



UNIVERSITÀ DI PADOVA
FACOLTÀ DI INGEGNERIA
DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE

Laurea Triennale in Ingegneria Elettronica

Legge di Moore e Tecnologie Microelettroniche Future

Relatore: Ch.mo Prof. Dr. Ing. Gaudenzio Meneghesso

Co-relatore: Ing. Alessio Griffoni

Candidato: Andrea Muraro

A.A. 2009-2010

Premessa

In questa tesi, prima, verranno analizzati i principali fattori tecnologici e fisici che hanno consentito lo sviluppo esponenziale della microelettronica in un arco temporale di oltre 40 anni. Successivamente, si descriveranno quali sono le più importanti iniziative in atto al fine di consentire il mantenimento di questo sviluppo nei prossimi 10-15 anni e i nuovi scenari che potrebbero consentire prestazioni superiori a quelle che la microelettronica ha già fornito e sta ancora fornendo. Ciò che mi ha spinto a scrivere una tesi su tali argomenti è stata l'attualità degli stessi e il desiderio di approfondire maggiormente i contenuti dei corsi seguiti finora.

Indice

Capitolo 1

- 1.1 Introduzione.....5
- 1.2 Legge di Moore.....6
- 1.3 Descrizione di un MOSFET.....8
- 1.4 Regole di Scaling.....11
- 1.5 Scaling dei MOSFET.....15
- 1.6 I limiti della legge di Moore.....18

Capitolo 2

- 2.1 Interconnessioni.....21
- 2.2 Strained Silicon.....25
- 2.3 Sviluppo e utilizzo di dielettrici high-k.....30
- 2.4 Tecnologia SOI.....35
- 2.5 Tecnologia FinFET.....40
- 2.6 Tecnologie non-CMOS.....47

Conclusioni.....50

Capitolo 1

1.1. Introduzione

Il Metal–Oxide–Semiconductor Field-Effect Transistor (MOSFET) è stata senza ombra di dubbio una delle scoperte più importanti del XX secolo. Grazie alle sue proprietà, infatti, si è potuti arrivare ad uno sviluppo tecnologico e ad una potenza di calcolo che fino alla metà del novecento sarebbe stata impensabile. Per rendersi conto di questo progresso basta pensare che la prima vera macchina da calcolo, l'ENIAC (Electronic Numerical Integrator And Computer), prodotta nel 1946, era composta da circa 18000 valvole termoioniche, occupava un'intera stanza, dissipava una potenza di circa 200 kW ed eseguiva un'istruzione ogni 0.5 ms. Oggigiorno, attraverso l'uso di microprocessori come il Core 2 Duo sviluppato da Intel nel 2007, si possono realizzare sistemi in grado di eseguire un'istruzione in meno di un ns e che dissipano potenze inferiori a 100 W. Nella società moderna un numero notevole e sempre crescente di prodotti, incorpora tecnologie elettroniche digitali, cioè basata su logica binaria, e quella CMOS (il cui mattone fondamentale è il MOSFET) ha soppiantato quasi ovunque tutte le precedenti tecnologie. Microprocessori, personal computer, telefoni cellulari, strumenti di misura o dispositivi per l'archiviazione dei dati sono solo alcuni dei molteplici esempi di applicazioni che hanno cambiato la vita quotidiana delle persone del nostro tempo.

1.2. Legge di Moore

Nel 1975, Gordon Moore, uno dei fondatori di Intel, formulò un'osservazione, che diventò poi nota con il nome di Legge di Moore, la quale affermava che il numero dei transistor che formavano un circuito integrato sarebbe raddoppiato ogni 12 mesi. La legge fu successivamente ridimensionata al raddoppio ogni 18 mesi prima e a 24 poi e stabilì una sorta di mappa guidata per l'evoluzione della tecnologia CMOS. La Figura 1.2.1 mostra quali erano le previsioni sul numero di transistor di Moore e qual'è stato negli anni il numero effettivo di MOSFET all'interno dei chip. Come si può notare le previsioni fatte sono state sempre rispettate e in alcuni casi, come quello dell'Intelium 2 sono state addirittura superate.

Comprendere le cause di tale sviluppo esponenziale non risulta difficile se si va a considerare il tipo di mercato a cui aziende come Intel, AMD, Motorola (ora Freescale), ST Microelettronics etc. si rivolgono che è in grado di assorbire una quantità estremamente elevata di micro-circuiti elettronici. Lo sviluppo dei circuiti integrati, infatti, ha consentito non solo di presentare nuovi prodotti, ma anche di sostituire quelli già esistenti riducendone il costo e aumentandone le prestazioni e la versatilità. Basti pensare al continuo sviluppo di beni di massa come computer, lettori mp3, telefoni cellulari, memorie flash USB etc..

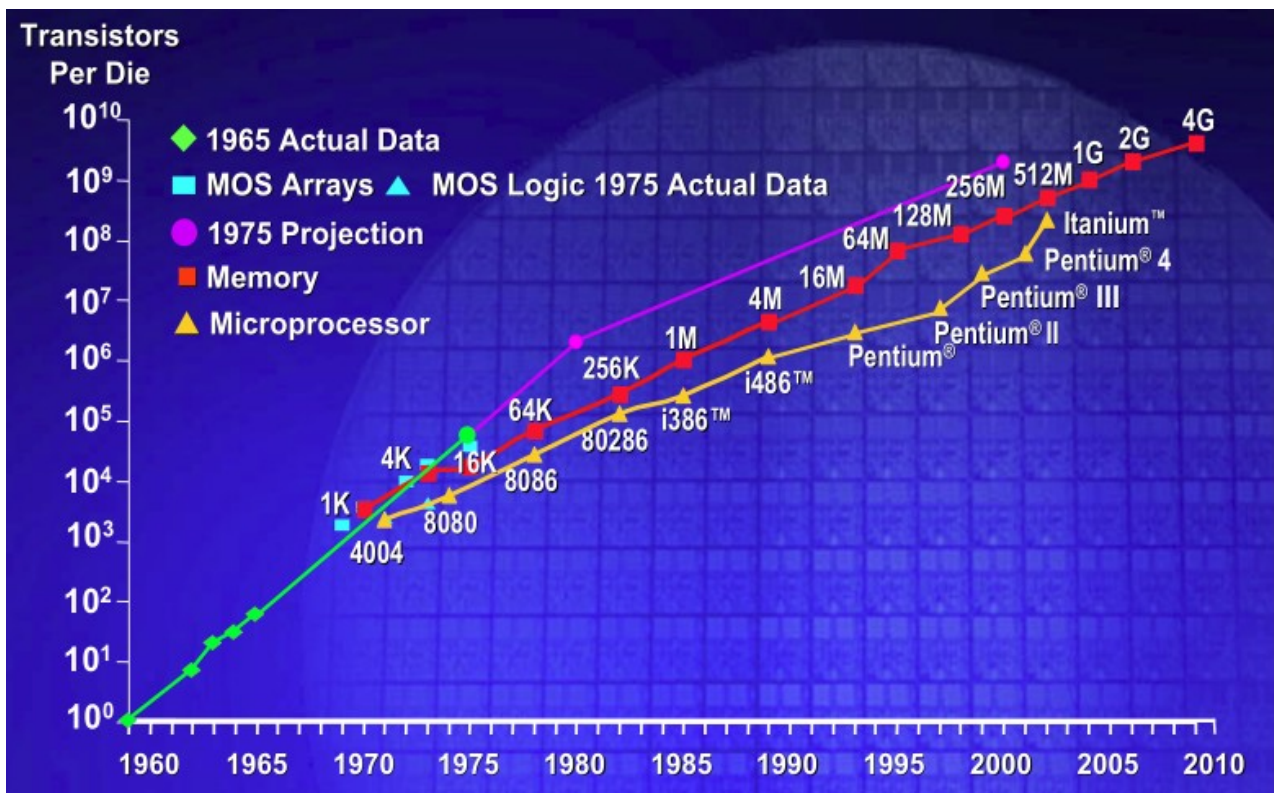


Fig 1.2.1: L'andamento nel tempo della legge di Moore [1].

Con l'aumentare dello sviluppo tecnologico tutti questi beni migliorano le loro prestazioni, inducendo ad acquistarne di nuovi e quindi, di fatto, il mercato non è mai saturo. L'aumento della densità di integrazione infatti è determinato soprattutto dalle risorse investite in ricerca e sviluppo, in quanto, pur esistendo barriere tecnologiche e fisiche non superabili, quando tali barriere stanno per essere raggiunte, altri materiali e altre tecnologie le spostano in avanti portando quindi ad uno sviluppo i cui limiti teorici e pratici non sono del tutto predicibili con esattezza. Per capire appieno quali siano gli aspetti tecnici sui quali si è basata la rivoluzione della microelettronica bisogna considerare l'elemento base che la costituisce, ossia il MOSFET, che sarà analizzato nel paragrafo successivo.

1.3. Descrizione di un MOSFET

La Fig 1.3.1 mostra la struttura di un MOSFET tradizionale (di tipo N) mettendo in evidenza tutte le sue regioni fondamentali.

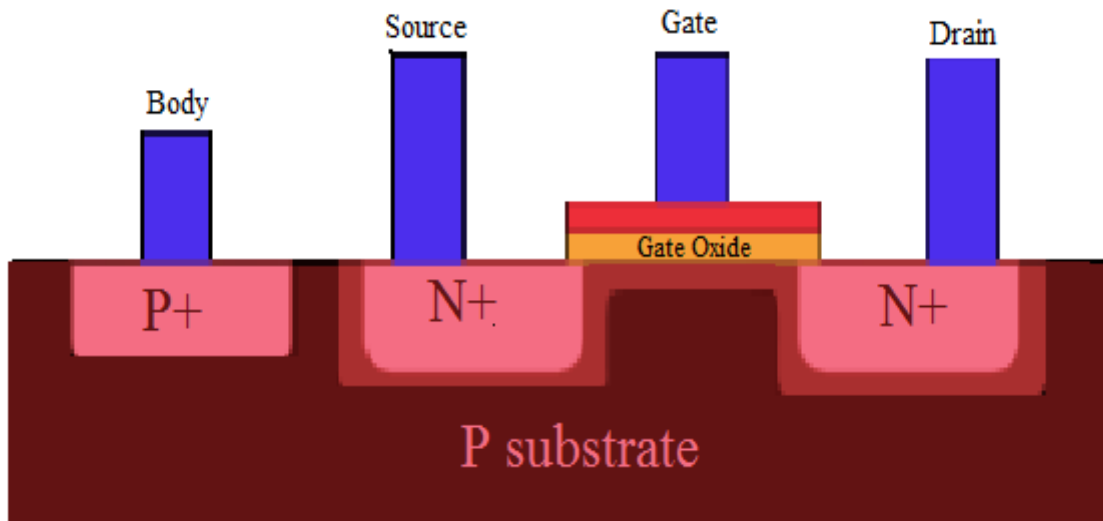


Fig 1.3.1:Modello di un MOSFET bulk

Come si può notare esso ha 4 terminali, ognuno dei quali controlla una regione del dispositivo. Source e drain sono costruiti in materiale semiconduttore fortemente drogato di tipo “N”, impiantati su un substrato semiconduttore di tipo “P” drogato debolmente. Le zone fortemente drogate sono a contatto con opportuni salicuri, per cui è possibile stabilire una differenza di potenziale tra le due (ed anche rispetto al body, anch’esso avente un contatto). La regione di gate invece giace sopra il substrato di tipo “P” ed è separata da quest’ultimo da un sottile strato di dielettrico, detto ossido di gate, ed è costruito in modo da sovrastare alla parte di substrato che separa il source dal drain, la quale verrà chiamata zona di canale. Quando il terminale di gate viene portato ad una tensione positiva superiore ad un valore critico, detto tensione di soglia, gli

elettroni vengono attirati nella zona di canale sottostante al gate, formando di fatto un percorso conduttivo tra il source ed il drain (detto canale), tra i quali può quindi fluire una corrente elettrica. Il MOSFET può quindi essere considerato come un interruttore, il cui stato (conduttivo o non conduttivo), è controllato elettricamente dalla tensione applicata sull'elettrodo di gate. Le prestazioni di un transistor MOS variano in base ai parametri geometrici (lunghezza L e larghezza W) e a quelli chiave, ossia la tensione di soglia V_{th} e la transconduttanza g_m . Per quanto riguarda MOSFET di tipo P (PMOS) invece il funzionamento è del tutto analogo a quello del transistor NMOS, senonché le polarità delle tensioni ed il tipo di droganti sono invertiti (drain e source fortemente drogati con impurità di tipo "P" costruiti su un substrato di tipo "N") e la corrente tra il source e il drain può fluire quando tra gate e body è applicata una differenza di potenziale negativa (inferiore alla tensione di soglia, anch'essa negativa).

Ciò che ha reso il MOSFET un componente così importante, come detto in precedenza, è stato lo sviluppo della tecnologia CMOS. Tale tecnologia si basa sull'utilizzo combinato di NMOS e PMOS, le cui caratteristiche sono complementari, in modo da ottenere delle funzioni logiche di interesse. Si consideri, ad esempio, la struttura logica più elementare, l'inverter CMOS, rappresentato nella Fig 1.3.2.

Essendo i gate dei due MOSFET connessi tra loro e al segnale di ingresso, quando esso è a V_{DD} (corrispondente al livello logico "1"), il transistor di tipo N sarà in conduzione, e avrà una bassa resistenza serie, mentre quello di tipo P sarà spento, e presenterà un'altissima resistenza, e quindi l'uscita sarà bassa. Il comportamento sarà l'opposto quando l'ingresso sarà a 0 V (corrispondente al

livello logico “0”).

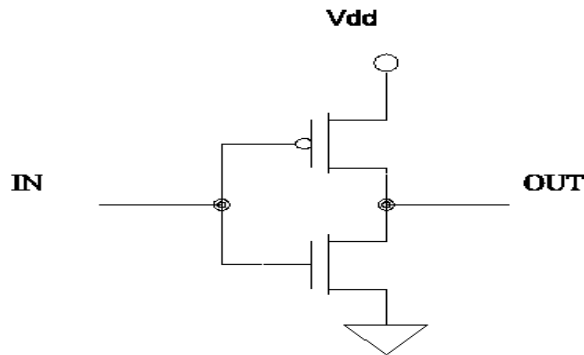


Fig 1.3.2: Schema circuitale di un inverter CMOS

Il circuito si comporta, quindi, come una serie di due interruttori e si ha passaggio di corrente solo durante le transizioni di stato logico. Una funzione logica (in questo caso, la negazione) realizzata in CMOS, ha quindi la proprietà di dissipare una bassa potenza quando i livelli di ingresso e uscita sono stabili (quindi il consumo di potenza statica non risulta elevato) e di dissipare una potenza non trascurabile solo durante la commutazione da uno stato logico all'altro (potenza dinamica).

