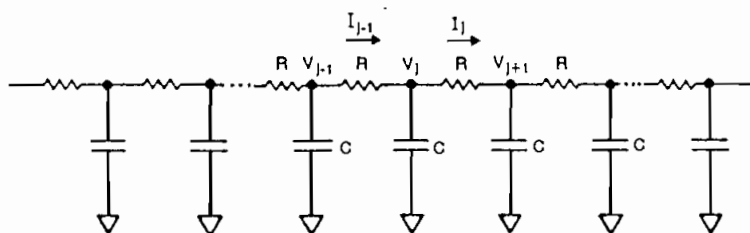


FIGURA 4.11 Representación de un camino largo con secciones RC distribuidas.

Si "n" se hace muy grande, la Ec. 4.22 se reduce a:

$$t_1 = \frac{rcl^2}{2}$$

[Ec. 4.23]

En donde:

l = la longitud del camino.

r = resistencia por unidad de longitud.

c = capacitancia por unidad de longitud.

Para reducir el retardo resultante de la señal en un camino largo de *polysilicon*, una posible estrategia es dividirlo en varias secciones e insertar *buffers* entre ellas. Por ejemplo, en la Fig. 4.12 se presenta un camino de 2mm de *polysilicon*, dividido en dos secciones de 1 mm. Para valores de  $r = 12 \Omega/\mu\text{m}$  y  $c = 4 \times 10^{-4} \text{ pf}/\mu\text{m}$ , la Ec. 4.23 entrega un resultado de 9.6 ns como retardo total.

Utilizando el *buffer*, en el camino de 1 mm se tiene un retardo de 2.4 ns, sumado el retardo del *buffer*. El valor del retardo del *buffer* está en el orden de los 2 ns. El retardo total en este caso es  $4.8 \text{ ns} + 2 \text{ ns} = 6.8 \text{ ns}$ .

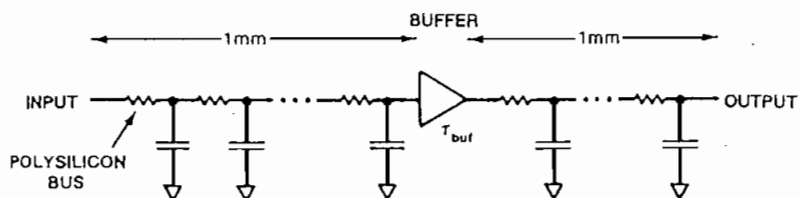


FIGURA 4.12 Segmentación de un camino de *polysilicio*.



El ejemplo presentado indica que con una adecuada segmentación de un camino y con la consecución de pequeños valores para el retardo del *buffer*, se pueden conseguir mejoras significativas.

Otra solución más satisfactoria la presentan tecnologías de fabricación que utilizan dos niveles de metal. Este segundo nivel de metal puede utilizarse para llevar las señales que originalmente lleva el *polysilicon*. El uso del *polysilicon* se reserva para las conexiones locales entre compuertas.

#### 4.1.6 Guía de diseño para la longitud de los caminos

Para propósitos de análisis de temporización un nodo puede definirse como la región a la cual se conectan diversos caminos y en el que el retardo de la señal es pequeño en comparación con los retardos de las compuertas.

Para caminos de longitudes lo suficientemente pequeños, los retardos producidos por los parámetros RC pueden ser despreciados. De esta manera, los caminos pueden tratarse como nodos y se los puede representar como simples cargas capacitivas. Por lo tanto, es útil definir reglas eléctricas sencillas que sirvan de guías para determinar las longitudes máximas de los caminos de interconexión para los diferentes materiales.

Para conseguir el objetivo planteado se requiere que el retardo del camino ( $\tau_w$ ) y el retardo de una compuerta ( $\tau_g$ ), cumplan la siguiente relación:

$$\tau_w \ll \tau_g$$

[Ec. 4.24]

Sustituyendo el valor de  $\tau_w$  de la Ec. 4.23, se obtiene:

$$l \ll \sqrt{\frac{2\tau_g}{RC}}$$

[Ec. 4.25]

La última relación establece un límite superior para la longitud permitida de los caminos de interconexión.

Por ejemplo, para un camino de interconexión de Aluminio, utilizando  $\tau_g$  con un valor de 2 ns y los valores de las Tablas 4.1 ( $R_{metal} = 0.03 \Omega/\square$ ) y 4.4 ( $C_{int} = 0.3 \times 10^{-4} \text{ pF}/\mu\text{m}^2$ ), tomados con respecto al valor  $l_{lambda}$ , se tiene:

$$l = \sqrt{\frac{2 \cdot 2.0 \cdot 10^{-9}}{\left(\frac{0.03}{\lambda}\right) \left(\frac{0.3 \cdot 10^{-16}}{\lambda}\right)}}$$

[Ec. 4.26]

que proporciona un valor de:

$$l=60000\lambda$$

[Ec. 4.27]

Estableciendo un límite muy conservador, se llega a que un camino de metal no debe exceder 20000  $\lambda$ .

En la Tabla 4.5 se señalan las reglas que deberían considerarse para determinar las longitudes máximas de los caminos con diferentes materiales para un proceso típico CMOS. Los valores pueden derivarse de manera similar a la realizada para el Al. Puede notarse claramente la diferencia entre los valores permitidos para caminos de metal, *polysilicon* y difusión. De ahí que el material utilizado preferentemente para el enrutamiento es el metal, y de ser necesario el *polysilicon*.

NIVEL	MAXIMA LONGITUD
METAL	20000 $\lambda$
POLYSILICON	200
DIFUSION	20

TABLA 4.5 Longitudes de los caminos para ignorar los retardos debido a RC

c) Distribución de  $V_{DD}$  y Gnd.

El requerimiento especial del enrutamiento de  $V_{DD}$  y Gnd es que todo componente activo del circuito debe tener una

conexión directa a estos caminos. Los caminos de  $V_{DD}$  y Gnd, nunca se realizan con *polysilicon*, debido a su alta resistividad que produce excesivas caídas de voltaje. Los dos caminos de polarización están hechos en metal y del ancho suficiente para evitar el efecto de migración.

Una estrategia muy común para enrutar los caminos de polarización es utilizar la "interdigitación", como se indica en la Fig. 4.13. En esa figura  $D_1, D_2, \dots, D_n$  son componentes activos. Con la estrategia planteada se puede distribuir la polarización a todo elemento que lo requiera.

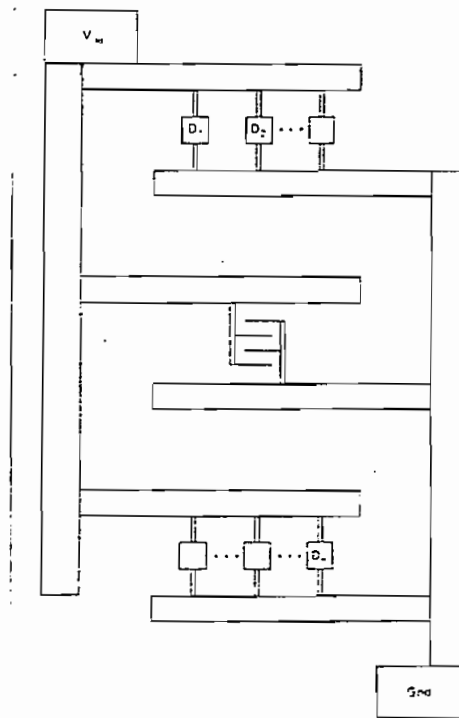


FIGURA 4.13 Estrategia para la distribución de  $V_{DD}$  y Gnd.

## 4.2 CARACTERISTICAS DE CONMUTACION Y RETARDOS

La velocidad de conmutación de una compuerta CMOS está limitada por el tiempo que toma cargar (hacia  $V_{DD}$ ) y descargar (a Gnd.) la capacitancia de carga ( $C_L$ ).

A continuación se desarrollan modelos que permiten determinar la característica de conmutación del inversor CMOS. Como primer paso se definen algunos términos:

- a) Tiempo de subida ( $t_r$ ) = tiempo requerido por una señal para subir del 10% al 90% de su valor estacionario.
- b) Tiempo de bajada ( $t_f$ ) = tiempo requerido por una señal para disminuir del 90% al 10% de su valor estacionario.
- c) Tiempo de retardo ( $t_d$ ) = el tiempo requerido para que una transición lógica pase de la entrada a la salida. Es la diferencia de tiempos entre la transición de entrada (50% del nivel) y el 50% del nivel de salida.

En la Fig. 4.14a se presenta un inversor CMOS con una capacitancia de carga ( $C_L$ ), conformada por las capacitancias de entrada de las otras compuertas, la de salida del inversor y la del enrutamiento. Para el análisis se aplica una forma de onda paso  $V_{in}(t)$  y a la salida se observa  $V_o(t)$ . Estas dos señales se presentan en la Fig. 4.14b.

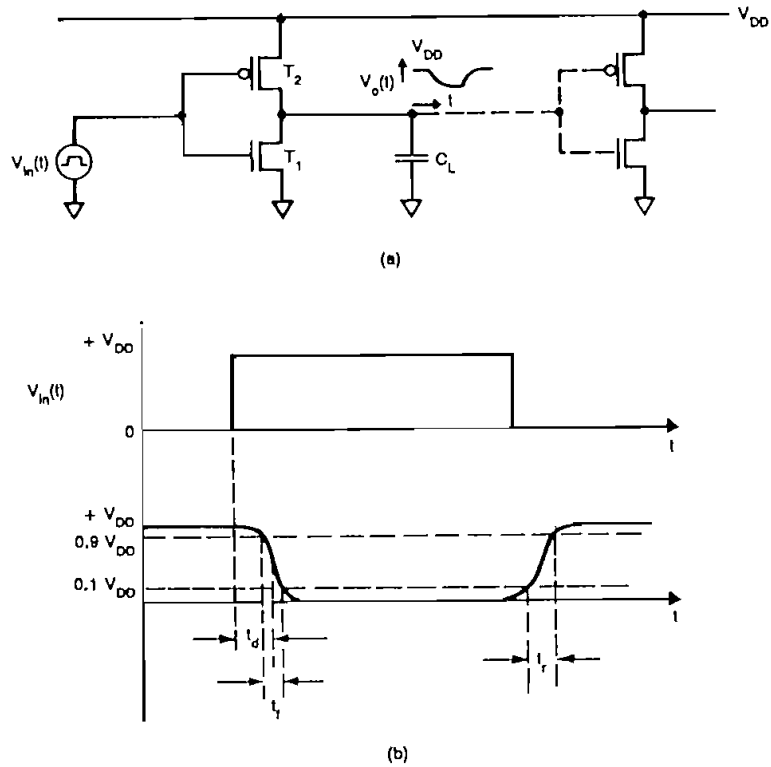


FIGURA 4.14 Característica de conmutación de un inversor CMOS.

