

1. Tendințe în evoluția semiconductoarelor și scalarea tehnologică

"What we didn't realize then was that the integrated circuit would reduce the cost of electronic functions by a factor of a million to one, nothing had ever done that for anything before", Jack Kilby

Circuitele integrate au apărut în anul 1959, fiind inventate simultan și independent de doi ingineri, Jack Kilby de la Texas Instruments și Robert Noyce de la Fairchild Semiconductor. Activitatea lor de cercetare a fost motivată de necesitatea de a rezolva problemele stringente legate de cerințele de complexitate tot mai accentuate ale circuitelor discrete, implementate la aceea oră cu tuburi, tranzistoare, rezistoare și condensatoare. Circuitul integrat s-a născut din ideea de a poziționa variante de dimensiuni reduse ale componentelor discrete pe același substrat semiconductor. Semiconductorul utilizat de Kilby a fost germaniul (Ge), iar Noyce a folosit siliciul.

Primul circuit integrat produs la scară industrială a apărut în anul 1961, fiind fabricată de compania Fairchild Semiconductor. Acesta conținea un singur tranzistor, trei rezistoare și un condensator, împreună cu traseele metalice de interconectare a acestora.

În următorii ani circuitele integrate au cunoscut o dezvoltare spectaculoasă urmînd a fi folosite la scară largă în aplicațiile de electronică, înlocuind tot mai frecvent componentele discrete. Revoluția adevărată în domeniul semiconductoarelor integrate s-a produs însă în anul 1967, când același Jack Kilby a inventat calculatorul portabil. Această invenție, care folosea circuite integrate ca și elemente de calcul, a fost urmată în anul 1968 de inventarea microprocesorului de inginerii companiei Intel, proaspăt fondate de Robert Noyce și Gordon Moore.

Gordon Moore este responsabil pentru ceea ce se numește legea lui Moore, o lege enunțată pe cale empirică, care guvernează sub o formă sau alta industria semiconductoarelor. Legea lui Moore a fost anunțată într-un articol în anul 1965, fiind valabilă cu ajustările de rigoare și în zilele noastre. Moore a analizat creșterea complexității circuitelor integrate în funcție de numărul de tranzistoare realizate pe același substrat semiconductor și a formulat o tendință care arată dublarea densității de integrare la fiecare interval de doi ani. Mai mult, legea lui Moore prezice continuarea acestei tendințe pentru un timp îndelungat.

În zilele noastre legea lui Moore a fost extinsă, luându-se în calcul, pe lângă densitatea de integrare și frecvența de funcționare și factorii care pot influența performanțele unui circuit integrat la scară foarte largă. De exemplu, între anii 2000 și 2009 frecvența semnalului de ceas al microprocesoarelor s-a dublat, crescând în medie de la 1,3GHz la 2.8GHz. Totodată, numărul de tranzistoare a crescut de la 37,5 milioane la 904 milioane. Chiar dacă nu respectă riguros dublarea numărului de componente integrate sau a frecvenței de funcționare, cei doi factori au dus la dublarea aproximativă a puterii de calcul, respectând în acest sens forma extinsă a legii lui Moore. La factorii frecvență și densitate de integrare se adaugă în zilele noastre și numărul unităților de procesare paralelă.

Conform predicției din industria semiconductoarelor, legea lui Moore va atinge o saturație atunci când scalarea tehnologică va aduce dispozitivele la un nivel de miniaturizare la care geometriile sunt comparabile cu dimensiunile atomice. În aceste condiții scalarea continuă pe calea clasică nu va mai fi posibilă. În practică limitările impuse de miniaturizarea extremă și scalarea tehnologiilor de fabricație se manifestă tot mai preponderent la procesele cu dimensiuni mai mici decât 0,18-0,13 μ m, punând numeroase probleme și necesitând soluții tot mai ingenioase pentru a păstra funcționalitatea și performanțele circuitelor.

Următoarea secțiune a acestui material descrie câteva dintre problemele tipice rezultate în urma scalării proceselor de fabricație și limitările introduse de acestea.

1.1. Scalarea tehnologiilor de fabricație CMOS

Miniaturizarea extremă și posibilitatea de scalare sunt proprietăți deosebit de importante ale proceselor moderne de fabricație CMOS, impuse de cerințele tot mai accentuate pentru putere de calcul, și reflectate direct în creșterea frecvențelor de tact și a complexității circuitelor.

În ultimii ani evoluția științei materialelor și a tehnologiilor de litografie a permis îmbunătățirea dramatică a performanțelor circuitelor integrate. Posibilitatea de a implementa circuite cu ajutorul proceselor cu geometria minimă din gama zecilor de nm duc la realizarea unor circuite monolitice având un număr impresionant de componente integrate. Rezultatele directe ale miniaturizării sunt creșterea numărului de compo-

mente și implicit a funcționalității circuitelor, alături de mărirea vitezelor de comutație. Adițional se obține reducerea drastică a puterii consumate, raportate la același număr de dispozitive. La baza acestor îmbunătățiri stă procesul de scalare tehnologică a tranzistoarelor MOS.

Scăderea geometriei duce la scăderea capacităților parazite asociate cu dispozitivele reale. Desemenea, scăderea lungimii canalului și a grosimii oxidului de sub poartă (t_{ox}) cresc capabilitățile în curent ale tranzistoarelor, având ca efect creșterea frecvențelor de tact asociate cu circuitele digitale.