Un fattore importante, almeno fino al raggiungimento del nodo tecnologico di 130 nm, è stato sicuramente la possibilità di ridurre le dimensioni del MOSFET senza richiedere particolari innovazioni del principio di funzionamento di base che contraddistingue i MOS di tipo bulk. Le dimensioni del transistor sono grandi rispetto a quelle per cui gli effetti quantistici divengono importanti e questo ha consentito di descrivere il MOSFET attraverso un insieme di equazioni differenziali e da una di queste, l'equazione di Poisson, si deduce un insieme di regole, dette regole di scaling, che sono state il cardine per lo sviluppo della microelettronica e che saranno analizzate nel prossimo paragrafo.

1.4. Regole di scaling

Le regole di scaling determinano il modo in cui i parametri fondamentali del MOSFET devono essere variati per scalare le dimensioni. Esse prendono il nome di scaling a campo costante, a tensione costante o generale a seconda del parametro che viene modificato.

Con lo scaling a campo costante (o scaling completo) tensioni e dimensioni sono scalate dello stesso fattore S . In questo modo le intensità dei campi elettrici nei dispositivi scalati, vengono mantenute invariate rispetto a quelli originali. Ciò garantisce l'integrità dei dispositivi ed evita fenomeni di breakdown o altri effetti secondari indesiderati. In questo modo vengono migliorate la densità di integrazione, i tempi di ritardo e il consumo di potenza. Come mostrato dalla Tabella 1.4.1, a causa della diminuzione della capacità C_{gate} e del valore di R mantenuto costante, la velocità del circuito cresce in maniera lineare mentre la potenza dissipata diminuisce con andamento quadratico.

Nella realtà però lo scaling a campo costante non è praticabile dato che, scalare le tensioni in modo arbitrario, obbligherebbe il costruttore a prevedere più tensioni di alimentazione per rendere compatibili i nuovi dispositivi con quelli già esistenti, introducendo, di fatto, notevoli costi aggiuntivi. Di conseguenza i progettisti sono portati a rispettare degli standard ben definiti per tensioni e livelli del segnale, riducendo in modo arbitrario solo le dimensioni.

Questo tipo di scaling viene detto a tensione costante. Fino ai primi anni '90 lo standard di tensione utilizzato era di 5 V e con l'avvento delle tecnologie CMOS da 0.5 μm cominciarono ad essere utilizzati gli standard a 3.3 V prima, 2.5 V

poi e via via la tensione di alimentazione venne abbassata con il passare dei nodi tecnologici come mostrato dalla Fig 1.4.2. Con lo scaling a tensione costante in dispositivi a canale corto, non vengono apportate migliorie alle prestazioni rispetto allo scaling a campo costante e, anzi, si possono notare dei peggioramenti in termini di potenza dissipata. Inoltre, mantenendo la tensione costante, aumentano i problemi legati a fenomeni fisici come l'effetto dei portatori caldi o il breakdown dell'ossido di gate. Ciò è causato dall'aumento dei campi elettrici all'interno dei dispositivi dovuto al fatto che le tensioni non scalano allo stesso modo delle dimensioni. Quando si è in presenza di campi elettrici maggiori di 10^6 V/ μm infatti l'effetto dovuto ai portatori caldi (o hot-carrier) può portare a seri problemi di affidabilità a lungo termine del dispositivo, nel senso che un circuito può risultare danneggiato dopo essere stato utilizzato per un certo periodo di tempo, degradando di fatto la caratteristica I-V del dispositivo come mostrato dalla Fig 1.4.3.

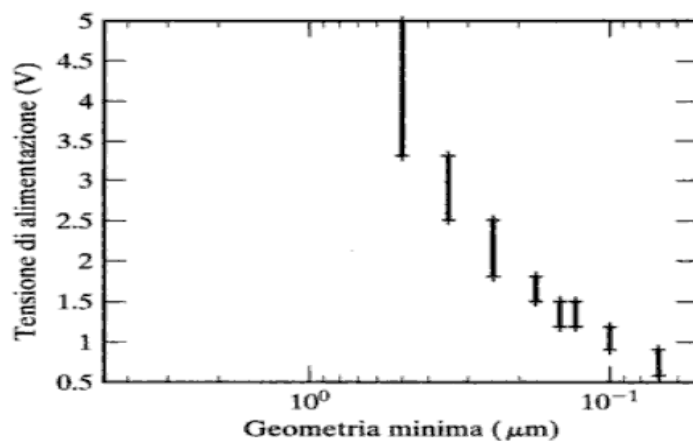


Fig 1.4.2: Evoluzione della tensione minima e massima in funzione della geometria minima [2]

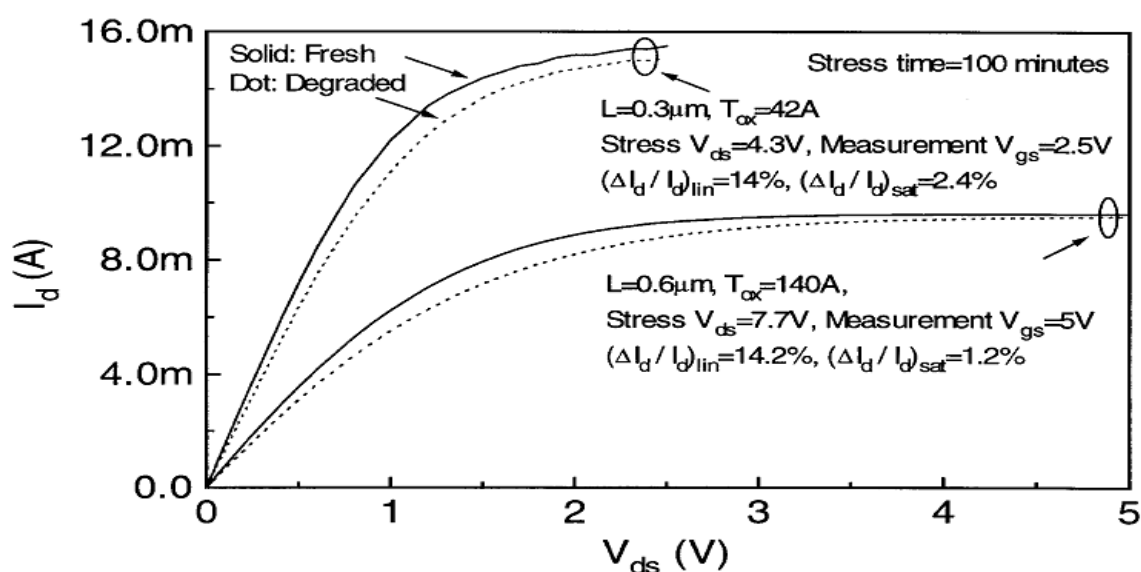


Fig 1.4.3: Degradazione della caratteristica I-V dovuta agli effetti da elettroni caldi.[3]

Tali problemi rendono quindi impraticabile la riduzione delle dimensioni a parità di tensione di alimentazione.

Lo scaling più usato risulta quindi quello generale, in cui le tensioni di alimentazione diminuiscono ma non con la stessa velocità della tecnologia. Per passare dalla tecnologia a $0.5\ \mu\text{m}$ a quella a $100\ \text{nm}$ ad esempio la tensione è passata da $5\ \text{V}$ a $1.5\ \text{V}$. A questo punto viene spontaneo chiedersi, perché non utilizzare sempre lo scaling a campo costante se mantenere le tensioni di alimentazione ad un valore maggiore non comporta nessun beneficio evidente? La risposta è data da due fenomeni che finora non sono stati tenuti in considerazione. Innanzitutto, abbassare la tensione di alimentazione, implica anche una riduzione della tensione di soglia del dispositivo quando queste diventano confrontabili. Ridurre la soglia sotto ad un certo valore rende difficile il completo spegnimento del dispositivo, aumentando il fenomeno delle correnti di sottosoglia (ossia la formazione di una corrente tra Drain e Source quando il dispositivo dovrebbe essere in interdizione). Come detto in precedenza, quando

la tensione applicata tra gate e source V_{GS} risulta minore della tensione di soglia V_{th} , in linea teorica il MOSFET si trova in uno stato di non conduzione, dato che il canale tra source e drain è chiuso. In realtà però, quando la tensione di soglia è ridotta a valori vicini allo zero, anche se la tensione V_{GS} è minore di della stessa, alcuni elettroni scorrono comunque dal source al drain dando quindi origine ad una corrente, detta appunto di sottosoglia, come mostrato dalla Fig 1.4.4. Questo comporta una dissipazione non trascurabile di potenza statica e quindi limita lo scalo delle tensioni.

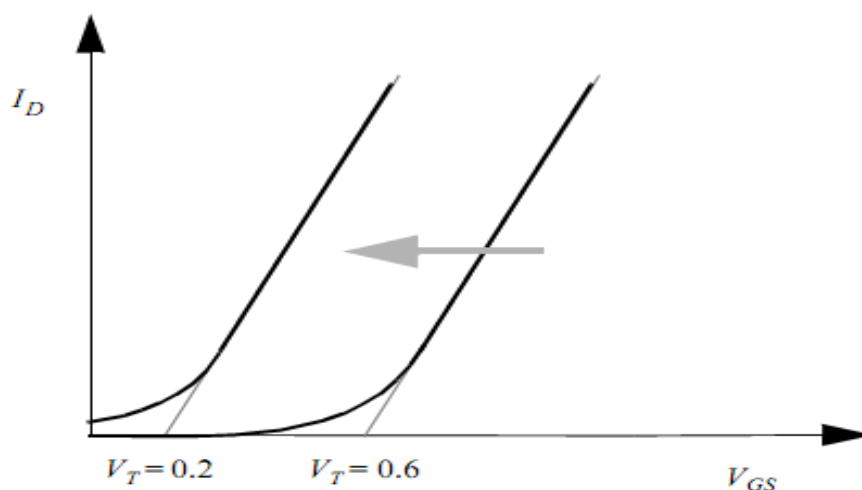


Fig 1.4.4: Diminuendo la tensione di soglia la corrente di sottosoglia aumenta [4]

Inoltre alcune tensioni intrinseche (come il bandgap del silicio e il potenziale intrinseco delle giunzioni) sono parametri caratteristici del materiale e quindi non possono scalare. Lo scaling di tipo generale è quindi fondamentale e, come si vede in Tabella 1.4.1, le dimensioni sono scalate di un fattore S mentre le tensioni di un fattore U . Con lo scaling generale le prestazioni vengono

migliorate dello stesso fattore rispetto agli altri scaling mentre in termini di dissipazione di potenza il modello generale si colloca in posizione intermedia rispetto agli altri due ($S < U < 1$).

Parametri	Espressioni	Scaling a campo costante	Scaling Generale	Scaling a tensione costante
W, L, t_{ox}		1/S	1/S	1/S
Vdd, V_t		1/S	1/U	1
Area/Dispositivo	WL	1/S ²	1/S ²	1/S ²
C_{ox}	1/ t_{ox}	S	S	S
C_{gate}	$C_{ox}WL$	1/S	1/S	1/S
I_{sat}	$C_{ox}WV$	1/S	1/U	1
R_{on}	V/ I_{sat}	1	1	1
Ritardo	$R_{on}C_{gate}$	1/S	1/S	1/S
P	$I_{sat}V$	1/S ²	1/U ²	1
Densità di Potenza	P/Area	1	S ² /U ²	S ²

Tab 1.4.1: Tipo di scaling tecnologico per dispositivi a canale corto

1.5. Scaling dei MOSFET

Con il passare degli anni il mercato dei microprocessori si è concentrato sempre più sulla miniaturizzazione dei MOSFET. Tutto ciò per far fronte alla domanda sempre più crescente di microprocessori aventi elevate prestazioni di calcolo e costi minori. MOSFET più piccoli aumentano la loro efficienza principalmente per due motivi: il primo è che, con la riduzione delle dimensioni, diminuiscono le capacità parassite. Questo si traduce a livello di prestazioni in un minor

tempo richiesto per far commutare lo stato del MOSFET da un livello logico all'altro e quindi il sistema risulterà più veloce. Il secondo è legato al fatto che dispositivi più piccoli permettono un maggior livello di integrazione, dando la possibilità ai progettisti di sviluppare sistemi più complessi sulla stessa superficie di silicio, e quindi di ridurre notevolmente i costi a parità di prestazioni. La riduzione delle dimensioni non comporta però solo fattori positivi. Oltre ai problemi di cui si è parlato in precedenza (correnti di sottosoglia, effetti legati ai portatori caldi e breakdown dell'ossido di gate) esiste un ulteriore problema di natura termica. Producendo MOSFET più piccoli infatti, come detto in precedenza, sulla stessa superficie di silicio si possono integrare molti più MOSFET, e questo ha l'effetto positivo di ridurre i costi ma quello negativo di aumentare la densità di potenza. Ciò comporta un aumento del calore, il quale ha un effetto estremamente negativo nei circuiti elettronici principalmente per due motivi. Il primo è che con l'aumentare della temperatura la mobilità di elettroni e lacune si riduce notevolmente, con un conseguente abbassamento delle prestazioni.

Il secondo è che temperature troppo elevate possono portare alla distruzione del circuito o comunque possono ridurre notevolmente il tempo di vita dello stesso. Per ovviare al problema i circuiti vengono dotati di sistemi di raffreddamento sempre più evoluti, in grado di dissipare elevatissime quantità di calore per unità di superficie. Per rendersi conto dell'efficienza di questi dispositivi basta pensare che i moderni dissipatori sono in grado di dissipare potenze superiori ai 100 W/cm^2 . La Fig 1.5.1 mostra un paragone tra la densità di potenza che dissipa una cpu e quella dissipata da apparati di altra natura ad alta potenza.