#### 4.2.1 Determinación del tiempo de bajada

En la Fig. 4.15 se presenta la trayectoria del punto de operación del transistor "n" cuando el voltaje de entrada cambia de 0V (X1) a  $V_{DD}$  (X2). Inicialmente el dispositivo "n" está cortado y  $C_L$  está cargado a  $V_{DD}$ , cuando el voltaje de entrada cambia a  $V_{DD}$ , el transistor "n" entra en saturación (X2); bajo estas condiciones, el transistor "p" está en corte. Desde X2, el punto de operación se mueve en la característica correspondiente a  $V_{gs} = V_{DD}$  hacia X3, en el origen.

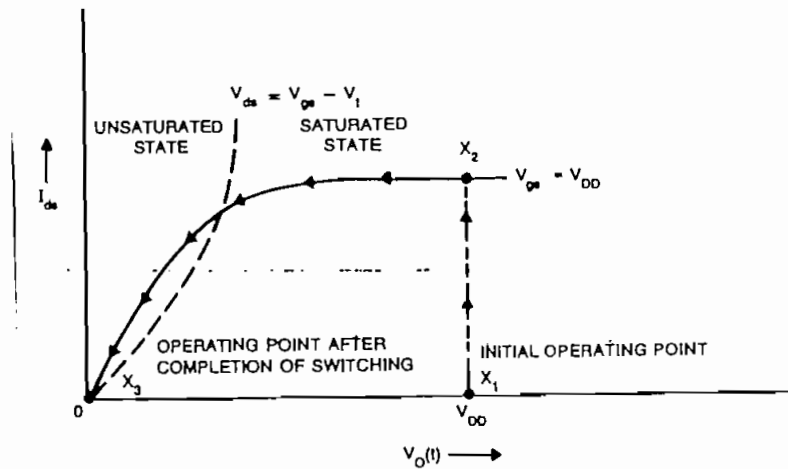


FIGURA 4.15 Trayectoria del punto de operación del transistor "n" durante la conmutación.

La definición de  $t_r$  está indicada en la Fig. 4.14. El valor  $t_r$  es el resultado de dos condiciones de operación del transistor "n". El voltaje de salida, debido a la descarga de  $C_L$ , determina que el transistor opere en la región de saturación y en la región lineal. Por lo tanto,  $t_r$  está dado por la suma de los tiempos en cada región:

- a)  $t_{r1}$ : es el tiempo durante el cual el voltaje del capacitor ( $V_o$ ) cae desde  $V1 = 0.9 V_{DD}$  (instante  $t1$ ), a  $V2 = V_{DD} - V_{tn}$  (instante  $t2$ ). En este rango de voltaje el transistor "n" está saturado, como se representa en la Fig. 4.16a. Utilizando la Fig. 4.16a se plantea la ecuación de corrientes:

$$C_L \frac{dV_o}{dt} + \frac{\beta_n}{2} (V_{DD} - V_{tn})^2 = 0$$

[Ec. 4.28]

Ordenando la Ec. 4.28 e integrando en los intervalos  $(t_1, t_2)$  y  $(V_1, V_2)$  se obtiene:

$$t_{f1} = \frac{2C_L (V_{tn} - 0.1V_{DD})}{\beta_n (V_{DD} - V_{tn})^2}$$

[Ec. 4.29]

b)  $t_{r2}$  es el tiempo durante el cual el voltaje del capacitor ( $V_o$ ) cae desde  $V_2 = V_{DD} - V_{tn}$  (en el instante  $t_2$ ), hasta  $0.1 V_{DD}$  (en el instante  $t_3$ ). En este rango del voltaje, el transistor "n" opera en la región lineal y la corriente de descarga no es constante. El circuito equivalente para esta condición se representa en la Fig. 4.16a. Planteando la ecuación de corrientes e integrando en los intervalos  $(t_2, t_3)$  y  $(V_2, V_3)$  se obtiene:

$$t_{r2} = \frac{C_L}{\beta_n (V_{DD} - V_{tn})} \ln \left( \frac{19V_{DD} - 20V_{tn}}{V_{DD}} \right)$$

[Ec. 4.30]

Para encontrar  $t_r$ , se suman los dos tiempos encontrados. Reemplazando en las Ec. 4.29 y 4.30 el valor típico del voltaje umbral ( $V_{tn} = 0.2 V_{DD}$ ), se obtiene un  $t_r$  de:



$$t_r \approx 4 \frac{C_L}{\beta_n V_{DD}}$$

[Ec. 4.31]

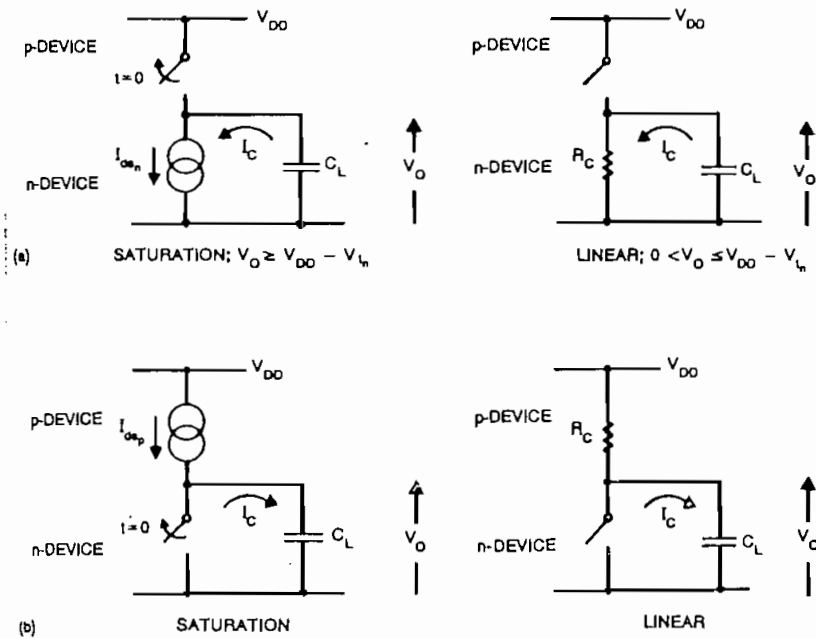


FIGURA 4.16 Circuitos equivalentes para la determinación de a)  $t_r$  y b)  $t_r$ .

#### 4.2.2 Determinación del tiempo de subida

Gracias a la simetría del circuito inversor CMOS, se sigue un proceso similar al seguido para el caso del tiempo de bajada. En la Fig. 4.16b se presentan los circuitos equivalentes para las dos condiciones de operación del transistor "p". El transistor "n" permanece en este caso cortado.

Las expresiones obtenidas son similares, y reemplazando el valor para el voltaje umbral de  $V_{tp} = -0.2 V_{DD}$ , se obtiene un valor de  $t_r$  igual a:

$$t_r = 4 \frac{C_L}{\beta_p V_{DD}}$$

[Ec. 4.32]

Para el caso en que los dos transistores tienen iguales dimensiones, se cumple que  $\beta_n = 2\beta_p$ , debido a que la movilidad de los electrones es aproximadamente el doble que la de los huecos. Relacionando las Ec. 4.31 y 4.32 se obtiene que:

$$t_f = \frac{t_r}{2}$$

[Ec. 4.33]

Lo que indica que el tiempo de bajada es menor que el tiempo de subida. Por lo tanto, si se desea tener iguales tiempos de subida y bajada para el inversor, se debe cumplir la relación  $\beta_n/\beta_p = 1$ . Lo que implica que el ancho del canal para el dispositivo "p" debe incrementarse en aproximadamente dos veces, de tal manera que:  $W_p = 2W_n$ .

Por ejemplo, se presenta el cálculo de  $t_r$  y  $t_f$  para un inversor que tiene las siguientes dimensiones:  $L = 2 \lambda$ ,  $W = 4 \lambda$ , con un valor típico de  $\beta_n = 35 (W/L) \mu A/V^2 = 2\beta_p$ . El valor de  $\beta_p$  es por lo tanto es  $35 \mu A/V^2$ . Con un valor

de  $C_L=0.049$  pF y  $V_{DD} = 5V$ , se tiene un valor de  $t_r$  de aproximadamente 1 ns y un valor de  $t_f$  de 2 ns, valores que en la práctica son incluso menores.

Nótese que para especificar exactamente la relación de los anchos de los transistores y lograr iguales tiempos de bajada y subida, debe conocerse exactamente la relación entre las movilidades de electrones y huecos, lo que depende de ciertos parámetros del proceso de fabricación.

#### 4.2.3 Tiempo de retardo

De acuerdo a la definición planteada anteriormente, se representa el tiempo  $t_d$  en la Fig. 4.14b, para una transición de entrada completamente perpendicular.

El tiempo de retardo de una sola compuerta está determinado principalmente por los tiempos de subida y bajada de la salida. El tiempo de retardo, para una transición de salida de alto a bajo, realizando una aproximación bastante gruesa, está dado por la mitad del tiempo de bajada ( $t_r$ ): (3)

$$t_{dr} = \frac{t_r}{2}$$

[Ec. 4.34]

---

3 "Principles of CMOS VLSI Design", Weste N. - Eshraghian K., pág. 141.