Scopul principal al scalării proceselor de fabricație CMOS este creșterea densităților de curent raportate la unitatea de lățime a unui tranzistor. Dacă se consideră că prin scalarea tehnologiei valorile capacităților parazite rămân aproximativ aceleași, atunci creșterea densităților de curent duce la un ciclu mai rapid de încărcare-descărcare a capacităților de sarcină văzute de circuitele digitale și implicit permit creșterea frecvenței de funcționare.

Metodele de scalare care permit mărirea densității de curent rezultă din examinarea atentă a ecuației specifice care definește curentul prin tranzistor în funcție de tensiunile de polarizare aplicate între terminale. La tehnologiile cu geometrie redusă aproximarea pătratică clasică nu mai funcționează datorită efectelor introduse de scurtarea canalului de conducție. Corecția necesară a expresiei curentului trebuie să țină cont și de câmpul electric longitudinal indus de potențialul drenei și implicit de saturația vitezei de deplasare a purtătorilor atunci când valoarea câmpului electric depășește un prag critic. Curentul prin tranzistor, definit de modelul cu canal scurt, este dat în următoarea ecuație:

$$I_D = \frac{\mu C_{ox} W}{L} \cdot \frac{\left(V_{GS} - V_{Th} - \frac{V_{DS}}{2} \right) V_{DS}}{1 + \frac{\mu V_{DS}}{L v_s}}, \quad (1.1)$$

unde $C_{ox} = \epsilon_{ox}/t_{ox}$ este capacitatea specifică a oxidului de sub grilă, definită ca și raportul dintre permitivitatea electrică și grosimea oxidului, μ este mobilitatea purtătorilor de sarcină, V_{Th} este tensiunea de prag, iar v_s este viteza de saturație a purtătorilor.

Condiția de saturație la limită a tranzistorului se obține în continuare calculând punctul de inflexiune al funcției $I_D = f(V_{DS})$. Derivarea curentului în raport cu tensiunea drenă-sursă duce la următoarea relație:

$$\frac{\partial I_D}{\partial V_{DS}} = \frac{\mu C_{ox} W}{L} \cdot \frac{(V_{GS} - V_{Th}) - V_{DS} - \frac{\mu V_{DS}^2}{L v_s}}{\left(1 + \frac{\mu V_{DS}}{L v_s} \right)^2} \quad (1.2)$$

Pentru a calcula punctul de inflexiune și totodată tensiunea de saturație a tranzistorului conform modelului cu canal scurt, se egalează derivata curentului cu zero, rădăcina V_{DS} a ecuației reprezentând tensiunea de saturație V_{DSat} . Din relația (1.2) se observă că ecuația va fi de gradul doi și va avea forma:

$$V_{DS}^2 + \frac{L v_s}{\mu} \cdot V_{DS} - (V_{GS} - V_{Th}) = 0 \quad (1.3)$$

Luând în considerare numai soluția pozitivă a acestei ecuații rezultă:

$$V_{DSat} = \frac{L v_s}{\mu} \cdot \left[\sqrt{1 + \frac{2\mu(V_{GS} - V_{Th})}{L v_s}} - 1 \right] = \frac{E_{Sat} L \cdot (V_{GS} - V_{Th})}{E_{Sat} L + (V_{GS} - V_{Th})} \quad (1.4)$$

În ecuația de mai sus $E_{sat} = v_s/\mu$ este câmpul longitudinal critic la care intervine saturația vitezei purtătorilor de sarcină (eng. *velocity saturation*). Acest câmp electric este direct proporțional cu tensiunea drenă-sursă și scade cu micșorarea geometriei minime a tranzistoarelor. Datorită saturației vitezei de deplasare a purtătorilor sub influența câmpului longitudinal critic, expresia curentului în regim saturat diferă de ecuația pătratică specifică modelului clasic care considera canalul de conducție mult mai lung decât geometria minimă permisă de tehnologie. Curentul se scrie

$$I_D = W v_s C_{ox} \cdot (V_{GS} - V_{Th} - V_{DSat}) \quad (1.5)$$

Prin examinarea atentă a ecuațiilor de mai sus se poate ajunge la concluzia că mărimile care pot fi implicate într-un proces de scalare tehnologică sunt lungimea canalului (L) și grosimea oxidului de sub terminalul de poartă (t_{ox}). În mod tipic, curentul crește cu reducerea lungimii canalului pentru aceleași condiții de polarizare. Totodată, se observă că beneficiile aduse de reducerea lungimii canalului sunt limitate de saturația vitezei de deplasare a purtătorilor care împiedică creșterea în continuare a curentului. Încercarea de a scala geometria tranzistoarelor și de a combate în același timp efectele nedorite, introduse de scurtarea canalului, duc la o abordare mai complicată a întregului proces în care o mărime, fie câmpul electric, fie tensiunea este menținută constantă, iar scalarea include și alte mărimi caracteristice procesului de fabricație, cum ar fi de exemplu adâncimea de dopare a joncțiunilor (X_j), adâncimea regiunii de golire de sub canal (W_d) și concentrația dopantului în substrat (N_A). În caz contrar, tranzistorul își poate pierde calitatea de sursă de curent comandată de tensiunea aplicată pe terminalul de poartă. Un exemplu tipic în acest sens este reducerea barierei de potențial a joncțiunii substrat-sursă (eng. *DIBL* sau *Drain Induced Barrier Lowering*) datorită lățimii excesive ale regiunii de golire din jurul zonei de difuzie a drenei. Reducerea barierei de potențial din jurul difuziei sursă duce la injecția accentuată a purtătorilor de sarcină în canal și la pierderi de curent (eng. *leakage*) sau chiar străpungerea regiunii dintre drenă și sursă (eng. *punch-through*).