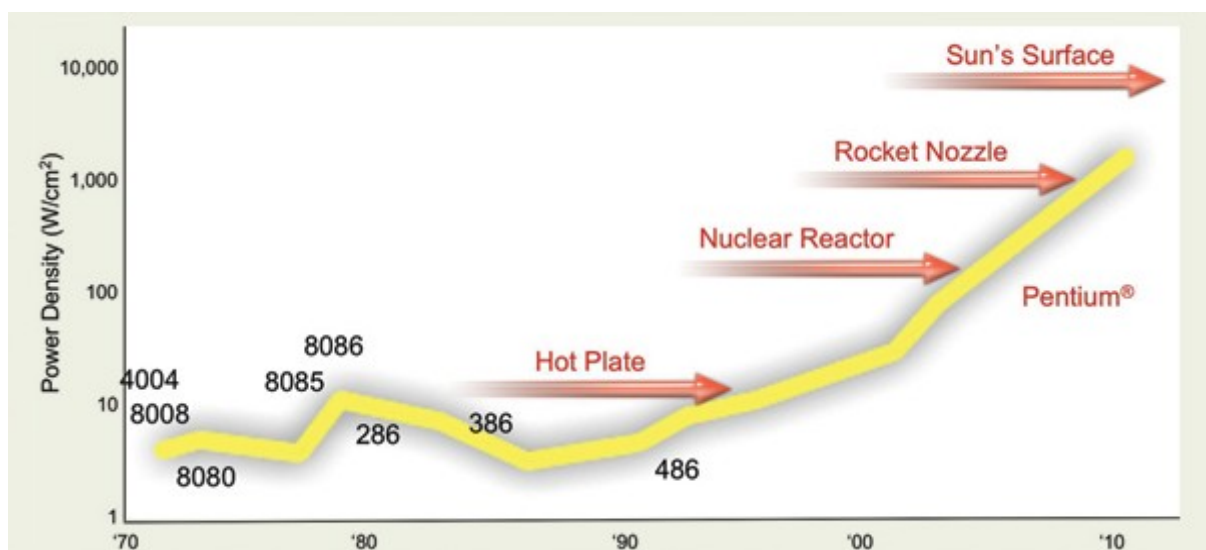


Fig 1.5.1: La densità di potenza nelle cpu moderne

è in continua crescita e oggi giorno è paragonabile addirittura a quella di un reattore nucleare.[5]

Esiste poi un problema di ritardi legato alle capacità di interconnessione. Storicamente il ritardo introdotto da una porta era direttamente proporzionale alle capacità di gate e non si teneva conto delle capacità che le connessioni tra i MOS introducevano. Con la miniaturizzazione dei circuiti però ci si trova in una situazione in cui nel singolo chip esistono molti più dispositivi e le connessioni tra essi sono naturalmente in numero maggiore. Il segnale, per propagarsi all'interno del chip, deve quindi attraversare molte connessioni e quindi le capacità da esse introdotte non possono più considerarsi trascurabili. In pratica non esiste più solo il contributo delle capacità di gate, ma ad esse devo sommare quelle di interconnessione, aumentando di fatto i tempi di propagazione e quindi abbassando le prestazioni.

1.6. I limiti della legge di Moore

Tenendo presenti le difficoltà di miniaturizzazione dei MOSFET viene ora spontaneo chiedersi fino a quando l'evoluzione esponenziale della microelettronica può proseguire e quali saranno i limiti che gli sviluppatori di circuiti integrati incontreranno o quali sono quelli che stanno già incontrando.

In un'analisi generale tali limiti si possono suddividere principalmente in quattro categorie:

- fisici,
- tecnologici,
- economici e
- affidabilità dei dispositivi.

Per quanto riguarda i limiti fisici ci si riferisce al fatto che i transistor non possono essere scalati sotto una certa dimensione, sia perché la lunghezza del canale non può essere troppo piccola, sia perché con le attuali tecniche di progettazione si andrebbe incontro all'impossibilità tecnologica di definire geometrie al di sotto di una certa dimensione. Esistono però delle tesi che dimostrano che tali limiti sono ancora lontani dall'essere raggiunti e quindi per almeno altre 2 o 3 generazioni tale crescita non si fermerà.

Diversa è invece la situazione dal punto di vista dei limiti tecnologici. Allo stato attuale risulta difficile pensare che molte aziende saranno in grado di sviluppare le nuove tecnologie dato che, la realizzazione delle stesse, dipende da molti fattori quali le attrezzature per la fabbricazione, materiali di estrema purezza o sistemi di misura in grado di garantire errori sotto il nanometro. Molte aziende del campo della microelettronica si sono allora attrezzate per formare un gruppo

di ricerca chiamato International Technology Roadmap for Semiconductors (ITRS) volto ad individuare i cosiddetti “colli di bottiglia”, ossia i punti deboli della produzione, e a fare delle previsioni sull'andamento dello sviluppo tecnologico. Hanno cioè formato una sorta di roadmap, ossia un percorso guidato globale che mira ad avere delle determinate tecnologie entro una data prefissata. Il compito dei ricercatori affiliati a tali aziende è quindi quello di trovare delle soluzioni ai problemi descritti in precedenza.

Il terzo limite incontrato come detto è quello economico. Da qui deriva una seconda legge di Moore che afferma che anche l'investimento necessario per sviluppare una nuova tecnologia, cresce in maniera esponenziale col tempo come mostrato in Fig 1.6.1. Stime fatte da esperti affermano che entro qualche anno la spesa per sviluppare una nuova tecnologia arriverà addirittura a toccare i 5 miliardi di dollari, il che fa presupporre che solo poche aziende potranno permettersi di sviluppare tali linee pilota così avanzate e che i guadagni, vista l'enorme spesa da affrontare, saranno notevolmente ribassati rispetto agli anni passati. Inoltre ci sarà una crescita esponenziale dei rischi connessi ad un investimento sbagliato, che colpirà soprattutto le società che sviluppano le attrezzature di produzione nel settore della microelettronica. In generale, si sta quindi assistendo ad un fenomeno di netto consolidamento del settore, con alte barriere di ingresso ed una forte riduzione della propensione a correre rischi.

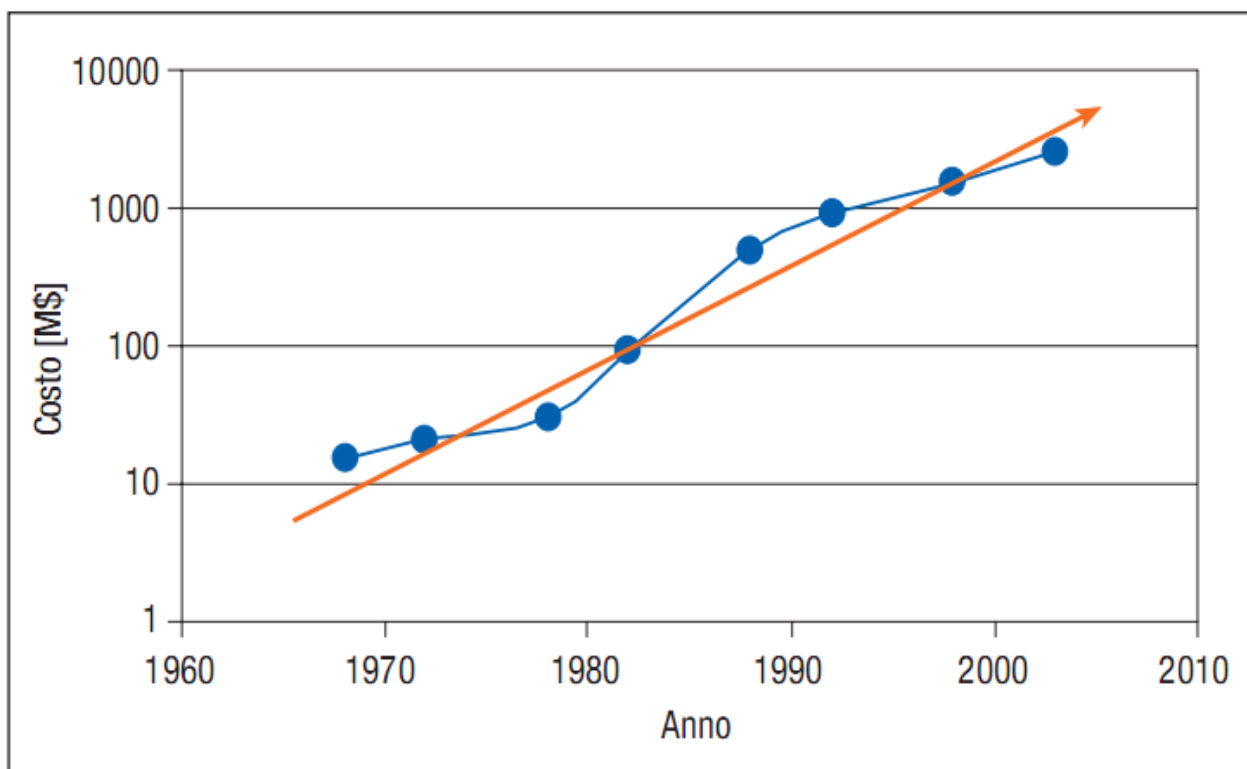


Fig 1.6.1: Andamento della seconda legge di Moore.[6]

Un altro fattore limitante è dato dalla naturale necessità di garantire un ritorno economico adeguato per gli investimenti fatti. Ogni nuova generazione tecnologica deve produrre abbastanza utili tali da ripagare le spese di sviluppo, e questo, viste le ingenti spese previste, sarà possibile solo se si apriranno dei nuovi mercati di massa. Potrebbe però arrivare un momento in cui non esisteranno più applicazioni di massa tali da giustificare economicamente lo sviluppo di tecnologie a prestazioni superiori.

Per quanto riguarda l'affidabilità, la miniaturizzazione dei dispositivi porta al verificarsi di fenomeni che degradano le prestazioni degli stessi. Tali fenomeni sono, ad esempio, la rottura dell'ossido di gate (breakdown), l'intrappolamento di carica (trapping), la generazione di portatori caldi o il verificarsi di correnti di leakage. Generalmente tutti questi fenomeni portano ad un degrado della

mobilità, un aumento della corrente di substrato e ad un aumento della tensione di soglia e, quindi, le prestazioni risultano notevolmente degradate. Tutto ciò porta inoltre ad una riduzione del tempo di vita del MOSFET e quindi del corretto funzionamento del circuito.

Capitolo 2

Come descritto in precedenza, le principali aziende produttrici di microprocessori, con il passare degli anni si sono trovate di fronte problemi sulla miniaturizzazione dei MOSFET che Moore non aveva considerato nella stesura della sua legge nel 1975. Tali aziende, come TSMC, motorola, Intel, ecc.. si sono allora attrezzate, o si stanno attrezzando, nella ricerca di nuove tecnologie, atte a far proseguire l'andamento che l'evoluzione della microelettronica ha avuto negli ultimi anni. Vediamo allora in dettaglio quali sono queste tecnologie.

2.1. Interconnessioni

Finora si è sempre parlato della riduzione delle dimensioni del MOSFET per incrementare le prestazioni dei circuiti integrati. In realtà con l'aumentare del processo di integrazione anche le interconnessioni tra i dispositivi sono aumentate notevolmente e questo ha influito pesantemente sulle prestazioni dei chip. Linee di interconnessione più strette comportano un aumento della resistenza R e l'avvicinamento delle stesse genera ulteriori capacità spurie C

che vanno ad incrementare il ritardo, dettato dal prodotto RC, che le interconnessioni introducono. Per far fronte al nuovo problema IBM nel 1997 introdusse per prima le interconnessioni in rame, al posto dell'alluminio, per ridurre la resistenza R che le linee stesse presentavano. Il rame offre infatti una resistenza specifica normalizzata inferiore a 2 Ohm per centimetro mentre l'alluminio, una resistenza superiore ai 3 Ohm per centimetro. Il passaggio al nuovo materiale comportò diverse difficoltà in quanto esso, così come l'oro, altera le proprietà del silicio. Per ovviare a tale problema IBM dovette inventarsi un procedimento di mascheramento del rame, com'è possibile vedere in Fig 2.1.1, dalla quale si evince che il rame è tenuto separato dall'ossido di silicio per mezzo di barriere di diffusione al tantalio.

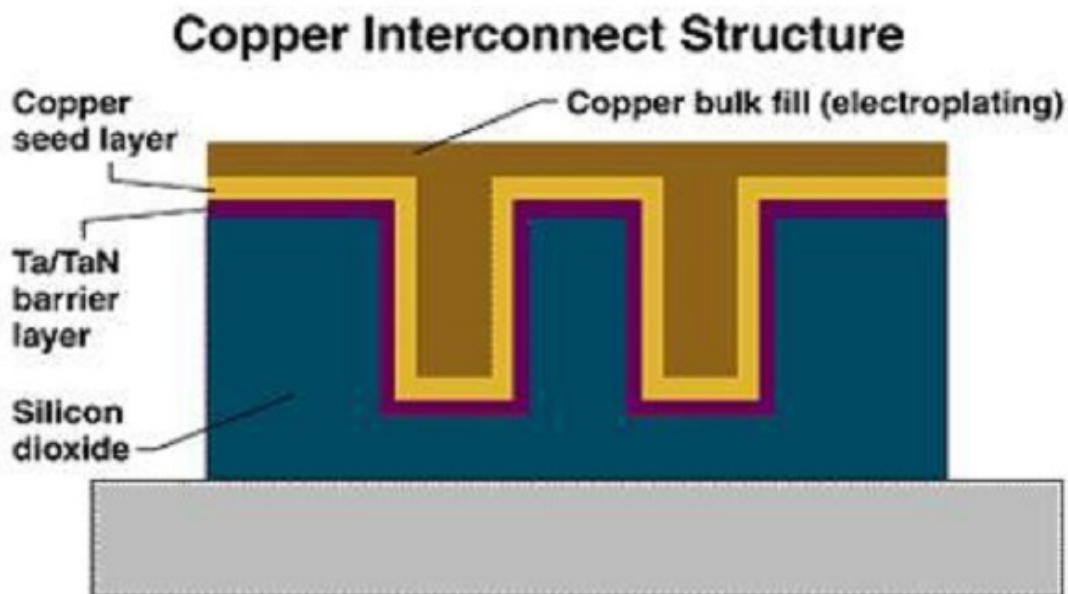


Fig 2.1.1: Con le interconnessioni in rame si deve introdurre un nuovo strato per evitare il contatto rame-silicio[7]

L'utilizzo del rame ha consentito di diminuire il numero di livelli di metallo, ovvero di utilizzare linee più sottili a parità di densità di corrente trasportata.

Da un altro punto di vista, ciò ha significato la riduzione dei consumi di circa il 30%. La tecnologia low-k è stata invece introdotta per ridurre la capacità C. Come si può vedere dalla figura 2.1.2, con il diminuire della distanza tra le interconnessioni sullo stesso livello, le capacità spurie crescono in maniera esponenziale. Come si può notare, utilizzando materiali aventi costante dielettrica k bassa rispetto all' SiO₂, le capacità spurie diminuiscono notevolmente e con esse diminuisce anche il ritardo RC.

