DESCRIPCIÓ DE L'APLICACIÓ

L'aplicació web actual té com a objectiu principal realitzar un seguiment tant per part de l'alumne com per part del profesor del grau de dedicació a l'assignatura de Disseny Microelectrònic I (DMI) de forma personalitzada.

L'aplicació web té dues zones molt ben diferenciades amb diferents privilegis d'administració: la **zona del profesor** o encarregat de l'assignatura que permet tenir un accés total a la base de dades sabent així la dedicació de cada alumne, i la **zona de l'alumne** que té accés limitat a la seva dedicació i no pot consultar la de la resta d'alumnes.

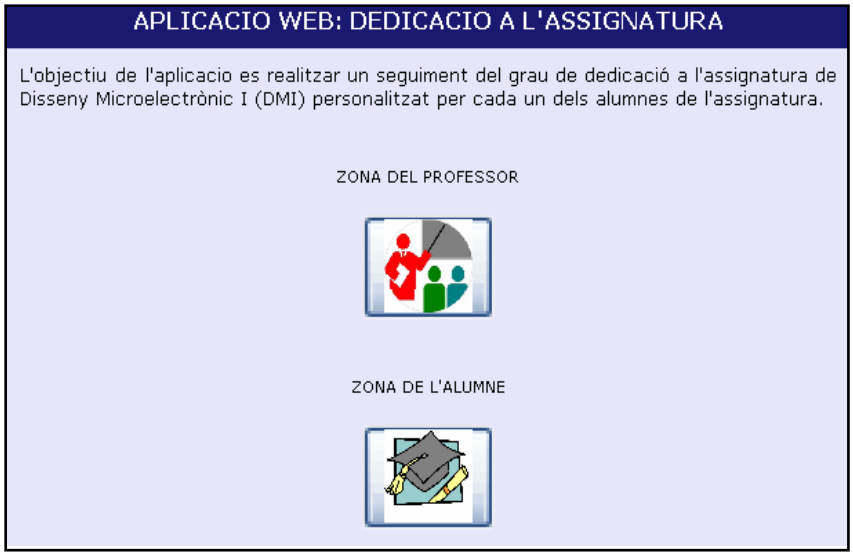


Figura 1: Menú principal

ZONA DE L'ALUMNE

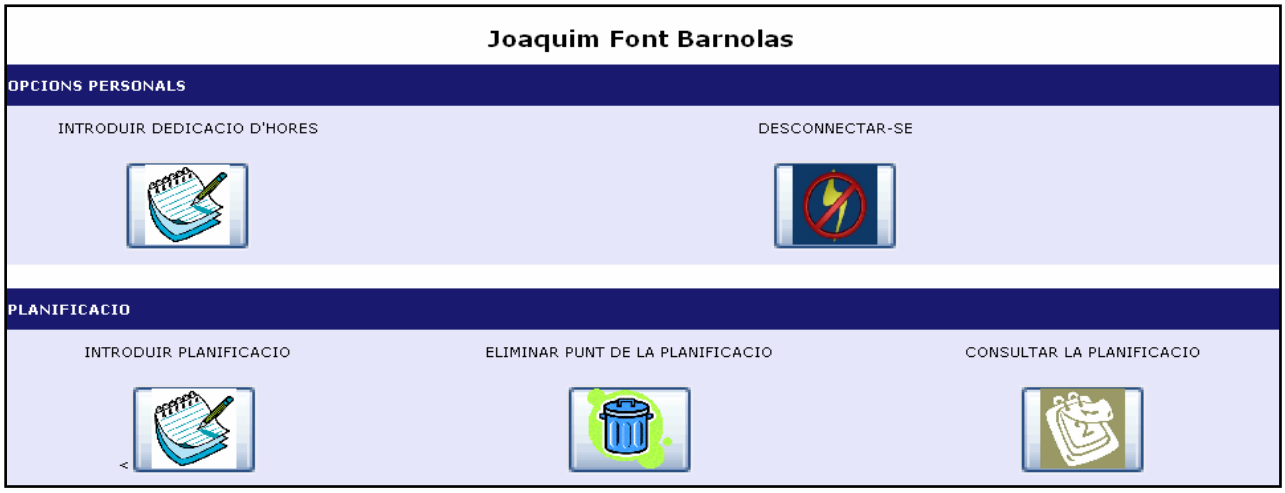


Figura 2: Zona de l'alumne

La zona de l'alumne permet introduir i consultar la dedicació d'hores realitzades en cada un dels temes de l'assignatura (també pot veure, en cas de que el professor ho desitgi, la dedicació mitja de tots els alumnes en cada un dels temes).
Per altra banda, l'alumne també disposarà d'una eina per realitzar una planificació de tasques en l'assignatura tals com elaboració de treballs, realització de pràctiques i activitats…

Figura 3: Zona del professor

La zona de l'alumne permet tenir un accés total a totes les dades de la base de dades, realitzant consultes així com realitzar modificacions en ella. Per tant permet realitzar consultes sobre la dedicació de cada un dels alumnes, així com modificar les seves dades i preferències. També permet realitzar modificacion sobre els punts del temari de l'assignatura.

Persones
Participants

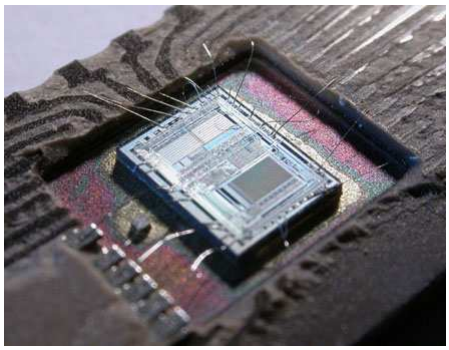
Activitats
Fòrums
Recursos

Cerca fòrums
Endavant
Cerca avançada

Administració
Activa edició
Paràmetres
Assignació de rols
Qualificacions
Competències
Grups
Còpia de seguretat
Restaura
Importa
Reinicia
Informes
Preguntes
Fitxers
Cancel·la la meva inscripció en 0809DMI5
Perfil

Els meus cursos
Disseny Microelectrònic I Intranet del Professorat
Circuits electrònics per a microsistemes
Circuits electrònics per a microsistemes
Disseny Microelectrònic I
Disseny Microelectrònic I
Projecte de Fi de Carrera
Tots els cursos ...

Disseny Microelectrònic I



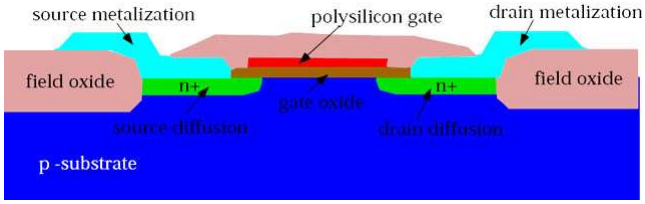
Al llarg dels últims anys, el sector de la microelectrònica ha evolucionat de manera espectacular gràcies al constant augment de la densitat d'integració dels circuits integrats, arribant a l'extrem d'empaquetar en un únic xip centenars de milions de components que implementen tots els subsistemes necessaris per una aplicació.

A l'assignatura de 'Disseny Microelectrònic I' es proporcionaran tots els coneixements necessaris per aprendre a dissenyar circuits integrats, especialment els digitals. Es presentaran diferents alternatives tecnològiques, parant especial atenció en la tecnologia CMOS, ja que és la més utilitzada avui en dia. També s'exposaran les diferents eines CAD que donen suport a les diverses etapes que formen part del disseny i que permeten agilitzar tot el procés. Finalment, es farà una introducció al test de sistemes electrònics.

Sinopsi
Fòrum de notícies

1

1. El procés de fabricació d'un circuit CMOS. Tecnologia microelectrònica.



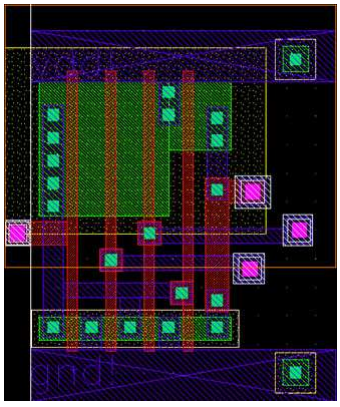
Objectius
1. Tipus de processos
2. Procés de fabricació

Per més informació
Transparències de classe 1
Transparències de classe 2
Llibre

Llibre bàsic del tema [Kang99]

2

2. Disseny físic de circuits integrats.



Objectius
1. Què és un layout?
2. Gràfic d'Euler
3. Diagrama de Sticks
4. Capacitats en referència al layout

Per més informació
Transparències de classe
Llibre

Llibres bàsics del tema [Uyemura88] i [Kang02]

3

3. El transistor MOS. Estructura i operació.

Últims notícies
Afegeix un nou tema...
(Encara no s'han enviat notícies)

Esdeveniments pròxims
Festiu: Nadal
dijous desembre 25
Festiu: Sant Esteve
divendres desembre 26
Festiu: Cap d'any
dijous gener 1
Festiu: Reis
dimarts gener 6
Vés al calendari...
Nou esdeveniment...

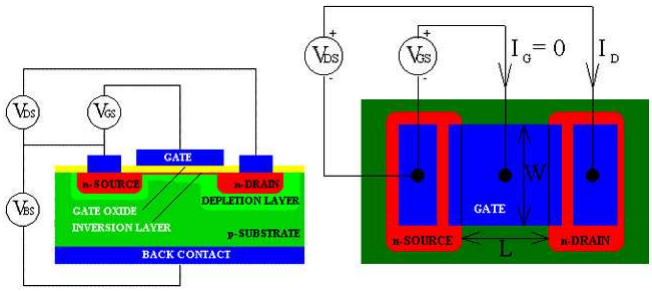
Activitat recent
Activitat des de dilluns, 15 desembre 2008, 15:01
Informe complet d'activitat recent...
Cap novetat des de l'última entrada

Missatges
No teniu missatges pendents
Missatges...

Calendari
◄ desembre 2008 ►

dl	dt	dc	dj	dv	ds	dg
1	2	3	4	5	6	7
8	9	10	11	12	13	14
15	16	17	18	19	20	21
22	23	24	25	26	27	28
29	30	31				

Tipus d'esdeveniments
Global Curs
Grup Usuari



Objectius

1. Estructura MOS
2. MOSFET d'enriquiment/empobriment de canal n/p
3. Regions de treball del transistor MOS
4. Efectes resistius, modulació del canal i efecte substrat

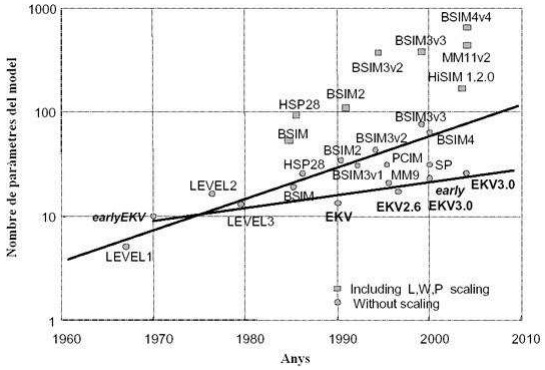
Per més informació

Transparències de classe
Llibre
Problemes

Llibres bàsics del tema [Rubio00] i [Rabaey02]

4

A1. Models del transistor MOS.



La simulació de circuits electrònics es realitza a partir de models que descriuen el comportament dels components que integren el circuit. Aquests models s'extreuen del comportament real de dispositius aïllats a través d'expressions amb base física o bé mitjançant ajustament empíric a les seves corbes de resposta. A continuació, es descriuen els més característics pel transistor MOS.

Objectius

1. Level 1
2. Level 2
3. Level 3
4. BSIM3v3
5. BSIM4
6. EKV

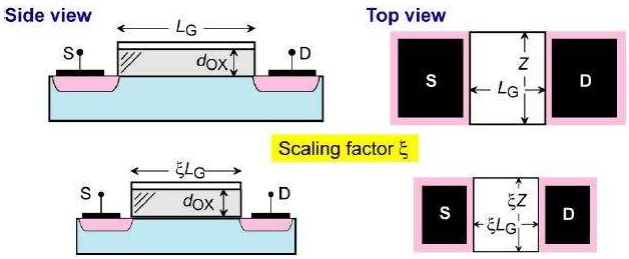
Per més informació

Llibre

Llibre bàsic del tema [Sischka]

5

A2. Escalat tecnològic.



L'escalat tecnològic és una tècnica que es basa en la reducció de les dimensions del MOSFET i d'altres característiques VLSI (gruix de l'òxid de porta o tensió d'alimentació, per exemple) per tal d'obtenir certes millores, com són l'augment de la densitat de transistors, l'augment de la freqüència de treball i la disminució de la potència dissipada.

Objectius

1. Tipus d'escalat
2. Avantatges
3. Limitacions
4. Futures solucions

Per més informació

Llibre

Llibre bàsic del tema [Chang00]

6

A3. Tècniques de connexionat i empaquetat dels CI.

Objectius

1. Procés de muntatge i empaquetat

2. Connexionat del CI
3. Tipus d'encapsulat

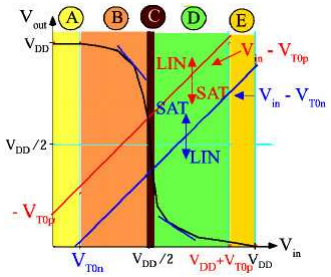
Per més informació

Llibre

Llibre bàsic del tema [Sze88]

7

4. Inversor CMOS. Característiques estàtiques.



L'estudi detallat de l'inversor CMOS permet entendre el funcionament de portes més complexes, d'aquí la seva importància.

Objectius

1. Taula de la veritat i esquemàtic
2. Equacions dels transistors de l'inversor per cada regió de treball
3. Corba de transferència de voltatge (CTV)
4. Paràmetres característics
5. Disseny

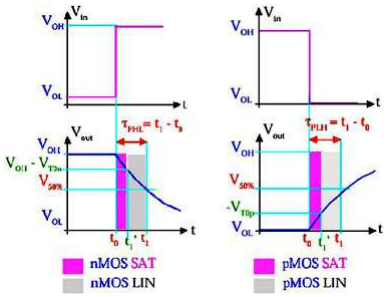
Per més informació

Transparències de classe
Llibre
Problemes

Llibres bàsics del tema [Weste93] i [Uyemura88]

8

5. Inversor CMOS. Característiques dinàmiques.



Objectius

1. Definició de paràmetres característics
2. Capacitats associades
3. Commutació
4. Disseny

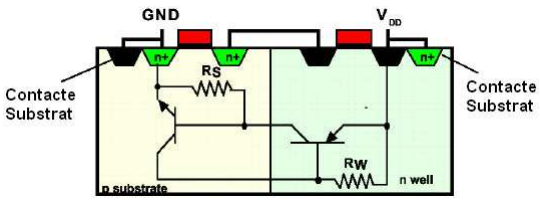
Per més informació

Transparències de classe
Llibre
Problemes

Llibres bàsics del tema [Rabaey02] i [Bota97]

9

6. Consum de corrents en CMOS.



La potència dissipada total és deguda a quatre contribucions, que són l'estàtica, la dinàmica, la de curtcircuit i la de corrent de fuites.

Objectius

1. Dissipació de potència estàtica
2. Dissipació de potència dinàmica
3. Dissipació de potència de curtcircuit
4. Dissipació de potència per corrent de fuites

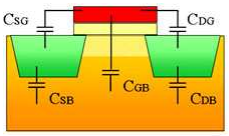
Per més informació

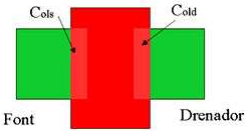
Transparències de classe
Llibre
Problemes

Llibre bàsic del tema [Rubio00]

10

7. Diseny de portes CMOS.





Objectius

1. Què és un transistor MOS?
2. Capacitats paràsites
3. Disseny dels blocs d'una porta lògica CMOS
4. Disseny de cada bloc a nivell de transistor
5. Pitjor cas de càrrega i descàrrega
6. Relació d'aspecte

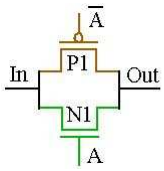
Per més informació

Transparències de classe 1
Transparències de classe 2
Llibre
Problemes

Llibres bàsics del tema [Uyemura94] i [Weste93]

11

8. Lògiques estàtiques alternatives a la lògica CMOS complementària.



Quan A='1', es permet el pas de '0' i '1'.
Quan A='0', la sortida es troba en estat d'alta impedància.

Objectius

1. Lògica pseudo-nMOS
2. Lògica BiCMOS
3. Lògica amb portes de pas
4. CVSL
5. CPL i DSL
6. SFPL

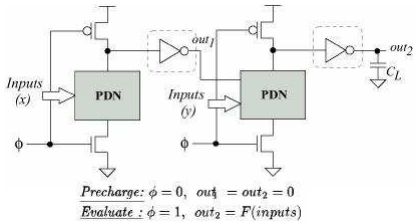
Per més informació

Transparències de classe 1
Transparències de classe 2
Llibre

Llibres bàsics del tema [Weste93] i [Uyemura94]

12

9. Lògica dinàmica.



Precharge: $\phi = 0,\ out_1 = out_2 = 0$
Evaluate : $\phi = 1,\ out_2 = F(inputs)$

Objectius

1. Precàrrega i avaluació (PE)
2. Lògica dominó
3. Lògica NP
4. Lògica C2MOS

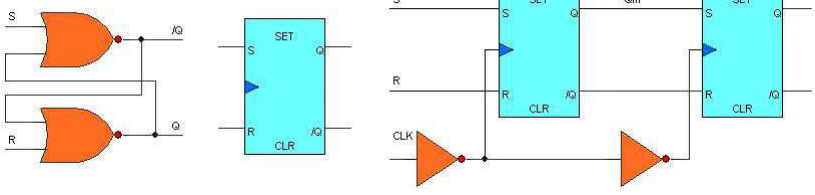
Per més informació

Transparències de classe 1
Transparències de classe 2
Llibre
Problemes

Llibres bàsics del tema [Rabaey02] i [Weste93]

13

10. Disseny de circuits seqüencials.



Objectius

1. Latches
2. Flip-flops
3. Temps en latches
4. Temps en flip-flops

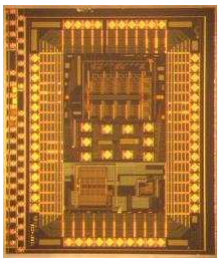
Per més informació

Transparències de classe
Llibre

Llibre bàsic del tema [Herms]

14

A4. Estructures avançades: lògica autotemporitzada.



Objectius

1. Disseny
2. Avantatges de la lògica asíncrona
3. Inconvenients de la lògica asíncrona
4. Detecció del final de l'operació
5. Aplicacions

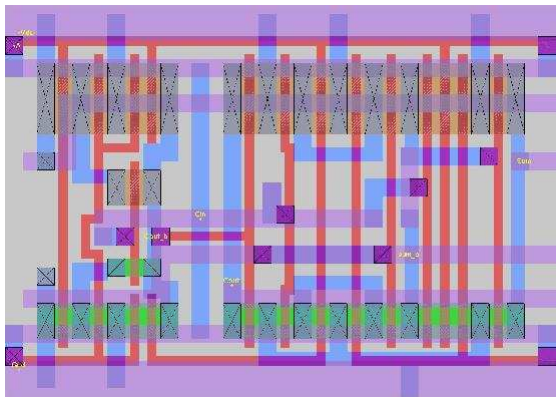
Per més informació

Llibre

Llibre bàsic del tema [Hauck95]

15

A5. Disseny de subsistemes digitals: sumadors.



Objectius

1. Sumador binari bàsic
2. Disseny en lògica CMOS
3. Disseny en lògica dinàmica
4. Sumadors de segmentació
5. Altres configuracions lògiques

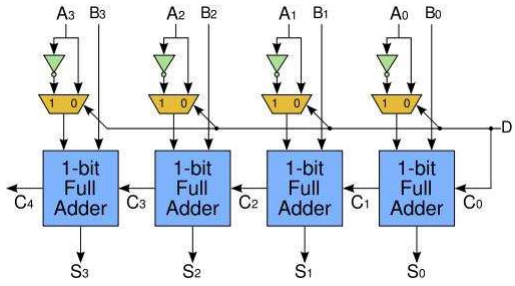
Per més informació

Llibre

Llibre bàsic del tema [Rabaey96]

16

A6. Disseny de subsistemes digitals: multiplicadors.



Objectius

1. Multiplicador fonamental
2. Classificador de multiplicadors
3. Implementació de multiplicadors

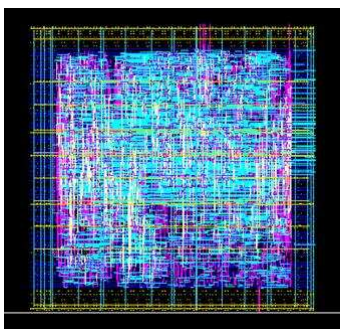
Per més informació

Llibre

Llibre bàsic del tema [Herms]

17

11. Disseny de sistemes VLSI.



Objectius

1. Disseny full custom
2. Disseny semi custom
3. FPGA
4. Comparació de tecnologies

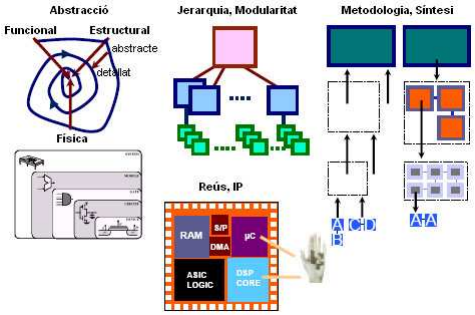
Per més informació

Transparències de classe 1
Transparències de classe 2
Llibre

Llibre bàsic del tema [Smith97]

18

12. Tractant la complexitat del disseny VLSI.



Objectius

1. Disseny multinivell
2. Dominis i nivells d'abstracció

Per més informació

Transparències de classe
Llibre

Llibre bàsic del tema [Rubio00]

19

13. Llenguatges d'alt nivell per disseny VLSI.

VHDL

```
library IEEE;
use IEEE.STD_Logic_1164, all;

entity LATCH_IF_ELSEIF is
  port (En1, En2, En3, A1, A2, A3: in std_logic;
        Y: out std_logic);
end entity LATCH_IF_ELSEIF;

architecture RTL of LATCH_IF_ELSEIF is
begin
  process (En1, En2, En3, A1, A2, A3)
  begin
    if (En1 = '1') then
      Y <= A1;
    elseif (En2 = '1') then
      Y <= A2;
    elseif (En3 = '1') then
      Y <= A3;
    end if;
  end process;

end architecture RTL;
```

Verilog

```
module LATCH_IF_ELSEIF  (En1, En2, En3, A1, A2, A3,Y);
    input En1, En2, En3, A1, A2, A3;
    output Y;

    reg Y;

    always @(En1 or En2 or En3 or A1 or A2 or A3)
       if (En1 == 1)
          Y =A1;
     else if (En2 == 1)
          Y = A2;
     else if (En3 == 1)
          Y = A3;

end module
```

Objectius

1. Concepte d'eines CAD-EDA
2. Llenguatges de descripció hardware

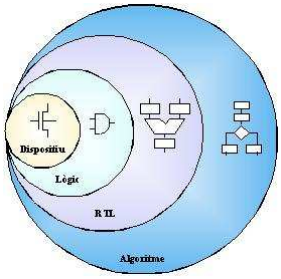
Per més informació

Llibre

Llibre bàsic del tema [Pardo99]

20

14. Síntesi.



La síntesi és el procés de transformació de la descripció conductual d'un disseny en un *layout*. Donat que assolir aquest objectiu en un sol pas és una tasca molt complexa, sinó impossible, se subdivideix el problema en diferents processos més simples.

Objectius

1. Síntesi funcional
2. Síntesi lògica
3. Síntesi física
4. Validació

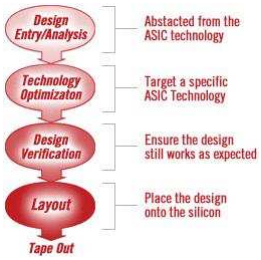
Per més informació

Transparències de classe
Llibre

Llibre bàsic del tema [Smith97]

21

A7. Eines comercials pel disseny VLSI.



Objectius

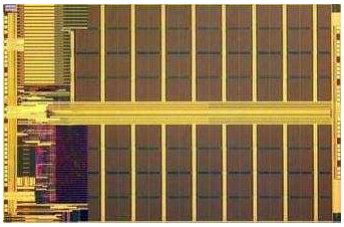
1. Eines per al flux de disseny
2. Eines comercials per al flux de disseny

Per més informació

Llibre

22

A8. Memòries VLSI.



Objectius

1. Arquitectura
2. Tipus de memòries
3. Circuiteria externa
4. Consum

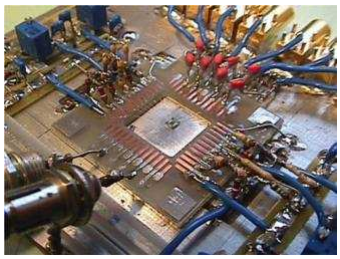
Per més informació

Llibre

Llibre bàsic del tema [Bota96]

23

15. Tècniques de test de circuits integrats.



El test és la tècnica que s'aplica després de la fabricació de cada circuit integrat per tal de determinar si aquest ha sigut fabricat correctament. Aquest propòsit pot dur-se a terme mitjançant diferents estratègies.

Objectius

1. Faltes i el seu origen
2. Model de faltes stuck-at
3. BIST
4. BILBO
5. Scan test

Per més informació

Transparències de classe 1
Transparències de classe 2
Llibre
Problemes

Llibres bàsics del tema [Abramovici90] i [Smith97]

24

Documentació de Moodle per a aquesta pàgina

Heu entrat com Dieguez Barrientos Angel (Sortida)

Inici

1.530482 secs
RAM: 21.1Mb
RAM peak: 21.2Mb
Included 91 files
ticks: 153 user: 25 sys: 3 cuser: 0 csys: 0
Load average: 1.02
Record cache hit/miss ratio : 0/0

ACTIVIDAD 2

Diseño a nivel lógico, transistor y físico del sistema de puertas $F = \overline{AB}$; $G = \overline{F + DE}$ con tecnología c35b4 de Austriamicrosystems. La capacidad en la salida de G es C_L=40fF. La capacidad de carga de F es la de entrada de G. Diseñad ambas puertas para que t_r=t_f=0.4 ns. Haced el layout (ambas celdas han de estar conectadas siguiendo una filosofia 'standard cell') lo mas compacto posible. Para el layout:
Altura celda=12µm. Ancho VDD/VSS=2µm. Posición VDD/VSS=0.4µm desde el borde.

Duración del Martes 04-11-08 al Miercoles 12-11-08
Sesión presencial (en aula S11): 04, 06, 10, 12 | **Presentación**: Miercoles 12-11-08

Definición de grupos para la realización de la actividad:

Grupo 1	Grupo 2	Grupo 3	Grupo 4

Metodología a seguir para el desarrollo de la actividad:

La misma que para la actividad anterior. Podeis cambiar el líder.

Material a entregar por cada grupo:

Cada grupo deberá entregar una carpeta con la siguiente información:

- ■ Resultados de la actividad
  - □ Esquema lógico
  - □ Esquema a nivel transistor indicando los tamaños de transistor[I]
  - □ Layout en papel milimetrado[II]
  - □ Datasheet de la puerta (con todos los parámetros necesarios). Incluid en el datasheet el error cometido en los tiempos si se tienen en cuenta todas las capacidades de la puerta después de hacer el layout.
  - □ Gráfico de análisis de la solución[III]
  - □ Comentar los puntos de mayor discusión o duda
  - □ Describir las tareas que ha realizado cada miembro del grupo[IV]
- ■ Autoevaluación de la actividad[V]
  - □ Cada miembro del grupo realizará una tabla donde describirá la/s actividad/es realizadas por los otros miembros del grupo y las evaluará de 0 a 10.
- ■ Encuesta de esfuerzo/valoración de la actividad (ver pagina siguiente)

I Detallad de forma separada todo el proceso de deducción de los tamaños de los transistores.
II Discutid brevemente cada una de las reglas de diseño aplicadas, el orden y el motivo de la existencia de la regla.
III En el gráfico de análisis debéis mostrar de forma libre y gráfica los pasos que han llevado a encontrar la solución propuesta de la forma más detallada posible. El gráfico no ha de ocupar más de una hoja.
IV Haced referencia al gráfico de análisis de la solución.
V La autoevaluación la enviará cada miembro del grupo a dieguez@el.ub.es y no se incluirá en la carpeta.

ENCUESTA DE ESFUERZO/VALORACIÓN

La encuesta de esfuerzo/valoración ha de presentarse en la carpeta por cada miembro del grupo. En la encuesta de esfuerzo debéis indicar el tiempo que os ha llevado individualmente cada parte de la actividad. En la encuesta de valoración debéis puntuar los aspectos indicados en la tabla.

ENCUESTA DE ESFUERZO – ACTIVIDAD 1	
Tiempo dedicado a reuniones de grupo (h)	
Tiempo dedicado a búsqueda de información (h)	
Tiempo dedicado a la redacción de resultados (h)	
Esquema lógico (h)	
Esquema a nivel transistor (h)	
Layout (h)	
Gráfico de análisis (h)	
Datasheet y cálculo de parámetros(h)	
Tiempo dedicado a la preparación de la presentación (h)	
Tiempo dedicado a la evaluación de otros miembros del grupo (h)	
Tiempo dedicado a la encuesta de esfuerzo/valoración (h)	
Otros (especificad) (h)	
Tiempo total de dedicación a la actividad (h)	

ENCUESTA DE VALORACIÓN – ACTIVIDAD 1	
Interés de la actividad en relación a los contenidos de la asignatura (0-10)	
Interés de la actividad en relación a la evaluación de la asignatura (0-10)	
La actividad aclara los conceptos que en ella se tocan? (0-10)	
La actividad te ha hecho observar tus capacidades de trabajo en grupo? (0-10)	
La actividad te ha hecho observar tus capacidades de liderazgo? (0-10)	
La actividad promueve la mejora del trabajo en grupo? (0-10)	
La actividad mejora la relación con tus compañeros? (0-10)	
Valora lo que has aprendido en relación a la clase magistral (0-10)	

DISSENY MICROELECTRÒNIC I

Angel Diéguez Barrientos
Eva Vilella Figueras

Departament d'Electrònica
Universitat de Barcelona

Continguts

Introducció 2

BLOC I

1. El procés de fabricació d'un circuit CMOS. Tecnologia microelectrònica 3
2. Disseny físic de circuits integrats 9
3. El transistor MOS. Estructura i operació 14
   - A1. Models del transistor MOSFET 21
   - A2. Escalat tecnològic 33
   - A3. Tècniques de connexionat i empaquetat dels CI 39
4. Inversor CMOS. Característiques estàtiques 42
5. Inversor CMOS. Característiques dinàmiques 48
6. Consum de corrents en CMOS 52
7. Disseny de portes CMOS 57
8. Lògiques estàtiques altenatives a la lògica CMOS complementària 65
9. Lògica dinàmica 70
10. Disseny de circuits seqüencials 73
    - A4. Estructures avançades: Lògica autotemporitzada 79
    - A5. Disseny de subsistemes digitals: sumadors 84
    - A6. Disseny de subsistemes digitals: multiplicadors 91

P1. Problemes 97

Introducció

L'objectiu principal d'aquest llibre és servir com a guia i punt de referència pels alumnes de l'assignatura 'Disseny Microelectrònic I', que s'imparteix a l'ensenyament d'Enginyeria Electrònica de la Universitat de Barcelona.

A l'esmentada assignatura es proporcionen tots els coneixements necessaris per aprendre a dissenyar circuits integrats, especialment els digitals. També es donen a conèixer diferents alternatives tecnològiques, parant especial atenció en la tecnologia CMOS, ja que és la més utilitzada avui en dia, i abarcant tant el disseny de circuits combinacionals com seqüencials.

Aquest text es presenta com un recull d'articles tècnics, els quals s'organitzen en temes independents segons l'estructura de la mateixa assignatura. Els articles se subdivideixen en obligatoris, imprescindibles pel correcte seguiment de l'assignatura, i d'ampliació, pensats per aquells alumnes que desitgin aprofundir els seus coneixements. A més, tot plegat es complementa amb una col·lecció de problemes resolts que, juntament amb la resta de material de classe, pretén servir tant pels alumnes de l'assignatura com per aquelles persones que vulguin conèixer o revisar algun dels aspectes que s'hi tracten.

L'elaboració d'aquest llibre s'ha dut a terme dins del projecte Ajut per la Millora de la Qualitat Docent a les Universitats de Catalunya (codi 2006MQD00067).

1. El procés de fabricació d'un circuit CMOS. Tecnologia microelectrònica.

***Abstract*—**En aquest article s'explicaran els diferents processos tecnològics que es porten a terme per obtenir circuits integrats, la totalitat dels quals constitueix el que s'anomena tecnologia planar. Així mateix, també es descriurà el procés complet de fabricació d'un circuit CMOS, concretament el d'un inversor de pou p, tot detallant les màscares que són aplicades en cada moment.

***Index Terms*—**Tecnologia planar, fotolitografia, màscara, oxidació, difusió, implantació iònica, dipòsit, epitàxia, *etching*, metal·lització, LOCOS, pou, àrea activa, polisilici, metall, contacte, via.

1. INTRODUCCIÓ

La fabricació d'un circuit integrat és una tasca complexa que consta d'una seqüència de processos independents, els quals s'anomenen fotolitografia, oxidació, dopatge (ja sigui per difusió o implantació iònica), dipòsit i *etching* o gravat. La totalitat d'aquests processos s'identifica mitjançant l'expressió tecnologia de procés planar.

D'altra banda, pot afirmar-se que un circuit CMOS és la interconnexió de transistors pMOS i nMOS i, per tant, un procés CMOS és la seqüència d'operacions tecnològiques que produeixen transistors pMOS i nMOS en una mateixa oblia.

2. TIPUS DE PROCESSOS

Per la creació de transistors a l'oblia de silici s'usen diferents processos tecnològics, els quals s'esmenten a continuació.

2.1. Fotolitografia

La fotolitografia és el procés de transferència de la geometria de cadascuna de les capes del circuit integrat sobre la superfície de l'oblia. Per obtenir un procés CMOS, es necessitaran aproximadament tantes màscares com capes de què consti aquest procés, tal com es mostra a la Taula 1. El nombre de processos tecnològics és molt més elevat, tenint de l'ordre de 600 etapes.

Màscara	Capa del procés CMOS
1	Àrea de pou n
2	Àrea activa
3	Polisilici
4	Implantació n^+
5	Implantació p^+
6	Contactes
7	Primer nivell de metall
8	Via
9	Segon nivell de metall
10	Passivació

Taula 1 Màscares bàsiques per la fabricació d'un procés CMOS.

En el procés fotolitogràfic, tota l'oblia es recobreix amb fotoreïna, aconseguint que aquesta presenti un gruix uniforme gràcies a un sistema de rotació per *spinner*. Tot seguit, s'aplica llum ultraviolada a través d'una màscara, la qual està formada per una placa de vidre de gran puresa. La màscara consta de zones transparents i d'altres zones opaques en resposta a la geometria que s'ha de transferir a l'oblia. Gràcies a aquestes zones transparents s'aconsegueix el revelat de la fotoreïna i el posterior gravat de la capa (fotoreïna positiva), mentre que les regions de l'oblia que es troben sota d'una regió transparent romanen intactes. Finalment, s'elimina la fotoreïna restant i es neteja.

Un aspecte molt important que ha de ser tingut en compte en aquesta part del procés és l'alineament entre la màscara i el substrat de l'oblia, cosa que s'aconsegueix mitjançant una màquina anomenada alineadora o *stepper*.

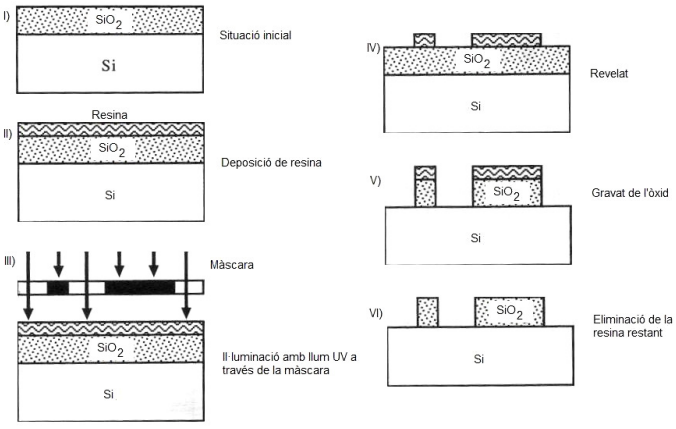


Fig. 1 Diferents etapes del procés fotolitogràfic.

2.2. Oxidació

L'oxidació és aquella tècnica que s'utilitza per fer créixer una capa de diòxid de silici (SiO_2) sobre l'oblia, el qual pot actuar com a màscara per la difusió, com a capa dielèctrica aïllant o com a font d'impureses. El procés d'oxidació, que pot tenir lloc gràcies a dues reaccions diferents, utilitza la calor com a catalitzador, pel que l'òxid

que creix rep el nom d'òxid tèrmic. Les dues reaccions possibles, que s'anomenen oxidació seca i oxidació humida, són les que es mostren tot seguit.

$$Si + O_2 \rightarrow SiO_2$$
$$Si + 2H_2O \rightarrow SiO_2 + 2H_2$$

Val a dir que l'oxidació seca, que produeix un SiO_2 més dens i de millor qualitat, és molt més lenta que la humida. Al forn on té lloc la reacció s'estableix una temperatura que pot variar entre 900 i 1200ºC. L'òxid de silici necessita Si per créixer (de fet es consumeix part del Si durant el creixement). Això vol dir que no creix si no hi ha Si a la superfície.

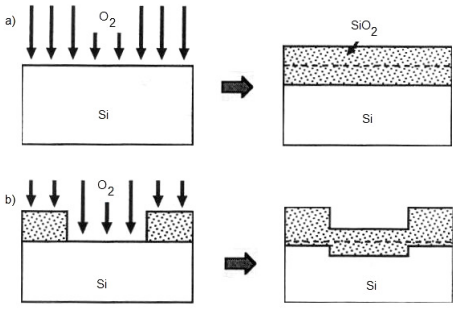


Fig. 2 Exemple del procés d'oxidació.

2.3. Dopatge

El dopatge és el procés mitjançant el qual s'afegeixen petites quantitats d'impureses en un material semiconductor per tal de modificar les seves propietats conductores. Aquest pot fer-se mitjançant dues tècniques, la difusió o la implantació iònica.

2.3.1. Difusió

La difusió era antigament el mètode més utilitzat per introduir, de forma controlada, impureses al substrat de silici i té lloc sempre que hi ha un procés a alta temperatura. Aquest procés està governat per dues lleis, anomenades lleis de Fick, les quals anuncien que el flux de partícules que difonen és proporcional al gradient de concentració i que la variació temporal de la concentració és funció de la concentració volúmica.

Generalment, la difusió es realitza en dues etapes. Durant la primera té lloc el predipòsit, que consisteix en introduir una quantitat específica de dopant a través de la superfície del silici, cosa que s'aconsegueix exposant l'esmentada superfície a un cert gas que conté material dopant. A la segona etapa, en canvi, té lloc la redifusió, en la qual les impureses són conduïdes a l'interior del material fins una profunditat considerable, sempre i quan l'oblia hagi sigut prèviament escalfada durant un període de temps llarg.

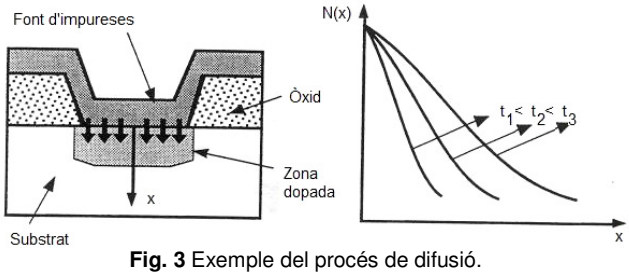


Fig. 3 Exemple del procés de difusió.

D'altra banda, durant el procés de difusió també sol tenir lloc la generació d'una capa prima d'òxid que serveix de passivació per processos posteriors. Per tant, hi ha consum superficial de silici, es modifica la distribució de dopants prop de la interfície amb l'òxid i pot haver-hi pèrdua d'impureses a través de la capa d'òxid.

2.3.2. Implantació iònica

La implantació iònica és el procediment utilitzat en tecnologies MOS per aconseguir regions amb perfil de dopatge tipus p o n. Consisteix en bombardejar la mostra amb ions accelerats de la impuresa adequada, els quals penetren el material. Per aquest propòsit, les impureses són primer ionitzades, posteriorment accelerades mitjançant camps elèctrics i finalment projectades sobre l'oblia, cosa que fa que xoquin de forma aleatòria amb els àtoms de silici fins que perden la seva energia i queden immobilitzats. Tot el procés té lloc al buit.

La profunditat de penetració de les impureses al substrat de silici pot controlar-se fàcilment a través de l'energia i la densitat de les partícules implantades, el que fa que aquesta tècnica sigui avantatjosa per sobre de la difusió. Com que els xocs són aleatoris, el perfil resultant correspon a una distribució gaussiana, tenint així poca dispersió lateral.

Per contra, és necessari implementar algun tipus de protecció per les zones que no han de veure's afectades pel procés de dopatge. Els materials habitualment utilitzats per aquesta finalitat són el SiO_2, el Si_3N_4 i, en algunes ocasions, la fotoreïna. A més, com a conseqüència que l'estructura cristal·lina del silici queda danyada a causa de les col·lisions de les impureses amb els àtoms del substrat, cal regenerar l'estructura mitjançant un breu tractament tèrmic.

2.4. Addició

L'addició és un procés fonamental en la fabricació de dispositius semiconductors. Tal com es veurà tot seguit, pot dur-se a terme mitjançant tècniques de dipòsit o bé d'epitàxia.

2.4.1. Dipòsit

Segons el tipus de material dipositat, es té dipòsit per procés no reactiu o bé dipòsit per procés reactiu. En el primer cas, el material dipositat no produeix cap reacció durant el procés, el qual s'anomena dipòsit físic en fase vapor (PVD). Les tècniques més habitualment utilitzades són l'evaporació i l'*sputtering*. La primera d'elles consisteix en escalfar el material que es vol dipositar fins que la pressió del vapor en permet una evaporació significativa, moment a partir del qual els àtoms

evaporats viatgen en línia recta fins al substrat. A l'*sputtering*, en canvi, els ions són accelerats gràcies a una tensió molt elevada i incideixen sobre el *target* del material a dipositar amb energia suficient com per arrancar-ne àtoms, els quals formen una capa uniforme que recobreix el substrat.

D'altra banda, en un procés reactiu el material que es diposita és el resultat d'una reacció química en fase vapor que té lloc sobre la superfície del substrat o prop d'aquest. Aquest tipus de processos són coneguts com processos de dipòsit químic en fase vapor (CVD). Aquesta tècnica presenta una sèrie d'avantatges, com són la possibilitat d'utilitzar una àmplia gamma de materials, el creixement en zones poc accessibles gràcies a una velocitat uniforme i el bon recobriment d'esglaons. Per dur a terme aquesta tècnica poden emprar-se diversos tipus de reactors.

2.4.2. Epitàxia

Consisteix en el creixement d'una capa prima de material semiconductor al damunt d'una altra capa, a partir d'una fase líquida o gasosa, tot seguint l'estructura de la de sota. L'epitàxia en fase vapor (VPE), que és la tècnica més important en dispositius de silici, habitualment utilitza l'hidrogen com a gas portador, encara que també pot emprar-se el clor. Durant el procés de creixement, el gas és degradat químicament obtenint així àtoms lliures, alguns dels quals es dipositen sobre el substrat cristal·lí. Tot el procés té lloc en un forn especial, anomenat reactor, on s'assoleixen temperatures elevades d'entre 1000 i 1200ºC. L'epitàxia en fase líquida, en canvi, s'utilitza principalment pel creixement d'una capa epitaxial de GaAs. En ambdós casos, les impureses de la capa de material que s'afegeix sobre el substrat queden uniformement distribuïdes, de manera que no es genera cap gradient de concentració.

2.5. Gravat (*Etching*)

Aquest procés permet eliminar el material no desitjat de les àrees de substrat no protegides per la fotoreïna o un altre material. És possible tenir una alta selectivitat, la qual permet atacar un material i no un altre, i anisotropia (una sola direcció) fent servir RIE (*Reactive Ion Etching*). Utilitzant mètodes químics l'atac no pot ser anisotròpic.

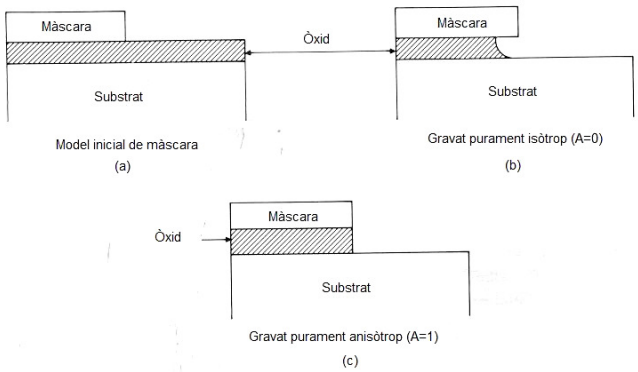


Fig. 4 Procés de gravat.

2.6. Metal·lització

Les capes de metall s'utilitzen per aconseguir connexions de baixa resistència mitjançant materials com l'alumini, tot i que també poden fer-se servir el coure o l'or. Aquestes capes metàl·liques s'obtenen a partir de la fusió del material i la seva posterior evaporació (gràcies a resistències calefactores, canons d'electrons o tècniques d'inducció de calor) en un sistema tancat al buit. Estrictament la metal·lització es un procés de dipòsit.

2.7. LOCOS

LOCOS, acrònim de *LOCal Oxidation of Silicon*, és un procés en el qual el diòxid de silici es forma sobre les àrees seleccionades de l'oblia de silici. Aquest creixement selectiu s'aconsegueix forjant les àrees actives amb nitrur de silici durant l'oxidació a una temperatura de 1000ºC.

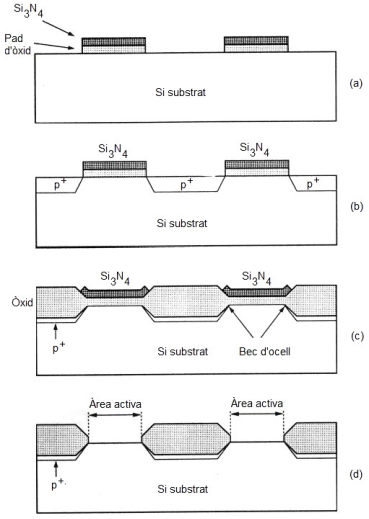


Fig. 5 Procés final LOCOS.

3. PROCÉS DE FABRICACIÓ

Per descriure el procés de fabricació d'un circuit CMOS s'analitzarà un inversor CMOS de pou p, tot i que aquesta dinàmica és generalitzable a altres tipus de circuits CMOS.

3.1. Oblia de silici

El primer pas consisteix en obtenir una oblia monocristal·lina amb un dopatge moderat, típicament de $10^{15}cm^{-3}$, mitjançant l'anomenat mètode de Czochralski. Els diàmetres típics de la barra d'on s'extrauran les oblies es troben entre els 10 i els 30cm, mentre que el seu gruix pot oscil·lar entre els 300 i els 600µm.

A continuació, té lloc el procés epitaxial per ampliar la capa de silici, però també per netejar la superfície de l'oblia, ja que això determinarà la qualitat dels dispositius. A més, sobre l'oblia s'aplica una capa de SiO_2 i, sobre aquesta, una capa de fotoreïna.

3.2. Àrea de pou n

La primera màscara fotolitogràfica defineix les regions de pou p. Llavors, s'elimina la fotoreïna i es crea el pou p

mitjançant implantació iònica amb un material acceptador, com per exemple el bor. La profunditat de la implantació no és excessiva, situant-se al voltant dels 3µm.

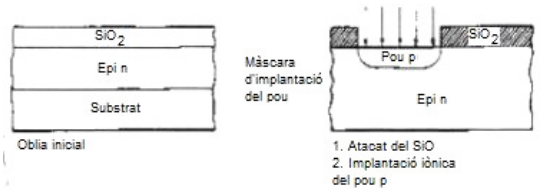


Fig. 6 Implantació de pou p.

A continuació, s'elimina la fotoreïna sobrant, així com també el diòxid de silici. Finalment, es crea una capa de nitrur (Si_3N_4) mitjançant un procés de dipòsit químic en fase vapor. El gruix aproximat d'aquesta capa és de 2µm.

3.3. Àrees actives

La segona màscara litogràfica defineix sobre la superfície de l'oblia les zones del canal, de *source* i de *drain*. També s'utilitza per delimitar les regions on actuen els anells de guarda, els quals s'encarreguen de prevenir l'efecte *latch-up*. S'irradia la superfície i, a continuació, s'eliminen el nitrur i la fotoreïna mitjançant atacs químics a les zones no delimitades per la màscara. El nitrur que no és eliminat serveix de protecció per les regions actives.

Seguidament, es fa créixer una capa d'òxid amb un gruix d'entre 0.8 i 1µm sobre tota la superfície, excepte en aquelles regions on hi ha nitrur. Aquest procés és l'anomenat LOCOS. Per entendre l'efecte d'aquesta màscara s'ha de recordar que no creix òxid allà on no hi ha silici. Per tant, la màscara d'àrea activa delimita les regions on no hi haurà òxid de camp (per aïllar dispositius), o el que es el mateix, on aniran els transistors.

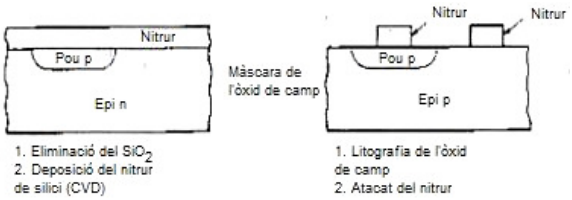


Fig. 7 Creació de les àrees actives.

Més tard, s'elimina el nitrur restant i té lloc el creixement de l'òxid de porta. Aquest, que es crea sobre tota la superfície de l'oblia, sol tenir un gruix petit d'entre 0.01 i 0.03µm, arribant a ser de 2nm en processos submicrònics de tecnologia de 100nm. Donat que aquest òxid ha de ser d'altíssima qualitat, ja que afecta directament la tensió llindar, les fuites i la capacitat dels transistors MOS, se sol utilitzar per la seva creació un procés en sec.

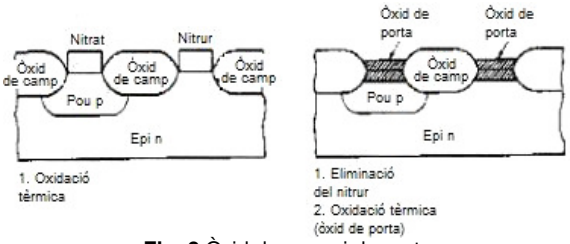


Fig. 8 Òxid de camp i de porta.

3.4. Formació de la porta de polisilici

Es dipositen una capa de polisilici i una de fotoreïna per tota la superfície de l'oblia. A continuació, s'aplica la tercera màscara litrogràfica, que defineix les regions de polisilici. La precisió d'aquesta etapa litogràfica és de crucial importància, ja que el polisilici serveix de màscara per les implantacions de *source* i *drain*. Variacions en l'amplada del polisilici, doncs, provocaran variacions en la longitud del canal del transistor. Al seu torn, aquest fenomen afectarà greument el rendiment del transistor.

Habitualment, després del dipòsit del polisilici, aquest és dopat amb fòsfor o bor per tal de disminuir la seva resistivitat. Aquest darrer punt és molt important, ja que el polisilici també és utilitzat en les interconnexions.

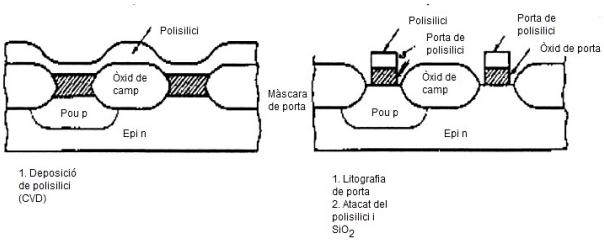


Fig. 9 Formació del polisilici.

3.5. Implantació de les àrees actives

En aquest procés tenen lloc la quarta i cinquena màscares, les quals delimiten les zones que seran exposades a la implantació iònica tipus n^+ i p^+, respectivament. Aquestes implantacions permeten la formació de les regions de *drain* i *source* dels transistors nMOS i pMOS, dels anells de guarda i dels contactes de polarització. Val a dir que en aquest procés només les regions actives pateixen la implantació. La resta de l'oblia es troba passivada.

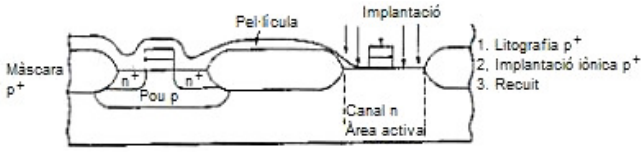


Fig. 10 Implantació de les àrees actives.

D'altra banda, mentre que la implantació n^+ es realitza mitjançant arsènic, la implantació p^+ s'efectua amb bor. La profunditat de penetració d'aquestes implantacions és de l'ordre de 0.5µm per les regions tipus n^+ i de 0.25µm per les regions tipus p^+.

3.6. Contactes

Després de la implantació iònica, tota la fotoreïna sobrant és eliminada i l'oblia es recobreix amb SiO_2 mitjançant un procés de dipòsit químic en fase vapor. Aquesta capa protectora, que sol tenir un gruix d'entre 0.25 i 0.5µm, serveix d'aïllament elèctric entre els nivells de polisilici i metall. Tot seguit, s'aplica fotoreïna i té lloc la sisena màscara, que delimita els diferents contactes dels dispositius (drenador, font, porta, substrat, contactes de resistències, etc.). S'ataca la capa protectora per la creació d'aquests contactes i després s'elimina la fotoreïna restant.

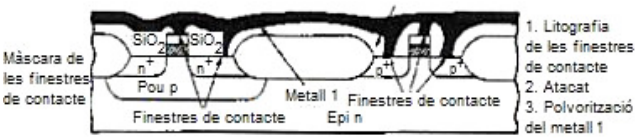


Fig. 11 Formació dels contactes.

3.7. Primer nivell de metall

En aquest procés, en el qual es creen les connexions entre els drenadors dels transistors pMOS i nMOS, així com les de les pistes de V_{DD} i GND, té lloc el dipòsit del metall, el qual presenta un gruix d'uns 0.8µm. Aquest metall (habitualment alumini) és dipositat mitjançant un sistema de *sputtering* i, tot seguit, s'aplica la setena màscara i la corresponent fotoreïna per tal d'eliminar tot el que sobra. A causa de la no planaritat de la superfície de l'oblia, tot el procés d'unió entre el metall i el silici és complex i es poden tenir ruptures, ja sigui durant el procés de fabricació o bé més tard a causa de l'electromigració (acumulació d'una quantitat important d'electrons en una pista de metall fins que provoquen la seva ruptura).

3.8. Vies i segon nivell de metall

Inicialment, es diposita una capa addicional de SiO_2 (o un altre òxid intermetàl·lic) a través d'un dipòsit químic en fase vapor. Tot seguit, es formen els contactes addicionals o vies, que serveixen de connexió entre dos nivells de metall, mitjançant la vuitena màscara i la corresponent fotoreïna. Finalment, es diposita la segona capa de metall i s'ataca amb la novena màscara i una nova fotoreïna.

Mentre que la capa de metall 1 s'utilitza per les interconnexions locals de les portes, la principal utilitat de la capa de metall 2 sol ser la distribució de les fonts d'alimentació. Si es necessiten més nivells de metall, es continuen afegint òxids intermetàl·lics, vies i capes de metall. Les capes més elevades solen tenir un gruix més elevat i es fan servir per les alimentacions.

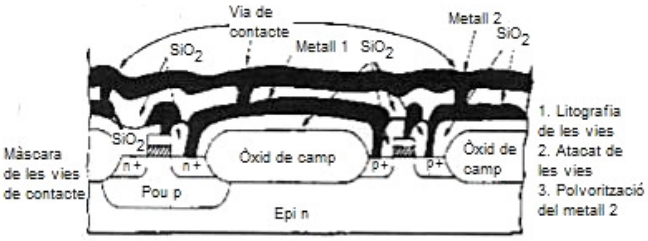


Fig. 12 Dipòsit de metall 1 i 2 i creació de vies.

3.9. Passivació

Després del dipòsit de l'última capa de metall, té lloc un procés de passivació en el qual es diposita una capa de SiO_2 mitjançant dipòsit químic en fase vapor, protegint així la superfície d'agents contaminants. En últim lloc, s'aplica la desena màscara per establir les àrees de soldadura o *pads* que s'utilitzen en el cablejat.

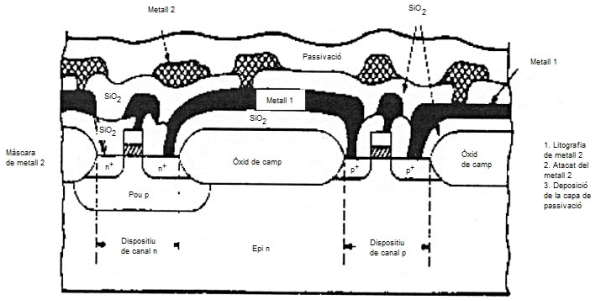


Fig. 13 Passivació. Dibuix final.

4. CONCLUSIONS

La tecnologia planar ofereix avantatges únics enfront d'altres mètodes de fabricació, com són la minimització de les pèrdues de corrent gràcies a la protecció de les unions semiconductores per capes d'òxid i la possibilitat de fabricar simultàniament un gran nombre de xips a la mateixa superfície de silici a causa de la fotolitografia i les tècniques de gravat. D'altra banda, la tecnologia CMOS és actualment la més utilitzada en la fabricació de circuits digitals a causa del seu baix consum de potència, de la bona immunitat enfront del soroll i de la particularitat que els valors de la tensió de sortida no depenen de les dimensions dels transistors.

5. REFERÈNCIES

[Kang99] S. M. Kang, Y. Leblebici. "CMOS Digital Integrated Circuits. Analysis and Design", Mc Graw-Hill, 2a edició, 1999.

[Johns97] D. A. Johns, K. Martin, "Analog Integrated Circuit Design", John Wiley & Sons, 1997.

[Maly87] W. Maly, "Atlas of IC Technologies. An introduction to VLSI Processes", The Benjamin/Cummings Publishing Company, 1987.

[Bota96] S. Bota, J. Carrabina, A. Herms, "Introducció al disseny CMOS VLSI", Textos docents núm 52, Edicions de la Universitat de Barcelona, 1996.

[Alcubilla95] R. Alcubilla, J. Pons, D. Bardés, “Disseny digital. Una perspectiva VLSI-CMOS”, Edicions UPC, 1995.
[Grebene84] A. B. Grebene, “Bipolar and MOS Analog Integrated Circuit Design”, John Wiley & Sons, 1984.
[Uyemura88] J. P. Uyemura “Fundamentals of MOS Digital Integrated Circuits”, Addison-Wesley, 1988.
[Gregorian86] R. Gregorian, G. C. Tems, “Analog MOS Integrated Circuits for signal processing”, John Wiley & Sons, 1986.

2. Disseny físic de circuits integrats.

***Abstract*—**En aquest article s'explicarà què és un *layout* i quin procediment se segueix per realitzar-lo, tant a nivell físic o de màscares com a nivell de disseny. Concretament, pel que fa a aquest últim aspecte, es donarà una metodologia que permet passar del *schematic* a nivell de transistors d'una porta lògica CMOS al *layout* realitzat amb *Cadence*, que ja pot ser enviat a la *foundry* per la seva fabricació. Finalment, es parlarà a nivell conceptual de les capacitats paràsites existents en un transistor MOS i es veurà com aquestes poden simplificar-se només pel sol fet de tenir en compte el *layout*.

Index Terms*—Layout*, màscara, pou, àrea activa, polisilici, difusió, metall, via, gràfic d'Euler, camí d'Euler, diagrama de Sticks, capacitat paràsita.

1. INTRODUCCIÓ

Realitzar el *layout* d'una porta significa representar la posició i mida (però no el gruix) de les diferents capes de materials que després es dipositaran a l'oblia de silici per fabricar el circuit integrat, o el que és el mateix, definir la forma de les principals màscares que intervenen en el procés de fabricació. Aquestes màscares bàsiques es dibuixen amb un programa de CAD.

Cal parar atenció en què el disseny respecti les regles de disseny especificades pel fabricant. De no ser així, podria ser que el que s'implementés realment fos diferent del que apareix perfectament definit a la pantalla de l'ordinador. L'eina DRC (*Design Rules Checker*), que estarà inclosa al programari que s'usarà, permet comprovar si se satisfan totes les regles exigides per la *foundry*. En general, les principals màscares que s'han de dibuixar són les següents:

• Màscara de pou. Al treballar amb transistors pMOS i nMOS alhora, s'ha de definir una zona amb dopatge oposat al del substrat (pou). Així, si es treballa sobre una oblia tipus p, s'ha de definir un pou tipus n per construir en aquest lloc el bloc p de la porta.

