

Il silicio



Frammento di una meteora contenente più del 50% di silicio



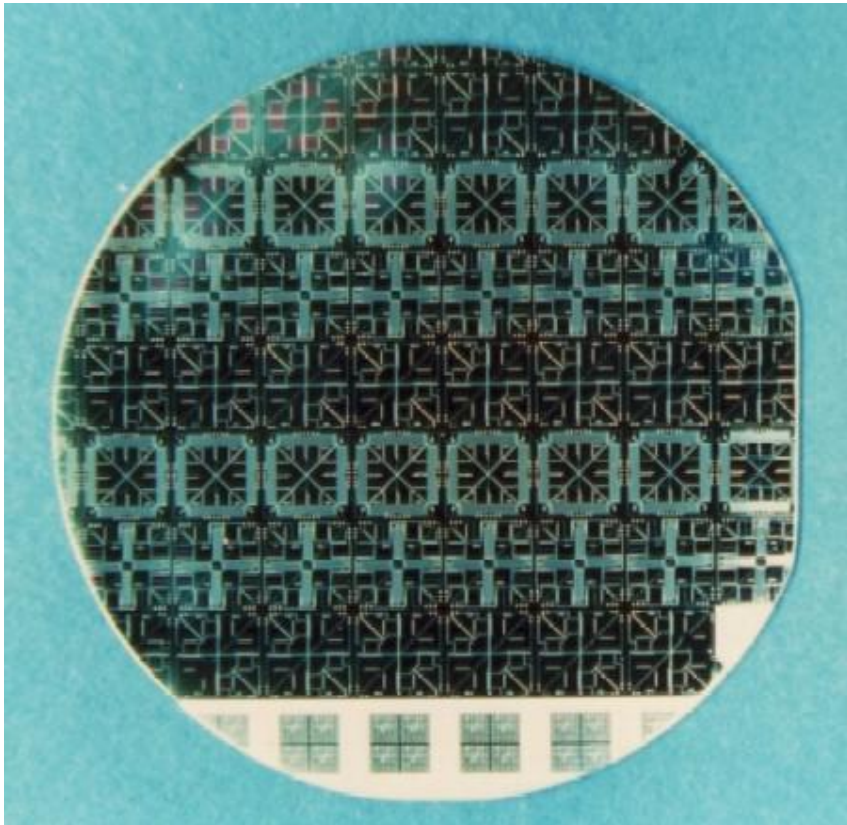
Un campione di silicio elementare

Numero atomico 14
Simbolo Si
Francese silicium
Inglese silicon
Tedesco Silicium
Italiano silicio
Spagnolo silicio

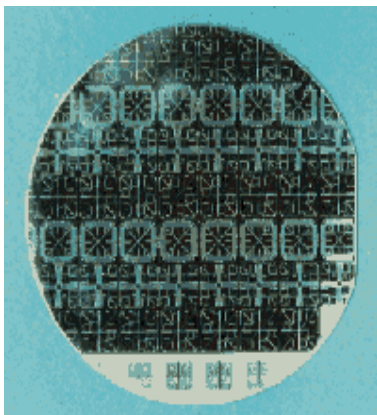
Costanti chimico-fisiche del silicio

Massa atomica relativa	28.0855 uma	Configurazione elettronica	[Ne] 3s ² 3p ²		
Valenza	2, 4	Numeri di ossidazione	+2, ±4	Gusci	2,8,4
Elettronegatività	1.80	Stato di aggregazione a 25 °C	Solido	Struttura cristallina	Cubico a facce centrate
Energia di prima ionizzazione	8.1517 eV	Energia di seconda ionizzazione	16.345 eV	Energia di terza ionizzazione	33.492 eV
Raggio atomico	1.46 Å	Raggio covalente	1.11 Å	Raggio ionico	0.26 (+4) Å
Volume atomico	12.1 cm ³ /mol	Calore specifico	0.71 J/gK	Calore di vaporizzazione	384.220 kJ/mol
Calore di fusione	50.550 kJ/mol	Conducibilità elettrica	2.52 ⁻¹² · 10 ⁶ /cm · ohm	Conducibilità termica	1.48 W/cmK
Temperatura di fusione	1410°C	Temperatura di ebollizione	3265°C	Densità a 293 K	2.33 g/cm ³

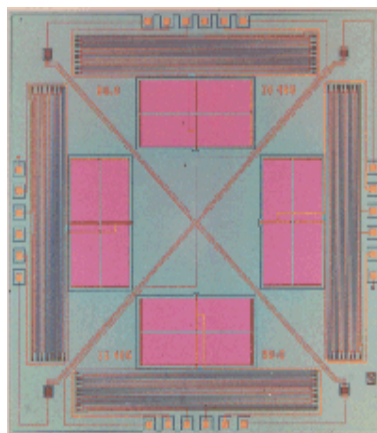
Sensori & Silicio



Un'immagine simbolo della nostra epoca. Un wafer di silicio su cui sono presenti tre tipi diversi di sensori

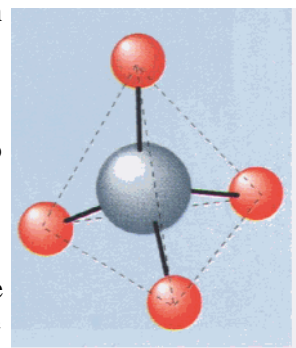


Il diametro reale del wafer è di due pollici



Questo è un sensore di flusso, sensibile alla direzione del flusso stesso. Le dimensioni reali sono di 6 mm x 6 mm

Il silicio fa parte dello stesso gruppo (il 4°) del Carbonio, quindi vi sono somiglianze tra i due elementi. La configurazione esterna del silicio è $3s^2 3p^2$. Come il carbonio ha la possibilità di eccitare un elettrone dall'orbitale s per raggiungere valenza 4 (configurazione sp^3). Il silicio non ha però le grandi possibilità del carbonio, in quanto tende ad ibridizzare **solo** sp^3 . Inoltre il silicio non forma le lunghe catene tipiche della chimica organica. Infatti il legame silicio-silicio ha un'energia troppo bassa che non permette una grande stabilità. Per questo motivo gli atomi di silicio sulla Terra tendono a legarsi con l'ossigeno atmosferico, formando catene Si-O-Si-O-.... in cui ogni atomo di silicio si lega a 4 atomi di ossigeno (fig. 1).

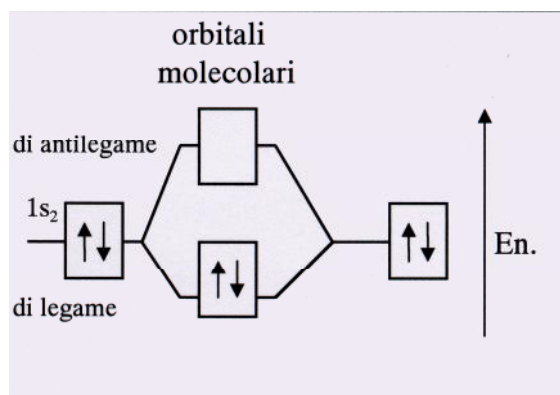


(figura 1)

Il silicio è stato scoperto negli anni '70 del secolo scorso e si è visto utile per l'elettronica in quanto è un **semiconduttore**, cioè non conduce, ma se si inseriscono nel cristallo di silicio piccole percentuali di elementi del 3° o del 5° gruppo si verifica la conduttività. Questo viene spiegato con una nuova teoria sui legami covalenti. **Block** è stato il primo a pensare a questa complessa teoria che spiega le caratteristiche fisiche che il modello classico non riusciva a spiegare.