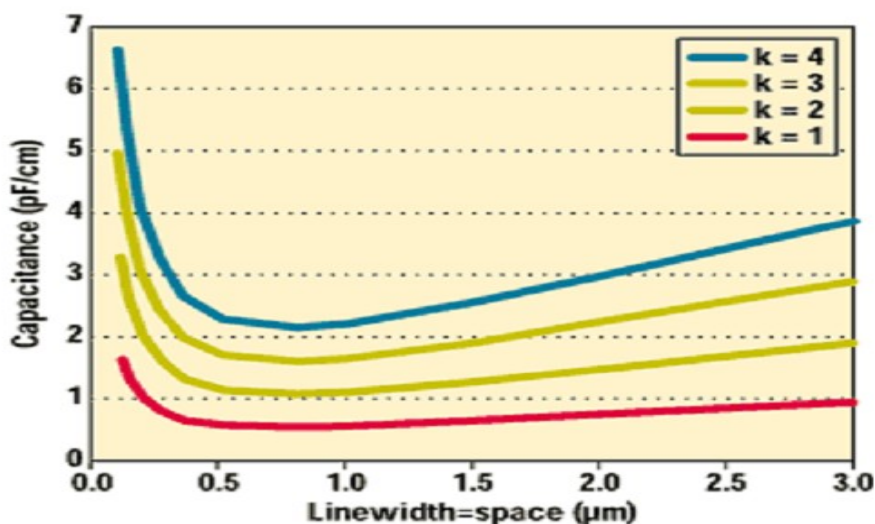


Fig 2.1.2: al decrescere della distanza tra le interconnessioni la capacità da esse introdotta aumenta notevolmente e l'utilizzo di materiali low-k permette di risolvere il problema [8]

Come mostrato da Fig 2.1.3 quando la distanza tra le interconnessioni diventa minore di 300 nm la capacità interstrato è quella dominante e rappresenta circa il 90% della capacità C totale. Vediamo che con il sistema tradizionale (Al/SiO₂) il ritardo in tecnologia 130 nm raggiunge 25 ps. Con l'adozione del sistema Cu/low-k si scende a 10 ps, il che equivale a dire che si guadagna un 60% in termini di prestazioni. I materiali low-k, oltre ad avere una bassa

costante dielettrica, devono anche soddisfare ad altre specifiche come l'alta resistenza meccanica, la stabilit  nelle dimensioni, l'alta stabilit  termica e presentare delle basse correnti di perdita. Attualmente i materiali utilizzati sono la silice nanoporosa, l'HSQ (usato dalla Texas Instruments) e il teflon-AF.

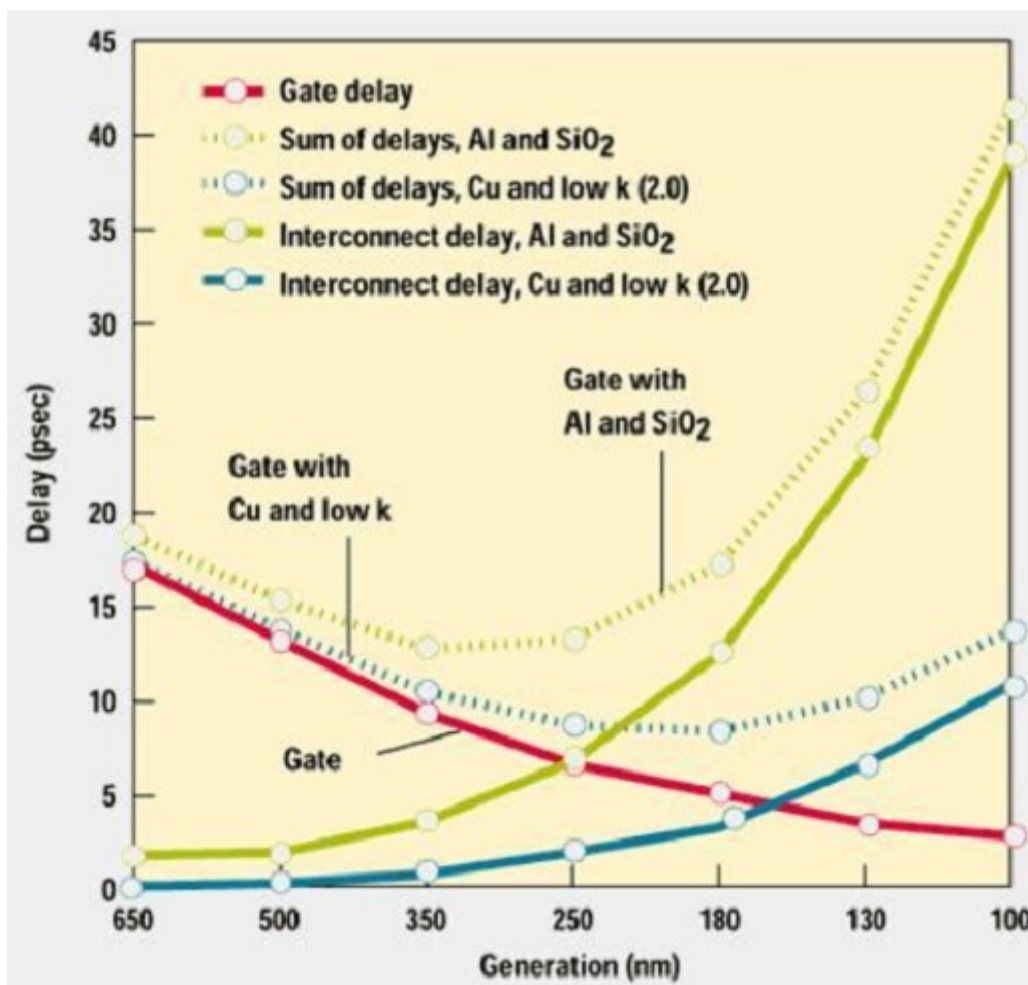


Fig 2.1.3: Attraverso l'utilizzo di piste in Cu/low-k si possono ridurre notevolmente i ritardi introdotti.[9]

Riassumendo possiamo quindi dire che l'abbassamento della resistenza delle interconnessioni attraverso l'uso del rame e, soprattutto, l'uso di materiali low-k per la riduzione delle capacit  parassite hanno migliorato in modo notevole le

prestazioni dei chip senza dover scalare ulteriormente le dimensioni.

2.2. Strained silicon

Dal 1987 IBM si è concentrata nello studio e nella produzione di integrati basati sulle proprietà del Silicio-Germanio (SiGe) tra cui spiccano l'elevata mobilità elettronica, facilità di integrazione e la possibilità di costruire dispositivi analogici come induttori, condensatori o sense amplifiers e di impacchettarli accanto a quelli digitali. Fino a qualche anno fa si accostava il SiGe a delle applicazioni di tipo analogico, ma dal 2001 Intel ha trovato molto interessante la possibilità di utilizzare il nuovo materiale per la produzione di circuiti integrati in modo da migliorarne le prestazioni e abbassarne i consumi. La tecnica conosciuta come "strained-silicon" (letteralmente silicio stirato) consiste nel depositare uno strato di silicio germanio sulla cima del bulk o corpo del wafer di silicio. Lo spessore di questo strato è di 2 μm e la concentrazione di Germanio è del 20%. La concentrazione di atomi di germanio non è uniforme, ma presenta una maggiore concentrazione di atomi dello stesso sulla cima. A questo punto uno strato di silicio, dello spessore di circa 20 nm, viene depositato sopra lo strato di Si-Ge. In questo modo gli atomi di silicio dello strato sovrastante tendono ad allinearsi con quelli dello strato di Silicio-Germano che, essendo più spesso, obbliga gli atomi di silicio a spaziarsi di una distanza analoga a quella degli atomi di Silicio-Germanio, come mostrato dalla Fig 2.2.1.

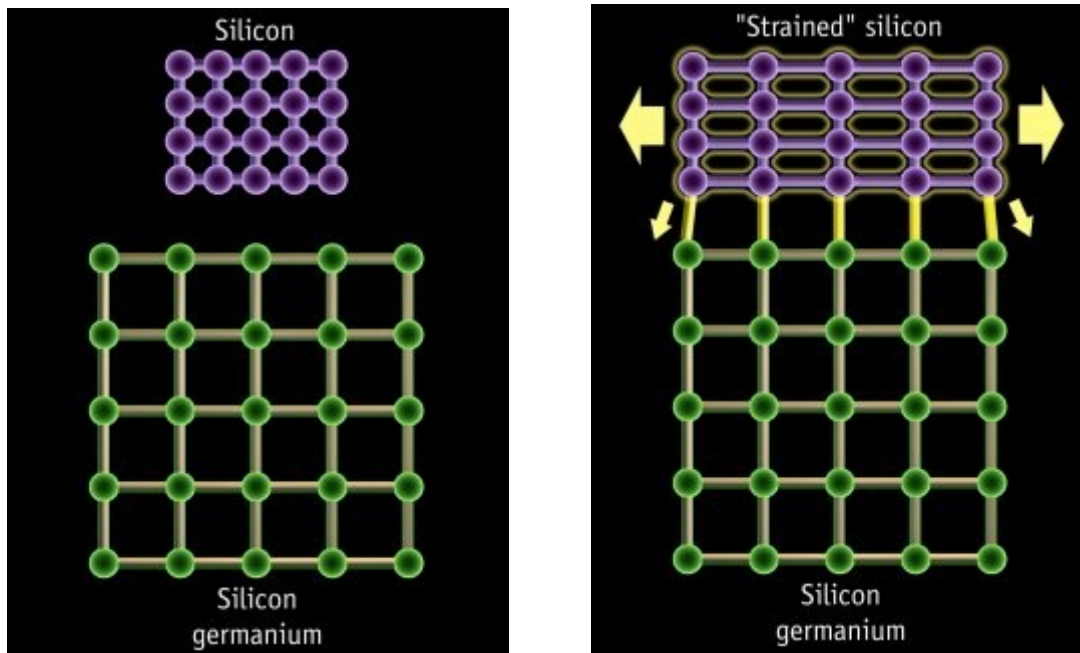


Fig 2.2.1: Quando lo strato di silicio viene depositato sopra quello di SiGe esso tende ad allinearsi con quest'ultimo allungandosi di fatto in direzione verticale ed orizzontale.[10]

Così facendo il reticolo cristallino del Silicio viene allungato di circa l'1% sia in direzione laterale che verticale e ciò permette un enorme incremento sulla mobilità dei portatori di carica, che incontrano una resistenza inferiore al loro passaggio e fluiscono fino al 70% più velocemente, cosa che rende più veloci i chip di circa il 30% senza bisogno di ulteriori miniaturizzazioni ed inoltre abbassa i consumi. Le prestazioni vengono ulteriormente migliorate attraverso l'utilizzo combinato della tecnologia Silicon On Insulator (SOI) e dello strained-silicon. In questo modo attraverso la SOI vengono notevolmente ridotte le capacità parassite, mentre lo strained silicon può aumentare del 50% la mobilità delle lacune e addirittura dell'85% quella degli elettroni rispetto alle normali condizioni.

Come mostra la Tabella 2.2.2, esistono però altri elementi semiconduttori (appartenenti al III e V gruppo della tavola periodica) che permettono una mobilità di elettroni e lacune superiore e quella del Si.

	Si	Ge	GaAs	InP	InAs	InSb
electron mob. (cm ² /Vs)	1600	3900	9200	5400	40000	77000
electron effective mass (/m ₀)	m _i : 0.19 m _i : 0.916	m _i : 0.082 m _i : 1.467	0.067	0.082	0.023	0.014
hole mob. (cm ² /Vs)	430	1900	400	200	500	850
Hole effective mass (/m ₀)	m _{HH} : 0.49 m _{LH} : 0.16	m _{HH} : 0.28 m _{LH} : 0.044	m _{HH} : 0.45 m _{LH} : 0.082	m _{HH} : 0.45 m _{LH} : 0.12	m _{HH} : 0.57 m _{LH} : 0.35	m _{HH} : 0.44 m _{LH} : 0.016
band gap (eV)	1.12	0.66	1.42	1.34	0.36	0.17
permittivity	11.8	16	12	12.6	14.8	17

Fig 2.2.2: Valori della mobilità, massa effettiva, permittività e band-gap di elettroni e lacune nel silicio e in altri semiconduttori [10]

E' importante osservare che l'InP e il GaAs presentano un band-gap elevato rispetto al Si e al Ge. Un elevato band-gap permette di diminuire la corrente di leakage e questo li rende particolarmente adatti ad essere utilizzati in applicazioni a bassa potenza. Osservando i dati relativi alla mobilità delle lacune, si può notare che il germanio è l'elemento che presenta il valore più alto. Questo lo rende particolarmente adatto ad essere utilizzato in dispositivi p-MOS ad alta velocità. D'altro canto però, lo stesso germanio, presenta un valore di band-gap basso rispetto agli altri elementi e questo, come visto in precedenza, aumenta la potenza statica dissipata, rendendolo inadatto per applicazioni a basso consumo di potenza. Per poter formare dei dispositivi basati sulla

tecnologia CMOS al germanio bisogna quindi risolvere alcuni problemi come la riduzione della corrente di leakage, il miglioramento delle prestazioni per quanto riguarda gli n-MOS, la formazione di giunzioni di source e drain che presentino una bassa resistività e prevedere delle strutture che aumentino l'immunità agli effetti di canale corto. Per risolvere i problemi legati all'immunità agli effetti di canale corto e alla corrente di leakage, si ricorre a strutture del tipo Si/Ge/Si. Per far questo si sfrutta il processo di condensazione del germanio, una tecnica relativamente semplice che sfrutta l'ossidazione ad alta temperatura. Partendo da un substrato di SOI standard con uno strato di SiGe posto sullo strato di Si del SOI, si procede all'ossidazione ad alta temperatura (come mostrato in Fig 2.2.3). Questo processo provoca la formazione di SiO₂ e gli atomi di Ge si diffondono nel substrato sottostante. Lo strato di GOI (German On Insulator) che si viene a creare può essere controllato attraverso lo spessore di SOI e di SiGe iniziale. Come si può notare in Fig 2.2.4 questo processo permette la creazione di substrati di SGOI particolarmente uniformi e, quindi, adatti ad essere utilizzati. Si può inoltre osservare che, la maggior parte degli atomi di germanio inizialmente presenti nello strato di SiGe, sono stati preservati e, la residua concentrazione di atomi di Si nello strato di GOI, pari allo 0,01%, conferma la qualità del processo. Lo strato di SiO₂ sarà utilizzato come ossido di gate nel dispositivo finale che presenterà prestazioni superiori rispetto ai normali dispositivi SOI in silicio soprattutto per quanto riguarda i p-MOS. Si può quindi pensare ad un nuovo tipo di tecnologia CMOS in cui, i p-MOS sono ottenuti mediante la tecnica appena descritta, e gli n-MOS sono ottenuti mediante tecniche standard basate sul silicio o attraverso l'impiego di altri materiali, in grado di migliorare la mobilità degli elettroni, e

quindi di incrementare le prestazioni degli n-MOS.