• Màscara d'àrea activa. Defineix la zona de l'oblia en la qual se situaran els transistors. La quantitat de dopatge usat servirà per definir algunes de les principals característiques dels dispositius, com per exemple la tensió *threshold*.

• Màscara de polisilici. El polisilici s'usa per definir les portes, i per tant els canals dels transistors. A causa de la seva conductivitat, també poden ser usats com a elements de connexió entre transistors. De totes maneres, en aquest últim cas s'ha de vigilar per on passen les línies de *poly* ja que cal tenir en compte que allà on un *poly* creua la zona activa, es té efecte transistor.

• Màscares de difusió. Primer es realitza la màscara per n^+ i després, per p^+. Serveixen per definir les regions de *source* i *drain* dels transistors n i p, respectivament. Convé tenir en compte que no és necessari que tinguin la forma exacta d'aquestes regions, ja que aquesta part del procés tecnològic va després de la creació dels *polys*, que s'encarreguen de protegir les regions de canal dels transistors.

• Màscara de contacte. Els contactes serveixen per unir el metall amb les diferents capes que es trobin per sota d'aquest.

• Màscara de metall 1. Aquesta màscara defineix les línies de metall que contacten amb els *polys* i les difusions. També es pot utilitzar per realitzar connexions amb metalls superiors.

• Màscara de metall 2...n. Es fa servir per fer connexions.

• Màscara de via. Les vies s'utilitzen per unir diferents capes de metall entre si.

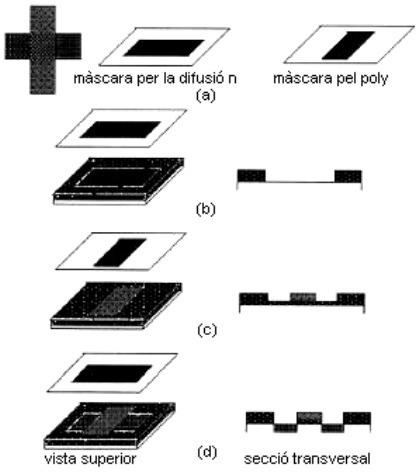
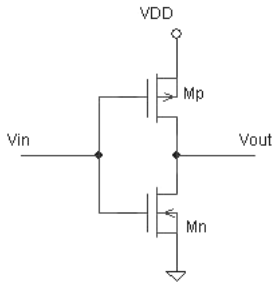


Fig. 1 Construcció d'un transistor nMOS.

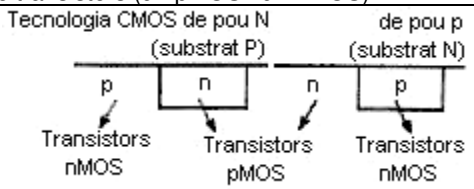
Com es pot observar a la Fig. 1, per construir un transistor nMOS, primer es tenen dues màscares (difusió n i polisilici). Després, la màscara de la difusió n s'usarà per crear la capa d'òxid de silici sobre l'oblia, cobrint les regions on es disposaran els transistors. La màscara de polisilici, en canvi, s'utilitzarà per col·locar la capa d'aquest material

sobre la capa d'òxid, per definir així les portes del transistor. Finalment, la regió de difusió n s'utilitzarà també per definir les fonts i els terminals de sortida del transistor.

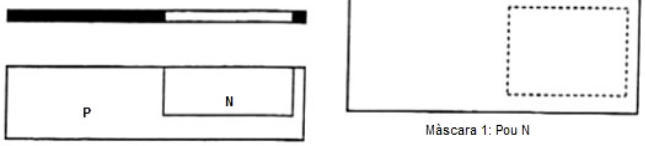
2. *LAYOUT* D'UN INVERSOR



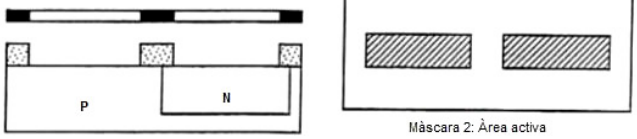
Per la fabricació d'un inversor, primer s'obté l'esquemàtic a nivell transistor. En aquest cas, només es necessiten un parell de transistors (un pMOS i un nMOS).



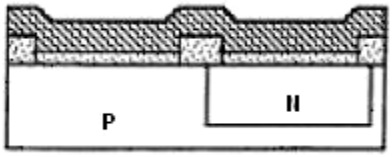
Per la creació dels transistors es pot utilitzar un pou de tipus n o tipus p (aquí es prendrà un pou tipus n).



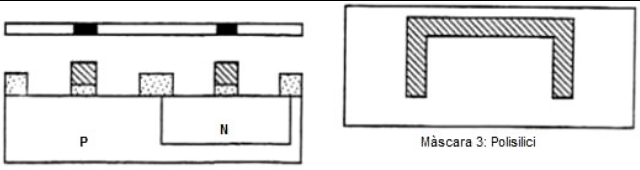
Creació del pou n, mitjançant implantació o difusió.



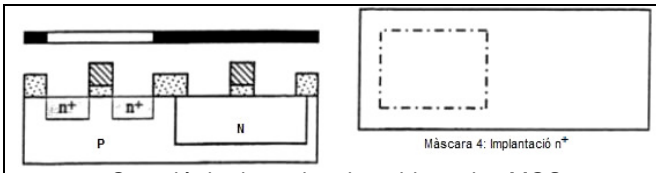
Creació de l'òxid de camp i delimitació de la zona activa.



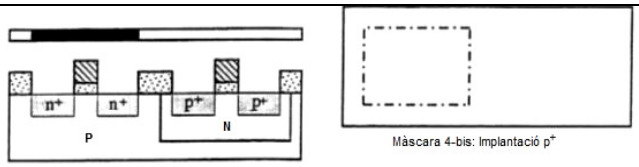
Creixement de l'òxid de la porta i disposició del polisilici.



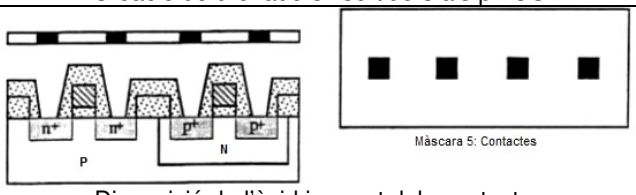
Gravat del polisilici i de l'òxid de la porta.



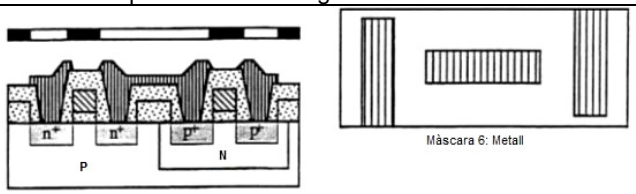
Creació de drenadors i sortidors als nMOS.



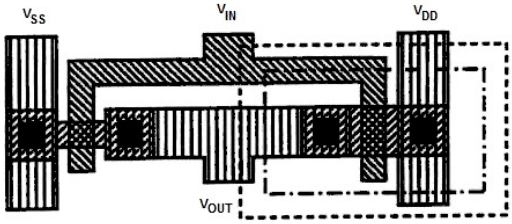
Creació de drenadors i sortidors als pMOS.



Disposició de l'òxid i gravat dels contactes.



Disposició de metall i gravat de les pistes. Ara només falten la segona capa de metal·lització i les vies.



Aquí es té la disposició del layout de l'inversor amb totes les seves capes representades.

3. GRÀFIC D'EULER I DIAGRAMA DE STICKS

El gràfic d'Euler és un mètode senzill per trobar una disposició òptima dels transistors en el cas de portes més complexes que un inversor. Permetrà dibuixar un *layout* clar i ordenat. Per traçar un gràfic d'Euler només cal tenir en compte unes quantes normes, que es detallen tot seguit.

En primer lloc, s'ha de tenir present que els nodes del circuit corresponen als vèrtexs del gràfic d'Euler. Hi ha un vèrtex corresponent al node de V_{DD} i un altre al de GND. Els altres vèrtexs corresponen als terminals de *source* i *drain* de cada transistor.

Els vèrtexs s'uneixen amb línies, que corresponen als transistors. Cada línia s'identifica amb el nom de l'entrada

del transistor i cada vèrtex, amb la tensió del node corresponent. En primer lloc es realitza un graf per la xarxa n i un altre per la xarxa p, tot seguint aquesta nomenclatura. Finalment, s'identifiquen els camins d'Euler. Són aquells que permeten recórrer cada graf corresponent a un bloc sense aixecar el llapis del paper i passant per tots els transistors (és a dir, passant una sola vegada per cada línia). Cada camí s'etiqueta amb els noms de les línies que recorre successivament. Es tracta de trobar un camí a la zona p i un altre a la zona n que segueixin la mateixa successió. Com que cada línia s'associa a una entrada, es col·locaran els transistors en el mateix ordre tant a la zona p com a la n. A més, els transistors connectats directament compartiran *source* i *drain*.

Així, per establir les bases del *layout* d'una porta OR3, per exemple, caldria en primer lloc partir del disseny de la porta a nivell de transistors i traçar el gràfic d'Euler. En aquest cas, existeixen dos camins d'Euler vàlids, serien ABC i CBA.

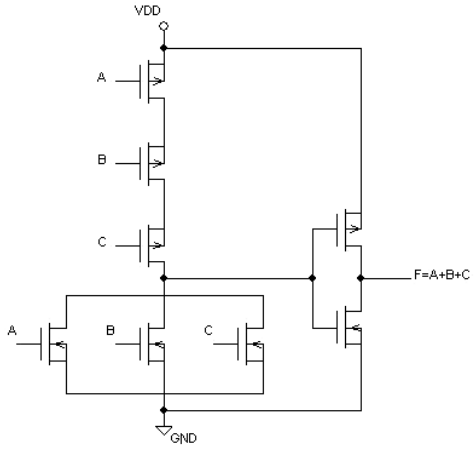


Fig. 2 Esquema a nivell de transistors per una OR3.

Cal parar compte que aquesta porta és no negada; com que la lògica CMOS és una lògica inversora, es tindrà un inversor connectat a la sortida de la porta NOR3. Aquest fet no afecta a nivell de gràfic d'Euler, però sí a nivell de *layout*. Dels dos camins d'Euler possibles, n'hi ha un que permetrà realitzar el *layout* d'una manera més còmoda. Aquest és el CBA, ja que es pot connectar l'inversor a continuació de la porta NOR3, obtenint així la porta OR3, sense haver de partir la difusió (veure Fig. 4).

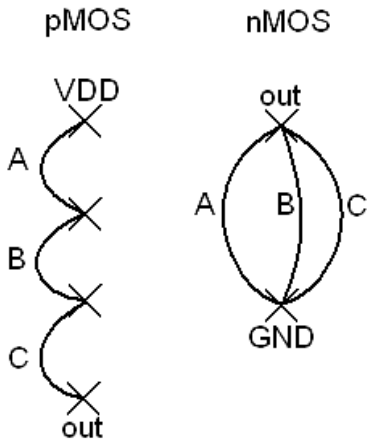


Fig. 3 Gràfic d'Euler d'una OR3.

Tot seguit, ja es pot dibuixar esquemàticament el *layout* seguint l'ordenació establerta pel camí d'Euler trobat. La millor manera és mitjançant l'anomenat diagrama de Sticks.

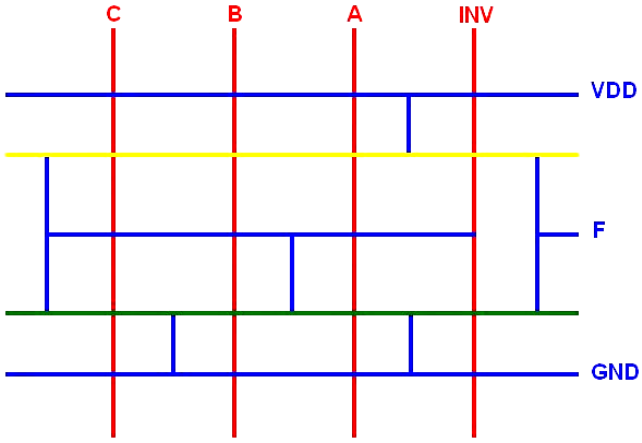


Fig. 4 Diagrama de Sticks d'una OR3.

En aquest diagrama, es representen cadascuna de les entrades de la porta mitjançant una línia vertical de polisilici (vermell). Alimentació (V_{DD}) i massa (GND) se simbolitzen a través de dues línies horitzontals (blau), de la mateixa manera que la difusió p (groc) i la difusió n (verd). La resta de línies que es veuen al diagrama (blau) són les interconnexions que cal realitzar per tal d'unir els diferents transistors que formen la porta. Per saber com s'han de dibuixar aquestes línies, només cal fixar-se en l'esquema a nivell de transistors (veure Fig. 2) i passar per tots els transistors seguint el camí d'Euler escollit. S'han de marcar les unions dels transistors amb V_{DD}, GND i la sortida. També s'han de connectar els transistors, els *polys* dels quals no apareixen de costat al diagrama de Sticks, però que a l'esquema a nivell de transistors sí que es troben units. Es connecten pel node que tinguin en comú, ja sigui

source o *drain*.

Aquest diagrama agrupa tota la informació relativa al disseny exceptuant les dimensions. Per tal de calcular les mides dels transistors, es poden seguir diferents criteris. Un d'ells consisteix en considerar quina és la xarxa equivalent en el pitjor cas i fer que els temps de pujada i de baixada de la porta siguin iguals.

4. REALITZACIÓ DEL *LAYOUT*

Una vegada s'ha completat el diagrama de Sticks, ja es pot procedir a realitzar el disseny físic o *layout* de la porta mitjançant un programa de disseny com el *Cadence*. Aquí es mostra com a exemple el *layout* de la porta OR3, realitzat amb una tecnologia de 0.35µm.

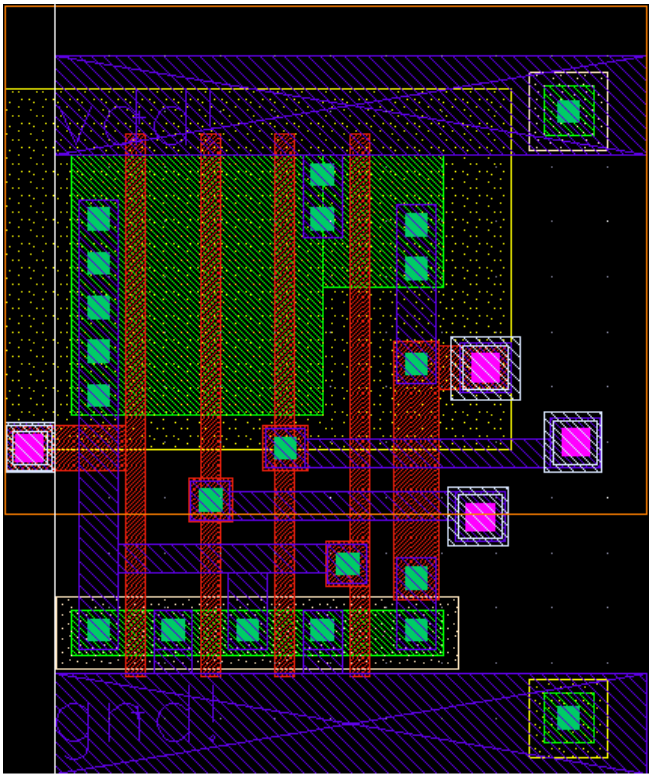


Fig. 5 *Layout* d'una porta OR3.

5. CAPACITATS EN REFERÈNCIA AL *LAYOUT*

En qualsevol porta lògica CMOS es tenen capacitats paràsites. Aquestes són les degudes a les unions pn de les regions de *source* i *drain* dels transistors, les de porta i la de sortida (veure Fig. 6). Donat que totes aquestes capacitats estan referenciades a massa, poden considerar-se en paral·lel.

En aquest apartat, es consideraran les capacitats paràsites que apareixen a *source* i *drain* de la porta OR3. Observant el *layout* de la porta, pot comprovar-se com el *source* del transistor C i el *drain* del transistor B, per exemple, comparteixen difusió. Per tant, les capacitats associades a aquestes regions presenten el mateix node i, a efectes pràctics, poden considerar-se com una sola capacitat. Ocorre el mateix amb el *source* del transistor B i el *drain* del transistor A, és a dir, com a conseqüència que aquestes regions comparteixen node al *layout*, les seves capacitats associades poden considerar-se com una de sola (veure Fig. 7).

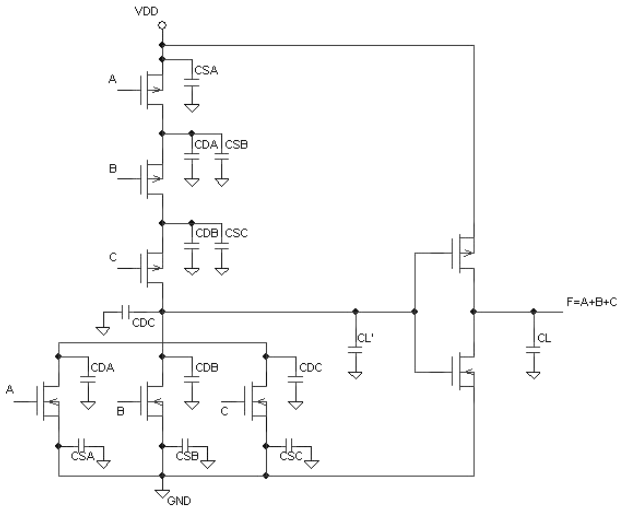


Fig. 6 Esquema a nivell de transistors per una OR3 amb totes les capacitats paràsites existents.

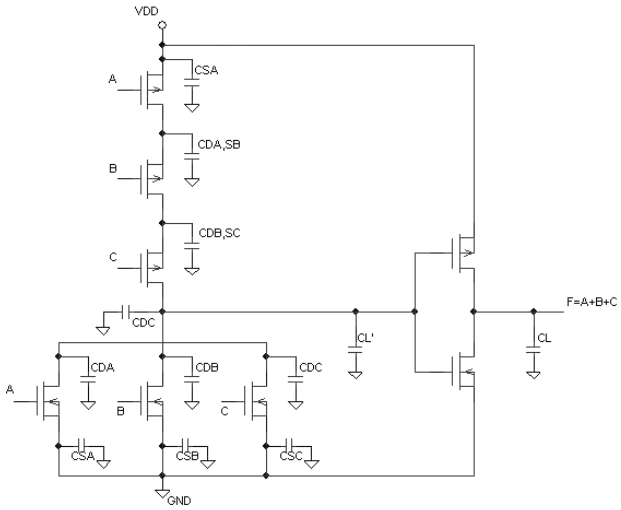


Fig. 7 Esquema a nivell de transistors per una OR3 on s'han unit les capacitats paràsites que comparteixen node.

6. CONCLUSIONS

Un *layout* és una representació exacta de la posició i mida dels materials que han de ser dipositats sobre una oblia de silici per la fabricació d'un circuit integrat. Per realitzar aquesta representació (mitjançant un programa de disseny com el *Cadence*), es parteix de l'esquema a nivell de transistors i s'obté l'anomenat gràfic d'Euler, que permet trobar els camins d'Euler. Aquests donen les disposicions vàlides dels diferents transistors que integren la porta, que hauran de ser tingudes en compte quan s'extregui el diagrama de Sticks, el qual és ja una representació esquemàtica del que posteriorment serà el *layout*.

Hi ha camins d'Euler més òptims que d'altres en el sentit que faciliten el diagrama de Sticks i, per tant, el disseny del *layout*. Serà l'experiència del dissenyador la que permetrà escollir a primera vista un camí d'Euler menys costós.

Finalment, cal esmentar que tenint en compte el *layout*, quan se segueix la metodología del diagrama de Sticks seguint camins d'Euler, es pot veure que el nombre de capacitats paràsites de difusió respecte l'*schematic* es redueix degut a que es comparteixen difusions.

7. REFERÈNCIES

[Uyemura88] J. P. Uyemura, "Principles of CMOS integrated circuits", Addison-Wesley, 1988.

[Kang02] S. M. Kang, Y. Leblebici, "CMOS Digital Integrated Circuits: Analysis and Design", McGraw Hill, 2002.

3. El transistor MOS. Estructura i operació.

***Abstract*—**En aquest article es presentarà el transistor d'efecte de camp MOS, també anomenat MOSFET (*Metal Oxide Semiconductor Field Effect Transistor*). Es veurà que és un dispositiu format per quatre terminals anomenats drenador o *drain* (D), porta o *gate* (G), font o *source* (S) i substrat o *bulk* (B), així com també que entre drenador i font circula un corrent que és controlat per la tensió aplicada a la porta. Es presentaran les expressions per aquest corrent per cadascuna de les regions de treball del transistor.

***Index Terms*—**Estructura MOS, transistor d'enriquiment, transistor d'empobriment, regió òhmica, regió de saturació, inversió de canal, inversió feble, inversió forta.

1. INTRODUCCIÓ

En aquestes línies es presenta un tipus de transistor, el funcionament del qual està basat en el transport de càrrega associat a un únic tipus de portadors (e^- o h^+). A causa d'això, també és conegut amb el nom de transistor unipolar, a diferència dels transistors bipolars (BJT), en els quals el transport de càrrega es realitza mitjançant ambdós tipus de portadors injectats a través de les unions pn.

Des del punt de vista físic, el principi de funcionament se centra en l'acció d'un camp elèctric sobre càrregues elèctriques, provocant el seu desplaçament i, en conseqüència, el corrent elèctric. D'aquí el seu nom genèric de FET (*Field Effect Transistor*).

S'han desenvolupat diverses estructures de transistors FET, segons la tecnologia i/o necessitats. Les més importants són les implementades amb tecnologies sobre silici com el JFET, o *Junction* FET, i el MOSFET, o *Metal Oxide Semiconductor* FET. En tecnologies d'arseniür de gal·li (GaAs) s'han implementat transistors MESFET o *Metal Semiconductor* FET. Finalment, en altres tecnologies s'han desenvolupat els coneguts HEMFET o *High Electron Mobility* FET. La majoria dels CI actuals, però, es realitzen sobre tecnologies de Si, on el transistor MOSFET és àmpliament el més utilitzat sobre els altres per posseir certes característiques que el fan avantatjós, fins i tot per sobre del transistor bipolar. Algunes d'aquestes característiques, que han impulsat el desenvolupament i ús dels transistors MOSFET, són les que es descriuen a continuació.

- El procés de fabricació és simple (menor nombre de passos).
- Mida reduïda, que condueix a densitats d'integració elevades.
- Es pot evitar l'ús de resistències, degut al fet que aquestes poden ser substituïdes per un MOSFET.
- Consum d'energia reduït (menor consum de potència).
- Poden implementar-se tant funcions analògiques com digitals i/o mixtes dins d'un mateix xip.

2. ESTRUCTURA *METAL OXIDE SEMICONDUCTOR*

El transistor MOSFET està basat en l'estructura MOS, que s'analitzarà a continuació. L'estructura MOS es compon de dos terminals i tres capes: un substrat de silici, pur o poc dopat p o n, sobre el qual es genera una capa d'òxid de silici (SiO_2) que posseeix característiques dielèctriques o aïllants. Finalment, sobre aquesta es col·loca una capa de metall (actualment polisilici), que posseeix característiques conductores. A la part inferior es col·loca un contacte òhmic, com es mostra a la Fig. 1.

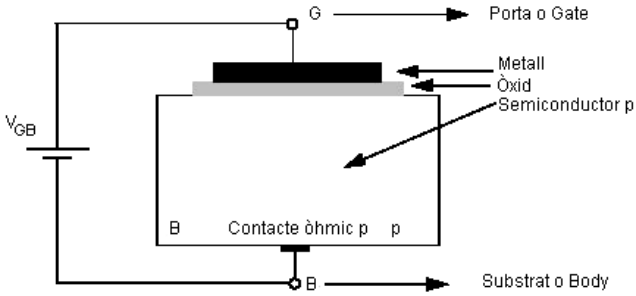


Fig. 1 Estructura MOS.

L'estructura MOS actua com un condensador de plaques paral·leles en el qual G i B són les plaques i l'òxid, l'aïllant. D'aquesta manera, quan $V_{GB}=0$, la càrrega acumulada és zero i la distribució de portadors, aleatòria. Correspon a l'estat d'equilibri al semiconductor. Si $V_{GB}>0$, apareix un camp elèctric entre els terminals de porta i substrat. La regió semiconductora p es comporta creant una regió d'empobriment de càrregues lliures p^+ (zona de deplexió), tal com passa a la regió p d'una unió pn quan està polaritzada negativament. Aquesta regió de ions negatius s'incrementa amb V_{GB}.

En arribar a una cota per V_{GB}, els ions presents a la zona semiconductora d'empobriment no poden compensar el camp elèctric i es provoca l'acumulació de càrregues negatives lliures (e^-) atretes pel terminal positiu. Es diu

llavors que l'estructura ha passat a l'estat d'inversió. El procés d'inversió s'identifica amb el canvi de polaritat del substrat sota la regió de porta. En inversió forta, es forma així un canal d'e^- lliures a les proximitats del terminal de *gate*.

La intensitat de porta, I_G, és zero, ja que en contínua es comporta com un condensador. Per tant, es pot dir que la impedància des de la porta al substrat és pràcticament infinita. Bàsicament, l'estructura MOS permet crear una densitat de portadors lliures suficient per mantenir un corrent elèctric.

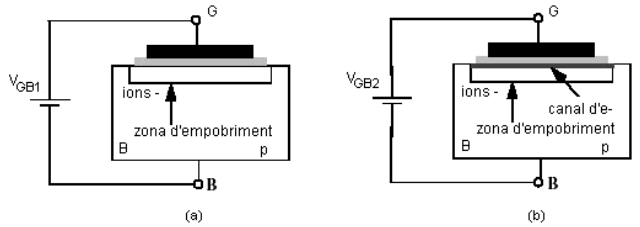


Fig. 2 Estructura MOS en inversió feble (a) i inversió forta (b).

3. PRINCIPIS BÀSICS DEL TRANSISTOR MOSFET

A continuació es distingeixen diverses estructures MOS similars, de les quals s'analitzarà en més detall l'anomenada MOSFET d'enriquiment de canal n, encara que el funcionament de totes elles és similar i es basen en el mateix principi d'operació.

3.1. MOSFET d'enriquiment de canal n

Es tracta d'una estructura MOS de quatre terminals, en la qual el substrat semiconductor és de tipus p poc dopat. A banda i banda de la interfície òxid semiconductor s'han practicat difusions de material n fortament dopat (n^+). Els quatre terminals de l'estructura de la Fig. 3 són:

- G, porta o *gate*
- B, substrat o *bulk*
- D, drenador o *drain*
- S, font o *source*

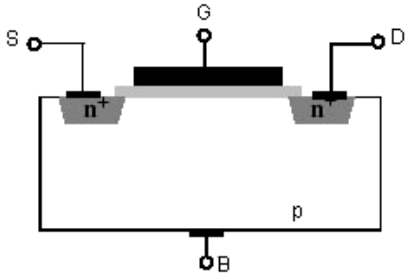


Fig. 3 Estructura MOSFET de canal n.

Els símbols més utilitzats per la seva representació a nivell de circuit es mostren a la Fig. 4. El terminal B sol estar col·locat a la tensió més negativa (referència o GND), pel que s'omet en alguns símbols. D'aquesta manera es garanteix que els díodes d'unió paràsits entre substrat, drenador i font, respectivament, sempre estan polaritzats negativament. La fletxa al terminal de font informa sobre el sentit del corrent. S'observa que I_G=0 i I_D=I_S. A continuació, es descriu el principi d'operació d'aquesta estructura de transistor.

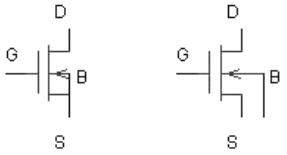


Fig. 4 Símbols pel transistor MOSFET de canal n.

Es considera l'estructura MOS de la Fig. 5. En ella apareixen diverses fonts de tensió polaritzant els diversos terminals: V_{GS} i V_{DS}. Els terminals de substrat i font s'han connectat a GND. D'aquesta manera, V_{BS}=0 i es diu que no existeix efecte substrat. Totes les tensions estan referenciades a la font (S).

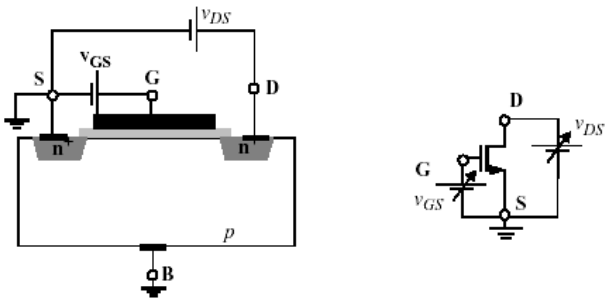


Fig. 5 Estructura MOS polaritzada.

Es consideren ara tres casos, segons els valors que prengui la tensió V_{GS}.

• V_{GS}=0. Aquesta condició implica que V_{GB}=0, ja que V_{BS}=0. En aquestes condicions, no existeix efecte de camp i no es crea el canal d'e^- sota la porta. Les dues estructures pn es troben tallades (B al terminal més negatiu) i aïllades. Aproximadament, I_{DS}=0, doncs s'alimenta de les intensitats inverses de saturació i el corrents subllindar.

$$V_{GS} = 0 \rightarrow I_{DS} = 0 \quad (1)$$

• V_{Tn}>V_{GS}>0, on V_{Tn} és la tensió llindar necessària per crear la inversió del canal. La tensió V_{GS} crea la zona d'empobriment o deplexió al canal. Es genera càrrega elèctrica negativa al canal a causa dels ions negatius de la xarxa cristal·lina (similar a la d'una unió pn polaritzada en inversa), donant lloc a la situació d'inversió feble anteriorment citada. L'aplicació d'un camp elèctric lateral (V_{DS}>0) no permet generar corrent elèctric I_{DS}.

• V_{GS}>V_{Tn}>0. La tensió V_{GS} dóna lloc a la inversió de canal i genera una població d'e^- lliures sota l'òxid de porta.

Es forma el canal n, o canal d'electrons, entre el drenador i la font (tipus n^+) que modifica les característiques elèctriques originals del substrat. Aquests electrons són càrregues lliures, de manera que en presència d'un camp elèctric lateral podrien veure's accelerats cap a D o S. No obstant això, existeix un valor mínim de V_{GS} perquè el nombre d'electrons pugui ser suficient per alimentar aquest corrent. Aquest valor mínim és V_{Tn}.

Per tant, es poden diferenciar dues zones d'operació per valors de V_{GS} positius: si $V_{GS}<V_{Tn}$, la intensitat $I_{DS}=0$ (en realitat només és aproximadament zero) i es diu que el transistor opera en inversió feble. En aquesta situació, els corrents són molt petits i la seva utilització s'emmarca en contextos de molt baix consum de potència. Es considerarà que el corrent és sempre zero. D'altra banda, si $V_{GS}>V_{Tn}$, llavors I_{DS} serà diferent de zero, sempre que V_{DS} sigui no nul. Es diu que el transistor opera en inversió forta.

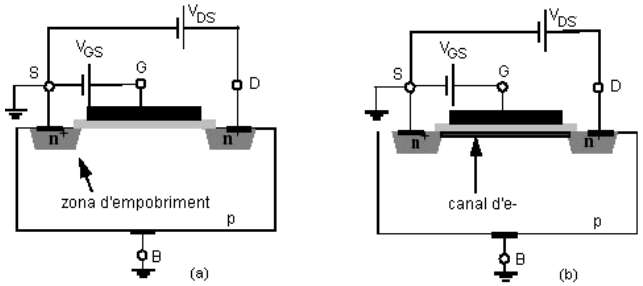


Fig. 6 Polarització del canal en inversió feble (a) i inversió forta (b).

Com més gran sigui el valor de V_{GS}, major serà la concentració de càrregues lliures al canal i per tant, el corrent I_{DS} serà superior. En ser la intensitat I_{DS} proporcional a V_{GS} i V_{DS}, es pot estudiar la relació paramètrica (I_{DS}, V_{DS}) amb V_{GS} com a paràmetre. S'obté la corba de la Fig. 7. En ella s'aprecia com a partir d'un valor donat de la tensió V_{DS}, la intensitat I_{DS} roman constant. Aquest efecte es pot explicar des del punt de vista de concentració d'e^- disponibles al canal.

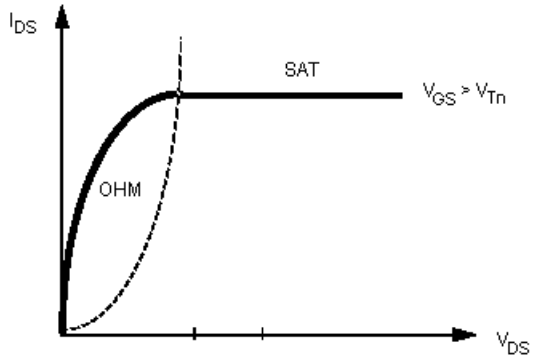


Fig. 7 Característica I-V del transistor MOS de canal n.

La Fig. 8(a) il·lustra la situació que ocorre quan s'aplica una tensió V_{DS} petita a un transistor en inversió forta. En estar més positivament polaritzada la regió del drenador respecte del substrat, la concentració d'e^- es fa major a les proximitats de la font. Per tensions V_{DS} grans, la regió de deplexió al voltant del D (si V_{DB} creix, la regió de deplexió també creix al voltant de DS; D i B formen una unió pn en inversa) fa que s'estrenyi el canal i el corrent de drenador pràcticament s'independitza de V_{DS}.

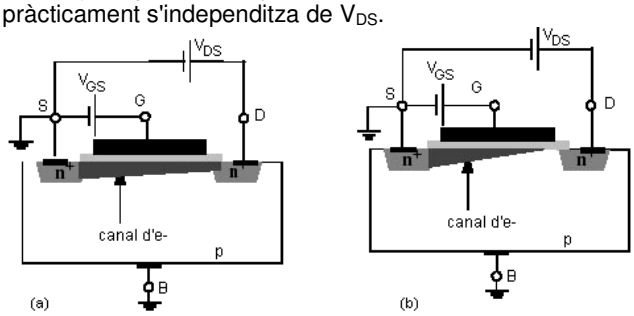


Fig. 8 Polarització en inversió forta per la regió òhmica (a) i per la regió de saturació (b).

I_{DS} s'incrementa amb V_{DS} des de petits valors. A partir d'un cert valor, aquest comportament canvia, i I_{DS} es fa constant. A la primera regió se l'anomena regió òhmica o lineal, mentre que a la segona se l'anomena regió de saturació. Per diferents valors de V_{GS}, es pot obtenir la família de corbes mostrada a la Fig. 9.

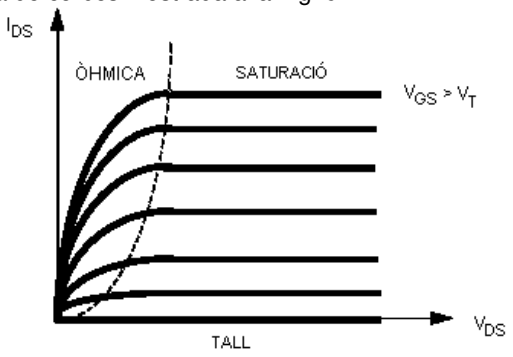


Fig. 9 Família de corbes del transistor MOS de canal n.

Es pot resumir l'exposat de la manera següent:

- Regió de tall

$$V_{GS} = 0 \rightarrow I_{DS} = 0 \quad (2)$$

- Regió de conducció

$$V_{GS} \geq V_{Tn} \rightarrow I_{DS} \neq 0 \text{ si } V_{DS} > 0 \quad (3)$$

Dins de la regió de conducció, es poden identificar dues possibles situacions pel MOS:

• Regió lineal. I_{DS} augmenta amb V_{DS}, és a dir, el MOSFET es comporta com un resistor (no lineal).

• Regió de saturació. I_{DS} és aproximadament constant amb V_{DS}. Es comporta com una font d'intensitat controlada

per tensió (V_{GS}=constant).

El límit entre la regió òhmica i la de saturació es troba a:

$$V_{GS} - V_{Tn} = V_{DS} \quad (4)$$

de manera que per a $V_{DS}<V_{GS}-V_{Tn}$ el transistor es troba a la regió òhmica i per $V_{DS}\geq V_{GS}-V_{Tn}$, a la regió de saturació.

Les expressions corresponents pel corrent de drenador, obtingudes mitjançant l'anàlisi del transport de càrrega des de la font cap al drenador per les diferents regions d'operació, es mostren a continuació (veure Uyemura per demostració).

$$I_{DS} = \frac{k_n}{2}\frac{W}{L}\left(2(V_{GS}-V_{Tn})V_{DS} - V_{DS}^2\right) \text{ òhmica} \quad (5)$$

$$I_{DS} = \frac{k_n}{2}\frac{W}{L}(V_{GS}-V_{Tn})^2 \text{ saturació} \quad (6)$$

on $k_n=\mu_n \cdot C_{ox}$ és la transconductància, μ_n la mobilitat dels e^- i C_{ox} la capacitat per unitat d'àrea de l'estructura MOS. Se sol definir també $\beta=K_n(W/L)$, que és un paràmetre que depèn tant de la geometria com dels paràmetres elèctrics de la tecnologia. Valors típics per k_n són de 20 a 70μA/V^2, i per k_p de 8 a 30μA/V^2. W i L defineixen l'àrea del canal i poden ser utilitzats pel dissenyador per ajustar les característiques del circuit a unes especificacions donades.

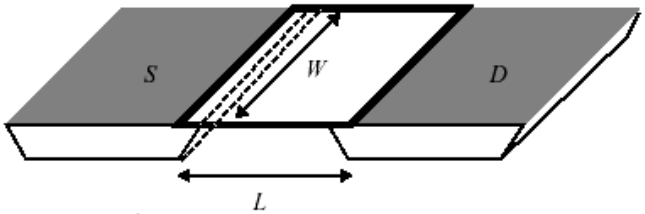


Fig. 10 Àrea del canal d'un transistor MOS de canal n.

Tot seguit, es discutiran les equacions que donen compte dels diferents fenòmens que tenen lloc en un transistor MOS de canal n.

3.1.1. Efectes resistius

L'expressió de la intensitat a la regió òhmica, si $V_{DS}<<1$, es pot aproximar a:

$$I_{DS} = k_n \frac{W}{L}(V_{GS}-V_{Tn})V_{DS} \quad (7)$$

que representa l'equació constitutiva d'una resistència de valor:

$$R^{-1} = k_n \frac{W}{L}(V_{GS}-V_{Tn}) \quad (8)$$

A partir d'aquesta expressió i mantenint fix el paràmetre tecnològic k_n, s'observa com disminueix la resistència, i per tant augmenta la intensitat, si s'augmenta l'amplada del canal W o es disminueix la longitud L.

En el límit de la regió òhmica, fent $V_{DS}=V_{GS}-V_{Tn}$, s'obté l'expressió per I_{DS} en saturació, que depèn quadràticament de V_{GS}.

3.1.2. Modulació del canal

Quan el transistor entra al règim de saturació, el canal desapareix a prop del contacte de drenador. Si la tensió aplicada V_{DS} augmenta, la regió estrangulada avança cap a la font. Aquest fenomen és conegut com a modulació de la longitud del canal. S'introdueix a l'expressió pel corrent de drenador mitjançant l'addició d'un terme multiplicatiu ($1+\lambda \cdot V_{DS}$), on λ és el paràmetre de modulació del canal.

$$I_{DS} = \frac{k_n}{2}\frac{W}{L}(V_{GS}-V_{Tn})^2(1+\lambda V_{DS}) \text{ saturació} \quad (9)$$

En general, el paràmetre λ modifica poc l'expressió de I_{DS} a la regió lineal. A la regió de saturació, és el responsable que el corrent no sigui constant i determina la resistència de sortida del transistor.

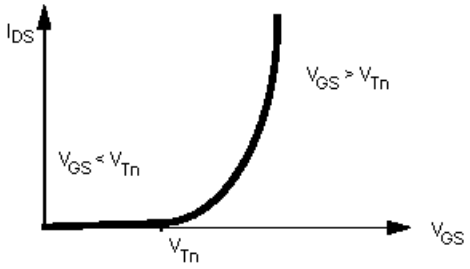


Fig. 11 Dependència de I_{DS} amb V_{GS} en saturació.

3.1.3. Efecte substrat

Quan V_{BS} no és zero, l'expressió de la tensió llindar queda modificada segons:

$$V_T = V_{Tn} + \gamma\left(\sqrt{2\phi_f + V_{BS}} - \sqrt{2\phi_f}\right) \quad (10)$$

on V_{Tn} és la tensió llindar nominal per $V_{BS}=0$, que sol oscil·lar entre 0.5 i 1.5V, i γ és el paràmetre d'efecte substrat (de l'ordre de 0.5V$^{1/2}$). Φ_f és una constant física que depèn del dopatge. Té per expressió:

$$\phi_f = \frac{KT}{q}.\ln\frac{N_D}{n_i} \quad (11)$$

3.2. MOSFET d'enriquiment de canal p

Respon a una estructura dual de la del MOS de canal n. S'intercanvien les regions dopades n per regions dopades p i viceversa. En aquest cas, el canal es forma gràcies a

l'existència de càrregues positives lliures (forats o p^+), però el funcionament és similar. És necessari col·locar el substrat a la tensió més positiva, formant-se el canal per valors de V_{GB} (V_{GS}) negatius, atraient les càrregues p^+. El corrent de drenador-font, I_{DS}, s'origina si $|V_{GS}|<|V_{Tp}|$.

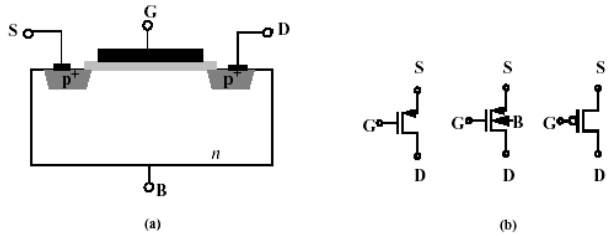


Fig. 12 Estructura MOSFET de canal p (a) i símbols (b).

Les corbes I-V característiques que s'obtenen es mostren a la Fig. 13.

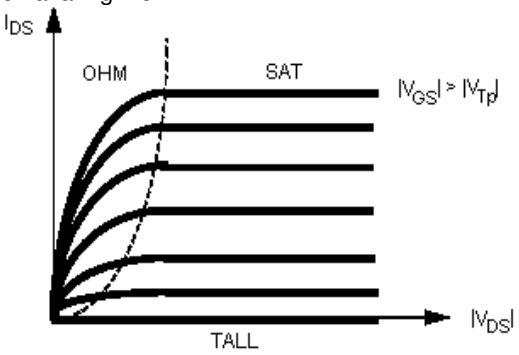


Fig. 13 Característica I-V per un transistor pMOS.

3.3. MOSFET d'empobriment o deplexió de canal n

L'estructura d'aquest MOS és similar a la d'enriquiment. No obstant això, durant el procés de fabricació s'ha d'afegir una implantació n^+ a la regió del canal (definida per W i L). Aquesta modificació permet incrementar el nombre de càrregues negatives al canal. D'aquesta manera pot existir corrent entre el drenador i la font per valors de V_{GS} nuls i inclús negatius (equivalents a l'existència de tensions llindars negatives).

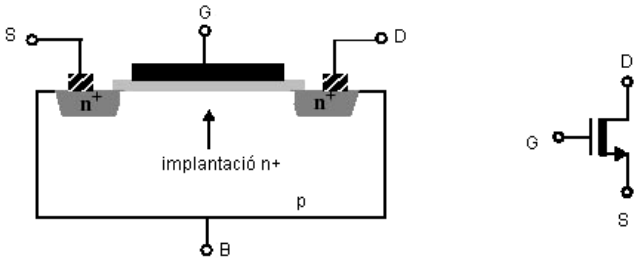


Fig. 14 Estructura MOS de canal n d'empobriment.

Les característiques I_{DS} i V_{DS} són molt semblants a les dels transistors d'enriquiment, distingint-se tres regions:

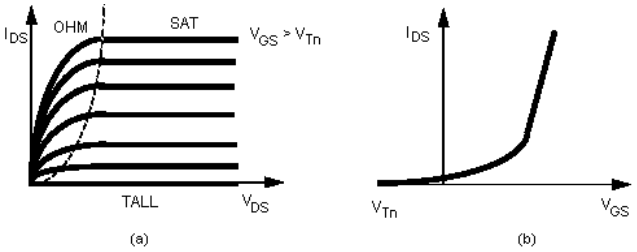


Fig. 15 Corbes I_{DS} vs V_{DS} (a) i I_{DS} vs V_{GS} (b).

- Tall

$$V_{GS} < V_{Tn} < 0 \quad (12)$$

$$I_{DS} = 0 \quad (13)$$

- Conducció

$$V_{GS} > V_{Tn} \quad (14)$$

$$I_{DS} > 0 \text{ si } V_{DS} > 0 \quad (15)$$

En aquest darrer cas es poden distingir dues regions d'operació. Per $V_{DS}<V_{GS}-V_{Tn}$ el transistor opera a la regió òhmica, on la intensitat respon a l'equació:

$$I_{DS} = \frac{I_{DSS}}{V_{Tn}^2}\left(2(V_{GS} - V_{Tn})V_{DS} - V_{DS}^2\right) \quad (16)$$

En aquesta, la funcionalitat de $I_{DS}=I_{DS}(V_{GS}, V_{DS}, V_{Tn})$ és la mateixa que pel transistor MOS d'empobriment. Es pot prendre $\beta/2=I_{DSS}/V_{Tn}$. Així mateix, si $V_{DS}<<1$ es pot aproximar per:

$$I_{DS} = \frac{2I_{DSS}}{V_{Tn}^2}(V_{GS} - V_{Tn})V_{DS} \quad \text{òhmica} \quad (17)$$

Per $V_{DS}\geq V_{GS}-V_{Tn}$, el transistor es troba a la regió de saturació. L'equació per la intensitat drenador-font és:

$$I_{DS} = \frac{I_{DSS}}{V_{Tn}^2}(V_{GS} - V_{Tn})^2 \quad \text{saturació} \quad (18)$$

on la intensitat és constant. D'aquesta equació es destaca que:

- Per $V_{GS}=0$, es troba $I_{DS}=I_{DSS}$ de manera que I_{DSS} representa la intensitat drenador-font per $V_{GS}=0$ en saturació.
- Per $V_{GS}-V_{Tn}=V_{DS}$, s'està a la frontera entre les regions òhmica i de saturació. Si $V_{GS}=0$ es verifica que $V_{DS}=-V_{Tn}$, és a dir, a la interfície entre regions la tensió V_{DS} és igual a la tensió llindar del transistor.

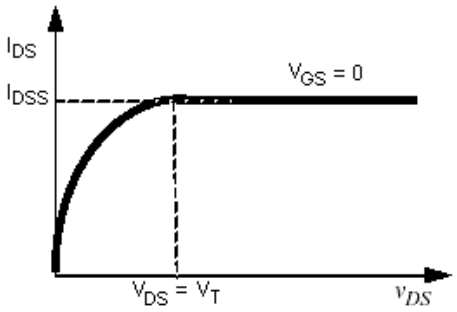


Fig. 16 Corba característica I-V.

3.4. MOSFET d'empobriment de canal p

És similar al MOS de canal n d'empobriment, però complementari respecte la funcionalitat de les regions n i p, així com també en el signe de les tensions i sentit de les intensitats.

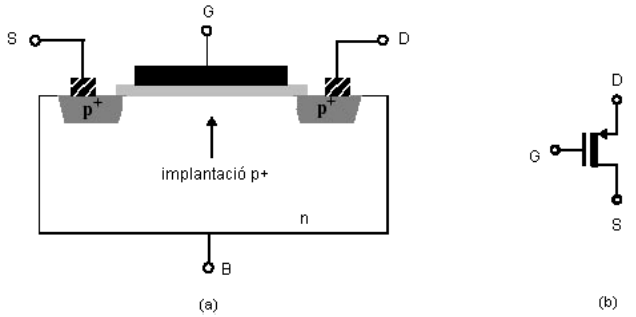


Fig. 17 Transistor MOS d'empobriment de canal p (a) i símbol (b).

- Tall.

$$|V_{GS}| < |V_{Tp}| \quad (19)$$

$$I_{DS} = 0 \quad (20)$$

- Conducció.

$$|V_{GS}| > |V_{Tp}| \quad (21)$$

$$I_{DS} > 0 \text{ si } |V_{DS}| > 0 \quad (22)$$

En aquest darrer cas es poden distingir dues regions d'operació. Per $|V_{DS}|<|V_{GS}|-|V_{Tp}|$, el transistor es troba a la regió òhmica, on la intensitat respon a l'equació:

$$I_{DS} = \frac{I_{DSS}}{V_{Tp}^2}\left(2\left(|V_{GS}| - |V_{Tp}|\right)|V_{DS}| - |V_{DS}|^2\right) \text{ òhmica} \quad (23)$$

En aquesta, la funcionalitat de $I_{DS}=I_{DS}(|V_{GS}|, |V_{DS}|, |V_{Tp}|)$ és la mateixa que pel transistor MOS d'empobriment. Així mateix, si $|V_{DS}|<<1$ es pot aproximar per:

$$I_{DS} = \frac{2I_{DSS}}{V_{Tp}^2}\left(|V_{GS}| - |V_{Tp}|\right)|V_{DS}| \text{ òhmica} \quad (24)$$

Per $|V_{DS}|\geq|V_{GS}|-|V_{Tp}|$ el transistor es troba a la regió de saturació. L'equació per la intensitat drenador-font és:

$$I_{DS} = \frac{I_{DSS}}{V_{Tp}^2}\left(|V_{GS}| - |V_{Tp}|\right)^2 \text{ saturació} \quad (25)$$

on la intensitat és constant. Finalment, es presenten dues taules a mode de resum amb les equacions per cadascuna de les regions d'operació pels transistors nMOS d'enriquiment i empobriment.

Regió	Condicions	Equacions
Tall	$V_{GS} \leq V_{Tn}$	$I_{DS} = 0$
Òhmica	$V_{GS} \geq V_{Tn}$ i $V_{GS} - V_{Tn} \geq V_{DS}$	$I_{DS} = \frac{k_n}{2}\frac{W}{L}[2(V_{GS} - V_{Tn})V_{DS} - V_{DS}^2]$
Saturació	$V_{GS} \geq V_{Tn}$ i $V_{GS} - V_{Tn} \leq V_{DS}$	$I_{DS} = \frac{k_n}{2}\frac{W}{L}(V_{GS} - V_{Tn})^2$

Taula 1 Regions d'operació i equacions pels transistors nMOS d'enriquiment.

Regió	Condicions	Equacions
Tall	$V_{GS} \leq V_{Tn}$	$I_{DS} = 0$
Òhmica	$V_{GS} \geq V_{Tn}$ i $V_{GS} - V_{Tn} \geq V_{DS}$	$I_{DS} = \frac{I_{DSS}}{V_{Tn}^2}[2(V_{GS} - V_{Tn})V_{DS} - V_{DS}^2]$
Saturació	$V_{GS} \geq V_{Tn}$ i $V_{GS} - V_{Tn} \leq V_{DS}$	$I_{DS} = \frac{I_{DSS}}{V_{Tn}^2}(V_{GS} - V_{Tn})^2$

Taula 2 Regions d'operació i equacions pels transistors nMOS d'empobriment.

4. CONCLUSIONS

Es pot concloure que el transistor MOSFET és un dispositiu format per quatre terminals (*drain*, *gate*, *source* i *bulk*), entre dos dels quals (*drain* i *source*) circula un corrent que és governat per la tensió aplicada a un tercer (*gate*).

Se n'han estudiat diverses estructures (canal p/canal n, enriquiment/empobriment) i segons la tensió aplicada als terminals, s'han distingit diferents regions de treball (tall, òhmica o lineal i saturació), per les quals s'han obtingut les equacions que descriuen la característica I(V) del transistor.

5. REFERÈNCIES

[Sedra91] A. S. Sedra, K. C. Smith, “Microelectronic Circuits”, Saunders Collegue Publishing, 3a edició, 1991.
[Schilling93] D. L. Schilling, Belove, “Circuits electrònics discrets i integrats”, McGraw-Hill, 3a edició, 1993.
[Ghausi87] M. S. Ghausi, “Circuits electrònics discrets i integrats”, Nueva Editorial Interamericana, 1987.
[Prat94] Ll. Prat, “Circuits i dispositius electrònics”, Politext, 1994.
[Rubio00] A. Rubio, “Disseny de circuits i sistemes integrats”, Politext, 2000.

A1. Models del transistor MOS.

***Abstract*—**En aquest article es detallen alguns dels models (*Level 1*, *Level 2*, *Level 3*, BSIM3v3, BSIM4 i EKV) que els simuladors comercials utilitzen per simular el comportament del transistor MOS. Així mateix, també es comenten les aproximacions realitzades a cada model per tal d'extreure els diferents paràmetres que donen compte dels fenòmens físics que tenen lloc en el citat transistor.

***Index Terms*—**Canal, ZCE, potencial de superfície, efectes de vora, tensió llindar, zona de deplexió, tensió de *flat-band*, *body factor*, corrent de *drain*, corrent de difusió, corrent de deriva, mobilitat, transconductància, modulació del canal, estrangulament, *pinch-off*, velocitat límit dels portadors, longitud efectiva, velocitat de saturació.

1. INTRODUCCIÓ

La simulació de circuits electrònics es realitza a partir de models que descriuen el comportament dels components que integren el circuit. Aquests models s'extreuen del comportament real de dispositius aïllats a través d'expressions amb base física o bé mitjançant ajustament empíric a les seves corbes de resposta.

Els simuladors elèctrics tipus SPICE són capaços d'implementar diferents models d'un mateix dispositiu, els quals difereixen en el grau de similitud amb el comportament real del dispositiu. La utilització d'un model o altre depèn de la precisió i temps de simulació desitjats. D'aquesta manera, els models més simples s'empren quan l'exactitud és menys important que el temps de *run*.

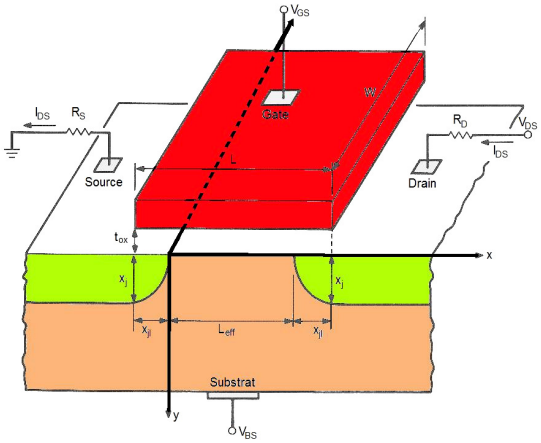


Fig. 1 Esquema d'un transistor MOS.

Donat que la tecnologia MOS és la que domina el mercat electrònic, també és la que disposa de més models de descripció. En aquest article, es descriuen els models *Level 1*, *Level 2* i *Level 3*, que es presenten ordenats de menor a major exactitud i complexitat. Aquests models es caracteritzen per utilitzar diferents expressions per una mateixa variable en funció del règim de funcionament del transistor MOS. Finalment es realitzen consideracions sobre models més avançats com són BSIM3v3, BSIM4 i EKV, els quals utilitzen equacions contínues, és a dir, vàlides per tots els règims d'operació.

2. *LEVEL 1*

Aquest model va ser proposat per Schichman i Hodges.

2.1. Formació del canal

La tensió V_{GS} aplicada a un transistor MOS pot ser expressada com:

$$V_{GS} = \phi_{ox} + \phi_s + \phi_{ms} \tag{1}$$

on Φ_{ox} és el potencial que cau a l'òxid, Φ_S és el potencial que cau a la zona de càrrega espacial (ZCE), també anomenat potencial de superfície, i Φ_{ms} correspon a la caiguda de tensió entre òxid i semiconductor. Per tal que es formi el canal, ha de passar que:

$$\phi_s = 2\phi_p - V_{BS} \tag{2}$$

on Φ_p és el potencial de la zona neutra, que depèn de la concentració de dopatge, tal com es mostra a continuació.

$$\phi_p = \frac{K_B T}{q} \ln\left(\frac{N_A}{n_i}\right) \tag{3}$$

La distribució de potencial expressada a l'eq. (1) és constant a la zona central del dipositiu, però presenta variacions als extrems en funció de la polarització. Donat que al *Level 1* no es consideren aquests efectes de vora, la profunditat de la ZCE ve donada per l'eq. (4), on s'ha considerat que la unió òxid-semiconductor és abrupta. En

aquesta equació, ε_s correspon a la permitivitat del semiconductor.

$$x_B = \sqrt{\frac{2\varepsilon_s}{qN_A}\phi_s} \tag{4}$$

2.2. Tensió llindar

S'anomena tensió llindar (V_T) la tensió V_{GS} per la qual la concentració de minoritaris s'iguala a la de majoritaris. El canal correspon a la regió de la ZCE que es troba a una tensió superior a V_T. A la resta de la ZCE, se l'anomena zona de deplexió.

Per tal de trobar l'expressió per la tensió llindar, cal aplicar diverses aproximacions. En primer lloc, cal tenir en compte que el dopatge N_A del substrat és uniforme. En aquesta situació, es té que el camp elèctric és totalment perpendicular a la interfície en qualsevol punt del canal i que les distribucions de potencial de superfície (Φ_s) i de càrrega al semiconductor (Q_s) són homogènies per tota la ZCE. A més, l'aproximació de deplexió considera que la concentració de portadors minoritaris a la ZCE és nul·la quan $V_{GS}<V_T$. Finalment, l'equació per la tensió llindar ve donada per:

$$V_{Tideal} = V_{FB} + 2\phi_p + \gamma\sqrt{2\phi_p - V_{BS}} \tag{5}$$

on V_{FB} és l'anomenada tensió de *flat-band* i correspon a la polarització externa que ha de ser aplicada per tal de contrarestar els efectes de diferència de funcions de treball entre òxid i semiconductor, i també de la càrrega acumulada a l'òxid i a la interfície. Finalment, γ és un paràmetre d'ajust i s'anomena paràmetre d'efecte substrat o *body factor*. Aquest paràmetre pot trobar-se a partir de l'expressió:

$$\gamma = \frac{\sqrt{2q\varepsilon_s N_A}}{C_{ox}} \tag{6}$$

2.3. Corrent de *drain*

El terme $V_C(x)$ es refereix a la distribució de potencial que hi ha entre un punt x qualsevol del canal i *source*, resultat d'aplicar una tensió V_{DS} diferent de zero. Així doncs, i donat que en aquest model totes les tensions estan referenciades a *source*, la tensió entre la porta i un punt qualsevol del canal és la suma de les tensions V_{GS} i $V_C(x)$. Tot seguit, es dedueix l'expressió pel corrent de *drain* a partir dels diferencials de càrrega que es tenen al canal. El valor d'aquesta càrrega es troba a partir de l'expressió:

$$dQ_I(x) = C_{ox}W\left[V_{GS} - V_T\gamma\left(\sqrt{2\phi_p - V_{BS} + V_c(x)} - \sqrt{2\phi_p - V_{BS}}\right) - V_c(x)\right]dx \tag{7}$$

Al *Level 1* es realitza l'aproximació de menysprear el terme $V_C(x)$ que es troba dins de l'arrel de l'equació anterior. Els dos mecanismes causants del moviment dels portadors de càrrega són l'arrossegament i la difusió, però en aquest nivell s'aproxima I_D com un corrent totalment degut a l'arrossegament de camp. L'origen d'aquest corrent d'arrossegament es troba en els portadors que es mouen a una velocitat que depèn del camp longitudinal i del potencial $V_C(x)$ tal com es mostra a l'eq. (8). Reagrupant els termes d'aquesta expressió es troba l'eq. (9).

$$v(x) = \frac{dx}{dt} = -\mu E_x(x) = \mu\frac{dV_c(x)}{dx} \tag{8}$$

$$\frac{1}{dt} = \mu\frac{dV_c(x)}{dx^2} \tag{9}$$

D'altra banda, la intensitat es defineix com:

$$I(x) = \frac{dQ_I}{dt} \tag{10}$$

Igualant les expressions (9) i (10) i integrant a banda i banda, s'obté:

$$\int_0^{L_{eff}} I(x)dx = \int_0^{L_{eff}} dQ_I\mu\frac{dV_c(x)}{dx^2}dx \tag{11}$$

En aquest punt del *Level 1* es considera que I(x) i la mobilitat són constants per qualsevol secció del canal. A més, el terme I(x) s'anomena I_D. Ara, substituint l'expressió de dQ_I a l'equació anterior, es té:

$$I_D = \mu C_{ox}\frac{W}{L}\int_0^{L_{eff}}\left[V_{GS} - V_T\gamma\left(\sqrt{2\phi_p - V_{BS} + V_c} - \sqrt{2\phi_p - V_{BS}}\right) - V_c\right]dx \tag{12}$$

Resolent la integral de l'equació anterior, finalment s'arriba a:

$$I_D = \beta\left[(V_{GS} - V_T)V_{DS} - \frac{V_{DS}^2}{2}\right] \tag{13}$$

on el coeficent β s'anomena transconductància i ve donat per:

$$\beta = \mu C_{ox}\frac{W}{L} \tag{14}$$

Els resultats experimentals demostren que existeix una dependència entre I_D i V_{DS} que no es veu reflectida a

l'expressió per I_D que s'ha desenvolupat en aquest nivell. Aquesta dependència, que s'anomena modulació del canal, es manifesta fortament al règim de saturació i es modela mitjançant el paràmetre de modulació de canal (λ). En models superiors al *Level 1* es proporciona una explicació física de la modulació del canal al règim de saturació, però en el present model es considera com una mera rectificació empírica.