Câteva modele de scalare din literatură definesc lungimea minimă a canalului care se poate obține pentru o configurație dată a parametrilor t_{ox} , X_j și W_d . Două dintre exemple sunt date după cum urmează:

$$\begin{cases} L_{\min} \propto t_{ox} \cdot X_j^{1/3} \\ L_{\min} \propto t_{ox}^{1/3} \cdot X_j^{1/3} \cdot W_d^{2/3} \end{cases} \quad (1.6)$$

În industrie se aplică două tipuri de scalare: scalarea în câmp constant și scalarea în tensiune constantă. În ambele cazuri scopul este menținerea unui parametru constant (câmpul electric sau tensiunea de alimentare) în timp ce geometria minimă a procesului scade, iar densitatea de curent raportată la unitatea de lățime crește.

- **Scalarea în câmp constant**

Scalarea în câmp constant implică reducerea lungimii minime a canalului de conducție, L_{\min} , prin scalarea cu un factor s subunitar. Celelalte mărimi scalate cu același factor sunt lățimea minimă admisă a tranzistorului, grosimea oxidului, concentrația de dopant în substrat și în regiunile de difuzie ale drenei/sursei și tensiunea de alimentare. Această schemă de scalare a caracteristicilor specifice procesului de fabricație permite păstrarea câmpului electric din tranzistor la valorile obținute înainte de scalare. Rezultatul net al scalării este reducerea puterii dinamice consumate cu un factor s^2 , a întârzierilor din circuit cu un factor s și creșterea frecvențelor de comutație cu un factor $1/s$. Această schemă de scalare presupune că atât câmpul transversal cât și cel longitudinal se mențin la valorile inițiale.

- **Scalarea în tensiune constantă**

Scalarea în tensiune constantă presupune menținerea tensiunilor din circuit la același nivel, asigurând astfel compatibilitatea circuitelor cu procesele mai vechi de fabricație. Dezavantajul major al scalării în tensiune constantă este creșterea semnificativă a intensității câmpului electric odată cu scăderea lungimii minime a canalului de conducție. Acest fapt conduce la efectele tipice asociate cu scurtarea canalului, cum ar fi de exemplu saturarea vitezei purtătorilor, degradarea mobilității, creșterea pierderilor de curent și scăderea tensiunilor critice de străpungere ale materialelor din componența dispozitivelor.

Efectele scalării asupra puterii consumate, întârzierilor din circuit și frecvențelor de comutație pot fi sintetizate în Tabelul 1.1. Din punctul de vedere al performanței în frecvență este de dorit ca tensiunile aplicate unui tranzistor să rămână cât mai mari pentru a păstra valori mari ale densității curentului. Totodată, cerințele de fiabilitate ale circuitelor împreună cu tendințele spre implementarea unor sisteme portabile, alimentate de la o singură baterie, impun reducerea tensiunilor de alimentare și a puterii consumate.

1.2. Scalarea tehnologiilor în tensiune

Calea cea mai eficientă pentru a reduce consumul de putere, impusă totodată și de câmpurile electrice tot mai reduse suportate de componentele extrem miniaturizate, este scăderea tensiunii de alimentare. Acest

fapt se datorează în mare măsură dependenței consumului total de putere (P_{tot}) al circuitelor CMOS digitale de tensiunea de alimentare. Dependența este dată de ecuația (1.7).

Parametru	Simbol	Scalare în câmp constant	Scalare în tensiune constantă	Scalare în tensiune constantă cu saturarea vitezei
lungime	L	s	s	s
lățime	W	s	s	s
câmpul electric	E	1	$1/s$	$1/s$
grosimea oxidului	t_{ox}	s	s	s
dopantul în substrat	N_A	$1/s^2$	$1/s^2$	$1/s^2$
capacitatea grilei	C_G	s	s	s
capacitatea oxidului	C_{ox}	$1/s$	$1/s$	$1/s$
timpul de tranzit	t_r	s^2	s^2	s^2
frecvența de tranzit	f_T	$1/s$	$1/s^2$	$1/s^2$
tensiunea de alimentare	V_{DD}	s	1	1
curentul de drenă	I_D	s	$1/s$	1
puterea dinamică	P_{din}	s^2	$1/s$	1

Tabelul 1.1. Scalarea parametrilor specifici procesului de fabricație și efectul asupra performanțelor la circuitele digitale

$$P_{tot} = P_{din} + P_{pierderi} = A \cdot f \cdot C_L \cdot V_{DD}^2 + I_{pierderi} \cdot V_{DD}, \quad (1.7)$$

unde f este frecvența de comutație, C_L este capacitatea de sarcină văzută de un circuit digital, A este coeficientul de activitate, iar $I_{pierderi}$ este curentul static de pierderi datorat efectelor de canal scurt (eng. *leakage*).

