

**Roberto Marani, Anna Gina Perri**  
 Dipartimento di Elettrotecnica ed Elettronica,  
 Laboratorio di Dispositivi Elettronici, Politecnico di Bari  
 Via E. Orabona 4, 70125, Bari, Italy  
 Phone: +39-80-5963314/5963427 Fax: +39-80-5963410 E-mail: perri@poliba.it  
 http://www.dispositivielettronici.com

## **I CRITERI DI PROGETTO DEI MOSFET SUBMICROMETRICI PER ALTE PRESTAZIONI**

**(DESIGN RULES OF NANOMOSFETS FOR HIGH PERFORMANCE)**

**S**ommario: in questo lavoro vengono esaminati i criteri di progetto dei MOSFET submicrometrici per alte prestazioni, illustrando le problematiche connesse ad una miniaturizzazione spinta del dispositivo e le soluzioni tecnologiche previste, allo scopo di ottimizzarne le caratteristiche e di mettere in evidenza i motivi per cui si adotta una determinata soluzione tecnologica.

**A**bstract: in this paper the design rules of nanomOSFETs for high performance are examined.

The problems dealing with nanodevices are illustrated to define the technological solutions to optimise the device characteristics.

### **I. Introduzione**

Il circuito integrato (C.I.) è stato inventato da Kilby nel 1958, I primi integrati furono degli oscillatori a sfasamento e dei flip-flop, prodotti su substrato di germanio.

Sin dai primi anni '70 la tecnologia ha prodotto circuiti integrati con meno di 30 dispositivi per chip (SSI, Small Scale Integration), per passare progressivamente alla MSI (Medium Scale Integration), caratterizzata da  $30 \div 10^3$  dispositivi per chip, alla LSI (Large Scale Integration), con  $10^3 \div 10^5$  dispositivi per chip, alla VLSI (Very Large Scale Integration), con  $10^5 \div 10^7$  dispositivi per chip, e per giungere, sin dai primi degli anni '90, alla ULSI (Ultra Large Scale Integration), caratterizzata da  $10^7 \div 10^9$  dispositivi per chip [1-3].

Questo ultimo tipo di tecnologia riguarda soprattutto i C.I. MOS, perché essi hanno raggiunto i migliori livelli per quanto riguarda la densità di integrazione, il basso consumo e le alte rese di produzione, tanto da assorbire attualmente l'88% del mercato dei C.I.

La rimanente parte del mercato è dei C.I. bipolari (8%), mentre il 4% riguarda i semiconduttori

III-V (MESFET ed HEMT), che per ora non vanno oltre la VLSI.

La miniaturizzazione dei dispositivi implica una maggiore densità di integrazione, una maggiore velocità e quindi funzionamento a frequenze sempre più elevate, una riduzione dei consumi di potenza e una riduzione dei costi.

I moderni dispositivi per ULSI possono effettuare commutazioni ad una velocità maggiore o uguale a 100 Gb/s, consumando una energia ( $\lambda \times P_D$ , dove  $\tau$  è il tempo di commutazione e  $P_D$  la potenza dissipata) di 6 volte inferiore rispetto a quella consumata dai primi C.I.

L'evoluzione della tecnologia dei circuiti integrati avvicina i dispositivi ai limiti imposti al loro funzionamento da leggi fisiche fondamentali.

Come conseguenza, fenomeni a livello microscopico che deteriorano l'affidabilità dei dispositivi hanno sostituito la risoluzione spaziale dei processi tecnologici nel determinare le dimensioni minime delle strutture da integrare.

In questo lavoro vengono esaminati i criteri di progetto dei MOSFET submicrometrici per alte prestazioni, illustrando le problematiche connesse ad una miniaturizzazione spinta del dispositivo e le

soluzioni tecnologiche previste, allo scopo di ottimizzarne le caratteristiche e di mettere in evidenza i motivi per cui si adotta una determinata soluzione tecnologica.

## 2. Evoluzione del MOSFET

La rapida evoluzione della tecnologia dell'elettronica dei semiconduttori è dovuta alla continua richiesta di migliori prestazioni e ad una feroce competizione internazionale volta ad una sempre maggiore integrazione dei dispositivi.

Per quanto riguarda la tecnologia MOS, questa evoluzione è misurata spesso in generazioni di tre anni: il tempo necessario per migliorare di un fattore 4 la capacità delle memorie e per aumentare la densità dei circuiti logici di un fattore compreso tra 2 e 3.

Questa continua evoluzione tecnologica è possibile solo attraverso una riduzione delle dimensioni dei circuiti MOS, che passa attraverso il miglioramento della tecnologia di fabbricazione, nonché della tecnica litografica e dunque della massima risoluzione con essa ottenibile.

La riduzione delle dimensioni dei MOSFET è stata notevolissima negli ultimi tre decenni. Iniziando da una lunghezza di gate di 10  $\mu\text{m}$  nel 1970 si è giunti, riducendo gradualmente le dimensioni, a lunghezze di gate di 65  $\mu\text{m}$ , con una riduzione risultante del 13% ogni anno.

Vista la notevolissima riduzione delle dimensioni dei dispositivi è importante tener conto della sempre crescente complessità dei processi di fabbricazione e dei moltissimi effetti del secondo ordine che si manifestano o che vengono ulteriormente amplificati nei MOSFET submicrometrici.

Questi effetti possono essere classificati in [4-5]:

- *Effetti dovuti a campi elettrici intensi.*
  1. effetto della saturazione della velocità dei portatori;
  2. effetti legati agli "elettroni caldi";
  3. effetto della degradazione della mobilità dei portatori;
  4. effetto di modulazione della lunghezza del canale.
- *Effetti del canale corto.*

Nel processo di miniaturizzazione del dispositivo, viene proposta tipicamente una metodologia, che prevede l'introduzione di alcune regole, deno-

minate regole di scaling.