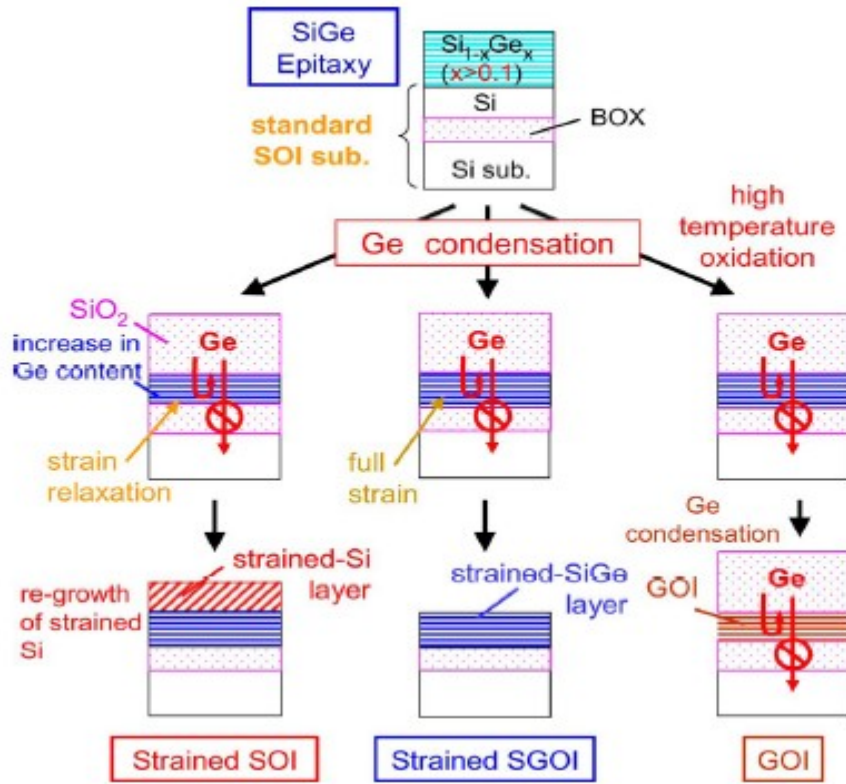


Fig 2.2.3: Schema di principio della condensazione del germanio[11]

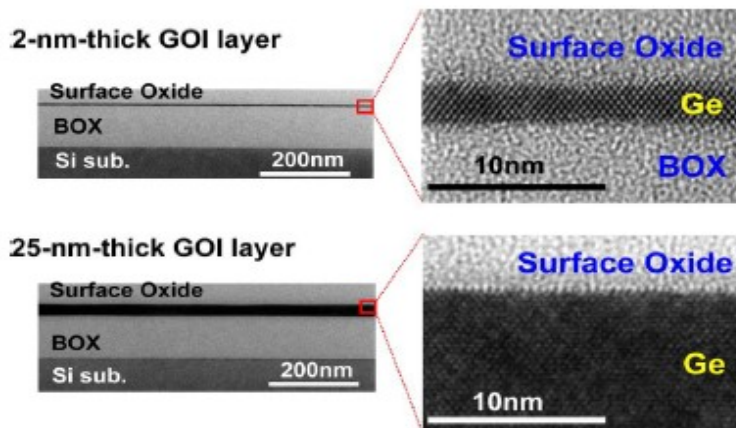


Fig 2.2.4: Foto di 2 strati GOI aventi spessore di 2 e 25 nm[12]

Per far questo l'attenzione è stata rivolta agli elementi del III e IV gruppo

presentati nella Tabella 2.2.2 dato che, l'elevata mobilità degli elettroni e la loro massa effettiva, può portare ad importanti miglioramenti degli n-MOS. Anche i dispositivi che utilizzano questi nuovi materiali devono soddisfare delle condizioni specifiche, ed in particolare i nuovi n-MOS devono poter essere formati su substrati di silicio e devono mantenere una buona immunità agli effetti di canale corto. Il principale problema introdotto dai materiali del III e V è legato al basso valore presentato dalla massa effettiva degli elettroni. A causa di ciò la mobilità viene degradata e di conseguenza sono degradate anche le prestazioni. In generale, comunque, gli studi fatti finora mostrano significativi aumenti in termini di prestazioni attraverso l'utilizzo di elementi del III e V gruppo (ed in particolare attraverso l'utilizzo del GaAs) per quanto riguarda gli n-MOS e attraverso l'utilizzo del germanio descritto in precedenza per i p-MOS.

2.3. Sviluppo e utilizzo di dielettrici high-k

Con il termine materiali high-k si intendono materiali aventi una costante dielettrica superiore a quella dell' SiO_2 (biossido di silicio) che è stato da sempre il materiale più usato per la formazione dell'ossido di gate. Sviluppando dispositivi sempre più piccoli i produttori hanno dovuto far fronte anche alla diminuzione dello spessore dell'ossido di gate e quando sono arrivati al nodo tecnologico di 90 nm lo spessore dello strato di SiO_2 era arrivato a toccare 1.4 nm (si era partiti da 100 nm). Quando lo spessore dell'ossido risulta minore di 2 nm le correnti di leakage che attraversano lo strato di ossido non sono più trascurabili e il consumo di potenza statica diventa molto più elevato (come

mostrato dalla figura 2.3.1).

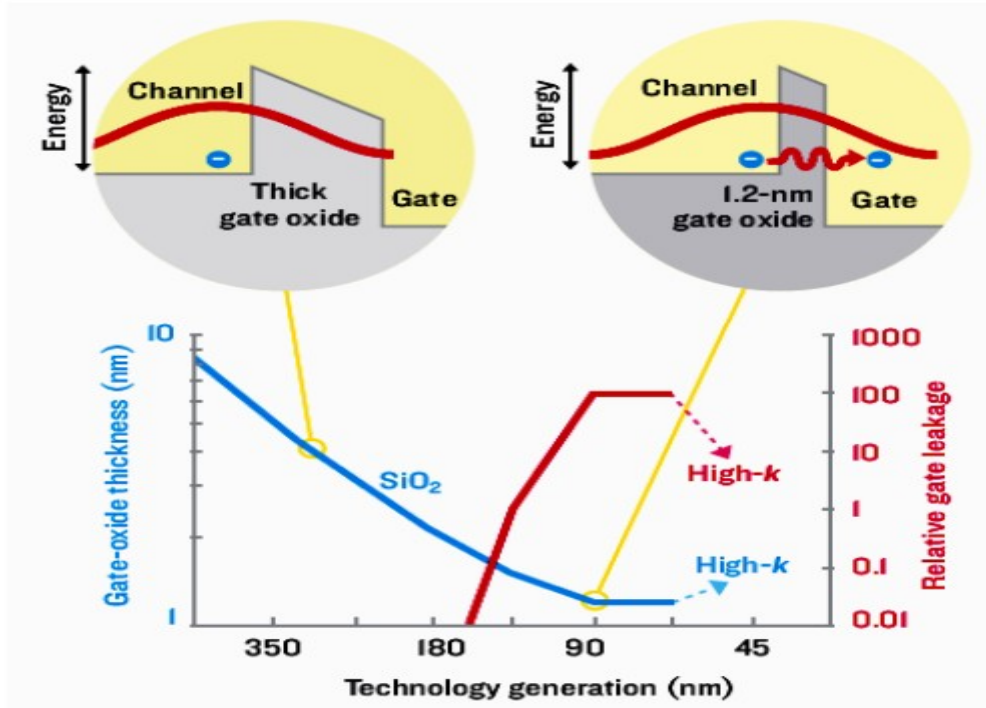


Fig 2.3.1: Andamento della corrente di leakage in funzione della diminuzione dello spessore dell'ossido[13]

Le aziende produttrici hanno quindi dovuto iniziare a pensare ad un nuovo materiale sostitutivo per poter continuare il trend dettato dalla legge di Moore. L'utilizzo di un materiale avente una costante dielettrica k più elevata permette di raggiungere campi elettrici sufficientemente elevati aumentando lo spessore dell'ossido e quindi evitando di innalzare le correnti di dispersione. Per poter confrontare le caratteristiche elettriche di un materiale high- k con quelle del biossido di silicio occorre introdurre il concetto di spessore d'ossido equivalente (EOT Equivalent-Oxide-Thickness) così definito:

$$EOT = \frac{\epsilon_{OX}}{\epsilon_{HK}} t_{HK} \quad C_{HK} = \frac{\epsilon_{HK}}{t_{HK}} = \frac{\epsilon_{OX}}{EOT} = C_{EQ}$$

dove $\epsilon_{ox}=3,9\epsilon_0$ ed è la costante dielettrica del biossido di silicio

$\epsilon_{HK}=k\epsilon_0$ ed è la costante dielettrica del materiale high-k

t_{HK} è lo spessore dell'high-k

Tale relazione dice che a parità di EOT anche la capacità è la stessa. Si può quindi affermare che a parità di capacità lo spessore dell'ossido formato da materiale high-k risulta essere più grande rispetto a quello formato da biossido di silicio come mostrato dalla figura 2.3.2.

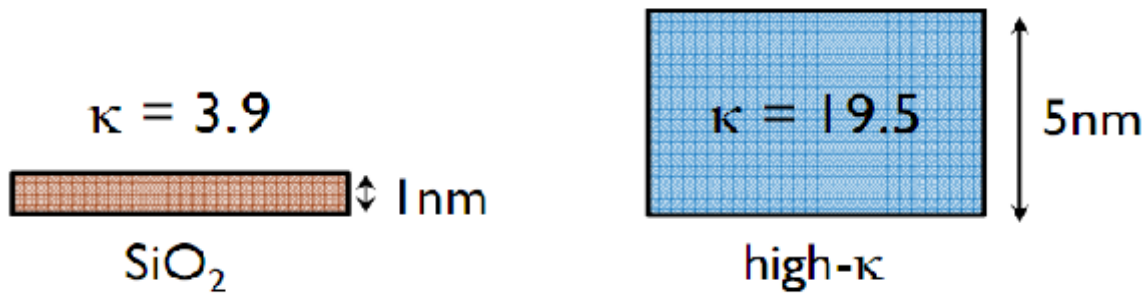


Fig 2.3.2: a parità di capacità il materiale high-k presenta uno spessore maggiore

Da un altro punto di vista si può quindi affermare che, a parità di spessore dell'ossido, attraverso l'utilizzo di dielettrici high-k la corrente di leakage viene fortemente ridotta, come mostra la Fig 2.3.3.

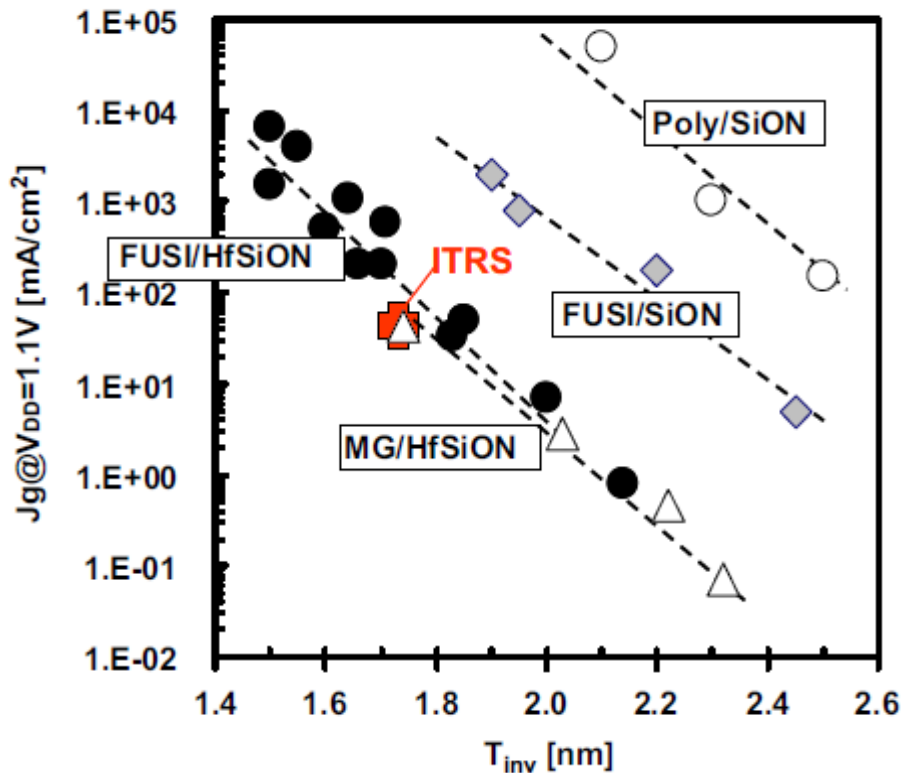
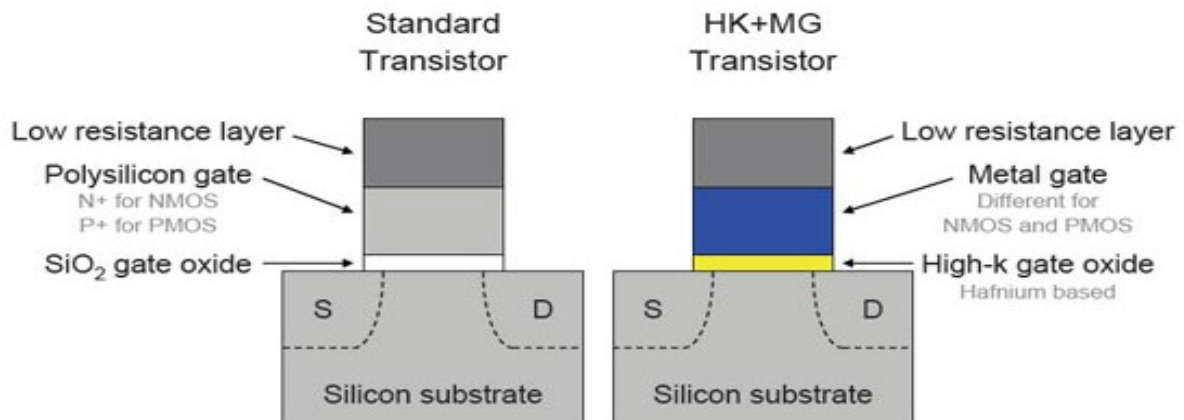


Fig 2.3.3: a parità di spessore dell'ossido di gate, attraverso l'utilizzo di dielettrici high-k, la corrente di leakage viene significativamente diminuita. [14]

Nel 2001 Intel iniziò a sviluppare MOSFET utilizzando il nuovo ossido ma i transistor ottenuti avevano caratteristiche elettriche pessime. I principali problemi erano legati alla difficoltà di accensione dei MOSFET e alla loro lentezza causata dalla scarsa mobilità e soprattutto dalle difficoltà di interazione tra il polisilicio di gate e il materiale high-k. Il successivo passo è stato quindi quello di trovare un modo per eliminare questi problemi. La soluzione trovata è stata quella di sostituire il gate in polisilicio con uno in materiale metallico (ritornando così al vero MOS). La Fig 2.3.4 mostra un transistor “classico” e uno “high-k”, così da evidenziare le differenze.