Curentul de pierderi are trei componente majore cunoscute, dependente de mecanismul care le induce. Cele trei componente apar în urma efectului de tunelare cuantică a purtătorilor liberi prin barierele de potențial din dispozitiv.

- *tunelarea prin oxid* – aceste pierderi sunt cele mai importante fiind totodată și cele mai studiate. Cauza este tunelarea de tip Fowler-Nordheim (la t_{ox} nu foarte reduse) sau directă (la t_{ox} mici) a purtătorilor prin oxidul subțire de sub poarta tranzistoarelor. Densitatea de curent depinde în mod tipic de tensiunea pozitivă grilă-sursă conform unei legi pătratice;
- *tunelarea prin substrat* – acest mecanism implică trecerea purtătorilor prin efect tunel de tip Zener prin barierele de potențial ale joncțiunilor sursă-substrat și drenă-substrat;
- *tunelarea directă sursă-drenă* – este efectul direct al reducerii barierei de potențial al joncțiunii sursă-substrat de către potențialul drenei și de extinderea regiunii sărăcite în sarcini din jurul drenei (DIBL).

Istoria circuitelor integrate este marcată de o continuă scădere a tensiunilor de alimentare. Pentru o lungă perioadă de timp tensiunea de alimentare standard a fost 5V, datorită în mare măsură dominației familiei TTL. Migrația din anii 1990 spre procese CMOS ieftine și tensiunile de alimentare de 3,3V asociate au declanșat o tendință de scădere a tensiunii de alimentare la fiecare trecere spre un proces de fabricație mai avansat, cu geometrie mai redusă. Tabelul 1.2 arată evoluția proceselor de fabricație într-o formă compactă.

L_{min} [μm]	V_{DD} [V]	t_{ox} [nm]	V_{Th} [mV]
1	5	250	950
0,8	5	200	850
0,5	3,3	135	730
0,35	3,3	100	590
0,25	2,5	60	520
0,18	1,8	50	420
0,12	1,2	42	320
0,10	1,2	36	310
0,07	0,9	30	300

Tabelul 1.2. Evoluția proceselor de fabricație în ceea ce privește geometria minimă, tensiunea de alimentare și tensiunile de prag asociate

Pentru a menține caracteristicile de comutație și marginea de zgomot a circuitelor digitale, este necesar ca tensiunea de alimentare să fie de câteva ori mai mare decât tensiunile de prag ale tranzistoarelor. Această cerință se păstrează și atunci când tehnologia trece la o nouă generație prin scalare. Din Tabelul 1.2 se observă că tensiunea de prag este scalată cu aproximativ același factor ca și tensiunea de alimentare. De exemplu, trecerea de la procesul cu geometria minimă de $1\mu\text{m}$ la procesul de $0,25\mu\text{m}$ implică înjumătățirea tensiunii de alimentare. Totodată, tensiunea de prag este scalată cu același factor aproximativ de 0,5.

O observație deosebit de importantă este aceea că scalarea tensiunii de prag poate fi făcută necondiționat numai la dimensiuni minime ale tranzistoarelor mai mari decât $0,25\mu\text{m}$. Sub această dimensiune funcționarea tranzistoarelor nu se mai conformează modelului convențional cu canal lung. Mai mult, scăderea tensiunii de prag poate conduce la probleme grave care se manifestă prin pierderi de curent static și imposibilitatea de a bloca complet un tranzistor.

Pentru a vedea motivul aplatizării curbei de scalare corespunzătoare tensiunii de prag la tehnologii cu geometrie redusă, trebuie examinată dependența curentului static de pierderi în regim subprag de tensiunea V_{Th} . La modelul cu canal lung curentul poate fi aproximat cu zero dacă tensiunea grilă-sursă este mai mică decât tensiunea de prag. La modelul cu canal scurt această afirmație nu mai este adevărată, curentul în regim subprag depinzând exponențial de tensiunea grilă-sursă. Ecuația specifică a curentului este:

$$I_{off} \propto e^{n \frac{V_{GS}}{V_T}}, \quad (1.8)$$

unde V_T este tensiunea termică, iar n este factorul de de pantă în regim subprag (eng. *subthreshold slope factor*). Pentru definiția factorului n se consideră secțiunea unui tranzistor din Figura 1.1 și capacitățile specifice, definite de polisiliciul grilei, izolatorul de sub grilă, canalul de conducție și regiunea sărăcită în sarcini dintre canal și substrat. Capacitatea $C_{ox} = \epsilon_{ox}/t_{ox}$ este capacitatea specifică a oxidului, iar $C_B = \epsilon_{Si}/W_d$ este capacitatea specifică dintre canal și substrat.