Para una transición bajo a alto en la salida, el tiempo de retardo está dado por la mitad del tiempo de subida ( $t_r$ ):

$$t_{dx} = \frac{t_r}{2}$$

[Ec. 4.35]

El retardo promedio de una compuerta para las transiciones de subida y bajada es:

$$t_{av} = \frac{t_{df} + t_{dx}}{2} = \frac{t_r + t_f}{4}$$

[Ec. 4.36]

Para  $t_r = 2$  ns y  $t_f = 1$  ns, el retardo promedio del inversor es 0.75 ns.

### 4.3 ESTIMACION DEL CONSUMO DE POTENCIA

Se deben considerar dos tipos de consumo de potencia en un circuito CMOS:

- a) Disipación de potencia estática, debida a las corrientes parásitas.
- b) Disipación de potencia dinámica, debida a dos factores:
  - i) Corrientes transitorias durante la conmutación.

- ii) Corrientes requeridas para la carga y descarga de las capacitancias de carga.

#### 4.3.1 Disipación estática ( $P_s$ ).

En la Fig. 4.17 se representa un inversor CMOS totalmente complementario, para los dos valores de entrada siempre uno de los transistores está abierto. Por lo tanto, ninguna corriente fluye en el circuito, no existe camino para DC entre  $V_{DD}$  y Gnd y el consumo de potencia  $P_s = 0$ .

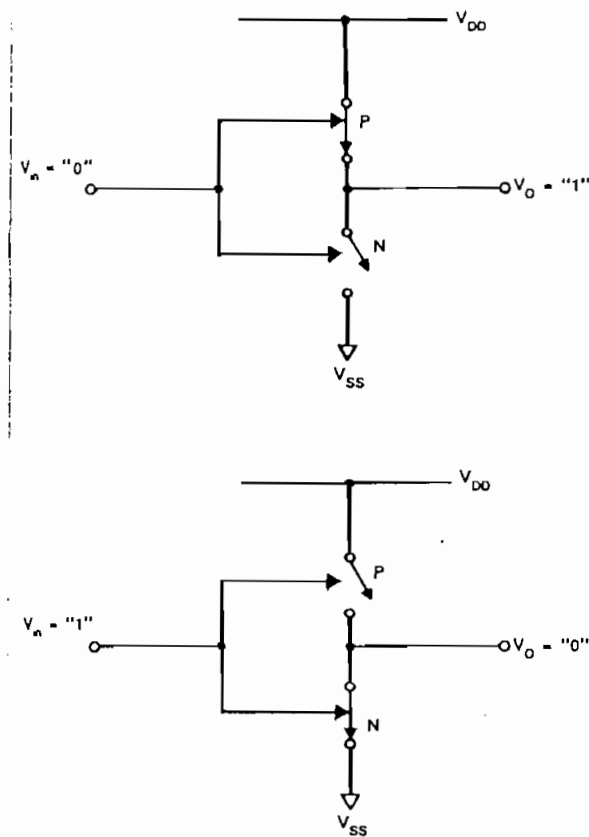


FIGURA 4.17 Circuitos equivalentes para la determinación de disipación estática.

En realidad si existe un pequeño consumo estático debido a las corrientes parásitas de polarización inversa. La causa del consumo son las junturas parásitas que se forman entre las diferentes regiones de difusión y el sustrato. En la Fig. 4.18 se representan las junturas formadas y las condiciones de polarización a las que están sometidos estos diodos. Puesto que los diodos están polarizados inversamente, solo la corriente parásita denominada corriente inversa de saturación contribuye a la disipación estática.

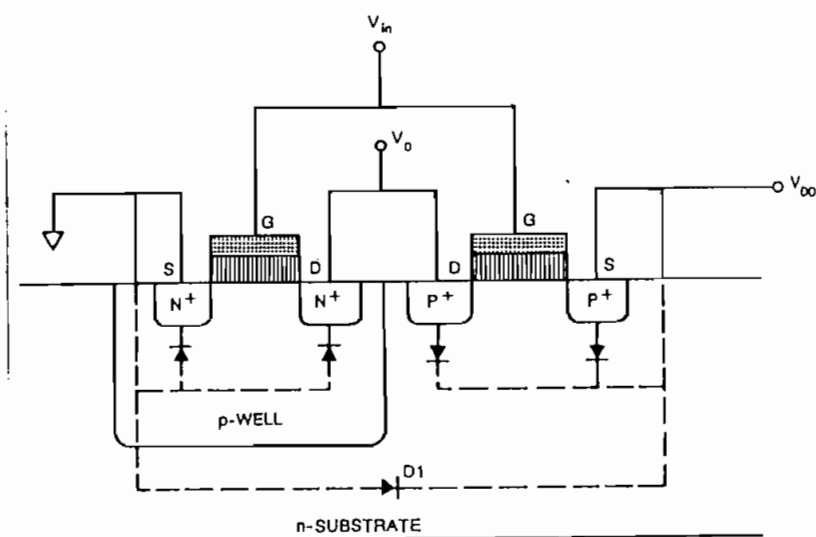


FIGURA 4.18 Modelo que representa las junturas parásitas formadas.

La disipación de potencia estática se estima como el producto de la corriente parásita del dispositivo y la fuente de polarización. Un valor práctico a ser asumido está en el rango de 0.1 nA a 0.5 nA por compuerta, a temperatura ambiente. La disipación total de potencia estática  $P_s$  se obtiene como:

$$P_s = \sum_1^n \text{corriente-parásita} * V_{DD}$$

[Ec. 4.37]

en donde n es el número de dispositivos. Por ejemplo, un valor típico de disipación estática para un inversor polarizado con 5V está entre 1 y 2 nW.

#### 4.3.2 Disipación dinámica (Pd).

Durante las transiciones de "0" a "1" o de "1" a "0", los dos transistores "n" y "p" conducen simultáneamente durante un corto periodo de tiempo, provocando pulsos de corriente que circulan entre  $V_{DD}$  y Gnd. También se requiere una corriente para cargar y descargar las capacidades de carga; al final, de las dos corrientes, esta última es la que predomina con su contribución al consumo de potencia. En el modelo escogido, para estimar la disipación dinámica, considera que el periodo de la señal aplicada a la compuerta es mucho mayor que los tiempos de subida y bajada requeridos. Las formas de onda que se utilizan para la determinación de la disipación dinámica se presentan en la Fig. 4.19.

La potencia dinámica promedio (Pd) disipada por una compuerta durante la conmutación, debido a una señal de entrada de onda cuadrada y frecuencia  $f_p = 1/t_p$ , con una carga  $C_L$  está dada por:

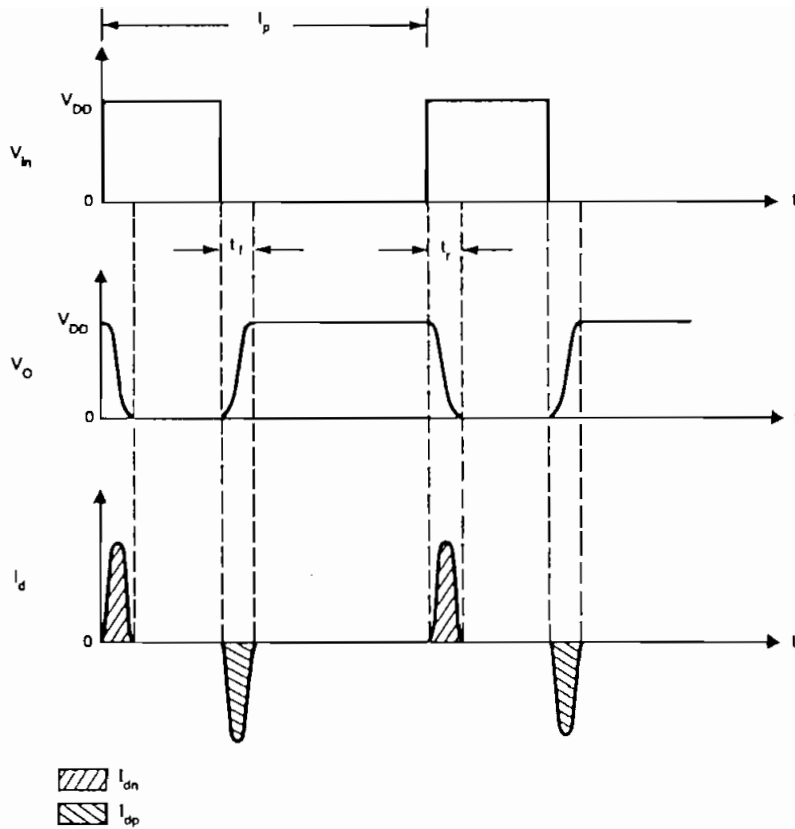


FIGURA 4.19 Formas de onda para la determinación de la disipación dinámica.

$$Pd = \frac{C_L V_{DD}^2}{\tau_p} = C_L V_{DD}^2 f_p$$

[Ec. 4.38]

De la Ec. 4.38 se estima la corriente promedio como  $f_p C_L V_{DD}$  y la energía disipada por ciclo como  $C_L V_{DD}^2$ . Por lo tanto, la potencia dinámica promedio disipada es proporcional a la energía requerida para cargar y descargar la capacitancia del circuito. Además,  $P_d$  es directamente proporcional a la frecuencia de conmutación, pero independiente de los parámetros del dispositivo. La fórmula planteada es aplicable a cualquier compuerta CMOS.