High-k + Metal Gate Transistors



r

Fig 2.3.4: MOS standard e high-k messi a confronto.[15]

I prototipi dei nuovi MOSFET furono realizzati nel 2003 e presentavano una soluzione leggermente diversa da come si era ipotizzato, dato che essi oltre al metal gate e al materiale high-k presentavano anche un piccolo strato di SiO₂ tra l'high-k e il substrato così da poter migliorare la qualità elettrica di tale interfaccia. La prima applicazione che sfruttava tale tecnologia fu presentata da Intel nel 2007: un processore a 45 nm con high-k (il materiale scelto è stato l'HfO₂) e metal gate in grado di essere 25% più veloce della precedente tecnologia a 65 nm a parità di leakage e di ridurre il leakage di un fattore 5 a parità di velocità.

2.4. Tecnologia SOI

La tecnologia SOI (Silicon On Insulator) è una tecnologia ideata e brevettata da IBM che permette, a parità di dimensioni, di aumentare del 25% la velocità di funzionamento e di ridurre del 30% i consumi. Come dice il nome stesso, i dispositivi basati su tale tecnologia presentano uno strato di ossido sepolto chiamato BOX (Buried OXide), come si vede nella rappresentazione schematica della figura 2.4.1. Attualmente molte fonderie hanno intrapreso questa strada: gli ultimi processori di AMD (Advanced Micro Devices), la serie Phenom, sono costruiti interamente in tecnologia SOI. Quest'ultima, appunto, permette di realizzare transistor caratterizzati da una buona immunità agli effetti di canale corto, soprattutto quando lo strato di silicio, o body, viene assottigliato, riducendo in questo modo anche le capacità di giunzione. Il problema principale, dovuto alla riduzione di spessore di SOI, è rappresentato dall'aumentato valore delle resistenze serie parassite di source e drain. Questo costituisce un nodo fondamentale dal quale dipende lo sviluppo della tecnologia di cui stiamo trattando. In base allo spessore di body, i dispositivi SOI vengono suddivisi in PD-SOI (Partially Depleted mostrato in Fig 2.4.2) e in FD-SOI (Fully-Depleted mostrato in Fig 2.4.3): se lo spessore è ridotto, tipicamente sotto i 100 nm, la zona di svuotamento si estenderà a tutto il film di silicio. Al contrario, se è maggiore di 100 nm, una parte di layer non sarà completamente svuotata di portatori maggioritari.

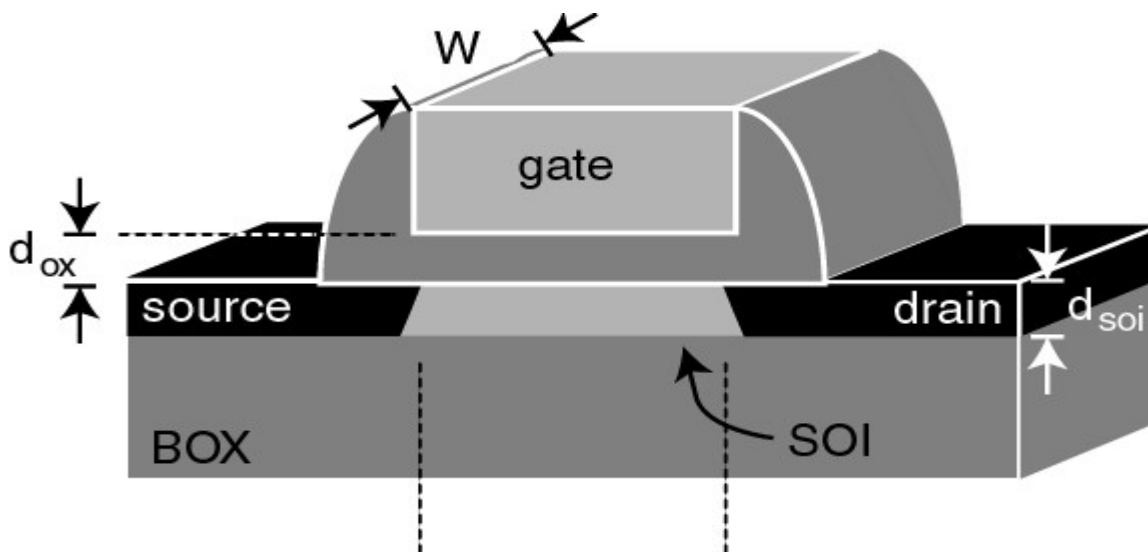


Figura 2.4.1: Modello di un dispositivo SOI-MOSFET nel quale vengono evidenziate le regioni fondamentali.[15]

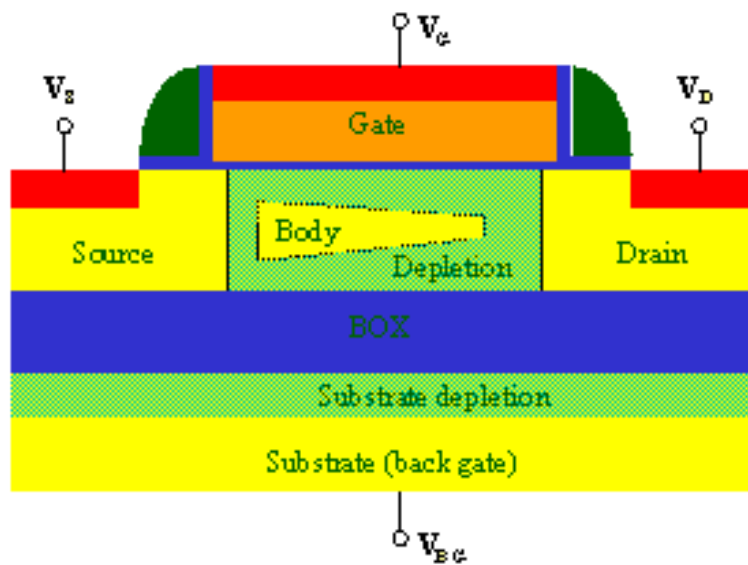


Fig 2.4.2: Struttura di un PD-SOI MOSFET

Evidentemente i due tipi di SOI presenteranno caratteristiche diverse. Il problema principale presentato dai PD-SOI, è da considerarsi il potenziale di body flottante che, causando una variazione di tensione di soglia, da origine ad alcuni fenomeni parassiti che riducono le prestazioni del dispositivo, come l'effetto kink, l'attivazione del BJT parassita, il secondo picco della

transconduttanza.

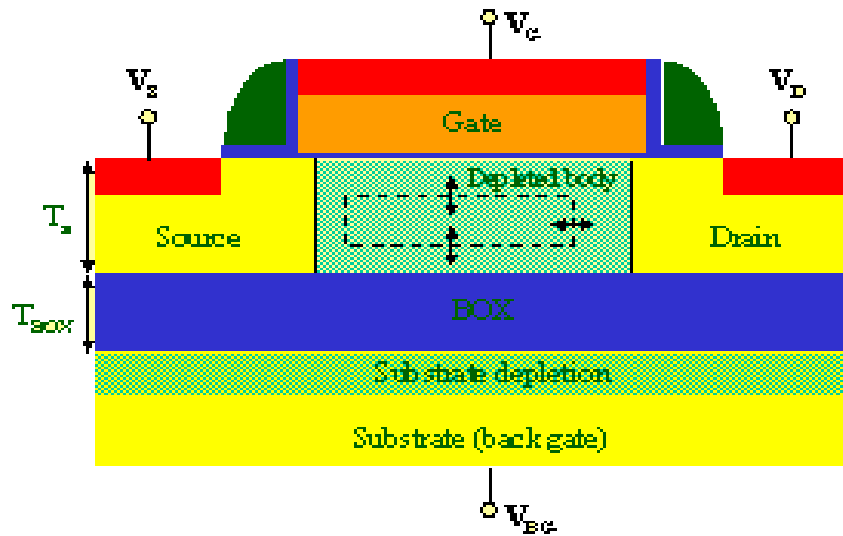


Fig 2.4.3: Struttura di un FD-SOI MOSFET

L'effetto kink è un aumento della corrente di saturazione dovuto alla generazione di coppie elettrone-lacuna indotta dagli elevati campi elettrici vicino alla giunzione di drain. Ciò si verifica per valori di V_{ds} elevati (come mostrato in Fig 2.4.4) e degrada di fatto le prestazioni del dispositivo. L'attivazione del BJT parassita si verifica quando il tempo di vita dei portatori di carica nel body è elevato. Anche questo provoca un aumento della corrente I_{ds} degradando le prestazioni del dispositivo.

Il secondo picco della transconduttanza è dovuto alla corrente di tunnel attraverso l'ossido di gate che polarizza positivamente il substrato. Questo provoca un aumento anomalo e non controllabile della corrente I_{ds} (come mostrato in Fig 2.4.5) che può ridurre le prestazioni e l'affidabilità del dispositivo.

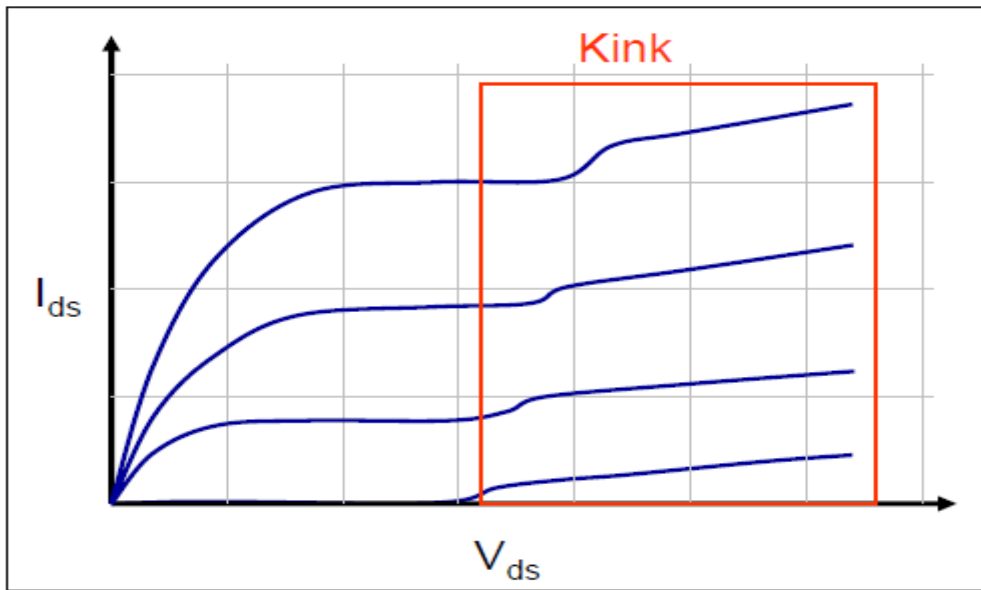


Fig 2.4.4: Aumento anomalo della corrente dovuto all'effetto "kink"[17]

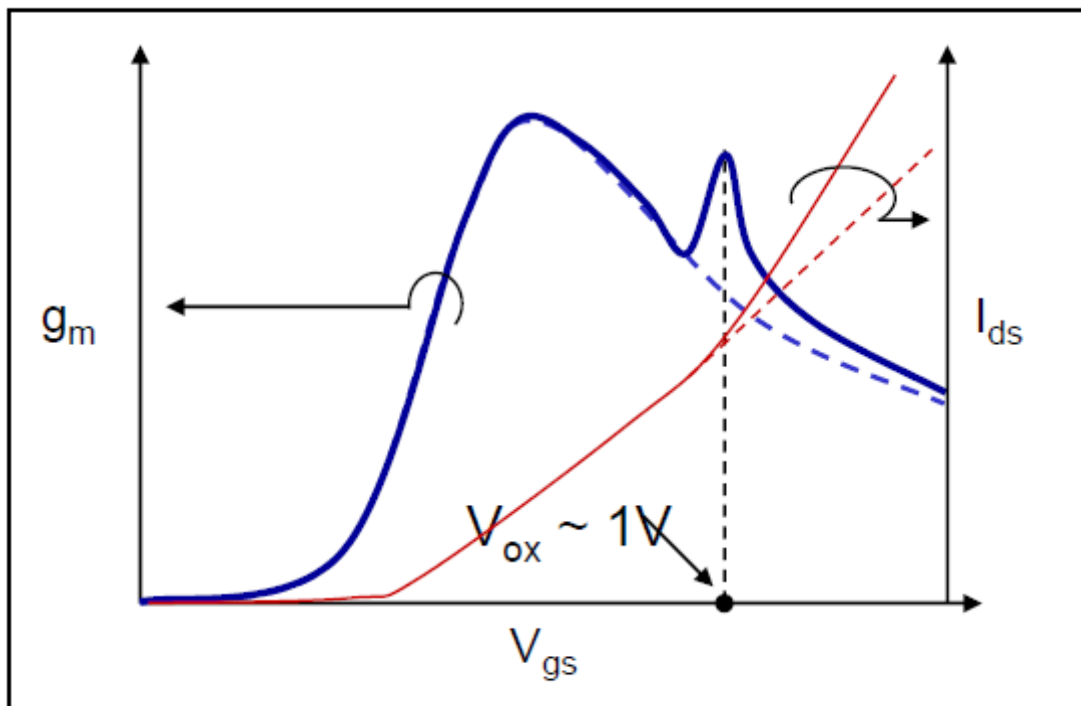


Fig 2.4.5: Il secondo picco della transconduttanza provoca un innalzamento della corrente I_{ds} [18]

Nei dispositivi FD-SOI lo strato di inversione sotto il canale si estende fino all'interfaccia con il BOX. Con una tale configurazione il film di silicio deve

essere molto sottile e va accuratamente progettato unitamente al drogaggio del body del transistor stesso affinché la condizione di completo svuotamento del film si verifichi effettivamente. Dovendo realizzare film di silicio di spessore così ridotto, il controllo dello spessore deve assicurare variazioni entro margini molto ristretti, tipicamente al di sotto del 10% dello spessore del film e questo comporta notevoli difficoltà nel processo di fabbricazione, come si vedrà nel seguito. Se lo spessore di body è appena più piccolo della zona di svuotamento, abbiamo una minore immunità agli effetti di canale corto, difficoltà nell'ottenere tensioni di soglia alte e inoltre si fanno sentire anche gli effetti parassiti di cui abbiamo parlato in precedenza a proposito dei PD-SOI. Si può affermare dunque, che i dispositivi FD-SOI presentano dei vantaggi quando lo spessore del layer di silicio diviene minore di 10 nm. Come detto vi sono differenze sostanziali anche nel processo di fabbricazione, il quale per i PD-SOI non presenta molti cambiamenti rispetto al processo standard mentre per i FD-SOI è costituito da un numero minore di passi, anche se, come detto, presenta maggiori difficoltà causate dal ridotto spessore di SOI. Tale processo necessita di una precisione estrema, dal momento che una disuniformità dello spessore del silicio incide pesantemente su molti parametri del dispositivo, il più importante dei quali è la tensione di soglia.