Peraltro uno scaling improprio delle sole dimensioni dei MOSFET può produrre la presenza o l'aumento dei summenzionati effetti del secondo ordine: basti pensare al notevole incremento del campo elettrico che si verifica quando le tensioni ai morsetti del dispositivo vengono lasciate immutate e invece la lunghezza di gate o lo spessore dell'ossido di gate vengono ridotti, con il conseguente aumento di fenomeni di leakage e degli effetti del canale corto.

Di fondamentale importanza è dunque effettuare un corretto scaling del transistor in modo da stabilire quali parametri (geometrici e/o tecnologici) del MOSFET possono essere aumentati e quali ridotti in modo da migliorare le prestazioni del dispositivo stesso.

## 3. Regole di scaling del MOSFET

Per quanto riguarda le regole di scaling [4-7], i vantaggi consentiti dai circuiti integrati, caratterizzati da una densità di componenti sempre più elevata, ha spinto i progettisti, come già detto, a pianificare MOSFET di dimensioni sempre più ridotte. Peraltro il MOSFET, essendo strutturalmente più semplice del BJT, meglio si presta ad una integrazione submicrometrica,

Se riducessimo però solo le dimensioni, non modificando opportunamente altri parametri del dispositivo stesso, molte delle proprietà del MOSFET risulterebbero degradate.

Infatti, ad esempio, qualora riducessimo la lunghezza del canale  $L$ , lasciando invariata la tensione applicata tra il drain e il source, il campo elettrico lungo il canale diverrebbe più intenso.

Un campo elettrico più intenso riduce il guadagno, a causa della velocità di saturazione, oppure rende più intensi gli effetti dovuti ai portatori caldi. Tale problema può essere risolto mediante una riduzione delle tensioni di polarizzazione, che dovrebbero essere scalate dello stesso fattore con cui sono ridotte le dimensioni del MOSFET.

Questo modo di procedere ha portato alla formulazione di una serie di regole di scaling, che consentono di migliorare la velocità e la densità d'integrazione, requisiti fondamentali nell'ambito della ULSI ma che devono essere verificate nel momento in cui si opera una riduzione delle dimensioni MOS.

Le regole non sono uniche, in quanto dipendono da quali simmetrie si vogliono conservare nel passaggio dal dispositivo MOS originario a quello ottenuto con la procedura di scaling.

Indicando con  $1/\lambda$  il fattore di scaling delle dimensioni ( $\lambda$  maggiore di 1) e con  $1/k$  quello delle tensioni ( $k$  maggiore di 1), abbiamo le seguenti regole di scaling del MOSFET:

Inoltre questo scaling è più vantaggioso perché, restando i potenziali invariati, anche la tensione di soglia  $V_T$  rimane costante, evitando così fenomeni di dispersione sottosoglia che si hanno quando  $V_T$  diminuisce.

Peraltro lo scaling a tensione costante pone seri problemi di affidabilità, cioè riduzione del tempo di vita medio del dispositivo, quando si lavo-

PARAMETRI	VARIABILI	FATTORE DI SCALING
Dimensioni	$W, L, x_{\text{oss}}, r_j$	$1/\lambda$ (riduzione)
Potenziali	$V_{\text{DS}}, V_{\text{GS}}$	$1/k$ (riduzione)
Conc. Drogante	$N_A, N_D$	$\lambda^2/k$ (aumento, se $\lambda > k^{1/2}$ )
Campo elettrico	$\mathcal{E}$	$\lambda/k$
Corrente	$I_{\text{DS}}$	$\lambda/k^2$
Ritardo di gate	$\tau_{\text{pd}}$	$k/\lambda^2$

avendo indicato con  $W$  la larghezza del MOSFET,  $L$  la lunghezza del canale,  $x_{\text{oss}}$  lo spessore dell'ossido e  $r_j$  la profondità di giunzione.

I valori scalati dei parametri  $\mathcal{E}$ ,  $I_{\text{DS}}$ ,  $\tau_{\text{pd}}$  sono ottenuti sostituendo nelle equazioni di base del MOSFET i valori scalati dei parametri dimensioni, potenziali e concentrazione di drogante.

I parametri  $\lambda$  e  $k$  possono essere variati indipendentemente tra loro, seguendo però due metodologie principali:

1) *Scaling a campo elettrico costante, cioè  $k = \lambda$*

Le dimensioni, i potenziali, la corrente e il ritardo di gate si riducono con conseguente aumento della velocità del dispositivo.

Peraltro, utilizzando le equazioni di base del MOSFET, si può dimostrare che la tensione di soglia  $V_T$  diminuisce di un fattore pari a circa  $1/k$ .

2) *Scaling a tensione costante, quindi  $k = 1$*

In questo caso i potenziali rimangono costanti ed aumenta la corrente, proporzionale in saturazione a  $(V_{\text{GS}} - V_T)^2$ , diminuisce il tempo  $\tau_{\text{pd}}$  ottenendo così una maggiore aumento della velocità del dispositivo rispetto al caso 1).

ra con campi elettrici molto elevati.

I MOSFET inizialmente operavano a 5.0V, poi a 3.3V, ora a 2.0V e la prospettiva è di passare a 1.5V fino a 0.9V.

Poiché la lunghezza  $L$  del dispositivo scalato è diminuita, il conseguente aumento del campo elettrico provoca l'insorgenza del fenomeno degli elettroni caldi, che però può essere risolto adottando soluzioni tecnologiche che complicano la struttura del dispositivo (cfr. *struttura LDD*).