La potencia total disipada se obtiene sumando las dos componentes encontradas:

$$P_{total} = P_s + P_d$$

[Ec. 4.39]

Como ejemplo, se estima la potencia total disipada en un sistema que utiliza N inversores, con los valores de capacitancia calculados para el transistor de la Fig. 4.7. Las condiciones de operación son:

$$f_p = 10 \text{ MHz}$$

$$V_{DD} = 5V.$$

$$\text{capacidad de salida} = 2 C_d = 2 * 40 \text{ fF}$$

$$\text{capacidad de entrada} = 2 C_g = 2 * 11.2 \text{ fF}$$

$$P_s = N (0.1 * 10^{-9} * 5) \text{ W} = N (0.5 * 10^{-9}) \text{ W}$$

$$P_d = N (2C_d + 2C_g) * 25 * 10 * 10^6 = N (25 * 10^{-6}) \text{ W}$$

#### 4.4 ESCALAMIENTO DE DIMENSIONES DEL TRANSISTOR MOS

Se presenta a continuación una visión general de los efectos en el comportamiento eléctrico que se esperan, según las dimensiones involucradas en la construcción de los dispositivos se reduzcan paulatinamente. El modelo utilizado para representar el efecto de la disminución de las dimensiones consiste simplemente en escalarlas por un factor

adimensional  $\alpha$ . Aunque es poco probable que en los procesos reales CMOS se escalen de una manera tan fácil como la expuesta, los resultados presentados dan una idea clara de lo que el diseñador debe esperar de las líneas futuras de fabricación.

#### 4.4.1 Principios de escalamiento

La Teoría de escalamiento de primer orden para los dispositivos MOS, basado en el modelo de "campo constante" de Dennard <sup>(\*)</sup>, señala que las características de un dispositivo MOS y sus características operacionales pueden conservarse si los parámetros críticos del dispositivo son escalados de acuerdo a un criterio dado, por un factor constante, de tal manera que todos los campos eléctricos del circuito permanezcan constantes.

Utilizando la idea propuesta se consigue que muchos de los factores no-lineales que afectan el comportamiento del dispositivo no cambien, como lo harían si se utilizase un esquema de escalamiento más complejo.

Esta teoría ha demostrado ser muy efectiva escalando características del rango de 5 a 10  $\mu\text{m}$  hacia el rango de 1 a 3  $\mu\text{m}$ .

---

\* "Solid State Circuits", Dennard, R.

Aunque la teoría no proporciona comportamientos optimizados de los dispositivos en pequeñas dimensiones, la técnica es muy poderosa para presentar las mejoras o limitaciones que pueden presentarse cuando se escalan los procesos.

Basicamente un dispositivo escalado se obtiene aplicando un factor adimensional  $\alpha$  a:

- a) Todas las dimensiones, incluidas las dimensiones perpendiculares a la superficie del dispositivo (longitud del canal  $L$ , ancho del canal  $W$ , espesor del óxido  $t_{ox}$ , profundidad de las difusiones  $X_d$ ).
- b) Los voltajes de los dispositivos.
- c) A la concentración del sustrato.

En la Fig. 4. 20 se ilustra el resultado de un escalamiento de primer orden en un dispositivo MOS.

En la Tabla 4.6 se resume el efecto del escalamiento de primer orden en las diferentes características de un dispositivo MOS.

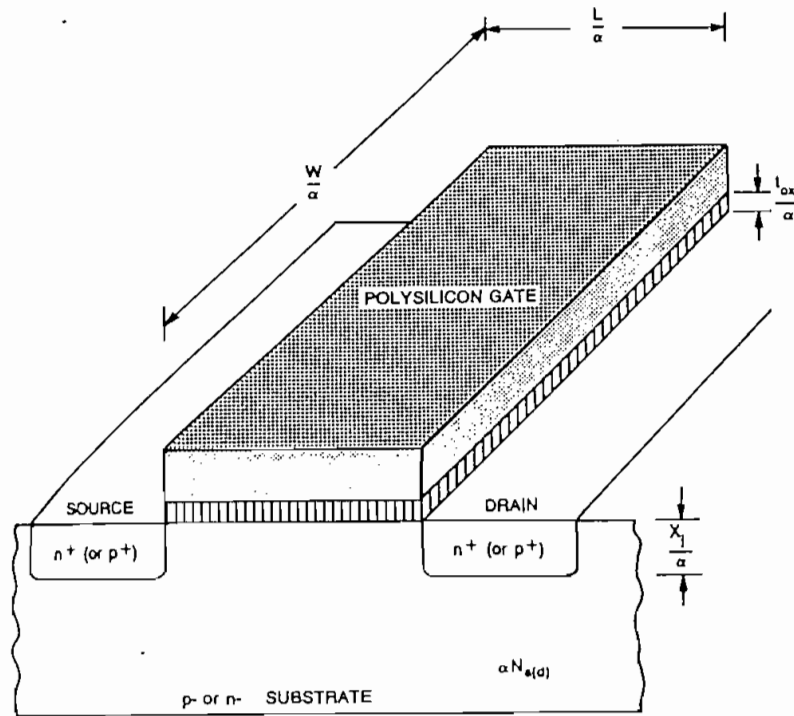


FIGURA 4.20 Dispositivo MOS escalado.

	PARAMETERS	SCALING FACTOR
DEVICE PARAMETERS	Length; $L$	$1/\alpha$
	Width; $W$	$1/\alpha$
	Gate oxide thickness; $t_{ox}$	$1/\alpha$
	Junction depth; $X_j$	$1/\alpha$
	Substrate doping; $N_a$ (or $d$ )	$\alpha$
	Supply voltage; $V_{DD}$	$1/\alpha$
	Electric field across gate oxide; $E$	1
RESULTANT INFLUENCE	Depletion layer thickness; $d$	$1/\alpha$
	Parasitic capacitance; $WL/t_{ox}$	$1/\alpha$
	Gate delay; $(VC/I)$	$1/\alpha$
	DC power dissipation; $P_s$	$1/\alpha^2$
	Dynamic power dissipation; $P_d$	$1/\alpha^2$
	Power-speed product	$1/\alpha^3$
	Gate area	$1/\alpha^2$
	Power density; $(VI/A)$	1
	Current density; $(I/A)$	$\alpha$
Transconductance; $g_m$	1	

TABLA 4.6 Influencia del escalamiento en las características de un dispositivo MOS

Se puede mencionar que al escalar los parámetros del dispositivo (dimensiones, voltajes y concentraciones), tiene las siguientes características:

- a) El valor del voltaje umbral ( $V_{t}$ ) y la corriente drenaje-fuente ( $I_{d_{on}}$ ) también disminuyen.
- b) La profundidad de las regiones de agotamiento de fuente y drenaje determinan la longitud mínima del canal, de tal manera que la compuerta controle la conductancia del canal. Para reducir la longitud del canal ( $L$ ), se debe reducir también la profundidad de las regiones de agotamiento ( $d$ ), lo que se consigue incrementando el nivel de dopaje del sustrato.
- c) A pesar que los voltajes son escalados, el campo eléctrico ( $E$ ) en el óxido de la compuerta permanece constante.
- d) El escalar las dimensiones del transistor por  $1/\alpha$ , produce que la corriente  $I_{d_{on}}$  se reduzca en  $\alpha$  y la densidad de integración se incremente en  $\alpha^2$ . El resultado neto es que la densidad de corriente aumente en  $\alpha$ , lo que implica caminos de metal más anchos para estructuras de mayor densidad de integración.
- e) La disipación de potencia estática ( $P_s$ ) y dinámica ( $P_d$ ) disminuyen en el factor  $1/\alpha^2$ ; sin embargo, puesto que

los dispositivos por unidad de área se incrementa en  $\alpha^2$ , el efecto resultante es que la densidad de potencia permanece constante.

Debe mencionarse que los valores presentados en la Tabla 4.6 son solamente aproximaciones de primer orden. Además, puede tenerse la impresión equivocada que con el escalamiento de primer orden puede llegarse a dimensiones tendientes a cero o a voltajes umbrales de cero. Consideraciones teóricas y prácticas no permiten alcanzar este límite.

#### 4.4.2 Escalamiento de los caminos de interconexión

A pesar de que el escalamiento de primer orden proporciona algunas ventajas, hay algunos parámetros del circuito que sufren degradación debido al escalamiento. Así, las caídas de voltaje, retardos de propagación en los caminos de interconexión, densidad de corriente, etc.

El incremento de la densidad de corriente, por ejemplo, hace que el efecto de migración del metal requiera especial atención. Se debe entonces buscar nuevos esquemas de metalización para manejar los valores más altos de la densidad de corriente.

En la Tabla 4.7 se resumen los efectos del escalamiento en algunos de los parámetros indicados.

PARAMETERS	SCALING FACTOR
Line resistance; $r$	$\alpha$
Line response; $rc$	1
Normalized line response	$\alpha$
Line voltage drop; $V_d$	1
Normalized line voltage drop	$\alpha$
Current density; $J$	$\alpha$
Normalized contact voltage drop; $V_c/V$	$\alpha^2$

TABLA 4.7 Influencia del escalamiento en los medios de interconexión

#### 4.5 ESTRATEGIAS DE TEMPORIZACION DE LOS SISTEMAS

El éxito del diseño de un CI VLSI no se centra solamente en los aspectos concernientes con su proceso de fabricación, sino también en la operación del circuito en el dominio del tiempo.

Como se ha explicado en este capítulo, los valores de capacitancia asociados a transistores y caminos de enrutamiento (capacitancias parásitas), tienen un rol fundamental para la operación de los circuitos.

Los valores lógicos "1s" y "0s" son tan solo una abstracción de los voltajes reales del circuito; las transiciones entre "1s" y "0s" representan el proceso físico de carga y descarga de capacitancias parásitas. Por lo tanto, los voltajes toman un tiempo finito para llegar a sus valores finales, contribuyendo al retardo del circuito.