2.4. Consideracions pel règim de saturació

Quan la tensió V_{DS} esdevé més gran que V_{GS}-V_T, es diu que el transistor ha entrat al règim de saturació i la seva resposta I(V) és diferent a la descrita fins ara. Per continuïtat, es pren:

$$I_D = I_D(\lambda = 0)(1 + \lambda V_{DS}) \tag{15}$$

La tensió V_{DS} que permet passar al règim de saturació és V_{Dsat}. Amb V_{DS}=V_{Dsat} la saturació es dóna a l'alçada del *drain*, mentre que per V_{DS} majors té lloc el fenomen de l'estrangulament del canal. L'origen d'aquest rau en el fet que la caiguda de tensió a la secció del canal entre *drain* i un punt anomenat *pinch-off* no és més gran que la tensió llindar. Entre *drain* i el punt de *pinch-off* no existeix canal.

En general, la tensió V_{DS} no és exactament V_{Dsat} en el moment en què comença a produir-se el fenomen de l'estrangulament del canal. Aquesta petita diferència, que s'anomena efecte de la velocitat límit dels portadors, no es té en compte al *Level 1*, considerant que el règim de saturació es dóna per aquella V_{DS} que estrangula el canal.

3. *LEVEL 2*

Aquest model, que va ser proposat per Grove i Frohman, és el més comú en tots els simuladors.

3.1. Tensió llindar

En aquest nivell es tenen en compte els efectes de vora quan el terme x_B és comparable a la longitud efectiva i amplada del canal, per la qual cosa els termes Φ_S i Q_I no presenten una distribució homogènia. A més, com que es modifica la distribució de la ZCE, apareixen en la tensió llindar dependències amb les polaritzacions. Es vol remarcar que els efectes de vora només són considerats a la zona de deplexió i no a la resta del canal.

Tot seguit, es presenta l'equació pel paràmetre γ, que correspon a l'expressió que es tenia al *Level 1*, a la qual se li ha aplicat una correcció. A més, també es donen les expressions per W_S, W_D i X_D, paràmetres que es troben representats a la Fig. 2.

$$\gamma = \gamma(Level1)\left[1 - \frac{x_j}{2L_{eff}}\left(\sqrt{1+\frac{2W_s}{x_j}} + \sqrt{1+\frac{2W_D}{x_j}} - 2\right)\right] \tag{16}$$

$$W_s = x_D\sqrt{2\phi_p - V_{BS}} \tag{17}$$

$$W_D = x_D\sqrt{2\phi_p - V_{BS} + V_{DS}} \tag{18}$$

$$x_D = \sqrt{\frac{2\varepsilon_s}{qN_A}} \tag{19}$$

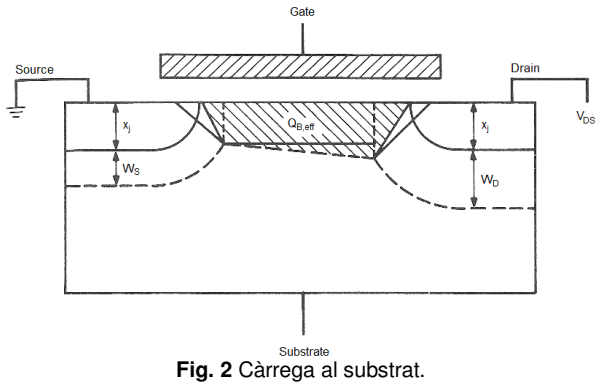


Fig. 2 Càrrega al substrat.

Tot i utilitzar la γ de l'eq. (16), l'expressió per la V_T del *Level 2* té una aparença molt semblant a la del *Level 1*. Tal com s'ha anunciat, apareixen dependències amb V_{BS} i V_{DS}.

$$V_T = V_{Tideal} + \frac{\varepsilon_s \delta\pi}{2C_{ox}}\left(2\phi_p - V_{BS}\right) \tag{20}$$

Encara que aquesta expressió per V_T és més precisa que la del *Level 1*, presenta una sèrie de limitacions. En primer lloc, se sobreestima la reducció de V_T per L_{eff} petites. A més, no s'obté una bona dependència amb V_{DS}. Per una millor coincidència amb els resultats experimentals, s'utilitzen variacions en els valors de N_A i x_j, encara que aquesta solució no proporciona bons resultats si el rang de valors és gran.

3.2. Mobilitat

En aquest nivell, es té en compte la dependència amb la polarització a l'hora de determinar la mobilitat dels portadors, tal com es mostra a la següent expressió.

$$\mu = \mu_0\left(\frac{\varepsilon_s}{\varepsilon_{ox}}\frac{U_c t_{ox}}{V_{GS} - V_T - U_t V_{DS}}\right)^{U_e} \tag{21}$$

Aquí intervenen els paràmetres camp elèctric entre porta i canal (U_c), la contribució de la tensió de *drain* al camp U_c (U_t), el camp elèctric promig perpendicular al canal (V_{GS}-V_T-U_t·V_{DS}) i, finalment, el coeficient experimental per la mobilitat (μ_e). Aquest model dóna molt bons resultats en

absència d'efectes de vora.

3.3. Corrent de difusió

Al *Level 2* es considera no nul·la la contribució del corrent de difusió a I_D. Aquesta pot implementar-se de diverses formes, però aquí es proporcionarà un model que considera I_D total com la suma d'una contribució deguda a la difusió i una altra deguda a la deriva. A l'eq. (22) es proporciona una expressió pel corrent I_D de difusió.

$$I_{Ddif} = \frac{qD_n x_c W_{eff}}{L_n \tanh\left(\frac{L_{eff}}{L_n}\right)} \frac{n_s n_x}{(n_s + n_x)} \left(1 - e^{-qV_{DS}/KT}\right) \quad (22)$$

A l'equació anterior intervenen la constant de difusió (D_n), la longitud de difusió (L_n), el gruix promig del canal (x_c) i el valor assimptòtic de la concentració de portadors al canal en situació de feble inversió (n_x). També intervé el factor n_s, que ve donat per l'expressió següent, on α és un paràmetre que s'inclou per incrementar el pendent de $\ln(I_D)$ respecte V_{GS} quan $V_{BS}>0$.

$$n_s = N_{\sup} \exp\left[\frac{(1 - \alpha V_{BS})(V_{GS} - V_T)}{\left(1 + q\frac{N_{FS}}{C_{ox}}\right)\left(\frac{KT}{q}\right)}\right] \quad (23)$$

3.4. Corrent de deriva

En aquest apartat es donarà una expressió per I_D que contempla exclusivament la contribució dels efectes de deriva. La seva deducció és anàloga a la del *Level 1*, amb l'excepció que el terme $V_c(x)$ que es troba dins de l'arrel quadrada no s'anul·la (veure eq. (7)). El *Level 2* comparteix les deduccions del *Level 1* fins al resultat de l'eq. (11). A partir d'aquest punt, es diferencien en la dQ_I utilitzada. L'expressió pel corrent de deriva ve donada per l'eq. (24).

$$I_{Dder} = \beta\left\{\left(V_{GS} - V_{FB} - 2\phi_p - \frac{V_{DS}}{2}\right)V_{DS} + \right.$$
$$\left. + \frac{2}{3}\gamma\left[\left(V_{DS} - V_{BS} - 2\phi_p\right)^{1.5} - \left(-V_{BS} + 2\phi_p\right)^{1.5}\right]\right\} \quad (24)$$

3.5. Velocitat de saturació dels portadors

Quan s'arriba al règim de saturació, es té una densitat de càrrega de canal dQ_I nul·la. No obstant això, s'ha de tenir en compte que en saturació es té una circulació de portadors, tant per la zona estrangulada com per la zona que no ho està. D'altra banda, el tipus de semiconductor i la conductivitat determinen la velocitat màxima dels portadors ($v_{màx}$).

Amb aquests condicionants, la densitat de càrrega que es té al punt de *pinch-off* ve donada per l'expressió que es mostra tot seguit.

$$Q_I(V_{Dsat}) = \frac{I_{Dsat}}{W v_{màx}} \quad (25)$$

3.6. Modulació del canal al règim de saturació

La modulació del canal al *Level 2* és considerada de diferent manera que al nivell anterior, tal com es mostra a l'eq. (26). A més, el denominador d'aquesta expressió pot obtenir-se a partir de l'eq. (27).

$$I_D = \frac{I_D(\lambda = 0)}{1 - \lambda V_{DS}} \quad (26)$$

$$\frac{1}{1 - \lambda V_{DS}} = 1 + \lambda V_{DS} + O^2(V_{DS}) \quad (27)$$

D'altra banda, la relació entre la longitud efectiva del canal no estrangulat (L') i V_{DS} ve donada per la següent equació, a partir de la qual s'obté una expressió que permet calcular el valor del paràmetre λ.

$$L' = L_{eff}(1 - \lambda V_{DS}) \quad (28)$$

$$\lambda = \frac{L_{eff} - L'}{L_{eff} V_{DS}} \quad (29)$$

Un model més realista per L' és el que es planteja a continuació, on es té en compte l'efecte de la velocitat de saturació dels portadors (v_{sat}).

$$L_{eff} - L' = x_D \sqrt{\left(\frac{x_D v_{màx}}{2\mu_0}\right)^2 + V_{DS} - V_{Dsat}} - \frac{x_D^2 v_{màx}}{2\mu_0} \quad (30)$$

$$x_D = \sqrt{\frac{2\varepsilon_B}{qN_A N_{eff}}} \quad (31)$$

Aquí N_{eff} és un paràmetre d'ajust.

4. *LEVEL 3*

El *Level 3*, que va ser desenvolupat per Liu el 1981, és computacionalment més eficient que el seu predecessor, tot i que és totalment coincident amb aquest en algunes regions.

4.1. Tensió llindar

L'expressió que s'utilitza per la tensió llindar en aquest

model, que sorgeix d'hipòtesis molt semblants a les emprades al *Level 2*, pot veure's a l'eq. (32), en la qual apareixen diversos termes. Els dos primers sumands són idèntics als dels nivells anteriors, mentre que els dos últims donen compte dels efectes de vora, tal com es feia al *Level 2*. El tercer sumand, en canvi, modela la variació de V_T amb V_{DS}, on el factor σ és empíric i ve donat per l'expressió (33), solucionant així problemes anteriors. El paràmetre γ té la mateixa expressió que al *Level 1*.

$$V_T = V_{FB} + 2\phi_p - \sigma V_{DS} + \gamma F_s \sqrt{2\phi_p - V_{BS}} + F_n\left(2\phi_p - V_{BS}\right) \quad (32)$$

$$\sigma = \eta \frac{8.15 \cdot 10^{22}}{C_{ox} L_{eff}^2} \quad (33)$$

El paràmetre η, que apareix a l'equació anterior, típicament val 1, tot i que és ajustable al model. D'altra banda, els efectes de canal s'introdueixen amb el factor F_S.

$$F_s = 1 - \frac{x_j}{L_{eff}} \left(\frac{x_{jl} + W_c}{x_j} \sqrt{1 - \left(\frac{W_p}{x_j + W_p} \right)^2} - \frac{x_{jl}}{x_j} \right) \quad (34)$$

El terme W_p és el gruix de la zona de deplexió a la regió plana de la unió.

$$W_p = x_D \sqrt{\phi_j - V_{BS}} \quad (35)$$

$$x_D = \sqrt{\frac{2\varepsilon_s}{qN_A}} \quad (36)$$

El paràmetre W_c pot trobar-se empíricament a partir de l'equació que es mostra tot seguit.

$$\frac{W_c}{x_j} = 0.0631353 + 0.8013292 \frac{W_p}{x_j} - 0.01110777 \left(\frac{W_p}{x_j} \right)^2 \quad (37)$$

4.2. Mobilitat

En aquest nivell, la mobilitat dels portadors ve donada per l'eq. (38), on el terme μ_S s'obté a partir de l'expressió (39).

$$\mu = \mu_{eff} \frac{\mu_s}{1 + \mu_s \dfrac{V_{DS}}{v_{màx} L_{eff}}} \quad (38)$$

$$\mu_s = \frac{\mu_0}{1 + \theta\left(V_{GS} - V_T\right)} \quad (39)$$

En aquestes expressions, que són empíriques, es modelen les dependències de la mobilitat amb V_{GS} i I_D amb major exactitud i simplicitat.

4.3. Corrent de *drain*

La contribució del corrent de difusió es modela de la mateixa manera que al *Level 2*. Per la contribució del corrent de deriva, en canvi, es realitza una aproximació de Taylor de primer ordre respecte la càrrega del substrat a l'expressió per I_D ja proporcionada al *Level 2*. D'aquesta manera, es tenen equacions més manejables.

$$I_{Dder} = \beta \left(V_{GS} - V_T - \frac{1 + F_B}{2} V_{DS} \right) V_{DS} \quad (40)$$

$$F_B = \frac{\gamma F_s}{4\sqrt{2\phi_p - V_{BS}}} + F_n \quad (41)$$

A l'equació anterior, es manifesten els efectes de vora en ser F_s i F_n els mateixos que s'han utilitzat anteriorment a les expressions per V_T i γ.

4.4. Tensió de *drain* en saturació

La tensió V_{DS} de saturació s'obté a partir de l'eq. (42), on els termes V_a i V_b corresponen a la tensió V_{DS} de saturació sense incloure efectes de $v_{màx}$ i a la tensió que sí que inclou $v_{màx}$, respectivament.

$$V_{Dsat} = V_a + V_b - \sqrt{V_a^2 + V_b^2} \quad (42)$$

$$V_a = \frac{V_{GS} - V_T}{1 + F_B} \quad (43)$$

$$V_B = \frac{v_{màx, L_{eff}}}{\mu_s} \quad (44)$$

Aquesta expressió presenta continuïtat en la transició de V_{DS} a V_{Dsat}.

4.5. Modulació del canal en règim de saturació

En aquest nivell, la dependència de la longitud efectiva del canal (L') amb V_{DS} es modela de diferent manera que al *Level 2*, tal com es mostra tot seguit.

$$L' = L_{eff} + \sqrt{\left(\frac{E_p x_D}{2} \right)^2 + K x_D \left(V_{DS} - V_{Dsat} \right)} - \frac{E_p x_D^2}{2} \quad (45)$$

$$E_p = \frac{I_{Dsat}}{g_{Dsat} L_{eff}} \quad (46)$$

El paràmetre K pren el valor típic 1. D'altra banda, el terme E_p correspon al camp lateral al punt de *pinch-off*. A

més, es compleixen (47) i (48).

$$g_{Dsat} = \left.\frac{\partial I_D}{\partial V_{DS}}\right|_{V_{DS}=V_{Dsat}} \tag{47}$$

$$I_{Dsat} = I\left(V_{DS} = V_{Dsat}\right) \tag{48}$$

5. BSIM3v3

La miniaturització dels sistemes requereix que els transistors MOS treballin sota condicions molt restrictives, caracteritzades pels baixos estats de polarització, els reduïts òxids de porta i les elevades freqüències. Al seu torn, aquest fenomen exigeix models pel MOS cada vegada més exactes. Això s'aconsegueix mitjançant la utilització de nous paràmetres, els quals permeten modelar tots els efectes que tenen lloc en un transistor MOS.

El BSIM3v3, que va ser desenvolupat per la Universitat de Berkeley a la dècada dels 90 i actualment constitueix l'estàndard de la indústria, és el primer dels tres models avançats que seran descrits en aquest article. Aquest model també és el primer que utilitza expressions contínues per les seves magnituds, no definides a trams. Encara que els efectes que s'exposen a continuació siguin explicats per trams, al final s'apliquen funcions de suavitzat.

5.1. Tensió llindar

Un model acurat de la tensió llindar (V_T) és un dels requeriments més importants per una descripció precisa de les característiques elèctriques d'un dispositiu. L'eq. (5), que modela la tensió llindar pel nivell més simple dels descrits en aquest article, assumeix que el canal és uniforme i utilitza l'aproximació unidimensional de Poisson en la direcció vertical del canal. Aquesta expressió és vàlida només quan la concentració de dopatge del substrat és constant i el potencial és uniforme al llarg del canal, condicions que no sempre són certes. Tot seguit, es presenta un conjunt de situacions que han de ser tingudes en compte per obtenir un model més exacte per la tensió llindar.

5.1.1. Dopatge no uniforme en la direcció vertical

La concentració de dopatge és habitualment més elevada a la interfície que no a la profunditat del substrat. Aquesta no uniformitat provoca que el paràmetre d'efecte substrat depengui del voltatge aplicat. Si l'amplada de la regió de deplexió és menor que x_t (veure Fig. 3), el terme N_A de l'equació pel coeficient d'efecte substrat (eq. (6)) és igual a N_{ch}. En cas contrari, val N_{sub}.

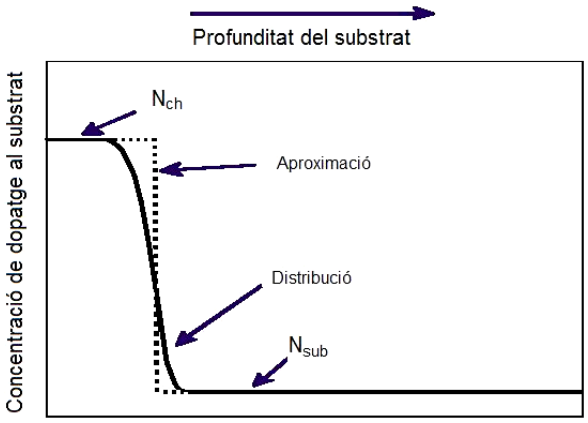


Fig. 3 Perfil real i aproximat del dopatge vertical.

Per tal de tenir en compte aquesta no uniformitat en el dopatge del substrat, es proposa la següent equació per la tensió llindar:

$$V_T = V_{Tideal} + K_1\left(\sqrt{\phi_s - V_{BS}} - \sqrt{\phi_s}\right) - K_2 V_{BS} \tag{49}$$

que coincideix amb l'eq. (5) quan el potencial del substrat és nul. Els paràmetres K_1 i K_2 poden ser determinats tenint en compte que V_T i la seva derivada respecte V_{BS} haurien de presentar el mateix valor a V_{bm}, on V_{bm} és la màxima tensió del substrat. Així doncs, K_1 i K_2 vénen donats per les expressions:

$$K_1 = \gamma_2 - 2K_2\sqrt{\phi_s - V_{bm}} \tag{50}$$

$$K_2 = \left(\gamma_1 - \gamma_2\right)\frac{\sqrt{\phi_s - V_{bx}} - \sqrt{\phi_s}}{2\sqrt{\phi_s}\left(\sqrt{\phi_s - V_{bm}} - \sqrt{\phi_s}\right) + V_{bm}} \tag{51}$$

on γ_1 i γ_2 representen el coeficient d'efecte substrat quan la concentració de dopatge al substrat és N_{ch} o N_{sub}, respectivament.

$$\gamma_1 = \frac{\sqrt{2q\varepsilon_{Si}N_{ch}}}{C_{ox}} \tag{52}$$

$$\gamma_2 = \frac{\sqrt{2q\varepsilon_{Si}N_{sub}}}{C_{ox}} \tag{53}$$

Finalment, V_{bx} és la polarització del substrat quan l'amplada de deplexió és X_t. En aquesta situació, V_{bx} pot calcular-se tal com es mostra a continuació.

$$\phi_s - V_{bx} = \frac{qN_{ch}x_t^2}{2\varepsilon_{Si}} \tag{54}$$

5.1.2. Dopatge no uniforme en la direcció horitzontal

En algunes tecnologies, la concentració de dopatge és major a la regió de *drain* o *source* que no a la zona central del canal, fenomen conegut com dopatge lateral no uniforme. Aquest dopatge lateral no uniforme provocarà que la tensió llindar augmenti, ja que la concentració promig de dopatge al substrat també és més gran. El dopatge promig al canal es pot calcular segons l'eq. (55).

$$N_{eff} = N_a\left(1+\frac{2L_x}{L}\frac{N_{ds}-N_a}{N_a}\right) \equiv N_a\left(1+\frac{N_{lx}}{L}\right) \tag{55}$$

Tenint en compte aquest fenomen, l'expressió per la tensió llindar esdevé:

$$V_T = V_{Tideal} + K_1\left(\sqrt{\phi_s - V_{BS}} - \sqrt{\phi_s}\right) - K_2 V_{BS} + \\ + K_1\left(\sqrt{1+\frac{N_{lx}}{L_{eff}-1}}\right)\sqrt{\phi_s} \tag{56}$$

5.1.3. Efectes de canal curt

Els efectes de canal curt han de ser tinguts en compte en l'expressió de la tensió llindar que modela dispositius que presenten dimensions per sota de la micra. Aquests efectes es modelen afegint a l'eq. (56) el sumand $-\Delta V_T$, segons el model desenvolupat per Liu et al.

$$\Delta V_T = \theta_t(L)\left[2\left(V_{bi}-\phi_s\right)+V_{DS}\right] \tag{57}$$

A l'equació anterior, V_{bi} és el potencial de construcció de la junció pn entre substrat i *source*. Es calcula com:

$$V_{bi} = \frac{K_B T}{q}\ln\left(\frac{N_{ch}N_d}{n_i^2}\right) \tag{58}$$

Aquí, N_d i N_{ch} corresponen a la concentració de dopatge de *source* i de substrat, respectivament. D'altra banda, el terme $\theta_t(L)$ presenta una gran dependència amb la longitud del canal i pot calcular-se com:

$$\theta_t(L) = \left[\exp\left(-L/2l_t\right) + 2\exp\left(-L/l_t\right)\right] \tag{59}$$

En aquesta expressió, l_t fa referència a la longitud característica i x_{dep} és l'amplada a la regió de deplexió del substrat, que és més important prop de la regió de *drain*.

$$l_t = \sqrt{\frac{\varepsilon_{Si} t_{ox} X_{dep}}{\varepsilon_{ox}\eta}} \tag{60}$$

$$x_{dep} = \sqrt{\frac{2\varepsilon_{si}\left(\phi_s - V_{BS}\right)}{qN_{ch}}} \tag{61}$$

5.1.4. Efectes de canal estret

La regió de deplexió real al canal és sempre més gran que el que habitualment s'assumeix. En aquest model, es dóna compte del canvi que aquest fenomen suposa en la tensió llindar afegint un nou terme a l'equació per V_T.

$$\left(K_3 + K_{3b}V_{BS}\right)\frac{t_{ox}}{\left(W_{eff}^{'} + W_0\right)}\phi_s \tag{62}$$

Aquí s'introdueixen els paràmetres K_3, K_{3b} i W_0. W'_{eff} és l'amplada efectiva del canal i no presenta dependència amb les polaritzacions. Finalment, l'equació per la tensió llindar pel model BSIM3v3 és la que es mostra tot seguit.

$$V_T = V_{Tideal} + K_1\left(\sqrt{\phi_s - V_{BS}} - \sqrt{\phi_s}\right) - K_2 V_{BS} + \\ + K_1\left(1+\frac{N_{lx}}{L_{eff}-1}\right)\sqrt{\phi_s} - \theta_t(L)\left[2\left(V_{bi}-\phi_s\right)+V_{DS}\right] + \\ + \left(K_3 + K_3 V_{BS}\right)\frac{t_{ox}}{\left(W_{eff}^{'} + W_0\right)}\phi_s \tag{63}$$

5.2. Mobilitat

Modelar correctament la mobilitat dels portadors és necessari per aconseguir exactitud. Així doncs, és necessari tenir en compte els molts paràmetres i condicions de polarització de què depèn aquesta. L'expressió empírica per la mobilitat ve donada per l'eq. (64), on el valor per E_{eff} es troba a partir de l'eq. (65).

$$\mu_{eff} = \frac{\mu_0}{1+\left(E_{eff}/E_0\right)^v} \tag{64}$$

$$E_{eff} = \frac{Q_B + Q_{n/2}}{\varepsilon_{Si}} \tag{65}$$

En últim lloc, es vol destacar que els paràmetres μ_0, E_0 i v presenten valors fixos en funció del tipus de portador. A la taula següent, i segons Liang et al. i Toh et al., es proporcionen els valors d'aquestes constants.

Paràmetre	e-	h+
μ_0 (cm^2/V)	670	160
E_0 (MV/cm)	0.67	0.7
v	1.6	1.0

Taula 1 Valors per μ_0, E_0 i v segons Liang et al. i Toh et al.

5.3. Velocitat de deriva

El BSIM3v3 utilitza un model simple i semiempíric per

modelar la velocitat de deriva:

$$v = \begin{cases} \dfrac{\mu_{eff} E}{1 + (E/E_{sat})}, E < E_{sat} \\ v_{sat}, E > E_{sat} \end{cases} \quad (66)$$

on E_{sat} correspon al camp elèctric al qual la velocitat dels portadors esdevé saturada. Per tal que el model presenti continuïtat, ha de satisfer-se:

$$E_{sat} = \frac{2 v_{sat}}{\mu_{eff}} \quad (67)$$

5.4. Efecte de càrrega substrat

Quan la tensió de *drain* és gran i/o la longitud del canal és important, el gruix de la regió de deplexió del canal no és uniforme. Aquest fenomen, que s'anomena efecte de càrrega substrat, provoca la variació de V_T al llarg del canal i és modelat mitjançant el paràmetre A_{bulk}, que es presenta tot seguit.

Per modelar aquest efecte, en el BSIM3v3 s'utilitza el paràmetre A_{bulk}, que es presenta tot seguit.

$$A_{bulk} = \left(1 + \frac{K_1}{2\sqrt{\phi_s - V_{BSeff}}}\left\{\frac{A_0 L_{eff}}{L_{eff} + 2\sqrt{x_j x_{dep}}}\left[1 - A_{GS} V_{GSeff}\left(\frac{L_{eff}}{L_{eff} + 2\sqrt{x_j x_{dep}}}\right)^2\right] + \frac{B_0}{W_{eff} + B_1}\right\}\right)\frac{1}{1 + K_{ETA} V_{BSeff}} \quad (68)$$

A l'expressió anterior, els coeficients A_0, A_{GS}, K_1, B_0, B_1 i K_{ETA} són determinats experimentalment. A més, es dedueix que A_{bulk} augmenta a mesura que la longitud del canal també ho fa.

5.5. Corrent de *drain* pel règim lineal

El càlcul de I_D es basa en el mateix procediment que pels models anteriors: s'integra la densitat de càrrega al canal i s'aplica l'aproximació de deplexió. Ara, però, s'utilitza una expressió per $dQ_I(x)$ més exacta, ja que es té en compte l'efecte de la càrrega del substrat, així com també un model més sofisticat per la velocitat d'arrossegament. D'aquesta manera, l'expressió pel corrent de *drain* en règim lineal queda com:

$$I_D = \mu_{eff} C_{ox} \frac{W}{L} \frac{1}{1 + (V_{DS} L)/E_{sat}} \left(V_{GS} - V_T - \frac{A_{bulk} V_{DS}}{2}\right) V_{DS} \quad (69)$$

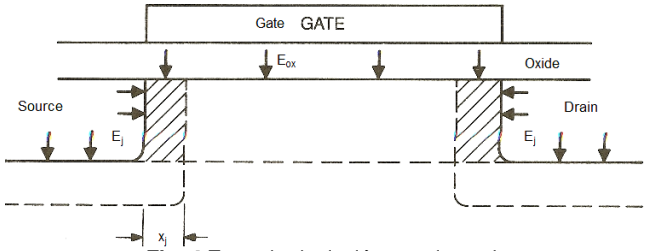


Fig. 4 Zona de deplexió sota el canal.

5.6. Corrent de *drain* pel règim de saturació

Hi ha tres mecanismes físics que afecten al corrent de *drain* en el règim de saturació: la modulació de la longitud del canal (CLM), la disminució de la barrera induïda pel *drain* (DIBL) i el corrent de substrat induït per l'efecte substrat (SCBE). Cadascuna d'aquestes contribucions domina en diferents trams de V_{DS}, tal com s'indica a la Fig. 5.

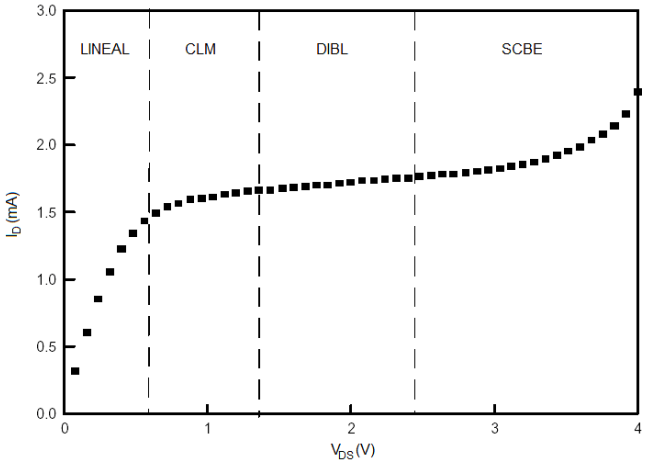


Fig. 5 Característica $I_D(V_{DS})$ per un transistor MOS.

Generalment, el corrent de *drain* depèn de les tensions de *gate* i *drain*. A la regió de saturació, però, I_D presenta una dependència molt petita amb V_{DS} i pot expressar-se com:

$$I_D \equiv I_{Dsat}\left(1 + \frac{V_{DS} - V_{Dsat}}{V_A}\right) \quad (70)$$

on

$$I_{Dsat} = W v_{sat} C_{ox}\left(V_{GSt} - A_{bulk} V_{Dsat}\right) \quad (71)$$

$$V_A = I_{Dsat}\left(\frac{\partial I_D}{\partial V_{DS}}\right)^{-1} \quad (72)$$

El paràmtre V_A s'anomena tensió Early i s'assumeix que

cadascun dels mecanismes citats anteriorment hi contribueix de forma independent. Així doncs, les tres contribucions poden ser calculades separadament, les expressions de les quals seran proporcionades tot seguit.

5.6.1. Modulació de la longitud del canal (CLM)

Quan la modulació de la longitud del canal és l'únic mecanisme que es considera, la tensió Early pot calcular-se com:

$$V_{ACLM} = \frac{A_{bulk} E_{sat} L + V_{GSt}}{A_{bulk} E_{sat} L} \tag{73}$$

5.6.2. Descens de la barrera induïda pel *drain* (DIBL)

Aquest és l'efecte produït per la caiguda de V_{DS} al canal, que modifica la distribució de càrrega al substrat i ve donat per una nova dependència de V_T amb V_{DS}. Aquesta contribució, que és major per L i V_{DS} grans, només es té en compte per I_{Dsat}. Llavors, si només es considera aquest efecte, la tensió Early pot calcular-se com:

$$V_{ADIBL} = \frac{\left(V_{GSeff} + 2v\right)}{\theta_{rout}\left(1 + P_{DIBLCB} V_{BSeff}\right)}\left(1 - \frac{A_{bulk} V_{Dsat}}{A_{bulk} V_{Dsat} + V_{GSeff} + 2v}\right) \tag{74}$$

on θ_{rout} s'obté a partir de:

$$\theta_{rout}(L) = P_{diblc1}\left[\exp\left(-D_{rout} L / 2l_t\right) + 2\exp\left(-D_{rout} L / l_t\right)\right] + P_{diblc2} \tag{75}$$

Aquí apareixen els termes P_{diblc1}, P_{diblc2} i D_{rout}. Aquests són paràmetres de correcció per la regió de forta inversió.

5.6.3. Corrent de substrat (SCBE)

El corrent de substrat és degut als parells electró-forat que apareixen gràcies al gran camp creat per V_{DS}. El corrent de substrat es computa mitjançant un nou sumand a l'expressió pel corrent de *drain*, obtenint així l'equació per I_D real (veure eq. (77)). V_{ASCBE} és la tensió Early deguda al corrent de substrat induït per l'efecte substrat i pot expressar-se com:

$$\frac{1}{V_{ASCBE}} = \frac{P_{SCBE2}}{L}\exp\left(-\frac{P_{SCBE1} l}{V_{DS} - V_{Dsat}}\right) \tag{76}$$

on les variables P_{SCBE1} i P_{SCBE2} es determinen experimentalment. Finalment, el corrent de *drain* total, després d'haver considerat els fenòmens CLM, DIBL i SCBE, ve donat per:

$$I_D = W v_{sat} C_{ox}\left(V_{GSt} - A_{bulk} V_{Dsat}\right)\left(1 + \frac{V_{DS} - V_{Dsat}}{V_A}\right)\left(1 + \frac{V_{DS} - V_{Dsat}}{V_{ASCBE}}\right) \tag{77}$$

5.7. Corrent de *drain* subllindar

En aquest model, es considera I_D subllindar no nul i es calcula segons:

$$I_{sub} = I_{S0}\left(1 - \exp\left(-\frac{V_{DS}}{V_t}\right)\right)\exp\left(\frac{V_{GS} - V_T - V_{off}}{n V_t}\right) \tag{78}$$

$$I_{S0} = \mu_0 \frac{W}{L}\sqrt{\frac{q\varepsilon_{Si} N_{ch}}{2\phi_s}} V_t^2 \tag{79}$$

Aquí, $V_t = K_B T/q$ és el voltatge tèrmic, V_{off} és el voltatge d'*offset* que es té quan V_{GS} és nul i n és el paràmtre de *swing*. Aquest últim depèn de la longitud del canal.

6. BSIM4

El model BSIM4 és una extensió del BSIM3v3, però està especialment enfocat a tecnologies menors de 100nm. Introdueix diversos efectes avançats per tal de donar cobertura als successius efectes d'escurçament dels transistors MOS.

Alguns dels models que utilitza el BSIM4 són iguals que els del seu predecessor, d'altres han sigut millorats. En aquest apartat, no es descriuran els models que són compartits amb el BSIM3v3, sinó només els que són exclusius del BSIM4.

6.1. Efecte del dopatge lateral no uniforme

Durant el procés de fabricació, es pot implantar un dopatge de canal superior a les proximitats de *drain* i *source* amb la finalitat d'aconseguir una disminució de V_T. La no uniformitat del dopatge en la direcció horitzontal es modela amb una contribució additiva i independent de V_{DS} a V_T. D'altra banda, en aquest model s'afegeix un sumand per l'efecte de la variació de la tensió llindar induït pel *drain* (DITS).

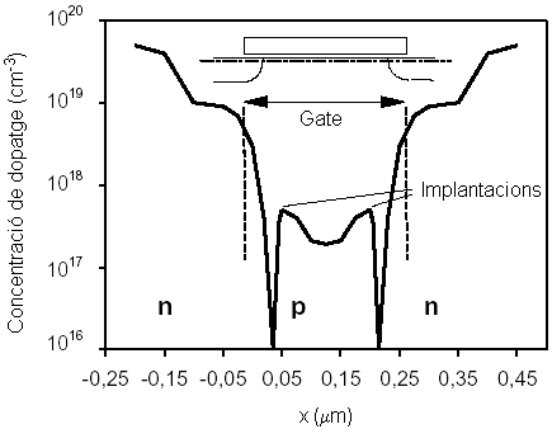


Fig. 6 Perfil de dopatge de les implantacions.

6.2. Corrent de *gate*

En tecnologies avançades s'aconsegueixen òxids de porta amb gruixos inferiors als 3nm. En aquesta situació, existeixen corrents que circulen a través de l'òxid de porta, que són modelats pel BSIM4.

Es distingeix entre cinc tipus de corrents diferents: corrents túnel entre *gate* i *bulk* (I_{GB}), entre *gate* i canal a la regió de *drain* (I_{GCD}), entre *gate* i canal a la regió de *source* (I_{GCS}), entre *gate* i difusió de *drain* (I_{GD}) i entre *gate* i difusió de *source* (I_{GD}). El corrent entre gate i canal (I_{GC}) correspon a la suma de I_{GCD} i I_{GCS}. La contribució d'aquests corrents a la intensitat total és additiva.

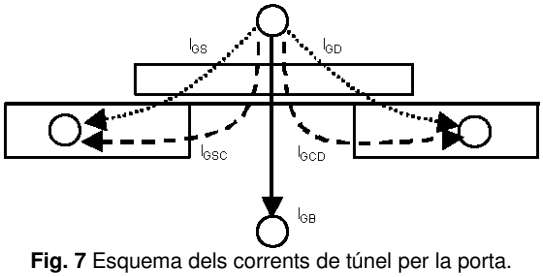


Fig. 7 Esquema dels corrents de túnel per la porta.

6.3. Corrent de *drain*

La diferència clau entre les expressions que utilitzen els models BSIM3v3 i BSIM4 pel corrent de drenador rau en el fet que pel primer model I_D és proporcional a C_{ox}, mentre que pel segon és proporcional a $C_{ox,eff}$. Aquests paràmetres donen compte del gruix d'òxid que realment és utilitzat en cada cas: gruix físic ($t_{ox,ph}$) en el primer model i gruix elèctric ($t_{ox,e}$) en el segon. Donat que aquesta diferència és més acusada en utilitzar òxids de porta menors, només és considerada al model BSIM4.

D'altra banda, es tenen en compte nous efectes en el càlcul del corrent de *drain* per la regió de saturació. El primer és degut al *pocket implant* i es modela afegint un corrector lineal per la variació de V_T amb V_{DS}. En segon lloc, també es té en compte el fenomen de la velocitat d'*overshoot*, que consisteix en la superació de la velocitat de saturació per part dels portadors a causa de les dimensions tan reduïdes. L'últim fenomen que es té en compte és l'anomenat *source end velocity*, que consisteix en considerar que la velocitat dels portadors no és constant a l'inici del seu recorregut pel canal.

7. EKV

El model EKV, que deu el nom als seus autors (Enz-Krummenacher-Vittoz) i va ser desenvolupat al *Swiss Federal Institute of Technology*, es caracteritza per ser continu per totes les regions de treball dels transistors MOS. Donat que les seves equacions són vàlides per tots els règims d'operació, ja sigui per sobre o per sota de la tensió llindar, el model EKV constitueix una eina potent per la simulació de tot tipus de circuits, però està especialment indicat per aplicacions MOS de baix consum.

Presenta la característica principal de referenciar tots els voltatges del dispositiu respecte el substrat en lloc de fer-ho respecte el terminal de *source*. Això condueix a expressions que presenten simetria respecte les tensions de *drain* i *source*.

7.1. Càrrega per la regió d'inversió

La càrrega corresponent a la regió d'inversió ve donada per l'eq. (80). En aquesta, apareix el paràmetre d'efecte substrat, que depèn de N_s segons l'eq. (81). N_s és el dopatge del substrat i s'assumeix constant al llarg de tot el canal.

$$Q_I = C_{ox}\left(V_G - V_{FB} - \phi_s - \gamma\sqrt{\phi_s}\right) \tag{80}$$

$$\gamma = \frac{\sqrt{2q\varepsilon_{Si}N_s}}{C_{ox}} \tag{81}$$

Per una tensió donada corresponent a la regió d'inversió, Q_I és pràcticament una funció lineal de Φ_S. En aquesta situació, Q_I pot expressar-se com:

$$Q_I \cong n_q C_{ox}\left(\phi_s - \phi_{s0}\right) \tag{82}$$

on Φ_{s0} és el valor de Φ_s pel qual Q_I val 0. La seva equació és:

$$\phi_{s0} = V_G - V_{FB} + \gamma_s\left(\frac{\gamma_s}{2} - \sqrt{\left(\frac{\gamma_2}{2}\right)^2 + V_G - V_{FB}}\right) \tag{83}$$

D'altra banda, n_q es defineix com:

$$n_q = 1 + \frac{\gamma_s}{2\sqrt{\phi_s}} \tag{84}$$

7.2. Corrent de *drain*

En aquest model, el corrent de *drain* ve donat per l'eq. (85):

$$I_D = \mu_n W\left(-Q_I \frac{dQ_S}{dx} + V_t \frac{dQ_I}{dx}\right) \tag{85}$$

$$I_D = \mu_n W(-Q_I)\frac{dV_x}{dx} \tag{86}$$

on μ_n és la mobilitat dels portadors, que s'assumeix constant al llarg del canal, W és l'amplada del canal, Q_I és la càrrega corresponent a la regió d'inversió, Φ_S és el potencial de superfície i V_t és el voltatge tèrmic. D'altra banda, V_x és el potencial que es té per cada diferencial del canal. Val V_S al terminal de *source* i V_D al terminal de *drain*.

Gràcies a la linealització anterior manifestada en l'eq. (83), la derivada del potencial de superfície es pot escriure com:

$$\frac{d\phi_s}{dx} = \frac{1}{n_q C_{ox}} \frac{dQ_I}{dx} \tag{87}$$

i permet expressar el corrent de *drain* com una funció de la càrrega per la regió d'inversió, obtenint:

$$I_D = 2n\mu_n C_{ox} \frac{W}{L}\left(\frac{K_B T}{q}\right)^2 \left[\left\{\ln\left[1+\exp\left(\frac{V_p - V_s}{2K_B T/q}\right)\right]\right\}^2 - \right.$$

$$\left. - \left\{\ln\left[1+\exp\left(\frac{V_p - V_D}{2K_B T/q}\right)\right]\right\}^2\right] \tag{88}$$

on

$$V_p = \frac{V_G - V_T}{n} \tag{89}$$

és el voltatge de saturació que hauria de ser aplicat al canal per anul·lar l'efecte del voltatge de porta. El terme n correspon al *body effect*:

$$n = \frac{1}{1 - \dfrac{\gamma}{2\sqrt{V_G - V_T + \left(\dfrac{\gamma}{2} + \sqrt{2\phi_p}\right)^2}}} \tag{90}$$

on

$$\gamma = \frac{\sqrt{2q\varepsilon_{Si}N_A}}{C_{ox}} \tag{91}$$

$$V_{T0} = \phi_{ms} + 2\phi_p + \frac{\sqrt{4q\varepsilon_{Si}N_A\phi_p}}{C_{ox}} \tag{92}$$

8. CONCLUSIONS

Cadascun dels diferents models descrits al llarg d'aquest article és utilitzat en diferents situacions. Així, a causa de la seva senzillesa, *Level 1* s'empra quan l'exactitud és menys important que el temps de simulació. Aquest model no inclou efectes de vora, cosa que *Level 2* sí que té en compte. *Level 3*, que va ser especialment creat per tenir en compte efectes de canal curt, presenta una exactitud similar a la del seu predecessor, però s'executa molt més ràpidament. D'altra banda, els models BSIM3v3 i BSIM4 són els més estesos gràcies a la seva exactitud. Pel que fa al model EKV, encara que aquest és molt exacte per la regió subllindar, només funciona apropiadament per dispositius de canal llarg. La seva principal aplicació són els dissenys de baixa potència.

En l'extracció de paràmetres, el model EKV és molt més ràpid que BSIM3v3 o BSIM4. Això no només es deu al fet que BSIM3v3 està format per un elevat nombre de paràmetres, sinó també a la gran correlació que aquests presenten entre si. El model EKV és el més fàcil d'utilitzar perquè aconsegueix una gran exactitud amb un reduït nombre de paràmetres.

9. REFERÈNCIES

[Sischka] F. Sischka, "Modelling tutorial. Characterization Handbook", DMLTUTR.doc.

[Massobrio96] G. Massobrio, P. Antonietti, "Semiconductor modeling with SPICE", McGraw-Hill, 2a edició, 1996.

[Algelov04] G. Algelov, T. Takov, S. Ristic, "Mosfet models at the edge of 100nm sizes", 24th International conference on microelectronics, vol 1, Sèrbia i Montenegro, 2004.

[Saijets] J. Saijets, M. Aberg, M. Andersson, "CMOS RF Models for Mobile Radio IC Design", Nokia Research

Group, Finlàndia.
[Zhou] X. Zhou, B. Chiah, “XSIM/DOUST: A compact model for design and optimization of ultra small transistors”, http://ntu.edu.sg/home/exzhou/Research/DOUST/.
[Cheng] I. Cheng, M. Chan, K. Hui et al., “BSIM3v3 manual”, Department of Electrical Engineering and Computer Sciences, UCLA, Berkeley, Califòrnia.
[Xi] X. Xi, M. Dunga, J. He, W. Liu et al., “BSIM4 Mosfet model”, Department of Electrical Engineering and Computer Sciences, UCLA, Berkeley, Califòrnia.
[Klein] P. Klein, S. Chaldek, “A New Model for Implanted Quarter Micron n-Mosfets and below”, Universitat de Munich.
[Entz] C. Entz, M. Butcher, A. Porret et al., “The foundations of EKV MOS transistor Charge-Based-Model”, Swiss Federal Institute of technology, Lausanne.
[Vittoz] I. Vittoz, S. Enz, F. Krummenacher, “A Basic of MOS transistors and its Circuits Implications”, Swiss Federal Institute of technology, Lausanne.
[Steer] M. B. Steer, C. Christoffersen, S. Velu, N. Kripani, “Global Modelling of Microwave Circuits”.
[Altet00] J. Altet, X. Aragonés, J. Gonzalez, D. Mateo, F. Moll, A. Rubio, “Disseny de circuits i sistemes integrats”, Edicions UPC, 2000.

A2. Escalat tecnològic.

***Abstract*—**En aquest article es farà una breu introducció al model teòric en el qual se sustenta l'escalat tecnològic i posteriorment es parlarà dels diferents tipus d'escalat existents segons el factor d'escala utilitzat. També es veuran els principals avantatges i inconvenients que presenta aquesta tècnica, tot fent especial esment del fenomen de canal curt. Finalment, es consideraran algunes de les solucions proposades per millorar l'escalat en un futur immediat.

***Index Terms*—**CMOS, *threshold*, *latch-up*, *bulk*, SOI, implantació, *scaling*, zona de càrrega espacial, longitud de Debye, *bandgap*, capacitat paràsita, *hot carrier*, canal curt, *pinch-off*, llei de Moore.

1. INTRODUCCIÓ

El concepte escalat tecnològic fa referència a la reducció de les dimensions del MOSFET i a d'altres característiques VLSI. La reducció de les dimensions d'aquest transistor ha sigut espectacular durant les últimes tres dècades, en les quals la longitud de porta ha passat dels 10μm el 1970 als 0.15μm de l'any 2000 i als 45nm del 2008. No obstant això, la reducció de la longitud i de l'amplada de porta en l'escalat del MOSFET no és suficient, sinó que també és necessària una reducció de totes les magnituds, incloent el gruix de l'òxid de porta i l'amplada de les capes de deplexió. Aquesta darrera, a la vegada, implica l'escalat de la densitat de dopatge del substrat.

Durant els anys 70 i 80 la tecnologia de l'escalat es basava en reduir les dimensions del MOSFET, principalment l'amplada, la longitud del canal i el gruix de l'òxid de porta. Mitjançant aquesta tècnica, que continua essent vigent avui en dia, s'arribava a dos tipus de transistors, els de canal ample i els de canal estret.

Actualment, es busquen altres formes d'escalat com la reducció de la tensió d'alimentació fins a un límit proper a 1V, cosa que implicaria la disminució de la tensió llindar fins a unes poques desenes de mV. També es busca canviar l'òxid de silici de la porta per altres aliatges amb una permitivitat major, com el tàntal, tot mantenint el mateix gruix.

Al final, l'objectiu de totes les tendències és aconseguir millores substancials com són l'augment de la densitat de transistors, l'increment de la freqüència de treball o la disminució de la potència dissipada.

2. MODEL TEÒRIC

La disminució de mida del transistor no només afecta els elements geomètrics, sinó també altres variables com la tensió aplicada o la concentració d'impureses. La base teòrica es formula considerant l'equació de Poisson i de continuïtat de corrent.

$$\frac{\partial^2\phi}{\partial x^2}+\frac{\partial^2\phi}{\partial y^2}+\frac{\partial^2\phi}{\partial z^2}=-\frac{q}{\varepsilon_s}\left(N_D-N_A+p-n\right) \quad (1)$$

$$\nabla\left(-q\mu_n n\nabla\phi+qD_n\nabla n\right)=0 \quad (2)$$

En condicions inferiors a V_{TH}, la concentració d'electrons és menyspreable, de manera que es realitza l'estudi oblidant l'equació de continuïtat de corrent. Es considera el següent canvi de variables.

$$\phi'=\phi/k \quad (3)$$

$$\left(x',y',z'\right)=\left(x,y,z\right)/\alpha \quad (4)$$

$$\left(n',p',N'_A,N'_D\right)=\frac{\left(n,p,N_A,N_D\right)}{k}\lambda^2 \quad (5)$$

Fent α=λ, l'equació de Poisson queda:

$$\frac{\partial^2 k\phi'}{\partial(\lambda x')^2}+\frac{\partial^2 k\phi'}{\partial(\lambda y')^2}+\frac{\partial^2 k\phi'}{\partial(\lambda z')^2}=-\frac{q}{\varepsilon_s}\left(N'_D-N'_A+p'-n'\right)\frac{k}{\lambda^2} \quad (6)$$

que fàcilment es redueix a:

$$\frac{\partial^2\phi'}{\partial x'^2}+\frac{\partial^2\phi'}{\partial y'^2}+\frac{\partial^2\phi'}{\partial z'^2}=-\frac{q}{\varepsilon_s}\left(N'_D-N'_A+p'-n'\right) \quad (7)$$

Així, s'obté novament l'equació de Poisson, però aquesta vegada per unes variables escalades sota els canvis indicats anteriorment. Si les condicions de contorn (potencial de font, drenador, substrat i porta) es redueixen proporcionalment al factor k, la solució de l'equació de Poisson difereix de la solució de l'equació escalada de Poisson només en un factor d'escala. En aquesta situació, la forma del camp elèctric és la mateixa en ambdós dispositius, presentant doncs el mateix comportament. En canvi, el que sí que varia és la intensitat del camp elèctric amb el factor ε=λ/k. Aquest increment de la intensitat del

camp com a resultat de l'escalat suposa un problema, atès que implica un deteriorament del dispositiu. Amb l'objectiu de disminuir aquest impacte, s'han portat a terme nombroses investigacions, els detalls de les quals són esmentats al següent apartat.

3. TIPUS D'ESCALAT

3.1. Factor d'escala amb camp elèctric constant

El primer d'ells, es basa en mantenir el camp elèctric constant. Per aquest cas, es té que ε=λ/k és igual a la unitat. Així doncs, λ=k i fent α=λ, s'obté k=α. Si es reprenen els canvis de variable assenyalats, es troba pel potencial:

$$\phi' = \phi / \alpha \tag{8}$$

Cosa que implica que per la tensió aplicada:

$$V' = V / \alpha \tag{9}$$

És el tipus d'escalat ideal, on totes les dimensions del dispositiu es redueixen en un factor α. Presenta l'inconvenient que no és fàcil de realitzar.

3.2. Factor d'escala generalitzat

La segona solució passa per tenir un factor d'escala generalitzat. En aquest cas, no es té cap tipus de restricció pel que fa a l'escalat, ja que no hi ha l'obligació de mantenir ε=1. Això facilita que els paràmetres físics puguin ser escalats d'una manera més selectiva. Per contra, el camp elèctric es veu incrementat en λ/k (o en α/k). Aquí el voltatge passa a ser:

$$V' = V / k = V \cdot \frac{\varepsilon}{\lambda} = V \cdot \frac{\varepsilon}{\alpha} \tag{10}$$

Pel cas particular en què k=1, és a dir, quan ε=α, es té que V'=V. Aquest és l'escalat per tensió constant, recurs que és bastant utilitzat ja que manté la compatibilitat en tensió entre les diferents tecnologies i generacions de xips, encara que els problemes associats a l'increment de ε fan que l'escenari més comú sigui el d'un compromís intermig entre V i E constants. Mentre que al cas anterior es té un dopatge incrementat en α, en aquest cas s'incrementa en εα, cosa que comporta problemes associats al desgast del dispositiu.

3.3. Factor d'escala generalitzat selectiu

Aquest darrer model encara és més genèric que l'anterior. En els dos casos anteriors, la geometria disminuïa en un factor α, indistintament de la variable espacial. Existia doncs un lligam geomètric on $\alpha_x=\alpha_y=\alpha_z$. Ara, però, es distingeix entre α_d i α_w, que fan referència a la reducció d'alçada i d'amplada, respectivament, del dispositiu. D'aquesta manera, s'aconsegueix controlar de forma independent un major nombre de paràmetres físics. Permet també modificar la resistència del dispositiu, paràmetre que no era modificable sota la condició de geometria restrictiva. Així, la modificació de RC no depèn exclusivament de la capacitat.

A la següent taula es poden veure alguns dels paràmetres escalats.

Paràmetre físic	$\vec{E}$ ctt	Generalitzat	Selectiu
x_j	1/α	1/α	$1/\alpha_d$
t_{ox}	1/α	1/α	$1/\alpha_d$
W i L	1/α	1/α	$1/\alpha_w$
E	1	ε	ε
V	1/α	ε/α	ε/α_d
Dopatge	α	$\varepsilon\cdot\alpha$	$\varepsilon\cdot\alpha_d$
Àrea	$1/\alpha^2$	$1/\alpha^2$	$1/\alpha^2{}_w$
C	1/α	1/α	$1/\alpha_w$
Retard	1/α	1/α	$1/\alpha_d$
P	$1/\alpha^2$	ε^2/α^2	$\varepsilon^2/\alpha_w\cdot\alpha_d$
Densitat de P	1	ε^2	$\varepsilon^2\cdot\alpha_w/\alpha_d$

Taula 1 Paràmetres físics per cadascun dels tres casos.

Una de les solucions adoptades en els últims anys per modificar la resistència ha sigut la substitució de l'alumini pel coure com a metall de contacte, donada la seva menor resistivitat (fins a un 40%). Aquest tipus de solucions són de difícil aplicació, doncs necessiten d'una tecnologia totalment nova.

4. AVANTATGES

Des de la construcció del primer transistor MOS, s'ha reduït la mida d'aquest en gairebé tres ordres de magnitud. Les motivacions són diverses. En primer lloc, s'ha de tenir en compte que aquesta reducció no és exclusiva de la tecnologia CMOS, sinó que altres tecnologies com la TTL també se'n beneficien. Així doncs, els avantatges de la reducció de mida són comuns en diferents tecnologies.

Històricament, el procés de reducció s'ha centrat en l'aplicació a la fabricació CMOS, ja que ha estat, i segueix sent, la tecnologia més utilitzada. Les tècniques desenvolupades, però, són en la seva majoria de fàcil aplicació a altres processos tecnològics, pel que es pot tractar l'escalat dels dispositius CMOS com un procediment genèric. A continuació, es descriuen els avantatges de la reducció de mida dels transistors.

4.1. Augment de la densitat de transistors

És el que també permet la disminució del cost de producció, ja que aquest va associat a la superfície total utilitzada pel CI. És a dir, permet augmentar la complexitat de la lògica sense augmentar l'àrea. A la Fig. 1 es mostra com es redueix la mida d'un transistor pel model d'escalat de camp elèctric constant. L'àrea efectiva d'aquest transistor es veu reduïda en $1/\alpha^2$.

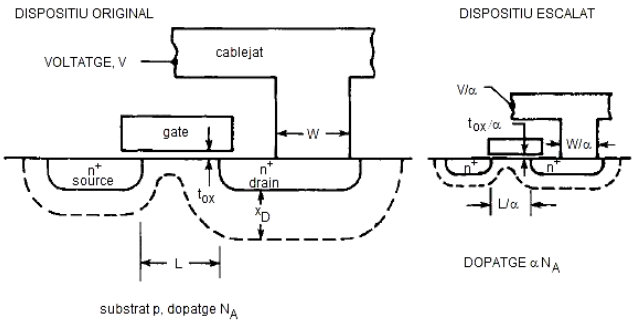


Fig. 1 Principis d'escalat amb camp elèctric constant per transistors MOS i circuits integrats.

4.2. Increment de la freqüència de treball

Quan es redueix la mida del transistor MOS en una determinada escala, també s'està disminuint la longitud entre font i drenador. Donat que es disminueix la tensió aplicada, també s'escurça la zona de càrrega espacial de la font i la capacitat es redueix proporcionalment al factor d'escalat. Pel que fa a la resistència, aquesta es manté pràcticament igual, ja que els components resistius no només disminueixen longitudinalment sinó que també ho fan en la seva amplitud. Es té, doncs, que el valor RC disminueix linealment en un factor g de manera que R'C'=RC/g. D'aquesta manera, la freqüència es veu incrementada en f'=g·f.

Una de les estratègies per seguir augmentant la freqüència mitjançant la reducció de mida és fer un escalat asimètric respecte la longitud i l'amplada. Així, es té que R'≠R.

4.3. Disminució de la tensió aplicada

Aquest és un altre dels avantatges de la reducció de mida dels dispositius. En disminuir la mida aplicant el model de camp elèctric constant, es redueix en el mateix factor la tensió aplicada. Atès que el consum respon a l'expressió:

$$P = KCV^2 f \qquad (11)$$

on K és un coeficient de proporcionalitat, es té que la disminució de la tensió aplicada incideix quadràticament en la disminució de la potència dissipada.

La disminució de la tensió aplicada és un altre dels objectius de l'escalat, ja que permet dissenyar electrònica de baix consum. Actualment i en funció dels resultats que es volen obtenir, existeixen dues opcions pel que fa a l'escalat: augment de freqüència o disminució de consum. Es dissenyen dispositius amb diferents criteris d'escalat, que permeten optimitzar un paràmetre o altre.

5. LIMITACIONS I EFECTES DE CANAL CURT

En aquest apartat es tracten els principis físics i tècnics que limiten el model d'escalat teòric.

5.1. Quantitats físiques no escalables

El model teòric d'escalat no pot ser aplicat amb total llibertat als paràmetres longitud de Debye, dopatge i tensió *threshold*. L'explicació teòrica d'aquest fenomen es presenta tot seguit.

La longitud de Debye és un paràmetre que presenta una sèrie de problemes, tot i que inferiors als d'altres magnituds, que incideixen en l'amplada de la zona de deplexió. Aquesta no decreix al mateix ritme que la resta del dispositiu i això fa que els portadors recombinin a un ritme diferent al que marca l'escalat, donant un increment de la concentració. Com a conseqüència, s'incrementa L_D.

El diagrama de bandes del silici és intrínsec, de manera que no varia excepte quan ho fa el nivell de Fermi. Si s'augmenta la concentració tenint un *bandgap* constant, a causa del potencial de junció tan elevat que apareix, es té efecte túnel i també trencament Zener. Aquest fenomen limita l'aplicabilitat de la teoria d'escalat per alts dopatges.

El pendent *subthreshold*, finalment, ve determinat per l'estadística termodinàmica dels portadors (Fermi-Dirac o Maxwell-Boltzmann), de manera que també és independent de l'escalat del dispositiu. Com a conseqüència, la tensió *threshold* no es pot reduir proporcionalment.

5.2. El problema de les interconnexions

Un problema degut a la reducció de mida és el del paràmetre RC associat a les interconnexions. Per freqüències superiors als 500MHz, aquest paràmetre indueix l'aparició d'un retard als dispositius que va en augment a mesura que s'incrementa la freqüència. La solució a aquest problema passa pel disseny de connexions mitjançant una jerarquia de nivells o la utilització de materials com el coure.

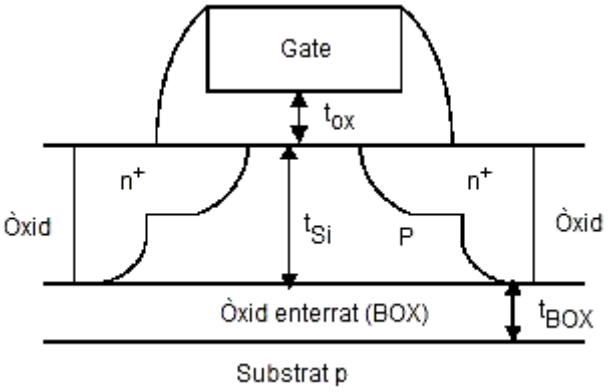


Fig. 2 Esquemàtic d'una secció transversal d'un dispositiu SOI.

5.3. Capacitats paràsites

Un altre dels problemes associats a la disminució de mida és l'aparició de capacitats paràsites i fenòmens com el *latch-up*. Una de les solucions implementades de tipus estructural és el SOI (*silicon on insulator*), que substitueix

l'estructura del *bulk* i permet un major control de les capacitats creant barreres de protecció (veure Fig. 2). L'avantatge del SOI sobre altres solucions similars és la senzilla implantació en el procés de fabricació.

5.4. *Hot carriers*

A mesura que els dispositius MOSFET van disminuint de mida, la necessitat de preservar el camp elèctric implica un augment en la concentració del dopatge. Aquesta concentració fa que el camp intrínsec a causa de la ZCE sigui molt elevat a les juncions dels transistors, propiciant així l'aparició de portadors altament energètics. Descrits mitjançant un model de temperatura efectiva, aquests passen a ser anomenats portadors calents o *hot carriers*. Els efectes derivats d'aquests portadors són el deteriorament del dispositiu a mesura que l'escalat va en augment, introduint noves limitacions en el seu disseny així com altres efectes paràsits.