Se viene diminuita solo la lunghezza del gate, senza adottare le regole di scaling, il MOSFET opera in regime di canale corto.

In tale situazione, con riferimento alla Fig. 1, possiamo affermare che:

- l'impedenza di uscita del MOSFET è più piccola;
- la corrente sottosoglia viene a dipendere da  $V_{\text{DS}}$ , mentre nel modello per MOSFET a canale lungo dipende da  $V_{\text{GS}}$  secondo la nota

$$I_{D_{st}} \approx e^{\frac{q(V_G - V_T)}{kT}}$$

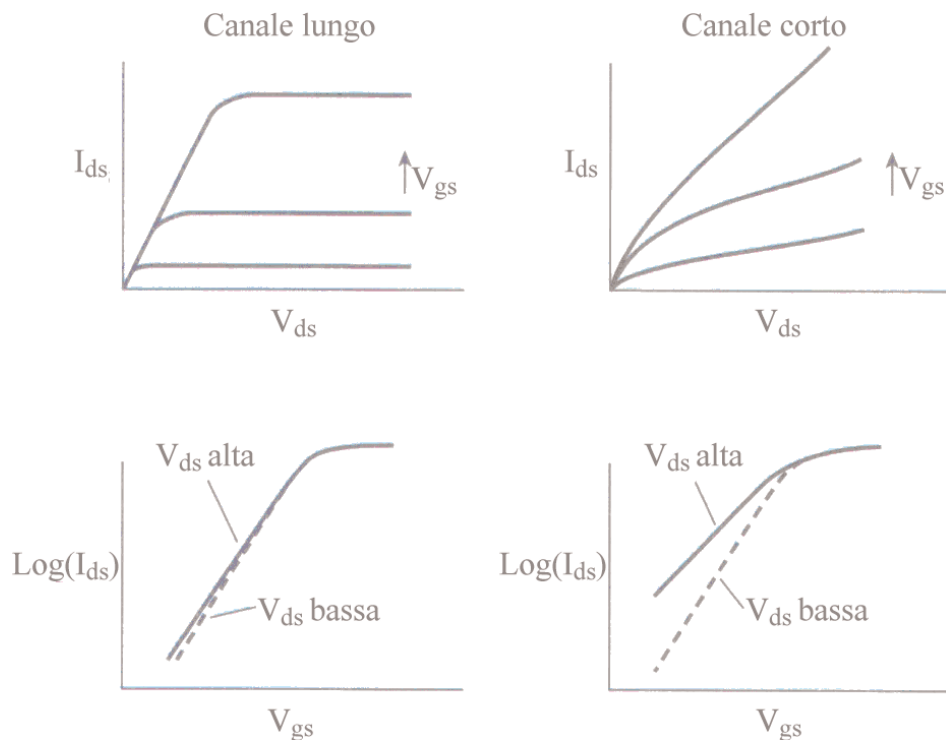


Fig. 1 - Confronto delle caratteristiche I-V di due MOSFET a canale lungo e a canale corto.

relazione

- la  $V_T$  si riduce drasticamente con la lunghezza di gate (effetto di canale corto).

In definitiva gli effetti di canale corto determinano un peggioramento della transconduttanza  $g_m$  del dispositivo perché c'è un controllo meno efficace della tensione  $V_{GS}$  sulla corrente  $I_{DS}$ .

Peraltro un'ipotesi fondamentale, ai fini della deduzione del modello matematico del MOSFET, è che la carica nella regione di carica spaziale superficiale, sia unicamente funzione della tensione tra gate e substrato.

Tale ipotesi viene meno in prossimità della regione di carica spaziale di source e di drain, dove le giunzioni p-n, così come il gate, influenzano la carica spaziale superficiale.

Inoltre, quando la lunghezza  $L$  del canale è molto maggiore dello spessore delle regioni di svuotamento, questa ipotesi causa errori del tutto trascurabili.

Però, riducendo la lunghezza del canale, si possono commettere errori nel valutare la tensione di soglia a causa di una quantità di carica nella regio-

ne di carica spaziale, collegata al gate, significativamente più piccola di quella prevista dalla teoria del primo ordine.

La rimanente carica è collegata alle regioni di svuotamento della giunzione p-n (substrato-canale).

Con riferimento alla Fig. 2, esiste una formula empirica che permette di calcolare il limite inferiore per  $L$ , valida anche per dispositivi submicrome-

$$L_{\min} = 0.4 \left[ r_j x_{\text{oss}} (X_D + X_S)^2 \right]^3$$

trici:

in cui:

- \*  $r_j$  indica la profondità di giunzione, in  $\mu\text{m}$ ;
- \*  $x_{\text{oss}}$  lo spessore dell'ossido, in  $\text{\AA}$ ;
- \*  $X_D$  e  $X_S$  sono rispettivamente l'estensione della regione di svuotamento sotto il drain ed il source, in  $\mu\text{m}$ .