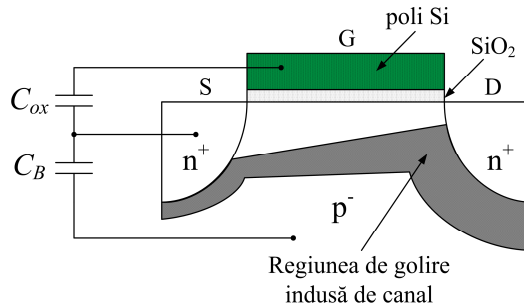


Figura 1.1. Secțiunea unui tranzistor NMOS și capacitățile specifice ale oxidului și substratului

Factorul n se definește în funcție de raportul celor două capacități

$$n = 1 + \frac{C_B}{C_{ox}} \quad (1.9)$$

Expresia curentului în regim subprag poate fi reprezentată grafic ca și o exponențială dacă axa Ox este liniară, sau ca și o dreaptă dacă axa Ox este logaritmică (Figura 1.2). Dacă se consideră reprezentarea curentului pe scară logaritmică și se ține cont de dependența curentului de pierderi de V_{GS} din ecuația (1.8), panta drepte se calculează

$$\frac{1}{S} = \frac{\partial(\lg I_{off})}{\partial V_{GS}} = \frac{\partial}{\partial V_{GS}} \left(\lg e^{n \frac{V_{GS}}{V_T}} \right) = \frac{\partial}{\partial V_{GS}} \left(\frac{\ln e^{n \frac{V_{GS}}{V_T}}}{\ln 10} \right) = \frac{\partial}{\partial V_{GS}} \left(\frac{V_{GS}}{n V_T \ln 10} \right) = \frac{1}{n V_T \ln 10} \quad (1.10)$$

Excursia logaritmică a curentului de pierderi ca efect al variației tensiunii grilă-sursă este definit de inversa S a pantei conform ecuației (1.11).

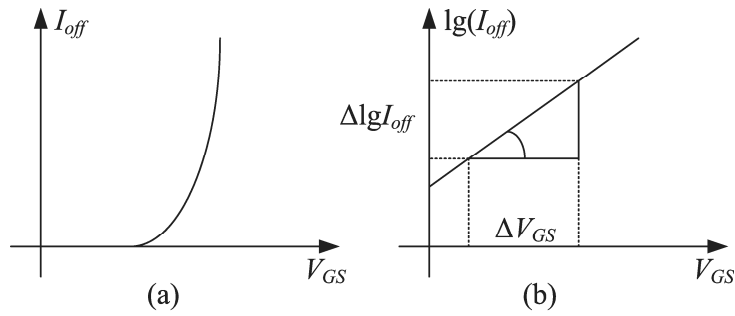


Figura 1.2. Curentul I_{off} în funcție de tensiunea grilă-sursă: scară liniară (a) și scară logaritmică (b)

$$S = nV_T \ln 10 = n \cdot \frac{kT}{q} \cdot \ln 10 \quad (1.11)$$

Din ecuația de mai sus se observă că, dacă se impune temperatura de funcționare a tranzistorului la o anumită valoare, iar factorul n se consideră egal cu valoarea minimă posibilă (unitară), atunci se poate obține un minim pentru panta inversă S în regim subprag. La temperatura camerei (300K) această valoare minimă este aproximativ 60mV/decadă. Ținând cont de efectul substratului asupra sarcinilor transferate prin tranzistor, factorul n devine cu siguranță supraunitar, iar S crește.

Știind panta inversă în regim subprag, poate fi definită legătura directă dintre curentul de pierderi prin tranzistor și tensiunea de prag:

$$I_{off} \cong I_{V_{Th}} \cdot 10^{-\frac{V_{Th}}{S}} = I_{V_{Th}} \cdot 10^{-\frac{qV_{Th}}{nkT \ln 10}} \quad (1.12)$$

În ecuația de mai sus $I_{V_{Th}}$ este curentul nominal la care se definește tensiunea de prag. Din această ecuație se observă că scăderea excesivă a tensiunii de prag duce la creșterea exponențială a curentului static de pierderi în regim subprag respectiv blocat. O soluție care ar permite scăderea tensiunii V_{Th} fără a mări prea mult pierderile de curent ar fi răcirea extremă a circuitelor. Această soluție implică costuri de implementare foarte mari și nu este practică la circuitele utilizate în sistemele comerciale actuale. În aplicațiile de joasă și medie putere valorile tipice pentru parametrii din ecuația (1.12) sunt:

$$\begin{cases} S \cong 90mV / dec \\ I_{V_{Th}} \cong 100mA / cm \\ I_{off} \cong 100nA / cm \end{cases} \quad (1.13)$$

Valoarea corespunzătoare a tensiunii de prag, care la temperatura camerei permite obținerea parametrilor de mai sus, este aproximativ 360mV. Concluzia la această analiză este că pentru a menține controlul asupra pierderilor statice de curent în regim blocat este necesar ca tensiunea de prag să se mențină în gama (300 ÷ 400)mV. Mai mult, aceste valori implică cel mai defavorabil caz. În realitate însă trebuie ținut cont și de toleranțele asociate procesului de fabricație care pot înrăutăți curentul de pierderi și duc implicit la un consum accentuat al circuitelor în regim static. Drept urmare, curenții de pierderi impun limita fundamentală de scalare a tensiunii de prag la procesele CMOS construite pe un substrat semiconductor slab dopat.

Domeniul de scalare a tensiunii de prag poate fi extins dacă panta inversă în regim subprag S este scăzută. Acest lucru este posibil în tehnologii mai costisitoare, cum ar fi de exemplu procesele de fabricație la care substratul semiconductor slab dopat este înlocuit cu un izolator (eng. *SOI* sau *Silicon On Insulator*). În aceste procese contribuția substratului la divizorul capacitiv $C_{ox}-C_B$ este eliminată, permițând astfel obținerea unui factor S egal cu valoarea minimă posibilă la o temperatură dată.