5.5. Efectes de canal curt

Una de les conseqüències directes del *scaling* és el distanciament de les condicions d'idealitat que permeten suposar el model de canal llarg sota el qual està dissenyat el transistor. Aquest model és substituït pel de canal curt, que es basa en el fet que l'amplada de la zona de deplexió és del mateix ordre de magnitud que la longitud del canal. La mínima longitud del canal per evitar problemes és:

$$L_{\min} = 0.4\left[x_j \cdot t_{ox} \cdot (W_S + W_D)^2\right]^{1/3} \qquad (12)$$

on x_j és la profunditat de la junció pn, t_{ox} és l'alçada de l'òxid i W és l'amplada de la zona de deplexió o ZCE entre font i drenador.

Una primera conseqüència de tenir canal curt és que el corrent de drenador s'incrementa més enllà del voltatge de *pinch-off*, cosa que entra en contrast amb la característica I(V) pròpia de canal llarg, on s'arriba a la saturació. A més, aquest corrent no es fa zero per potencial de porta nul i potencial de drenador elevat, pel que la porta perd la capacitat de tancar el dispositiu. Com que aquest és un comportament semblant al del BJT, aquesta tensió V_{DS} s'anomena tensió Early.

Una segona conseqüència de tenir canal curt és el comportament del règim *subthreshold*. Per potencials pròxims al de *threshold*, la característica és la que s'ha vist fins ara, on per un cert voltatge V_{GS} no s'obtenia corrent a la sortida perquè desapareixia el canal de conducció. Ara, per potencials petits, el perfil comença a variar i el corrent de drenador ja no es pot fer zero. Existeix una contribució del potencial de difusió que no s'havia considerat fins ara. Si el canal encara es fa més petit, arriba un instant en què V_{GS} no controla el corrent. És a partir d'aquesta longitud que el transistor no pot ser utilitzat com a interruptor.

La tercera característica és el canvi de valor de la tensió *threshold*. Pel model de canal llarg el valor *threshold* és independent de la longitud del canal, però pel cas de canal curt aquest potencial decreix amb la longitud amb una forta caiguda per a L<2µm.

6. FUTURES SOLUCIONS

La situació actual de l'escalat permet preveure la continuïtat de la tecnologia CMOS en els pròxims anys. Malgrat això, limitacions com les esmentades a l'apartat anterior, que augmenten progressivament, així com l'aparició de nous fenòmens a mesura que es redueix la mida dels transistors, fa preveure un final a mitjà termini que ningú és capaç d'estimar amb certesa.

A la Fig. 3 es poden veure les diferents fronteres amb les quals la tecnologia hauria d'enfrontar-se en els pròxims anys si es continua aplicant la llei de Moore.

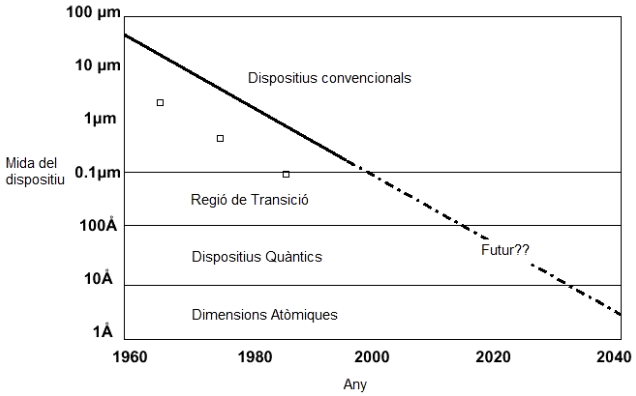


Fig. 3 Escalat dels dispositius en funció del temps segons la llei de Moore.

Algunes de les solucions proposades pel futur immediat passen per l'ús de diferents materials que permetrien un augment de les possibilitats dels dispositius, com per exemple el que es presenta a la Fig. 4, on es troba un SOI CMOS en el qual s'utilitza tant silici com germani.

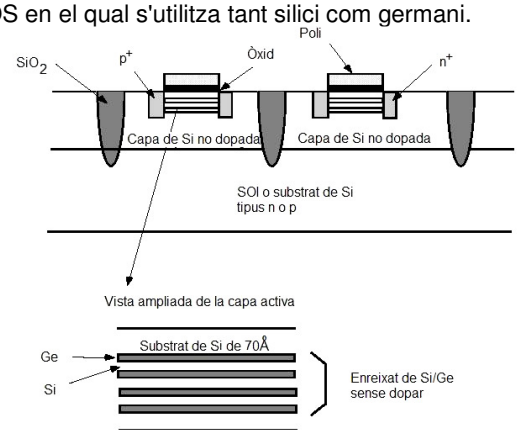


Fig. 4 Transistor CMOS de SiGe.

Una altra de les solucions proposades es basa en el disseny de dispositius en configuració vertical, tal com es

mostra a la Fig. 5. Aquesta configuració permet una densitat de dispositius 4 vegades superior i una menor longitud de canal, així com un major control d'aquest i una conductància més gran. Al mateix temps obre les portes a una altra de les tecnologies en vies de desenvolupament, com és el de circuits integrats tridimensionals, on els transistors queden col·locats l'un sobre l'altre en diferents plànols, augmentant de manera considerable la capacitat de connexionat dels mateixos, així com la seva integrabilitat.

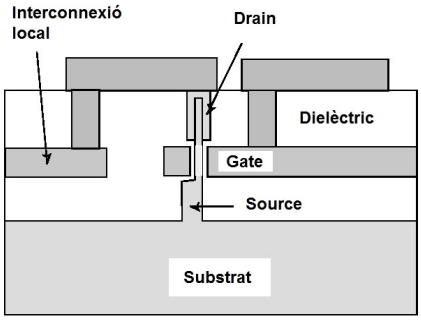


Fig. 5 Disseny d'un transistor CMOS vertical.

7. CONCLUSIONS

L'escalat tecnològic consisteix en la reducció de totes les magnituds del MOSFET, incloent el gruix de l'òxid de porta i l'amplada de les capes de deplexió.

Segons el tipus de factor d'escala utilitzat, es distingeixen tres tipus d'escalat. El primer d'ells és l'escalat amb camp elèctric constant, on totes les dimensions del dispositiu es redueixen en el mateix factor. Constitueix l'escalat ideal. El segon tipus d'escalat és el que utilitza factor d'escala generalitzat, en què l'escalat no està restringit però presenta l'inconvenient que el camp elèctric es veu incrementat. Per últim, es té l'escalat generalitzat selectiu, que és el menys restrictiu i permet escalar diferents variables mitjançant diferents escalats.

L'escalat tecnològic presenta un conjunt d'avantatges, d'entre els quals destaca la disminució del cost com a conseqüència de l'augment de la densitat de transistors, l'augment de la freqüència de treball donat que el producte RC disminueix i la disminució de la tensió aplicada si s'utilitza el model de camp elèctric constant. Aquest últim fenomen implica la disminucó quadràtica de la potència consumida. D'altra banda, la tècnica de l'escalat es veu limitada pel fet que no totes les magnituds físiques són escalables, l'aparició de retards per freqüències superiors a 500MHz i l'existència de capacitats paràsites i portadors altament energètics que provoquen el deteriorament prematur del dispositiu.

Una altra conseqüència de l'escalat és el fenomen de canal curt, que ocorre quan l'amplada de la zona de deplexió és del mateix ordre de magnitud que la longitud del canal. En aquesta situació, el corrent de drenador s'incrementa més enllà del voltatge de *pinch-off*, cosa que implica l'aparició d'una tensió Early. A més, en anar disminuint el canal, arriba un moment en què V_{GS} no controla I_D i el transistor no pot ser utilitzat com a interruptor. Un últim inconvenient d'aquest fenomen és que el potencial *threshold* no és independent de la longitud.

Finalment, tot i que actualment s'utilitzen estructures SOI i dissenys en configuració vertical per tal de pal·liar els inconvenients de l'escalat tecnològic, ja es preveu un límit en l'evolució d'aquest.

8. REFERÈNCIES

[Chang00] C. I. Chang, S. M. Sze, "ULSI Devices", Wiley-Interscience, 1a edició, 2000.

[Wang] K. L. Wang, "Issues of CMOS scaling - On and off the roadmap", arxiu pdf de la Universitat de Califòrnia, Los Angeles.

[Dabari95] B. Dabari, R. H. Dennard, G. G. Shahidi, "CMOS scaling for high performance and low power", Proceedings of the IEEE, vol 83, núm 4, abril 1995.

[Mistry00] K. Mistry, T. Ghani, M. Armstrong, S. Tyagi, P. Packan, S. Thomson, S. Yu, M. Bohr, "Scalability Revisited: 100nm PD-SOI transistors and implications for 50nm devices", arxiu pdf de Portland technology development, INTEL Corporation, 2000.

[Das] K. K. Das, "SOI overwiev on the CMOS scaling: technology and circuit design issues", arxiu pdf de la Universitat de Michigan.

[Gonzalez97] R. Gonzalez, B. M. Gordon, M. A. Horowitz, "Supply and threshold voltage for low power CMOS", IEEE Journal for solid-state circuits, vol 32, núm 8, agost 1997.

[Kim02] S. D. Kim, C. M. Park, J. C. S. Woo, "Advanced Model and analysis of Series Resistance for CMOS scaling into nanometer regime - Part I: Theoretical derivation", IEEE transactions on electron devices, vol 49, núm 3, març 2002.

[Moore 65] G. I. Moore, "Cramming more components onto integrated circuits", Electronics, vol 38, núm 8, abril 1965.

[Allan02] A. Allan, D. Edenfeld, W. H. Jorney Jr., A. B. Khang, M. Rodgers, Y. Zoriant “2001 technology Roadmap For Semiconductors”, IEEE 2002.

[Shimohigashi93] K. Shimohigashi, K. Seiki, “Low Voltage ULSI design”, IEEE Journal for solid-state circuits, vol 28, núm 4, abril 1993.

A3. Tècniques de connexionat i empaquetat dels CI.

***Abstract*—**En aquest article es descriurà el principal procés de muntatge i empaquetat d'un circuit integrat. Seguidament, s'explicaran els diferents mètodes que existeixen per tal de realitzar el connexionat elèctric. Per últim, es farà esment dels diferents tipus d'encapsulat que poden ser utilitzats per protegir el xip.

***Index Terms*—**Oblia, *die*, *pad*, *leadframe*, connexionat, *wire bonding*, *taped automated bonding*, *flip chip*, encapsulat.

1. INTRODUCCIÓ

Els CIs de silici en forma de *die*, encara que sempre porten una capa protectora, són fràgils i difícils de manejar. A més, els diminuts *pads* o contactes dificulten la seva connexió. En conseqüència, per protegir el *die* i facilitar les connexions, els CIs són empaquetats.

Aquí s'explicarà el procés de muntatge d'un circuit integrat, així com els principals mètodes de connexionat que existeixen.

2. PROCÉS DE MUNTATGE I EMPAQUETAT

La seqüència d'operacions pel procés d'empaquetat general és la que es descriu tot seguit.

Després del testejat de l'oblia, aquesta es col·loca en una estructura adhesiva que la subjecta per la part del darrere, és a dir, per la cara on no hi ha el circuit imprès. A l'espai que es troba entre cadascun dels *dies* que formen l'oblia no hi ha cap circuiteria marcada. D'aquesta manera, es té un camí per tal que un autòmat d'alta velocitat amb un diamant molt fi pugui serrar l'oblia tot separant els *dies*. L'estructura adhesiva serveix per subjectar individualment els *dies* al seu lloc després del tall amb la serra.

Cada *die* bo de l'oblia és retirat de l'estructura adhesiva per una màquina de *pick and place* i col·locat en un marc de metall o *leadframe*. Arribat aquest punt, es realitza el connexionat del CI seguint qualsevol dels mètodes explicats a l'apartat següent, tot i que cal remarcar que el mètode escollit dependrà del tipus d'encapsulat que s'utilitzi i del cost del conjunt. Aquestes connexions elèctriques uneixen cada *pad* del *die* amb el corresponent pin del *leadframe*.

Un cop s'han realitzat les connexions, ja es pot procedir a encapsular el circuit. Moltes vegades, però, aquest es recobreix abans amb una capa de silicona o una làmina d'alumini per augmentar-ne la protecció. A més, també és convenient netejar tota la superfície del xip de possibles restes de soldadures o impureses. Tot seguit, té lloc l'encapsulat, que té com a principal objectiu protegir el CI de possibles agressions externes durant el seu temps de vida. S'aconsegueix modelant una àrea rectangular d'epoxi negre al voltant del *die* i del *leadframe*, tot deixant sobresortir els pins d'aquest darrer.

Finalment, una eina mecànica agafa cada CI embalat individualment, tot doblegant els pins fins la seva configuració final. Les unitats són marcades amb el nom del fabricant, sèrie identificativa o data sobre el plàstic d'epoxi per tal de reconèixer-les.

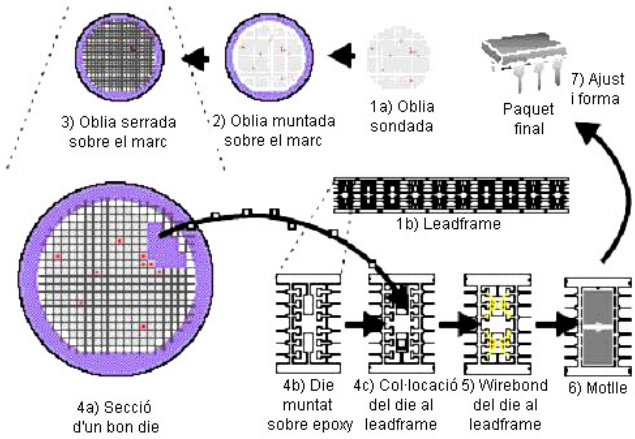


Fig. 1 Procés d'embalatge de plàstic.

3. CONNEXIONAT DEL CI

La principal finalitat de l'encapsulat és proporcionar la connexió elèctrica entre els terminals d'E/S del xip i la placa de circuit imprès. Hi ha, per tant, tota una estructura d'interconnexió a l'encapsulat, a la qual el xip es connecta mitjançant diversos mètodes. Tot seguit, es veuran les diferents formes de connexió del xip.

3.1. *Wire bonding*

Aquest és el mètode més utilitzat en el qual s'utilitza un cable fi (de 0.2 a 0.4 mm), típicament d'or o d'alumini, tot i que el coure ha guanyat una atenció considerable a causa del seu baix preu i la seva forta resistència, per unir el *pad* del xip al *pad* de l'encapsulat. Aquest procés, que té lloc

després de l'adhesió del propi xip al substrat de l'encapsulat, pot realitzar-se mitjançant tècniques de termocompressió, soldat ultrasònic o termosònic. La connexió es realitza seqüencialment per tots els *pads* del xip.

La unió per termocompressió va ser la primera que van presentar els laboratoris Bell el 1957 i va ser àmpliament usada fins a l'aparició de la tecnologia d'unió ultrasònica a mitjans dels anys 60. La unió per termocompressió requereix exercir una gran força sobre la superfície a altes temperatures (prop dels 300ºC). El material del cable forçosament ha de ser d'or, però el *pad* pot ser d'or o d'alumini. La tècnica d'unió per ultrasons, en canvi, és la tecnologia més usada per la unió de cables d'alumini i es porta a terme a temperatura ambient. Finalment, la tècnica termosònica s'utilitza en cables d'or i actualment comprèn aproximadament el 90% de totes les unions de cable. Es porta a terme a una temperatura d'entre 100 i 240ºC.

A la taula següent es mostra una comparació dels tres tipus de tecnologies d'unions de cable.

Wire bonding	**T (ºC)**	**Materials**	**Categoria**
Termocompressió	300-500	Au, Al-Au	Energètic
Ultrasons	25	Au, Al, Al-Au	Pressió baixa
Termosons	100-240	Au, Cu, Al-Au	Pressió baixa

Taula 1 Comparació de les tècniques de *wire bonding*.

En el passat, una gran proporció dels fracassos en el camp dels semiconductors i dispositius electrònics va ser causada per les unions de cable, tot i que gràcies als avenços recents, la fiabilitat del *wire bonding* està augmentant. Malgrat això, actualment encara ocorren un cert nombre de fracassos. Les causes més comunes d'aquests són la ruptura d'un *pad* d'unió, la fractura del cable i les soldadures pobres o emplaçades incorrectament que provoquen unions febles o curtes.

Aquests fracassos han de ser avaluats, tot seguint un criteri que varia en funció dels requisits de l'aplicació. Els mètodes d'avaluació poden trobar-se en diversos estàndards publicats, però el més comú és el MIL-STD-883. Aquest llista diferents mètodes de test, com són proves d'estirament de la unió, acceleració contínua, vibració aleatòria, commoció mecànica, model d'estabilització o resistència a la humitat.

3.2. *Taped automated bonding* (TAB)

Aquesta tècnica es basa en la fabricació d'una làmina de polímer, similar a una pel·lícula fotogràfica, sobre la qual s'imprimeixen línies de coure mitjançant un procés de dipositació. La connexió o soldat es realitza al mateix temps per tots els *pads* del xip. Posteriorment es realitza pels de l'encapsulat. Això significa que es requereix la fabricació específica de les interconnexions de la làmina per un disseny concret. D'altra banda, aquest mètode permet una distància menor entre *pads* i minimitza les distàncies de les connexions.

3.3. *Controlled collapse chip connection (flip chip)*

Els pins relativament grans degraden els senyals d'alta freqüència. Recentment, s'han desenvolupat noves tècniques per produir paquets més petits i, per tant, permetre connexions que puguin treballar amb senyals de freqüències molt elevades. Les tècniques més noves es basen en referir l'embalatge a l'escala del xip, és a dir, que el paquet tot just sigui una mica més gran que el xip de silici. Una d'aquestes tècniques és coneguda com *flip chip*.

Als *pads* del xip, que poden estar distribuïts per tota la superfície i no només a la perifèria, es col·loca una bola de soldadura. Posteriorment, es gira el xip tot deixant la part superior encarada al substrat de l'encapsulat i seguidament se solden totes les connexions a la vegada. A més de la rapidesa del mètode i independentment del nombre de terminals d'E/S, la densitat d'interconnexió és molt elevada i la longitud de les interconnexions, la mínima possible. D'aquesta manera, es garanteix que els components paràsits seran molt petits. Cal destacar que els productes de *flip chip* apropiadament muntats presenten una fiabilitat elevada.

Addicionalment, en la unió per cable, els *pads* han d'estar a la vora del CI per minimitzar la longitud del mateix cable. En bastants casos, però, el nombre de *pads* requerit per les connexions elèctriques excedeix el nombre que cap a la perifèria del CI. En aquest cas, la mida del CI ha de ser incrementada precisament per complaure el número de *pads*, incrementant així el cost. Amb la tècnica de *flip chip*, els *pads* poden ser posats en matriu en qualsevol lloc sobre el CI, cosa que minimitza el cost.

4. TIPUS D'ENCAPSULAT

A causa de l'increment de la velocitat en els circuits integrats, així com també dels marges de soroll potencialment reduïts, el disseny de l'encapsulat ha de ser tingut en compte amb molta cura. Pel que fa a aquest, existeixen moltes alternatives. La selecció d'una o altra depèn del nombre de terminals, la forma i els materials (plàstic o ceràmic) de l'encapsulat, el tipus de muntatge en placa o el disseny de la refrigeració.

Existeixen diferents denominacions comercials pels encapsulats com són DIL (*Dual In Line*), que té un màxim de 48 terminals i la distància entre ells és de 2.5mm, SOIC (*Small Outline Integrated Circuit*), que permet distàncies menors entre terminals contigus, LCC (*Leadless Chip Carrier*) i QFP (*Quad Flat Package*), que evolucionen del SOIC per aprofitar els quatre costats de l'encapsulat, augmentant així el nombre de terminals per unitat d'àrea. Finalment, es tenen els encapsulats PGA (*Pin Grid Array*) i BGA (*Ball Grid Array*), en els quals s'aprofita tota la superfície per treure terminals.

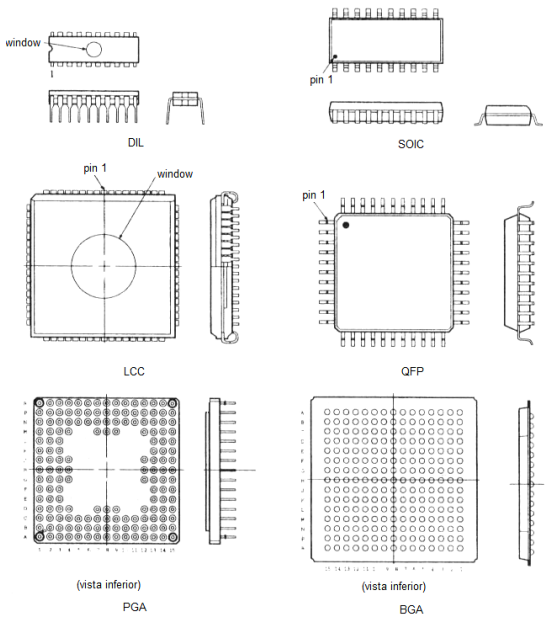


Fig. 2 Diferents tipus d'encapsulats comercials més utilitzats.

5. CONCLUSIONS

Després de ser impresos sobre l'oblia, els diferents circuits integrats han de ser separats, connectats elèctricament amb el *leadframe* i finalment, encapsulats.

Avui dia, el connexionat dels CI es realitza mitjançant tres mètodes diferents: *wire bonding*, *taped automated bonding* i *flip chip*. El darrer presenta avantatges importants pel que fa a la mida i qualitat del senyal a altes freqüències en comparació amb el *wire bonding*, a més de presentar una robustesa mecànica major. Però també resulta més car per xips de poques connexions, pel que actualment se segueix utilitzant *wire bonding* en aquelles aplicacions en les quals l'estalvi de l'empaquetatge no és significatiu i la freqüència de treball no és crítica. La tècnica *atuomated taped bonding* es caracteritza per minimitzar les distàncies entre *pads*.

Finalment, es vol destacar que existeixen molts tipus d'encapsulats diferents. L'elecció d'un o altre va en funció de les característiques del circuit integrat.

6. REFERÈNCIES

[Tummala88] R. Tummala, E. Rymaszewski, "Microelectronics Packaging Handbook", 1988.

[Sze88] S. M. Sze, "VLSI Technology", McGraw-Hill, 2a edició, 1988.

[Harmon97] G. Harmon, "Wire Bonding in Microelectronics: Materials, Processes, Reliability and Yield", McGraw Hill, 2a edició, 1997.

[Lau00] J. H. Lau, "Low Cost Flip Chip Technologies for DCA, WLCSP and PBGA Assemblies", McGraw-Hill, 2000.

[Rubio00] A. Rubio, J. Altet, X. Aragonés, J. L. González, D. Mateo, F. Moll, "Diseño de circuitos integrados", Edicions UPC, 2000.

[01/10/07] http://www.flipchips.com/SMErev01.htm

[01/10/07] http://www.empf.org/empfasis/oct03/703mmicprog.htm

[01/10/07] http://www.twi.co.uk/

[01/10/07] http://nepp.nasa.gov/wirebond/

4. Inversor CMOS. Característiques estàtiques.

***Abstract*—**L'estudi de les característiques de les portes lògiques bàsiques en tecnologia CMOS ha estat indispensable pel desenvolupament i millora del disseny VLSI. L'inversor CMOS, que és la porta més senzilla, és el punt de partida per l'anàlisi de portes més complexes. En aquest article es recullen les idees sobre les característiques estàtiques de l'esmentat dispositiu i com aquestes en determinen el seu disseny físic.

***Index Terms*—**Característiques estàtiques, consum, CTV, disseny, inversor CMOS, inversor simètric, lògica CMOS, NOT, porta inversora.

1. INTRODUCCIÓ

La porta lògica inversora és una de les portes més simples que existeixen i és el dispositiu que implementa la funció de negació NOT a nivell microelectrònic. Existeixen diferents tipus d'inversors, entre ells el CMOS.

Aquest inversor es caracteritza per incloure en la seva estructura dos transistors MOSFET, un de canal n i un altre de canal p, de tal manera que amb la mateixa entrada de porta comuna s'aconsegueix que treballin de forma complementària. Les característiques d'aquests transistors, tal i com s'explica més endavant, determinen el disseny d'aquest dispositiu.

Els circuits CMOS normalment utilitzen un pou n sobre substrat p per tal d'obtenir el transistor pMOS del parell de transistors que constitueix la cel·la bàsica CMOS. No obstant, és possible trobar tecnologies de pou p i, molt habitualment, de doble pou.

2. LÒGICA DE L'INVERSOR CMOS

La Taula 1 mostra la taula de la veritat necessària per implementar un inversor lògic.

Input	*Output*
0	1
1	0

Taula 1 Taula de la veritat d'un inversor.

Si s'examina aquesta taula es pot veure que quan hi ha un '0' a l'entrada, apareix un '1' a la sortida. Això suggereix un *p-switch* o interruptor p connectat entre una font de valor '1' (V_{DD}) i la sortida. Quan hi ha un '1' a l'entrada, és un '0' el que s'ha de connectar a la sortida. Això suggereix un *n-switch* o interruptor n entre la sortida i una font de valor '0' (GND). S'ha de notar que el transistor nMOS només es necessita per posar un '0' a la sortida, mentre que el transistor pMOS funciona pel cas contrari. L'esquema a nivell de transistors i el símbol lògic de la porta es mostren a la Fig. 1. En general, doncs, una porta CMOS de lògica complementària sempre disposa d'un interruptor n connectat a '0' (GND) i d'un interruptor p connectat a '1' (V_{DD}).

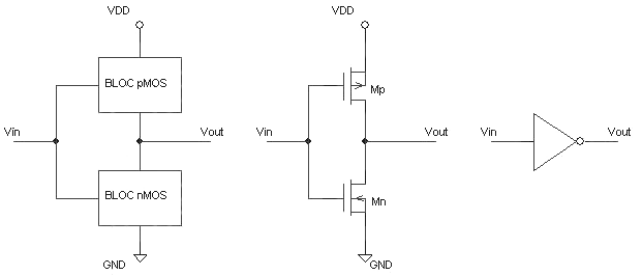


Fig. 1 Inversor CMOS.

S'ha de destacar, però, el fet que per un inversor CMOS es té la mateixa entrada per tots els transistors. Això és el que porta altra vegada a la Taula 1, perquè aquesta representa els únics estats possibles de sortida de la porta inversora sense deixar lloc a possibles estats indeterminats, ja que quan un transistor està en conducció, l'altre està en tall i viceversa.

3. CARACTERITZACIÓ D'UN INVERSOR CMOS EN ESTÀTIC

3.1. Corba de transferència de voltatge (CTV)

La corba de transferència de voltatge (CTV) és una representació gràfica del canvi en la tensió de sortida quan es modifica la tensió d'entrada. Ve caracteritzada per l'estat de conducció en què es trobin els transistors que formen la porta lògica.

La corba presentada a la Fig. 3 és la CTV per un inversor típic. S'observa que s'ha assenyalat en aquesta gràfica una sèrie de paràmetres. Són V_{OH} o senyal de tensió de sortida alt, V_{OL} o senyal de tensió de sortida baix, V_{IL} o màxim senyal d'entrada que és interpretat com un '0' lògic, V_{IH} o mínim senyal d'entrada que és interpretat com un '1' lògic i, finalment, V_{th} o tensió de commutació. Per un inversor

CMOS, $V_{OH}=V_{DD}$ i $V_{OL}=0$, tal com es demostrarà més endavant.

D'altra banda, també poden definir-se els marges de soroll. NM_H és el marge de tensió que pot tenir un '0' lògic; NM_L, en canvi, és el marge de tensió que pot tenir un '1' lògic. Les seves expressions són les que figuren tot seguit.

$$NM_H = V_{OH} - V_{IH} \quad (1)$$

$$NM_L = V_{IL} - V_{OL} \quad (2)$$

3.2. CTV de l'inversor CMOS

L'inversor CMOS bàsic es mostra a la Fig. 2. La tipologia del circuit és tal que les portes dels dos transistors estan connectades formant una única entrada de valor:

$$V_{in} = V_{GSn} = V_{DD} - \left|V_{GSp}\right| \quad (3)$$

mentre que els dos drenadors dels transistors p i n estan al mateix punt, que constitueix el voltatge de sortida donat per:

$$V_{out} = V_{DSn} = V_{DD} - \left|V_{DSp}\right| \quad (4)$$

La complementarietat d'aquesta porta es pot veure reflectida a l'eq. (3). Mentre $V_{in}=V_{GSn}$ augmenta, $|V_{GSp}|$ decreix i viceversa. Per aquest inversor el concepte de *driver* i *load*, com poden tenir altres inversors (p.e. l'inversor nMOS), no existeix ja que V_{in} controla la porta d'ambdós MOSFETs. Notar que no existeixen efectes de substrat, ja que $V_{BSn}=|V_{BSp}|=0$.

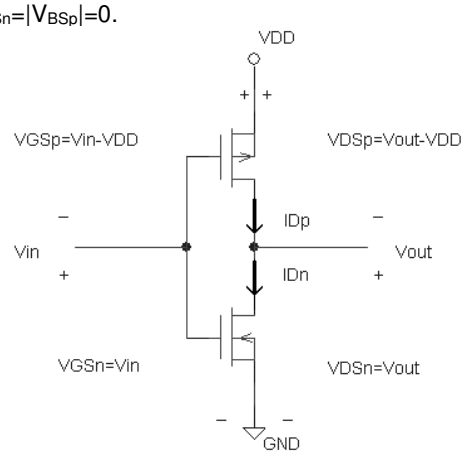


Fig. 2 Esquema a nivell de transistor d'un inversor CMOS, indicant els corrents i els voltatges de polarització.

Per tal de construir la CTV d'un inversor CMOS, es comença amb el voltatge d'entrada $V_{in}<V_{Tn}$. En aquesta situació el transistor nMOS està en tall i per tant $I_{Dn}=I_{Dp}=0$. Tanmateix, el transistor pMOS té la seva capa d'inversió formada ja que $|V_{GSp}|$ és alt. Aleshores, pel règim lineal es pot escriure:

$$I_D = (\beta_p/2)\left(\left|V_{GSp}\right| - \left|V_{Tp}\right|\right)\left|V_{DSp}\right| \quad (5)$$

Quan $I_D=0$, forçosament ha de passar que $|V_{DSp}|=0$ o bé que $|V_{GSp}|-|V_{Tp}|=0$, tal com es dedueix de l'equació anterior. Donat que la darrera opció no és vàlida, es té que $|V_{DSp}|=0$. Això implica que:

$$V_{out} = V_{OH} = V_{DD} \quad (6)$$

Quan V_{in} augmenta a un nivell lleugerament superior a V_{Tn}, el transistor nMOS entra a la regió de saturació ($V_{DSn}\geq V_{GSn}-V_{Tn}$) i el pMOS, a la regió de conducció lineal ($|V_{DSp}|<|V_{GSp}|-|V_{Tp}|$). Fent $I_{Dn}=I_{Dp}$, es troba l'equació corresponent a aquest tram de la CTV:

$$\beta_n (V_{in} - V_{Tn})^2 = \beta_p \left[2\left(V_{DD} - V_{in} - \left|V_{Tp}\right|\right)\left(V_{DD} - V_{out}\right) - \left(V_{DD} - V_{out}\right)^2\right] \quad (7)$$

on s'han utilitzat les equacions (3) i (4) combinades amb l'expressió de I_{Dp} (veure Taula 2). Si V_{in} continua creixent, V_{out} continua comportant-se segons l'eq. (7) fins que:

$$\left|V_{DSp}\right| \geq \left|V_{GSp}\right| - \left|V_{Tp}\right| \quad (8)$$

$$\left|V_{out} - V_{DD}\right| \geq \left|V_{in} - V_{DD}\right| - \left|V_{Tp}\right| \quad (9)$$

se satisfà. És quan es dóna aquesta condició que ambdós transistors estan en saturació. Per això, al tram central de la CTV s'obté que:

$$\beta_n (V_{in} - V_{Tn})^2 = \beta_p \left(V_{DD} - V_{in} - \left|V_{Tp}\right|\right)^2 \quad (10)$$

Aquesta equació prediu una caiguda vertical de la CTV. Quan V_{out} cau a un nivell on:

$$V_{out} < V_{in} - \left|V_{Tp}\right| \quad (11)$$

el transistor pMOS entra en una situació de saturació, mentre que el nMOS passa al règim lineal. Aleshores, fent $I_{Dn}=I_{Dp}$, l'equació per la CTV queda:

$$\beta_n \left[2(V_{in} - V_{Tn})V_{out} - V_{out}^2\right] = \beta_p \left(V_{DD} - V_{in} - \left|V_{Tp}\right|\right)^2 \quad (12)$$

Finalment, quan V_{in} augmenta fins al punt on:

$$\left|V_{in} - V_{DD}\right| < \left|V_{Tp}\right| \quad (13)$$

el transistor pMOS entra en tall, de manera que $I_{Dn}=I_{Dp}=0$.

El transitor nMOS encara conserva la capa d'inversió (canal format), per tant:

$$V_{out} = V_{OL} = V_{DSn} = 0V \tag{14}$$

Una CTV típica obtinguda a partir de les equacions (3)-(14) es mostra a la Fig. 3.

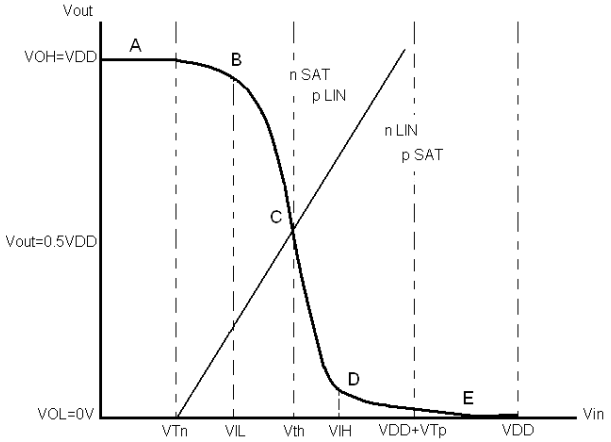


Fig. 3 Representació de la CTV segons els estats dels transistors per un inversor CMOS.

Per completitud, es mostren les Taules 2 i 3 com a resum de les diferents situacions de l'inversor.

pMOS		
tall	$\|V_{GSp}\|<\|V_{Tp}\|$ $\|V_{in}-V_{DD}\|<\|V_{Tp}\|$	
lineal	$\|V_{GSp}\|>\|V_{Tp}\|$ $\|V_{in}-V_{DD}\|>\|V_{Tp}\|$	$\|V_{DSp}\|<\|V_{GSp}\|-\|V_{Tp}\|$ $\|V_{out}-V_{DD}\|<\|V_{in}-V_{DD}\|-\|V_{Tp}\|$
saturació	$\|V_{GSp}\|>\|V_{Tp}\|$ $\|V_{in}-V_{DD}\|>\|V_{Tp}\|$	$\|V_{DSp}\|\geq\|V_{GSp}\|-\|V_{Tp}\|$ $\|V_{out}-V_{DD}\|\geq\|V_{in}-V_{DD}\|-\|V_{Tp}\|$

Taula 2 Resum de situacions pel pMOS de l'inversor.

nMOS		
tall	$V_{GSn}<V_{Tn}$ $V_{in}<V_{Tn}$	
lineal	$V_{GSn}>V_{Tn}$ $V_{in}>V_{Tn}$	$V_{DSn}<V_{GSn}-V_{Tn}$ $V_{out}<V_{in}-V_{Tn}$
saturació	$V_{GSn}>V_{Tn}$ $V_{in}>V_{Tn}$	$V_{DSn}\geq V_{GSn}-V_{Tn}$ $V_{out}\geq V_{in}-V_{Tn}$

Taula 3 Resum de situacions pel nMOS de l'inversor.

3.3. Consum

De la representació gràfica de la Fig. 3 s'extreu que a les regions on es compleix $V_{in}<V_{Tn}$ o $|V_{in}-V_{DD}|<|V_{Tp}|$ (regions A i E de la Fig. 3), no flueix càrrega entre V_{DD} i terra. Com que aquestes dues zones corresponen a '0' i '1' lògics a l'entrada, respectivament, un circuit CMOS només admet consum degut a corrents de fuites quan els *inputs* són estables. La pèrdua d'energia o circulació de corrent significativa es dóna a les etapes de canvi de valor d'entrada o estats de transició (regions B, C i D). Concretament, el punt de màxima dissipació de potència correspon al punt de commutació de l'inversor. A la Fig. 4 es modelitzen aquestes regions de transició per millorar la comprensió del seu comportament.

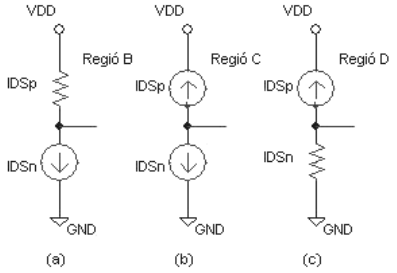


Fig. 4 Circuits equivalents per les regions de transició d'un inversor CMOS.

3.4. Anàlisi de paràmetres característics

Amb $V_{in}<V_{Tn}$ el transistor nMOS està en tall mentre que el pMOS té la capa d'inversió creada. Aleshores, igual que a (3):

$$V_{OH} = V_{DD} \tag{15}$$

Quan $|V_{in}-V_{DD}|<|V_{Tp}|$, el pMOS està en tall mentre que el nMOS té format el canal que connecta la sortida cap a terra. Aleshores:

$$V_{OL} = 0V \tag{16}$$

L'entrada de baixa tensió V_{IL} es produeix quan el transistor nMOS està saturat i el pMOS està en conducció i no saturat. En aquest punt de la CTV, la derivada de la gràfica és -1. Fent $I_{Dn}=I_{Dp}$ i derivant a banda i banda de la igualtat, s'obté:

$$\left(\frac{\partial I_{Dn}}{\partial V_{in}}\right)dV_{in} + \left(\frac{\partial I_{Dn}}{\partial V_{out}}\right)dV_{out} = \left(\frac{\partial I_{Dp}}{\partial V_{in}}\right)dV_{in} + \left(\frac{\partial I_{Dp}}{\partial V_{out}}\right)dV_{out} \tag{17}$$

Tenint en compte que la derivada de I_{Dn} respecte V_{out} és nul·la i reagrupant termes, finalment es té:

$$\frac{dV_{out}}{dV_{in}} = \frac{\left(\partial I_{Dn}/\partial V_{in}\right) - \left(\partial I_{Dp}/\partial V_{in}\right)}{\partial I_{Dp}/\partial V_{out}} = -1 \tag{18}$$

Avaluant les derivades que apareixen a l'expressió anterior s'arriba a la següent equació per V_{IL}:

$$V_{IL}\left(1+\beta_n/\beta_p\right) = 2V_{out} + \left(\beta_n/\beta_p\right)V_{Tn} - V_{DD} - \left|V_{Tp}\right| \tag{19}$$

La combinació de les equacions (7) i (19) dóna una solució quadràtica per V_{IL}. Quan $V_{in}=V_{IH}$, el transistor nMOS

està en situació de no saturació mentre que el pMOS està saturat. De manera anàloga al cas anterior, en aquest punt de la CTV el valor del pendent de la gràfica és -1.

$$\frac{dV_{out}}{dV_{in}} = \frac{\left(\partial I_{Dp}/\partial V_{in}\right) - \left(\partial I_{Dn}/\partial V_{in}\right)}{\partial I_{Dn}/\partial V_{out}} = -1 \qquad (20)$$

Avaluant les derivades a la regió de treball corresponent de cada transistor i substituint V_{in} per V_{IH}, s'obté l'equació per V_{IH}:

$$V_{IH}\left(1+\beta_n/\beta_p\right) = 2V_{out} + V_{Tn} + \left(\beta_p/\beta_n\right)\left(V_{DD} - \left|V_{Tp}\right|\right) \qquad (21)$$

El punt de commutació per l'inversor ($V_{th}=V_{in}=V_{out}$) ve donat quan ambdós transistors estan en saturació. Estant en aquesta regió, si es té en compte l'eq. (10), s'arriba a la següent expressió:

$$V_{th}\left(1+\sqrt{\beta_n/\beta_p}\right) = V_{Tn} + \sqrt{\beta_n/\beta_p}\cdot\left(V_{DD} - \left|V_{Tp}\right|\right) \qquad (22)$$

que dóna la intersecció de la CTV amb la línia de guany unitari $V_{in}=V_{out}$.

4. DISSENY D'UN INVERSOR CMOS

Una primera aproximació pel disseny de l'inversor CMOS consisteix en determinar el punt de commutació donat per l'eq. (22). Concretament, els paràmetres geomètrics que caracteritzen el disseny estan continguts en les variables específiques β_n i β_p a través de:

$$\beta_p/\beta_n = \mu_p\left(W/L\right)_p / \mu_n\left(W/L\right)_n \qquad (23)$$

Aquí s'assumeix que C_{ox} és igual pels dos transistors, ja que els òxids de porta han crescut al mateix temps. El valor requerit per aquest factor per tal d'establir el punt de commutació de l'inversor s'obté de:

$$\sqrt{\beta_p/\beta_n} = \left(V_{th} - V_{Tn}\right) / \left(V_{DD} - V_{th} - \left|V_{Tp}\right|\right) \qquad (24)$$

En aquest punt és interessant remarcar que un inversor simètric té $V_{th}=V_{DD}/2$. Utilitzant aquest valor a l'eq. (22), es fixen els paràmetres físics segons:

$$\sqrt{\beta_p/\beta_n} = \left(V_{DD}/2 - V_{Tn}\right) / \left(V_{DD}/2 - \left|V_{Tp}\right|\right) \qquad (25)$$

Per un inversor simètric també es té que $V_{Tn}=|V_{Tp}|$ i, per tant, $\beta_n=\beta_p$. A partir d'aquesta darrera expressió, s'obté:

$$\left(W/L\right)_p / \left(W/L\right)_n = \mu_n/\mu_p \qquad (26)$$

Per una tecnologia amb $\mu_n/\mu_p\approx 3$, es té que $(W/L)_n\approx 1$ i $(W/L)_p\approx 3$ per una àrea mínima. A partir de la Fig. 5, es veu com per un inversor simètric ($\beta_n/\beta_p=1$) es té una CTV completament simètrica respecte $V_{DD}/2$.

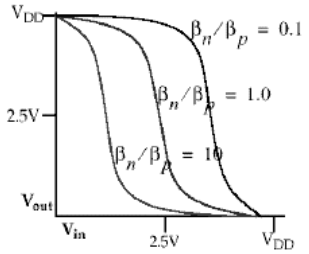


Fig. 5 Influència del factor β_n/β_p per les característiques de l'inversor.

Es pot trobar V_{IL} reduint l'eq. (19) a la forma:

$$V_{out} = V_{IL} + V_{DD}/2 \qquad (27)$$

vàlida per aquest cas; reutilitzant-la juntament amb l'eq. (7) per V_{IL} i arreglant els resultats en l'equació lineal s'arriba a:

$$V_{IL} = \left(V_{Tn} + 3V_{DD}/2\right)/4 \qquad (28)$$

El voltatge d'entrada de nivell alt es troba d'una manera similar:

$$V_{IH} = \left(5V_{DD}/2 - V_{Tn}\right)/4 \qquad (29)$$

Si se sumen les equacions (28) i (29) es troba la relació abastament útil:

$$V_{IL} + V_{IH} = V_{DD} \qquad (30)$$

Aquesta equació demostra que la forma de la CTV d'un inversor és invariant respecte el valor que s'escull per V_{DD}. Aleshores, aquests circuits poden treballar amb un ampli rang de voltatges d'alt nivell. El mínim valor absolut teòric de V_{DD} és $(V_{Tn}+|V_{Tp}|)=2V_{Tn}$. Tanmateix, $V_{DD}=3V_{Tn}$ és un valor més pràctic si es vol tendir a una transició d'amplada zero (i.e. $V_{IL}\cong V_{IH}$).

En el disseny d'aquest dispositiu és interessant calcular els marges de soroll. En el cas simètric vénen donats segons:

$$NM_H = NM_L = V_{IL} = \left(V_{Tn} + 3V_{DD}/2\right)/4 \qquad (31)$$

D'aquí es veu que incrementant V_{DD} es millora la

resposta del sistema davant la influència de soroll, tot i que aleshores augmenta la potència de consum del circuit.

Si es fixa V_{DD} però s'altera la condició de simetria $\beta_n=\beta_p$, l'inversor augmenta el seu marge de soroll a nivell alt NM_H si el factor β_n/β_p també augmenta. De manera anàloga, si el factor disminueix, el marge de soroll a nivell baix NM_L és el que augmenta. Això es pot entendre també veient que el punt de commutació és més gran.

5. EXEMPLES DE DISSENY

5.1. Inversor CMOS simètric

Utilitzant la condició de simetria $\beta_n=\beta_p$, juntament amb els valors de tall dels transistors $V_{Tn}=|V_{Tp}|=0.8V$ fixats i una tensió d'alimentació $V_{DD}=5V$, es pot dissenyar un inversor CMOS d'àrea mínima amb el corresponent càlcul de paràmetres.

$$V_{IL}=\frac{1}{4}(0.8+15/2)=2.075V \quad (32)$$

$$V_{IH}=\frac{1}{4}(25/2-0.8)=2.925V \quad (33)$$

Prenent el punt de commutació a $V_{DD}/2$, es té:

$$V_{th}=\frac{0.8+1(5-0.8)}{2}=2.5V \quad (34)$$

Es pot comprovar que aquest disseny compleix $V_{IL}+V_{IH}=V_{DD}$ i que té uns marges de soroll $NM_L=NM_H=V_{IL}$.

5.2. Inversor CMOS no simètric

En aquest cas s'agafa una relació de $\beta_n/\beta_p=2.5$ amb uns valors de tall $V_{Tn}=1V$ i $V_{Tp}=-0.9V$ donats pel procés, i una tensió d'alimentació $V_{DD}=5V$. Aleshores, substituint V_{IL} per V_{in} a (7), es té:

$$2.5(V_{IL}-1)^2=\left[2(4.1-V_{IL})(5-V_{out})-(5-V_{out})^2\right] \quad (35)$$

mentre que de l'eq. (14) s'obté $V_{out}=1.75V_{IL}+1.7$, que si se substitueix, s'arriba a una equació quadràtica per V_{IL}:

$$V_{IL}^2+2.133V_{IL}-6.628=0 \quad (36)$$

de solució:

$$V_{IL}=1.72V \quad (37)$$

$$V_{out}=4.71V \quad (38)$$

Anàlogament, substituint V_{in} per V_{IH} a (12) i utilitzant l'eq. (21), es té:

$$2.5\left[2(V_{IH}-1)V_{out}-V_{out}^2=\left[16.81\right]-8.2V_{IH}+V_{IH}^2\right] \quad (39)$$

$$V_{out}=0.7V_{IH}-1.32 \quad (40)$$

Amb les expressions anteriors s'arriba a la solució quadràtica de resultat aproximat:

$$V_{IH}\approx 2.48V \quad (41)$$

Finalment, el punt de commutació per aquest disseny té el següent valor:

$$V_{th}=\frac{1+\sqrt{0.4}(5-0.9)}{1+\sqrt{0.4}}\cong 2.2V \quad (42)$$

6. CONCLUSIONS

L'inversor CMOS és el dispositiu que implementa la funció de negació NOT. Aquest està format per dos transistors MOSFET, un de canal n, que s'encarrega de connectar la sortida amb GND quan a l'entrada hi ha un '1' lògic, i un altre de canal p, que connecta la sortida amb V_{DD} quan es té un '0' lògic a l'entrada. Convé remarcar que l'entrada de l'inversor és la mateixa pels dos transistors.

La corba de transferència de voltatge (CTV) d'una porta lògica ve caracteritzada per l'estat de conducció en què es trobin els dos transistors que el formen. Es distingeixen diferents punts importants V_{OH}, V_{OL}, V_{IL} i V_{IH}. En el cas de l'inversor realitzat amb tecnologia CMOS, V_{OH} coincideix amb V_{DD} i V_{OL} amb 0V. Els altres dos valors, en canvi, depenen de la tecnologia emprada. Per determinar-los, s'han d'igualar les expressions pel corrent de *drain* dels transistors que constitueixen l'inversor. Per obtenir V_{IL}, només cal tenir en compte que el transistor nMOS es troba en règim de saturació i el pMOS, a la regió lineal. Per V_{IH} la situació és just al revés, és a dir, el transistor nMOS es troba en règim lineal i el pMOS, en saturació. Cal considerar que el pendent de la CTV en V_{IL} i V_{IH} es fa igual a -1.

Al tram central de la CTV ambdós transistors es troben en saturació. Igualant les seves expressions pel corrent de *drain* es troba l'expressió per la tensió de commutació, que es produeix quan $V_{in}=V_{out}=V_{th}$.

7. REFERÈNCIES

[Uyemura88] J. P. Uyemura, “Fundamantals of MOS Digital Integrated Circuits. A Systems Perspective”, Addison-Wesley, 1988.

[Weste93] W. E. Weste, K. Eshranghian, “Principles of CMOS VLSI design”, Addison-Wesley, 2a edició, 1993.

[Glasser88] L. A. Glasser, D. W. Dobberpuhl, “The Design and Analysis of VLSI Circuits”, Addison-Wesley, 1988.

[Bota97] S. Bota, J. Carrabina, A. Herms, “Introducció al disseny CMOS VLSI”, Textos docents núm 52, Edicions de la Universitat de Barcelona, 1997.

[Bracho99] S. Bracho del Pino, “La Ingeniería Microelectrónica ante el cambio del Milenio”, Servicios Publicitarios de la Universidad de Cantabria, 1999.

[Rubio00] A. Rubio, J. Altet, X. Aragonés, J. L. González, D. Mateo, F. Moll, “Diseño de circuitos y sistemas integrados”, Edicions UPC, 2000.

5. Inversor CMOS. Característiques dinàmiques.

***Abstract*—**En aquest article es presenta l'anàlisi dinàmic de l'inversor CMOS. Per tant, s'introdueixen les capacitats paràsites de l'inversor i es dedueixen les expressions que permeten calcular el temps de propagació, el temps de commutació i el temps de retard. L'anàlisi és el punt de partida per l'estudi de portes més complexes.

***Index Terms*—**Capacitats paràsites, capacitats de superposició, capacitats de porta, capacitats d'unió, capacitat de cablejat, commutació de l'inversor, temps de propagació, temps de commutació, retard.

1. INTRODUCCIÓ

L'inversor CMOS és un dels circuits més senzills possibles. El seu estudi detallat permet entendre fàcilment el disseny de portes més complexes.

Aquí s'analitzaran les característiques dinàmiques de l'inversor CMOS. Per tal de fer aquest estudi és necessari entendre les capacitats que intervenen en la commutació de l'inversor.

2. PARÀMETRES CARACTERÍSTICS EN LA COMMUTACIÓ D'UN INVERSOR

En aquest apartat es defineixen els principals paràmetres temporals que intervenen en la commutació d'un inversor. Aquests paràmetres són el temps de propagació i el temps de commutació, que poden veure's representats gràficament a la Fig. 1. També es definirà el temps de retard.

El retard d'una porta, el que es coneix com a temps de propagació (t_p), és el promig dels temps de propagació en la commutació a la sortida de nivell baix a alt i d'alt a baix. El temps de propagació de pujada (t_{pLH}) és el retard existent entre $V_{1/2}$ del senyal d'entrada i $V_{1/2}$ del senyal de sortida quan aquesta commuta de V_{OL} a V_{OH}. El temps de propagació de baixada (t_{pHL}), en canvi, és el retard entre $V_{1/2}$ del senyal d'entrada i $V_{1/2}$ del senyal de sortida quan aquesta commuta de V_{OH} a V_{OL}.

Els temps de commutació es defineixen en CMOS entre el 10 i el 90% del senyal de sortida. Així doncs, el temps de commutació de pujada (t_r o t_{LH}) és el temps que triga la sortida en passar de V_0 a V_1 quan l'entrada commuta de V_{OH} a V_{OL}. El temps de commutació de baixada (t_f o t_{HL}), en canvi, és el temps que triga la sortida en passar de V_1 a V_0 quan l'entrada commuta de V_{OL} a V_{OH}.

Els nivells de tensió que es defineixen pel càlcul dels temps descrits són:

$$V_0 = V_{OL} + 0.1\left(V_{OH} - V_{OL}\right) \quad (1)$$

$$V_1 = V_{OL} + 0.9\left(V_{OH} - V_{OL}\right) \quad (2)$$

$$V_{1/2} = V_{OL} + 0.5\left(V_{OH} - V_{OL}\right) \quad (3)$$

Per un inversor CMOS V_{OL} és 0V i V_{OH}, V_{DD}. Així doncs, és evident que els nivells de tensió aquí presentats corresponen al 10, 90 i 50% de l'alimentació de l'inversor.

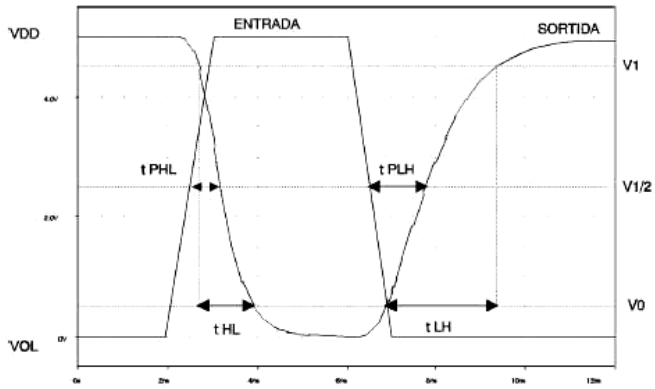


Fig. 1 Commutació de l'inversor.

3. CAPACITATS ASSOCIADES

El càlcul dels temps de propagació i commutació requereix l'anàlisi del corrent a través de la capacitat de la sortida de la porta. Per tant, per calcular aquests temps és necessari conèixer les capacitats que intervenen en la commutació.

Totes les capacitats paràsites existents associades a l'inversor es mostren a la Fig. 2. A efectes pràctics, però, moltes d'aquestes capacitats poden no ser considerades, bé perquè estan connectades a un potencial constant (C_{GSp}, C_{GSn}) o bé perquè són molt petites i poden despreciar-se després de considerar el teorema de Miller (C_{GDn}, C_{GDp}).

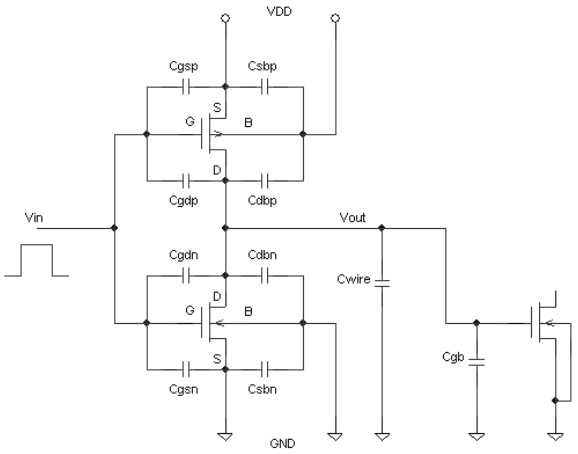


Fig. 2 Inversor amb totes les seves capacitats paràsites.

Les capacitats que sí que s'han de tenir en compte i que constitueixen la capacitat de càrrega de l'inversor són, doncs:

$$C_L = C_{GB} + C_{SBn} + C_{SBp} + C_{DBn} + C_{DBp} + C_{wire} \quad (4)$$

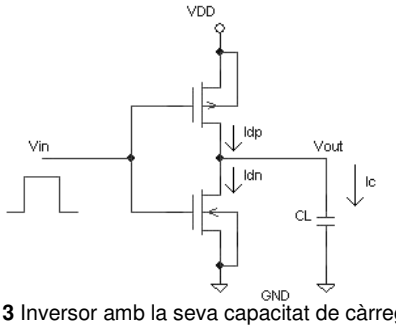


Fig. 3 Inversor amb la seva capacitat de càrrega.

Cal remarcar que la capacitat de *gate-bulk* fa referència a les portes connectades a la sortida. A continuació, s'explica com obtenir cadascuna de les capacitats anteriors.

3.1. Capacitats d'òxid

Les capacitats d'òxid se subdivideixen en dos grups: les de superposició i les de porta. Les capacitats de superposició són les que s'associen al drenador i a la font (C_{OLD}, C_{OLS}) i depenen de la geometria del transistor. La longitud del canal, en realitat, és menor que la longitud de la porta de polisilici a causa de les difusions laterals que poden experimentar aquestes zones. L'expressió per aquestes capacitats és:

$$C_{OLD} = C_{OLS} = L_{OL} \cdot W \cdot C_{ox} \quad (5)$$

on L_{OL} és la longitud d'àrea activa superposada (OL=*overlap*), W és l'amplada del canal i C_{ox}, que equival a ε_{ox}/t_{ox}, és la capacitat de l'òxid. Donat que la mida de l'àrea activa superposada (L_{OL}) és la mateixa per *source* i *drain*, C_{OLD} i C_{OLS} són iguals.

Les capacitats de porta són les associades al contacte de porta del transistor i se subdivideixen en capacitat de porta-substrat (C_{GB}), de porta-font (C_{GS}) i capacitat de porta-drenador (C_{GD}). Totes aquestes capacitats depenen de la regió de treball en què es trobi el transistor, és a dir, de les tensions de polarització, tal com es mostra tot seguit.

Capacitat	Tall	Lineal	Saturació
C_{GB}	$C_{ox}WL$	0	0
C_{GD}	$C_{ox}WL_D$	$0.5C_{ox}WL+C_{ox}WL_D$	$C_{ox}WL_D$
C_{GS}	$C_{ox}WL_S$	$0.5C_{ox}WL+C_{ox}WL_S$	$2/3C_{ox}WL+C_{ox}WL_S$

Taula 1 Capacitats de porta segons la regió de treball.

La capacitat total de porta pot expressar-se com:

$$C_G = C_{OLD} + C_{OLS} + C_{GB} + C_{GS} + C_{GD} \quad (6)$$

Encara que a efectes pràctics de càlcul, es pot considerar que aquesta capacitat pren el valor:

$$C_G = C_{GB} = C_{ox} \cdot W \cdot L \quad (7)$$

3.2. Capacitats d'unió

Són les capacitats que es formen entre *source-bulk* i *drain-bulk* com a conseqüència del fet que les regions de drenador són unions pn polaritzades en inversa. Les capacitats d'unió es deuen a la càrrega que s'acumula a la zona d'empobriment que es forma a la frontera entre les dues regions. Les expressions per aquestes capacitats, que presenten una primera contribució associada a l'àrea de la difusió i una segona associada al perímetre, són les que es mostren tot seguit.

$$C_{SB} = C_{j0} \cdot A + C_{jsw} \cdot P = C_{j0} \cdot L_S \cdot W + C_{jsw} \cdot 2(W + L_S) \quad (8)$$

$$C_{DB} = C_{j0} \cdot A + C_{jdw} \cdot P = C_{j0} \cdot L_D \cdot W + C_{jdw} \cdot 2(W + L_D) \quad (9)$$

$$C_{j0} = \sqrt{\frac{\varepsilon_{Si} \cdot q}{2\phi_B}\left(\frac{N_A N_D}{N_A + N_D}\right)} \quad (10)$$

$$C_{jsw} = C_{jdw} = C_{j0w} \cdot x_j \quad (11)$$

$$C_{j0w} = \sqrt{\frac{q \cdot \varepsilon_{Si}}{2\phi_{0w}}\left(\frac{N_{Aw} N_D}{N_{Aw} + N_D}\right)} \quad (12)$$

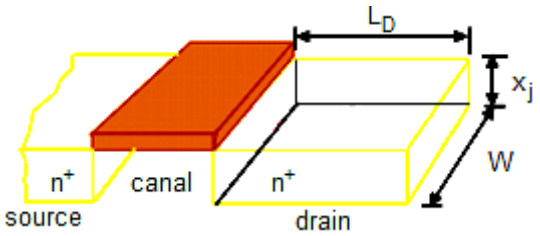


Fig. 4 Regió de *drain* on s'indica el valor de cada dimensió.

Habitualment les longituds de les difusions de *source* i *drain* (L_S, L_D) seran coincidents i, per tant, els valors de C_{SB} i C_{DB} seran idèntics. Els paràmetres C_{j0} i C_{jsw}/C_{jdw} corresponen a la capacitat d'unió per unitat d'àrea i a la capacitat per unitat de longitud associada al perímetre de les regions de *source* i *drain*, respectivament. Aquests paràmetres vénen donats pels *datasheets* de totes les tecnologies. x_j és la profunditat de les difusions.

3.3. Capacitat de cablejat

Depèn de la longitud i amplada dels cables amb què es fan les connexions.

4. COMMUTACIÓ DE L'INVERSOR

4.1. Temps de propagació

Per realitzar el càlcul dels temps de retard, es pren el senyal d'entrada V_{in} com un senyal quadrat ideal on els temps de pujada i baixada són zero (veure Fig. 5). En un inversor CMOS, quan l'entrada commuta de nivell baix a nivell alt, passa el contrari a la sortida, però amb un cert temps de retard, a causa de les capacitats paràsites que s'han comentat anteriorment. Així doncs, caldrà fixar-se en aquell punt en què el senyal d'entrada passa de nivell baix a nivell alt i viceversa.

Convé tenir present que per fer aquests càlculs s'assumeix que les transicions del senyal d'entrada són abruptes, però en realitat aquest senyal canvia gradualment.