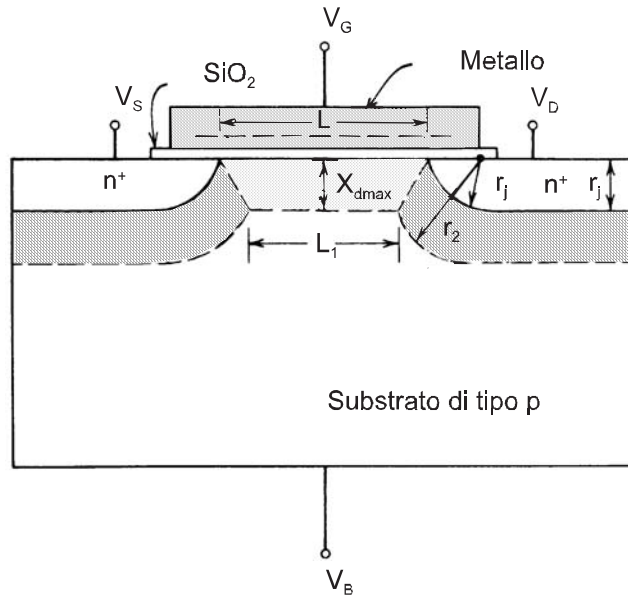


Fig. 2 - Sezione trasversale di un MOSFET a canale corto; la carica indotta nella regione di canale spaziale di gate è supposta contenuta nella regione punteggiata.

Le regioni di svuotamento  $X_D$  e  $X_S$  si riducono perché nel processo di scaling del dispositivo il drogaggio va aumentato.

Lo spessore  $x_{oss}$  dell'ossido può essere ridotto per fabbricazione a meno di  $50\text{\AA}$ , anche se questo si può ottenere tramite processi delicati e costosi.

Il problema maggiore è la riduzione della profondità di giunzione  $r_j$  perché implica un aumento delle resistenze serie parassite  $R_D$  e  $R_S$  del dispositivo. Occorre, dunque, comprendere gli effetti causati dalle resistenze serie. A tal fine è opportuno introdurre il concetto di **fattore  $\theta$  di riduzione della mobilità**. Se il campo elettrico nel canale aumenta, aumenta anche la densità dei portatori nel canale e ciò produce un aumento dello scattering dei portatori tra loro e con l'interfaccia Ossido-Silicio, con conseguente degradazione della mobilità  $\mu$ .

E' stata proposta in letteratura una espressione empirica che modella la degradazione della mobilità:

$$(1) \quad \mu_{eff} = \frac{\mu_n}{1 + \theta(V_{GS} - V_T)}$$

in cui  $\mu_n$  indica la mobilità degli elettroni nel canale quando  $V_{GS}$  è eguale alla tensione di soglia  $V_T$ .

Per un MOSFET ideale (cioè senza resistenze

parassite e senza degradazione della mobilità) che stia funzionando in zona lineare, la corrente di drain è data dalla ben nota relazione:

$$(2) \quad I_{DS} = \mu_n C'_{oss} \frac{W}{L} (V_{GS} - V_T) V_{DS}$$

Se al MOSFET ideale aggiungiamo le resistenze serie parassite  $R_D$  e  $R_S$ , considerandole esterne al dispositivo ideale, e portiamo in conto anche la degradazione della mobilità, l'eq.(2) può essere

$$I_{DS} = C'_{oss} \frac{W}{L} \frac{\mu_n}{1 + \theta(V_{GS} - R_S I_{DS} - V_T)} \cdot [(V_{GS} - R_S I_{DS}) - V_T] [V_{DS} - (R_S + R_D) I_{DS}]$$

riscritta nel modo seguente:

ovvero:

$$(3) \quad I_{DS} = C'_{oss} \frac{W}{L} \mu_n \cdot \frac{(V_{GS} - V_T) V_{DS}}{1 + \left[ \theta + C'_{oss} \frac{W}{L} \mu_n (R_S + R_D) \right] (V_{GS} - V_T)}$$

L'eq. (3) può essere scritta come:

avendo posto:

$$I_{DS} = C'_{oss} \frac{W}{L} \mu_n \frac{(V_{GS} - V_T) V_{DS}}{1 + [\theta + \Delta\theta](V_{GS} - V_T)}$$

In definitiva l'aumento delle resistenze parassi-

$$\Delta\theta = C'_{oss} \frac{W}{L} \mu_n (R_S + R_D)$$

te  $R_D$  e  $R_S$  produce nel complesso un aumento del fattore  $\theta$  di riduzione della mobilità, con conseguente degradazione della trasconduttanza del dispositivo.

L'effetto è particolarmente evidente per piccole lunghezze di gate, in quanto il  $\Delta\theta$  è proporzionale a  $1/L$ .

Il secondo problema che degrada le prestazioni di un MOSFET submicrometrico è la presenza della resistenza di contatto  $R$  (*spreading resistance of the contact*).

Per i transistori bipolari e gli altri dispositivi con un flusso di corrente verticale, la resistenza di contatto effettiva è proporzionale alla resistenza di contatto specifica  $R_C$  ( $\Omega \cdot \text{cm}^2$ ) secondo la relazione:

assumendo che la dimensione del contatto non sia molto piccola.

$$R = \frac{R_C}{Area}$$

Nel caso di un dispositivo MOS, tuttavia, assumendo che il contatto si estenda per l'intera larghezza  $W$  del dispositivo, si può calcolare la resistenza di contatto effettiva  $R$  utilizzando l'espressione dedotta dal modello unidimensionale delle linee di trasmissione:

$$(4) \quad R = \frac{\sqrt{R_C \rho_D}}{W} \coth \left( L_C \sqrt{\frac{\rho_D}{R_C}} \right)$$

in cui  $\rho_D$  indica la resistenza di strato ( $\Omega/\square$ ) e  $L_C$  la lunghezza del contatto.

L'eq. (4) può essere approssimata nei seguenti

due casi, con una accuratezza del 10% circa:

o

$$R = \frac{R_C}{WL_C} \quad \text{se } L_C < 0.6 \sqrt{\frac{R_C}{\rho_D}}$$

Per dispositivi MOS submicrometrici è ovvia-

$$R = \frac{\sqrt{R_C \rho_D}}{W} \quad \text{se } L_C > 1.5 \sqrt{\frac{R_C}{\rho_D}}$$

mente necessario realizzare contatti piccoli e questo comporta un indesiderato aumento della resistenza parassita di contatto  $R$ .