Il nuovo modello di Block

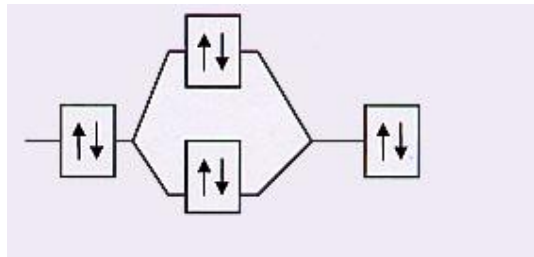
Se si immaginano due atomi di idrogeno, questi avranno un orbitale con un certo quid di energia. Quando i due atomi si avvicinano cambia il loro contenuto energetico **senza sovrapposizione orbitalica** e, dei due orbitali molecolari che si formano, uno ha contenuto energetico più alto, l'altro più basso. Gli elettroni occupano prima gli orbitali di livello più basso (detto "**di legame**"). L'altro orbitale rimane, in questo caso, vuoto e viene chiamato orbitale **di antilegame**. Nel caso dell'elio si spiega l'improbabilità di ottenere una molecola He_2 in quanto la molecola non si stabilizza (vedi fig. 2), dovendo occupare anche un livello di energia più alto.



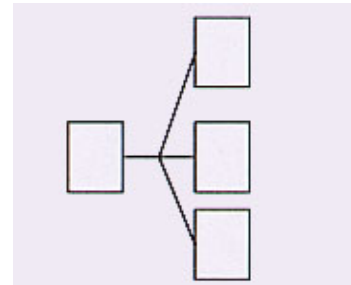
(figura 2)

Il legame metallico

Immaginiamo due atomi di litio (fig. 3): quando questi si avvicinano, i due orbitali 2s si dividono a due differenti livelli energetici. Ma se gli atomi di litio sono 3 si avranno 3 nuovi livelli di energia (fig. 4). Quando gli atomi sono n si avranno quindi n diversi livelli energetici che formeranno una **banda di energia continua** in cui gli elettroni si disporranno nella metà del livello energetico inferiore. Ognuno degli orbitali è però in comune con tutto il campione. La conducibilità è quindi data dalla semplicità del "salto" energetico degli elettroni (**transizione elettronica**). In un cattivo conduttore la banda è riempita completamente e, per dare la conducibilità, si deve dare una notevole quantità di energia, necessaria per passare ad un livello più alto.



(figura 3)



(figura 4)

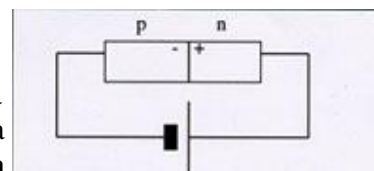
Il caso del silicio

Il silicio è, in condizioni normali, un cattivo conduttore. Ma se nel cristallo di silicio vengono inserite piccole percentuali di elementi del 3° o del 5° gruppo il composto risulta conduttore. Infatti l'elettrone del silicio non utilizzato nel legame con l'elemento estraneo nel caso di drogaggio con elemento del 3° gruppo si posiziona tra due bande (in una situazione instabile). In questa situazione l'elettrone viene eccitato mediante una piccola quantità di energia che permette la conducibilità. Se il drogaggio viene effettuato con un elemento del 5° gruppo sarà la lacuna positiva ad essere "spostata" facilmente.

Semiconduttori

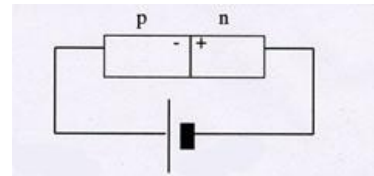
Se si collegano due piastre diversamente drogate il "gas di elettroni" si diffonde nelle due lastre. Ma gli elettroni sono carichi e, spostandosi, lasciano una carica positiva sulla lastra che hanno abbandonato.

La distribuzione delle cariche rallenta il processo di diffusione. Ma nella lastra p vi sono lacune che, colmate dagli elettroni della banda n, si spostano in questa. Attorno alla giunzione vi è quindi una **zona di svuotamento** (depletion zone) in cui i portatori maggioritari diffondono. Se si collega alla lastra una pila come in figura 1, si rivelerà che la giunzione tende a non condurre in quanto la zona attorno alla giunzione non permette lo scorrimento dei portatori.



(figura 1)

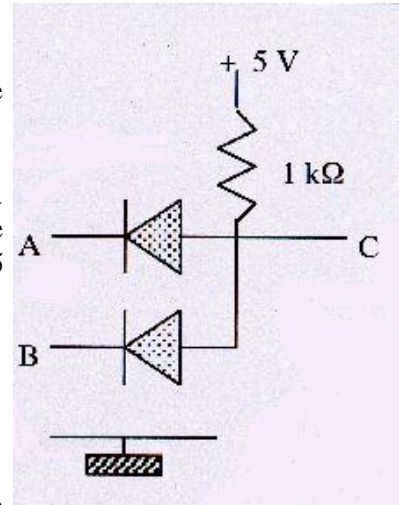
Se invece si collega la pila nel verso opposto (fig. 2) i portatori diffondono ulteriormente grazie al campo elettrico. Si è ottenuto un **diodo**.



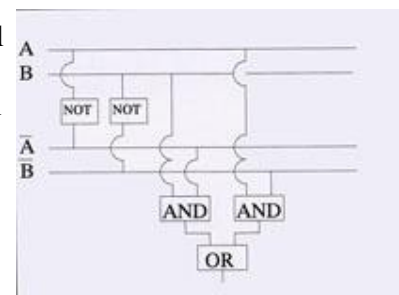
(figura 2)

Circuiti logici

Tramite gli operatori AND, OR, NOT si può creare tutta la logica. I segnali logici vengono trasportati attraverso delle differenze di potenziale: si può usare, ad esempio, una differenza di potenziale di 5 V per il valore 1 e 0 V per il valore 0. Se si utilizzano segnali in ingresso logici e alcuni diodi è possibile creare i primi circuiti logici. Due ingressi A e B sono collegati a due diodi; si ha una filo collegato a terra che rappresenta il segnale di valore 0 (0 V). Il segnale di valore 1 è dato da un generatore a 5 V (vedi fig. 4). Se si collega il punto A al valore 1 e il punto B al valore 0 si ottiene una differenza di potenziale tra C (risultato) e la terra di 0,7 V. Infatti la parte del circuito collegata a B conduce, mentre quella collegata ad A no. Il diodo in B dà una caduta di potenziale di 0,7 V ed è proprio quella che determina il risultato. Dopo alcune porte il risultato non è più affidabile, dato che ogni porta modifica leggermente il segnale. Per calcolare un operatore OR esclusivo (XOR) si utilizza la forma normale disgiuntiva (FND): $A \cdot B + \bar{A} \cdot \bar{B}$ (le due proposizioni vere collegate mediante operatore OR). Il circuito potrebbe essere quindi quello in fig. 3. L'operatore XOR è utile come operatore **somma** tra due bit (nel caso dei valori 1 - 1 si traslascia il riporto, per questo l'operatore XOR è detto **semisommatore**). Aggiungendo un AND tra A e B si ottiene anche il riporto (**carry**).