4.1.1. Temps de propagació de pujada

Si l'entrada és zero, el transistor pMOS és el que condueix (nMOS en tall) i per això V_{out} està a nivell alt. Però quan V_{in} commuta a nivell alt, llavors és el nMOS el que condueix i C_L es descarrega (s'havia carregat mentre conduïa el pMOS). Es considera que V_{out} ha passat a nivell baix quan s'ha sobrepassat $V_{1/2}$ del senyal de sortida. Durant aquesta descàrrega, però, es poden diferenciar dues zones de funcionament del transistor segons el nivell de tensió: saturació i lineal. Així doncs, si el nivell de tensió es troba entre V_{OH} i V_{OH}-V_{Tn}, el transistor treballa en saturació, i si es troba entre V_{OH}-V_{Tn} i $V_{1/2}$, en zona lineal. El corrent de sortida serà el que passarà a través de la càrrega C_L. La seva expressió és la següent:

$$I_C = C_L \cdot (dV_{out}/dt) \qquad (13)$$

on I_c és la diferència entre I_{Dp} i I_{Dn}, però com que el pMOS està en tall, només quedarà -I_{Dn}. Utilitzant aquesta equació i els valors de tensió límit per cada zona de treball, es pot trobar el temps t_1 (veure Fig. 5) on es passa de saturació a lineal. Una vegada trobat aquest temps, ja es pot calcular t_{pHL}. Integrant l'expressió i aïllant:

$$t_{pHL} = \frac{C_L}{\beta_n (V_{OH} - V_{Tn})}\left[\frac{2V_{Tn}}{(V_{OH} - V_{Tn})} + \ln\left(\frac{4(V_{OH} - V_{Tn})}{(V_{OH} + V_{OL})} - 1\right)\right] \qquad (14)$$

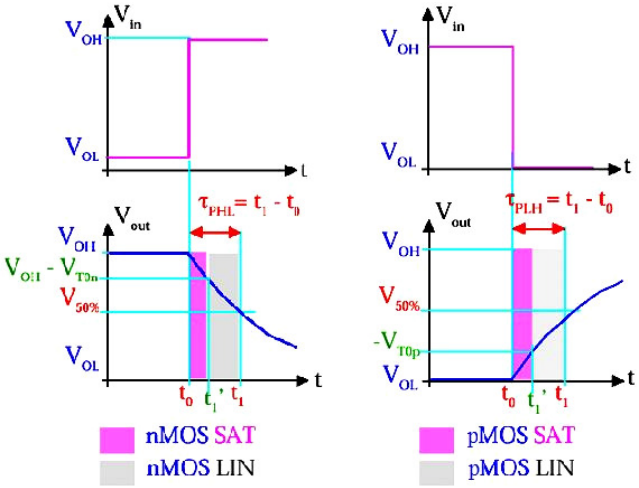


Fig. 5 Càlcul de t_{pHL} i t_{pLH}, respectivament.

4.1.2. Temps de propagació de baixada

Per t_{pLH}, el procés és el mateix però al revés. En aquest cas, V_{in} commuta de nivell alt a nivell baix. Llavors, V_{out} tendirà a V_{DD} des de 0V. Igual que en el cas anterior, fins que V_{out} hagi arribat al valor de $-V_{Tp}$, l'inversor es trobarà a la zona de saturació. I, des d'aquest punt fins que hagi sobrepassat el valor $V_{1/2}$, en zona lineal. Cal recordar que aquest és el cas en què condueix el pMOS i el nMOS està en tall, ja que V_{in} passa a ser 0V.

Seguint el mateix procediment que abans, s'obtindrà la relació per t_{pLH}, només que aquesta vegada, com que condueix el pMOS, I_c serà igual a I_{Dp}. L'equació buscada és la que segueix:

$$t_{pLH} = \frac{C_L}{\beta_p (V_{OH} - |V_{Tp}|)}\left[\frac{2|V_{Tp}|}{(V_{OH} - |V_{Tp}|)} + \ln\left(\frac{4(V_{OH} - |V_{Tp}|)}{(V_{OH} + V_{OL})} - 1\right)\right] \qquad (15)$$

4.2. Temps de commutació

Aquests temps es calculen de la mateixa manera que els temps de propagació, només cal tenir en compte que cal canviar els límits d'integració. Les expressions pels temps de pujada i baixada, respectivament, són les que es

mostren a continuació.

$$t_r = \frac{C_L}{\beta\left(V_1 - |V_{Tp}|\right)}\left[\frac{2|V_{Tp}|}{\left(V_1 - |V_{Tp}|\right)} + \ln\left(\frac{2\left(V_1 - |V_{Tp}|\right)}{V_0} - 1\right)\right] \quad (16)$$

$$t_f = \frac{C_L}{\beta_n\left(V_1 - V_{Tn}\right)}\left[\frac{2V_{Tn}}{\left(V_1 - V_{Tn}\right)} + \ln\left(\frac{2\left(V_1 - V_{Tn}\right)}{V_0} - 1\right)\right] \quad (17)$$

4.3. Retard

El temps de propagació total o retard de l'inversor ve donat per la mitjana aritmètica de t_{pLH} i t_{pHL}.

$$t_p = \frac{t_{pLH} + t_{pHL}}{2} \quad (18)$$

5. TÈCNIQUES DE DISSENY

El temps total de propagació d'un inversor CMOS es pot reduir seguint les indicacions següents.

5.1. Reduir la capacitat de càrrega C_L

Els tres factors més importants que contribueixen a la capacitat de sortida són les capacitats de *gate*, les d'unió i les de cablejat. Un disseny acurat del *layout* ajuda a reduir les capacitats de *gate* i cablejat. Convé que l'àrea de difusió de *drain* i *source* sigui com més petita millor.

5.2. Incrementar la relació W/L dels transistors

Augmentar la relació W/L permet disminuir la resistència del transistor. El retard ve donat pel producte RC i, per tant, aquesta és l'optimització més efectiva que el dissenyador pot dur a terme. De totes maneres, s'ha d'anar amb molt de compte, ja que quan s'augmenta la mida del transistor, també s'incrementa el valor de C_L a causa d'un increment de les capacitats d'unió. De fet, quan aquestes comencen a dominar sobre el *fan-out*, per exemple, augmentar la mida de la porta ja no ajuda a reduir el retard. Simplement fa l'àrea dels transistors més gran, cosa que afecta negativament a la seva velocitat.

5.3. Incrementar V_{DD}

El retard de porta es pot modular modificant la tensió d'alimentació però, a certs nivells, la millora és mínima i aquest canvi hauria de ser evitat.

6. CONCLUSIONS

L'inversor CMOS constitueix una peça bàsica en el disseny de circuits digitals complexos. En l'estudi de les seves característiques dinàmiques cal tenir en compte els diferents tipus de capacitats paràsites que existeixen: les d'òxid (que se subdivideixen en capacitats de superposició i de porta), les d'unió i les de cablejat. El conjunt de totes aquestes capacitats forma el que s'anomena capacitat de càrrega C_L.

La commutació de l'inversor depèn de la capacitat de carregar/descarregar o de la mida/resistència dels transistors de càrrega/descàrrega. Finalment, val a dir que tot el que s'ha vist per l'inversor és extrapolable a qualsevol porta CMOS complexa.

7. REFERÈNCIES

[Rabaey02] J. M. Rabaey, A. Chandrakasan, "Digital Integrated Circuits", University of California, 2a edició, 2002.

[Bota97] S. Bota, J. Carrabina, A. Herms, "Introducció al disseny CMOS VLSI", Textos docents núm 52, Edicions de la Universitat de Barcelona, 1997.

[Prat94] Ll. Prat, R. Bragós, J. A. Chávez, M. Fernández, V. Jiménez, J. Madrenas, E. Navarro, J. Salazar, "Circuitos y dispositivos electrónicos. Fundamentos de electrónica", Edicions UPC, 1994.

[Rubio00] A. Rubio, J. Altet, X. Aragonés, J.L. González, D. Mateo, F. Moll, "Diseño de circuitos y sistemas integrados", Edicions UPC, 2000.

[Benini95] L. Benini, "Lectures: Digital Integrated Circuits", University of Bologna, 1995.

6. Consum de corrents en CMOS.

***Abstract*—**En aquest article es parlarà del consum de corrents present en qualsevol circuit integrat CMOS. Es veurà com la dissipació de potència total és deguda a tres contribucions (estàtica, dinàmica i de curtcircuit) i també les diferents opcions existents per tal de minimitzar-la.

***Index Terms*—**Potència estàtica, potència dinàmica, potència de curtcircuit, corrent de fuites, corrent de saturació, corrent subllindar, factor d'activitat, capacitat de càrrega.

1. INTRODUCCIÓ

En un circuit integrat existeixen tres tipus de consum: l'estàtic, el dinàmic i el de curtcircuit. El primer és causat bàsicament pels corrents de fuites de les unions i subllindar dels MOSFETs. El segon és degut a la càrrega i descàrrega de les capacitats durant les commutacions dels transistors. El tercer és degut a la no nul·litat dels temps de pujada dels transistors, cosa que fa que dos transistors puguin conduir alhora.

A causa de la gran demanda d'equips electrònics portàtils, el consum de corrents s'ha convertit en un factor a optimitzar a l'hora de dissenyar circuits integrats, amb una importància similar a la que tenen la velocitat o l'àrea. La minimització del consum del circuit és una tasca que ha d'enfocar-se des dels diferents nivells d'abstracció utilitzats en el disseny microelectrònic: algorítmic, estructural, lògic, circuital, físic i tecnològic.

2. POTÈNCIA DISSIPADA EN CMOS

En qualsevol porta lògica, el consum total pot dividir-se en els termes que apareixen a l'eq. (1), que s'estudiaran tot seguit:

$$P_{total} = P_s + P_d + P_{sc} \tag{1}$$

2.1. Dissipació de potència estàtica

El consum per corrent de fuites o *leakage* és degut a l'existència de fuites als transistors. Aquests corrents donen lloc a un consum estàtic:

$$P_{leakage} = I_{leakage} \cdot V_{DD} \tag{2}$$

Aquest corrent de fuites presenta dos components. D'una banda es tenen els corrents causats per les unions pn (*drain*/*source* i pou/substrat) polaritzades en inversa, i per l'altra, el corrent subllindar dels transistors.

$$I_{leakage} = I_{pn} + I_{subllindar} \tag{3}$$

Es tractaran en primer lloc els corrents deguts a les unions pn en inversa. A la Fig. 1 es mostra una secció transversal d'un inversor CMOS convencional amb les unions pn paràsites que es formen. Les unions de la font i del drenador del transistor nMOS amb el substrat p formen dues unions pn, així com el drenador i la font amb el pou n. A més, també s'ha de considerar la unió que formen directament el pou n i el substrat p. Perquè tots aquests díodes paràsits no entrin en directa, el pou n es polaritza a V_{DD} (la tensió més alta que pot haver-hi dins del CI) i el substrat p a GND (la més baixa), mitjançant contactes n^+ i p^+. D'aquesta forma, es garanteix que els díodes no estaran mai en directa, però encara així existeix el corrent de saturació en inversa que caracteritza tota unió pn. L'expressió pel corrent d'una unió pn és:

$$I_{pn} = I_S \cdot \left(e^{qV/kT} - 1\right) \tag{4}$$

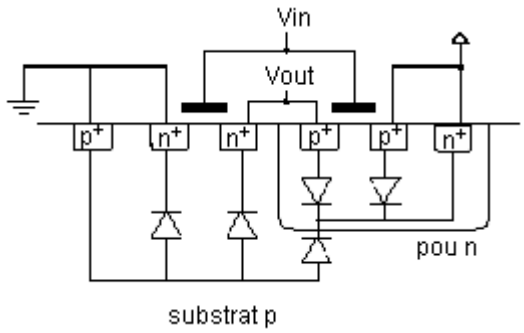


Fig. 1 Tall transversal d'un inversor CMOS sobre tecnologia de pou n.

En quant al consum a causa del corrent subllindar, per tensions de control V_{GS} menors que la llindar V_T, el MOS continua conduint, encara que en menor proporció. És el que s'anomena conducció subllindar.

$$P_{subllindar} = I_{subllindar} \cdot V_{DD} \quad (5)$$

L'equació que permet trobar el valor del corrent subllindar d'un nMOS és:

$$I_{subllindar} = I_{D0}\left(\frac{W}{L}\right)e^{(V_{GS}-V_T)/(n\phi_T)}\left(1-e^{-V_{DS}/\phi_T}\right) \quad (6)$$

Utilitzant valors d'una tecnologia de 0.25µm (I_{D0}=3µA, V_T=0.25V i n=1.5), suposant que V_{DS} és suficientment gran com per poder menysprear el terme de la dreta i una tensió de control V_{GS} nul·la, s'obté el valor del corrent subllindar.

$$I_{subllindar} \approx 23nA \quad (7)$$

Per una alimentació de 2.5V el consum d'un CI d'un milió de portes és:

$$P_{subllindar} \approx 0{,}3W \quad (8)$$

Si se suposa la mateixa tecnologia, però operant a una tensió d'alimentació de només 1.5V i utilitzant una tensió llindar de valor V_T=0.15V, el consum augmenta exponencialment:

$$P_{subllindar} \approx 1{,}4W \quad (9)$$

S'observa que el consum a causa del corrent subllindar pot arribar a ser significatiu. Val la pena comentar que en el cas de tecnologies que no siguin CMOS existeix una contribució molt important al consum degut a l'existència d'algun camí conductiu de baixa impedància entre V_{DD} i GND. En aquest cast:

$$P_s = I_s \cdot V_{DD} \quad (10)$$

on I_s és el corrent consumit en condicions estàtiques, incloent-hi les contribucions comentades anteriorment. Un exemple clar son les lògiques nMOS, en les quals existeix un consum de corrent no nul quan una porta dóna un '0' a la seva sortida. En lògica CMOS complementària, és a dir, en aquella lògica on en cada moment només una de les dues xarxes nMOS/pMOS es troba activa, el corrent consumit en condicions estàtiques és nul, ja que en aquestes condicions mai hi haurà cap connexió entre V_{DD} i GND. Ha d'evitar-se la utilització de famílies lògiques que tinguin consum en condicions estàtiques, ja que en algunes ocasions pot ser important.

2.2. Dissipació de potència dinàmica

El consum dinàmic és degut a les commutacions dels nodes del circuit. Per canviar el valor de tensió de qualsevol node, es requereixen desplaçaments de càrrega a través d'un medi dissipatiu (els transistors), que consumeix energia. Per estimar aquesta dissipació s'analitzarà la commutació de l'inversor de la Fig. 2 i de la Fig. 3, que es considera carregat a la seva sortida amb una capacitat C_L.

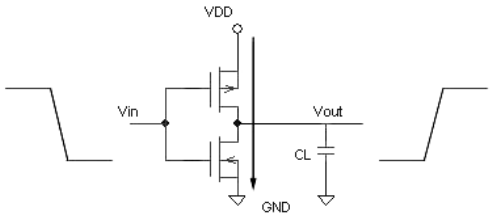


Fig. 2 Càrrega de la C_L d'un inversor CMOS.

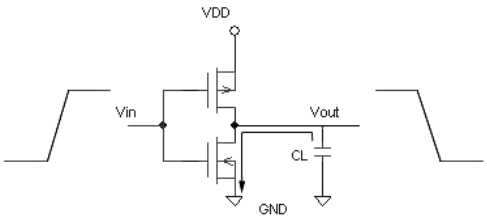


Fig. 3 Descàrrega de la C_L d'un inversor CMOS.

Se suposa que inicialment l'entrada val '1' i, per tant, la sortida pren el valor '0' i la capacitat C_L es troba descarregada. Si l'entrada commuta d''1' a '0', el transistor nMOS entra en tall i s'activa el pMOS, cosa que implica que la sortida es connecta a V_{DD} i C_L comença a carregar-se. Si pel cas contrari l'entrada valgués inicialment '0' i commutés a '1', el pMOS entraria en tall i el nMOS s'activaria, començant a descarregar el condensador C_L.

Se suposa que l'inversor treballa en un sistema que funciona a una freqüència de rellotge f_{CLK} i que la probabilitat que el node de sortida hagi de commutar en un període de rellotge és γ (factor d'activitat). Llavors, la potència dinàmica que l'inversor consumeix en les esmentades condicions és:

$$P_d = \gamma \cdot C_L \cdot V_{DD}^2 \cdot f_{CLK} \quad (11)$$

La dissipació de potència dinàmica, l'expressió de la qual és aplicable a qualsevol porta lògica CMOS, constitueix la contribució més important de tot el consum en qualsevol circuit digital. Per tant, és també la que més s'ha de procurar disminuir.

2.3. Dissipació de potència de curtcircuit

A l'apartat anterior, s'ha suposat que el senyal d'entrada precisava d'un temps nul per realitzar la commutació. Això és una idealització, ja que els senyals reals requereixen d'un cert temps no nul (anomenat temps de pujada o de baixada) per commutar. Se suposa ara un senyal d'entrada com el de la Fig. 4 i una tensió d'alimentació tal que:

$$V_{DD} > V_{Tn} + \left|V_{Tp}\right| \tag{12}$$

És evident que existeix un període de temps en què la tensió d'entrada V_{in} es troba a la regió:

$$V_{DD} - \left|V_{Tp}\right| > V_{in} > V_{Tn} \tag{13}$$

Així doncs, durant el període de temps en què V_{in} es troba a la regió anterior, ambdós transistors condueixen i això provoca l'existència d'un corrent, anomenat de curtcircuit, entre V_{DD} i GND.

Tot seguit, es calcularà l'expressió per la dissipació de potència de curtcircuit. En primera aproximació, se suposarà que la capacitat de càrrega és nul·la, que els transistors tenen unes longituds mínimes i unes amplades tals que $\mu_p W_p = \mu_n W_n$. També se suposaran tensions llindars d'igual mòdul, és a dir, $V_{Tn} = |V_{Tp}|$.

Entre els punts τ_1 i τ_2 (veure Fig. 4), es compleix l'expressió (13) i, per tant, es tindrà conducció de curtcircuit. El transistor nMOS començarà a conduir a l'instant τ_1, trobant-se al principi en saturació. Quan l'entrada passi per $V_{DD}/2$, aproximadament, aquest transistor deixarà d'estar en saturació. L'expressió pel corrent de curtcircuit entre τ_1 i τ_2:

$$I_{sc} \approx I_{nMOS,SAT} = (\beta_n/2)(V_{in} - V_{Tn})^2 \tag{14}$$

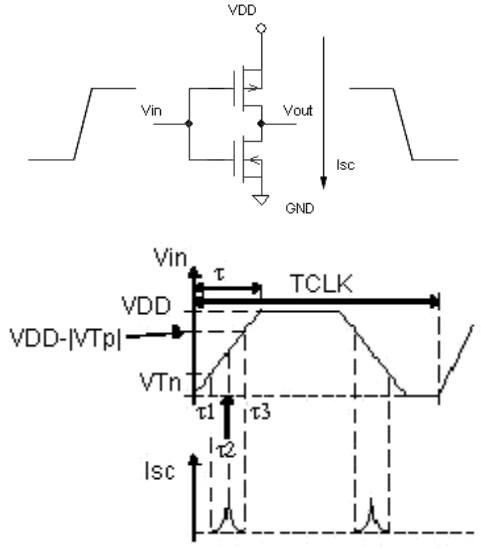


Fig. 4 Corrent de curtcircuit.

En el moment en què V_{in} sigui igual a $V_{DD}/2$, el transistor nMOS passarà a la regió òhmica i el pMOS entrarà en saturació. Per calcular la potència de curtcircuit total dissipada en una commutació de la sortida, es calcula el corrent mig d'un període:

$$I_{sc,mean} = 2\left[\frac{1}{T_{CLK}}\int_{\tau_1}^{\tau_2} I(t)dt + \frac{1}{T_{CLK}}\int_{\tau_2}^{\tau_3} I(t)dt\right] \tag{15}$$

I donat que els corrents de saturació d'ambdós transistors són iguals:

$$I_{sc,mean} = 2\left[\frac{2}{T_{CLK}}\int_{\tau_1}^{\tau_2} \frac{\beta}{2}\left(V_{in}(t) - V_T\right)^2 dt\right] \tag{16}$$

Finalment, tenint en compte que:

$$V_{in}(t) = (V_{DD}/\tau)t \tag{17}$$

$$\tau_1 = (V_T/V_{DD})\tau \tag{18}$$

$$\tau_2 = \tau/2 \tag{19}$$

es troba l'expressió pel corrent mig:

$$I_{sc,mean} = \frac{\beta}{12 \cdot V_{DD}}\left(V_{DD} - 2V_T\right)^3 \frac{\tau}{T_{CLK}} \tag{20}$$

La potència de curtcircuit queda com:

$$P_{sc} = \frac{\beta}{12}\left(V_{DD} - 2V_T\right)^3 \frac{\tau}{T_{CLK}} \tag{21}$$

Per tant, la potència de curtcircuit és, per tensió d'alimentació i freqüència fixes, proporcional a l'amplada dels transistors i al temps de pujada de l'entrada, ja que com més gran és aquest, major és el temps durant el qual circula corrent.

En cas de tenir una capacitat suficientment gran a la sortida, el corrent de curticircuit que passarà serà molt petit. En el límit, per una capacitat de càrrega tendint a infinit, el corrent de curtcircuit serà aproximadament zero.

Per tal de disminuir el corrent de curtcircuit, només cal tenir uns temps de commutació d'entrada i de sortida similars.

3. DISMINUCIÓ DEL CONSUM DEGUT ALS CORRENTS

3.1. Minimització de la potència estàtica

El nivell de disseny que afecta de forma directa al consum estàtic és el circuital, ja que segons quina família lògica s'utilitzi es tindrà o no aquest tipus de consum. Si es desitja un disseny amb baixa dissipació de potència estàtica, la lògica escollida ha de ser tal que en condicions estàtiques no tingui cap camí de baixa impedància entre alimentació i massa. Per això, han d'evitar-se les lògiques nMOS, pseudo-nMOS i derivades. Les lògiques que no

presenten consum estàtic són la complementària i les basades en transistors de pas.

La contribució més important al corrent de fuites és la deguda als corrents subllindar. Aquests corrents poden entendre's com la conseqüència de l'existència de camins d'alta impedància entre alimentació i massa. Per tant, una forma de minimitzar els corrents de fuites per conducció subllindar és utilitzant lògiques de transistors de pas, en les quals hi ha molt poques xarxes connectades entre V_{DD} i GND.

En cas d'utilitzar la lògica estàtica convencional o lògiques dinàmiques, existeixen solucions que es basen en realitzar un control dinàmic de la tensió llindar V_T. Augmentant el mòdul d'aquesta tensió, s'aconsegueix disminuir el corrent de fuites per conducció subllindar. A la Fig. 6 es mostra com es realitza l'esmentat control dinàmic. En una porta convencional, el terminal de substrat dels transistors està connectat a V_{DD} en el cas dels transistors pMOS i a GND en el dels nMOS. El control de V_T es fa tenint el terminal de substrat accessible com un senyal més. D'aquesta forma, un augment de la tensió de substrat provoca un augment de la tensió llindar corresponent (V_{Tn} o $|V_{Tp}|$), amb el que es disminueix el corrent subllindar dels nMOS o pMOS respectivament. El control de la tensió de substrat ha de realitzar-se tant sobre els transistors nMOS com sobre els pMOS, ja que una de les dues xarxes estarà sempre en conducció.

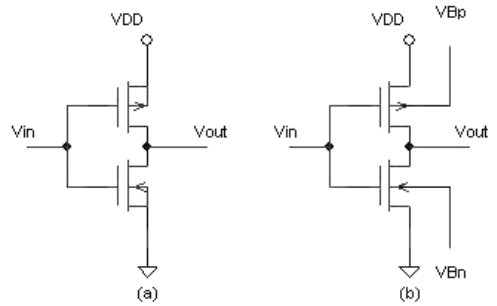


Fig. 6 Connexió normal de les polaritzacions de substrat (a) i reducció del consum mitjançant el control dinàmic (b).

Un augment de la tensió llindar, però, provoca un augment del temps de retard de la porta. És per aquest motiu que a l'hora de realitzar el control de V_T cal fer una partició del sistema en blocs, de manera que quan un determinat bloc no sigui utilitzat o bé no s'esperi d'ell una gran velocitat de resposta, es pugui augmentar el valor de la seva tensió de substrat per tal de disminuir la potència dissipada per corrent subllindar.

Pel que fa al nivell tecnològic, és possible reduir tant els corrents inversos de saturació de les unions mitjançant el control dels perfils de dopatge, com els corrents deguts a la conducció subllindar mitjançant la variació de les tensions llindar.

Finalment, ha de tenir-se en compte que els corrents de fuites, tant els deguts als corrents inversos de saturació com als de conducció subllindar, tenen una dependència tipus exponencial amb la temperatura. Sistemes elèctrics que treballin a baixa temperatura mitjançant refrigeració poden ser una possible alternativa per minimitzar aquest consum.

3.2. Minimització de la potència dinàmica

La potència dinàmica representa una bona part del consum total i, per tant, es dediquen molts esforços a intentar minimitzar-la. Per reduir aquest tipus de consum existeixen diferents alternatives, consistents totes elles en disminuir algun o varis dels factors que intervenen en l'expressió de la potència dinàmica (veure eq. (3)), ja sigui el factor d'activitat, la capacitat de càrrega o la tensió d'alimentació. La freqüència del rellotge és un paràmetre que no s'utilitza per reduir el consum, ja que la seva disminució incrementa el temps del procés.

Una opció per reduir el consum d'un circuit integrat consisteix en minimitzar la seva activitat a la imprescindible, és a dir, reduir el factor d'activitat. El primer nivell de disseny on ha de considerar-se la minimització de l'activitat és al nivell algorítmic. L'elecció d'un algoritme de treball o altre pot ser de gran importància a l'hora de minimitzar el nombre d'operacions a realitzar pel circuit, ja que si s'aconsegueix reduir el nombre de commutacions, també es redueix l'activitat del circuit i es minimitza la potència dinàmica.

A nivell d'arquitectura es poden reduir les transicions innecessàries o *glitchos*, que poden representar entre el 20 i el 70% del consum dinàmic total. El nombre de transicions indesitjades pot reduir-se minimitzant la profunditat d'un bloc, tot augmentant el seu paral·lelisme i utilitzant *pipeline*.

Una segona manera de minimitzar el consum dinàmic consisteix en reduir la capacitat que ha de commutar. A nivell de *layout*, un bon disseny de les cel·les permet reduir les capacitats paràsites. Les eines del CAD usualment realitzen la col·locació tot minimitzant la longitud de les interconnexions, cosa que equival a minimitzar la capacitat de cada línia. Durant la síntesi lògica es té en compte el consum i es poden mapejar nodes importants (capacitivament parlant) cap a nodes interns fent servir portes complexes tipus AOI.

Finalment, el tercer factor que es pot disminuir és la tensió d'alimentació. A la vegada, la disminució d'aquest factor també comporta una disminució del possible consum estàtic, del consum de curtcircuit i del consum a causa del corrent de fuites.

Com a contrapartida, una disminució de la tensió d'alimentació provoca un augment del temps de resposta. Per solucionar aquest inconvenient es pot augmentar la velocitat del sistema mitjançant paral·lelisme, compensant així l'augment del temps de resposta a costa d'augmentar l'àrea i el consum. Malgrat aquests inconvenients, s'aconsegueix minimitzar la potència dinàmica.

3.3 Minimització de la potència de curtcircuit

La minimització de la potència deguda al corrent de curtcircuit ha de ser considerada als nivells de disseny circuital i tecnològic. En *standard cells* el que es fa és dissenyar les portes de forma que els temps de commutació siguin aproximadament els mateixos a totes les commutacions. D'aquesta forma, el consum de curtcircuit es manté per sota del 20% del dinàmic.

A nivell circuital, el tipus de lògica escollida afecta, presentant les lògiques dinàmiques un corrent de curtcircuit inferior o nul. També existeix la possibilitat de treballar amb una tensió d'alimentació que no compleixi l'eq. (12). D'aquesta forma es garanteix que el corrent de curtcircuit serà nul, ja que els transistors nMOS i pMOS mai es trobaran actius al mateix temps, ni tan sols en el cas de commutacions extremadament lentes.

4. CONCLUSIONS

El consum de corrents en lògica CMOS és degut a tres components. El consum estàtic apareix com a conseqüència de l'existència de camins conductius de baixa impedància entre V_{DD} i GND. Pot arribar a ser important en lògiques nMOS, però és molt baix en lògica CMOS complementària ja que mai es troben actives ambdues xarxes alhora, essent únicament degut al corrent de fuites, que depèn dels corrents causats per les unions pn polaritzades en inversa i del corrent subllindar dels transistors.

Un altre component és el consum dinàmic, que es manifesta com a conseqüència de la commutació de les entrades del circuit. És la contribució més important a la potència total dissipada i existeixen diferents opcions per minimitzar-la. Es pot disminuir la tensió d'alimentació, però això augmenta el retard, amb el que també es fa disminuir la tensió llindar, cosa que augmenta el consum de fuites. Ha d'haver-hi, llavors, un compromís entre els dos consums. També es pot minimitzar el factor d'activitat del circuit reduint les transicions innecessàries. L'última opció és disminuir la capacitat de càrrega optimitzant el circuit a nivell de *layout*.

El temps de commutació dels senyals d'entrada és no nul i, a causa d'aquest fet, hi ha un instant de temps en què ambdues xarxes condueixen, cosa que comporta la circulació d'un corrent que connecta V_{DD} i GND. Aquest és el corrent de curtcircuit. Per tal de minimitzar-lo, es poden atacar dos fronts: el nivell circuital o bé el tecnològic. A nivell circuital es pot escollir una lògica dinàmica, que presenta corrent de curtcircut nul, i a nivell tecnològic es pot disminuir la tensió d'alimentació.

5. REFERÈNCIES

[Rubio00] A. Rubio, J. Altet, X. Aragonés, J. L. González, D. Mateo, F. Moll, "Disseny de circuits i sistemes integrats", Edicions UPC, 2000.

[Veendrick84] H. J. Veendrick, "Short-Circuit Dissipation of Static CMOS Circuitry and Impact on the Design of Buffer Circuits", IEEE Journal of Solid State Circuits, vol 19, agost 1984.

[Bolet95] K. Bolet et al., "50% Active-Power Saving without Speed Degradation using Standby Power Reduction (SPR) Circuit", IEEE International Solid State Circuits Conference, San Francisco, 1995.

[Kuroda96] T. Kuroda et al., "A 0.9V, 150MHz, 4mm^2, 2-D Discrete Cosine Transform Core Procesor with Variable Treshold Voltage (V_T) Scheme", IEEE Journal of Solid State Circuits, vol 31, núm 11, novembre 1996.

[Chandrakasan95] A. P. Chandrakasan, R. W. Brodersen, "Low Power Digital CMOS Design", Kluwer Academic Publishers, 1995.

[Borgatti98] M. Borgatti, M. Felici, A. Ferrari, R. Guerrieri, "A Low Power Integrated Circuit for Remote Speech Recognition", IEEE Journal of Solid State Circuits, vol 33, núm 7, juliol 1998.

[Brodersen93] R. W. Brodersen, A. Chandrakasan, S. Sheng, "Design Techniques for Portable Systems", IEEE International Solid State Circuits Conference, 1993.

7. Disseny de portes CMOS.

***Abstract*—**En aquest text es detalla la metodologia a seguir a l'hora de dissenyar una porta lògica CMOS a nivell de transistor, parant especial atenció en l'obtenció del pitjor cas de càrrega i descàrrega de les capacitats i en la relació d'aspecte, de la qual s'extreu la relació que ha d'existir entre les amplades dels transistors que formen la porta. Així mateix, també s'ofereix una breu explicació sobre el mode d'operació i les característiques bàsiques d'un transistor MOS. Es comenta l'existència de capacitats paràsites i com aquestes afecten el funcionament de l'esmentat transistor.

***Index Terms*—**Transistor MOS, lògica CMOS, inversió de població, tensió llindar, disseny a nivell de transistor, capacitat paràsita, pitjor cas, temps de pujada i baixada, relació d'aspecte.

1. INTRODUCCIÓ

La fabricació de circuits electrònics integrats complexos sobre petites oblies de material semiconductor ha significat l'aparició de la microelectrònica. Els inicis de la microelectrònica es remunten a la invenció del transistor bipolar el novembre de 1947 per part de W. Shockley, J. Bardeen i W. Brattain dels laboratoris Bell. Des de llavors, l'electrònica ha evolucionat fins a nivells de miniaturització i densitat d'integració sens dubte inimaginables als seus inicis. Gran part d'aquest progrés deu el seu èxit als transistors d'efecte de camp i en particular al transistor MOS, el primer dels quals va ser fabricat el 1960 per D. Khang, també dels laboratoris Bell, i usat massivament per la producció de dissenys integrats des de mitjans de la dècada dels 70. El funcionament d'aquest dispositiu es basa en la capacitat per modular la concentració dels portadors lliures d'un semiconductor mitjançant un camp elèctric controlable. Això permet variar les seves característiques elèctriques, com la seva conductivitat, pel que els MOS poden ser considerats com interruptors elèctrics i, per tant, com les peces idònies per plasmar físicament els elements de l'àlgebra de Boole.

Els transistors MOS presenten interessants avantatges sobre els primers bipolars, com una menor grandària (permetent una major densitat d'integració) i un menor consum. Per contra, el seu funcionament és més lent encara que això es va compensant progressivament a causa de la constant reducció de la longitud del canal.

La forma de combinar, connectar i relacionar aquests dispositius defineix una lògica de funcionament, ja que amb les mateixes peces es poden inventar diferents mètodes de treball per construir circuits lògics digitals. La més usada en l'actualitat és la lògica CMOS, que combina l'ús de transistors nMOS i pMOS. Aquests es diferencien pels seus portadors majoritaris (tipus n o tipus p), el que els proporciona característiques de funcionament diferenciades. Els nMOS transmeten bé els valors baixos de tensió i els pMOS, els alts. D'aquesta forma la seva combinació en un mateix circuit permet prestacions elèctriques de funcionament molt bones i fiables, de les quals destaca sobretot el seu baix consum. En canvi, el nombre de transistors necessaris per fabricar una porta és més elevat que amb altres tecnologies. Si una porta CMOS té N entrades, el nombre total de transistors serà 2N si és negada, i 2N+2 si no ho és.

2. FUNCIONAMENT I CARACTERÍSTIQUES DEL TRANSISTOR MOS

Elèctricament, els transistors MOS són elements de 4 connexions, que s'anomenen *source* (S), *drain* (D), *gate* (G) i *bulk* (B).

Es pot entendre el seu funcionament a nivell digital com si es tractessin d'interruptors definits entre els terminals de *source* i *drain* (V_{DS}), controlats per les tensions entre *gate* i *bulk* (V_{GB}). En primera aproximació, els dos primers terminals són indiferents quant a construcció i funcionament, i només depenen de la seva polarització. Així, el *source* serà el terminal pel qual entren les càrregues elèctriques al dispositiu i el *drain* el terminal pel que surten. El valor de V_{GB} determinarà si l'interruptor està en *on* o bé en *off*. Als transistors nMOS, l'estat de conducció entre *drain* i *source* es tindrà sempre que la tensió d'entrada (V_{GS}) sigui major que la tensió llindar (V_{Tn}). En cas contrari, l'interruptor estarà en *off* i no conduirà. Pels pMOS, en canvi, es tindrà conducció entre *source* i *drain* quan la diferència entre la tensió d'entrada i la de *source* ($|V_{GS}|=|V_{in}-V_{DD}|$) sigui major que la llindar ($|V_{TP}|$).

Per analitzar aquest comportament amb major detall, es necessita conèixer la seva estructura interna. A la Fig. 1 es pot observar la construcció d'un nMOS.

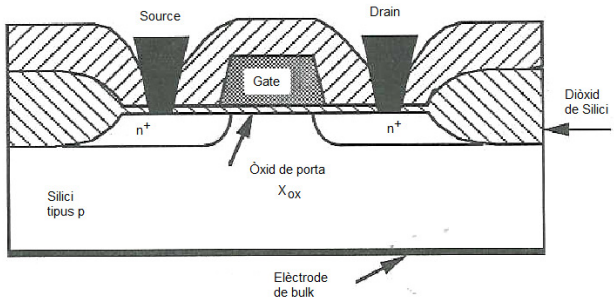


Fig. 1 Construcció d'un transistor nMOS.

Com es pot observar, entre els terminals de *drain* i *source* hi ha una regió tipus p, anomenada canal del transistor, de manera que apareixen dues unions pn. Ambdues unions es troben sempre en inversa, impossibilitant el pas de corrent cap al substrat, que està connectat a GND. El funcionament d'aquests dispositius es basa en la redistribució de portadors a la zona de canal situada entre *source* i *drain*, que pot ser induïda pel camp elèctric aplicat entre els terminals de *gate* i *bulk*. D'aquesta manera, es pot assolir una inversió de la població a la zona del canal, és a dir, es pot aconseguir que creixi la densitat dels portadors minoritaris en aquesta zona fins al punt que passin a ser majoritaris. Això es traduiria en el fet que el canal passaria de ser una zona p a una zona n i, per tant, desapareixerien les unions pn que trencaven la connexió entre *drain* i *source*. En aquesta situació, la circulació de càrregues entre aquests dos terminals és possible ja que el canal s'ha convertit en una simple resistència semiconductora de tipus n. A la Fig. 1 s'observa que entre *gate* i *bulk* hi ha una fina capa d'òxid de porta, que impedeix que hi hagi conducció entre aquests dos terminals. Aquesta estructura és la que dóna nom al dispositiu: *Metal Oxide Semiconductor Field Effect Transistor*. Un pMOS funcionaria de la mateixa forma, però el substrat seria de tipus n i estaria polaritzat a una tensió més alta (V_{DD}). Les regions de *drain* i *source* serien de tipus p.

Cal tenir en compte que, perquè s'arribi a produir la inversió de població, és necessari aplicar un voltatge mínim entre *gate* i *bulk* anomenat voltatge *threshold* o llindar (V_T). Per sota d'aquest voltatge, que és positiu pels n i negatiu pels p, es diu que l'interruptor està obert (*off*).

Una vegada assolida la situació d'inversió, la conductivitat del canal augmenta a mesura que es va superant el nivell de *threshold*. Això depèn no només de la tensió V_{GB}, sinó també de la tensió V_{DS}. Una vegada superat el *threshold* i per V_{DS} petits, es diu que el transistor es troba en règim de funcionament lineal ja que el seu canal es comporta com una simple resistència. A mesura que V_{DS} creix, s'arriba a un punt en què la intensitat que circula pel canal és màxima i, idealment, és constant. Aquesta és l'anomenada regió de saturació.

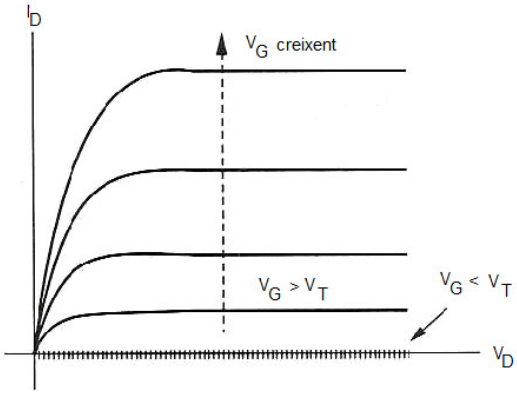


Fig. 2 Evolució del corrent en funció de V_{DS} per diferents valors de V_{GB} en un transistor nMOS.

Les equacions més simples que descriuen aquest funcionament (i que per l'anàlisi que es realitzarà seran suficientment adequades) són les que consten a continuació, on (1) correspon a l'equació per la regió de tall, (2) per la regió lineal i (3), per la de saturació.

$$V_{GS} < V_{Tn} \rightarrow I_D = 0 \quad (1)$$

$$\left.\begin{array}{l} V_{GS} > V_{Tn} \\ V_{DS} < V_{GS} - V_{Tn} \end{array}\right\} \rightarrow I_D = \beta_n \left[(V_{GS} - V_{Tn}) V_{DS} - \frac{1}{2} V_{DS}^2 \right] \quad (2)$$

$$\left.\begin{array}{l} V_{GS} > V_{Tn} \\ V_{DS} \geq V_{GS} - V_{Tn} \end{array}\right\} \rightarrow I_D = (\beta_n / 2)(V_{GS} - V_{Tn})^2 \quad (3)$$

En aquestes expressions, el terme β_n és igual a $K_n \cdot (W/L)_n$, amb $K_n = \mu_n \cdot C_{ox}$. μ_n és la mobilitat dels portadors majoritaris (electrons en aquest cas) una vegada establert el canal i C_{ox} fa referència a la capacitat de l'estructura MOS per unitat d'àrea. El seu valor es pot aproximar per l'expressió $C_{ox} = \varepsilon_{ox}/t_{ox}$, on t_{ox} és el gruix de l'òxid. W és l'amplada del canal i L la seva longitud.

L'equació per la regió de saturació dibuixa una paràbola i és vàlida fins que s'arriba al seu màxim, és a dir, fins al punt de saturació definit per $V_{DS} = V_{GS} - V_{Tn}$. A partir d'aquí, el corrent deixa de créixer amb V_{DS} i es manté aproximadament constant. Si no se supera el nivell de *threshold*, el transistor està en tall i no hi ha circulació de corrent pel seu canal. Les expressions anteriors serveixen tant pel nMOS com pel pMOS, ja que només cal tenir en compte que en el darrer cas les tensions aplicades es troben en valor absolut, tal com es mostra a continuació.

$$|V_{GS}| < |V_{Tp}| \rightarrow I_D = 0 \quad (4)$$

$$\left.\begin{array}{l} |V_{GS}| > |V_{Tp}| \\ |V_{DS}| < |V_{GS}| - |V_{Tp}| \end{array}\right\} \rightarrow I_D = \beta_p \left[(|V_{GS}| - |V_{Tp}|) |V_{DS}| - \frac{1}{2} |V_{DS}|^2 \right] \quad (5)$$

$$\left.\begin{array}{l}|V_{GS}|>|V_{Tp}| \\ |V_{DS}|\geq|V_{GS}|-|V_{Tp}|\end{array}\right\}\rightarrow I_D=(\beta_p/2)(|V_{GS}|-|V_{Tp}|)^2 \quad (6)$$

Convé notar que el corrent és proporcional a la mobilitat dels portadors. Normalment la dels electrons és el triple que la dels forats, i això és el que fa que siguin més ràpids els transistors tipus n.

A continuació, s'analitzen les capacitats paràsites associades al disseny vist a la Fig. 1. En total hi ha 5 capacitats, que són la de *gate-bulk*, *gate-drain*, *gate-source*, *source-bulk* i *drain-bulk*, segons s'observa a la Fig. 3.

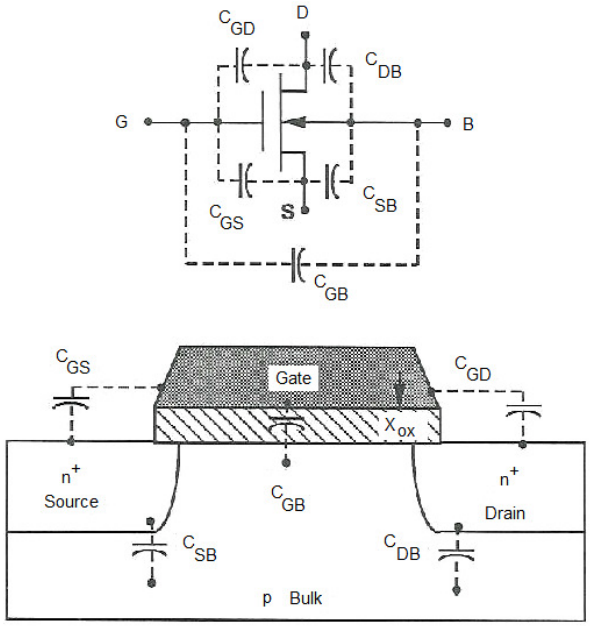


Fig. 3 Capacitats del MOS.

Convindrà fixar-se principalment en les tres capacitats referides a *bulk*, que s'anomenen capacitats de *gate*, *source* i *drain*. La primera i més important és la capacitat de porta. És evident que l'estructura MOS equival a l'estructura d'un condensador de làmines paral·leles, ja que té dos materials conductors (metall i semiconductor) separats per un dielèctric (òxid). Així doncs, en aplicar un voltatge V_{GB} s'emmagatzema una càrrega elèctrica al transistor, que és la responsable de la inversió de població. Les altres dues capacitats (de *source* i *drain*) apareixen a causa de la polarització en inversa de les unions pn. És a dir, són les capacitats d'una unió pn. Aquestes són les responsables de l'asimetria de prestacions entre les diferents commutacions possibles d'una porta CMOS. Els seus valors vénen determinats per les equacions:

$$C_{GB}=C_{ox}\cdot W\cdot L \quad (7)$$

$$C_{SB}=C_{j0}\cdot L_S\cdot W+C_{jsw}\cdot 2(W+L_S) \quad (8)$$

$$C_{DB}=C_{j0}\cdot L_D\cdot W+C_{jdw}\cdot 2(W+L_D) \quad (9)$$

on C_{j0} és la capacitat de la unió per unitat d'àrea, i C_{jsw} i C_{jdw} són les capacitats per unitat de longitud associades al perímetre de les regions de *source* i *drain*, respectivament. W i L són l'amplada i la longitud del canal; L_S i L_D corresponen a la longitud inferior de *source* i *drain*.

Els paràmetres C_{ox}, C_{j0}, C_{jsw} i C_{jdw} apareixen calculats als *datasheets* de totes les tecnologies. Les altres capacitats existents són negligibles per qualsevol porta lògica, tal i com s'explicarà més endavant.

3. DISSENY D'UNA PORTA CMOS A NIVELL DE TRANSISTOR

L'esquema bàsic general d'una porta CMOS es pot observar a la Fig. 4:

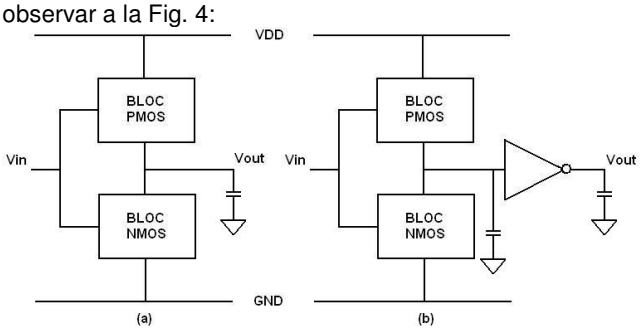


Fig. 4 Esquema d'una porta CMOS.

La Fig. 4(a) representa una funció negada i la 4(b) una funció sense negar. Per una banda es té un bloc format per transistors tipus p i per l'altra, un de format per transistors tipus n. El bloc p està connectat a V_{DD} per un extrem i per l'altre a la sortida de la porta, ja que és un bon transmissor d'uns lògics. El bloc n, en canvi, està connectat entre GND i la sortida perquè transmet bé els zeros lògics.

El funcionament de la porta és tal que quan la sortida hagi de ser 1, el bloc p conduirà i permetrà que es carregui la capacitat de sortida fins a V_{DD}, mentre que el bloc n estarà en tall i, per tant, serà com si no existís. Quan la sortida hagi de ser 0 la situació s'invertirà, descarregant-se la capacitat a través del bloc n. L'estat de cadascun dels blocs vindrà definit pels valors de les entrades, que aniran connectades a les portes dels transistors de cada bloc. Convé notar que mentre que un dels dos blocs condueix l'altre està en tall, de manera que en situació estacionària mai es crea un camí entre V_{DD} i GND. Aquest és el motiu pel qual aquesta tecnologia es diu lògica MOS complementària (CMOS), ja que el comportament lògic de cada bloc és el complementari de l'altre. Una vegada la porta ha commutat, no hi ha consum, essent aquest el principal punt fort d'aquest tipus de dissenys. També cal notar que com que el consum només es produeix durant el

canvi d'estat de les portes, als sistemes síncrons de tecnologia CMOS el consum total s'associarà a la freqüència de rellotge. A major freqüència més canvis d'estat, més corrent i major consum.

Es diu d'aquesta lògica que és estàtica perquè quan s'ha produït l'activació d'un dels dos blocs, la sortida queda connectada a V_{DD} o GND. Això significa que aquest tipus de sistemes són molt fiables, ja que no es produeixen atenuacions dels senyals que puguin traduir-se en errors i pèrdues d'informació.

Tot seguit, s'explica com es dissenya cada bloc. En primer lloc cal conèixer la lògica associada als transistors nMOS i pMOS. Un nMOS deixa passar el corrent quan hi ha un 1 a la porta i està en tall quan hi ha un 0. Així, dos transistors nMOS en sèrie donaran una funció NAND de les seves entrades, ja que només es tindrà un 0 a la sortida quan els dos nMOS condueixin. Només en aquest cas existirà connexió entre la sortida i GND.

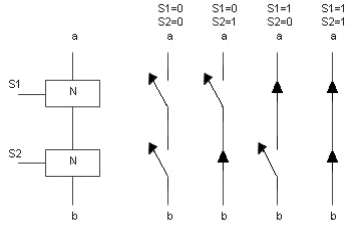


Fig. 5 Connexió en sèrie d'interruptors n.

Dos transistors nMOS en paral·lel, en canvi, donaran una funció NOR ja que es tindrà un 1 a la sortida només en el cas en què totes dues entrades siguin 0. Per la resta de casos, sempre existirà una connexió entre GND i la sortida, cosa que farà que aquesta darrera sigui 0.

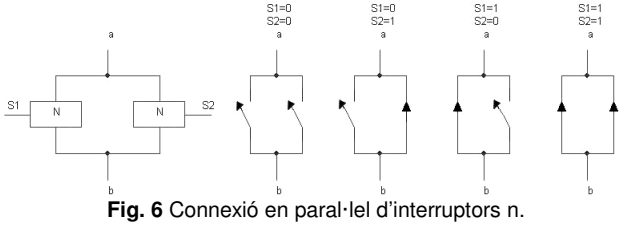


Fig. 6 Connexió en paral·lel d'interruptors n.

A diferència dels nMOS, els pMOS són actius quan a la seva porta hi ha un 0 (una entrada que connecta un nMOS desconnecta un pMOS). Així, combinar pMOS en sèrie o paral·lel farà les funcions AND i OR de les seves entrades negades, cosa que segons el teorema de Morgan és equivalent a dir que dos pMOS en paral·lel fan una NAND i dos pMOS en sèrie fan una NOR.

S'estableix la següent estratègia de disseny d'una porta CMOS a nivell de transistors:

• Tota variable d'entrada va a un transistor nMOS i a un pMOS. Una porta amb N entrades tindrà 2N transistors.

• Com que la lògica CMOS és una lògica inversora, si el que es vol és implementar una funció que no sigui negada, s'hauria d'afegir a la sortida de la porta un inversor (veure Fig. 4 (b)). Llavors, la porta tindria 2N+2 transistors, ja que l'inversor consta de dos transistors.

• El bloc pMOS és el dual del nMOS, és a dir, les connexions en sèrie dels nMOS es tradueixen en connexions en paral·lel pels pMOS i viceversa.

Seguint aquests 3 passos es pot dissenyar qualsevol funció lògica. A continuació es tenen alguns exemples.

3.1. Inversor

És la porta més senzilla. El seu disseny a nivell de transistor és:

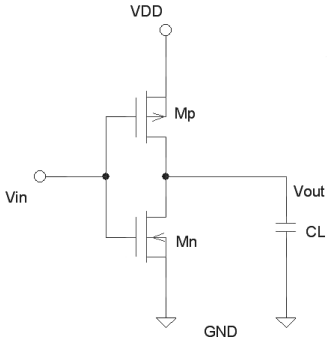


Fig. 7 Disseny d'un inversor.

C_L simbolitza la capacitat de sortida, i generalment serà la capacitat de porta del següent sistema. Si l'entrada és 1, el bloc n condueix i la sortida és 0; quan l'entrada és 0, el bloc p és el que condueix i la sortida és 1.

3.2. NAND2

Una NAND és una AND negada. Una AND al bloc n es construeix amb dos transistors en sèrie. Això significarà dos transistors en paral·lel al bloc p.

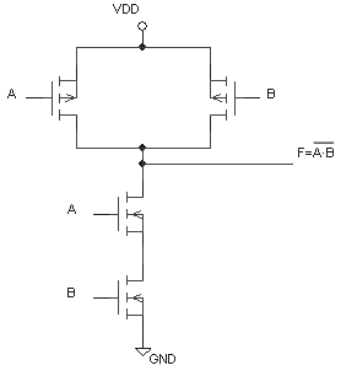


Fig. 8 Disseny a nivell de transistors per una NAND2.

3.3. OR3

Una OR3 és una OR amb tres entrades. El bloc p es construeix amb tres transistors en sèrie, cosa que implica que el bloc n es construirà amb tres transistors en paral·lel. Com que aquesta porta és no inversora, s'ha d'afegir un inversor a la sortida, ja que sinó la porta que es tindria seria una NOR3.

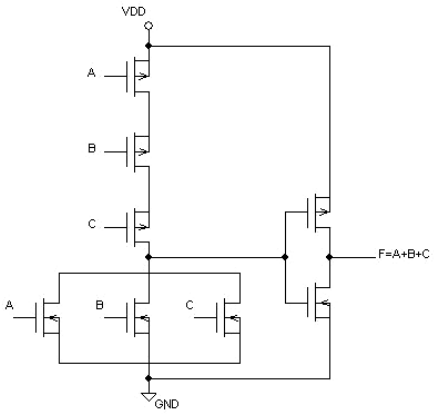


Fig. 9 Esquema a nivell de transistors per una OR3.

3.4. AO43

Una forma comuna de referir-se a una porta és amb les sigles AO, AOI, OA i OAI que fan referència a portes que implementen una suma de productes (AO) o un producte de sumes (OA). La I final fa referència a si la funció és negada o no. Així doncs, la porta que s'ha d'implementar correspon a:

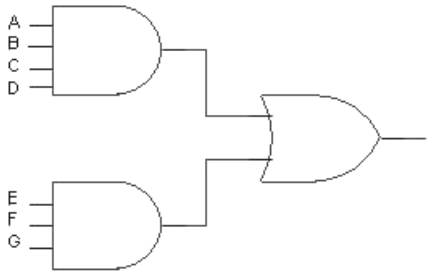


Fig. 10 Disseny a nivell de porta per una AO43.

En el cas concret de la porta AO43 i d'acord amb la Fig. 10, la funció a implementar serà F=(ABCD)+(EFG). Així doncs, per representar aquesta porta en un esquema a nivell de transistors, només cal traduir directament els productes per transistors en sèrie i la suma per connexions en paral·lel al bloc n, afegint un inversor a la sortida per desnegar la funció. El bloc pMOS, com sempre, serà el mateix que el nMOS canviant sèrie per paral·lel i viceversa.

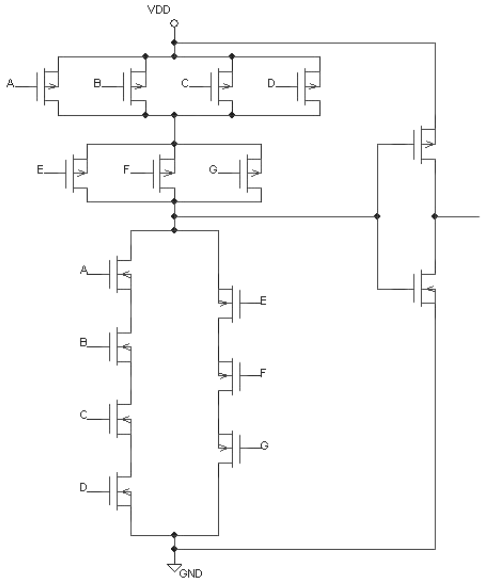
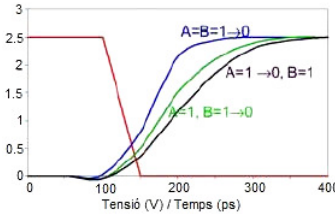


Fig. 11 Disseny a nivell de transistors per una AO43.

4. PITJOR CAS DE CÀRREGA I DESCÀRREGA

En general, els circuits digitals funcionen al ritme que marca el senyal de rellotge. A causa de la naturalesa dels transistors, els processos de commutació dels blocs combinacionals duren el temps en què totes les capacitats (la de sortida i les paràsites internes) triguen en carregar-se o descarregar-se completament. No obstant això, no totes les commutacions degudes a les variacions de les entrades duren el mateix. És sabut que la durada de la càrrega/descàrrega d'una capacitat depèn del seu valor.



Commutació a l'entrada	Retard (ps)
A=0, B=0→1	67
A=1, B=0→1	64
A=0→1, B=1	61
A=1, B=1→0	45
A=1, B=1→0	80
A=1→0, B=1	81

Fig. 12 Retard a la sortida en funció de la commutació a l'entrada.

Les capacitats que es consideraran seran les paràsites degudes a les unions pn de les regions de *source* i *drain* dels transistors, les de porta i la de sortida. Totes aquestes capacitats són respecte massa i, per tant, es poden considerar connectades en paral·lel. A la Fig. 13 i a mode d'exemple, es poden observar les capacitats existents en una porta OR3.

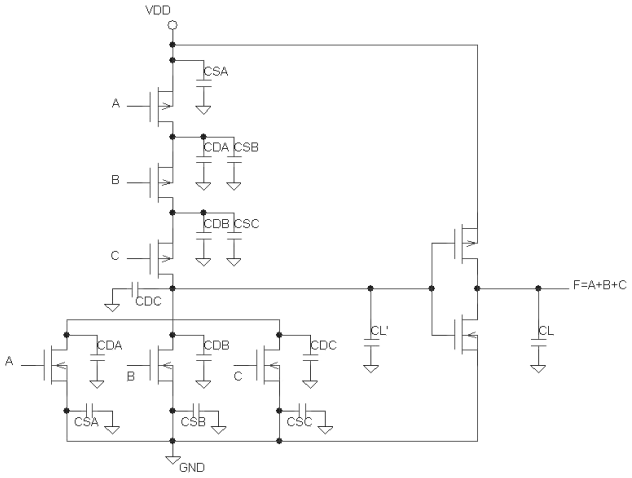


Fig. 13 Esquema a nivell de transistors amb les capacitats paràsites existents.

Es vol fer notar que les capacitats que es troben connectades a un potencial constant (V_{DD} o GND) no s'han de tenir en compte a l'hora buscar el pitjor cas de càrrega i descàrrega. A la figura anterior, aquestes capacitats són C_{SA} al bloc p i C_{SA}, C_{SB} i C_{SC} al bloc n. Així mateix, les capacitats de *drain* i *source* de transistors diferents que comparteixen node, a efectes pràctics poden considerar-se com una capacitat de valor doble.

D'altra banda, com més capacitats paràsites hi hagi, major serà el temps emprat per carregar-les. Les connexions en paral·lel, en canvi, facilitaran la càrrega de la capacitat de sortida. Així, la pitjor commutació d'una porta serà la que involucri un major nombre de transistors en sèrie, i en cas que hi hagi diverses opcions, la que impliqui més capacitats paràsites.

Un transistor MOS pot ser considerat com una resistència R, que depèn de les dimensions del propi transistor. Així doncs, el retard associat a la càrrega/descàrrega del MOS és proporcional al producte RC.

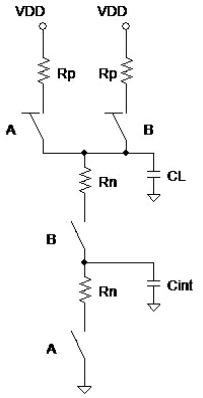


Fig. 14 Model RC corresponent a una NAND2.

Tot seguit, s'exposa el cas de la NAND2. Aquesta porta pot modelitzar-se tal com es mostra a l'esquema de la Fig. 14, on cada transistor es representa mitjançant una resistència. Per calcular el temps de baixada, s'han de considerar ambdós transistors A i B del bloc n, i l'expressió per calcular el temps és la que es mostra a l'eq. (10). Per obtenir el temps de pujada, en canvi, s'ha de tenir en compte que poden conduir ambdós transistors o bé un de sol (A o B). En cada cas, la resistència i la capacitat a considerar seran diferents, tal com es pot veure a les equacions (11), (12) i (13).

$$\tau_{HL} = \left(C_{\text{int}} + C_L\right)\left(R_{nA} + R_{nB}\right) \tag{10}$$

$$\tau_{LH} = C_L \cdot R_{pA} R_{pB} \Big/ \left(R_{pA} + R_{pB}\right) \tag{11}$$

$$\tau_{LHB} = C_L R_{pB} \tag{12}$$

$$\tau_{LHA} = C_L R_{pA} \tag{13}$$

A continuació, s'estudia el pitjor cas de càrrega i descàrrega de la porta OR3. Al bloc p hi ha un sol cas de càrrega, que involucra els tres transistors. Al bloc n, en canvi, el pitjor cas de descàrrega té lloc quan només un dels tres transistors condueix. No obstant això, els tres casos possibles (que només condueixi A o B o C) no són iguals. Quan el procés és a través d'A, al bloc p es té que els transistors B i C condueixen, i per tant estan disponibles les seves capacitats de *drain* i *source*, i la de *drain* del transistor A, fent més lent el procés de descàrrega de la capacitat de sortida. Aquest és el pitjor cas de descàrrega. Quan al bloc n condueix només C, per exemple, al bloc p C està en tall i només està disponible la seva capacitat paràsita de *drain*, cosa que fa que el procés sigui més ràpid. Així, la commutació que servirà per definir les característiques de la porta i les dimensions dels transistors serà la que correspondrà al pas de A=B=C=0 (pitjor cas de càrrega) a A=1, B=C=0 (pitjor cas de descàrrega).