Sono stati sviluppati diversi processi tecnologici [8-9] per risolvere il suddetto problema.

La più diffusa è la tecnica della **salicidazione autoallineata** dell'area di contatto, con cui si riesce ad aumentare l'area di contatto.

Innanzitutto per diffusione vengono realizzate le regioni di source e di drain. Successivamente si incapsula il gate con ossido.

Prima di realizzare le metallizzazioni, si deposita sul silicio un metallo, tipicamente Titanio o Platino, e lo si fa reagire (*sinterizzazione*) con il silicio, che avrebbe dovuto costituire il contatto. La reazione produce uno strato di **silicide** cioè di siliciuro di titanio ( $\text{TiSi}_2$ ) o di platino, avendo così realizzato, con la successiva fase della metallizzazione, una maggiore area di contatto con un materiale (silicide) più conduttivo del semplice silicio drogato.

Quello della salicidazione è un processo delicato e costoso, che riduce però drasticamente le resistenze parassite del MOSFET. Tipicamente per MOS 0.25  $\mu\text{m}$  viene impiegato siliciuro di titanio o di cobalto, con particolari accorgimenti nello sputtering del metallo e nel successivo annealing per realizzare il siliciuro. Si rimanda ai testi in bibliografia per ulteriori approfondimenti.

#### 4. Metodi per ridurre gli effetti da elettroni caldi: "drain engineering"

Eseguendo lo scaling a potenziale costante, il campo elettrico nel canale del dispositivo aumenta

notevolmente.

Quando è talmente alto che, con riferimento, ad esempio, ad un NMOSFET, l'energia guadagnata dagli elettroni tra collisioni successive (libero cammino di circa 100 Å) raggiunge valori di energia superiori all'energia di soglia di ionizzazione da impatto (**impact-ionization threshold energy**), questi elettroni fortemente energizzati (perciò detti **elettroni caldi** [10]) generano una coppia elettrone-lacuna per effetto dell'impatto con altre particelle (in questa situazione gli elettroni caldi riescono ad evitare lo scattering fononico).

Uno dei possibili effetti è che molti di questi elettroni guadagnano abbastanza energia e quantità di moto, lungo la direzione del biossido di silicio, da permettere a tali elettroni di superare la barriera di potenziale all'interfaccia Si/SiO<sub>2</sub> e di danneggiare l'ossido, o di creare cariche positive sia per ionizzazione da impatto nell'ossido sia per formazione di ioni H<sup>+</sup> (incamerati durante i vari processi tecnologici) all'elettrodo di gate

Poiché questa carica può essere intrappolata nell'ossido, i principali effetti deleteri sono uno shift della tensione di soglia V<sub>T</sub> e una diminuzione della transconduttanza g<sub>m</sub> del MOSFET.

La creazione di ioni H<sup>+</sup> all'interfaccia gate-biossido di silicio è senz'altro uno dei meccanismi che più pesantemente degrada le prestazioni del dispositivo.

Infatti questi ioni passano l'interfaccia creando legami non compensati, che possono essere scissi con appena 0.3 eV di energia, creando stati superficiali o trappole, che degradano le prestazioni del MOSFET, in quanto possono catturare portatori.

Sperimentalmente si è visto che questo effetto è più forte al drain, soprattutto quando il MOSFET funziona in saturazione e con tensione V<sub>GS</sub> ≈ V<sub>DS</sub>/2.

Peraltro, effettuando lo scaling a campo elettrico costante, si è osservato che il suddetto effetto diventa meno importante per lunghezze di gate L ≤ 0.2 μm a causa delle basse tensioni utilizzate.

In definitiva il fenomeno degli elettroni caldi va a limitare la tensione di alimentazione massima utilizzabile.

Per far fronte a questo problema sono state sviluppate regole di scaling finalizzate a mantenere costante il campo elettrico massimo  $\mathcal{E}_{\max}$ , ma non

quello medio.

Pertanto, a parità di caduta di tensione progressiva tra drain e source, occorre aumentare la distanza effettiva su cui avviene questa caduta.

In particolare, dalla parte del drain, in saturazione, la giunzione drain-substrato si comporta come una giunzione p-n polarizzata inversamente, per la quale, come è noto,  $\mathcal{E}_{\max}$  è in corrispondenza della sezione metallurgica.

Inoltre un elettrone deve percorrere al più 4 liberi cammini medi per acquisire una energia pari a quella per produrre ionizzazione da impatto, e diventare "caldo".

Ne consegue che per prevenire il fenomeno degli elettroni caldi occorre agire al drain (**drain engineering**) dove si localizza  $\mathcal{E}_{\max}$  (anche se poi i processi tecnologici impiegati vengono applicati anche al source per motivi di simmetria), o:

- aumentando il numero di urti dei portatori, riducendone così il libero cammino medio, oppure
- rendendo graduale il drogaggio della giunzione drain-substrato, con conseguente riduzione di  $\mathcal{E}_{\max}$ .

Il primo tentativo di drain engineering è il **double-diffused drain** [9], tecnica nella quale vengono impiantati nel substrato dapprima atomi di fosforo e successivamente atomi di arsenico a concentrazione più elevata per realizzare le regioni di source e di drain.

Poiché il fosforo diffonde più rapidamente dell'arsenico, vicino al bordo del canale il profilo predominante è quello del fosforo e la giunzione drain-substrato (drogato con atomi di boro) non è troppo brusca.