Possiamo affermare, dunque, che non esiste una soluzione migliore in assoluto, ma esistono soluzioni più o meno "ad hoc" a seconda delle esigenze. I criteri sui quali viene basata la scelta di un'alternativa piuttosto che un'altra, sono la velocità, la dissipazione di potenza e il costo. Sharp Electronics ad esempio, impegnata nelle comunicazioni a bassa potenza, ha scelto i FD-SOI, caratterizzati da un'ottima pendenza sottosoglia e quindi da limitate correnti di

perdita. D'altro canto IBM i cui sforzi si dirigono verso la realizzazione di microprocessori per servers e workstation, dove il consumo di potenza è di secondaria importanza rispetto alla velocità, si è indirizzata verso i PD-SOI, dove si raggiungono elevate correnti di conduzione e quindi elevate velocità. Abbiamo sin qui descritto i vantaggi della tecnologia SOI rispetto a quella classica, quando le dimensioni diventano nanometriche. Ci chiediamo allora, visto che l'innovazione non è così radicale, dato che è "bastato aggiungere un pò di ossido", il motivo della sua tardiva introduzione dato che l'idea di principio della SOI era nota ai progettisti di IBM da quasi trent'anni. La risposta sta nel fatto che, fino a pochi anni fa, i bulk soddisfacevano a pieno le esigenze e quindi gli investimenti venivano fatti in quella direzione: non si rischiava di intraprendere una via nuova, quando la vecchia dava ottimi risultati. Ancora oggi INTEL, leader mondiale nella produzione di microprocessori, continua a considerare la tecnologia classica la migliore e la più affidabile. E' molto difficile inoltre crescere silicio su un materiale diverso, visto che le proprietà cristalline, come la costante reticolare ad esempio, sono diverse.

2.5. Tecnologia FinFET

I MOSFET durante tutta loro evoluzione hanno sempre mantenuto una forma planare come quella descritta in precedenza. Negli ultimi anni però le aziende produttrici stanno riscontrando dei problemi (legati principalmente agli effetti di canale corto e alla dissipazione di potenza) che solo una riprogettazione fondamentale del transistor può risolvere. Qualche anno fa un transistor si

poteva pensare come un interruttore ideale (a meno di perdite trascurabili) in cui, almeno nello stato off, il dispositivo consumava una potenza praticamente nulla. Oggi invece, come già detto, con la diminuzione delle dimensioni e l'aumento del numero di transistor sullo stesso chip, le correnti di perdita non si possono più considerare trascurabili e anche quando il dispositivo è nello stato off dissipa potenza statica, riducendo di gran lunga l'autonomia dei dispositivi a batterie e provocando un riscaldamento generale del circuito. Risulta quindi evidente che, per poter arrivare a produrre transistor aventi una lunghezza di gate L_g particolarmente piccola (inferiore ai 20 nm), sia necessario ridurre quanto più possibile il consumo di potenza statica. Con la classica struttura Bulk-Si, per produrre transistor aventi una L_g molto piccola, servono alti drogaggi ($\geq 10^{18} \text{ cm}^{-3}$) per poter ridurre le correnti di perdita. Questo comporta vari svantaggi come la fluttuazione casuale del numero di droganti, provocando variazioni non controllabili della tensione di soglia e quindi aumenti indesiderati della corrente di leakage. Inoltre aumenta il numero di impurità ionizzate all'interno del silicio e questo va ad influire pesantemente sulla mobilità dei portatori di carica e quindi degrada le prestazioni del componente.

Una possibile soluzione a questo tipo di problema sta nell'utilizzo di transistor che possano diminuire le correnti di perdita senza variare il livello di drogaggio delle parti. Transistor aventi source, drain e canale sufficientemente sottili (come quello mostrato in Fig 2.4.1) permettono di controllare in modo efficiente le correnti di perdita senza il bisogno di aumentare il drogaggio.

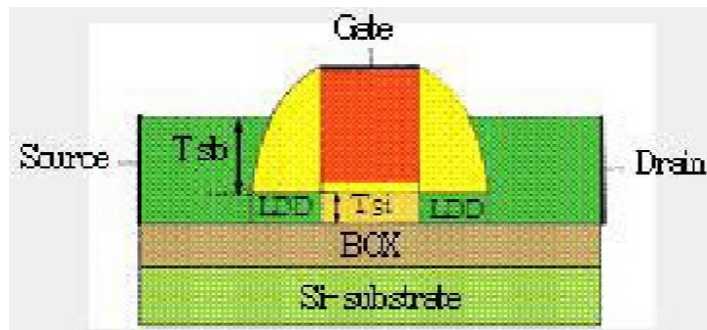


Fig 2.4.1: "thin-body" transistor

In particolare si può agire in modo che la percentuale di drogaggio non superi i 10^{15} cm^{-3} (ossia intrinseco) così da poter rendere immune il dispositivo alle variazioni della tensione di soglia.

Per poter migliorare ulteriormente le prestazioni del MOSFET un gruppo di ricerca della UC Berkeley hanno ideato il FinFET, ossia un transistor che presenta le caratteristiche di un "thin-body", ma avente il canale, il source e il drain innalzati rispetto al substrato (da qui il nome fin dato che il canale rialzato assomiglia ad una specie di pinna) e un gate fatto in modo che questo vada a coprire il canale su più superfici, come mostrato dalla Fig 2.4.2.

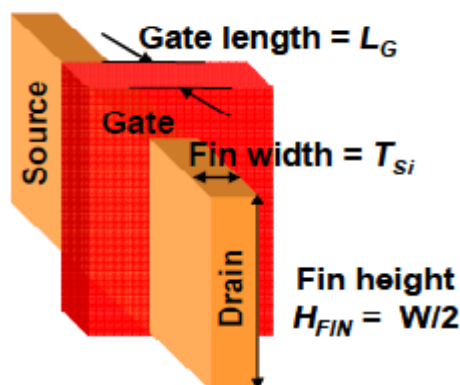


Fig 2.4.2: Struttura di base di un FinFET. In questo caso il gate copre il canale su 3 lati differenti.[19]

In un FinFET, come si può notare, esistono più parametri geometrici rispetto al MOSFET tradizionale. Tali parametri sono la lunghezza di gate L , l'altezza del fin H_{fin} e la larghezza del fin W_{fin} . Nei MOSFET tradizionali la dimensione minima è legata alla lunghezza di gate L . Nei FinFET invece la dimensione minima è legata a W_{fin} . Se $W_{fin} < 0.7 L$ gli effetti di canale corto e la corrente di leakage sono notevolmente ridotti. In particolare più piccola risulta W_{fin} e minore sarà la corrente di sottosoglia, come si può notare in Fig 2.4.3, tanto da poter assumere la pendenza di sottosoglia quasi ideale.

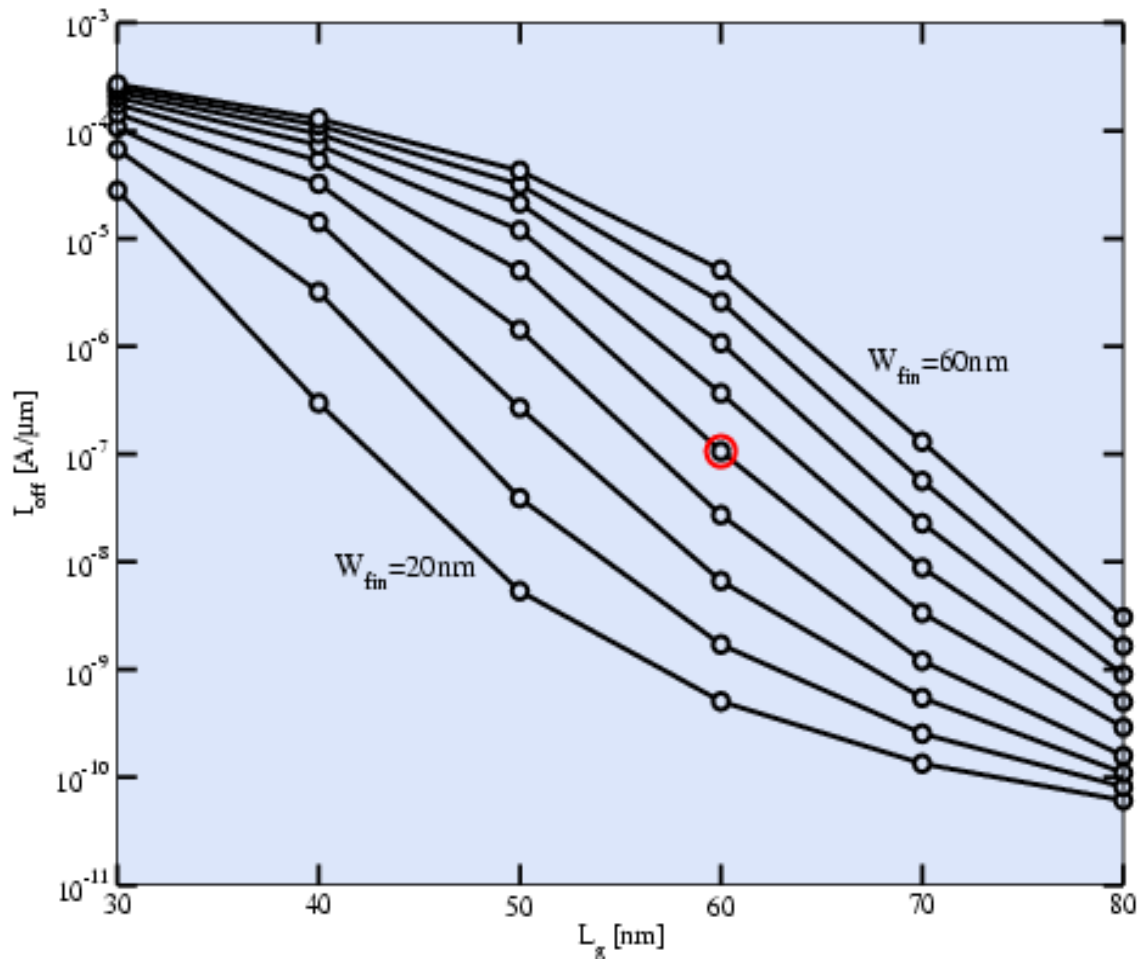


Fig 2.4.3: Diminuzione della corrente di sottosoglia al variare di W_{fin} e L_g [20]

Il principio di funzionamento del FinFET si basa sul fatto che la corrente può passare solo attraverso il canale rialzato e, quindi, gli elettroni non hanno più un percorso attraverso il quale si possono avere le correnti di perdita. Inoltre, essendo il gate avvolto attorno al canale, si può esercitare un maggior controllo sullo svuotamento e sul riempimento dello stesso. Lo svantaggio principale riscontrato in questo tipo di dispositivi è legato all'aumento del valore delle resistenze serie parassite di source e drain. Questo problema può però essere risolto attraverso l'utilizzo di particolari materiali per la formazione di S/D e con l'utilizzo di materiali aventi una bassa resistività ($\rho_c < 10^{-8} \Omega \text{ cm}^2$) per i contatti tra il S/D e il canale. Per rendere indipendente il controllo sul canale si può agire (attraverso delle maschere) sulla parte di gate sopra il canale rimuovendola, così da ottenere 2 gate separati per il controllo del canale (come mostrato dalla Fig 2.4.4) e gestire in modo più efficiente l'energia in gioco e i ritardi introdotti dal dispositivo.

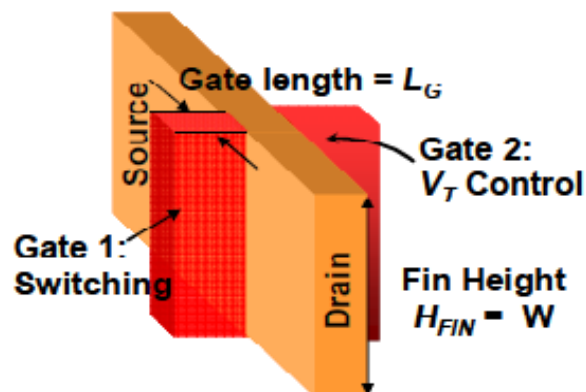


Fig 2.4.4: FinFET con doppio gate.[21]

Finora sono stati sperimentati 2 tipi di FinFET, SOI e bulk (mostrati in Fig 2.4.5 e 2.4.6).

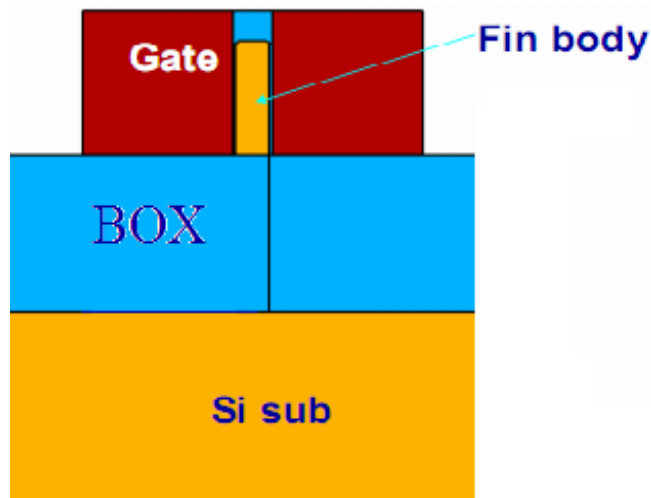


Fig 2.4.5: SOI-FinFET

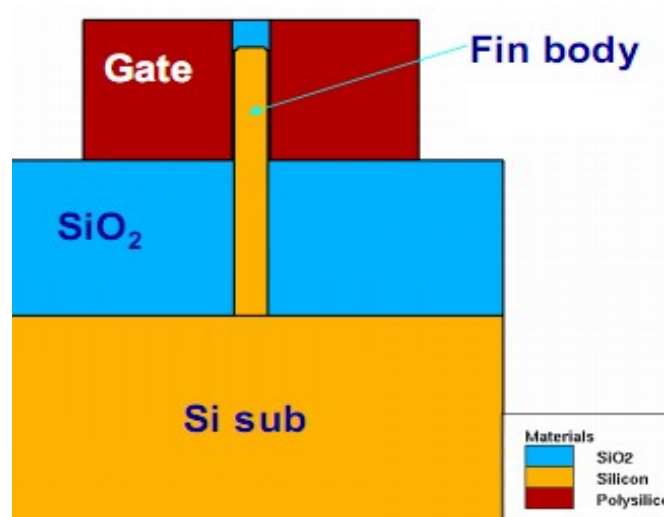


Fig 2.4.6: Bulk- FinFET

La differenza tra i due tipi di dispositivi, come si può notare, sta nella diversa formazione del fin. Per quanto riguarda le prestazioni, i test svolti dimostrano che non ci sono grandi differenze tra l'una e l'altra tipologia. I FinFET di tipo bulk richiedono, però, un alto livello di drogaggio per contenere la corrente di leakage (come mostrato in Fig 2.4.7). Ciò comporta una maggiore difficoltà nella fabbricazione, dato che il livello di drogaggio deve essere il più uniforme

possibile per poter evitare delle variazioni della tensione di soglia.