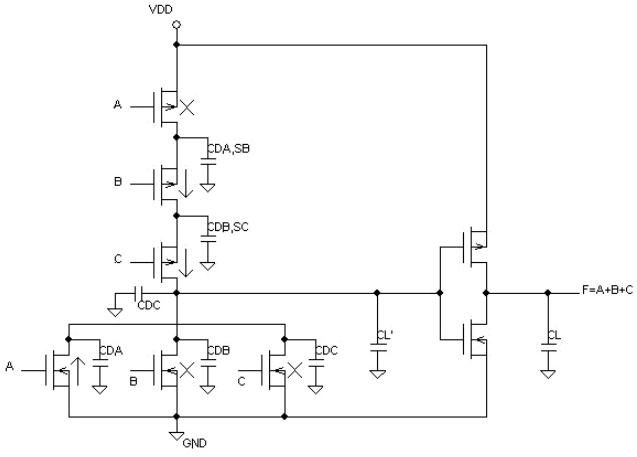


Fig. 15 Pitjor cas de descàrrega.

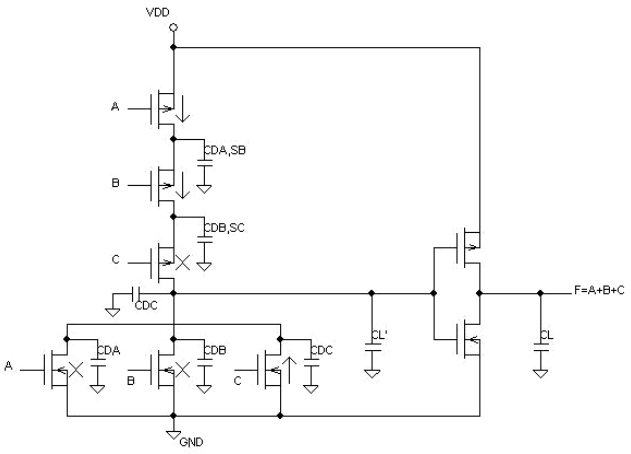


Fig. 16 La conducció del transistor C al bloc n no constitueix el pitjor cas de descàrrega.

5. RELACIÓ D'ASPECTE

Les prestacions i característiques d'una porta es defineixen sobre la commutació d'entrades que duen del pitjor cas de càrrega al pitjor cas de descàrrega. Un criteri a seguir a l'hora de dissenyar la porta és establir que el temps de durada de la pitjor càrrega sigui igual al de la pitjor descàrrega. D'aquí s'extreuen les dimensions dels transistors de cada bloc.

Seguint amb l'exemple de la porta OR3, pel pitjor cas de càrrega, al bloc p es tenen tres transistors en sèrie conduint. Per tal de calcular el temps de pujada, cal considerar la β equivalent d'aquesta xarxa.

$$\frac{1}{\beta_p^{eq}} = \frac{1}{\beta_p} + \frac{1}{\beta_p} + \frac{1}{\beta_p} \tag{14}$$

$$\beta_p^{eq} = \frac{\beta_p}{3} \tag{15}$$

Al bloc n, en canvi, en el pitjor cas de descàrrega només es té un sol transistor conduint. Per tant, la β equivalent d'aquesta xarxa serà senzillament β_n.

Les expressions pels temps de pujada i baixada són les que es mostren a continuació. Aquestes corresponen a la commutació d'un inversor CMOS substituint β_p per β_p^{eq} i β_n per β_n^{eq}.

$$t_r = \frac{C_L}{\beta_p^{eq}\left(V_1 - \left|V_{Tp}\right|\right)}\left[\frac{2\left|V_{Tp}\right|}{\left(V_1 - \left|V_{Tp}\right|\right)} + \ln\left(\frac{2\left(V_1 - \left|V_{Tp}\right|\right)}{V_0} - 1\right)\right] \tag{16}$$

$$t_f = \frac{C_L}{\beta_n^{eq}\left(V_1 - V_{Tn}\right)}\left[\frac{2V_{Tn}}{\left(V_1 - V_{Tn}\right)} + \ln\left(\frac{2\left(V_1 - V_{Tn}\right)}{V_0} - 1\right)\right] \tag{17}$$

C_L és la capacitat de càrrega, β_p^{eq} i β_n^{eq} són les β equivalent per cada xarxa i la resta de termes són una constant, que s'agrupen sota un únic paràmetre per cada tipus de transistor, α_p pels p i α_n pels n. V_0 i V_1 corresponen al 10 i al 90%, respectivament, de la diferència entre la tensió més elevada i la tensió més baixa aplicada al circuit.

$$V_0 = V_{GND} + 0.1\left(V_{DD} - V_{GND}\right) \tag{18}$$

$$V_1 = V_{GND} + 0.9\left(V_{DD} - V_{GND}\right) \tag{19}$$

A les equacions (18) i (19) també pot utilitzar-se V_{SS} en lloc de V_{GND}. Igualant les expressions pels temps de pujada i baixada, s'obté el que s'anomena relació d'aspecte (W_p/W_n) per la porta NOR3.

$$t_r = t_f \tag{20}$$

$$\frac{3\alpha_p C_L}{\mu_p C_{ox}\left(W/L\right)_p} = \frac{\alpha_n C_L}{\mu_n C_{ox}\left(W/L\right)_n} \tag{21}$$

$$\frac{\left(W/L\right)_p}{\left(W/L\right)_n} = 3\frac{\mu_n \alpha_p}{\mu_p \alpha_n} \tag{22}$$

En el cas particular en què $|V_{Tp}|$ sigui igual a V_{Tn}, es té que $\alpha_p=\alpha_n$. Llavors:

$$\frac{W_p}{L_p} = 3\frac{\mu_n}{\mu_p}\frac{W_n}{L_n} \tag{23}$$

I finalment, si es considera que $L_p=L_n$ ($L_p=L_n=0.35$ μm), s'obté:

$$W_p = 3\frac{\mu_n}{\mu_p}W_n \tag{24}$$

que constitueix la relació d'aspecte per la porta NOR3. Es vol remarcar que aquesta expressió no dóna les dimensions dels transistors de cada bloc, sinó la relació que ha d'existir entre elles. A continuació, s'obté la mateixa relació per l'inversor. En aquest cas, donat que només hi ha un sol transistor a cada bloc, només hi ha un possible cas de càrrega i de descàrrega, essent la relació d'aspecte la relació entre mobilitats.

$$W_p = \frac{\mu_n}{\mu_p}W_n \tag{25}$$

Agafant la mobilitat dels electrons com tres vegades la dels forats ($\mu_n=3\mu_p$), es poden calcular alguns exemples numèrics. Prenent $W_n=0.8$μm i d'acord amb l'eq. (24), es té

W_p=7.2µm per la NOR. Per l'inversor, en canvi, prenent la mateixa W_n, es té W_p=2.4µm.

És fonamental tenir en compte que la disposició dels transistors ve determinada per la funció lògica que es vol realitzar, però moltes vegades es tenen diverses situacions possibles dels transistors MOS que corresponen a la mateixa funció. S'optarà per la situació per la que el cas més desfavorable ho sigui el menys possible. En portes complexes això no és fàcil de veure i aquí entra en joc l'experiència del dissenyador. Òbviament, també convé usar les regles de l'àlgebra de Boole per simplificar al màxim la funció que es vol implementar.

6. CONCLUSIONS

La lògica CMOS implementa funcions lògiques fent ús de dos tipus de transistors, els nMOS i els pMOS. Els nMOS s'activen en presència d'un 1 lògic o senyal alt de tensió a la seva porta; els pMOS, en canvi, deixen d'estar en tall quan tenen un 0 lògic o senyal baix de tensió a la seva entrada. Els nMOS transmeten bé els zeros i els pMOS, els uns. És per aquest motiu que és necessària la presència dels dos tipus de transistors per tal d'implementar una funció lògica. D'altra banda, com que la lògica CMOS és inversora, cal afegir un inversor a la sortida si la porta que es vol dissenyar és no inversora.

A l'hora de realitzar una porta CMOS, cal considerar el pitjor cas de càrrega i descàrrega, que contempla el temps màxim que trigarà en carregar-se i descarregar-se completament la capacitat de sortida en la situació més desfavorable. Aquesta situació es dóna quan hi ha més capacitats paràsites pel camí i menys branques que condueixen.

Un dels possibles criteris a seguir a l'hora de determinar les dimensions dels transistors (W_p i W_n) que constitueixen una porta CMOS, és imposar que el temps de pujada i de baixada siguin iguals en el pitjor cas. A partir d'aquí, s'obté el que s'anomena relació d'aspecte, que proporciona la relació que ha d'existir entre les amplades dels transistors p i n que formen la porta.

7. REFERÈNCIES

[Uyemura94] J. P. Uyemura, "Circuit Design for CMOS VLSI", Kluwer Academic Publishers, 1994.

[Baker97] R. J. Baker, H. W. Li, D. E. Boyce, "CMOS circuit design, layout and simulation", IEEE Press Series on Microelectronic Systems: S. K. Tewksbury, Series Editor, 1997.

[Bota96] S. Bota, J. Carrabina, A. Herms, "Introducció al disseny CMOS VLSI", Textos docents núm 52, Edicions de la Universitat de Barcelona, 1996.

[Weste93] N. Weste, K. Eshragian, "Principles of CMOS VLSI Design", Addison-Wesley, 1993.

[Pierret] R. F. Pierret, "Dispositivos de efecto de campo", Addison-Wesley Iberoamericana, 2a edició.

[Brinkman97] W. F. Brinkman, D. E. Haggan, W. W. Troutman, "A History of the Invention of the Transistor and Where It Will Lead Us", IEEE Journal of Solid-State Circuits, vol 32, núm 12, 1997.

8. Lògiques estàtiques alternatives a la lògica CMOS.

***Abstract*—**La lògica CMOS és la més usada en la gran majoria dels dissenys VLSI, per la seva simplicitat, el seu baix consum i la gran disponibilitat d'eines de disseny. Existeixen altres lògiques que ofereixen alternatives, en àrees com la densitat d'integració, la dissipació de potència o la velocitat de commutació, encara que la millora d'alguna d'aquestes prestacions sovint va en detriment d'una altra.

***Index Terms*—**Lògica pseudo-nMOS, lògica BiCMOS, porta de pas, CVSL, CPL, DSL, SFPL.

1. INTRODUCCIÓ

Al llarg d'aquest text es tractaran les lògiques estàtiques més significatives alternatives a la lògica CMOS. Entre les que tenen un ús més comú es troba la lògica pseudo-nMOS, que proporciona una major densitat d'integració, encara que presenta dissipació de potència estàtica. La lògica BiCMOS aconsegueix una millora en la velocitat i en la capacitat de càrrega en dispositius que necessitin bons *drivers*, combinant transistors CMOS amb transistors BJT en les etapes de sortida del circuit.

Una altra de les lògiques que s'utilitza com a alternativa al disseny amb lògica CMOS és la lògica amb transistors de pas, que constitueix una bona alternativa per la millora de prestacions que ofereix, especialment en estructures específiques com multiplexors i memòries RAM o ROM, que precisen d'una molt bona optimització d'àrea.

La lògica CVSL (*Cascode Voltage Switch Logic*) no s'utilitza tant com les descrites anteriorment, però ofereix una bona velocitat de commutació i una gran flexibilitat com a conseqüència de proporcionar com a sortida una funció i la seva complementària. Aquest fet, però, va acompanyat d'un augment d'àrea i, per tant, d'una disminució en la densitat d'integració.

Existeixen altres lògiques, com la CPL (*Complementary Pass-transistor Logic*) i la DSL (*Differential Split Level*), que són versions millorades de la lògica CVSL. Una altra lògica alternativa a la CMOS és la SFPL (*Source Follower Pull-up Logic*), que es caracteritza per basar-se en una estructura semblant a la lògica pseudo-nMOS. També existeixen lògiques dinàmiques com a alternativa, però aquestes no són objecte d'aquest text.

2. LÒGICA PSEUDO-nMOS

2.1. Funcionament

La lògica pseudo-nMOS és una estructura formada per un bloc nMOS, que realitza la funció lògica desitjada, i un únic transistor pMOS amb el contacte de porta polaritzat permanentment a GND per generar l'1 lògic.

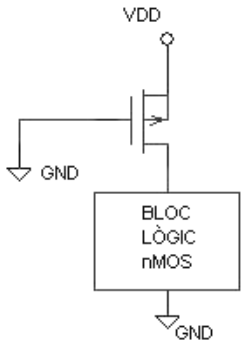


Fig. 1 Circuit basat en lògica pseudo-nMOS.

Aquesta lògica presenta un avantatge sobre la lògica CMOS, i és que es necessiten n+1 transistors per implementar una funció lògica de n variables. Així doncs, ocupa menys àrea i, per tant, ofereix major densitat d'integració i menor capacitat de càrrega. A més, aquesta lògica és compatible amb la CMOS.

En contraposició, a diferència de les estructures realitzades amb lògica CMOS complementària, té consum estàtic, ja que el pMOS està conduint permanentment.

El procés de disseny mitjançant la utilització d'estructures pseudo-nMOS resulta més complex si es desitja que V_{OL} sigui acceptablement baix, ja que aquest valor depèn de les dimensions dels transistors nMOS i pMOS (β_n/β_p), i com a conseqüència les seves característiques dinàmiques són asimètriques ($t_{pHL}<<t_{pLH}$).

Atès que el transistor pMOS està sempre conduint, $V_{GSp}=V_{DD}$. En conseqüència, V_{OL} no és 0, però es pot aconseguir un valor raonable dimensionant els transistors de forma adequada. Això implica una major complexitat en el disseny ja que, en un inversor, per exemple, per aconseguir un valor de V_{OL} prou petit, és necessari que $\beta_n/\beta_p >>1$. Concretament:

$$\frac{\beta_n}{\beta_p} = \frac{\left(V_{DD} - \left|V_{Tp}\right|^2\right)}{2\left(V_{DD} - V_{Tn}\right)V_{OL} - V_{OL}^2} \quad (1)$$

2.2. Aplicacions

Es poden trobar estructures realitzades amb lògica pseudo-nMOS principalment en dispositius com PLAs i ROMs. Amb aquesta lògica també se solen realitzar estructures de tipus NOR amb elevat *fan-in*.

3. LÒGICA BiCMOS

Aquest tipus de lògica està orientat principalment a incrementar la capacitat de càrrega a la sortida d'una porta realitzada amb lògica CMOS complementària. Això és possible mitjançant l'ús de transistors BJT. Aquest tipus de lògica s'empra típicament en circuits que requereixen bones prestacions com a *drivers*.

3.1. Funcionament

Com a exemple es descriu el funcionament del circuit de la Fig. 2.

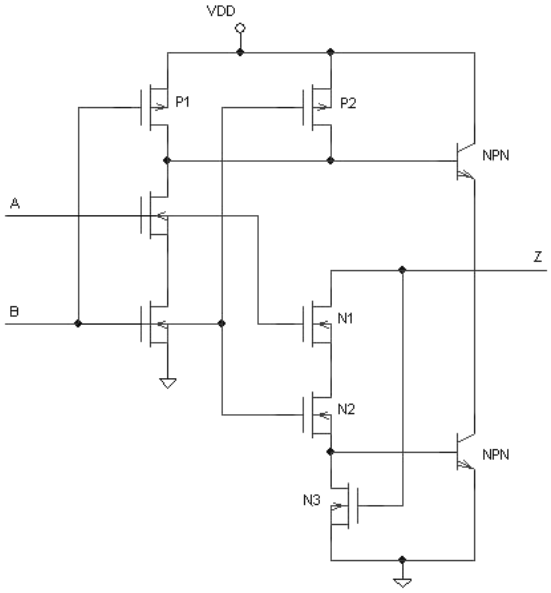


Fig. 2 Circuit exemple basat en lògica BiCMOS.

Els transistors N_1 i N_2 alimenten el transistor NPN *pull-down* quan l'entrada es troba a nivell alt, i N_3 fa el mateix quan la que es troba a nivell alt és la sortida. De la mateixa manera, P_1 i P_2 alimenten el transistor NPN *pull-up*.

Una estructura diferent a l'anterior és la que es mostra a la Fig. 3. En aquest cas es repeteix l'estructura NAND dels transistors nMOS al *path* de *pull-down* del circuit amb l'objectiu d'aconseguir un bon nivell V_{OL}, encara que no és convenient tenir molts transistors en sèrie per tal d'evitar que la velocitat de *pull-down* sigui superior a la de *pull-up*.

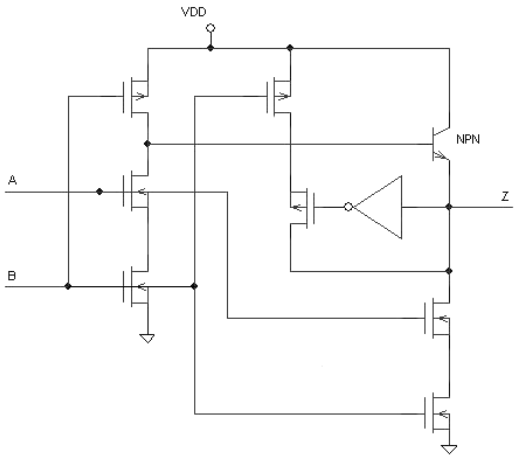


Fig. 3 Circuit exemple basat en lògica BiCMOS.

Un altre tipus d'estructura consisteix simplement en realitzar la lògica del circuit amb lògica CMOS i utilitzar com a *drivers* de sortida transistors BJT.

3.2. Aplicacions

La lògica BiCMOS se sol utilitzar en aplicacions on es requereixi un bon comportament del circuit com a *driver*, ja sigui aquest de bus o bé d'entrada/sortida. També té aplicació en xips amb senyal mixt, amplificadors operacionals de grans prestacions i altres circuits lineals.

4. LÒGICA AMB PORTES DE PAS

4.1. Funcionament

Una porta de pas és equivalent a un commutador controlat per tensió. Per dissenyar-la, generalment es disposa un transistor nMOS i un transistor pMOS en paral·lel, doncs així s'aconsegueix una bona transmissió de 0 (nMOS) i 1 (pMOS) lògics. Mitjançant aquest tipus de lògica es pot implementar qualsevol funció. L'estructura d'una porta de pas s'il·lustra a la Fig. 4 i un model simple del seu comportament es mostra a la Fig. 5.

La resistència associada a la porta de pas (R_{TG}), encara essent una funció no lineal, pot aproximar-se a un valor constant determinat per:

$$R_{TG} = \frac{V_{TG}}{I_{Dn} + I_{Dp}} \qquad (2)$$

on V_{TG} és el voltatge. També es pot estimar el valor de la resistència equivalent de la porta de pas com el paral·lel de la resistència dels transistors nMOS i pMOS. La capacitat de sortida C_L es pot aproximar per l'expressió següent:

$$C_L = \left(C_{GDn} + C_{GSp}\right) + k\left(V_l\right)\left[C_{DBn} + C_{SBp}\right] + C_{wire} + C_{in} \quad (3)$$

on C_{in} és la capacitat de la cel·la anterior.

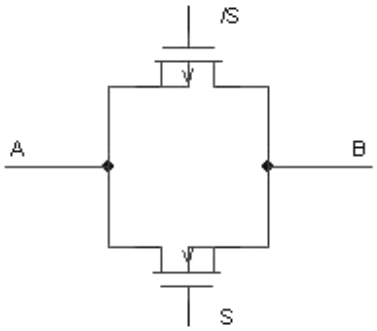


Fig. 4 Símbol d'una porta de pas.

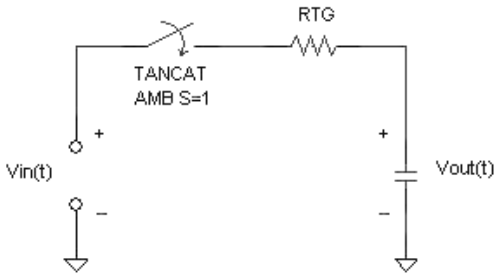


Fig. 5 Esquema elèctric equivalent d'una porta de pas.

La lògica que se serveix de portes de pas aconsegueix avantatges en velocitat, sempre que les estructures realitzades limitin els transistors en sèrie a un nombre raonable. Si s'implementen circuits amb portes de pas a partir de portes NAND i NOR en estructura CMOS, és possible reduir el nombre de transitors i, per tant, millorar l'àrea i les prestacions.

Com a desavantatge, però, cal considerar el caràcter bidireccional de la porta de pas, que comporta un procés de redistribució de càrregues, ocasionant variacions de tensió en el senyal transmès. Un altre inconvenient de les estructures dissenyades amb portes de pas és que els marges de soroll empitjoren quan s'implementen funcions complexes.

4.2. Aplicacions

Malgrat que la lògica CMOS és la més utilitzada en el disseny de sistemes VLSI, la lògica amb portes de pas ofereix més flexibilitat i millors prestacions en alguns casos concrets, com per exemple en multiplexors de portes XOR i XNOR. També és possible obtenir avantatges en la implementació de *latches* i *flip-flops*.

Una altra gran utilitat que presenten les portes de pas és que poden emprar-se com a elements *tri-state* de propòsit general.

5. CVSL

La lògica CVSL és aquella lògica diferencial que utilitza i genera senyals lògics, i també els seus complementaris. Permet implementar qualsevol funció.

Està formada per dos blocs nMOS complementaris, que proporcionen la funció de sortida, i per dos transistors pMOS creuats, que formen un *latch*. A la vegada, aquests fan la funció de *pull-up* de les sortides. La interconnexió entre els transistors pMOS proporciona un increment de la velocitat de commutació del circuit, encara que no necessàriament superior a l'obtinguda utilitzant lògica CMOS convencional.

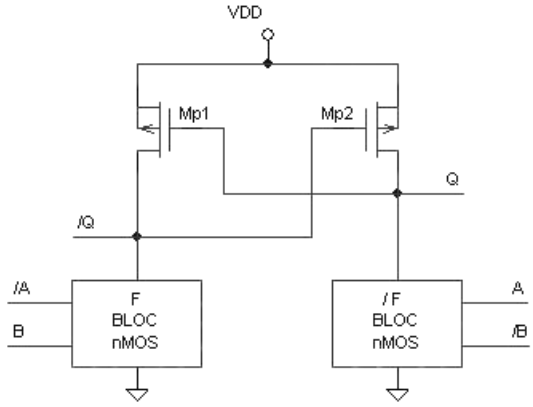


Fig. 6 Circuit basat en la lògica CVSL.

5.1. Funcionament

Se suposa el cas més simple, que consisteix en tenir un primer transistor nMOS al bloc F (M_{n1}) i un segon al bloc /F (M_{n2}). Quan el transistor M_{n1} té un 0 a l'entrada, es troba tallat, i el transistor M_{n2} condueix. Per tant, a la sortida del bloc /F es té un 0 i el transistor M_{p1} condueix, a diferència del M_{p2}, que es troba en tall. Quan l'entrada s'inverteix, la cel·la canvia el seu estat.

Normalment, els dos transistors pMOS es dissenyen amb la mateixa β_p. Cal tenir en compte que es tenen valors grans de $(W/L)_p$ i, per tant, es poden aconseguir capacitats de càrrega ràpides. Aquest fet, però, incideix negativament en la commutació entre els transistors pMOS.

En funcions complexes, els blocs F i /F poden tenir característiques de commutació diferents, circumstància que provoca que la resposta transitòria i la potència dissipada siguin pitjors del que s'hauria d'esperar.

El principal avantatge d'aquesta lògica respecte la lògica CMOS convencional és la seva flexibilitat, com a conseqüència de generar la funció de sortida i la seva complementària. Això mateix constitueix un inconvenient quant a àrea. A més, cal considerar el major consum de la cel·la, ja que en la commutació hi ha sempre un pMOS que condueix. Per la mateixa raó es produeix un retard en la commutació i tot el circuit és més sensible al soroll.

5.3. Aplicacions

Aquesta lògica s'utilitza per aconseguir dissenys que requereixin una alta velocitat de treball.

6. CPL

Constitueix una versió millorada de la lògica CVSL, ja que en minimitza els inconvenients. Presenta una estructura similar a l'anterior (veure Fig. 7), però elimina els transistors pMOS amb la finalitat d'incrementar la velocitat de commutació. Incorpora *buffers* inversors a les sortides que actuen com a restauradors de nivell. Els *latches* pMOS poden ser usats a les sortides per reduir el consum.

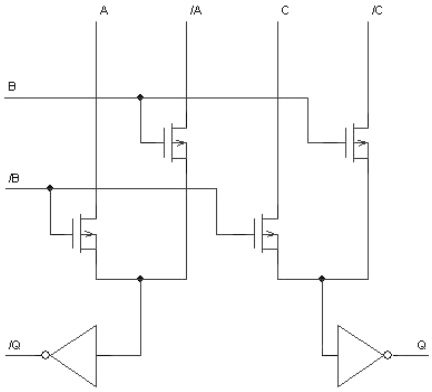


Fig. 7 Circuit basat en lògica CPL.

Un dels problemes d'aquesta lògica és la cadena de MOSFETs que es forma, que actua com a línia RC, i que ha de carregar-se i descarregar-se.

7. DSL

Aquesta és una altra millora de la lògica diferencial CVSL. S'afegixen dos transistors nMOS entre la sortida i la seva complementària i també entre els nodes interns.

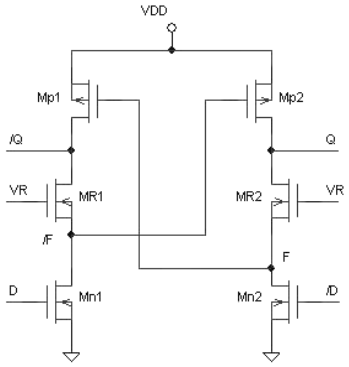


Fig. 8 Circuit basat en lògica DSL.

Presenta diversos avantatges, com són millors prestacions quant a commutació i una reduïda tensió *drain-source*, que permet utilitzar transistors més petits. Com a inconvenient, cal tenir en compte que la tensió de control dels transistors nMOS introduïts produeix dissipació de potència.

8. SFPL

Aquest tipus de portes es basen en una estructura semblant a la lògica pseudo-nMOS, excepte pel *pull-up*, que es controla amb les entrades. Millora la velocitat de la porta i ofereix millors prestacions per estructures amb un elevat *fan-in*, a costa d'augmentar el consum DC.

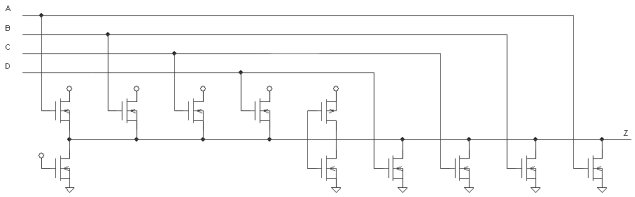


Fig. 9 Circuit basat en lògica SFPL.

9. CONCLUSIONS

Les lògiques estàtiques que s'han presentat en aquest text apareixen com una alternativa a la lògica CMOS complementrària convencional, tot procurant aconseguir millors prestacions pel que fa a l'àrea, el consum i la velocitat de commutació.

Les més utilitzades són la lògica BiCMOS i les portes de pas. La lògica pseudo-nMOS no constitueix una bona alternativa ja que, malgrat aconseguir reduir l'àrea del circuit a n+1 transistors, presenta consum estàtic.

Altres lògiques alternatives són la CVSL i les seves versions millorades, la CPL i la DSL. La primera es caracteritza per la seva elevada flexibilitat, ja que genera senyals lògics i els seus complementaris, però com a inconvenient presenta àrea, consum i retard en la commutació elevats. Per intentar solventar aquests problemes apareixen les lògiques CPL i DSL, que aonsegueixen incrementar la velocitat de commutació, però també agumenten el consum.

Finalment, s'ha estudiat la lògica SFPL, que presenta una estructura semblant a la lògica pseudo-nMOS, tot millorant-ne la velocitat a costa d'augmentar el consum.

10. REFERÈNCIES

[Weste93] N. H. I. Weste, K. Esharagian, “Principles of CMOS VLSI design: a systems perpective”, Addison-Wesley, 2a edició, 1993.

[Uyemura92] J. P. Uyemura, “Circuit design for CMOS VLSI”, Kluwer Academic Publishers, 1992.

[Bota96] S. Bota, J. Carrabina, A. Herms, “Introducció al disseny CMOS VLSI”, Textos docents núm 52, Edicions de la Universitat de Barcelona, 1996.

[Rubio00] A. Rubio, J. Altet, X. Aragonés, J. L. González, D. Mateo, F. Moll, “Disseny de circuits i sistemes integrats”, Edicions UPC, 2000.

9. Lògica dinàmica.

***Abstract*—**En aquest article es parlarà dels diferents tipus d'estructures que corresponen a la lògica dinàmica. Concretament, es descriuran les principals característiques de la lògica de precàrrega i avaluació, la lògica *dominó*, la lògica NP i la lògica *clocked* CMOS o C2MOS. També es detallaran els principals avantatges i inconvenients que presenten cadascuna d'aquestes estructures.

***Index Terms*—**Lògica dinàmica, precàrrega i avaluació, lògica *dominó*, lògica NP, lògica C2MOS.

1. INTRODUCCIÓ

La lògica dinàmica és aquella lògica en la qual pot representar-se un valor lògic mitjançant la tensió associada a la càrrega emmagatzemada en un node. D'ara en endavant, el node a estudiar serà el de sortida i la càrrega serà de tipus capacitiu. Aquesta capacitat de càrrega correspon a la suma de les capacitats de *drain* dels transistors nMOS i pMOS, de *gate* de la porta següent i de *wire*. Aquesta darrera serà menyspreable en el cas que ambdues portes es trobin molt properes entre si. Així doncs, la resposta dinàmica ve donada per les capacitats.

Hi ha diferents tipus de lògica dinàmica: de precàrrega i avaluació, *dominó*, NP i *clocked* CMOS o C2MOS. En els següents apartats se'n veuran les característiques.

2. PRECÀRREGA I AVALUACIÓ (PE)

És la família més senzilla, on el senyal de rellotge s'encarrega del control de la porta (veure Fig. 1). A la fase de precàrrega, el *clock* es troba a 0 i, per tant, és el transistor pMOS el que està actiu. En aquestes condicions, la capacitat de sortida comença a carregar-se fins que assoleix la tensió d'alimentació. La fase d'avaluació, en canvi, es té quan el *clock* està a 1. En aquest cas, s'avalua el bloc lògic nMOS i la sortida es posa a 0 o bé a 1 en funció de les entrades.

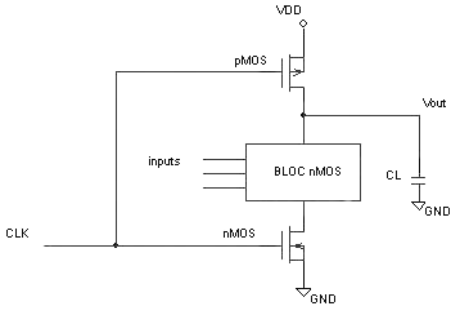


Fig. 1 Estructura basada en lògica PE.

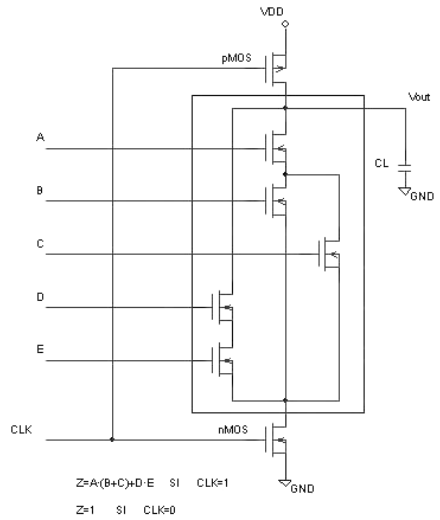


Fig. 2 Exemple de funció dissenyada amb lògica PE.

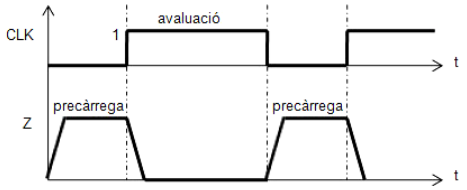


Fig. 3 Cronograma amb els senyals de *clock* i de sortida.

D'altra banda, en lloc d'utilitzar un bloc nMOS per analitzar la funció, pot fer-se servir un bloc pMOS. En aquest cas es variarà l'estructura i els transistors tindran major àrea a causa de la poca mobilitat dels forats.

Aquest tipus de lògica presenta una sèrie d'avantatges, com són la reducció de l'àrea de la porta fins al 50% aproximadament (només s'empren N+2 transistors), la millora del temps de pujada t_{LH} a causa de la precàrrega i la reducció de les capacitats, i un conseqüent augment de velocitat. A més, el consum estàtic es pràcticament 0.

Per contra, també es manifesten alguns inconvenients que s'oposen a la seva utilització. El més important de tots

ells és el problema de la redistribució de càrrega. Aquest fenomen es basa en el fet que si durant l'avaluació el primer transistor nMOS té un 1 com a entrada, també cal considerar la seva capacitat de *source*. Llavors, la tensió de sortida serà inferior, ja que s'haurà de repartir entre dues capacitats. És per aquest motiu que la sortida podria ser errònia. Una possible solució per evitar aquest problema és fer que la capacitat de *source* sigui molt menor que la de sortida.

Un altre inconvenient que es manifesta és que no es poden fer connexions en cascada, ja que la segona porta podria començar a avaluar-se en el moment en què la primera encara no ha lliurat un valor correcte a la seva sortida i tenir així un resultat final erroni. Se suposa que es tenen dues portes NAND connectades en cascada (veure Fig. 4), per exemple, i que la sortida de la primera ha de valdre 0 durant la fase d'avaluació. La descàrrega de la capacitat paràsita requereix un cert temps i, per tant, la segona porta NAND pot començar a avaluar-se amb un 1 en lloc d'un 0 a la seva entrada C. Si l'entrada del transistor D és un 1, la capacitat de sortida pot descarregar-se i a la sortida es tindrà un resultat incorrecte. Una possible solució és la utilització de senyals de rellotge retardats o la utilització de lògica *dominó*.

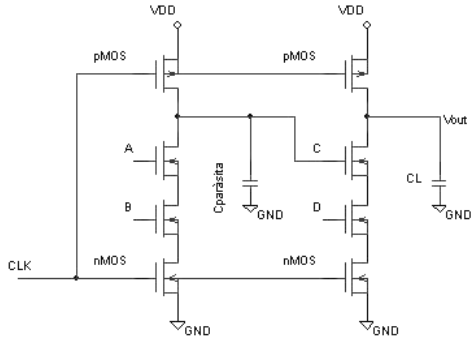


Fig. 4 Portes NAND basades en lògica PE connectades en cascada.

Altres problemes són l'existència d'un corrent de fuites com a conseqüència de l'existència d'una unió drenador-substrat en inversa i el *clock-feedthrough*, fenomen que produeix un sobrepic en la tensió de sortida com a conseqüència de l'existència d'una capacitat paràsita entre porta i drenador del transistor p.

3. LÒGICA *DOMINÓ*

Aquest tipus de lògica presenta una estructura i un funcionament igual a la lògica de precàrrega i avaluació, però afegint un inversor a la sortida. Aquest inversor soluciona els problemes de redistribució de càrrega i permet fer connexions en cascada. La fase de precàrrega de la primera porta sempre donarà un 0 com a sortida out_1, evitant així que el primer transistor del bloc lògic nMOS de la segona porta s'activi quan no ha de fer-ho.

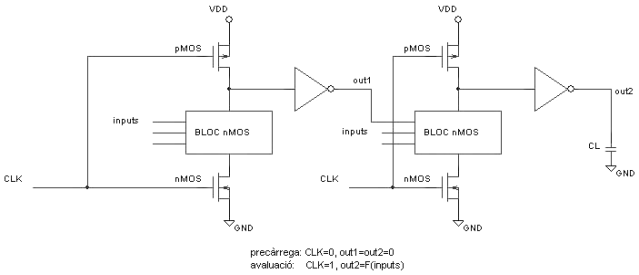


Fig. 5 Estructura amb lògica *dominó*.

La lògica *dominó*, com a conseqüència de la presència d'un inversor addicional per cada etapa, presenta un augment de l'àrea del xip i també dels retards. A més, tan sols poden implementar-se funcions no inversores. Una possible solució és l'aplicació de les lleis de Morgan per passar de NOR a AND i de NAND a OR pel cas d'entrades negades.

4. LÒGICA NP

Aquest tipus de lògica permet fer connexions en cascada sense haver d'incorporar l'inversor, ja que evita les carreres, és a dir, que una porta s'activi quan no és el moment.

A la segona etapa, és un bloc lògic pMOS el que analitza la funció. Per aquest motiu, el retard en la transició d'1 a 0 a la sortida de la primera etapa no produeix resultats erronis. Malgrat aquesta millora, presenta els inconvenients de ser un sistema més lent a causa de la incorporació de transistors pMOS i de necessitar un senyal de rellotge amb dues fases (el *clock* i el seu complementari).

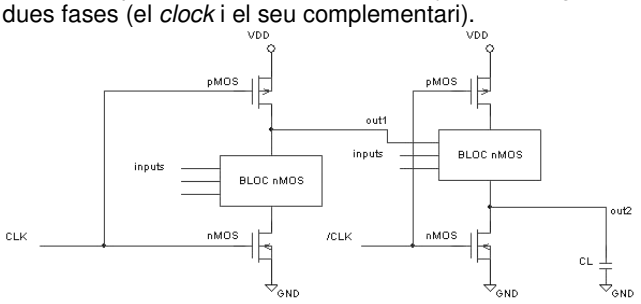


Fig. 6 Estructura amb lògica NP.

5. LÒGICA C2MOS

Aquesta lògica, també anomenada *clocked* CMOS, consisteix en una xarxa estàtica convencional juntament amb dos transistors nMOS i pMOS controlats per un senyal de *clock* i el seu complementari, respectivament. Presenta un avantatge molt important respecte totes les lògiques anteriors ja que minimitza el consum dinàmic que apareix

en la lògica CMOS complementària. Quan les entrades canvien, la sortida es troba en alta impedància (el *clock* és 0). En aquestes condicions no hi ha pas de corrent entre V_{DD} i GND i, per tant, tampoc consum dinàmic. En l'actualitat, però, el consum de la porta CMOS s'ha reduït molt i la utilització de la porta C2MOS ja no és necessària. A més, aquesta porta té una major àrea com a conseqüència de la presència de més transistors.

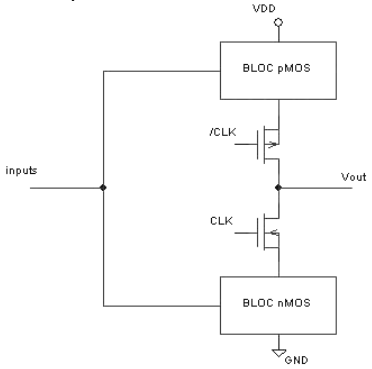


Fig. 7 Estructura amb lògica C2MOS.

6. CONCLUSIONS

La lògica dinàmica, generalment, només utilitza una xarxa de transistors a cada etapa per tal d'obtenir una funció. Així doncs, empra menys transistors que en el cas de la lògica estàtica convencional. Això últim implica menor àrea i menor retard, pel que és recomanable l'ús d'aquesta lògica en sistemes que requereixin altes prestacions en velocitat.

La lògica de precàrrega i avaluació, que està governada per un senyal de rellotge, proporciona el resultat de la funció després de cada cicle de dues etapes. A la primera, es precarrega la sortida a 1 i a la segona, s'avalua la funció en si. Aquesta estructura, però, presenta una sèrie d'inconvenients, d'entre els quals destaquen el problema de la repartició de càrrega i el fet que no es poden realitzar connexions en cascada, pels quals se'n desaconsella l'ús. La lògica dominó és una versió millorada de l'anterior que, pel sol fet d'incloure un inversor a la sortida de cada etapa, aconsegueix eliminar-ne els problemes.

La lògica NP alterna la utilització d'una etapa on el bloc que realitza l'anàlisi és tipus n amb una de tipus p. Així se soluciona el problema de la redistribució de càrrega sense haver d'afegir inversors i, per tant, sense haver d'augmentar l'àrea. L'última lògica estudiada, la C2MOS, destaca per minimitzar el consum dinàmic.

7. REFERÈNCIES

[Diéguez] A. Diéguez, "Lògica dinàmica", Transparències de l'assignatura de Disseny Microelectrònic I.

10. Disseny de circuits seqüencials.

***Abstract*—**Aquest text és una introducció al disseny de circuits seqüencials. Concretament, es donarà una visió general dels circuits bàsics més emprats, que són *latches* i *flip-flops*, i s'explicarà la dinàmica de funcionament de cadascuna de les seves variants, tot detallant-ne l'equació característica. Finalment, es definiran els paràmetres de restricció temporal temps de *set-up* i temps de *hold.*

***Index Terms*—**Senyal d'excitació d'entrada, estat actual, circuit seqüencial síncron/asíncron, *flip-flop*, *latch*, unió pn, efecte *leakage*, *buffer*, temps de *set-up*, temps de *hold*.

1. INTRODUCCIÓ

Un circuit seqüencial està determinat pels senyals d'entrada i per l'estat actual, tot i que l'estat de transició i l'estat següent també tenen una importància rellevant. Cal tenir en compte, però, que els circuits seqüencials poden ser de dos tipus: síncrons o bé asíncrons. Els primers, a diferència dels segons, estan controlats per un senyal de rellotge.

Una representació d'aquest model de circuit seqüencial es troba a la Fig. 1, on les $x_1,\ldots,x_n$ són l'entrada, les $z_1,\ldots,z_m$ són la sortida, les $y_1,\ldots,y_r$ són l'estat present i les $Y_1,\ldots,Y_r$ representen l'estat següent.

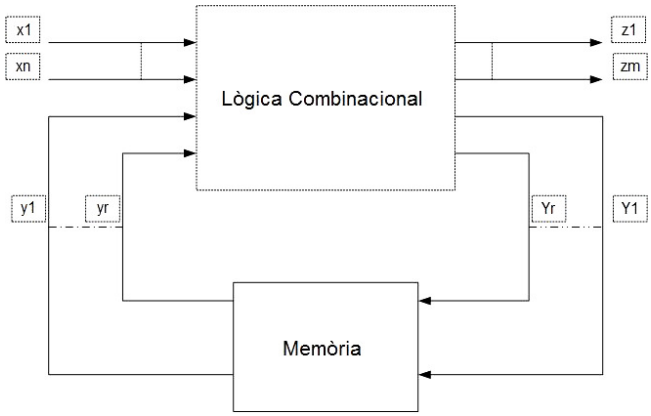


Fig. 1 Model de circuit lògic seqüencial.

Els dispositius de memòria del diagrama de blocs de la Fig. 1 poden ser de diversos tipus. Aquí es veuran dispositius com *flip-flops* o *latches*.

2. DISPOSITIUS DE MEMÒRIA

Una part important d'un circuit seqüencial és la unitat de memòria, que s'encarrega d'emmagatzemar les dades binàries. Aquesta unitat de memòria passa a l'estat 0 per emmagatzemar un 0 i a l'estat 1 per emmagatzemar un 1. La sortida $y_1,\ldots,y_r$ del circuit indica l'estat present de la memòria.

Els dos tipus de memòria normalment utilitzats són els *latches* i els *flip-flops*. L'estat d'un *latch* ve controlat pels senyals d'excitació d'entrada. Si la sortida es veu obligada a assumir el valor 1, el dispositiu és un *latch set*. Si, en canvi, la sortida assumeix el valor 0, el dispositiu és un *latch reset*. També existeix la possibilitat que coexisteixin els senyals *set* i *reset*. En aquest cas, es tractarà d'un *latch set-reset*.

Un *latch* canvia d'estat immediatament, segons els seus senyals d'excitació d'entrada. Un *flip-flop*, en canvi, està governat per un senyal de control anomenat rellotge, al qual ha d'esperar abans de canviar d'estat. L'estat final d'un *flip-flop* ve determinat pels valors que prenen els senyals d'excitació en l'instant en què es presenta el senyal de rellotge.

3. *LATCHES*

3.1. *Latch* simple

Es considera la porta OR de la Fig. 2(a), suposant que ambdues entrades es troben a l'estat 0 lògic. Si es connecta la sortida de nou a una de les entrades (Fig. 2(b)), la porta s'estabilitza amb una sortida de 0. Si es connecta un 1 lògic a l'entrada no retroalimentada S, la porta canvia la seva sortida a un 1 lògic. Si es torna a canviar l'entrada S a un 0 lògic, la sortida continua essent un 1 lògic. Aquest dispositiu s'anomena *latch set*. Igualment es podria fer un *latch reset*, que s'estabilitza a un valor de 0 lògic.

Els dispositius que s'estabilitzen a un sol estat no són molt útils, excepte en situacions de disseny poc usuals. Combinant les característiques dels dos *latches* anteriors, s'obté el *latch set-reset* o *latch SR*.

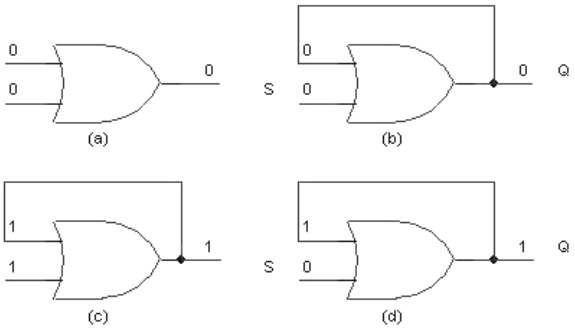


Fig. 2 Retroalimentació en un *latch* simple.

3.2. *Latch* dinàmic CMOS

Aquest *latch* simple es pot realitzar de moltes maneres, però aquí s'estudiarà l'estructura a nivell de transistors del que està format per una porta de pas CMOS amb un inversor a la sortida, tal com es pot observar a la Fig. 3. Aquest sistema és desaconsellable, excepte quan es treballa amb senyals elevats de rellotge. El motiu és la falta d'alimentació de la capacitat d'entrada de l'inversor, que guarda la dada, ja que una vegada aquesta s'ha emmagatzemat, pot descarregar-se a través de les unions pn paràsites (efecte *leakage*) o per l'efecte d'interferències externes.

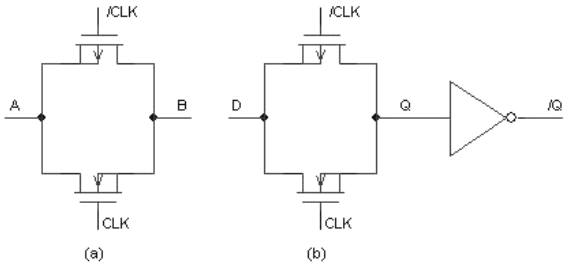


Fig. 3 Porta de pas CMOS (a) i *latch* dinàmic CMOS (b).

3.3. *Latch D*

Per solucionar el problema del *latch* anterior, es realimenta el nus Q a partir de la sortida /Q, tal com es veu a la Fig. 4(a). D'aquesta forma, es restaura el valor del nus que guarda la dada i es minimitza l'efecte *leakage*. Aquesta estructura constitueix el que és conegut com a *latch D*.

L'estructura de realimentació es repeteix en gairebé totes les cel·les de memòria. Les dimensions de l'inversor de realimentació han de ser menors que les de l'inversor de sortida, per contrarestar així la pròpia realimentació. També s'han de tenir en compte les dimensions dels transistors de la porta de pas, per no dificultar la captura d'una nova dada des de l'entrada.

Els problemes sobre el dimensionament de l'inversor de realimentació es poden reduir si es col·loca una segona porta de pas a la sortida d'aquest inversor, de tal manera que sincronitzi la realimentació (veure Fig. 4(b)). Es recomana utilitzar *buffers* de sortida per evitar efectes relacionats amb la redistribució de càrrega al nus de sortida (veure Fig. 4(c)). L'expressió que permet obtenir el valor de sortida d'aquest dispositiu és la que es mostra a continuació.

$$Q^* = DC + \overline{CQ} \tag{1}$$

Una altra possible implementació del *latch D* és amb inversors *tristate*. El seu funcionament és idèntic a l'estructura formada per un inversor seguit d'una porta de pas, però presenten menys capacitats paràsites i un disseny més compacte.

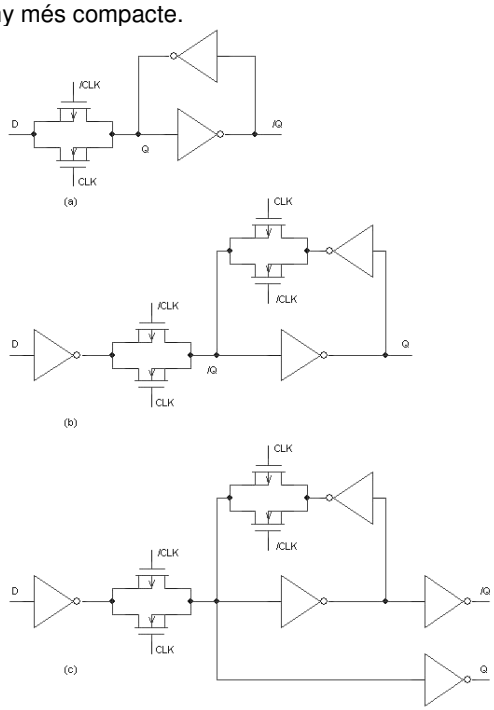


Fig. 4 *Latch D* amb realimentació (a), amb realimentació sincronitzada (b) i amb realimentació sincronitzada i *buffers* a la sortida (c).

3.4. *Latch SR*

Un *latch SR* disposa d'una entrada S realimentada a la sortida, que normalment dóna un 1 lògic, i d'una entrada R, que permet realitzar un *reset* al dispositiu posant la sortida a 0 lògic. Això es pot realitzar fàcilment mitjançant portes simples NOR i NAND, que són fàcils de realitzar amb tecnologia CMOS. Una altra forma d'implementar els *latches* és amb un senyal d'activació, que inhibeix les entrades o les activa segons convingui (veure Fig. 5(c)).

Les expressions que es presenten tot seguit descriuen el funcionament d'un *latch SR* i d'un *latch SR* amb senyal d'activació, respectivament.

$$Q^* = S + \overline{SR} \tag{2}$$

$$Q^* = SC + \overline{QR} + \overline{CQ} \tag{3}$$

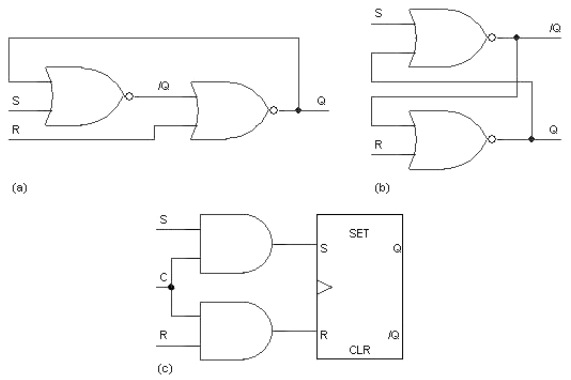


Fig. 5 *Latch SR* (a), *latch SR* (b) i *latch SR* amb senyal d'activació (c).

4. *FLIP-FLOPS*

Els circuits vistos fins ara no són adequats per ser usats en circuits seqüencials síncrons. Quan el senyal d'activació està actiu, les entrades d'excitació s'envien directament a la sortida Q. Per això, qualsevol canvi a l'entrada d'excitació provoca un canvi immediat a la sortida del *latch*. Es resoldrà aquest problema amb un senyal de control de temps anomenat rellotge, que limita els instants en què poden canviar els estats dels elements de memòria.

4.1. *Flip-flop SR*

Una solució al problema d'inestabilitat anterior és utilitzar dos *latches* en configuració mestre-esclau (veure Fig. 6). Els senyals d'activació dels dos *latches* són controlats per senyals complementaris de rellotge. Quan el senyal CLK està a nivell baix, el *latch* mestre està actiu i l'esclau en retenció. Els canvis dels senyals d'excitació S i R són enviats cap al *latch* mestre, mentre que el *latch* esclau ignora qualsevol canvi en les seves entrades.

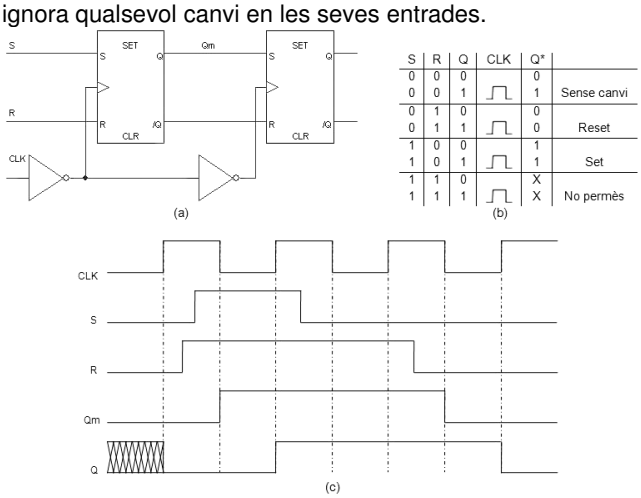


Fig. 6 Esquema d'un *flip-flop SR* (a), taula d'excitació (b) i diagrama temporal (c).

Quan el rellotge canvia a nivell alt, els dos *latches* intercanvien les seves funcions. Aquesta combinació de *latches* s'anomena *flip-flop SR* o *flip-flop master-slave*. Es vol fer notar que aquí la combinació S=R=1 no està permesa.

A partir de la taula d'excitació (veure Fig. 6(b)), es dedueix l'equació característica del *flip-flop SR*:

$$Q^* = S + \overline{RQ} \tag{4}$$

Si aquest dispositiu s'utilitzés en un circuit seqüencial síncron, no podria haver-hi cap oscil·lació inestable ja que, en tot moment, el *latch* mestre i el *latch* esclau estarien en mode de retenció, bloquejant el comportament inestable.

4.2. *Flip-flop D*

Es pot construir un *flip-flop D* mitjançant dos *latches D* units en estructura mestre-esclau, tal com es mostra a la Fig. 7. A la taula d'excitació d'aquesta mateixa figura, s'aprecia que quan CLK=0, el mestre està activat, de manera que la seva entrada passa a l'esclau. En la transició de 0 a 1 del CLK, el mestre captura el valor d'entrada de D i el conserva. Quan CLK=1, l'esclau es connecta i el valor capturat pel mestre passa a la sortida Q. En la transició de 1 a 0 del CLK, l'esclau captura la dada del mestre.

El comportament del *flip-flop D* es pot resumir dient que Q assumirà el valor de D en la transició ascendent del rellotge (veure equació (5)). Aquest procés també es pot realitzar per la transició descendent del *flip-flop* i en aquest cas es diria que el dispositiu és actiu per flanc de baixada.

$$Q^* = D \tag{5}$$

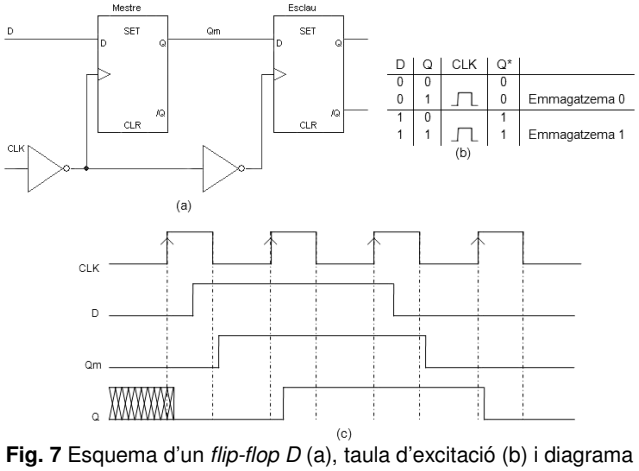


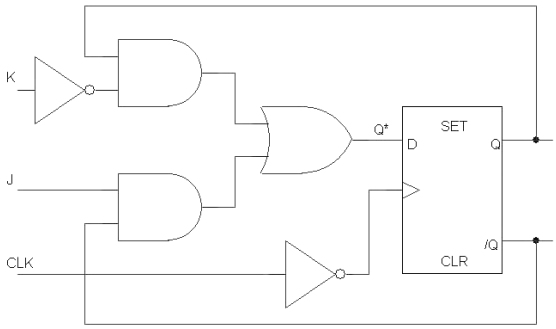
Fig. 7 Esquema d'un *flip-flop D* (a), taula d'excitació (b) i diagrama temporal (c).

4.3. *Flip-flop JK*

El *flip-flop JK* opera de la mateixa manera que el *flip-flop SR*, tot fent l'assignació J=S i K=R. No obstant això, la

combinació S=R=1, que al SR no està permesa, aquí s'utilitza pel mode d'operació alternat. L'estat es canvia de 0 a 1 i de 1 a 0 quan J=K=1. A continuació es presenten l'equació característica, l'estructura i la taula d'excitació d'aquest dispositiu.

$$Q^* = \overline{K}Q + J\overline{Q} \tag{6}$$



(a)

J	K	Q	CLK	Q*	
0	0	0		0	
0	0	1	⎍	1	Sense canvi
0	1	0		0	
0	1	1	⎍	0	Reset
1	0	0		1	
1	0	1	⎍	1	Set
1	1	0		1	
1	1	1	⎍	0	Alternança

(b)

Fig. 8 Esquema d'un *flip-flop JK* (a) i taula d'excitació (b).

4.4. *Flip-flop T*

El *flip-flop T* dóna com a estat futur de sortida el complementari de l'estat present només quan l'únic senyal d'excitació d'entrada (T) és 1. Una configuració senzilla d'aquest dispositiu s'obté a partir d'un *flip-flop JK*, tot imposant J=K=1 (veure Fig. 9(a) i Fig. 9(b)). Algunes versions de *flip-flop T* operen sota el control dels polsos de rellotge (Fig. 9(c)).

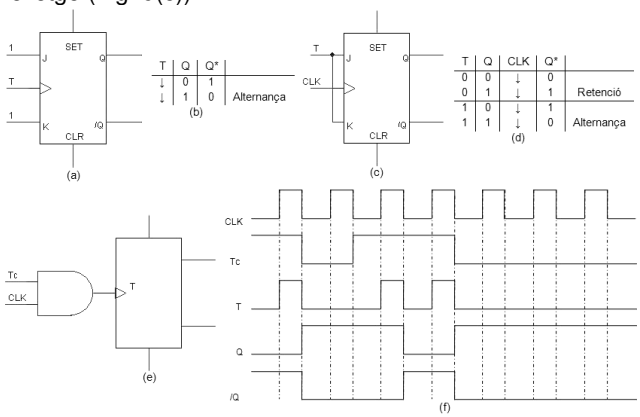


Fig. 9 Esquema d'un *flip-flop T* a partir d'un *flip-flop JK* (a), taula d'excitació (b), *flip-flop T* operant sota control de rellotge (c), taula d'excitació (d), esquema de configuració de *flip-flop T* sota control de rellotge de forma selectiva (e) i diagrama de temps (f).

El funcionament d'aquests dos tipus de *flip-flop T* ve descrit per les expressions següents.

$$Q^* = \overline{Q} \tag{7}$$

$$Q^* = T\overline{Q} + \overline{T}Q \tag{8}$$

Una estructura comuna als circuits lògics seqüencials que compta els polsos en una línia de senyal és el *flip-flop T*. S'utilitza amb freqüència per construir mòduls comptadors.

5. ESQUEMA DE TEMPS EN *LATCHES* I *FLIP-FLOPS*

5.1. Temps en *latches*

Per garantir que un valor específic a l'entrada d'excitació D determini el valor final del *latch*, D no pot variar al voltant de l'instant en què el senyal d'activació passa d'alt a baix. Es defineixen dos paràmetres de restricció de temps per cada *latch*, amb la finalitat d'identificar quan ha de romandre constant l'entrada per tal de garantir el bon fucionament del dispositiu.

Es defineix temps de configuració o de *set-up* del *latch* (t_{su}), com el període immediatament anterior a la transició del senyal d'activació, durant el qual l'excitació d'entrada ha de ser estable. Es pot afirmar, doncs, que l'excitació d'entrada ha de ser configurada almenys un temps t_{su} abans de la transició del senyal d'activació.

D'altra banda, també es defineix el temps de retenció o de *hold* del *latch* (t_h) com el període de temps immediatament posterior a la transició del senyal d'activació, durant el qual D no pot canviar. Per tant, l'entrada d'excitació ha de romandre constant almenys un temps t_h després de la transició del senyal d'activació per garantir que el *latch* mantindrà el valor correcte.

Aquests paràmetres es poden veure representats a la Fig. 10, on s'han utilitzat temps de propagació nuls per simplificar el diagrama. S'observen dues violacions de temps. Un canvi en D durant la transició de 0 a 1 massa proper al pols del rellotge representa una violació del temps de *set-up*, i la sortida del *latch* podria no canviar de 0 a 1 tal com seria d'esperar. D'igual manera, un canvi en D durant la transició de 1 a 0 posterior al pols de rellotge, però molt proper a aquest, pot provocar un estat impredictible.

A més de les restriccions de temps, molts *latches* necessiten un ample de pols mínim a l'entrada activa per poder garantir un canvi correcte d'estat. Aquest ample de pols mínim s'anomena t_w. Qualsevol pols d'ample inferior a t_w no provocaria el canvi d'estat desitjat.