Problema tensiunii de prag scalate diferit față de tensiunea de alimentare afectează circuitele digitale din punctul de vedere al pierderilor statice, al excursiei semnalului și implicit al marginii de zgomot. Aceste probleme pot fi parțial sau complet rezolvate cu ajutorul unor tehnici de proiectare orientate spre ajustarea dinamică a tensiunii de prag sau prin utilizarea unor procese de fabricație care să suporte tranzistoare MOS cu multiple tensiuni de prag.

1.3. Soluții pentru menținerea funcționalității tranzistoarelor în domeniul digital

Condiția esențială pentru funcționarea corectă a unui circuit digital din punctul de vedere al marginii de zgomot și al integrității semnalelor este ca tensiunea de alimentare să fie de câteva ori mai mare decât tensiunea de prag. Datorită faptului că tensiunea de alimentare se scalează odată cu geometria minimă a procesului de fabricație, singura soluție posibilă rămâne reducerea temporară sau permanentă tensiunii de prag. Această scalare trebuie făcută astfel încât curentul static de pierdere în tranzistoarele blocate să nu crească semnificativ. În industrie s-au raportat câteva soluții posibile la această problemă. Printre aceste soluții pot fi amintite utilizarea unui substrat izolator în locul celui semiconductor slab dopat, implementarea unor tranzistoare cu grilă flotantă, polarizarea dinamică a terminalului de substrat și extinderea tehnologiei CMOS clasice pentru tensiuni multiple de prag.

Prima opțiune, utilizarea unei tehnologii SOI, nu este în totdeauna convenabilă datorită costului mai ridicat. În mod tipic, aplicațiile comerciale de joasă și medie putere destinate publicului larg, cum ar fi terminalele de comunicații, sistemele de navigație sau echipamentele electronice de consum, nu justifică în totdeauna utilizarea unui proces bazat pe un substrat izolator. Celelalte tehnici pot fi considerate ca și extensii ale tehnologiilor CMOS convenționale, fiind aplicabile fără a crește semnificativ costurile de producție.

1.3.1. Tranzistoare cu poartă flotantă (Floating Gate-MOS sau FG MOS)

Pentru a reduce problemele legate de tensiunea de prag, V_{Th} ideală ar trebui să tindă la zero în condițiile în care curentul de pierdere nu crește semnificativ. Această cerință ar impune introducerea unor tranzistoare cu canal inițial, dispozitive care nu apar de regulă în procesele CMOS ieftine deoarece necesită măști suplimentare și pași adiționali de fabricație, ridicând implicit costurile asociate procesului.

Prin implementarea tranzistoarelor cu grilă flotantă se dorește imitarea funcționării unui tranzistor cu canal inițial adăugând o regiune flotantă de polisiliciu între poarta de control și canalul de conducție. Structura convențională a unui tranzistor MOS nu trebuie modificată, dispozitivul fiind astfel compatibil cu procesele CMOS de cost redus, similar cu tehnologiile de implementare ale memoriilor nevolatile EEPROM. Structura principală a unui tranzistor NMOS cu poartă flotantă este dată în Figura 1.3

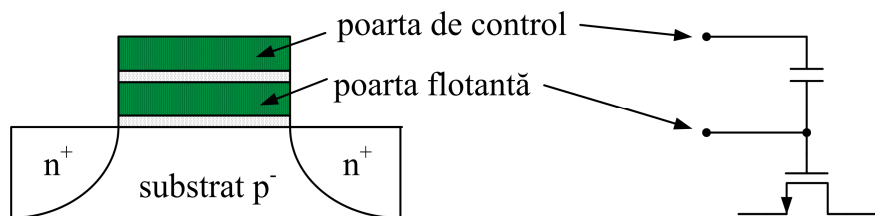


Figura 1.3. Tranzistorul cu poartă flotantă – structura principală și schema echivalentă

Din structura tranzistorului cu poartă flotantă se observă că poarta de control, a cărei funcționalitate este aceeași ca și la tranzistoarele simple, este cuplată capacitiv la poarta flotantă, transferul de sarcini făcându-se prin efect de tunelare de tipul Fowler-Nordheim prin stratul subțire de oxid. Sarcina pozitivă acumulată pe poarta flotantă produce apariția unei regiuni de golire sau chiar a unui canal de conducție inițial dacă diferența de potențial este suficient de mare. Efectul acumulării de sarcini este prepolarizarea tranzistorului și deplasarea valorii efective a tensiunii de prag spre valori mai scăzute. Această deplasare nu implică modificarea parametrilor de proces, drept urmare nu influențează semnificativ nici curentul de pierdere în regim blocat.

Utilizarea porții flotante permite reglajul precis al tensiunii de prag efective. În funcție de modul de stocare a sarcinilor, potențialul porții flotante poate fi reglat fie prin aplicarea efectivă a unei tensiuni, fie prin iluminare cu lungimi de undă din gama ultra-violetă.