Aunque la capacitancia es considerada como un factor negativo, que contribuye a que los circuitos sean más lentos, ésta tiene dos virtudes:

- a) Proporciona "inercia eléctrica"; es decir, todo cambio necesita de un tiempo finito de respuesta. Sin esta característica los circuitos serían extremadamente sensibles a pulsos espúreos. Por ejemplo, pueden ocurrir cambios momentáneos indeseados de una señal o puede ocurrir que durante una transición la señal va normalmente a su valor final, pero debido a causas extrañas regresa al valor inicial y finalmente retorna al valor final, al menos una vez. Estos dos efectos que producen pulsos espúreos se ilustran en la Fig. 4.21a y Fig. 4.21b, respectivamente, y pueden ser causados por los retardos en los caminos que llevan las señales.

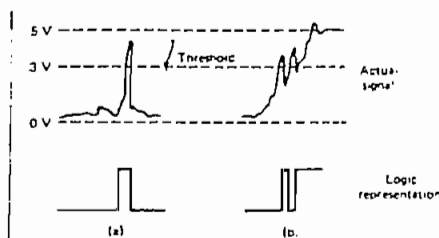


FIGURA 4.21 Efectos no deseados que pueden ser eliminados gracias a las capacitancias.

- b) Puede mantener carga eléctrica. Una carga puede representar información y un circuito digital manipula las cargas almacenadas para realizar funciones lógicas.



Si los procesos de carga y descarga ocurren en momentos discretos se obtiene control sobre el flujo de información. Para esto se utilizan señales de control, generalmente señales de reloj. En este caso se dice que el flujo de información se produce sincronicamente con un pulso de reloj y los circuitos se denominan temporizados (*clocked*) o síncronos.

Circuitos asincrónicos, por el contrario, son aquellos que funcionan libremente, sin el pulso de reloj. En los circuitos asincrónicos las relaciones de tiempo están determinadas por los retardos inherentes del circuito.

#### 4.5.1 Esquemas de temporización de circuitos síncronos

Para realizar el control de la información en los circuitos síncronos, se adoptan diferentes esquemas de temporización, entendiéndose por esquema de temporización la estrategia utilizada para definir los momentos durante los cuales la información se transfiere entre las diferentes etapas de un circuito y los intervalos de tiempo durante los cuales las etapas están aisladas entre sí.

Muchos esquemas de temporización son posibles y una gran variedad de ellos están siendo utilizados actualmente en el diseño de CI, otros esquemas no son prácticos y no se utilizan.

Los dos esquemas de temporización utilizados en el desarrollo de la presente tesis son:

- a) El esquema de temporización de una fase.
- b) El esquema de temporización de dos fases no-sobrelapadas.

Estos esquemas tienen amplia utilización y a partir de ellos se han desarrollado muchas variantes.

a) Esquema de temporización de una fase

El esquema de temporización de una fase consiste en utilizar una señal de reloj (CK), la cual está constituida por un tren de pulsos, secuencias entre "0s" y "1s". Cada periodo de CK ( $T$ ) consiste de un pulso de ancho  $W$  (en alto) y entre los pulsos se tiene un intervalo en bajo  $g$ . La forma de la señal de reloj descrita se representa en la Fig. 4.22.

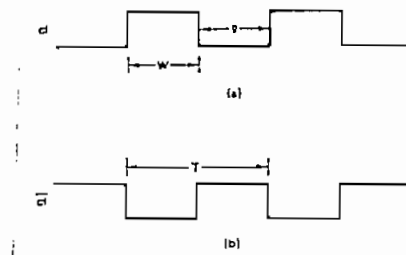


FIGURA 4.22 Señal de reloj de una sola fase y su complemento.

La señal de reloj de una fase define dos eventos separados: el flanco de subida o flanco positivo y el flanco de bajada o flanco negativo.

Estos eventos determinan los puntos en el tiempo cuando generalmente los valores de las señales, en diferentes partes del circuito deben cambiar, llevando a cabo una transición hacia un nuevo estado del circuito. También puede utilizarse dentro del esquema de temporización la señal de reloj complementada ( $\overline{CK}$ ), que es particularmente útil en circuitos CMOS.

b) Esquema de temporización de dos fases no-sobrelapadas

Las dos señales de reloj utilizadas para el esquema de temporización de dos fases no-sobrelapadas se esquematiza en la Fig. 4. 23.

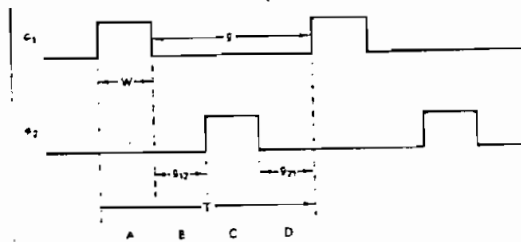


FIGURA 4.23 Señales de reloj para el esquema de dos fases no-sobrelapadas.

Como puede verse en la Fig. 4.23, para este esquema de temporización se requiere de dos señales de reloj:  $\phi_1$  y  $\phi_2$ , sincronizadas entre sí y cada una con un periodo  $T$ . Nótese

que las señales nunca se sobrelapan, en otras palabras, para todo tiempo  $t$ , se cumple que:

$$\phi_1(t) \cdot \phi_2(t) = 0$$

Realizando una observación conjunta de las dos señales, se observa que entre los pulsos de  $\phi_1$  y los de  $\phi_2$  existen espaciamientos iguales:  $t_{12} = t_{21}$ . Sin embargo, no es estrictamente necesario que siempre se cumpla que  $t_{12} = t_{21}$ , pero generalmente se asume que así ocurre.

De lo explicado se desprende que en el intervalo de tiempo  $T$ , se tienen dos intervalos durante los cuales una de las señales de reloj está en alto, de ahí el nombre de este esquema de temporización. Se dispone entonces de dos fases, en diferentes tiempos, para controlar diferentes eventos dentro de un circuito.

Las dos fases del reloj definen cuatro eventos separados: el flanco positivo de  $\phi_1$ , el flanco negativo de  $\phi_1$ , el flanco positivo de  $\phi_2$  y el flanco negativo de  $\phi_2$ .

En la Fig. 4.24 se presenta la configuración de un circuito a partir del cual se pueden obtener las señales de reloj no-sobrelapadas y sus complementos.

Como entrada el circuito requiere una señal de reloj de una sola fase, denominado reloj maestro. Las cuatro señales tienen aplicación en circuitos CMOS. En circuitos nMOS se utilizan solo  $\phi_1$  y  $\phi_2$ .

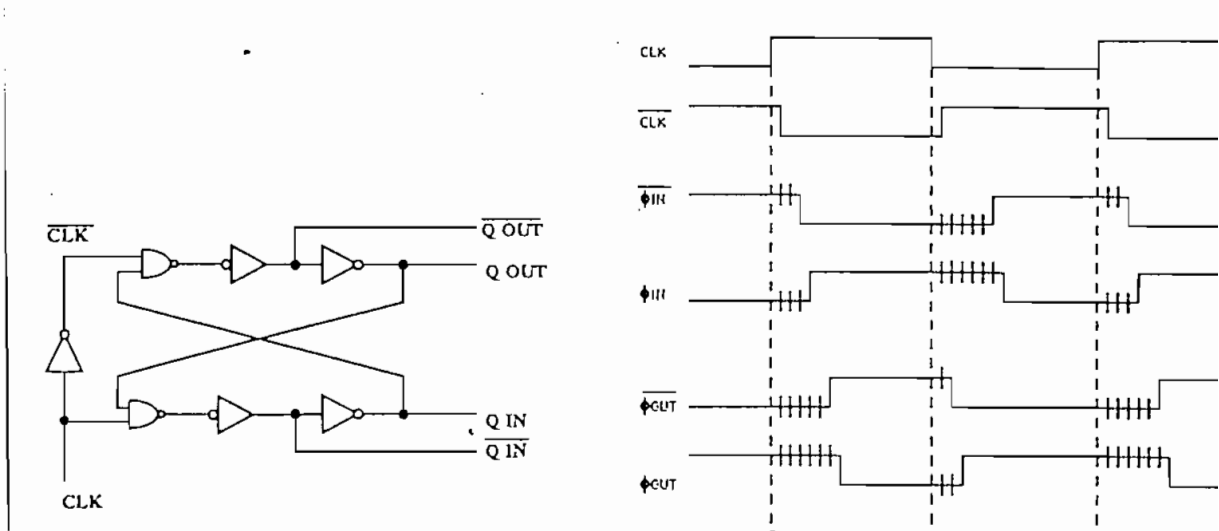


FIGURA 4.24 Generación de las señales de reloj para el esquema de dos fases no-sobrelapadas.

#### 4.5.2 Modelo lógico para la descripción de los esquemas de temporización

El modelo lógico que se utiliza para la descripción de las estrategias de temporización de los circuitos sincrónicos es el de las Máquinas de Estados Finitos (MEF). El modelo general de una MEF incluyen dos bloques fundamentales, un circuito combinacional (CC) y un camino de realimentación que utiliza elementos de almacenamiento (memoria), como se ilustra en la Fig. 4. 25.

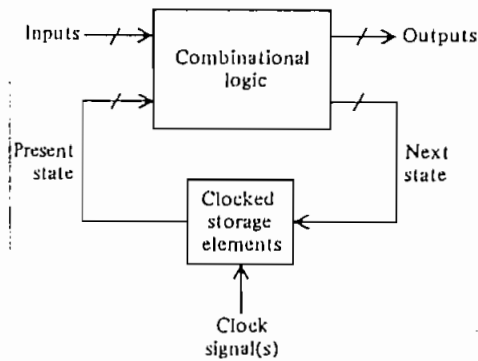


FIGURA 4.25 Modelo de una MEF.

El CC no requiere de mayor explicación; sin embargo, los elementos de almacenamiento, sobre los que actúan las señales de reloj, de acuerdo a los esquemas de temporización, para determinar los intervalos en los que se realiza el almacenamiento y transferencia de la información de los estados presente y siguiente, requieren de una explicación introductoria.

a) Elementos de memoria

Basicamente todos los elementos de memoria son dinámicos o estáticos.