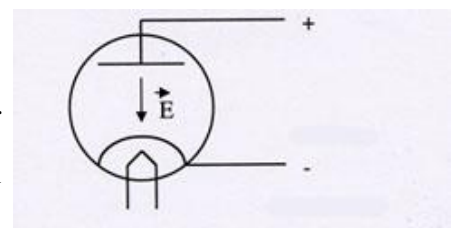
(figura 3)



(figura 4)

Tubi a vuoto

Un filamento in cui scorre corrente riscalda anche un catodo che, in questo modo, può emettere elettroni (effetto termoionico). Infatti alcuni elettroni particolarmente energetici (cioè veloci) possono sfuggire al reticolo metallico. Il gas di elettroni che si è formato troverà un equilibrio con il metallo. Se gli elettroni sono costretti all'interno del tubo si può costruire un **diodo a vuoto**. Se il tubo a vuoto ha due poli (catodo e anodo o placca) si può creare un campo elettrico che spinga gli elettroni verso la placca (fig. 5). Se il campo elettrico è invertito gli elettroni non si muovono perché il catodo continua a emettere altri elettroni.



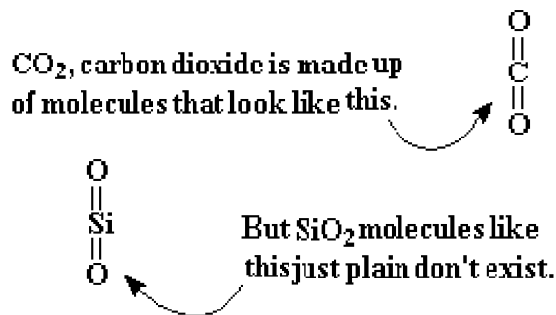
(figura 5)

IL VETRO E' UN POLIMERO?

Capita di parlare del vetro quando si parla di polimeri, specialmente quando si parla di materiali compositi. Le fibre di vetro vengono spesso usate per rinforzare i polimeri. Cosa è questo materiale chiamato vetro? Lo usiamo molto con i poli meri, naturalmente, ma il vetro in se è un polimero?

Prima di affrontare questa questione vediamo cosa è il vetro. La qualità più alta di vetro ha come formula chimica SiO_2 . Può trarre in inganno. Questa formula fa venire in mente piccole molecole di biossido di silicio, ma le molecole di biossido di silicio non esistono.

Il biossido di carbonio CO_2 è fatto di molecole come questa. Le molecole di SiO_2 come questa non esistono.

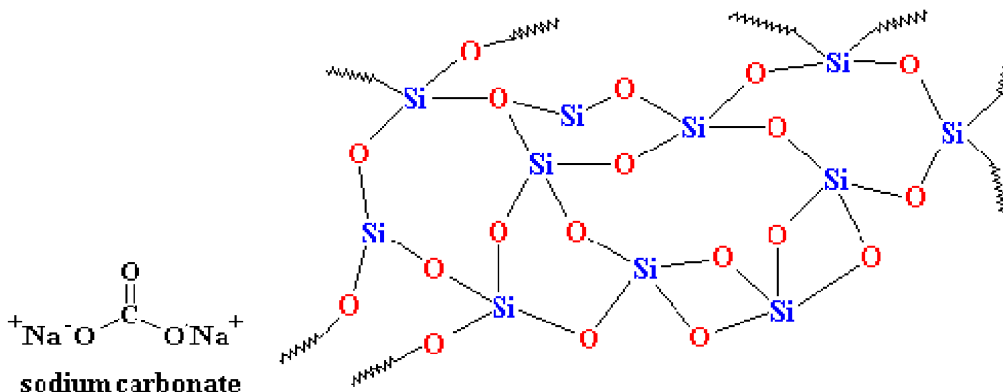


SiO_2 nella sua forma cristallina, il quarzo.

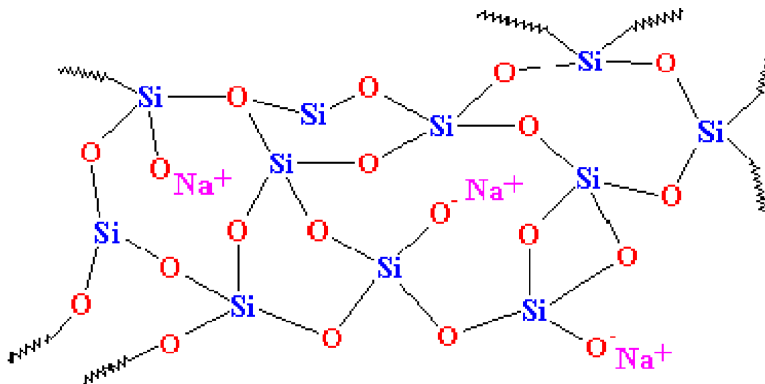


Al contrario in natura SiO_2 si trova spesso come solido cristallino, con una struttura come quella che potete vedere sulla destra. Ogni atomo di silicio è legato a quattro atomi di ossigeno, in forma tetraedrica, naturalmente; ed ogni atomo di ossigeno è legato a due atomi di silicio. Quando SiO_2 è nella sua forma cristallini viene detto silice. Avete già visto la silice. Quando si hanno grandi cristalli di silice abbiamo il quarzo. Quando si hanno tanti piccoli cristalli di silice abbiamo la sabbia.. Questa silice però non è vetro. Bisogna fare ancora qualche cosa prima di trasformarla in vetro. E' necessario scaldarla fino a quando si fonde e si raffredda velocemente. Quando si fonde gli atomi di silicio e di ossigeno sfuggono dalla loro struttura crist allina. Se lo raffreddiamo lentamente, gli atomi lentamente si allineano nuovamente nella loro disposizione cristallina. (Ricordate, il calore non è altro che il movimento casuale di atomi e molecole. Gli atomi caldi si muovono molto, gli atomi freddi si muovono molto poco).

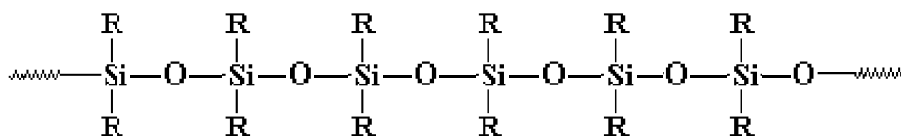
Se la raffreddiamo abbastanza velocemente, gli atomi della silice verranno fermati nella loro traiettoria. Non avranno il tempo di allinearsi, e rimarranno bloccati nelle precedenti disposizioni. Saranno così:



Come potete vedere non c'è alcun ordine nella disposizione degli atomi. I materiali come questo sono detti amorfi. Questo è il vetro utilizzato per le lenti telescopiche e simili. Ha ottime proprietà ottiche, ma è fragile. Per un utilizzo quotidiano abbiamo bisogno di qualche cosa di più resistente. La maggior parte del vetro è realizzata con la sabbia, quando la sabbia viene fusa, normalmente si aggiunge il carbonato di sodio. Si ottiene in questo modo un vetro più resistente con una struttura come questa:



Questo è il vetro che vedete tutti i giorni, nei barattoli e nelle finestre, è il vetro utilizzato nei compositi. È un polimero oppure no? Normalmente viene considerato tale. Perché? Alcuni potrebbero dire che è inorganico, ed i polimeri normalmente sono organici. Ci sono però anche molti polimeri inorganici. Cosa ne dite ad esempio dei polisilossani? Questi polimeri lineari ed inorganici hanno una struttura molto simile al vetro, e sono considerati polimeri. Date un'occhiata ad un polisilossano:



Il vetro può essere considerato un polisilossano altamente reticolato. Normalmente però non la pensiamo così. Perché? Non sono così sicuro. Avete qualche idea?