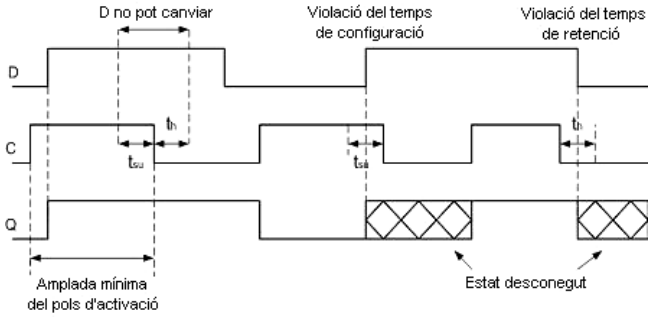


Fig. 10 Representació del temps de *set-up* i de *hold* en un diagrama de temps per un *latch D*.

5.2. Temps en *flip-flops*

El funcionament del temps en un *flip-flop SR* és semblant al d'un *latch*. Les definicions de temps de *set-up* i de *hold* són idèntiques.

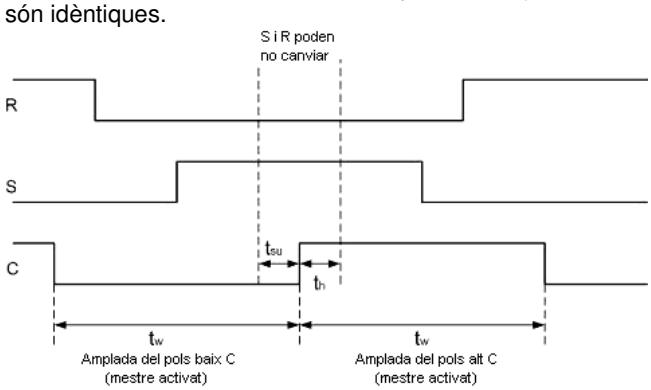


Fig. 11 Representació del temps de *set-up* i de *hold* en un diagrama de temps per un *flip-flop SR*.

Les entrades de S i R del *latch* mestre han de ser estables abans de la transició de rellotge que porta el mestre al mode de retenció. A la Fig. 11 es pot observar el diagrama de temps d'un *flip-flop SR*. Com que les entrades de configuració només afecten al *latch* mestre, es definiran els temps de *set-up* i de *hold* relatius a la transició ascendent del senyal de rellotge, que és la transició de rellotge en què canvia el mode del *latch* mestre. Les entrades d'excitació del *latch* esclau no es veuen afectades per les entrades d'excitació del dispositiu.

A més, el paràmetre mínim per l'ample del pols és l'ample mínim necessari pel funcionament adequat del *latch* mestre, mentre que el paràmetre màxim per l'ample de pols és l'ample mínim necessari pel *latch* esclau. La suma d'aquestes dues amplades de pols determina el període mínim de qualsevol senyal de rellotge que serà utilitzat pel *flip-flop*.

6. CONCLUSIONS

S'han vist els dispositius de memòria *latches* i *flip-flops*. Els primers s'utilitzen principalment en situacions en què les dades es capturen des de línies de senyal i es guarden. El *latch SR* captura polsos aleatoris a les seves entrades S i R, atès que a cada pols reestableix el seu estat. Els *latches SR* i *D* amb senyals d'activació canvien d'estat només als instants en què el *latch* està habilitat. Per tant, els *latches* amb senyals d'activació poden ser utilitzats per capturar els senyals que arriben i s'estabilitzen abans del final d'un pols d'activació.

Els *flip-flops*, en canvi, s'utilitzen per dissenys de circuits seqüencials on tots els canvis d'estat han de sincronitzar-se amb les transicions d'un senyal de rellotge. En la major part d'aquests circuits s'utilitzen *flip-flops JK* o *D*. Els *flip-flops SR* s'utilitzen poc, atès que els *flip-flops JK* proporcionen els mateixos modes d'operació i, a més, afegeixen el mode d'alternança. Així s'elimina el problema que es tenia en el cas S=R=1. Els *flip-flops T* s'utilitzen bàsicament en el disseny de comptadors.

Per garantir el bon funcionament de *latches* i *flip-flops* és imprescindible respectar els temps de *set-up* i de *hold*. El temps de *set-up* (*hold*) és l'interval de temps previ (posterior) a la transició del senyal d'activació durant el qual l'entrada ha de romandre constant. A més, aquests dispositius precisen d'una amplada mínima del pols per assegurar un estat futur correcte.

7. REFERÈNCIES

[Herms] A. Herms, S. Bota, S. Bosch, J. M. López Villegas, "Disseny Digital Bàsic", Textos docents núm 126, Edicions de la Universitat de Barcelona.

[Bota96] S. Bota, J. Carrabina, A. Herms, "Introducció al disseny CMOS VLSI", Textos docents núm 52, Edicions de la Universitat de Barcelona, 1996.

[Nelson96] V. P. Nelson, H. Troy, B. D. Carrol, J. D. Irwin, "Anàlisi i Disseny de Circuits Lògics Digitals," Prentice Hall, 1996.

[Wolf94] W. Wolf, "Modern VLSI Design", Prentice Hall, 1994.

[Uyemura94] J. P. Uyemura, "Circuit Design for CMOS VLSI", Kluwer Academia Publishers, 1994.

[Hollstein] T. Hollstein, M. Petrov, C. Schlachta, O. Soffke, "VLSI-Design of Integrated Circuits", Darmstadt University of Technology, Institute of Microelectronic Systems, http://www.microelectronic.e-technik.tu-darmstadt.de/lectures/winter/vlsi/vlsi_vorlesung_ef.html.

[Plusquellic04] J. Plusquellic, “Advanced VLSI Design”, http://www.cs.umbc.edu/~plusquel/.
[06/09/07] Escola Pértiga, “Apunts de Sistemes Electrònics”, http://www.pertiga.biz/oposicions/ejemse.pdf.

A4. Estructures avançades: lògica autotemporitzada.

***Abstract*—**Aquest document pretén ser una ràpida introducció als sistemes asíncrons. Es compara aquesta estructura amb els sistemes clàssics amb rellotge i es veuen els seus principals avantatges, així com els problemes que apareixen en implementar-la. Finalment, es veuen les possibles aplicacions de la lògica asíncrona en sistemes on, a priori, pugui produir millors resultats.

***Index Terms*—**Lògica síncrona, lògica asíncrona, rellotge, *glitch*, atzar, diagrama de Karnaugh, exclusió mútua, xarxa de Petri, *dual-rail signaling*, *transition signaling*, LEDR, *current-sensing completion detection*.

1. INTRODUCCIÓ

La majoria de dissenys lògics actuals es basen en dos punts fonamentals: tots els senyals són binaris i el temps és discret. El primer punt permet la utilització de la lògica de Boole sense cap mena de restricció. El segon, en canvi, implica que tots els fenòmens són transitoris. Així doncs, es poden menysprear els canvis d'estat indesitjats i les realimentacions, ja que el sistema s'espera que el senyal sigui correcte per tal d'avaluar-lo.

2. DISSENY

En la lògica asíncrona, les entrades del sistema són constantment avaluades per obtenir un resultat, de forma que són sempre importants. D'altra banda, en un sistema síncron les entrades només tenen sentit al flanc o nivell de rellotge en què es realitza l'avaluació. Això implica que s'han d'evitar els canvis a l'entrada en aquests instants de temps.

Es defineix *glitch* com una transició no final d'una sortida, és a dir, el valor que assoleix una variable abans d'arribar al valor final. Els atzars (*hazards*), es defineixen com la possibilitat de tenir un *glitch* en una entrada. Els atzars són un perill donada la naturalesa de contínua avaluació dels sistemes asíncrons, ja que aquests no poden distingir entre un atzar i dues transicions ràpides de les entrades.

Si es restringeixen les entrades a un canvi d'un sol bit cada vegada i s'assumeix que aquests canvis són prou lents com per estabilitzar la sortida del sistema abans de tornar a canviar, és possible implementar funcions lògiques arbitràries amb lògica combinacional i sense atzars.

Si es considera la funció F=AB+/BC, per exemple, en una transició de l'entrada ABC=111 a ABC=101, el primer terme esdevé fals mentre que el segon esdevé vertader. En funció de les característiques de retard del circuit, les dues entrades de la porta OR final poden ser momentàneament falses i produir un 0 momentani a la sortida.

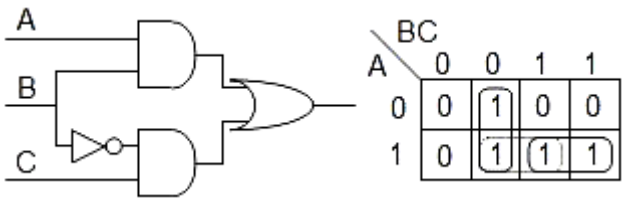


Fig. 1 Taula de Karnaugh i implementació de la funció amb atzars.

Els atzars es poden entendre fàcilment analitzant un diagrama de Karnaugh. Són aquelles sortides que presenten un mateix valor, però que provenen de termes diferents. Els atzars que realitzen una sola transició invàlida abans d'arribar al seu valor final s'anomenen atzars estàtics. Si la sortida ha de valdre 1 però presenta una transició temporal a 0, es té un atzar estàtic 1. En cas contrari, es té un atzar estàtic 0. També existeixen els atzars dinàmics, on el senyal que ha de fer un únic canvi de valor en fa tres o més abans d'assolir el valor final.

Cal tenir en compte que si es permet una variació en més d'un bit a l'entrada no es poden eliminar els atzars. En el cas que es vulgui que el sistema pugui canviar més d'una de les seves entrades a la vegada, es pot fer una lectura diferent de la convencional. Si un circuit no veu de la mateixa manera una transició en una variable d'entrada que en una altra i es permet la variació de les dues variables, s'ha de fer el possible perquè el circuit només vegi un dels canvis, és a dir, s'ha de forçar una exclusió mútua entre les entrades. Per aconseguir-la s'utilitzen els anomenats àrbitres, que forcen l'exclusió mútua entre les entrades del sistema a nivell funcional. Habitualment, els circuits d'arbitratge es basen en el circuit d'exclusió mútua de la Fig. 2. Bàsicament, aquest bloc és un *flip-flop* amb un detector analògic de metaestabilitat a la sortida.

En aplicar dos canvis a l'entrada al mateix temps, sempre guanyarà una de les dues sortides. D'això es desprèn que és indiferent quina de les dues guanyi. Amb els punts anteriors i aquest sistema d'arbitratge ja es poden aconseguir circuits sense atzars.

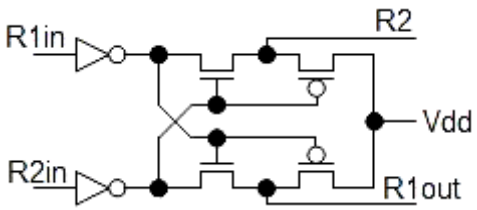


Fig. 2 Circuit d'exclusió mútua d'entrades.

3. SÍNTESI

Una vegada es té l'arquitectura del sistema a dissenyar, cal trobar un mètode sistemàtic pel seu disseny. Hi ha dues aproximacions. La primera és essencialment una versió de les màquines d'estat síncrones i s'anomena xarxa de Petri. La segona aproximació és un sistema algebraic basat en els processos seqüencials de les comunicacions de Hoare, que inicialment es van proposar pel disseny concurrent de *software*.

Un element utilitzat sovint en sistemes síncrons és l'anomenat element C. Aquest element té dues entrades i una sortida. Sempre que les dues entrades són iguals, la sortida és igual a les entrades, mentre que quan les entrades són diferents l'element manté el valor anterior.

4. AVANTATGES DE LA LÒGICA ASÍNCRONA

Els circuits asíncrons presenten senyals d'entrada binaris, però no tenen en compte la condició de temps discret. Per aquest motiu, també se'ls anomena autotemporitzats. Aquest canvi obre tot un ventall de noves possibilitats, que es descriuen tot seguit.

4.1. No retard en el senyal de rellotge

Generalment, en un disseny complex síncron el mateix senyal de rellotge s'ha de transmetre a diferents punts del circuit, provocant un retard en la transmissió, de manera que s'acostuma a disminuir la velocitat en funció del retard màxim. Donat que els dissenys asíncrons no tenen un rellotge comú, no existeix la possibilitat que aquest arribi retardat als diferents punts del circuit.

4.2. Menys potència

Els circuits síncrons han de carregar i descarregar punts del circuit que no estan treballant en el processat de la informació que interessa en aquell moment. Tot i que un disseny asíncron pot necessitar més commutacions per resoldre un càlcul, aquestes es produiran només en l'àrea encarregada del processat que interessa. Per tant, la potència total consumida en un circuit asíncron serà menor.

4.3. Limitacions no degudes al pitjor cas

Els circuits síncrons han d'esperar que el camí més lent hagi acabat abans de tenir en compte el resultat, així que la seva velocitat està limitada pel pitjor cas. Molts circuits asíncrons, en canvi, poden detectar quan s'ha realitzat l'operació, de manera que la velocitat no està limitada pel pitjor cas sinó per un cas intermig.

4.4. Optimització temporal global

En sistemes síncrons, el rellotge i la capacitat del sistema estan limitats pel camí més crític, així que moltes parts del circuit s'han d'optimitzar al màxim per tal d'obtenir el rellotge més ràpid possible. Els sistemes asíncrons funcionen a la velocitat del camí que s'està processant, de manera que les parts poc utilitzades poden deixar-se sense optimitzar, sense que això comporti una baixada considerable en el rendiment del conjunt.

4.5. Migració tecnològica

Els circuits integrats poden ser implementats en diferents tecnologies al llarg del seu cicle de vida. A causa de la limitació del camí més crític, la millora del rendiment d'un sistema síncron està lligada a la millora de tots i cadascun dels components en la nova tecnologia. En sistemes asíncrons, en canvi, millorant només el camí més utilitzat es millora el rendiment mitjà del conjunt. A més, les parts amb diferents retards es poden intercanviar sense que això afecti les altres estructures, ja que la majoria de circuits dissenyats mitjançant lògica asíncrona detecta el final de les operacions.

4.6. Adaptació automàtica a limitacions físiques

El retard en un circuit pot variar segons els paràmetres constructius, tensió d'alimentació o temperatura. Els sistemes síncrons han de tenir en compte que funcionen en la pitjor combinació d'aquestes condicions i utilitzar un rellotge d'acord amb aquest fet. En canvi, la majoria de circuits asíncrons detecten el final de les operacions, de manera que sempre funcionaran al màxim de les seves possibilitats temporals.

4.7. Exclusió mútua i monitorització d'entrades

Els elements que han de mantenir exclusió mútua per senyals independents, així com la sincronització entre les entrades i el rellotge, estan subjectes a metaestabilitat. Un estat metaestable és un estat inestable en equilibri. Donat que els circuits síncrons necessiten acabar l'operació en un temps determinat, l'exclusió mútua pot fallar en alguns sistemes. Els sistemes asíncrons, en canvi, poden esperar el temps que faci falta per acabar l'operació permetent una exclusió mútua segura.

5. INCONVENIENTS DE LA LÒGICA ASÍNCRONA

Amb tots aquests avantages, un es pregunta el perquè del predomini de la lògica síncrona. El motiu és degut a que els sistemes asíncrons presenten una sèrie d'inconvenients, dels quals se'n destaca la dificultat de ser dissenyats seguint un mètode sistemàtic i automàtic. En un

sistema síncron, el dissenyador només ha de definir els estats combinacionals necessaris i implementar-los, escollint un rellotge suficientment lent perquè funcioni correctament, mentre que en sistemes asíncrons s'ha de tenir en compte el funcionament dinàmic del circuit i evitar la introducció de canvis d'estat indesitjats durant les operacions.

Desafortunadament, els sistemes asíncrons no poden utilitzar eines CAD. Molts mètodes de disseny asíncron permeten només manipulacions algebraiques a partir de les lleis de Morgan i de les propietats commutativa i associativa.

Finalment, tot i que sembla que els sistemes asíncrons han de permetre un millor rendiment, no és evident que siguin més ràpids, ja que necessiten temps extra degut a la monitorització dels senyals propis, incrementant així el temps de retard mitjà.

6. DETECCIÓ DEL FINAL DE L'OPERACIÓ

Com es pot desprendre de comentaris anteriors, en un sistema asíncron és fonamental assolir un mètode eficient i ràpid per la detecció de la fi d'avaluació de la part en funcionament, per tal de poder passar a la següent etapa.

Una metodologia molt utilitzada en aquesta part del disseny és la que s'anomena de doble via (*dual-rail signaling*), on cada variable lògica es codifica utilitzant dues línies de senyal que s'anomenen parell de codificació. Els sistemes de doble via utilitzen tres estats lògics: 0, 1 i invàlid. El protocol d'aquests sistemes obliga a la variable a passar a l'estat invàlid després de tenir un 0 o un 1. D'aquesta manera s'introdueix un temps d'espera entre els valors correctes de les variables, permetent així la detecció de la finalització d'una operació.

En altres sistemes s'introdueix una variable de control per poder realitzar un *reset* o una precàrrega de totes les portes abans d'acceptar la següent dada com a vàlida. Aquest tipus de lògica amb precàrrega permet reduir el retard introduït entre dues dades vàlides.

Un altre mètode és la utilització de la senyalització de transicions (*transition signaling*), que s'implementa de forma que un canvi en la línia x_i (i=0, 1) d'un parell de codificació, s'interpreta com una nova dada de valor i. Aquest mètode elimina la necessitat d'espais d'espera entre dades vàlides. Desafortunadament, és molt difícil d'implementar.

Un sistema amb una implementació menys complicada que l'anterior és la codificació per nivell de dues fases i doble via (LEDR). És un sistema de doble via amb dues fases possibles: parell o senar. Cada dada ha de tenir la fase oposada a la dada anterior. Aquest mètode permet evitar els temps d'espera entre senyals.

Tots els mètodes anteriors requereixen una implementació amb un cost en àrea de silici considerable, a més de necessitar una lògica molt robusta per tal de tenir un funcionament correcte. El circuit de detecció augmenta de mida amb el nombre de variables de sortida. Aquestes limitacions han restringit el nombre d'aplicacions comercials que utilitzen lògica asíncrona.

Una altra alternativa molt més eficient és la que s'anomena detecció per monitorització de corrent o *current-sensing completion detection* (CSCD), que és aplicable a circuits CMOS autotemporitzats. Aquest sistema monitoritza constantment el corrent que es consumeix, de manera que es pot arribar a detectar la finalització d'una operació si es té suficientment caracteritzat el consum del sistema. Així, s'aconsegueix una implementació que utilitza una sola línia de codificació de fi per cada variable, com en sistemes síncrons. Els blocs d'aquesta alternativa no han de ser tan robustos com els esmentats en altres sistemes, de manera que acostumen a tenir la meitat de línies i transistors que un bloc de doble via.

7. APLICACIONS

Actualment, la indústria dels semiconductors té molt present els circuits asíncrons com una tecnologia a utilitzar. En general, aquests circuits s'han utilitzat només com a subcircuits o perifèrics de controladors. Els exemples més habituals són comptadors, temporitzadors, circuits d'arbitratge, controladors d'interrupcions, *fifos*, controladors de bus i circuits d'*interface*.

Durant les últimes dècades, s'han realitzat esforços considerables per aconseguir la producció de sistemes i circuits VLSI asíncrons. Els primers circuits comercials totalment asíncrons van aparèixer a finals dels anys 90. Un sistema asíncron pot ser utilitzat en aplicacions on es requereixi alta velocitat, baixa potència, poc soroll i poca radiació electromagnètica.

En un disseny asíncron, la següent operació pot començar immediatament després d'acabar l'anterior, ja que no s'ha d'esperar la transició del rellotge. Això comporta que els circuits asíncrons siguin potencialment més ràpids, però, com ja s'ha dit, el retard produït degut a la detecció de finalització del circuit l'acostuma a fer més lent. En el cas d'un sumador de N bits amb *carry*, per exemple, amb la lògica convencional es té un retard $\log_2 N(4)$. En canvi, si aquest mateix sumador s'implementa mitjançant la lògica asíncrona el retard és molt menor. En circuits en els quals el camí o la longitud del camí depenen de les dades, els circuits asíncrons poden arribar a millorar els circuits convencionals.

Pel que fa a la baixa potència, els circuits asíncrons només consumeixen quan estan realitzant un càlcul i no necessiten la constant aportació d'energia per generar el rellotge.

En el camp del soroll i l'emissió electromagnètica, si s'estudia l'espectre electromagnètic radiat per un sistema síncron, es comprova que conté un conjunt de freqüències múltiples de la freqüència de treball i, per tant, es té una certa potència de soroll. En canvi, en analitzar l'espectre electromagnètic radiat per una implementació del mateix

sistema basada en lògica asíncrona, s'obté una potència de soroll molt més baixa.

Per aquests motius aquí esmentats, s'han realitzat dissenys complets de cirucits asíncrons que poden millorar o almenys competir amb les característiques dels predecessors síncrons (Fig. 3 i Fig. 4).

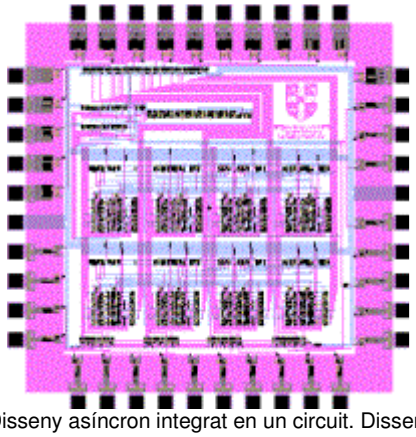


Fig. 3 Disseny asíncron integrat en un circuit. Dissenyat per University of Cambridge.

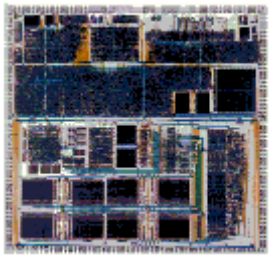


Fig. 4 Disseny asíncron integrat en un circuit de comunicacions (DRACO). Dissenyat per University of Manchester.

8. CONCLUSIONS

En resum, doncs, sembla evident que si la tecnologia asíncrona no s'ha utilitzat més és degut a la complexitat del seu disseny, verificació i implementació automàtica. Els coneixements actuals sobre el funcionament i les limitacions dels sistemes síncrons estan impulsant, però, el desenvolupament de solucions alternatives com poden ser els circuits asíncrons.

D'altra banda, les noves tecnologies, barrejant senyals de diferent naturalesa i buscant sempre el menor consum, fan de la lògica asíncrona una solució ideal per ser utilitzada en futures aplicacions. Tot i això, cal tenir present que encara falta molt desenvolupament i recerca sobre els sistemes d'implementació de sistemes asíncrons, per arribar a fer aplicacions comercialment competitives i que puguin millorar els sistemes síncrons, ja que de moment només s'acosten a les seves prestacions i introdueixen molta complexitat en el disseny.

Paral·lelament a l'evolució del disseny, s'ha d'impulsar el desenvolupament de les eines d'automatització i disseny del tipus CAD per poder aproximar-se als sistemes convencionals.

9. REFERÈNCIES

[Hauck95] S. Hauck, "Asynchronous Design Methodologies: An Overview", Proceedings of the IEEE, vol 83, núm 1, gener 1995.

[Chaney73] T. J. Chaney, C. E. Molnar, "Anomalous Behaviour of Syncchronizers and Arbiters", IEE Transactions on Computers, vol C22, pp 421–422, abril 1973.

[Meald80] C. Meald, L. Conway, "Introduction to VLSI Systems", Addison-Wesley, 1980.

[Seitz] C. Seitz, "System timing. An Introduction to VLSI Systems", Addison-Wesley.

[Dean94] M. E. Dean, D. L. Dill, M. Horowitz, "Self-timed logic using Current-Sensing Completion Detection (CSCD)", Journal of VLSI signal processing, núm 7, pp 7-16, 1994.

[Anantharaman86] T. S. Anantharaman, "A delay insensitive regular expression", IEEE VLSI Technicall Bulletin, 1986.

[David89] I. David, R. Ginosar, M. Yoeli, "An efficient implementation of Boolean functions as self-timed circuits", Technion and Israel Institute of Technology, 1989.

[Martin89] A. Martin, "On the existence of delay-insensitive circuits", MIT Converence on Advanced Research in VLSI, març 1989.

[Singh81] N. P. Singh, "A design methodology for Self-timed systems", M. Sc. Thesis, MIT Laboratory for Computer Science Technical Report TR-258 MIT, Cambridge, febrer 1981.

[Sutherland89] I. Sutherland, "Micropipelines", Communications of the ACM, vol 32, pp 720-738, 1989.

[Jacobs88] G. Jacobs and R. Brodersen, “Self-timed integrated circuits for digital signal processing applications”, Proceedings of Third workshop on VLSI signal processing, setembre 1988.

[Johnson90] M. Johnson, N. Jouppi, “Notes for EE371, Advanced VLSI Design”, Stanford University, Spring Quarter, 1990.

[Meng88] T. Meng, “Asynchronous design for programmable digital signal processors”, Ph.D. Thesis, UC Berkeley, 1988.

[Dean91] M. E. Dean, T. E. Williams, D.L. Dii, “Efficient self-timing with level-encoded 2-phase dual-rail (LEDR)”, MIT Conference on Advanced Research in VLSI, març 1991.

[Izosimov90] O. A. Izosimov, I. I. Shagurin, V. V. Tsylyov, “Physical approach to CMOS module self-timing”, Electronic Letters, vol 26, núm 22, octubre 1990.

[van Berkel] C. H. van Berkel, M. B. Josephs, S. M. Nowick, “Scanning the Technology: Applications of Asynchronous Circuits”.

[McPeak99] S. McPeak, “Asynchronous Logic”, UC Berkeley, desembre 1999.

[Davis97] A. Davis, S. M. Nowick , “An Introduction to Asynchronous Circuit Design”, setembre 1997.

[Unger69] S. H. Unger, “Asynchronous Sequential Switching Circuits”, Wiley-Interscience, 1969.

[Seger89] C. J. Seger, “A bounded delay race model”, Proceedings of the IEEE International Conference on Computer-Aided Design, pp 130-133, IEEE Computer Society Press, novembre 1989.

[Miller65] R. E. Miller, “Switching Theory Vpolume II: Sequential Circuits and Machines”, John Wiley and Sons, 1965.

[Clark67] W. A. Clark, “Macromodular computer systems”, Proceedings of the Spring Joint Computer Conference AFIPS, abril 1967.

[Udding86] J. T. Udding, “A formal model for defining and classifying delay-insensitive circuits and systems”, Distributed Computing, pp 197-204, 1986.

[Dill89] D. L. Dill, “Trace Theory for Automatic Hierarchical Verification of Speed-Independent Circuits”, MIT Press, Cambridge, 1989.

[Burns96] S. M. Burns. “General condition for the decomposition of state holding elements”, Proceedings of the International Symposium on Advanced Research in Asynchronous Circuits and Systems, pp 48-57, IEEE Computer Society Press, novembre 1996.

[Martin90] A. J. Martin, “Programming in VLSI: From communicating processes to delay-insensitive circuits”, Developments in Concurrency and Communication, pp 1-64, Addison-Wesley, 1990.

[van Berkel93] K. van Berkel, “An asynchronous architecture for VLSI programming International Series on Parallel Computation 5”, Cambridge University Press, 1993.

[Nowick91] S. M. Nowick, D. L. Dill, “Automatic synthesis of locally-clocked asynchronous state machines”, Proceedings of the ICCAD, pp. 192-197, 1991.

[Murata89] T. Murata, “Petri nets: properties, analisys and applications”, Proceedings of the IEEE, vol 77, núm 4, pp. 541-580, 1989.

A5. Disseny de subsistemes digitals: sumadors.

***Abstract*—**En aquest text es farà una introducció als subsistemes digitals, desenvolupant el tema dels sumadors. Es podrà observar el disseny d'una sèrie de sumadors i com petites modificacions d'aquests serveixen per realitzar millores, sobretot pel que fa a la velocitat de propagació de la suma.

***Index Terms*—**Optimització, *full-adder* o sumador complet, *carry*, lògica CMOS, sumador simètric, lògica dinàmica, generador de *carry* Manchester, segmentació, *carry-bypass*, selecció lineal, selecció per arrel quadrada, càlcul anticipat.

1. INTRODUCCIÓ

Normalment, un dissenyador compta amb una llibreria de cel·les amb blocs relativament complexos que ja han estat dissenyats, provats i optimitzats amb anterioritat per tal d'alleujar el seu treball. D'aquesta manera, el dissenyador ja pot integrar-los al seu disseny amb total seguretat. Un exemple d'aquests blocs són els sumadors.

La velocitat dels sumadors és el tema principal que s'ha d'optimitzar, ja que la majoria de les vegades, posteriors operacions que es vulguin realitzar vindran influenciades pel retard que s'introdueixi en l'operació de sumar dos nombres. Per exemple, si es desitja realitzar l'operació:

$$y = \frac{a+b}{2} \tag{1}$$

Primer s'ha de realitzar la suma dels termes a i b. Com que s'està treballant en binari, després cal desplaçar tots els bits a la dreta per fer la divisió entre dos, operació que no es pot efectuar si no s'ha finalitzat la suma. Si el sumador comporta un retard elevat, tota l'operació acumula un gran retard. Per tot això és essencial millorar la seva velocitat.

Optimitzacions típiques d'un sumador són reescriure les equacions Booleanes per obtenir un circuit més ràpid, ja que en simplificar una funció, per exemple, es pot arribar a reduir el nombre de transistors en sèrie i així escurçar el camí de càrrega. Això es té esquematitzat a la Fig. 1, on la càrrega superior és més ràpida que la inferior en tenir menys termes RC en sèrie. El cas de la descàrrega és anàleg al de la càrrega.

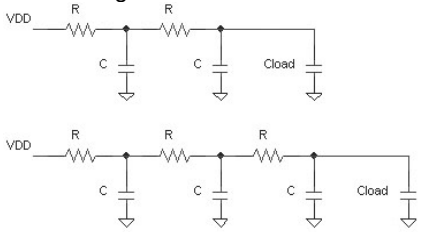


Fig. 1 Optimització d'un sumador.

A continuació es descriurà el funcionament del sumador binari bàsic i alguns exemples de les millores que s'aconsegueixen introduint petits canvis a aquest.

2. SUMADOR BINARI BÀSIC

Es parteix de la base, és a dir, del *full-adder* o sumador complet d'un bit.

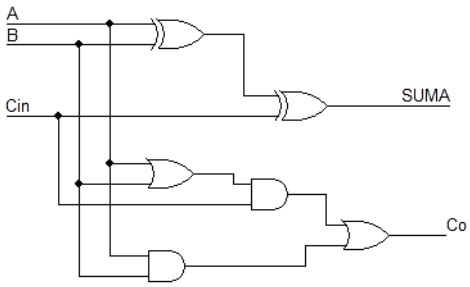


Fig. 2 Sumador complet d'un bit.

Mitjançant les equacions (2) i (3), s'obté el resultat de la suma (S) i el valor del *carry* (C_o).

$$S = A \oplus B \oplus C_{in} \tag{2}$$

$$C_o = AB + BC_{in} + AC_{in} \tag{3}$$

Es tenen entrades pels dos bits i pel *carry* d'entrada, que poden estar connectats a un altre sumador complet d'un bit i així poder fer la suma de nombres de n bits.

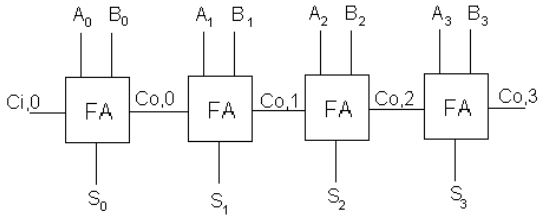


Fig. 3 Sumador complet de quatre bits.

Cada bloc FA de la Fig. 3 correspon a un sumador complet d'un bit com el descrit a la Fig. 2. En una configuració de n bits, es pot donar el cas de tenir un *carry* generat pel bit menys significatiu i que aquest hagi de ser arrossegat pels n sumadors d'un bit fins al bit més significatiu, és a dir, que hagi de recórrer tot el camí. Aquest retard és proporcional al nombre de bits i és aproximadament:

$$t_{sumador} \approx (n-1)t_{carry} + t_{suma} \quad (4)$$

on t_{carry} és el temps que triga el senyal en propagar-se des de C_{in} a C_o i t_{suma} és el que es triga en sumar el bit A al B.

3. DISSENY EN LÒGICA CMOS

Simplificant les equacions (2) i (3) s'obtenen les equacions (5) i (6), a partir de les quals es pot implementar un sumador complet d'un bit en lògica CMOS, tal com es mostra a la Fig. 4. Aquí s'han utilitzat inversors per no obtenir les funcions S i C_o negades, ja que la lògica CMOS és negada.

$$C_o = AB + BC_i + AC_i \quad (5)$$

$$S = ABC_i + \overline{Co} \cdot (A + B + C_i) \quad (6)$$

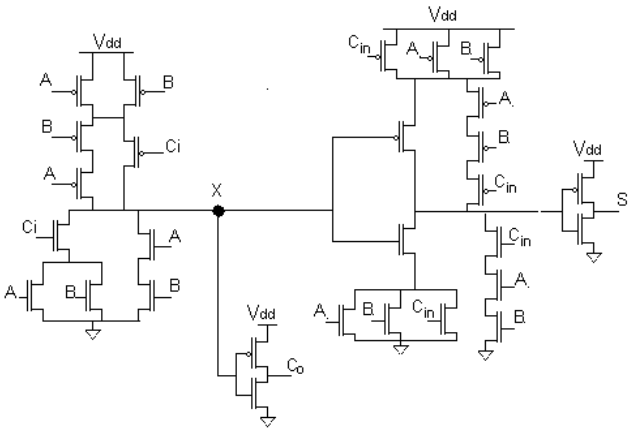


Fig. 4 Disseny que utilitza 28 transistors.

En aquest disseny es requereixen 28 transistors i s'ocupa força àrea. D'altra banda, donat que es requereixen tres pMOS i dos nMOS en sèrie per avaluar la funció X, i dues etapes inversores per bit pel circuit de generació de *carry*, existeixen un bon nombre de capacitats paràsites que comporten un retard considerable en la propagació de X. Finalment, la capacitat de càrrega del C_o és gran ja que està formada per dues capacitats de difusió i sis de porta que s'han de sumar a la de *wire*, que encara que és teòricament menyspreable, existeix i pot influenciar.

En relació al que s'ha descrit anteriorment, es té una primera millora fent ús simplement de les propietats dels sumadors. La de la inversió, per exemple, que diu que invertint les entrades es tindran invertides les sortides. Això permet eliminar l'inversor en la propagació del *carry*, disminuint el retard i augmentant la velocitat, però obligant a alternar diferents cel·les per cada cicle de propagació, cosa que és més incòmode que utilitzar sempre les mateixes cel·les.

Si se suprimeixen els inversors, quan s'obté el resultat del primer bloc FA de l'exemple de la Fig. 3, aquest es farà amb el C_o negat i, per tant, s'haurà d'usar un FA amb les entrades negades a continuació per propagar correctament el valor del *carry*. És per això que es necessitarà alternar aquestes dues configuracions per eliminar un inversor de cada etapa i aconseguir que el funcionament a nivell lògic sigui el correcte.

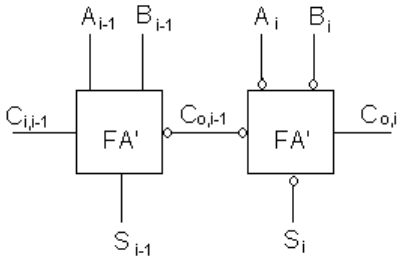


Fig. 5 Sumador de dos bits.

A la Fig. 5 es mostra com seria un sumador de dos bits. Pel cas de quatre bits, només caldria connectar en cascada els dos bits següents.

Una altra millora bastant útil és definir S i C_o com funcions d'alguns senyals intermitjos, com poden ser G (generar), D (esborrar) o P (propagar). G=1 (D=1), per exemple, indica que es generarà un bit de *carry*. D'altra banda, quan P=1 es garanteix que es propagarà un bit de *carry* cap al següent bloc. Es poden deduir les expressions de G, D i P de la taula de la veritat d'un *full-adder*.

A	B	C_i	S	C_o
0	0	0	0	0
0	0	1	1	0
0	1	0	1	0
0	1	1	0	1
1	0	0	1	0
1	0	1	0	1
1	1	0	0	1
1	1	1	1	1

Taula 1 Taula de la veritat d'un full-adder.

$$G = A \cdot B \quad (7)$$

$$D = \overline{A} \cdot \overline{B} \quad (8)$$

$$P = A + B \quad (9)$$

I llavors, S i C_o es poden escriure com:

$$C_o(G,P) = G + P \cdot C_i \quad (10)$$

$$S(G,P) = P \oplus C_i \quad (11)$$

A continuació, es presenta l'anomenat sumador simètric o de mirall. La primera etapa dóna com a resultat $/C_o$. Quan P=0 (no es propaga), A i B han de ser zero. Llavors G=0 i no es genera *carry*. Mirant la Fig. 6 es veu que es té un camí de càrrega, és a dir, $/C_0$ serà igual a 1 i, per tant, no es tindrà *carry* de sortida. En canvi, quan P=1 la capacitat de càrrega que hi ha en $/C_0$ s'ha de descarregar per tenir un 0, que pel *carry* de sortida serà un 1. Es tenen tres possibilitats:

- A=B=1. Això implica una descàrrega (*carry* de sortida a 1), sigui quin sigui el valor de C_{in}. Es verifica el que s'ha descrit a l'eq. (6).
- A=0, B=1. En aquest cas es té una descàrrega quan C_{in}=1. Si C_{in}=0, el *carry* de sortida és zero. Per tant, es comprova que si P=1, al *carry* de sortida es té el valor que hi ha al carry d'entrada.
- B=1, A=0. Aquest cas és anàleg a l'anterior.

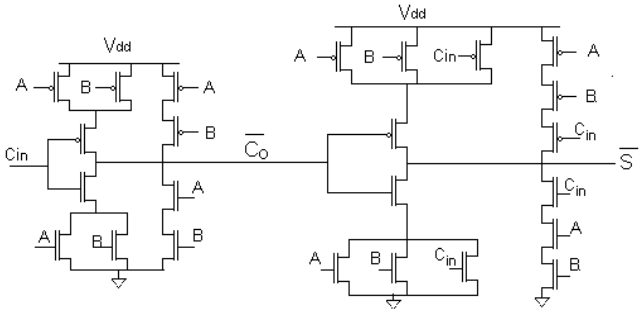


Fig. 6 Sumador simètric o de mirall.

Així doncs, s'obté un disseny que presenta un seguit d'avantatges, dels quals es destaca el reduït espai que ocupa, ja que s'ha disminuït el nombre de transistors. Com a conseqüència de tenir menys transistors, també es tindran menys capacitats de difusió, cosa que comporta una minimització de la capacitat de càrrega al node del *carry* de sortida. A més, donat que en aquest darrer punt es tindran dos pMOS en sèrie com a màxim, s'aconsegueix que tot el disseny sigui més ràpid.

4. DISSENY EN LÒGICA DINÀMICA

Els circuits descrits anteriorment poden ser implementats també en estils de lògica dinàmica, com poden ser la lògica NP o la *dominó*.

Es mostra un exemple d'un sumador basat en lògica NP a la Fig. 7. Quan la zona marcada en gris estigui avaluant, realitzarà la propagació de dos bits de *carry*. Si es compara la Fig. 7 amb la Fig. 6, es veu que el bloc de propagació de cada *carry* no deixa de ser un bloc pMOS o nMOS, però sense el seu complementari nMOS o pMOS, respectivament. Aquests blocs complementaris són substituïts per transistors controlats pel rellotge, col·locats de manera que tinguin lloc les fases de precàrrega i avaluació. El funcionament del circuit és el que es descriu a continuació.

Si el rellotge està a nivell baix, es realitza una descàrrega del node C_{i2} a 0 (etapa de precàrrega) i el node $/C_{i1}$ efectua una precàrrega a 1. Quan el rellotge passi a un nivell alt, els dos blocs realitzaran l'avaluació, ja que els pMOS que pertanyen al bloc del node C_{i2} deixaran passar el corrent i permetran tenir aquest node a 1 o 0 segons el que es tingui a les entrades. Al bloc de $/C_{i1}$, els nMOS del rellotge deixaran passar el corrent i avaluaran, si s'escau, la descàrrega.

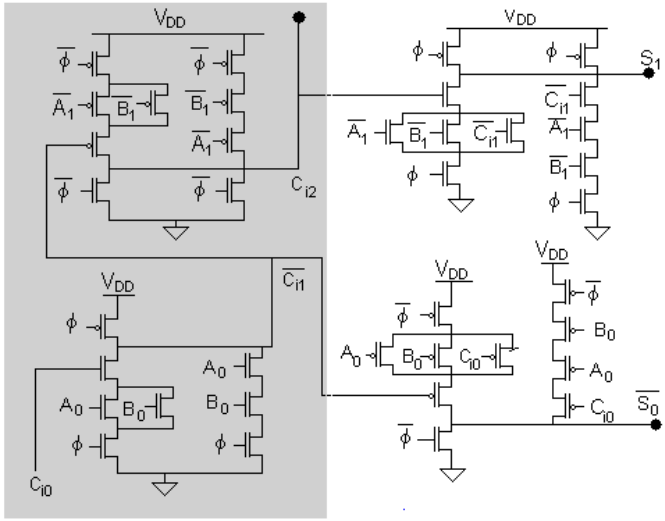


Fig. 7 Sumador basat en lògica NP.

A partir del muntatge anterior i tenint en compte les equacions (7) i (9), es mostra la Taula 2 per tal de veure l'evolució dels nodes durant l'avaluació.

A_k	B_k	$C_{i,k-1}$	G_k	P_k	$C_{0,k}=C_{i,k}$	$/C_{i,k}$
0	0	X	0	0	0	1
0	1	0	0	1	0	1
0	1	1	0	1	1	0
1	0	0	0	1	0	1
1	0	1	0	1	1	0
1	1	X	1	1	1	0

Taula 2 Taula de la veritat d'un sumador basat en lògica NP.

En definitiva, pel fet d'utilitzar un rellotge s'ha aconseguit l'estalvi d'un bloc pMOS o nMOS i, per tant, una reducció de les capacitats paràsites i una major velocitat de propagació.

D'altra banda, la lògica dinàmica també permet realitzar diagrames alternatius de disseny que en lògica estàtica són

molt complicats. Un exemple seria el bloc generador de *carry* Manchester, que utilitza una sèrie de transistors de pas en cascada pel càlcul del *carry*. El seu funcionament és el que es descriu tot seguit.

Els senyals de propagació (P) i generació (G) del *carry* controlen els transistors de pas, com es pot observar a la Fig. 8. Llavors, el rellotge a nivell baix precarrega tots els nodes a V_{DD} i el rellotge a nivell alt, avalua. Durant la precàrrega posa M_k a V_{DD}, és a dir, posa el *carry* de sortida de l'últim bit a zero (hi ha un inversor). Però quan el rellotge passa a 1, el camí de descàrrega queda obert. Els nodes M_k es poden descarregar en funció del valor de G_{k-1}.

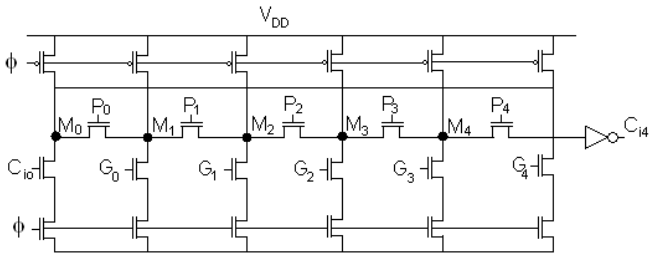


Fig. 8 Bloc generador de *carry* Manchester de cinc bits.

Com a exemple, s'explicarà el cas en què es vol propagar C_{i0}=1 a la sortida C_{i4}=1. Com que es vol avaluar, s'estarà en ϕ=1. A més, P_k=1 perquè els transistors de pas juntament amb C_{io}=1 deixin passar el zero, ja que pel primer transistor es descarreguen tots els nodes precarregats a V_{DD}. Aquest últim pas implica que el corrent que es té a cada node no serà el mateix, ja que tots els nodes es descarreguen per un sol transistor.

$$I_{Mo} > I_{M1} > I_{M2} > I_{M3} > I_{M4} \quad (12)$$

Per aquesta relació, si es trien transistors amb dimensions que variïn progressivament en funció de si han de suportar major o menor corrent, es tindrà una millora en quant a la velocitat del circuit. Pel node que suporti un major corrent, el transistor de pas tindrà unes dimensions majors que per un altre que suporti un corrent menor. Això també és correcte pels transistors que fan la descàrrega i estan connectats a les funcions P i G. S'aconsegueix una millora en el temps de retard per cada bit.

El retard d'aquest pitjor cas de propagació el marcarà un circuit com el de la Fig. 9.

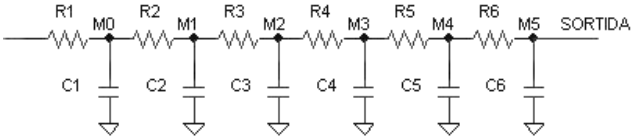


Fig. 9 Circuit que representa el retard pel pitjor cas de propagació.

A l'eq. (13) es té que C_i és la capacitat entre cada node i terra. R_j, en canvi, és la suma de resistències entre aquest node intermig i el node pel qual es descarreguen totes les capacitats. Si s'analitza l'eq. (14), es veu que R_1 es repeteix fins a sis vegades. Per tant, minimitzant aquesta contribució fent el primer transistor més ample (com més ample, menys resistència presenta ja que R∝1/W), i així successivament per tots, s'aconseguirà minimitzar el temps de propagació i, per tant, es tindrà un bloc més ràpid. S'ha de tenir en compte que pot passar que es redueixi molt el retard però que l'àrea augmenti massa i el cost sigui excessiu. S'ha de prendre un terme mig, en funció del cost.

$$t_{propagació} \propto \sum_{i}^{N} C_i \cdot \left(\sum_{j}^{N} R_j \right) \quad (13)$$

$$t_{propagació} \propto \begin{pmatrix} C_1 \cdot R_1 + C_2 \cdot (R_1 + R_2) + C_3 \cdot (R_1 + R_2 + R_3) + \\ C_4 \cdot (R_1 + R_2 + R_3 + R_4) + C_5 \cdot (R_1 + R_2 + R_3 + \\ R_4 + R_5) + C_6 \cdot (R_1 + R_2 + R_3 + R_4 + R_5 + R_6) \end{pmatrix} \quad (14)$$

5. SUMADORS DE SEGMENTACIÓ

Un altre tipus d'implementació de sumadors és el que utilitza la segmentació. La segmentació consisteix en utilitzar recursos del circuit que queden inactius durant un cert espai de temps, degut a que estan actuant altres parts.

A la Fig. 11 es mostra un exemple d'un sumador de segmentació. En aquest, el bloc de *carry* del bit i efectua una precàrrega quan el rellotge està a nivell baix. En canvi, quan el rellotge es troba a nivell alt, és el bloc de *carry* del bit i+1 el que efectuarà una precàrrega. Mentre un fa una precàrrega, l'altre avalua, i viceversa. Això és la segmentació.

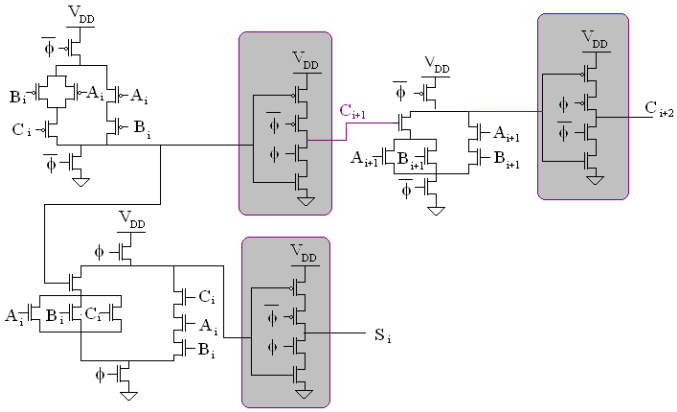


Fig. 11 Sumador de segmentació.

Es té una sèrie de transistors marcats en gris que deixaran passar el *carry* i de la suma quan el rellotge es trobi a nivell alt. En canvi, quan el rellotge està a nivell baix, deixarà propagar el *carry* i+1, ja que en aquest moment s'estarà avaluant. Aquests blocs fan que el resultat es

propagui només quan el bloc que els precedeix està avaluant, és a dir, fins que no es té un cicle complet, no es propaga tota la informació. Així es dóna temps perquè els blocs i i i+1 efectuïn la seva etapa corresponent d'avaluació. En canvi, si no s'utilitza la tècnica de la segmentació, per obtenir el *carry* i, i+1 i la suma, es necessitarien dos cicles de rellotge, un per obtenir el valor de la suma i el *carry* i un altre cicle per obtenir el del carry i+1.

Així doncs, en N cicles es tindrà el resultat correcte de tots els *carrys*. Un *carry* necessita del bit anterior per tenir un valor correcte. Per tant, en cada cicle de rellotge s'anirà progressant un bit. A la Fig. 12 es té un resum del seu funcionament.

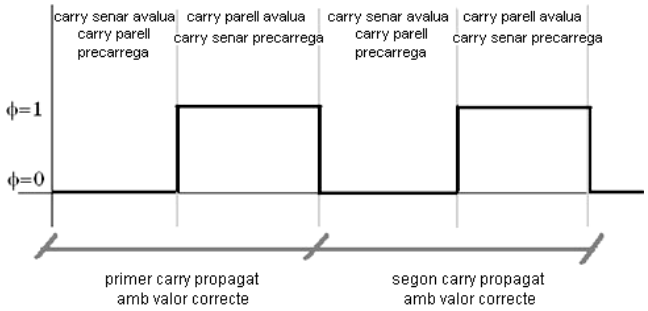


Fig.12 Funcionament d'un sumador de segmentació.

6. ALTRES CONFIGURACIONS LÒGIQUES

Els sumadors que s'han vist fins ara són únicament pràctics per paraules de longitud curta, ja que en el cas d'un processador que realitzi operacions amb paraules d'un gran nombre de bits, el retard en realitzar la suma pot ser massa gran. A continuació es mostren alguns exemples de configuracions que disminueixen aquest retard.

6.1. Sumador *carry-bypass*

Perquè un bit es propagui des del principi fins al final ha de passar que P_k=1. Per tant, es pot definir una funció BP= $P_0 \cdot P_1 \cdot \ldots \cdot P_{k-1} \cdot P_k$, que en aquest cas concret serà 1.

Així, si es pensa en l'exemple de la Fig. 8, es pot seguir un camí alternatiu entre el primer *carry* d'entrada i l'últim de sortida mitjançant un transistor de pas nMOS controlat per la funció BP (veure Fig.13).

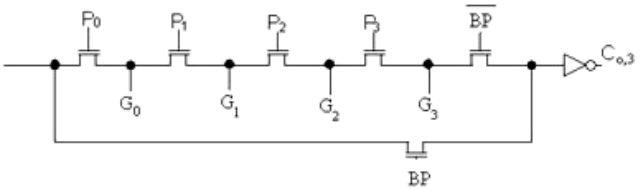


Fig. 13 Sumador *carry-bypass*.

El retard d'un sumador de n-bits consta de t_{setup} o temps per crear els senyals de generació, t_{carry} o retard d'un bit, t_{bypass} o retard en la propagació pel camí alternatiu i t_{suma} o temps en generar la suma.

$$t_p = t_{setup} + m \cdot t_{carry} + \left(\frac{n}{m} - 1\right) \cdot t_{bypass} + t_{suma} \qquad (15)$$

S'ha suposat que el sumador total s'ha dividit en (n/m) etapes de *bypass* d'igual longitud i que cadascuna conté m bits, on n és el nombre total de bits. Tot seguit, es descriu un exemple.

Es vol fer la suma de 16 bits de 4 en 4 bits. Per tant, es té n/m=4. Llavors, es té el temps de generar P i G, propagar els m bits de *carry* de cada bloc i generar la suma. També cal tenir en compte que es perdrà temps en camins alternatius (tres en aquest cas). A la Fig. 14 es té una comparació del temps de retard entre un sumador com el dels apartats anteriors i l'explicat ara mateix.

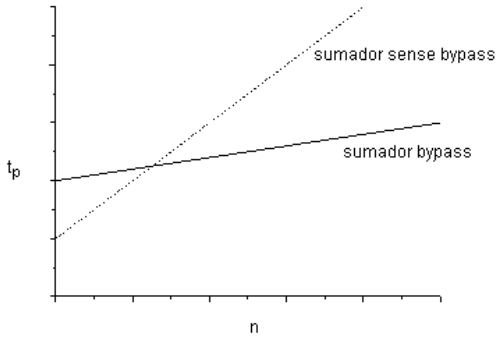


Fig. 14 Comparació entre un sumador amb i sense *bypass*.

Com es pot observar, per paraules de pocs bits (entre 4 i 8) el sumador sense *bypass* és més ràpid, però quan el nombre de bits es fa molt gran la diferència és abismal.

6.2. Sumador de selecció lineal de *carry*

En una configuració habitual, cada *full-adder* ha d'esperar que entri el *carry* per generar el seu. Una forma d'evitar que s'hagi d'esperar és anticipar les dues possibilitats i avaluar el resultat que el *carry* d'entrada sigui 0 o 1. Una vegada que es coneix el valor del *carry*, se selecciona el valor correcte mitjançant un multiplexor.

Se suposa que el nombre total de bits és n i que el nombre de bits per etapa és m. A més, s'ha de suposar el pitjor cas per la propagació. Això ocorre quan s'han d'avaluar els quatre multiplexors. El temps de retard ve donat per l'eq. (16).

$$t_p = t_{setup} + m \cdot t_{carry} + \left(\frac{n}{m}\right) t_{mux} + t_{sum} \qquad (16)$$

L'únic temps nou que apareix a l'equació anterior és el del multiplexor t_{mux}. S'ha d'afegir que, a nivell de circuit, cada bloc té dos camins de propagació de *carry* (un pel 0 i

un altre per l'1), i un multiplexor que tria un resultat o altre en funció del senyal de selecció, que vindrà donat pel *carry* real d'entrada.

6.3. Sumador de selecció de *carry* per arrel quadrada

Es parteix del sumador lineal i se li fan algunes modificacions. Es considera el multiplexor de l'última fase de la suma. Les entrades d'aquest són els resultats dels seus dos *carrys* corresponents i d'un senyal del multiplexor de l'etapa anterior. Els resultats del *carry* són estables un temps després de la seva arribada al multiplexor. Això pot donar la idea d'equilibrar el retard fent que els blocs, d'una forma progressiva, vagin afegint bits a cada etapa. Per tant, si se suposa que hi ha P etapes, que la primera etapa suma m bits i que en cada etapa s'afegeix un bit:

$$n = m + (m+1) + ... + (m+P-1) = \frac{P^2}{2} + P \cdot \left(m - \frac{1}{2} \right) \qquad (17)$$

Com que normalment m<<n, es té:

$$P \approx \sqrt{2 \cdot n} \qquad (18)$$

Per comparar aquest retard amb el del sumador amb selecció de *carry* lineal, només cal relacionar $\sqrt{2 \cdot n}$ amb (n/m). Per n gran, el temps de retard pel sumador amb selecció de *carry* per arrel quadrada creix molt a poc a poc, mentre que pel lineal ho fa a un ritme major (veure Fig. 15).

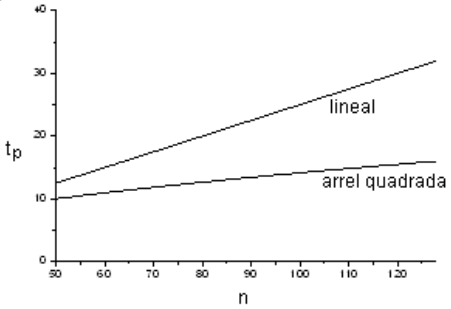


Fig.15 Temps de retard per n majors de 50.

6.4. Sumador de càlcul anticipat de *carry*

Aquest tipus de sumadors ofereixen la possibilitat de barrejar les propietats dels sumadors de *bypass* i de selecció de *carry*. La funció d'un *carry* de sortida es pot expressar a partir d'un terme generador i un de propagador. Per tant, s'arriba a una expressió com la de l'eq. (17). Aquesta relació de recurrència permet calcular qualsevol bit de *carry* a partir dels bits dels nombres a sumar i del *carry* inicial.

$$C_{o,i} = G_{i-1} + P_{i-1} \cdot C_{o,i-1} \qquad (17)$$

Aquests blocs generadors de *carry* són només útils per sumes de 4 bits com a màxim. Això és degut al fet que per generar el *carry* es necessita una porta OR de i+1 entrades. Això fa que el *fan-in* sigui massa gran i disminueix la velocitat del circuit. El *fan-out* tendeix a créixer excessivament, cosa que també fa que el circuit sigui més lent. Tot això és degut al fet que el temps de retard d'una porta és proporcional a la seva càrrega; com més entrades, més càrrega és necessària. A la Fig. 16 es té l'esquema d'un bloc generador de *carry* de 4 bits.

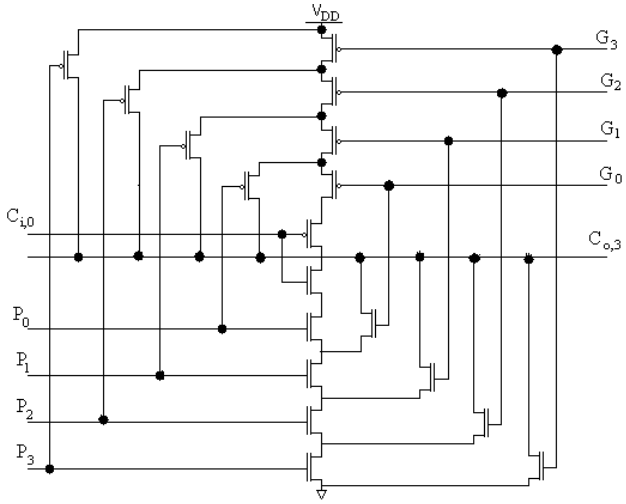


Fig. 16 Sumador de càlcul anticipat de *carry*.

7. CONCLUSIONS

Existeixen diferents estructures que permeten implementar un sumador binari. El sumador complet o *full-adder* constitueix l'estructura bàsica, a partir de la qual es realitzen modificacions per tal d'optimitzar l'àrea i la velocitat de propagació.

Una de les alternatives consisteix en dissenyar el sumador mitjançant lògica CMOS, tot definint la suma i el *carry* com funcions d'uns senyals intermitjos anomenats generar, esborrar i propagar. Seguint aquest esquema es pot dissenyar el sumador simètric o de mirall, que presenta l'avantatge de tenir un reduït nombre de transistors i, per tant, menys capacitats paràsites i més velocitat.

Una segona opció es basa en dissenyar el sumador amb lògica NP, que redueix considerablement el nombre de transistors en substituir blocs sencers per transistors controlats per un rellotge, de manera que puguin tenir lloc les fases de precàrrega i avaluació. Mitjançant la lògica dinàmica també es pot dissenyar el generador de *carry* Manchester, que permet millorar la velocitat del circuit implementant transistors de diferents dimensions en funció de la quantitat de corrent que suportin.

Una altra tècnica per optimitzar el temps de retard en sumadors binaris és la segmentació, que es basa en utilitzar recursos del circuit en situacions en què

normalment queden inactius. Finalment, existeixen altres configuracions lògiques que disminueixen el retard i permeten efectuar sumes de paraules llargues, com són el sumador *carry-bypass*, el sumador de selecció lineal de *carry*, el sumador de selecció de *carry* per arrel quadrada i el sumador de càlcul anticipat de *carry*.

8. REFERÈNCIES

[Rabaey96] J. M. Rabaey, “Digital integrated circuits. A design perspective”, Prentice-Hall, 1996.

[Herms] A. Herms, S. Bota, S. Bosch, J. M. López Villegas, “Disseny Digital Bàsic”, Textos docents núm 126, Edicions de la Universitat de Barcelona.

A6. Disseny de subsistemes digitals: multiplicadors.

***Abstract*—**L'objectiu d'aquest document és estudiar una de les aplicacions dels sistemes digitals, els multiplicadors. Concretament, es veurà la seva estructura bàsica, així com les diferents tècniques que existeixen per tal d'optimitzar la velocitat d'execució. Finalment, també es comentaran diversos mètodes que permeten la seva implementació

***Index Terms*—**Multiplicació binària, multiplicador, multiplicand, producte parcial, producte entre bits, algoritme de Booth, arbre de Wallace, esquema de multiplicació Dadda, dígit, subparaula, paraula, *full-adder*, *array*, FPGA, LUT.

1. INTRODUCCIÓ

Una de les aplicacions més importants dels sistemes digitals és el càlcul d'operacions aritmètiques tals com sumar, restar, multiplicar o dividir. Els circuits que realitzen aquestes operacions sovint s'utilitzen com a mòduls bàsics per la implementació d'operacions i algoritmes de major complexitat.

Tot i que la multiplicació digital actualment segueix sent un tema d'investigació, les aportacions més importants en aquest camp es van realitzar a la dècada dels 60. La major part dels treballs que es van fer en aquella època segueixen tenint validesa, ja que la seva utilitat va més enllà del marc tecnològic en el qual van ser concebuts.