Il successivo impianto di atomi di arsenico, secondo il profilo mostrato in Fig. 3, produce in sostanza uno spostamento verso il drain della sezione metallurgica e quindi anche parte della regione di carica spaziale della giunzione drain-substrato si sposta verso il drain.

In particolare, se la concentrazione di atomi di fosforo è opportunamente calcolata, si può fare in modo che la regione di svuotamento dalla parte del drain comprenda solo gli atomi di fosforo, per cui la regione effettiva di drain è quella drogata con

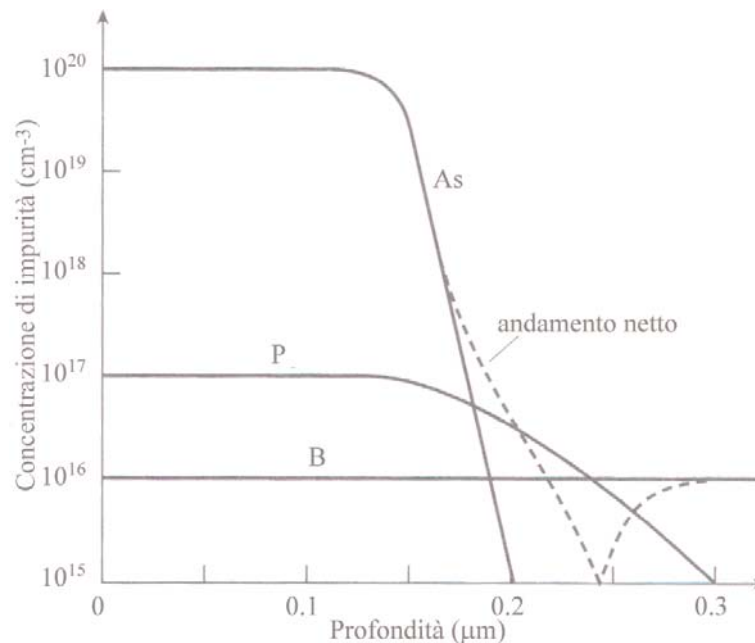


Fig. 3 - Profilo di drogante in un MOSFET con double-diffused drain

atomi di arsenico.

Anche sotto il gate, a causa della diffusione laterale, si crea un analogo profilo di drogaggio.

In tal modo si ottiene una riduzione di circa il 20% del campo elettrico massimo, ma con lo svantaggio di aumentare la profondità delle giunzioni drain-substrato e source-substrato, con conseguente insorgenza degli effetti di canale corto.

In alternativa viene impiegata la tecnica del Lightly Doped Drain (LDD [11]), ovvero la regione di drain (e di source) viene realizzata con una prima diffusione più estesa di atomi a concentrazione  $N_D^-$ , e poi con una seconda diffusione a concentrazione  $N_D^+$  meno estesa, ma più profonda.

L'effetto è sempre quello di realizzare un profilo graduale del drogaggio della regione di drain (e di source) in modo da ridurre il campo elettrico massimo.

La diminuzione dell'intensità dei campi elettrici è illustrata chiaramente in Fig. 4, che riporta l'andamento del campo elettrico all'interfaccia  $\text{SiO}_2$  in funzione della profondità di giunzione, ottenuto

attraverso una simulazione bidimensionale, sia per un dispositivo convenzionale che per uno realizzato con la tecnica LDD.

In definitiva il MOSFET LDD è più affidabile in relazione al fenomeno degli elettroni caldi, ma, tuttavia, è più lento, in quanto la struttura LDD incrementa le resistenze parassite: occorre dunque cercare un compromesso tra prestazioni ed affidabilità del dispositivo.

Infine, un'altra tecnologia, peraltro molto costosa ma specifica per la ULSI, usata per minimizzare l'effetto degli elettroni caldi è la tecnica **Large Angle of Tilt Implanted Drain (LATID)**, ovvero la sacca  $n^+$  viene impiantata verticalmente, mentre quella  $n^-$  viene impiantata con angoli compresi tra  $30^\circ$  e  $60^\circ$ : in tal modo, cioè con un impianto  $n^-$  inclinato (*tilt*), è possibile realizzare la sacca  $n^-$  anche sotto il gate.

I vantaggi che si ottengono con questa tecnica sono essenzialmente:

1. le capacità (*di overlap*) tra gate e regione  $n^-$  sono più piccole rispetto al caso in cui la zona  $n^+$  finisce per errore sotto il gate, e possono inoltre essere controllate;
2. le resistenze parassite sono minori;