Le frontiere del silicio

Introduzione

Le dimensioni dei gate dei transistor si sono ridotte di circa 90 volte negli ultimi 30 anni, passando dai 12 micron degli ultimi anni '60 agli attuali 0.13 micron dei processori più recenti.

Questo processo di miniaturizzazione, al di là di accorgimenti architetturali quali pipeline con numerosi stadi caratterizzati ognuno da poca logica o di implementazioni superscalari, è quello che ha consentito l'incremento esponenziale della frequenza di clock, che oramai raddoppia ogni due anni.

Il problema principale che si trovano a dover affrontare gli ingegneri del silicio attualmente è dato dalla **dissipazione termica** e dalle **correnti di perdita** delle più avanzate CPU. Più un chip è piccolo, maggiore è la densità di potenza per unità di superficie che sviluppa e più difficile risulta mantenere il silicio ad una temperatura di funzionamento decente senza dover ricorrere a sistemi di raffreddamento ad acqua come i vecchi calcolatori facenti uso di transistor bipolari anziché CMOS.

Sono quindi allo studio numerose e complesse tecniche che devono far fronte a questo problema: una di queste l'abbiamo già studiata ed è la SOI, della quale tanto si parla oggigiorno e che noi di Lithium abbiamo ampiamente trattato l'anno scorso in questo articolo.

In un'altra sezione abbiamo anche discusso delle questioni inerenti l'utilizzo di packages sempre più complessi e costosi, dietro i quali si celano tecnologie all'avanguardia come la **BBUL** di Intel: anche in questo caso, vi consigliamo di leggere l'articolo a quest'altro indirizzo.

In quest'occasione, ci soffermeremo invece su tre nuove tecnologie promosse da IBM e che promettono di essere presto implementate nelle prossime CPU ed integrati vari di Intel, IBM e Motorola: la tecnologia **Low-K dielectric**, lo **Strained Silicon** e i **Transistor Multigate**.

Inutile dire che si tratta di tecnologie estremamente complesse e di difficile messa a punto: pare addirittura che uno dei motivi per il **ritardo** del rilascio della prossima GPU di nVidia sia da addebitare ai problemi riscontrati da TSMC nella produzione del chip NV30 in tecnologia 0.13 utilizzando collegamenti in rame e materiali low-K! La rivale ATI invece ha prodotto il suo R300 basandosi su tecnologie più consolidate: 0.15 micron per i transistor e piste in rame senza l'uso di materiali low-K.

Fondamenti

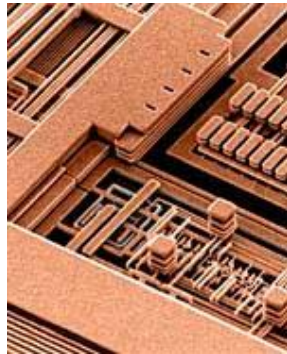
Come ogni articolo tecnico che si rispetti, non si possono trattare questioni complesse senza prima aver chiari i fondamenti della materia che si sta trattando.

Innanzitutto, occorre capire come è costituito un chip: ogni chip è visibile come un palazzo, in cui al piano terra c'è una base di milioni di transistor, mentre i piani superiori ospitano una ragnatela complicatissima fatta di connessioni metalliche che servono a far comunicare i transistor del piano terra.

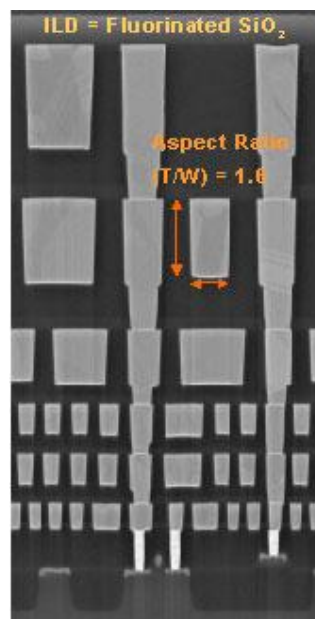
In genere i livelli di connessioni metalliche arrivano fino a sei (e già sono tantissimi!), ma pare che AMD abbia prodotto le ultime CPU Athlon Throughbred facendo uso di ben 8 strati di metallizzazione. Capiremo poi i perchè di tale scelta.

I transistor del piano terra sono complesse costruzioni di silicio, metallo e impurità (i droganti) che sono depositati con precisione elevatissima per creare milioni di minuscoli interruttori on-off.

Di seguito potete ammirare lo spaccato di un chip, che evidenzia i diversi strati di metallizzazione. Sembra in effetti di osservare una fotografia scattata dall'alto e raffigurante una città con case, palazzi e strade.



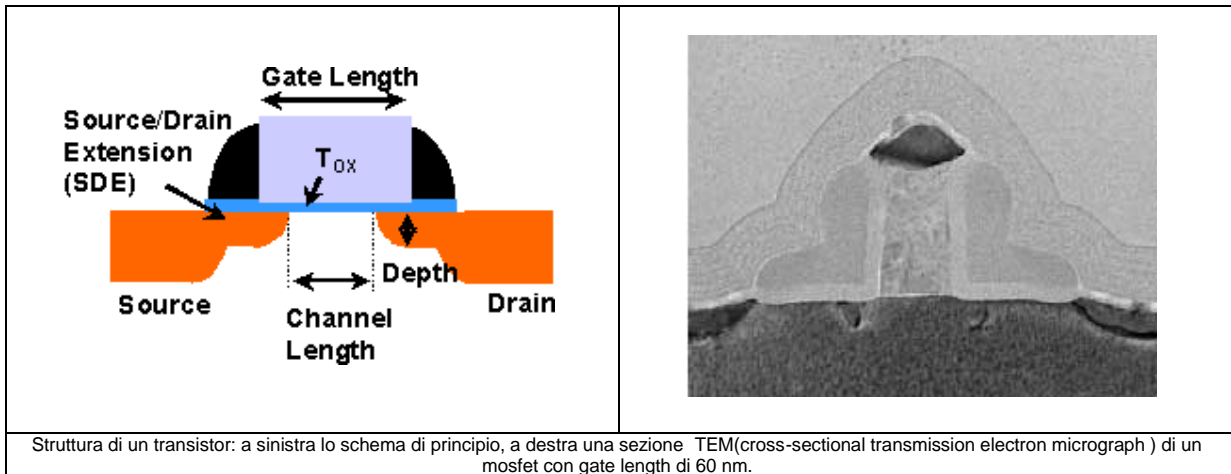
In quest'altra foto possiamo invece osservare i 6 diversi strati di metallizzazione che sprofondano verso i gate dei transistor. In bianco vi sono i contatti con i gate fatti in tungsteno.