2. MULTIPLICADOR FONAMENTAL

Com a exemple de multiplicador senzill, s'analitzarà el cas d'una multiplicació binària de dos números de 2 bits. Per veure com es pot implementar, cal desenvolupar pas a pas l'operació de multiplicar.

		B_1	B_0
	X	A_1	A_0
		A_0B_1	A_0B_0
	A_1B_1	A_1B_0	
C_3	C_2	C_1	C_0

Fig. 1 Multiplicació binària de dos números de 2 bits.

La multiplicació queda descomposta en productes AND formats pels bits del multiplicador i del multiplicand, en tants nivells com bits tingui el multiplicador. Per tant, per J bits del multiplicador i K bits del multiplicand seran necessàries JxK portes AND i (J-1) sumadors de K bits per tal de produir un producte de J+K bits. Per tant, una possible implementació d'aquest senzill multiplicador serà:

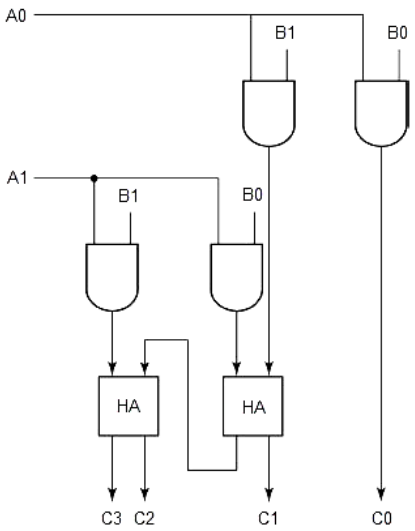


Fig. 2 Implementació d'un multiplicador 2x2.

Aquest senzill exemple és útil per definir alguns conceptes. S'entén per producte parcial el resultat de multiplicar el multiplicand per un bit del multiplicador, i s'anomena producte entre bits al producte entre un bit del multiplicand i un altre bit del multiplicador.

Els dos paràmetres més importants en el disseny d'un multiplicador són la velocitat d'execució i l'àrea ocupada. Segons les necessitats requerides en cada cas, caldrà arribar a un compromís entre aquests dos factors. Per tal de maximitzar la velocitat existeixen tres metodologies de disseny, que s'analitzaran a continuació.

2.1. Accelerar la formació de productes parcials

La primera opció condueix a un *array* de portes AND que realitza els n productes parcials de forma simultània. Aquesta idea de disseny va ser la que es va implementar en els primers circuits multiplicadors. És el cas del multiplicador de *ripple-carry*, l'esquema del qual és el següent.

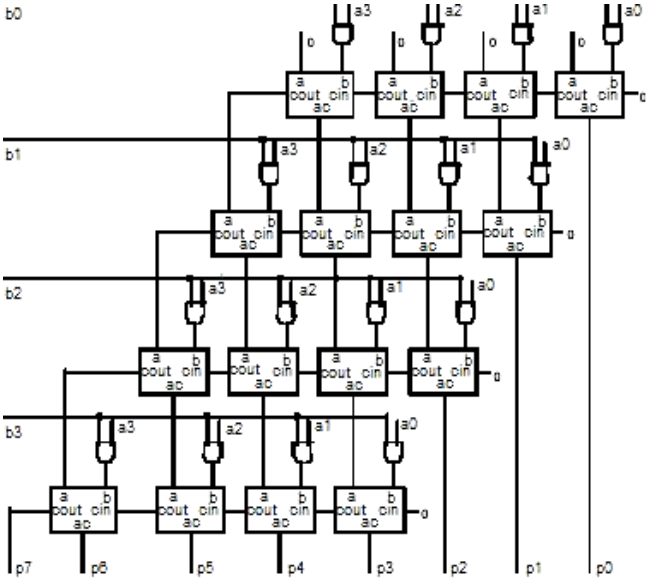


Fig. 3 Multiplicador de *ripple-carry*.

2.2. Reduir el número de productes parcials

Aquesta alternativa de disseny consisteix en recodificar un dels operands per reduir el número de productes parcials. El problema de la recodificació va ser abordat per Booth l'any 1951 i en l'actualitat s'utilitza una variació d'aquell algoritme modificat, anomenat MBA (*Modified Booth Algorithm*). L'operand que es recodifica es divideix en grups de 3, 4 o més bits solapats entre si en un bit i que es converteixen, segons una taula de veritat, a dígits amb signe que indiquen les transformacions a realitzar sobre l'altre operand. La suma de les successives transformacions sobre el segon operand dóna el producte. L'algoritme anterior també és vàlid per operands codificats en complement a dos.

Tot i que és possible utilitzar grups d'un major número de bits, és preferible recodificar l'operand dividint-lo en grups de tres bits ja que així els dígits recodificats depenen tan sols d'un petit grup de bits de l'operand que es recodifica. A més, en aquesta situació les transformacions a realitzar sobre l'altre operand són molt més simples (desplaçar, complementar, etc).

L'algoritme de Booth presenta com a contrapartida un augment de la longitud dels productes parcials a sumar i una pèrdua de regularitat.

2.3. Accelerar la suma de productes parcials

La tercera alternativa per augmentar la velocitat dels multiplicadors consisteix en accelerar la suma dels productes parcials. L'esquema més utilitzat per aconseguir-ho és l'arbre de Wallace, format per un conjunt de *full-adders* (també anomenats sumadors *carry-save* o CSA), on el retard de la suma és proporcional al logaritme del número de sumands. L'inconvenient principal d'aquesta estructura és la seva poca regularitat. L'esquema d'un arbre de Wallace amb sis operands és el que es mostra a continuació.

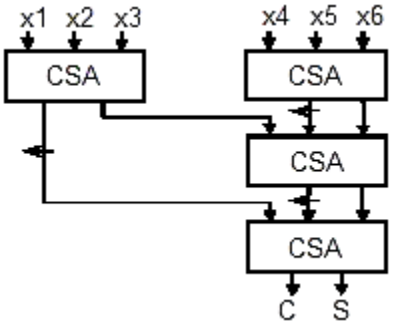


Fig. 4 Arbre de Wallace.

Una altra manera d'accelerar la suma de productes parcials s'aconsegueix mitjançant l'esquema de multiplicació de Dadda. La idea és utilitzar comptadors paral·lels per tal d'obtenir la suma dels bits que formen les columnes de la matriu de productes parcials. Es recorda que un comptador paral·lel és un circuit combinacional la sortida del qual és el número d'uns que hi ha a la seva entrada.

La diferència entre els procediments de Wallace i Dadda és que mentre el primer es fixa en les línies de la matriu de productes parcials, el segon mètode es fixa en les seves columnes.

3. CLASSIFICACIÓ DE MULTIPLICADORS

3.1. Segons l'estructura de paraula dels operands

Per estructura de paraula dels operands s'entén la distribució espai-temps dels bits que formen el multiplicand i el multiplicador. Segons com es combinin el paral·lelisme espacial amb el paral·lelisme temporal (*pipelining*), s'aconseguiran diferents distribucions i, per tant, diferents estructures de paraula. S'anomenarà dígit el grup de bits que es transmeten al mateix temps pels diferents fils del subsistema multiplicador; diferents dígits es transmetran l'un a continuació de l'altre pel mateix grup de fils. Una subparaula és senzillament un grup de dígits. Diferents subparaules estan separades en l'espai (es troben en diferents grups de fils) i en el temps (apareixen en intervals de temps no superposats entre ells).

D'aquesta manera, l'estructura d'una paraula de n bits queda determinada per tres paràmetres: número de bits per dígit (b), número de dígits per subparaula (d) i número de subparaules per paraula (s). El producte d'aquests tres paràmetres proporciona el número de bits de la paraula. Si ara s'agrupen els tres paràmetres anteriors en un vector <b,d,s> i es representa en un sistema de coordenades logarítmic, la constel·lació de possibles estructures estarà sobre un triangle equilàter. A la Fig. 5 es té una representació per una paraula de 8 bits.

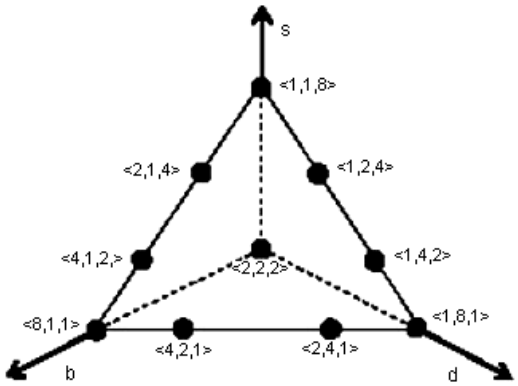


Fig. 5 Constel·lació d'estructures per una paraula de 8 bits.

S'ha de tenir en compte que b i s fan referència a la dimensió espacial del multiplicador i d a la dimensió temporal. El producte b·s és igual al número total de fils necessaris per la paraula.

Una paraula de 8 bits pot acomodar-se en 8 fils i processar-se en un únic cicle de rellotge. La seva estructura seria doncs <8,1,1>. Aquesta no és l'única solució, ja que també es pot ocupar un sol fil i executar el procés en 8 cicles de rellotge, és a dir <1,8,1>. Totes les solucions intermitges poden ser igualment vàlides. Fins i tot, en un *pipeline* de 8 etapes, els 8 fils estan permanentment ocupats per 8 subparaules corresponents als diferents operands. En aquest cas el vector seria <1,1,8>.

Els vèrtexs del triangle de la Fig. 5 donen lloc a tres famílies molt diferenciades de multiplicadors:

• Bit-sèrie. La seva estructura genèrica és la de <n,1,1>. El processat és bit a bit. Minimitzen el paral·lelisme espacial i, en conseqüència, presenten una gran economia en àrea però velocitat limitada.

• Bit-paral·lel. La seva estructura genèrica és <1,n,1>. Presenten paral·lelisme espacial, per la qual cosa obtenen majors velocitats a canvi d'un major consum en àrea.

• Bit sèrie-paral·lel. És una barreja entre els dos tipus anteriors. La seva estructura és la que es veu a la Fig. 6.

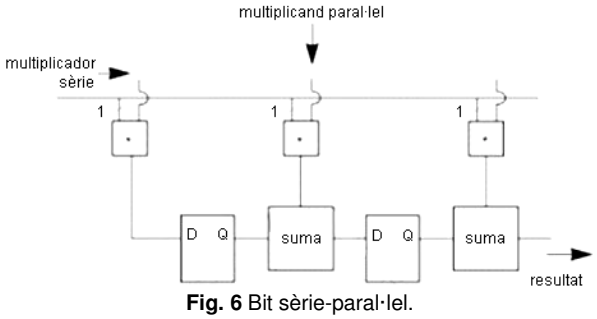


Fig. 6 Bit sèrie-paral·lel.

• Bit-paral·lel amb *pipeline*. La seva estructura genèrica és <1,1,n>. En aquest cas, el processat es realitza en grups de n bits. Es maximitzen els paral·lelismes espacial i temporal, per la qual cosa s'obté la màxima velocitat a canvi d'un gran consum en àrea.

3.2. Segons l'algoritme de multiplicació

Existeixen dues classes de multiplicadors segons l'agoritme utilitzat. El primer tipus de multiplicadors són aquells que en una primera etapa utilitzen l'algoritme de Booth per tal de reduir el nombre de productes parcials a sumar. Després, es realitza la suma d'aquests productes amb *full-adders* o bé amb comptadors paral·lels. El cor de l'estructura bàsica d'un multiplicador que implementi l'algoritme de Booth es mostra a la Fig. 7.

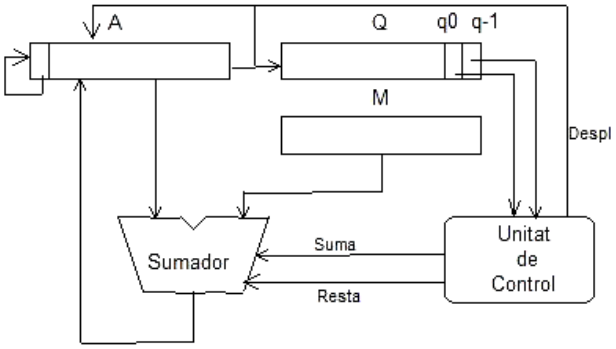


Fig. 7 Algoritme de Booth.

Un multiplicador de n bits que utilitza l'algoritme de Booth realitza la recodificació que es descriu tot seguit. Primer, se suma A+M a A si q_0q_{-1}=01; en canvi, si aquests dos bits valen 10, es resta A-M a A. Després, es desplaça AQ una posició a la dreta. Aquest procés es repeteix n vegades. En el cas de l'algoritme de Booth modificat, la recodificació és la de la taula següent:

$Y_i\ Y_{i-1}\ Y_{i-2}$	increment
000	0
001	x
010	x, i. e. 2x-2=x
011	2x
100	-2x
101	-x, i. e. -2x+x=-x
110	-x
111	0

Taula 1 Recodificació segons MBA.

Un multiplicador que utilitza repetidament algoritmes de desplaçament i suma per calcular els productes és un multiplicador iteratiu. Aquest segon tipus de multiplicadors es construeix a partir d'un processador elemental, que constitueix una cel·la bàsica del multiplicador i es repeteix per formar un *array* regular. Si cada cel·la només es comunica amb les veïnes, es diu que el multiplicador és local. Si, a més, es comunica amb altres cel·les serà global.

3.3. Segons el sistema de representació numèrica

D'acord amb el sistema de representació numèrica

utilitzat per representar els operands i els resultats, es poden classificar els multiplicadors en quatre tipus: signe i magnitud (S+M), complement a dos, *residue numbering system* (RNS) i *signed-digit*. A continuació, es descriurà cadascun d'ells.

Els multiplicadors S+M són aquells que treballen en valor absolut i calculen de forma separada el signe del resultat. Es caracteritzen per aprofitar al màxim la regularitat de la multiplicació. D'altra banda, és possible transformar un multiplicador iteratiu per números sense signe en un multiplicador que operi en complement a dos mitjançant algoritmes adequats. Això pot resultar interessant en alguns casos, però s'ha de tenir en compte que aquesta transformació disminuirà la regularitat del circuit multiplicador. El més adequat en aquest cas és utilitzar l'algoritme de Booth modificat.

Els multiplicadors RNS representen un número sencer mitjançant els mòduls de la divisió sencera d'aquest número amb un conjunt de números primers entre ells. Utilitzant aquesta tècnica és possible realitzar multiplicacions sense haver de propagar cap informació de *carry* entre els diferents dígits. Amb freqüència aquests multiplicadors s'implementen amb *look-up tables* (LUTS) tot i que també poden implementar-se d'altres maneres. Finalment, existeix un darrer sistema de numeració anomenat *signed-digit*, que permet que cadascun dels dígits d'un número tingui associat un signe positiu o negatiu. En aquest mètode no existeixen cadenes de propagació pel *carry*.

4. IMPLEMENTACIÓ DE MULTIPLICADORS

Principalment, existeixen tres formes d'implementar sistemes multiplicadors. El fet de decidir-se per una d'aquestes implementacions en concret dependrà del grau de prestacions desitjat i dels aspectes econòmics involucrats en el disseny i en la producció del multiplicador (temps de disseny, àrea de silici, volum de producció, etc).

4.1. Circuits VLSI específics

Els multiplicadors combinacionals solen utlitzar-se només en sistemes d'altes prestacions, especialment els dedicats al processat de senyal, on cal fer moltes multiplicacions i molt ràpidament. El dissenyador realitza tots els passos de màscara seguint unes regles de disseny en concret i envia a la *foundry,* utilitzant el suport informàtic adient, la descripció del circuit. Degut al seu disseny específic, la seva optimització en àrea per realitzar multiplicacions binàries és molt elevada, però per contrapartida exigeix un temps de disseny superior a les altres implementacions. Implementar multiplicadors amb circuits VLSI és adequat per grans volums de producció o per aplicacions on el grau de prestacions sigui crític.

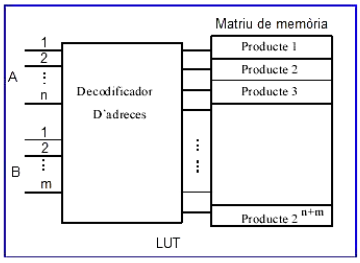
4.2. FPGAs

Aquesta alternativa és una altra manera d'implementar un multiplicador dissenyat per una aplicació en concret, és a dir, que substitueix l'arquitectura genèrica dels multiplicadors. Fa alguns anys era impossible per volums petits o mitjans, però actualment la reducció de costos introduïda per les FPGAS l'ha convertit en una implementació factible de màquines de càlcul realitzades completament a mida de l'aplicació, inclús per volums de producció petits.

El mètode de disseny és el que es descriu a continuació. S'utilitza una descripció VHDL, generada automàticament per algun sistema generador de multiplicadors, per cada tipus de multiplicador. D'aquestes descripcions VHDL en RTL dels multiplicadors se n'extreu, mitjançant el *software* adequat, un arxiu que conté una llista amb totes les connexions. Finalment, aquests arxius s'introdueixen en un altre *software* específic per tal de realitzar els processos de *place&route*, optimització i anàlisi de retards. El pas següent és analitzar les materialitzacions del multiplicador en termes d'àrea, velocitat, freqüència màxima de funcionament, portes lògiques i quantitat de *flip-flops*. El *software* utilitzat és específic de cada empresa fabricant.

4.3. Taules LUT

En aquest cas, es genera la taula de veritat dels n+m bits dels factors A i B i del seu producte. El valor del producte entre aquests dos factors es guarda en una memòria (una ROM, per exemple) i s'hi accedeix introduint-hi una adreça. A la memòria hi ha emmagatzemada la taula de multiplicar de totes les possibles combinacions.



A B	P
0000 0000	00000000
0000 0001	00000000
....	...
0010 0001	00000010
0010 0010	00000100
0010 0011	00000110
...	...
1111 1101	11000011
1111 1110	11010011
1111 1111	11100001

Fig. 8 Implementació d'un multiplicador en una LUT.

Se suposa, per exemple, una multiplicació on els dos operands són de n i m bits. Es disposa d'una ROM amb n+m entrades i 2^{n+m} posicions de memòria. La capacitat total de la ROM serà $(n+m)\cdot 2^{n+m}$. Si els operands són de vuit bits, la capacitat és 2^{20}=1Mb. És important reduir les mides de les taules al màxim per tal de reduir el tamany de la memòria. Per fer-ho, només cal aprofitar les propietats de la multiplicació, com són la commutativa, la multiplicació per zero, etc. El cost d'això és que cal introduir un circuit combinacional per preprocessar els operands. Per aprofitar la propietat commutativa de la multiplicació, per exemple, cal introduir un bloc lògic que ordeni els operands de manera que el menor sigui sempre el primer.

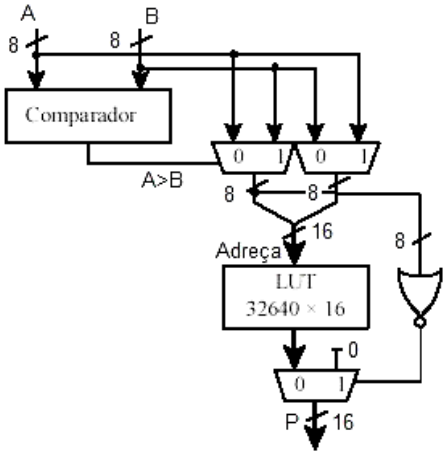


Fig. 9 Tècniques de reducció de memòria.

En aquestes condicions la taula LUT requerirà aproximadament la meitat de la memòria que necessitava abans. També poden aplicar-se tècniques d'arrodoniment o de partició en productes parcials per reduir encara més la memòria necessària. El producte anterior es pot efectuar dividint la LUT en d'altres LUTS més petites. Per això, es parteixen el multiplicand i el multiplicador en seccions disjuntes i es fan els productes parcials amb LUTS. En el cas d'una multiplicació de dos números de vuit bits, A i B, aquests es poden partir en dues parts tal com s'explica a continuació.

Inicialment, es tenen els dos números de vuit bits $A=\{a_8a_7a_6a_5a_4a_3a_2a_1\}$ i $B=\{b_8b_7b_6b_5b_4b_3b_2b_1\}$. Tot seguit, es modifiquen aquests nombres en $A_{58}=\{a_8a_7a_6a_50000\}$ i $A_{14}=\{0000a_4a_3a_2a_1\}$, $B_{58}=\{b_8b_7b_6b_50000\}$ i $B_{14}=\{0000b_4b_3b_2b_1\}$, on $A=A_{58}+A_{14}$ i $B=B_{58}+B_{14}$. Aleshores, $AxB=(A_{58}+A_{14})(B_{58}+B_{14})=A_{58}B_{58}+A_{58}B_{14}+A_{14}B_{58}+A_{14}B_{14}$. En aquesta nova situació, caldran 4 LUTS per efectuar els productes parcials i sumar-los adequadament, amb 4 sumadors de 4 bits, tal com es mostra a la Fig. 10.

El tamany de memòria necessària és de $(4+4)\cdot 2^8=2Kb$ per cada LUT. Per tant, en total calen 8Kb. Si, a més, s'aplica la reducció de la propietat commutativa i la multiplicació per zero, queda una mida total de memòria requerida de 3.75Kb. Aplicant les tècniques de reducció anteriors, s'ha passat d'1Mb inicial a 3.75Kb.

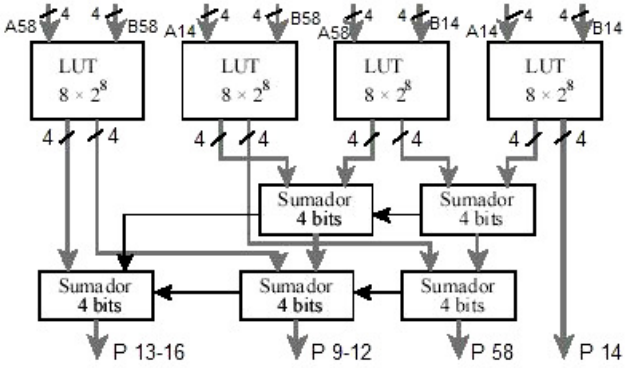


Fig. 10 Divisió en LUTS més petites.

5. CONCLUSIONS

Un multiplicador fonamental efectua l'operació de multiplicar amb JxK portes AND i (J-1) sumadors de K bits, on J i K són el número de bits del multiplicador i del multiplicand, respectivament. A partir d'aquest multiplicador fonamental es poden fer diverses modificacions per tal d'optimitzar la velocitat d'execució, com són l'acceleració de la creació o de la suma de productes parcials. El primer mètode precisa d'un *array* de portes AND per realitzar simultàniament tots els productes parcials; el segon mètode, en canvi, suposa l'addició de sumadors si s'utilitza l'arbre de Wallace o de comptadors paral·lels, en el cas que es faci servir l'esquema de multiplicació Dadda. Ambdues tècniques, però, comporten un increment de l'àrea ocupada. Existeix un tercer mètode que es basa en la reducció del nombre de productes parcials a partir de l'algoritme de Booth. Aquesta tècnica implica una pèrdua de regularitat.

Un multiplicador pot ser implementat segons tres estructures diferents. Els circuits VLSI específics ofereixen unes prestacions molt elevades, però requereixen un alt temps de disseny, cosa que fa que els multiplicadors implementats amb aquesta estructura siguin adequats només per grans volums de producció. Les FPGAS, en canvi, són una alternativa rentable inclús a petita escala. Les taules LUT, que constitueixen la darrera opció, guarden en memòria el resultat dels productes de totes les combinacions possibles entre dos factors. És molt important reduir al màxim el nombre de termes a emmagatzemar per tal de disminuir el tamany de la memòria. S'accedeix al producte desitjat mitjançant un descodificador d'adreces.

6. REFERÈNCIES

[Herms] A. Herms, S. Bota, S. Bosch, J. M. López Villegas, "Disseny Digital Bàsic", Textos docents núm 126, Edicions de la Universitat de Barcelona.

[Boemo93] E. I. Boemo , E. Juárez, J. Meneses, "Taxonomía de Multiplicadores", Published at the Proc. VIII DCIS Conference, Universidad de Málaga, 1993.

[Acosta] N. Acosta, C. Collado, K. Larsen, “Métricas para la generación de Multiplicadores paralelos en FPGA”, INCA – Departamento de Computación y Sistemas, Facultad de Ciencias Exactas, Universidad Nacional del Centro de la Provincia de Buenos Aires.

[Herms] A. Herms, “Sistemes Digitals”, Textos docents núm 172, Edicions de la Universitat de Barcelona.

P1. Problemes

1. La funció F=AB+AC+BC correspon a la funció *carry* complementat d'un sumador de dos bits. Dissenyeu una porta CMOS que realitzi aquesta funció.

Cal tenir en compte que, com que la lògica CMOS és inversora i la funció F és no negada, caldrà incloure un inversor a la sortida.

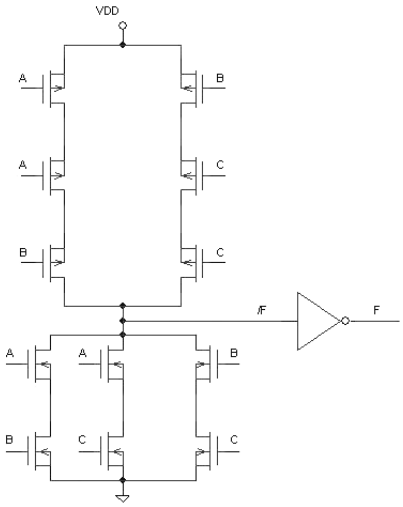


Fig. 1 Esquema a nivell de transistors.

2. Trobeu la longitud L per l'amplada mínima permesa per cadascuna de les capes conductores d'un circuit. Feu-ho pels tres casos següents:

(i) la resistència paràsita de la capa és R=100Ω.
(ii) la capacitat paràstia és C=100fF.
(iii) es té un producte RC=1ns.

Utilitzeu l'expressió $C=C_{j0}\cdot L\cdot W+C_{jw}\cdot 2(W+L)$ per calcular les capacitats de les difusions. Utilitzeu també les dades de la taula adjunta.

Capa	Amplada mínima (μm)	R□ (Ω /□)	C_s (μF/m^2)	C_j (F/m^2)	C_{sjw} (F/m)
Àrea activa N	1	55	-	$350\text{x}10^{-6}$	$450\text{x}10^{-12}$
Àrea activa P	1	70	-	$540\text{x}10^{-6}$	$760\text{x}10^{-12}$
Polisilici	1	33	58	-	-
Metall 1	1.5	$60\text{x}10^{-3}$	29	-	-
Metall 2	1.5	$30\text{x}10^{-3}$	17	-	-

Taula 1 Dades del circuit.

(i) $R = R'\dfrac{L}{W} \rightarrow L = W\dfrac{R}{R'}$, on R'=R□.

(ii) Per les àrees actives N i P: $C = C_{j0}\cdot L\cdot W + C_{jw}\cdot 2(W+L) \rightarrow L = \dfrac{C-2C_{jw}W}{C_{j0}W+2C_{jw}}$

Per M1, M2 i el polisilici: $C = C_{ox}\cdot W\cdot L \rightarrow L = \dfrac{C}{C_{ox}W}$

(iii) Pel cas en què es té un producte RC de 1ns, només cal multiplicar per R l'expressió per les capacitats de les difusions. Resolent l'equació de segon grau que s'obté i substituint els paràmetres pels seus valors numèrics, es troba la longitud L demanada.

$$RC = \left(R' \frac{L}{W} \right) \left[C_{j0} \cdot L \cdot W + C_{jw} \cdot 2(W + L) \right] = R' C_{j0} L^2 + 2R' C_{jw} L + \frac{2R' C_{jw} L^2}{W}$$

3. Obteniu el corrent màxim que pot circular per un transistor nMOS de mides mínimes (L=1µm, W=1µm) i també per un transistor pMOS. Suposeu que la tensió de funcionament pot variar entre 0 i 5V.

El corrent màxim que circula per un transistor MOS es donarà a la regió de saturació. Així doncs, substituint els valors de l'enunciat a l'expressió per la intensitat de saturació:

(i) Pel nMOS:

$$I_{D,sat} = \frac{\beta_n}{2} (V_{GS} - V_{Tn})^2 \text{, on } \beta_n = K_n (W/L)_n = \mu_n C_{ox} (W/L)_n = 20 \frac{\mu A}{V^2}$$

$$I_{D,sat} = \frac{20 \mu A/V^2}{2} (5 - 0.75V)^2 = 180.6 \mu A$$

(ii) Pel pMOS:

$$I_{D,sat} = \frac{\beta_p}{2} \left(|V_{GS}| - |V_{Tp}| \right)^2 \text{, on } \beta_p = K_p (W/L)_p = \mu_p C_{ox} (W/L)_p = 8 \frac{\mu A}{V^2}$$

$$I_{D,sat} = \frac{8 \mu A/V^2}{2} (5 - 0.9V)^2 = 67.24 \mu A$$

4. Per un cert procés CMOS es demana que:
(i) Dissenyeu un inversor CMOS que verifiqui V_{th}=2.5V quan la polarització és de 5V.
(ii) Quin és el nou valor de V_{th} si es canvia la polarització a 10V?
(iii) Calculeu V_{IL} i V_{IH}.
Dades: V_{Tn}=0.8V, V_{Tp}=-0.9V, $K_n=\mu_n \cdot C_{ox}$=40µA/V², $K_p=\mu_p \cdot C_{ox}$=16µA/V².

(i) Es demana obtenir la relació entre W_p i W_n per tal que V_{th}=2.5V, que és la tensió de commutació, quan V_{DD}=5V. En el punt de commutació, tant el pMOS com el nMOS es troben en règim de saturació. Igualant les expressions per les seves intensitats de *drain*:

$$I_{Dn} = I_{Dp}$$

$$\frac{1}{2} \beta_n (V_{GSn} - V_{Tn})^2 = \frac{1}{2} \beta_p \left(|V_{GSp}| - |V_{Tp}| \right)^2$$

$$\frac{1}{2} \beta_n (V_{th} - V_{Tn})^2 = \frac{1}{2} \beta_p \left(V_{DD} - V_{th} - |V_{Tp}| \right)^2$$

$$V_{th} = \frac{V_{Tn} + \sqrt{\beta_p / \beta_n} \left(V_{DD} - |V_{Tp}| \right)}{\left(1 + \sqrt{\beta_p / \beta_n} \right)}$$

Substituint numèricament els termes de l'equació anterior es troba el valor de β (on $\beta=(\beta_p/\beta_n)^{1/2}$). Suposant que les longituds de les difusions n i p són iguals, i tenint en compte el valor de β, es troba la relació entre amplades demanada a l'enunciat.

$$\left.\begin{aligned} 2.5 &= \frac{0.8+4.1\beta}{1+\beta} \rightarrow \beta = 1.602 \\ \frac{\beta_p}{\beta_n} &= \frac{K_p(W/L)_p}{K_n(W/L)_n} = \frac{1}{2.5}\frac{W_p}{W_n} \end{aligned}\right\} \rightarrow \beta^2 = 1.602^2 = \frac{\beta_p}{\beta_n} = \frac{1}{2.5}\frac{W_p}{W_n} \rightarrow W_p = 2.82W_n$$

(ii) Si es canvia la polarització a 10V, per trobar la nova tensió de commutació només cal substituir valors a l'expressió per V_{th} obtinguda a l'apartat anterior, tenint en compte que ara V_{DD}=10V.

$$V_{th} = \frac{V_{Tn} + \sqrt{\beta_p/\beta_n}\left(V_{DD} - |V_{Tp}|\right)}{\left(1+\sqrt{\beta_p/\beta_n}\right)} = \frac{0.8 + 1.062(V_{DD} - 0.9)}{2.062} = 5.07V$$

(iii) Primer de tot, es calcula V_{IL}. El transistor nMOS es troba en règim de saturació i el pMOS, en règim lineal. Igualant les expressions pels seus corrents de *drain* i derivant a continuació es té:

$$\frac{1}{2}\beta_n(V_{GSn} - V_{Tn})^2 = \beta_p\left[\left(|V_{GSp}| - |V_{Tp}|\right)|V_{DSp}| - \frac{|V_{DSp}|^2}{2}\right]$$

$$\frac{1}{2}\beta_n(V_{IL} - V_{Tn})^2 = \beta_p\left[\left((V_{DD} - V_{IL}) - |V_{Tp}|\right)(V_{DD} - V_{out}) - \frac{(V_{DD} - V_{out})^2}{2}\right]$$

$$\frac{\partial I_{Dp}}{\partial V_{in}}dV_{in} + \frac{\partial I_{Dp}}{\partial V_{out}}dV_{out} = \frac{\partial I_{Dn}}{\partial V_{in}}dV_{in} + \frac{\partial I_{Dn}}{\partial V_{out}}dV_{out}$$

Com que el corrent de *drain* del transistor nMOS no depèn de la tensió de sortida, l'últim terme de l'expressió anterior és nul. Reagrupant termes i tenint en compte que en el punt en què $V_{in}=V_{IL}$, dV_{out}/dV_{in}=-1:

$$\frac{dV_{out}}{dV_{in}} = \frac{\frac{\partial I_{Dn}}{\partial V_{in}} - \frac{\partial I_{Dp}}{\partial V_{in}}}{\frac{\partial I_{Dp}}{\partial V_{out}}} = -1 \rightarrow -\frac{\partial I_{Dp}}{\partial V_{out}} = \frac{\partial I_{Dn}}{\partial V_{in}} - \frac{\partial I_{Dp}}{\partial V_{in}}$$

Ara només cal calcular les derivades i substituir-les a l'expressió anterior.

$$\begin{cases} \frac{\partial I_{Dp}}{\partial V_{out}} = -\beta_p\left(V_{DD} - V_{IL} - |V_{Tp}|\right) + \beta_p(V_{DD} - V_{out}) \\ \frac{\partial I_{Dn}}{\partial V_{in}} = \beta_n(V_{IL} - V_{Tn}) \\ \frac{\partial I_{Dp}}{\partial V_{in}} = -\beta_p(V_{DD} - V_{out}) \end{cases}$$

$$-\beta_p\left(V_{IL}+\left|V_{Tp}\right|-V_{out}\right)=\beta_n\left(V_{IL}-V_{Tn}\right)+\beta_p\left(V_{DD}-V_{out}\right)$$
$$-\left|V_{Tp}\right|+2V_{out}-V_{DD}+\left(\beta_n/\beta_p\right)V_{Tn}=V_{IL}\left(1+\beta_n/\beta_p\right)$$
$$V_{IL}=\frac{2V_{out}+\left(\beta_n/\beta_p\right)V_{Tn}-V_{DD}-\left|V_{Tp}\right|}{1+\beta_n/\beta_p}$$

Aquí s'ha trobat una solució per V_{IL} en funció de V_{out}. Per trobar el valor numèric de V_{IL}, cal conèixer també el valor de V_{out}. Per aquest propòsit, ha de resoldre's un sistema de dues equacions i dues incògnites tal com es mostra tot seguit:

$$-\left|V_{Tp}\right|+2V_{out}-V_{DD}+\left(\beta_n/\beta_p\right)V_{Tn}=V_{IL}\left(1+\beta_n/\beta_p\right)\rightarrow 1.88V_{IL}=2V_{out}-5.19$$
$$0.88\left(V_{IL}-0.8\right)^2=2\left(4.1-V_{IL}\right)\left(5-V_{out}\right)-\left(5-V_{out}\right)^2$$
$$V_{out}=4.58V$$
$$V_{IL}=2.11V$$

Tot seguit, es calcula V_{IH}. Aquí es té que el transistor nMOS es troba a la regió lineal i el pMOS, a la de saturació. Similarment al cas anteior, igualant les expressions pels seus corrents de *drain* i derivant, tot tenint en compte que $dV_{out}/dV_{in}=-1$ quan $V_{in}=V_{IH}$, s'obté:

$$\frac{\beta_n}{2}\left[2\left(V_{IH}-V_{Tn}\right)V_{out}-V_{out}^2\right]=\frac{\beta_p}{2}\left[V_{DD}-V_{IH}-\left|V_{Tp}\right|\right]$$
$$V_{IH}\left(1+\beta_p/\beta_n\right)=2V_{out}+V_{Tn}+\left(\beta_p/\beta_n\right)\left(V_{DD}-\left|V_{Tp}\right|\right)$$

Realitzant un sistema de dues equacions i dues incògnites i substituint numèricament, es troba primer la tensió de sortida quan $V_{in}=V_{IH}$, cosa que permet trobar la pròpia V_{IH}.

$$0.88\left[2\left(V_{IH}-0.8\right)V_{out}-V_{out}^2\right]=\left(4.1-V_{IH}\right)^2$$
$$2.128\cdot V_{IH}=2V_{out}+5.42$$
$$V_{out}=0.16V$$
$$V_{IH}=2.69V$$

5. Considereu el *layout* del transistor MOSFET de la Fig. 2. Calculeu C_0, C_{j0}, C_{jsw}, C_{SB}, C_{DB} i C_G. Dades: $N_A=10^{15}cm^{-3}$, $N_{Asw}=2\cdot10^{16}cm^{-3}$, $N_D=10^{20}cm^{-3}$, $d_{ox}=500Å$, $x_j=0.5\cdot10^{-6}m$, $L_{OL}=0.4\cdot10^{-6}m$.

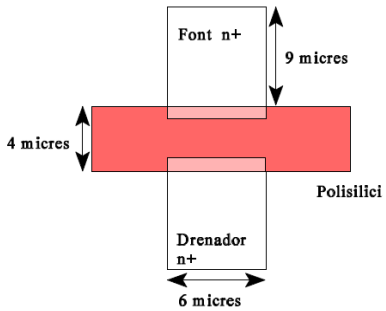


Fig. 2 *Layout* d'un transistor MOSFET.

(i) Primer de tot, es calcula C_o, que és el producte de la capacitat d'òxid per la longitud d'àrea activa superposada. També es calcularan les capacitats de superposició, tant per *source* com per *drain*, que seran idèntiques.

$$C_o = C_{ox} \cdot L_{OL}, \text{ on } C_{ox} = \frac{\varepsilon_{ox}}{d_{ox}} = \frac{\varepsilon_{SiO_2} \cdot \varepsilon_0}{d_{ox}} = \frac{3.9\left(8.854 \cdot 10^{-14}\right) F/cm}{0.05 \cdot 10^{-4}\, cm} = 6.91 \cdot 10^{-8}\ F/cm^2$$

$$C_o = \left(6.91 \cdot 10^{-8}\ F/cm^2\right)\left(0.4 \cdot 10^{-4}\, cm\right) = 0.276\ fF/cm$$

A partir de l'expressió per la capacitat de superposició i substituint:

$$C_{OLD} = C_{OLS} = C_{ox} \cdot L_{OL} \cdot W = \left(0.276\ fF/cm\right) 6 \cdot 10^{-4}\, cm = 1.66\, fF$$

(ii) Tot seguit, es calculen les capacitats de *source* i *drain* amb el substrat. Aquestes presenten una contribució deguda a l'àrea (tapa inferior) i una altra deguda al perímetre (lateral).

$$C_{SB} = C_{DB} = C_{j0} \cdot L_D \cdot W + C_{jsw} \cdot 2\left(W + L_D\right)$$

Per fer aquest càlcul, cal conèixer C_{j0}, que és la capacitat de la unió associada a l'àrea:

$$C_{j0} = \sqrt{\frac{\varepsilon_{Si} \cdot q}{2\phi_B}\left(\frac{N_A N_D}{N_A + N_D}\right)}, \text{ on } \phi_B = \frac{KT}{q} \cdot \ln\left(\frac{N_D \cdot N_A}{n_i^2}\right) = 0.878V$$

$$C_{j0} = 9.7\, nF/cm$$

i C_{jsw}, que és la capacitat de la unió associada al perímetre:

$$C_{jsw} = C_{jdw} = C_{j0w} \cdot x_j$$

$$C_{j0w} = \sqrt{\frac{q \cdot \varepsilon_{Si}}{2\phi_{0w}}\left(\frac{N_{Asw} N_D}{N_{Asw} + N_D}\right)} = 4.18 \cdot 10^{-18}\ F/cm^2$$

$$\phi_{0w} = \frac{KT}{q} \cdot \ln\left(\frac{N_D \cdot N_{Asw}}{n_i^2}\right) = 0.96V$$

$$C_{jsw} = 2.09\ pF/cm$$

Substituint C_{j0} i C_{jsw} a l'expressió per C_{SB} i C_{DB}, tot tenint en compte que L=9µm i W=6µm, es té:

$$C_{SB} = C_{DB} = 11.5\, fF$$

(iii) En aquest últim apartat, es calcula la capacitat de porta tal com es mostra a continuació. C_{ox} s'ha calculat a l'apartat (i).

$$C_G = C_{ox} \cdot L \cdot W = \left(6.98 \cdot 10^{-8}\right)\left(6 \cdot 10^{-4}\right)\left(4 \cdot 10^{-4}\right) = 16.58\, fF$$

6. Un inversor CMOS és dissenyat amb $\beta_n=\beta_p=35\mu A/V^2$ i $V_{Tn}=0.9V$, $V_{Tp}=-0.8V$. La capacitat de sortida de la porta és aproximadament igual a 125fF i $V_{DD}=5V$. Es demana:

(i) Calculeu t_{HL} i t_{LH}.

(ii) Calculeu t_p.

(iii) Estimeu PDP quan s'aplica una ona quadrada al circuit amb una freqüència $f=0.1f_{màx}$.

(i) Primer de tot, cal conèixer els valors de V_0, V_1 i $V_{1/2}$. Donat que V_{DD}=5V i que la tecnologia utilitzada és CMOS, V_{OL}=0V i V_{OH}=5V.

$$V_0 = V_{OL} + 0.1(V_{OH} - V_{OL}) = 0.5V$$
$$V_1 = V_{OL} + 0.9(V_{OH} - V_{OL}) = 4.5V$$
$$V_0 = V_{OL} + 0.5(V_{OH} - V_{OL}) = 2.5V$$

Per calcular el temps de commutació de baixada es parteix de l'expressió pel corrent de sortida:

$$I_C = -C_L(dV_{out}/dt)$$
$$\int_{t_1}^{t_2} dt = -\int_{V_1}^{V_0} C_L \frac{dV_{out}}{I_D}$$
$$t_{HL} = t_f = \frac{C_L}{\beta_n(V_1 - V_{Tn})}\left[\frac{2V_{Tn}}{V_1 - V_{Tn}} + \ln\left(\frac{2(V_1 - V_{Tn})}{V_0} - 1\right)\right]$$
$$t_{HL} = t_f = \frac{125fF}{126\mu A/V}\left[\frac{1.8}{3.6} + \ln 13.4\right] = 3.07ns$$

Similarment, pel temps de commutació de pujada:

$$t_{LH} = t_r = \frac{C_L}{\beta_p(V_1 - |V_{Tp}|)}\left[\frac{2|V_{Tp}|}{V_1 - |V_{Tp}|} + \ln\left(\frac{2(V_1 - |V_{Tp}|)}{V_0} - 1\right)\right]$$
$$t_{LH} = t_r = \frac{125fF}{129.5}\left[\frac{1.6}{3.7} + \ln 13.8\right] = 2.95ns$$

(ii) El retard és el promig del temps de propagació de pujada i del temps de propagació de baixada.

$$t_p = \frac{1}{2}(t_{pLH} + t_{pHL})$$

Pel temps de propagació de pujada:

$$t_{pHL} = -\int_{V_{OH}}^{V_{1/2}} C_L \frac{dV_{out}}{I_C}$$
$$t_{pHL} = \frac{C_L}{\beta_n(V_{OH} - V_{Tn})}\left[\frac{2V_{Tn}}{V_{OH} - V_{Tn}} + \ln\left(\frac{4(V_{OH} - V_{Tn})}{V_{OH} + V_{OL}} - 1\right)\right] = 1.10ns$$

Similarment, pel temps de propagació de baixada:

$$t_{pLH} = +\int_{V_{OL}}^{V_{1/2}} C_L \frac{dV_{out}}{I_D}$$

$$t_{pLH} = \frac{C_L}{\beta_p\left(V_{OH} - |V_{Tp}|\right)}\left[\frac{2|V_{Tp}|}{V_{OH} - |V_{Tp}|} + \ln\left(\frac{4\left(V_{OH} - |V_{Tp}|\right)}{V_{OH} + V_{OL}} - 1\right)\right] = 1.05ns$$

Finalment, substituint a l'expressió pel retard:

$$t_p = 1.075ns$$

(iii) Per calcular la potència total consumida pel circuit, cal tenir en compte que aquesta depèn de:

$$P_{total} = P_s + P_d + P_{sc}$$

En CMOS, la potència estàtica consumida és zero. La potència dinàmica, en canvi, ve donada per:

$$P_d = C_L \cdot V_{DD}^2 \cdot f_{CLK} = 0.3125 \cdot f_{màx}\, p(W \cdot s)$$

La potència de curtcircuit s'obté a partir de l'expressió:

$$P_{sc} = \frac{\beta}{12}\left(V_{DD} - 2V_T\right)^3 \frac{\tau}{T_{CLK}}$$

on V_T correspon al promig entre V_{Tn} i $|V_{Tp}|$, ζ és el promig dels temps de commutació de baixada i de pujada i T_{CLK} és el període del rellotge. Substituint, es troba que:

$$P_{sc} = \frac{35\,\mu A/V^2}{12}\left(5 - 2 \cdot 0.85V\right)^3 \frac{3.01ns}{10\left(f_{màx}\right)^{-1}} = 31.4 \cdot f_{màx} f\left(W \cdot s\right)$$

Sumant la potència dinàmica i la de curtcircuit s'obté la potència total consumida pel circuit. Això és:

$$P_{total} = 3.44 \cdot f_{màx}\, p(W \cdot s)$$

7. En el circuit de la Fig. 3, calculeu W_{n2} en funció de W_{n1} i de C_L per tal que el temps de propagació del sistema sigui mínim.
Dades: $L_n=L_p$=longitud mínima, $V_{Tn}=|V_{Tp}|$, $K_n=3K_p$, $C_L>>C_{wire}+C_{gb2}$.

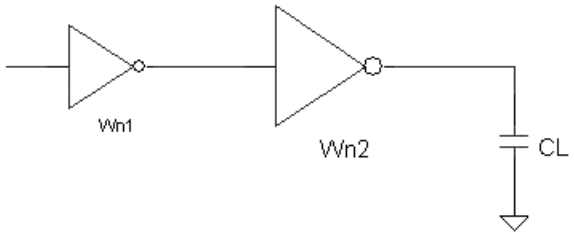


Fig. 3 Diseny a nivell de portes lògiques del *buffer*.

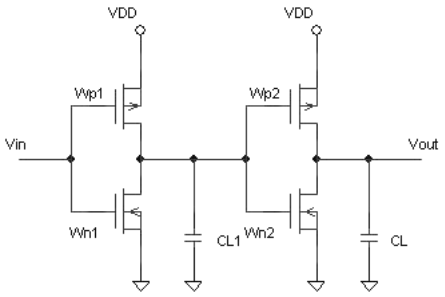


Fig. 4 Disseny a nivell de transistors del *buffer*.

Es tenen dos inversors de mides diferents. El primer, té les dimensions idònies per carregar ràpidament el segon, mentre que aquest té les dimensions adequades per carregar ràpidament la capacitat de càrrega. El temps de propagació total del sistema ve donat per:

$$t_p = t_{p1} + t_{p2}\text{, on } t_{pi} = \frac{1}{2}\left(t_{pHL} + t_{pLH}\right)$$

Els temps de propagació poden reescriure's tal com es mostra a continuació. Els termes R_{pHL} i R_{pLH} seran una constant pel circuit aquí estudiat.

$$t_{pHL} = \frac{C_L}{(W/L)_n} R_{pHL}\text{, on } R_{pHL} = \frac{1}{K_n\left(V_{OH} - V_{Tn}\right)}\left[\frac{2V_{Tn}}{V_{OH} - V_{Tn}} + \ln\left(\frac{4\left(V_{OH} - V_{Tn}\right)}{V_{OH} + V_{OL}} - 1\right)\right]$$

$$t_{pLH} = \frac{C_L}{(W/L)_p} R_{pLH}\text{, on } R_{pLH} = \frac{1}{K_p\left(V_{OH} - \left|V_{Tp}\right|\right)}\left[\frac{2\left|V_{Tp}\right|}{V_{OH} - \left|V_{Tp}\right|} + \ln\left(\frac{4\left(V_{OH} - \left|V_{Tp}\right|\right)}{V_{OH} + V_{OL}} - 1\right)\right]$$

Primer de tot, cal calcular el temps de propagació corresponent al primer inversor:

$$t_{p1} = \frac{1}{2}\left[\frac{R_{pHL}}{(W/L)_n} + \frac{R_{pLH}}{(W/L)_p}\right] C_{L1}$$

La capacitat de càrrega del primer inversor està formada per:

$$C_{L1} = C_{G2} + C_{D1} + C_{wire}$$

La capacitat associada al cable (C_{wire}) és molt petita i podrà despreciar-se. Les capacitats de *gate* i de *drain* vénen donades per:

$$C_{G2} = C_{ox} \cdot L \cdot \left(W_{n2} + W_{p2}\right)$$
$$C_{D1} = C_{j0} \cdot L \cdot \left(W_{n1} + W_{p1}\right) + C_{jsw} \cdot 2\left(W_{n1} + W_{p1} + L\right)$$

Com que $W_2 >> W_1$, $C_{G2} >> C_{D1}$. Per tant, la capacitat de càrrega del primer inversor pot reescriure's com:

$$C_{L1} \cong C_{G2}$$

Com que ambdós inversors presenten una relació $K_n = 3K_p$ i tenen la mateixa longitud de difusió, es tindrà que:

$$W_{p1} = 3W_{n1}$$
$$W_{p2} = 3W_{n2}$$

Així, substituint a l'expressió per t_{p1}, tot tenint en compte que aquí s'ha associat 3 a la constant γ:

$$t_{p1} = \frac{1}{2}\left[\frac{R_{pHL}}{(W/L)_{n1}} + \frac{R_{pLH}}{(W/L)_{p1}}\right]C_{L1} = \frac{1}{2}\left[\frac{R_{pHL}}{(W/L)_{n1}} + \frac{R_{pLH}}{(W/L)_{p1}}\right]C_{ox}\cdot L\left(W_{n2} + W_{p2}\right) =$$
$$= \frac{1}{2}\left[\frac{R_{pHL}}{W_{n1}} + \frac{R_{pLH}}{\gamma\cdot W_{n1}}\right]C_{ox}\cdot L^2\left(W_{n2} + \gamma\cdot W_{n2}\right) = \frac{1}{2}C_{ox}\cdot L^2\left(\gamma+1\right)\left[R_{pHL} + \frac{R_{pLH}}{\gamma}\right]\frac{W_{n2}}{W_{n1}} = a_1\frac{W_{n2}}{W_{n1}}$$

La capacitat de càrrega pel segon inversor és C_L. Així, el temps de propagació associat al segon inversor serà:

$$t_{p2} = \frac{1}{2}\left[\frac{R_{pHL}}{(W/L)_{n2}} + \frac{R_{pLH}}{(W/L)_{p2}}\right]C_{L2} = \frac{1}{2}L\left[\frac{R_{pHL}}{W_{n2}} + \frac{R_{pLH}}{3W_{n2}}\right]C_L = a_2\frac{C_L}{W_{n2}}$$

El temps total de propagació és la suma de t_{p1} i t_{p2}:

$$t_p = a_1\frac{W_{n2}}{W_{n1}} + a_2\frac{C_L}{W_{n2}}$$

Per trobar les dimensions òptimes de W_{n2} en funció de W_{n1}, cal derivar l'expressió anterior respecte W_{n2}:

$$\frac{dt_p}{dW_{n2}} = 0 \rightarrow \frac{a_1}{W_{n1}} - \frac{a_2\cdot C_L}{W_{n2}^2} = 0 \rightarrow W_{n2} = \sqrt{\frac{a_2 C_L}{a_1}W_{n1}}$$

A la Fig. 5 pot veure's una representació gràfica del temps de propagació en funció de W_{n2}. Pot apreciar-se la mida òptima d'aquest.

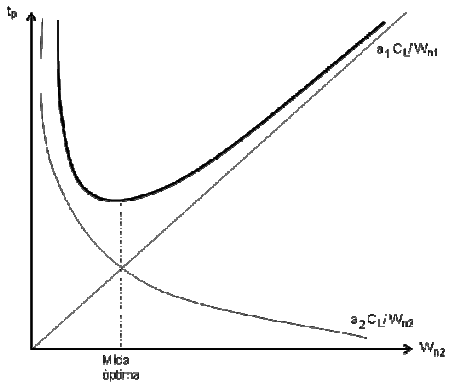


Fig. 5 Representació gràfica del temps de propagació en funció de W_{n2}.

8. Es vol dissenyar una biblioteca de cel·les estàndard. En càlculs previs s'ha comprovat que la cel·la més lenta és la AOI222. Calculeu quina alçada mínima haurà de tenir la biblioteca per aconseguir que el temps de pujada d'aquesta cel·la sigui de 1ns quan la seva capacitat de càrrega externa sigui de 0.3pF.
Dades: μ_n=690cm^2/V·A, μ_p=231cm^2/V·A, C_{ox}=1.37·10^{-3}F/m^2, V_{Tn}=0.82V, $|V_{Tp}|$=1.4V.

Per calcular l'alçada mínima de la biblioteca, cal conèixer l'amplada dels transistors pMOS i nMOS que integren la cel·la AOI222 en funció dels paràmetres especificats. Per calcular W_p i W_n s'utilitzen les expressions pels temps de pujada i baixada, respectivament, que depenen de β_p^{eq} i β_n^{eq}. Per això, cal estudiar primer quin és el pitjor cas de càrrega i descàrrega d'aquesta cel·la.

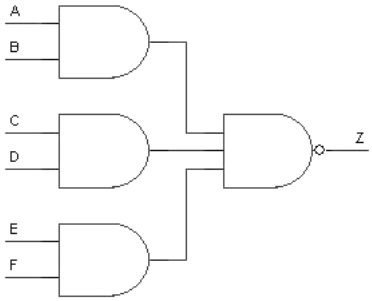


Fig. 6 Disseny de la cel·la AOI222 a nivell de portes lògiques.

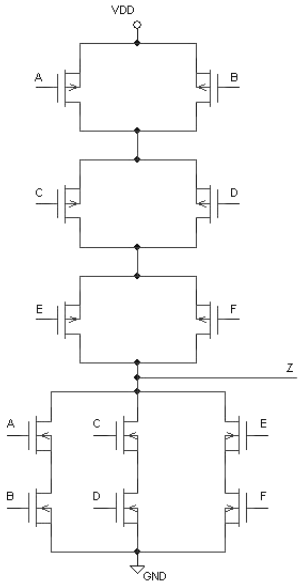


Fig. 7 Disseny de la cel·la AOI222 a nivell de transistors.

El pitjor cas al bloc p correspon a tenir 3 transistors pMOS en sèrie (B, D i F). En canvi, el pitjor cas al bloc n correspon a tenir 2 transistors nMOS en sèrie (A i B). Per tant, es tindrà $W_p^{eq}=W_p/3$ i $W_n^{eq}=W_n/2$. Ara, suposant que $L_n=L_p=1\mu m$:

$$t_f = \frac{C_L}{K_n\left(W_{eq}/L\right)_n\left(V_1 - V_{Tn}\right)}\left[\frac{2V_{Tn}}{\left(V_1 - V_{Tn}\right)} + \ln\left(\frac{2\left(V_1 - V_{Tn}\right)}{V_0} - 1\right)\right] = \frac{C_L}{W_n^{eq}\cdot K_n}\left(0.83V^{-1}\right)$$

$$t_r = \frac{C_L}{K_p\left(W_{eq}/L\right)_p\left(V_1 - \left|V_{Tp}\right|\right)}\left[\frac{2\left|V_{Tp}\right|}{\left(V_1 - \left|V_{Tp}\right|\right)} + \ln\left(\frac{2\left(V_1 - \left|V_{Tp}\right|\right)}{V_0} - 1\right)\right] = \frac{C_L}{W_p^{eq}\cdot K_p}\left(1.07V^{-1}\right)$$

Prenent $t_f=t_r=1$ns i substituint a les expressions anteriors, s'obté:

$$W_p = 30.43\mu m$$
$$W_n = 5.27\mu m$$

Tenint en compte que cal deixar una separació de 6μm entre el bloc p i el bloc n, l'alçada aproximada de la cel·la serà:

$$W = (30.43 + 5.27 + 6)\mu m = 41.70\mu m$$

9. Es considera el circuit dinàmic de la Fig. 8 i es demana:

(i) Quina funció implementa el circuit anterior?

(ii) Si l'entrada A commuta a un estat alt durant la fase d'avaluació, mentre B i C es troben a un estat baix, quins són els voltatges finals als nodes X i Y? Assumiu V_x=0 durant la precàrrega.

(iii) Repetiu l'apartat anterior en el cas en què C_x=10fF.

(iv) Si A, B i C són variables no correlacionades i tenen la mateixa probabilitat de commutar a 0 o a 1, calculeu la probabilitat amb què Y realitzarà una transició de 0 a 1.

(v) Si les entrades arriben a una freqüència de 10MHz, calculeu la dissipació de potència del circuit (assumiu C_x=0).

Dades: V_{Tn}=1V, V_{Tp}=-1V, C_x=25fF, C_y=50fF.

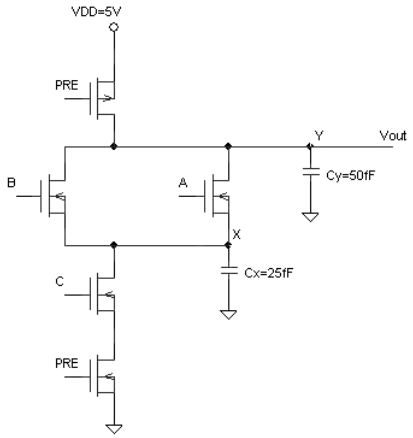


Fig. 8 Disseny a nivell de transistors d'una porta dinàmica.

(i) El circuit dinàmic de la Fig. 8 és un *tristate* que deixa passar o no la funció lògica Y. El circuit funciona de tal manera que a la fase de precàrrega, la sortida sempre és 1. Durant la fase d'avaluació, a la sortida hi haurà el valor lògic que doni la porta. Per determinar quina funció implementa, cal obtenir la seva taula de la veritat.

A	B	C	Y
0	0	0	1
0	0	1	1
0	1	0	1
0	1	1	0
1	0	0	1
1	0	1	0
1	1	0	1
1	1	1	0

Taula 1 Taula de la veritat de l'esquema implementat a la Fig. 8.

D'aquesta taula s'extreu que la funció que implementa el circuit dinàmic de la Fig. 8 és:

$$Y = \overline{\overline{a}\overline{b}} + \overline{c}$$

(ii) Durant la fase de precàrrega, C_y s'haurà carregat a V_{DD}. Si durant la fase d'avaluació A=1, B=0 i C=0, només el transistor A conduirà, i per tant la càrrega emmagatzemada en C_y es distribuirà entre C_x i C_y. Se suposa que a l'instant t=0, V_x=0 i V_y= V_{DD}. Es distingeixen dues situacions:

Si $\Delta V_y < V_{Tn}$:

$$C_y \cdot V_{DD} = C_y \cdot V_y(t) + C_x(V_{DD} - V_{Tn})$$
$$\Delta V_y(t) = V_{DD} - V_y(t) = (C_x / C_y)(V_{DD} - V_{Tn}) = 2V$$

Si $\Delta V_y \geq V_{Tn}$:

$$C_y \cdot V_{DD} = C_y \cdot V_y(t) + C_x \cdot V_y(t) = (C_x + C_y) V_y(t)$$
$$V_y(t) = \frac{C_y}{C_x + C_y} V_{DD} = 3.33V$$
$$\Delta V_y(t) = V_{DD} - V_y(t) = (5 - 3.33)V = 1.67V$$

Com que $\Delta V_y \geq V_{Tn}$, la situació que és vàlida és la segona. Per determinar $V_x(t)$:

$$V_x(t) = V_y(t) - V_{Tn} = 2.33V$$

(iii) Si C_x=10fF, només cal refer l'apartat anterior canviant el valor numèric d'aquesta capacitat. Donat que ara $\Delta V_y < V_{Tn}$, el mètode que ha d'utilitzar-se per fer el càlcul és el primer:

$$\Delta V_y(t) = (C_x / C_y)(V_{DD} - V_{Tn}) = 0.8V$$
$$V_y(t) = V_{DD} - \Delta V_y(t) = 4.2V$$
$$V_x(t) = V_y(t) - V_{Tn} = 3.2V$$

(iv) La probabilitat de passar de 0 a 1 ve donada per l'expressió:

$$P_{0\to1} = P_0 \cdot P_1 = (1 - P_1) P_1$$

Segons la taula de la veritat, de 8 transicions possibles, 3 donen com a resultat un 0 lògic i 5, un 1 lògic. Per tant, P_0=3/8 i P_1=5/8. Així doncs, la probabilitat que a la sortida hi hagi una transició de 0 a 1 serà:

$$P_{0\to1} = P_0 \cdot P_1 = (3/8)(5/8) = 15/64$$

(v) En una porta dinàmica, la potència dinàmica consumida ve donada per la probabilitat de les transicions de 0 a 1.

$$P_d = C_y \cdot V_{DD}^2 \cdot P_{0\to1} \cdot f = 0.293W$$