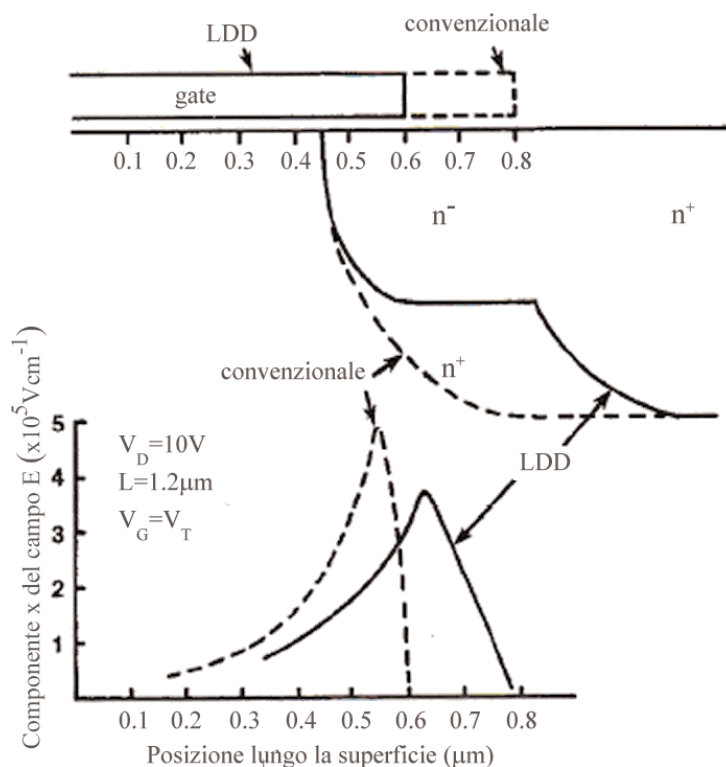


Fig. 4 - Campo elettrico all'interfaccia  $\text{SiO}_2$  in funzione della profondità per due dispositivi MOS (convenzionale e realizzato con la tecnica LDD) aventi  $L=1.2\mu\text{m}$ ,  $V_{DS}=8.5 \text{ V}$ ,  $V_{GS}=V_T$ .

3. il campo elettrico nel canale è più basso;
4. il dispositivo è autoallineato e presenta una lunghezza di canale minore.

### 5. MOSFET submicrometrici: struttura risultante

Un altro problema associato ai MOSFET submicrometrici è l'accrescimento di ossidi dallo spessore variabile tra 40 e 100 Å. Ossidi così sottili sono particolarmente vulnerabili al fenomeno degli elettroni caldi, così come alla presenza di altre cariche spurie, che possono addirittura provocare il fenomeno della rottura dielettrica.

La capacità dell'ossido a resistere ai suddetti fenomeni è nota come **robustezza dell'ossido**.

L'accrescimento di ossidi sottili è, tuttavia, un problema che si può risolvere solo raffinando i processi tecnologici.

Inoltre, al ridursi delle dimensioni del dispositivo, viene anche ridotta la profondità delle giunzioni di source e di drain, realizzando così **giunzioni**

### superficiali.

Inoltre, essendo drain e source molto vicini, avviene il fenomeno del **DIBL (Drain Induced Barrier Lowering, [4-5])**, cioè un abbassamento della barriera di potenziale dalla parte del gate verso il source, ad opera della polarizzazione di drain, consentendo ad un maggior numero di portatori di diffondere dalla regione di source verso il canale.

Questo si manifesta sottoforma di una forte degradazione della  $g_m$ , ossia nella necessità di escursioni maggiori della  $V_{GS}$  per modulare la  $I_{DS}$ .

Peraltro il fenomeno del DIBL provoca una riduzione  $\Delta V_T$  della tensione di soglia secondo la relazione:

$$(5) \quad \Delta V_T \propto \exp\left(-kLr_j^{-1/3}x_{oss}^{-1}\right)$$

in cui  $k$  è una costante empirica, dipendente dalla tecnologia.

Dunque, per non ridurre troppo la tensione di soglia  $V_T$ , occorre ridurre la profondità di giunzione  $r_j$  man mano che il dispositivo viene scalato.

Questo, però, implica un aumento della resistenza di strato delle regioni diffuse di source e di drain, prima della salicidazione.

La maggior parte delle tecnologie profondamente scalate presenta dei valori di resistenza di strato variabili da 500 a 1000  $\Omega/\square$ . In particolare è di grande importanza la concentrazione superficiale, perché da essa dipende la resistenza di contatto specifica.

Sagomare opportunamente il drogaggio, sia in profondità che per la lunghezza del canale, in modo da ottimizzare le prestazioni del dispositivo e/o da prevenire gli effetti di canale corto, è possibile ma con costi molto elevati e processi tecnologici complessi.

In Fig. 5 è mostrata la sezione trasversale di un MOSFET nanometrico, in cui sono evidenziate

ni, essendo n- in prossimità dell'interfaccia Si/SiO<sub>2</sub>, consente di evitare che le RCS delle giunzioni source-substrato e drain-substrato possano venire a contatto tra di loro, evitando cioè il *reachthrough*.

Peraltro conviene mantenere alto il drogaggio del substrato per evitare sia il reachthrough che l'effetto del corpo; inoltre il suo valore deve essere scelto in modo da minimizzare anche il fenomeno del latch-up [4-5].

Le prestazioni del dispositivo vengono migliorate sagomando il profilo di drogaggio nel canale ed intorno ad esso, introducendo uno **steep channel**, ottenuto con impianti che hanno il picco di concentrazione e profondità al di sotto di quelli delle regioni di source e di drain. Spesso per questi impianti si preferisce usare indio anziché boro perché l'indio ha un coefficiente di diffusione più basso rispetto al boro e quindi si riduce la profondità di penetrazione.