Tehnica tranzistoarelor cu poartă flotantă nu este foarte nouă. Ea a apărut la sfârșitul anilor '90, odată cu identificarea problemelor legate de scalarea proceselor tehnologice suportând geometrii mai mici decât $0,35\mu\text{m}$. În ciuda vechimii, avantajele circuitelor care utilizează tranzistoare cu poartă flotantă se mențin și pentru cele mai avansate tehnologii actuale. Aceste avantaje pot fi enumerate succint după cum urmează:

- *tensiunea de prag* – poate fi programată la valoarea dorită, simultan pentru toate tranzistoarele de pe același circuit integrat sau selectiv pentru grupuri de tranzistoare. Programarea nu necesită semnale sau circuite adiționale, redundante;
- *tensiunea de alimentare* – datorită scalării tensiunii de prag devine posibilă reducerea extremă a tensiunilor de alimentare, până în gama de valori sub 1V;
- *adaptivitatea* – tensiunile, de prag sau de alimentare, pot fi modificate din mers, astfel încât devine posibilă implementarea mai eficientă a soluțiilor pentru reducerea dinamică a puterii consumate (eng. *power saving*);
- *simplitate* – la implementarea circuitelor digitale nu mai este nevoie de scalarea geometriei pentru a reduce dezechilibrul mobilităților la dispozitivele NMOS/PMOS și pentru a egala timpii de tranziție. Această scalare este făcută direct la nivelul capacităților de cuplaj la poarta flotantă;
- *aria pe siliciu* – aria utilizată este mai redusă decât la circuitele digitale CMOS convenționale datorită posibilității de a efectua operația ȘI logic (tranzistoare serie în tehnologia CMOS clasică) direct la nivelul capacităților de cuplaj. Tranzistorul transmite direct semnalul rezultat din operația ȘI;
- *neîmperecherea tranzistoarelor* – problemele legate de neîmperechere pot fi reduse prin scalarea adaptivă a potențialului porții flotante și implicit a tensiunii de prag;
- *viteza de comutație* – datorită prepolarizării cu ajutorul porții flotante, devine posibilă implementarea unor circuite de putere foarte joasă fără a sacrifica viteza de comutație.

Figura 1.4 arată două exemple tipice de implementare a unor porți fundamentale cu ajutorul tranzistoarelor având poarta flotantă. Din această figură se poate remarca simplificarea căii de semnal serie care realizează operația ȘI logic.

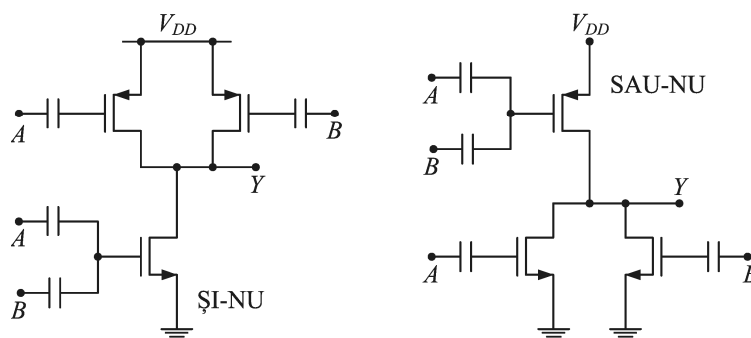


Figura 1.4. Porțile elementare ȘI-NU și SAU-NU cu tranzistoare având poarta flotantă

1.3.2. Tranzistoare MOS cu prag dinamic (Dynamic Threshold MOS)

Acest tip de tranzistoare folosesc terminalul de substrat sau o poartă adițională îngropată pentru a modula potențialul terminalului de substrat și implicit tensiunea de prag. Principiul este utilizat pentru a obține o tensiune de prag mare în stare de blocare și o tensiune de prag redusă dacă tranzistorul este polarizat în regim de conducție (dacă $|V_{GS}| \approx V_{DD}$). Cele două posibilități de a modula dinamic potențialul substratului sunt ilustrate în Figura 1.5.

Avantajul acestei abordări este acela că terminalele grilă și substrat sunt la același potențial și astfel se elimină efectul substratului asupra factorului de pantă n în regim subprag. Acesta fiind unitar, panta inversă S în regim subprag va fi minimă pentru o temperatură dată. Totodată, tensiunea de prag poate fi scăzută până la limita fundamentală impusă de valoarea curentului static de pierderi.

Dezavantajele acestei tehnici de implementare sunt posibilitatea limitată de a aplica modularea dinamică a tensiunii de prag la un singur tip de dispozitive la procesele CMOS convenționale și capacitatea grilă sursă accentuată datorită faptului că substratul contribuie activ la capacitatea totală văzută la terminalul grilă.

O observație importantă este aceea că la curenți și tensiuni mici tranzistoarele cu prag dinamic aproximează bine relația curent-tensiune exponențială a unei joncțiuni pn , astfel încât poate fi utilizat cu succes la implementarea unor blocuri analogice de foarte joasă tensiune care necesită diode bipolare. Exemple tipice în acest sens ar fi referințele de tip bandă interzisă, implementate în procese CMOS în care tranzistoarele bipolare laterale pot fi înlocuite cu succes de tranzistoare MOS cu prag dinamic.