Cominciamo ad esaminare gli "inquilini" del piano terra, i **transistor**. In realtà gli inquilini vanno presi, per così dire, a coppie, nel senso che per formare il circuito più semplice, un **inverter**, è necessaria una coppia di transistor, uno chiamato n-mos, l'altro p-mos. Nell'insieme, la logica che fa uso di questi due tipi di transistor viene chiamata logica C-MOS, ove la C sta per complementary.

Rimandiamo all'articolo sulla SOI per comprendere come funziona un SINGOLO transistor.

Per fissare meglio i concetti, riproponiamo di seguito la struttura di un transistor.



La larghezza di gate è quella tipicamente utilizzata per definire il tipo di tecnologia. Se i gate dei CMOS sono larghi 130 nanometri, si parlerà di tecnologia a 0.13 micron.

Tuttavia, gli elettroni fluiscono dal drain al source passando “sotto” il gate, attraverso il “channel” o canale che si viene a creare fra i due pozzi di droganti. Il canale si crea o meno a seconda della intensità tensione applicata sul gate. Per un transistor n-mos, se la tensione è bassa, non può scorrere corrente, se invece si supera una tensione detta di soglia (threshold voltage) allora la corrente può fluire. Agendo sul gate, insomma, il transistor funziona da interruttore.

I fattori primari che stabiliscono la velocità di commutazione di un transistor in sè, e quindi la velocità di clock di un processore per un dato design sono la lunghezza di canale e lo spessore dell'ossido di gate. Ad esempio, in tecnologia 0.13 micron per superare la barriera dei 3GHz la lunghezza di canale non deve superare 0.06 micron e lo spessore del diossido 0.0015 micron.

Ovviamente, il fatto che un transistor isolato riesca a raggiungere i 5Ghz, non vuol dire assolutamente che il chip sarà in grado di viaggiare a quella velocità.

Le performance di un chip sono infatti sempre più limitate dai ritardi RC delle interconnessioni: infatti, mentre i tempi di commutazione dei transistor si assottigliano con l'aumentare della miniaturizzazione, linee più strette comportano un aumento della resistenza e l'avvicinamento delle stesse generano ulteriori capacità spurie che finiscono per produrre effetti di cross-talk o diafonia fra le linee.

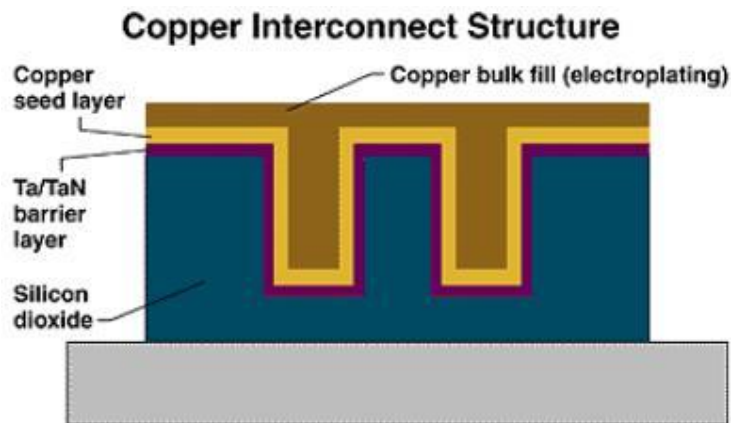
E' qui che entra in gioco la **tecnologia Low-K**.

Tecnologia Low-K

Qualche anno fa IBM introdusse la tecnologia al Rame (**Cu** il simbolo chimico, dal latino Cuprum) che permise di ridurre le resistenze delle interconnessioni al passaggio della corrente rispetto al consolidato alluminio.

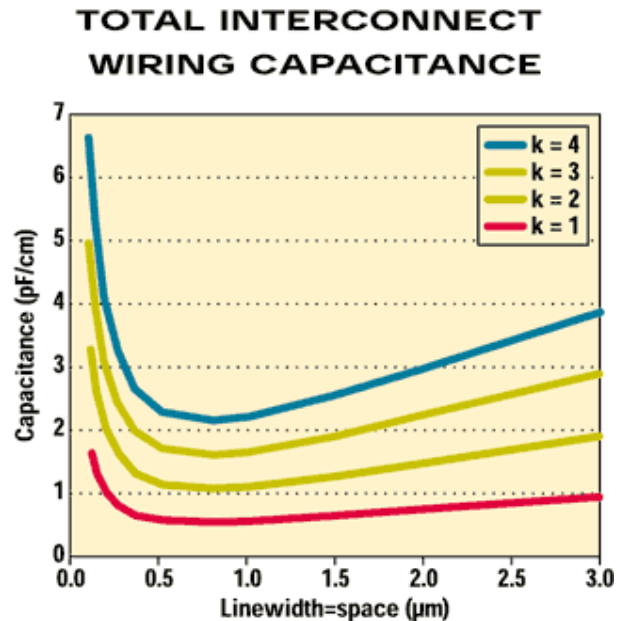
Il rame offre una resistenza specifica normalizzata inferiore a 2 Ohm per centimetro mentre l'alluminio, una resistenza superiore ai 3 Ohm per centimetro.

Il passaggio non fu indolore, perché il rame, così come l'oro, avvelena il silicio. IBM dovette inventarsi un procedimento di mascheramento del rame, com'è possibile vedere in questa figura, dalla quale si evince che il rame è tenuto separato dall'ossido di silicio per mezzo di barriere di diffusione al tantalio:



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 va a braccetto con quella al rame. Per capire la questione, osserviamo la figura sottostante:



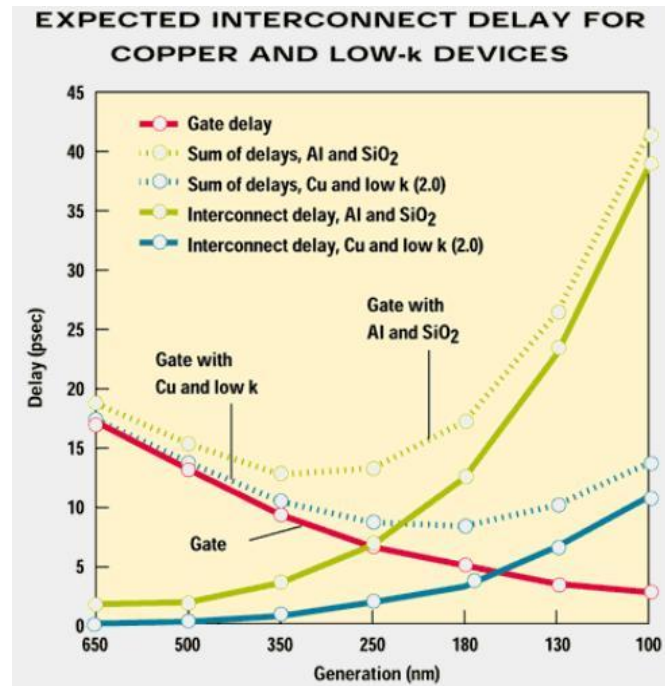
K è la costante dielettrica del materiale interposto fra i collegamenti in rame. Osserviamo che al diminuire della distanza fra i fili situati allo stesso livello, la capacità C sale vertiginosamente.

Un sistema per ridurre C consiste nell'adozione di materiali con bassa costante dielettrica K , **low-K** appunto.