En un elemento dinámico el almacenamiento se realiza en una capacitancia, que mantiene la información en base a la presencia o ausencia de carga eléctrica. El grave problema que presentan estos elementos es que la carga no puede mantenerse indefinidamente.

En un elemento de memoria estático, la información almacenada es debida a la conducción o no-conducción de transistores. La información puede ser mantenida indefinidamente. Un ejemplo típico de estos elementos es el flip-flop.

Los dispositivos estáticos consumen más potencia que los elementos dinámicos pero pueden mantener la información indefinidamente. Los elementos dinámicos requieren un refresco periódico de la carga almacenada pues ésta puede perderse debido a las corrientes parásitas.

Además, los elementos dinámicos son más sensibles al ruido de las fuentes de polarización, campos eléctricos y magnéticos externos y ambientes radioactivos. Cualquiera de los factores mencionados puede ser causa suficiente para añadir o retirar carga de un nodo y alterar su estado lógico. Como otra característica negativa, el refresco necesario en los elementos dinámicos obliga a operar los circuitos dinámicos a bajas frecuencias.

b) Elemento de memoria dinámico con un esquema de temporización de una fase.

En la Fig. 4.26 se presenta un "latch" formado por un transistor de paso y un inversor. El proceso de almacenamiento de la señal de entrada X consiste en poner en alto la señal de control (CK) y mantener la señal de entrada en un valor estable, de tal manera que la capacidad de la compuerta

del inversor y del transistor de paso (nodo U), pueda ser cargada al valor X, al tiempo empleado en este proceso se le denomina tiempo de retardo (*delay time*). La salida del inversor (V), luego de un tiempo finito de retardo, llega a su valor final.

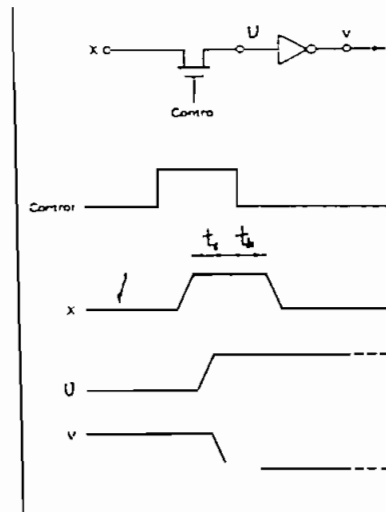


FIGURA 4.26 Latch dinámico nMOS.

Si se asume que la señal de control está en alto por un tiempo finito (no indefinido), la señal X debe mantenerse en un valor estable, antes del flanco negativo de la señal de control, por un tiempo mínimo denominado "*preset time*" o "*setup time*" ( $t_s$ ). Obviamente CK debe estar en alto un tiempo mayor al de establecimiento.

Una condición adicional se impone a la señal de entrada, la señal X no debe cambiar inmediatamente después del flanco negativo de la señal de control, debe mantenerse (*hold*)



estable al menos un tiempo  $t_h$ . Este tiempo  $t_h$  es necesario para permitir posibles retardos entre el flanco negativo del reloj maestro y el flanco negativo de la señal de reloj local que actúa como señal de control. Este retardo se conoce también como "clock skew", y se explica más adelante en este capítulo.

Si la señal de control cambia a cero, los nodos U y V mantienen sus valores, al menos hasta que se descarguen por efectos parásitos, a pesar que la señal X cambie de valor. Para este latch, se dice que la operación de almacenamiento tuvo lugar con el flanco negativo de la señal de reloj. El dispositivo se denomina dinámico puesto que la información se mantiene hasta que la carga del nodo U se disipe y el voltaje caiga bajo el voltaje umbral del inversor. En CMOS, el transistor de paso puede reemplazarse por una compuerta de transmisión complementaria, como se indica en la Fig. 4.27. En este caso la operación del Latch es similar a la discutida.

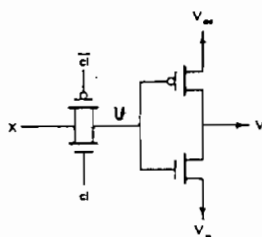


FIGURA 4.27 Latch dinámico CMOS.

### 4.5.3 Problemas con el esquema de temporización de una fase

Aunque muchos sistemas digitales utilizan el esquema de temporización de una fase, existe un gran número de fuertes limitaciones que deben considerarse con respecto a los retardos en los circuitos, periodos de las señales de reloj ( $T$ ) y ancho del pulso ( $W$ ) de la señal de reloj, para garantizar una correcta operación del circuito.

Para entender los problemas con el esquema de temporización de una fase, se utiliza la estructura de la MEF. En la Fig. 4.28, se presenta nuevamente la estructura de la MEF. Se asume que la realimentación se realiza con simples latches, la salida del CC define el estado siguiente y las salidas de los latches definen el estado presente.

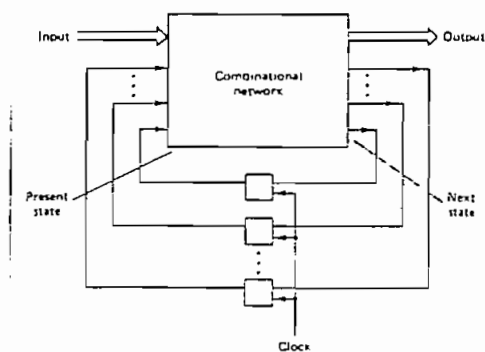


FIGURA 4.28 MEF con una señal de reloj de una fase.

El ancho del pulso de reloj ( $W$ ) debe ser más grande que el tiempo de retardo que introduce la circuitería del *latch*, pero no puede ser mayor que el retardo mínimo del CC (dado por cualquiera de las señales que entrega el CC), pues causaría una condición de "multistepping" o "racing" (competencia). Esto significa que los valores del estado presente cambian más de una vez durante el periodo de la señal de reloj.

Por ejemplo, si se supone que  $W = 5\text{seg}$ , que el retardo del *latch* es despreciable y que el retardo del CC es  $1\text{seg}$ . Si el estado siguiente es  $Q_0$ , luego de un corto tiempo, este valor pasa por el *latch* de realimentación y  $Q_0$  está presente en la entrada del CC como estado presente. Conjuntamente con el valor de las entradas el CC, se determina que el estado siguiente sea  $Q_1$ . Luego de  $1\text{seg}$ ,  $Q_1$  está presente a la entrada del *latch* y muy pronto está nuevamente a la entrada del CC. El CC determina en esta ocasión que el estado siguiente sea  $Q_2$ , y entrega este valor. Hasta el momento han transcurrido algo más que  $2\text{seg}$ , posiblemente hasta que transcurran los  $5\text{seg}$ , el estado siguiente tiene un valor  $Q_4$ , y no  $Q_1$  como se deseaba.

Si el estado final del circuito está de acuerdo con la transición especificada en el diagrama de estados, tal condición de "racing", se denomina "no-crítica". Durante la transición al nuevo estado, el circuito puede pasar por algunos estados intermedios, diferentes del estado final.

Si las salidas que entrega el CC en estos estados intermedios son diferentes de la salida del estado final esperado, se pueden producir señales espúreas. Si estas salidas no son sensibles a las señales espúreas, un "racing" no-crítico es aceptable.

Una condición más seria se presenta cuando el circuito se queda en un estado final incorrecto. Tal condición se denomina "racing" crítico, y debe ser eliminada.

También existen algunas limitaciones al periodo  $T$  de la señal de reloj. La primera señala que  $T$  debe ser mayor que el retardo más grande del CC, de tal manera que el cálculo del estado siguiente se propague completamente hasta la entrada del latch de realimentación, para que sea realimentado al CC en el siguiente pulso de reloj. La segunda señala que  $T$  no puede ser demasiado grande si se utilizan elementos de almacenamiento dinámicos en el camino de realimentación, esto debido a que la información almacenada se pierde por las corrientes parásitas. Para cumplir este objetivo, el periodo de la señal de reloj debe ser menor que el tiempo de refresco de los elementos dinámicos.

Algunos de los problemas presentados anteriormente, pueden eliminarse utilizando el esquema de una fase de reloj y su complemento, que incluye la señal de reloj y su complemento.

4.5.4 Esquema de temporización de una fase de reloj y su complemento

La forma general de un circuito secuencial utilizando este esquema de temporización y elementos de memoria dinámicos, se presenta en la Fig. 4.29a (solo un camino de realimentación se presenta).

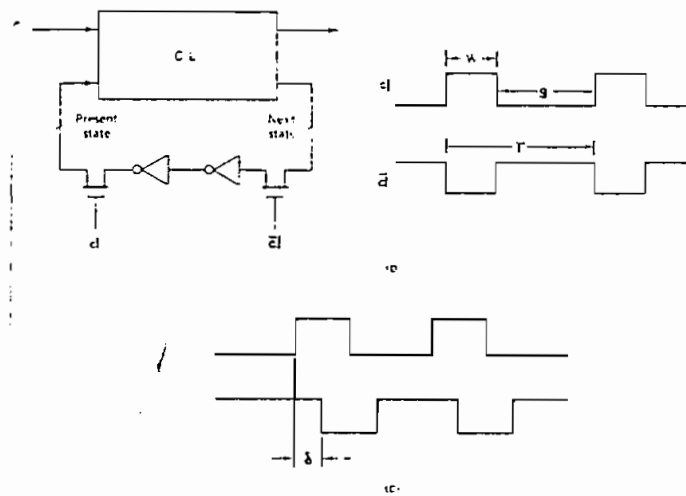


FIGURA 4.29 MEF con una señal de reloj de una fase y su complemento.

La información del estado siguiente se almacena cuando  $\overline{CK}=1$ , en la primera etapa del latch, pero no se permite su realimentación hasta que  $\overline{CK}=0$  y  $CK=1$ . Solamente bajo esta condición se entrega la información a la entrada del CC.

El ancho del pulso ( $w$ ) debe ser mayor que el tiempo de establecimiento ( $t_s$ ) del CC, pero ya no tiene que ser menor que el retardo mínimo ( $t_{min}$ ) del CC.