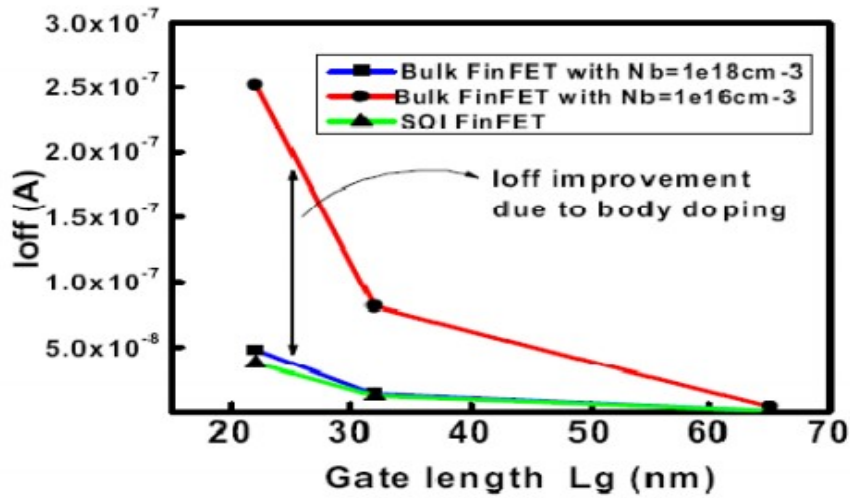


Fig 2.4.7: Per contenere la corrente di sottosoglia il livello di drogaggio nei Bulk-FinFET deve essere elevato[21]

Per produrre FinFET di tipo Bulk il substrato da utilizzare è più economico rispetto al substrato utilizzato nei SOI-FinFET ma servono un maggior numero di maschere ed un maggior numero di passi rispetto a quest'ultimi. In generale quindi, almeno per quanto riguarda le applicazioni digitali, SOI-FinFET e bulk-FinFET presentano prestazioni e costi molto simili.

I FinFET saranno utilizzati molto probabilmente dai produttori di celle SRAM dato che in esse i problemi legati alla dissipazione di potenza e alla stabilità delle celle stanno diventando così importanti da mettere in discussione la possibilità di produrre celle sotto agli attuali 45 nm. Nel campo delle memorie infatti, perdere un'eccessiva potenza può voler dire perdere inesorabilmente l'informazione contenuta nella cella. Attualmente si sta processando un array formato da 6 celle a FinFET e lo si sta confrontando a livello di prestazioni con lo stesso array sviluppato in tecnologia Bulk-Si. I risultati sono abbastanza sorprendenti dato che il margine di lettura alla stessa tensione di alimentazione (1 V) è di 175 mV contro i 135 mV e questo è dovuto

naturalmente alle minori correnti di leakage presenti.

2.6. Tecnologie non-CMOS

Tutte le iniziative considerate finora appartengono ad uno schema evolutivo volto a migliorare ed estendere prestazioni e funzioni dei circuiti integrati, sempre basati sulla tecnologia MOS. Sul lungo periodo, però l'approccio evolutivo è necessariamente destinato a fallire, e occorrerà sviluppare tecnologie alternative. Queste nuove tecnologie in via di sviluppo hanno come principale obiettivo quello di rimpiazzare nel tempo l'attuale tecnologia CMOS dato che essa è destinata a volgere al termine della sua evoluzione entro qualche decina d'anni, quando la complessità dei circuiti integrati sarà sull'ordine del miliardo di transistor. Fig 2.5.1 può aiutare a capire l'entità del problema, confrontando la densità di transistor nei circuiti integrati al variare delle varie epoche.

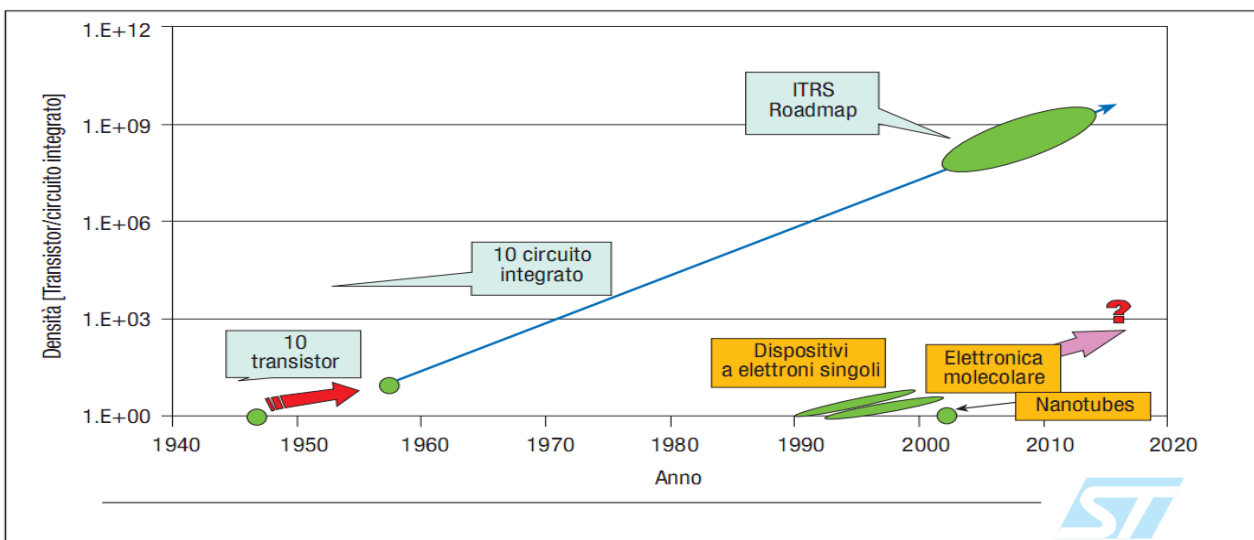


Fig 2.5.1: Evoluzione della densità di transistor nei circuiti integrati. [22]

In generale le attuali tecnologie sono volte allo sviluppo di transistor sempre più piccoli per aumentare la densità ma tendono ad ignorare tutti quei problemi che fino a qualche anno fa non venivano neanche presi in considerazione. Le tecnologie attuali utilizzano nei circuiti integrati un livello di transistor e dai 6 ai 8 livelli di interconnessioni e attualmente la gran parte della potenza dissipata viene spesa per caricare e scaricare le capacità parassite da esse introdotte. Una nuova tecnologia quindi, prima dello sviluppo di transistor di dimensioni minori rispetto a quelle odierne, dovrà risolvere il problema delle interconnessioni tra gli elementi attivi se vorrà far proseguire l'evoluzione dettata dalla legge di Moore. Attualmente sono già conosciute delle molecole dell'ordine dei 3 nm capaci di svolgere le funzioni base dell'elettronica (come la memorizzazione di dati o la rettificazione) ed è quindi ipotizzabile il loro utilizzo come dispositivi dal cui assemblaggio nascerà un circuito integrato. Tra tutte le molecole disponibili un gruppo in particolare hanno attirato l'attenzione su se stesse ed esse sono chiamate molecole coniugate(mostrate in Fig 2.5.2).

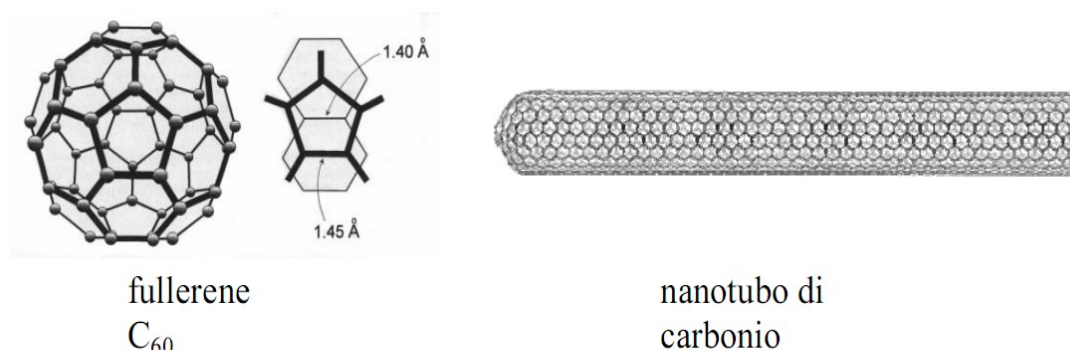


Fig 2.5.2: Le molecole coniugate[23]

Ciò che le rende particolarmente adatte ad essere utilizzate in campo elettronico

è la loro caratteristica di essere drogabili in maniera simile a quello che avviene con il silicio, così da poter effettuare la traduzione dell'elettronica a base di silicio al mondo del carbonio. Gli studi attuali si stanno concentrando sulla possibilità di utilizzare tali molecole per lo sviluppo di funzioni elettroniche complesse e i problemi riscontrati sono principalmente due. In primo luogo, è necessario lo sviluppo di una tecnologia “dal basso verso l’alto” (bottom-up) e cioè capace di disporre le molecole-dispositivo in maniera ordinata ed accessibile con tecniche autoassemblanti. Inoltre tale assemblaggio non può essere condizionato dalle caratteristiche intrinseche delle molecole, ma deve corrispondere ad un ordine arbitrario imposto dall’esterno. Il problema da risolvere è quindi quello di arrivare ad una traduzione del progetto del circuito integrato nella corrispondente disposizione delle molecole, senza far uso delle costose tecniche litografiche attualmente utilizzate per lo sviluppo dei chip di ultima generazione. Si deve inoltre far fronte al problema costituito dall’estrazione a livello macroscopico (e quindi accessibile all’uomo) dell’informazione contenuta negli stati microscopici di atomi o molecole che richiede sistemi di amplificazione all’avanguardia ma che sono già attualmente realizzabili attraverso l’uso di componenti microelettronici. In via di principio quindi la soluzione più logica è quella di sviluppare una tecnologia ibrida micro-nanoelettronica, ossia di mettere sullo stesso chip molecole-dispositivo (assemblate in modo tale da svolgere le funzioni di manipolazione dei dati) e componenti tipici della microelettronica per svolgere funzioni di amplificazione, indirizzamento e alimentazione. In questo modo si dovrebbe coniugare le tecnologie di miniaturizzazione (top-down) proprie della microelettronica a quelle autoassemblaggio (bottom-up) proprie della nascente

elettronica molecolare.

Conclusioni

Lo sviluppo della microelettronica, sostenuto da fattori tecnici ed economici e reso possibile dalla lontananza di limiti fisici fondamentali, è avvenuto su base spontanea ed in maniera esponenziale per oltre 40 anni. Le difficoltà progressivamente crescenti descritte in precedenza a cui si sta andando incontro nel processo di miniaturizzazione suggeriscono che tale sviluppo spontaneo è destinato a cessare in tempi brevi. Al fine di non perdere i vantaggi che lo sviluppo esponenziale della microelettronica offre, non solo al proprio settore ma all'intera economia mondiale, si è passati da una crescita spontanea ad una assistita, attraverso una chiara identificazione dei problemi che occorre risolvere, affinché la legge di Moore continui a valere. È presumibile che la focalizzazione della ricerca sui colli di bottiglia allo sviluppo, e la concertazione delle attività da parte di diversi operatori, riusciranno a sostenere la tendenza attuale almeno un'altra decina d'anni. Sul fatto che la tendenza alla miniaturizzazione possa continuare oltre il 2020 o che lo sviluppo dell'elettronica passi attraverso nuovi tracciati tecnologici (come l'autoassemblaggio di dispositivi molecolari) resta argomento di intensa discussione.

Referenze:

- [1] http://www.cmg.org/measureit/issues/mit41/m_41_2/plot.png
- [2], [3], [4] "Circuiti integrati digitali: l'ottica del progettista" Di Jan M. Rabaey, Anantha P. Chandrakasan, Bora Nicolić pag. 132, 122, 240
- [5] http://www.hardwarezone.com/img/data/articles/2004/1011/power_density.jpg
- [6],[22],[23] "La legge di Moore e lo sviluppo dei circuiti integrati" Livio Baldi, Gianfranco Cerofolini
- [7], [8], [9], [13], [15], [15] "Le frontiere del silicio" L. Marchetti
- [10], [11], [12] "Carrier-Transport-Enhanced Channel CMOS for Improved Power Consumption and Performance" Shinichi Takagi
- [14] "Advanced CMOS device technologies for 45nm node and below" A. Veloso
- [16] http://www.fz-juelich.de/ibn/datapool/page/608/abb1_k.jpg
- [17],[18] "Slide corso nanoelettronica" Andrea Cester
- [19], [21] "FinFETs for Nanoscale CMOS Digital Integrated Circuits" Tsu-Jae King
- [20] "Fabrication and Characterization of bulk FinFETs for Future NanoScale CMOS Technology" Jong-Ho Lee
- [21] "Comparison study of FinFETs: SOI vs. Bulk" SOI consortium

Materiale Utilizzato:

- "Circuiti integrati digitali: l'ottica del progettista" Di Jan M. Rabaey, Anantha P. Chandrakasan, Bora Nicolić
- "Slide corso nanoelettronica" Andrea Cester
- "FinFETs for Nanoscale CMOS Digital Integrated Circuits" Tsu-Jae King

“Carrier-Transport-Enhanced Channel CMOS for Improved Power Consumption and Performance” Shinichi Takagi

“La legge di Moore e lo sviluppo dei circuiti integrati” Livio Baldi, Gianfranco Cerofolini

“Le frontiere del silicio” L. Marchetti

“Frontiers of silicon-on-insulator” G. K. Celler

“Novel Gate Concepts for MOS Devices” Jean-Pierre Colinge

“Multi-gate devices for the 32 nm technology node and beyond” N. Collaert

“Innovative Materials, Devices, and CMOS Technologies for Low-Power

“Mobile Multimedia” Thomas Skotnicki

“Advanced CMOS device technologies for 45nm node and below” A. Veloso