Se con il rame abbassiamo la resistenza R , con i materiali low - K abbassiamo la capacità C , a tutto vantaggio dell'abbassamento del **prodotto RC**, il quale ha le dimensioni di un tempo, si misura in nanosecondi e costituisce il ritardo che accusa un segnale quando passa dallo stato alto a quello basso e viceversa nell'attraversare la pista. Gli ossidi di silicio convenzionali presentano una $K = 3.9-4.2$, mentre materiali innovativi presentano un $K < 3.0$.

Riepilogando, il ritardo RC si combatte abbassando R (uso del rame) e C (uso dei materiali low - K). C è proporzionale a K , e quando lo spazio fra le interconnessioni è $< 0.3\mu m$, la capacità interstrato è quella dominante (costituisce il 90% della C totale). Analogamente a quanto detto prima, l'uso dei materiali low- K permette di ridurre il numero degli strati di metallo abbattendo i costi e garantendo rese per wafer più alte.

Passando dall'alluminio al rame si passa da 14 a 12 strati; se poi in luogo dell'ossido di silicio si usa anche un low-K, si possono raggiungere soli 6 strati di metallo! La figura che segue riassume tutte le parole spese finora:



Vediamo che nel sistema tradizionale(Al/SiO₂) il ritardo in tecnologia 0.13u raggiunge 25 picosecondi. Con il sistema Cu/low-K si scende a 10 picosecondi, cioè si guadagna un 60%. I materiali low-K devono sottostare a specifiche stringenti, che qualitativamente possiamo così riassumere:

- Alta resistenza meccanica
- Stabilità nelle dimensioni
- Alta stabilità termica
- Facilità ad essere disegnati
- Buona aderenza
- Basse correnti di perdita

Alcuni esempi di materiali low-K sono la **silice nanoporosa** (nanoporous silica), l'**HSQ** (idrogenosilquioxano, lo usa la Texas Instruments) e il **teflon-AF**.

Strained Silicon

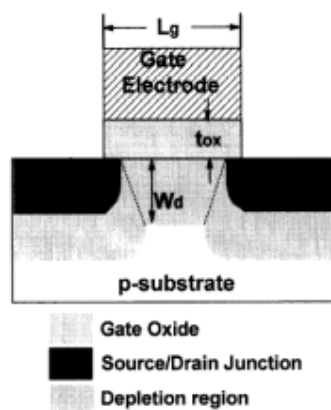
Ci occuperemo di alcune tecnologie sviluppate nei laboratori IBM in questi ultimi anni: l'utilizzo del **rame** al posto dell'alluminio nelle piste per l'abbattimento della resistenza al passaggio della corrente e l'uso di dielettrici **low-K** che diminuiscono il cross-talk e le perdite spurie fra i fili metallici quando questi si trovano a convivere nel medesimo layer a pochi sub-micron di distanza. Cronologicamente, il rame venne introdotto da IBM nel 1997, mentre i materiali low-K hanno fatto il loro debutto a metà dell'anno 2000.

Ora invece parleremo di un'altra tecnologia, sempre promossa da IBM (sebbene non manchino le polemiche a tal proposito, come scopriremo alla fine) che va sotto il nome di **Strained Silicon**, ovvero silicio "stirato". Tale tecnica, *annunciata nel giugno del 2001*, promette così bene che Intel nell'ultimo Intel development Forum, ha dichiarato che intende costruire i prossimi processori nel 2003 in tecnologia 0.09 micron, con wafer da 300 mm e sfruttando la tecnica strained silicon, che permetterebbe un incremento del 20 – 30 % della frequenza di commutazione dei circuiti digitali a parità di altri fattori.

Insomma, se con la tecnologia SOI già si guadagna a parità di design un 20% in frequenza, lo stesso incremento pare sia ascrivibile anche all'impiego del silicio stirato, oltre che da materiali a bassa costante dielettrica.

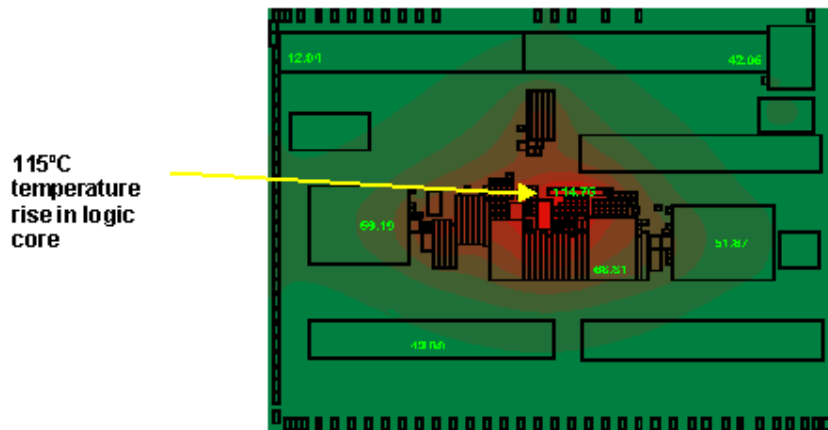
Ma perché non basta più scalare le dimensioni del transistor? I motivi sono molteplici, ma in primo luogo occorre menzionare il **costo delle procedure fotolitografiche** per la produzione degli integrati. All'aumentare della risoluzione ottica, il costo degli impianti sale esponenzialmente. Non solo: scalando le dimensioni del transistor, diminuisce la tensione di alimentazione ovvero la tensione che commuta il dispositivo dallo stato di on (corrente che passa) a quello di off (corrente bloccata): tale diminuzione purtroppo aumenta la possibilità che fluisca corrente anche quando il transistor dovrebbe stare spento. Si capisce che milioni di **transistor deboli** ("weak") fanno passare un sacco di corrente aumentando il consumo medio del dispositivo.

Non è finita: come le dimensioni della lunghezza del gate del transistor si avvicinano agli 0.1 micron, lo **spessore del diossido di silicio** che isola il gate dal canale sottostante si assottiglia fino a raggiungere i 2 nanometri: ciò equivale a 5 strati atomici di diossido di silicio! In queste condizioni, è evidente che la tensione di breakdown oltre la quale si ha la rottura dell'ossido è un problema assai rilevante. Nella Figura sottostante, che rappresenta il solito transistor in sezione, lo spessore dell'ossido è indicato con t_{ox} .



Senza entrare nel merito, dico solo che sono allo studio materiali **Hi-K**, vale a dire con alta costante dielettrica da usare al posto del solito ossido di silicio, ovvero procedimenti per estrarre silicio purissimo, privo di isotopi o di impurità che ne degradano le caratteristiche dielettriche quando ossidato. Con materiali Hi-K sarà possibile stabilizzare per un pò le problematiche connesse alla sottigliezza dell'isolante di gate.

Da ultimo, ma più importante in assoluto, c'è la **dissipazione di potenza**. Osserviamo la figura sottostante: è una termografia "ritoccata" di un processore IBM estratto da un wafer da 200 mm i cui numeri indicano, espressi in gradi celsius, gli innalzamenti della temperatura in diverse zone del chip rispetto alla temperatura ambiente: quindi nella zona centrale il chip raggiunge i 150 gradi.



Quest' area comprende solo il 5-10 % dell'area totale del chip, ed è la zona in cui la logica lavora di più. Se le dimensioni del chip diminuiscono, è più difficile estrarre il calore prodotto.

Tenete conto che la *Semiconductor Industry Association* prevede che entro il 2005 i processori più avanzati arriveranno a consumare 200W @ 5 GHz! A quel punto, se non si escogitano soluzioni mirate, sarà necessario ricorrere al raffreddamento a liquido nei normali pc da desktop.