La ventaja expuesta se consigue ya que la señal tiene que pasar primero por el transistor de paso manejado por  $\overline{CK}$ , que está "apagado" cuando el otro transistor está "encendido". De esta manera las condiciones de "racing" o "multistep-ping" se eliminan completamente.

La separación entre pulsos ( $g$ ) debe ser suficiente para que la información del siguiente estado se propague por el transistor de paso y los dos inversores. Al igual que en el caso anterior, el periodo ( $T = W + g$ ) debe ser mayor que el retardo máximo del CC ( $\delta_{max}$ ), de tal manera que la información del estado presente esté lista para entregarse en el siguiente pulso de reloj. Todas las condiciones mencionadas pueden cumplirse si se hacen  $W$  y/o  $g$  más grandes.

a) Fenómeno de "Clock skew".

La razón fundamental que le permite al circuito anterior eliminar la condición de "racing" es que el camino de realimentación nunca está cerrado. La información se transfiere a través de dos etapas, aisladas una de la otra gracias, a que las señales de control son complementarias.

El esquema presentado funciona adecuadamente si los flancos de  $CK$  y  $\overline{CK}$  están perfectamente alineados, de manera complementaria. Una condición de "racing" puede presentarse en este circuito si los flancos en mención están desalineados, debido al fenómeno denominado "clock skew". Este fenó-

Debido a los problemas enunciados, el esquema de temporización de una fase no es el preferido de los diseñadores. Sin embargo, utilizarlo conduce a circuitos de bajo costo, pues el distribuir un menor número de señales implica menor área y menor complejidad en la búsqueda de caminos de enrutamiento.

La mayoría de los problemas encontrados con este esquema de temporización se solucionan utilizando un esquema de temporización diferente, aunque más costoso. Este esquema es el de dos fases no-sobrelapadas.

En los circuitos CMOS, se utilizan los esquemas de temporización de una fase y el de dos fases no sobrelapadas, incluidas las señales de reloj complementadas.

#### 4.5.5 Esquema de temporización de dos fases no-sobrelapadas.

En la Fig. 4.30 se presenta la estructura de una MEF y una nueva estructura del *latch* de realimentación. El *latch* representado en este caso y la operación del circuito es muy similar a la descrita para el esquema anterior. En este caso se garantiza el completo aislamiento entre las dos etapas del *latch* de realimentación. En la figura se representan también las señales de reloj utilizadas y se aprecian cuatro intervalos de tiempo:  $\phi_1$ ,  $t_{12}$ ,  $\phi_2$  y  $t_{21}$ .

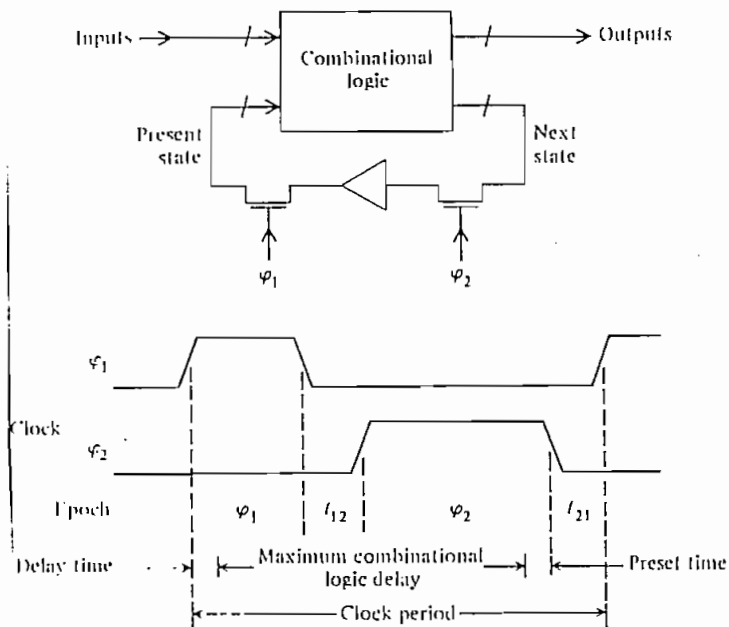


FIGURA 4.30 MEF con un esquema de temporización de dos fases no-sobrelapadas.

Durante  $\phi_1$ , la información previamente almacenada se presenta al CC como la información del estado presente.  $\phi_1$  debe permanecer en alto el tiempo suficiente para que la información atraviese el transistor de paso y se carguen las capacitancias de entrada del CC con el valor adecuado (*delay time*). EL CC inicia el cálculo de los valores de las salidas y el estado siguiente, independientemente del instante en que  $\phi_1$  cambie de alto a bajo.

$t_{12}$  se requiere para que no exista el sobrelapamiento de las dos fases de reloj. Se puede demostrar que es factible reducir el valor de  $t_{12}$  hasta incluso permitir un sobrelapamiento, siempre y cuando no se exceda el retardo del CC. Sin embargo, como la tecnología busca la reducción de los tiempos de retardo, es conveniente hacer que  $t_{12}$  siempre sea mayor que cero.



En el intervalo durante el cual  $\phi_2$  está en alto, se muestrea el valor del CC que contiene la información del estado siguiente, las señales del CC muestreadas deben estar estables al menos por el tiempo de establecimiento (*preset time*) antes que  $\phi_2$  cambie de alto a bajo. Obviamente  $\phi_2$  debe estar en alto un tiempo mayor al de establecimiento.

El siguiente intervalo  $t_{21}$  es de no-sobrelapamiento. Este intervalo de tiempo debe establecerse de tal manera que se eviten los efectos del *skew clock*.

El periodo mínimo del reloj maestro para una correcta operación está dado por la suma del retardo máximo del CC, el máximo tiempo de retardo, el máximo tiempo de establecimiento (*preset*) y por un adecuado valor de  $t_{21}$ .

En el caso de utilizar elementos de memoria estáticos, cualquiera de las condiciones establecidas se consiguen simplemente incrementando cualquiera de los 4 intervalos de tiempo. Si se utilizan elementos de memoria dinámicos se debe considerar el periodo de refresco necesario.

#### 4.5.6 Ejemplos de aplicación de las estrategias de temporización a circuitos CMOS.

La presentación de los esquemas de temporización se realizó utilizando transistores de paso nMOS como parte de los

*latches* de realimentación. Los misma estructura de estos circuitos son válidos para CMOS, si se utilizan compuertas de transmisión e inversores CMOS.

En este numeral se presenta la estructura de algunos elementos de memoria básicos utilizados en CMOS. Nuevamente se recuerda que además de la lógica totalmente complementaria, existen configuraciones lógicas CMOS alternativas. Muchas de estas alternativas dan uso de los conceptos que se acaban de presentar, por lo que se recomienda su estudio en los diferentes textos de diseño VLSI para tecnología CMOS.

a) Circuitos con esquema de temporización de dos fases no-sobrelapadas.

En estos circuitos se dispone de  $\phi_1$ ,  $\phi_2$ ,  $\bar{\phi}_1$  y  $\bar{\phi}_2$ . Es decir se dispone de hasta cuatro fases de reloj para ser distribuidas en el circuito.

En la Fig. 4.31 se presenta la estructura de un *latch* (DFF1) y su principio de operación. Durante  $\phi_1$  la compuerta de transmisión de la primera etapa se cierra y permite el almacenamiento del nivel de la entrada en C1. El estado de la segunda etapa se almacena de manera similar en C2. Cuando  $\phi_2=1$ , la compuerta de transmisión de la primera etapa se abre y el valor inverso de la capacitancia C1 se almacena en C2. Por la presencia de las compuertas de transmisión deben utilizarse las cuatro señales de reloj disponibles.

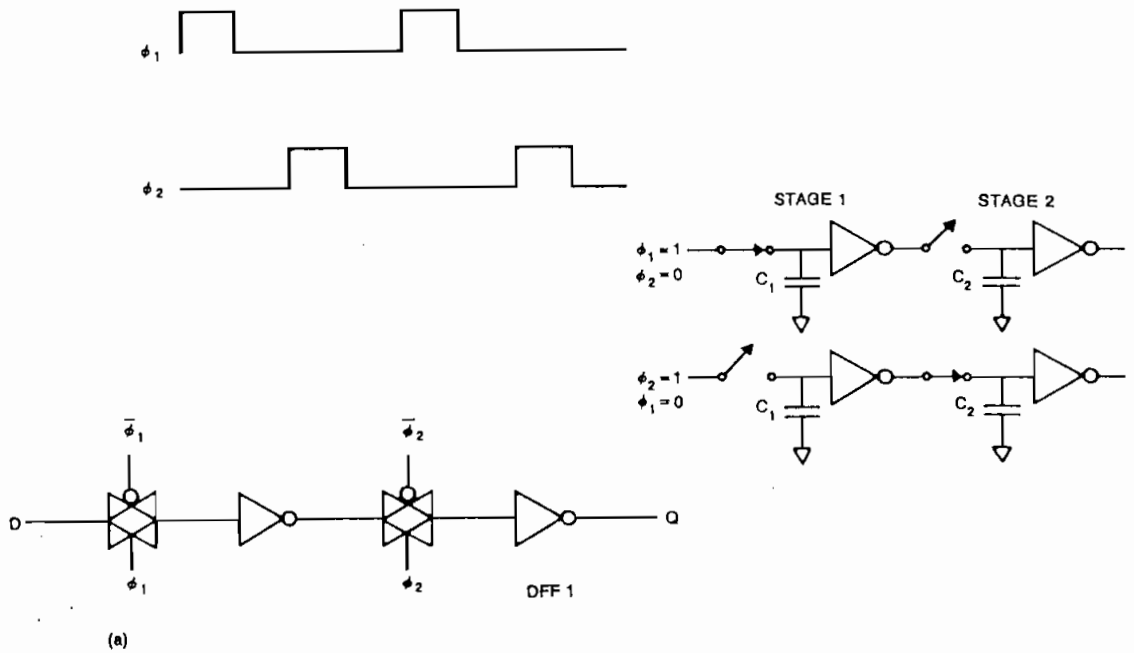


FIGURA 4.31 Latch DFF1 con su principio de operación y señales de reloj utilizadas.