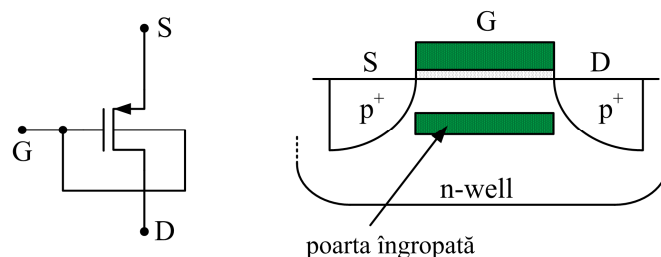


Figura 1.5. Tranzistorul MOS cu prag dinamic

1.3.3. Tehnologii CMOS cu multiple tensiuni de prag

Tranzistoarele cu prag multiplu sunt implementate în tehnologia CMOS convențională, dar în același proces pot fi realizate mai multe tipuri de tranzistoare, fiecare având tensiunea de prag modificată în scopul implementării unor sarcini precise în circuite. Această tehnică este relativ populară deoarece permite compatibilitatea cu pașii tehnologici din procesele CMOS simple. În general, ajustarea tensiunii de prag se face prin modificarea concentrației dopantului și implantarea selectivă a ionilor.

O alternativă la modificarea din fabricație a tensiunilor de prag prin reglajul parametrilor de proces este utilizarea unor circuite speciale care să controleze potențialul substratului. Astfel, dacă tranzistorul este în repaus (regim static) atunci se polarizează invers dioda substrat-sursă, iar tensiunea de prag este mărită semnificativ. Dacă tranzistorul trebuie să comute, circuitul de control reduce tensiunea V_{BS} la zero sau chiar polarizează direct dioda substrat-sursă, astfel încât pe timpul comutației tensiunea de prag este redusă și se asigură mărirea vitezei de comutație.

Principiile de proiectare ale circuitelor cu tensiuni de prag multiple se bazează pe faptul că tranzistoarele cu tensiune de prag mare, poziționate strategic, întrerup căile de scurgere ale curenților statici de pierderi în regim blocat, în timp ce tranzistoarele cu tensiune de prag redusă asigură viteze mari de comutație. Acest principiu este ilustrat în Figura 1.6.

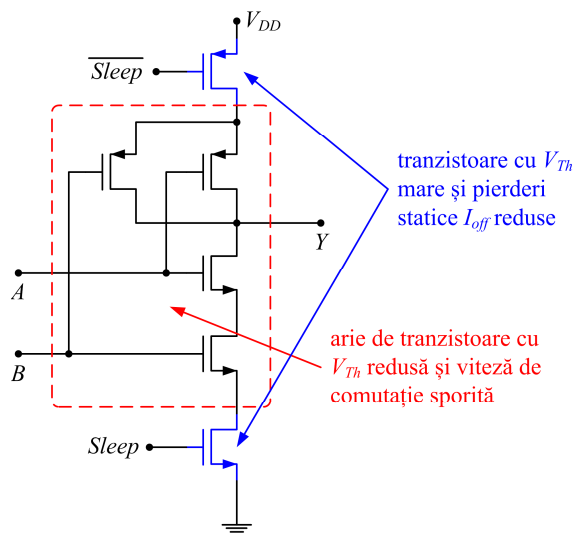


Figura 1.6. Principiul de proiectare a circuitelor digitale CMOS utilizând tensiuni de prag multiple

Tranzistoarele cu tensiunea de prag redusă sunt utilizate pentru a implementa circuitele logice cu specificații stringente de frecvență. Circuitele cu V_{Th} mare sunt în mod tipic utilizate pentru a deconecta de la liniile de alimentare blocurile rapide atunci când acestea sunt în stare statică sau de veghe. Totodată, este întreruptă calea de scurgere a curenților de pierderi datorată tensiunii de prag reduse din blocurile rapide. Tranzistoarele cu tensiune de prag mare pot fi utilizate și la căile de semnal care nu sunt critice în ceea ce privește viteza de comutație. Acest lucru înseamnă că utilizarea tehnologiei cu praguri multiple implică optimizarea și partiționarea simultană a circuitelor la nivel arhitectural pentru a elimina toate căile posibile de pierderi și pentru a obține un consum de putere cât mai redus. Importanța metodelor de partiționare și de conservare a puterii crește cu scalarea geometriei și diminuarea tensiunii de alimentare. În literatura de specialitate sunt

raportate diferite abordări pentru procedura de optimizare, în funcție de constrângerile de viteză, consum, arie și circuite redundante necesare pentru controlul stării de veghe.

Bibliografie

1. Jan M. Rabaey, Anantha Chandrakasan, Borivoje Nikolic – *Digital Integrated Circuits*, Second Edition, Prentice Hall, 2003;
2. J.F. Wakerly, *Digital Design: Principles and Practices Package*, 4th Edition, Prentice Hall, 2005;
3. N.H.E. Weste and D. M. Harris, *CMOS VLSI Design: A Circuits and Systems Perspective*, 4th Edition, Addison-Wesley, 2010;
4. R.J. Baker, *CMOS Circuit Design, Layout, and Simulation*, 3rd Edition, Wiley-IEEE Press, 2010;
5. A. Hastings, *The Art of Analog Layout*, Prentice Hall, 2005
6. documentare pe web – International Technology Roadmap for Semiconductors - <http://www.itrs.net/>