Per concludere questa carrellata di motivazioni (peraltro nemmeno completa!), c'è da aggiungere la questione **hot spots**, ben nota agli assidui overclockers che spesso ne parlano nei forum di hardware. I normali processi di fabbricazione generano imperfezioni nei transistor a causa di variazioni nella dimensione del drain e source, o per le fluttuazioni dello spessore del wafer, o per variazioni nella concentrazione dei droganti o ancora per le differenze fra le resistenze dei contatti, etc etc. Queste variazioni a loro volta inducono fluttuazioni della densità di corrente e "punti caldi" locali nel circuito integrato.

A causa di questi hot spots, la mobilità degli elettroni decresce (a parità d'intensità di campo elettrico "corrono di meno") e peggiorano sensibilmente le prestazioni del chip all'aumentare della temperatura. Questa è una delle ragioni per cui i processori estratti dal medesimo wafer sono "spee d sorted": alcuni raggiungono i 2 Ghz in totale affidabilità, altri stanno 200MHz sotto. Di conseguenza la resa diminuisce e con essa anche i profitti dell'azienda.

Si capisce quindi che parallelamente alla miniaturizzazione dei transistor occorra anche concentrarsi sullo studio e l'impiego di materiali di nuova concezione rivolti al superamento dei problemi di cui sopra.

Il Si-Ge

Il Germanio (**Ge**) è un elemento che presenta caratteristiche da semiconduttore al pari del silicio o di altri materiali ben più costosi come l'Arsenuro di Gallio (**GaAs**) o il fosfuro di Indio (**InP**) che vengono attualmente impiegati per produrre circuiti ad altissime velocità da utilizzare in tecnologie fotoniche.

A partire dal 1987, IBM si è concentrata nello studio e nella produzione di integrati basati sulle proprietà del **Si-Ge**, fra cui spiccano l'elevata mobilità elettronica, facilità di integrazione e, ultimo ma non meno importante, come vedremo fra poco, la possibilità di costruire dispositivi analogici come sense amplifiers e di impacchettarli accanto a quelli digitali. Fino a qualche anno fa nessuno avrebbe scommesso nella possibilità di costruire amplificatori integrati.

Quanto detto è fondamentale per immaginare i chip del futuro: chip con antenna integrata che si sintonizzano sul segnale tramite circuiti risonanti INTEGRATI su germanio-silicio (addio vecchi condensatori e bobine o componentistica esterna!).

In effetti, è questa la strada che Intel vuole intraprendere nel prossimo futuro: la totale integrazione di memoria, circuiti digitali e analogici in un unico chip. Si parla del 2006 come data limite per questo obiettivo.

Per inciso, IBM in febbraio prima e in giugno poi ha annunciato la sintesi di transistor bipolari (non MOS, quindi), basati su tecnologia Si-Ge, che correvano alla mirabolante velocità di 110 e 210 GHz. Altro aspetto rilevante del SiGe è che consuma meno corrente rispetto al silicio standard: ciò consente di avere sistemi portatili con maggiore autonomia o chip che dissipano meno.

In questo periodo il SiGe è comunemente impiegato nella produzione di tutti quei circuiti che lavorano ad altissima frequenza come set-top box satellitari, sistemi radar per automobili, wireless Lan card, chip per cellulari ma anche PDA o articoli di vasto consumo elettronici che devono essere stipati nel minor spazio possibile.

Come avrete immaginato la tecnologia **Strained-Silicon** di Intel si basa sull'adozione del SiGe. Apparentemente uno sarebbe indotto a credere che Intel parta svantaggiata rispetto a concorrenti che abbiano investito per anni nel Si-Ge. Ma al solito occorre ricordare le capacità finanziarie di cui la Azienda di Santa Clara dispone. Infatti, pare che solo nel 2003 abbia in programma *investimenti di 5 miliardi di dollari in nuovi impianti e macchinari e di 4 miliardi di dollari in ricerca e sviluppo*. E' chiaro che nessun'altra azienda del pianeta è in grado di dispiegare simili "armate finanziarie" nel reparto semiconduttori a meno di formare consorzi. Non solo, negli ultimi anni Intel ha acquistato dozzine di imprese del ramo delle telecomunicazioni, e si capisce bene perché: il Silicio-Germanio può tollerare frequenze di funzionamento altissime, che gli permettono di essere usato presso antenne in apparati radio sofisticati come telefoni cellulari o stazioni di ricetrasmisione wireless.

Insomma, il futuro a breve termine per Intel è l'integrazione del digitale con l'analogico: i circuiti analogici convertono l'input prodotto ad esempio da un'onda radio agganciata da una PLL in un cellulare, in "uni" e "zeri" portati da correnti elettriche in forma binaria. Questi bit possono poi essere processati da normali circuiti digitali.

Non solo: Intel è leader nella fabbricazione di integrati con wafer da 300 mm e tra non molto anche in tecnologia 0.09micron. Si capisce quindi come possa sperare in rese elevatissime che le consentano di essere assai competitiva sul mercato rispetto a soluzioni separate, offerte dalla concorrenza, che prevedano una scheda madre ove inserire chip per l'elaborazione numerica e separatamente circuiti analogici per la ricetrasmissione del segnale.

Strained-Silicon

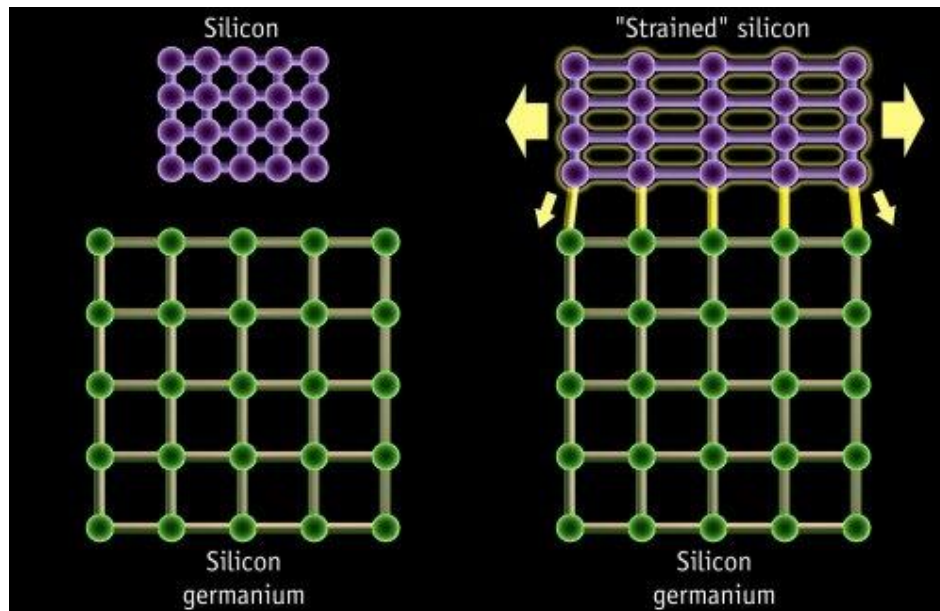
Compresi i punti salienti delle stime di mercato che hanno indotto Intel ad insistere sul Si -Ge, passiamo ora alla personale implementazione del Si -Ge nei circuiti MOS promossa da Intel e che va sotto il nome di strained silicon. Intel prevede di commercializzare i primi processori con questa tecnologia a partire già dalla fine del 2003 sfruttando un nuovo processo di fabbricazione a 0.09micron denominato **D1-C**.