Una forma alternativa del flip-flop DFF1 se presenta en la Fig. 4.32a (DFF2). DFF2 utiliza la estructura básica del inversor CMOS y su funcionamiento es muy similar al de DFF1. Cuando  $\phi_1=1$ , el nodo n1 va a un estado alto o bajo, de acuerdo a la señal de entrada. Cuando  $\phi_1=0$ , este valor es almacenado y los transistores manejados por la señal  $\phi_1$  y su complemento se "apagan". La segunda etapa opera de manera similar, pero controlada por  $\phi_2$  y su complemento.

En la Fig. 4.33a se presenta una posible configuración para un latch dinámico tipo D (DFF3). La entrada se denota como D y la salida como Q, la señal LD determina los momentos de "carga" de los datos.

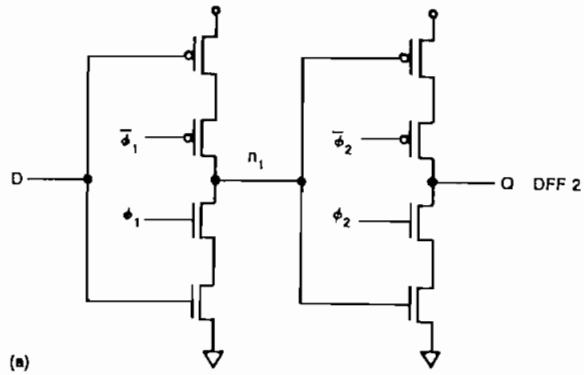


FIGURA 4.32 Latch DFF2 basado en la estructura del inversor CMOS.

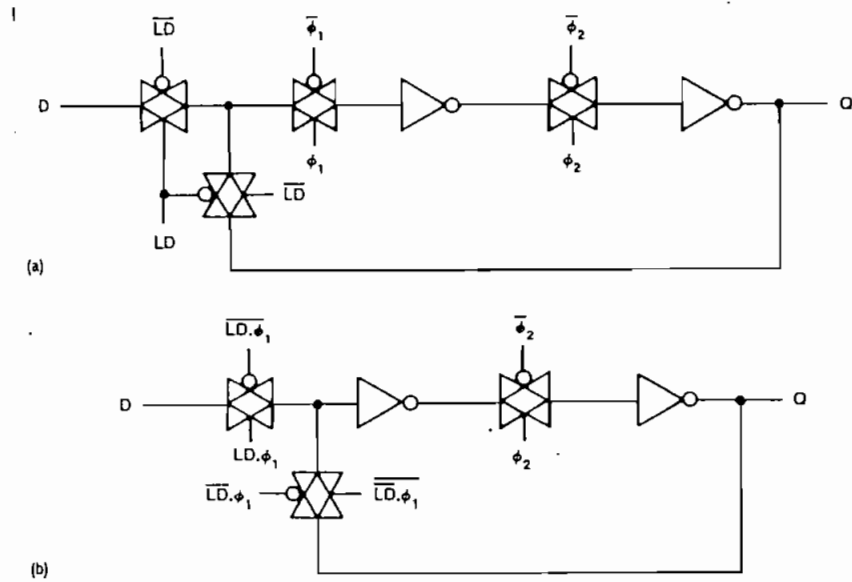


FIGURA 4.33 Latch dinámico tipo "D" (DFF3).

Las ecuaciones características que determinan la operación de este latch se escriben como:

$$[ Q(t) = D(t) \text{ si } LD=1 ] \quad \text{y} \quad [ Q(t) = Q(t-1) \text{ si } LD=0 ] .$$

En la Fig. 4. 33b se presenta una estructura alternativa (DFF4), en la que se realiza previamente operaciones lógicas con las señales LD y  $\phi_1$ .

b) Esquema de temporización de una fase

Cuando sea necesario reducir el número de caminos destinados a la distribución de las señales de reloj, se utiliza el esquema de temporización de una fase complementada, que utiliza las señales  $\phi$  y  $\bar{\phi}$ .

En la Fig. 4.34 se presenta la estructura de un flip-flop dinámico (DFF5), utiliza compuertas de transmisión e inversores. DFF5 tiene una estructura muy similar a DFF1, pero usa solamente  $\phi_1$  y su complemento.

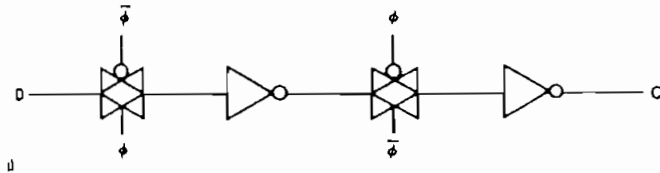


FIGURA 4.34 Latch dinámico DFF5.

En la Fig. 4.35 se presenta una alternativa para un latch estático (DFF6) utilizando este esquema de temporización.

Las dos estructuras presentadas (DFF5 y DFF6) tienen los problemas mencionados anteriormente para este tipo de esquema de temporización, principalmente debido al "clock skew".

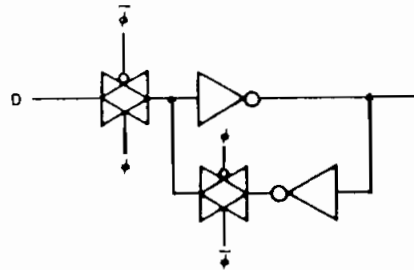


FIGURA 4.35 Latch estático DFF6.

En la Fig. 4.36 se presentan la estructura de un flip-flop estático (DFF7), del tipo maestro-esclavo, incluidas las señales de SET y RESET, y utiliza el esquema de temporización de una fase.

La principal característica del flip flop maestro-esclavo es que es inmune a los problemas de temporización de una fase, gracias a su estructura y funcionamiento. Las dos etapas de este flip flop pueden aislarse como en el latch dinámico de la Fig. 4.29. La desventaja de este flip flop es el gran número de transistores que requiere y por ende de un área considerablemente mayor que utiliza.

De esta última estructura (DFF7) hay variaciones que utilizan las señales  $\phi$  y  $\bar{\phi}$ , y que tratan de mantener sus características, pero a la vez optimizar el área utilizada.

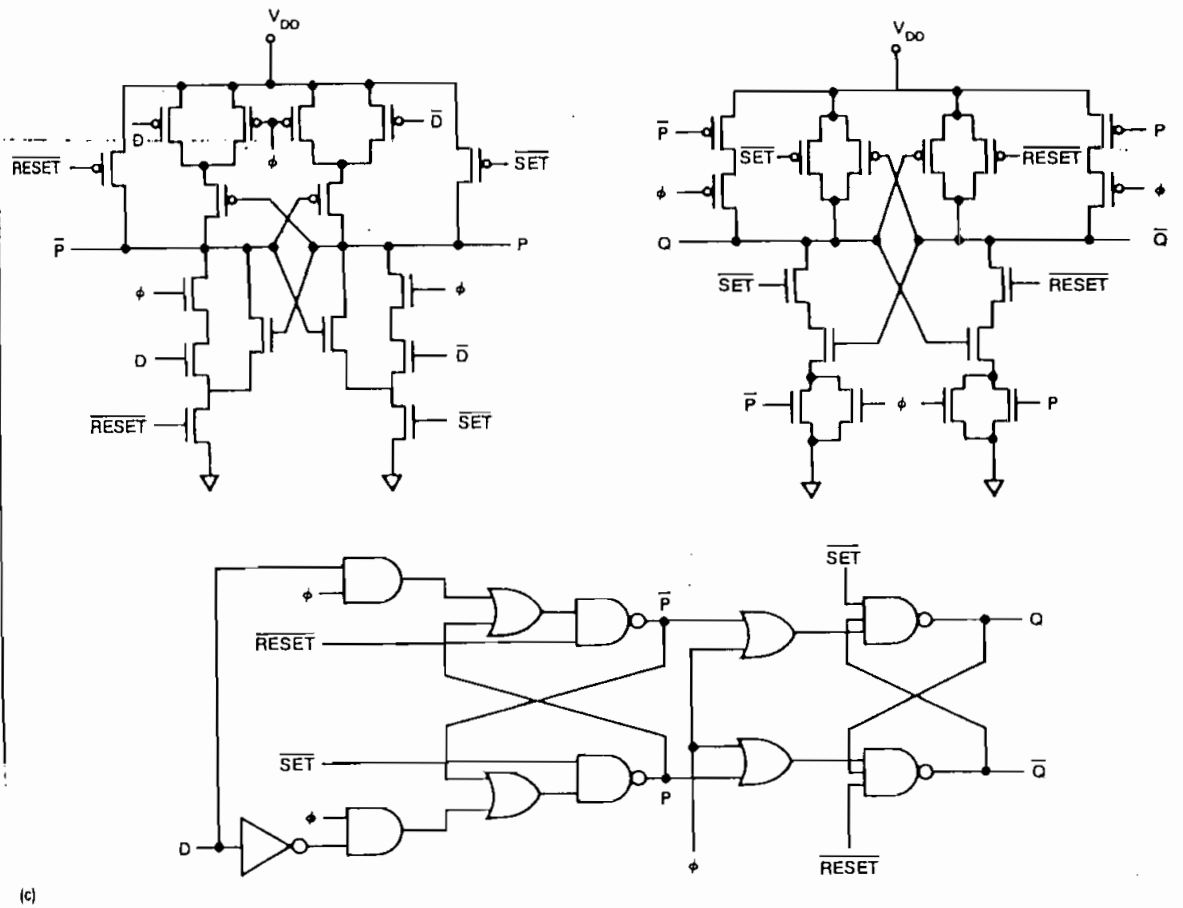


FIGURA 4.36 Flip-flop estático maestro-esclavo DFF7.