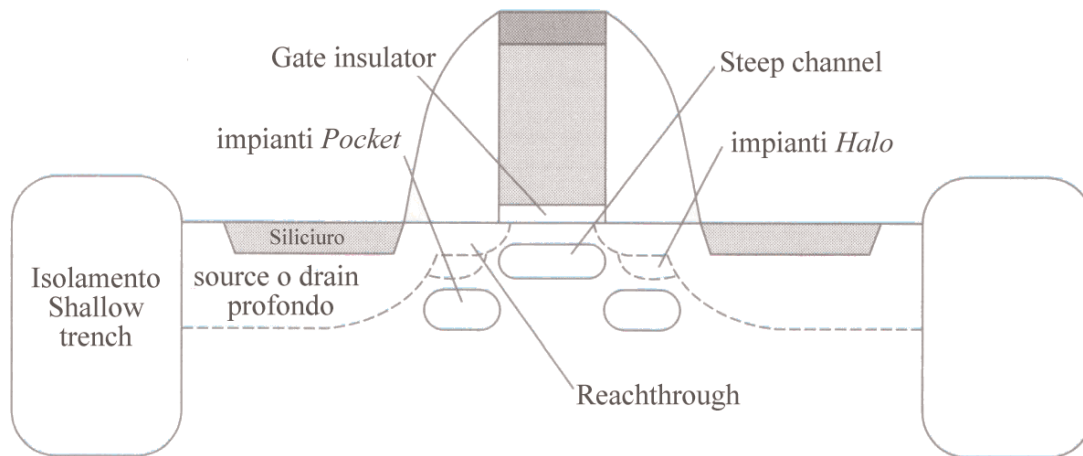


Fig. 5 - Sezione trasversale di un MOSFET nanometrico

tutte le soluzioni tecnologiche adottate per migliorare le condizioni di funzionamento del dispositivo.

In particolare, gli **strati di silicide** sulle regioni di source, drain e sul gate di polisilicio, sono stati inseriti per diminuire le resistenze parassite.

Le regioni di source e di drain sono state realizzate secondo la tecnica LDD o varianti, allo scopo, come si è già detto nel paragrafo precedente, di minimizzare gli effetti da elettroni caldi. Inoltre il profilo di drogaggio delle suddette regio-

Il motivo dell'introduzione dello **steep channel** è quello di creare un gradiente di concentrazione di cariche (positive), e quindi un campo elettrico nel canale che, allontanando le cariche (negative), produce una loro riduzione con conseguente riduzione dello scattering e quindi aumento della mobilità.

Gli **halo implant** (n-) prevengono gli effetti di canale corto, in quanto schermano il canale dalle

regioni di source e di drain più profonde ( $n^+$ ).

I **pocket impiant** ( $n^-$ ) diminuiscono l'estensione delle regioni di svuotamento di source e di drain, riducendo così il fenomeno del DIBL, cioè riducendo la dipendenza della tensione di soglia  $V_T$  dalla lunghezza  $L$  di gate, e la dipendenza della corrente sottosoglia dalla tensione di polarizzazione drain-source  $V_{DS}$ .

Infine le shallow trench hanno lo scopo di garantire l'isolamento elettrico tra dispositivi adiacenti e prevenire il fenomeno del latch-up.

## 6. Conclusioni

La contrazione delle geometrie dei dispositivi (attivi, passivi e linee di interconnessione), che caratterizza lo sviluppo della microelettronica, comporta inevitabilmente un aggravamento delle condizioni operative delle strutture con preoccupanti riflessi sulle loro caratteristiche di affidabilità.

Questo processo è inevitabile dal momento che la tecnologia evolve rapidamente verso i limiti ultimi posti al buon funzionamento dei dispositivi da leggi fisiche fondamentali e che l'avvicinamento di tale condizione riduce, per definizione, i margini di sicurezza delle passate generazioni di componenti.

Inoltre la natura sempre più microscopica dei fenomeni che portano alla rottura dei dispositivi e, conseguentemente, alla crescente difficoltà di riconoscerli e studiarli, indurrà ad un uso sempre più sistematico di sofisticate tecniche di analisi dei complessi circuiti ULSI, richiedendo che tutti i grossi utilizzatori di componenti elettronici sviluppino ed acquisiscano competenze molto vaste e diversificate (dalla chimica e chimico-fisica dei processi fino alle tecniche sperimentali e all'architettura dei sistemi) nonché mezzi ed attrezzature sempre più simili a quelle fino ad oggi utilizzate unicamente dai costruttori di circuiti integrati.

**BIBLIOGRAFIA**

- [1] C.Y. Chang e S.M. Sze: "ULSI Devices"; John Wiley and Sons, 2000.
- [2] C.Y. Chang e S.M. Sze: "ULSI Technology"; McGraw-Hill, 1996.
- [3] Y. Taur e T.H. Ning: "Fundamentals of Modern VLSI Devices"; Cambridge University Press, 1998.
- [4] A. G. Perri: "Introduzione ai dispositivi micro e nanoelettronici"; vol. 1, Edizioni Biblios, Bari, pp. 1 - 473 , settembre 2007, ISBN 978-88-6225-000-9.
- [5] A. G. Perri: "Introduzione ai dispositivi micro e nanoelettronici"; vol. 2, Edizioni Biblios, Bari, pp. 1 - 356, settembre 2007, ISBN 978-88-6225-000-9.
- [6] R. Yan, A. Ourmazd e K. Lee.: "Scaling the Si MOSFET: from Bulk to SOI to Bulk"; IEEE Transactions on Electron Devices, 1992.
- [7] D. Frank, Y. Taur e H.-S.P. Wong: "Generalized scale length for two-dimensional effects in MOSFET's"; IEEE Transactions on Electron Devices Letters, 1993.
- [8] S.A. Campbell: "The Science and Engineering of Microelectronic Fabrication"; Oxford University Press, seconda edizione, 2001.
- [9] G.S. May e S.M. Sze: "Fundamentals of Semiconductor Fabrication"; John Wiley and Sons, 2004.
- [10] K. Mistry e B. Doyle: "How Hot Carriers Degrade N-Channel MOSFETs?" Circuits & Devices, 1995.
- [11] S. Ogura, P.J. Tsang, W.W. Walker, D.L. Critchlow e J.F. Shepard: "Design And Characteristics Of The Lightly Doped Drain-source (LDD) Insulated Gate Field Transistor"; IEEE Journal Of Solid-State Circuits, vol. sc-15, n.4, 1980.