Il principio di funzionamento del mosfet con silicio stirato è semplice: uno strato di silicio germanio viene depositato sulla cima del bulk o corpo del wafer di silicio. Lo spessore di questo strato è di 2 micron e la concentrazione di Germanio è del 20%. La concentrazione di atomi di germanio non è uniforme, cioè lo strato è "graded", con una maggiore concentrazione di atomi di germanio sulla cima. A questo punto, un sottilissimo strato di silicio, dello spessore di circa 20 nm, viene depositato sopra lo strato di Si-Ge.

Questa tecnica allunga il reticolato cristallino del silicio di circa l' 1% sia in direzione laterale che verticale, ma questo piccolo cambiamento impatta enormemente sulla mobilità dei portatori di carica, specialmente gli elettroni, che incontrano una resistenza inferiore al loro passaggio e fluiscono fino al 70% più velocemente., il che porta a chip più veloci del 30% senza dover ricorrere ad un ulteriore procedimento di miniaturizzazione.

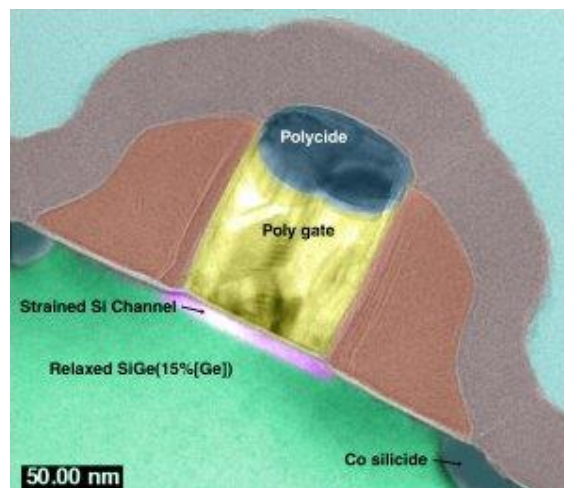
Il principio che sta dietro tutto questo è che gli atomi di silicio dello strato sovrastante tendono a allinearsi con quelli dello strato sottostante di Si-Ge che, essendo più spesso, "vince" e obbliga gli atomi di silicio a spaziarsi in una distanza analoga a quella degli atomi di Si -Ge.

Di seguito potete osservare due figure esplicative.



In quella di sinistra, potete vedere i reticoli cristallini di silicio e silicio-germanio separati. In quella di destra il silicio stirato. Il disaccoppiamento (mismatch) che si viene a creare fra i due reticolati è compensato dallo stiramento dello strato di silicio sovrastante.

Di seguito invece una TEM in falsi colori di un MOSFET in strained silicon.



Conclusioni

La decisione di Intel di tuffarsi nel silicio stirato è stata per certi versi un po' una sorpresa, visto che per anni era rimasta in un atteggiamento "wait and see" verso il reparto R&D di IBM. Intel pare abbastanza sfiduciata dalla SOI, abbracciata invece da AMD dietro licenza dell'onnipresente IBM, mentre afferma che l'utilizzo dello strained silicon incrementerebbe i passi di produzione dei chip solo del 2%.

E non è tutto: IBM prevedeva l'utilizzo del silicio stirato nei circuiti integrati con prima del 2005, all'affacciarsi della tecnologia a 0.065 micron. Intel invece sembra precorrere i tempi.

Insomma, i tempi sono maturi per vedere alla luce nel tardo 2003 processori che combinano piste in rame schermate con materiali low-K e transistor 0.09 micron costruiti con tecnologia strained silicon.

Nel 2005 dovrebbero affacciarsi sul mercato i primi chipset per le tele comunicazioni "all in 1", che renderanno possibili cellulari e apparati wireless ultra-compatti.

Per ora è tutto. Nella terza parte di questo speciale sulle nuove frontiere del silicio ci occuperemo dei transistor **Multi-Gate**, che sembra promettano molto bene nella sempre più affannosa corsa verso le più avanzate frontiere del silicio.

Multi-gate.

L'argomento non è per neofiti e chiedo ai lettori un minimo di pazienza per introdurli al complesso e affascinante mondo delle giunzioni p-n dei semiconduttori. Per quanti non l'avessero già fatto, consiglio la lettura degli articoli sulla **SOI** e sullo **strained silicon** prima di proseguire.

Una volta compreso il meccanismo di funzionamento di elementi come Diodi e MOSFET saremo in grado di comprendere i segreti che stanno dietro ai nuovissimi FINFET transistor di IBM o AMD .

Prima di andare avanti, soffermiamoci un'ultima volta sulle cause che rendono necessario il ricorso a tecnologie così complesse.

Le dimensioni del gate dei transistor sono ormai giunte alla soglia dei 0.1 micron. Per ridurre ulteriormente i mosfet, occorre superare ostacoli come i **tunnel quantistici** che si vengono a creare attraverso l'ossido di gate(essendo questo sottilissimo - pochi atomi di silicio di spessore- le cariche fluiscono anche se non dovrebbero), **fluttuazioni stocastiche** della concentrazione dei droganti(basta una leggera disomogeneità nella concentrazione dei dopanti nel drain e nel source che si generano imperfezioni macroscopiche nel comportamento del transistor), **effetti di canale corto**, etc etc.

Non solo: è più difficile mantenere basse le correnti di perdita perchè è più difficile garantire che un transistor permanga nello stato di off completo quando questi deve commutare miliardi di volte in un secondo.

Chiaro quindi che per superare queste difficoltà si deve ricorrere a materiali diversi oppure adottare accorgimenti del tutto nuovi nella costruzione dei transistor.

Quanto segue è per necessità di natura didattica e imprecisa. Una trattazione più approfondita avrebbe richiesto la trattazione di argomenti quali bande di valenza, di conduzione, livelli di Fermi, densità di probabilità e altre chicche che fanno parte dell'abc della meccanica quantistica applicata ai semiconduttori. Tutto ciò esula chiaramente dai nostri scopi.

Le basi: il silicio drogato

Tutto il mondo dell'elettronica riposa sulla omogiunzione p-n.

Per giunzione si intende il contatto di due elementi, omogiunzione è quando i due elementi appartengono allo stesso materiale (silicio, ad esempio), in caso contrario si parla di eterogiunzione (nello **strained silicon** il germanio e il silicio costituiscono una eterogiunzione, per l'appunto).

Ora, per **silicio drogato di tipo p** si intende un silicio nel quale vengono aggiunte impurità chiamati droganti accettori come l'alluminio. Per **silicio drogato di tipo n** un silicio in cui le impurità invece sono droganti donatori come il fosforo. In maniera molto rude, possiamo dire che l'alluminio si prende un elettrone, il fosforo invece rilascia un elettrone.

La figura sottostante dovrebbe chiarire le idee.

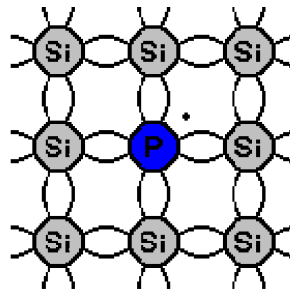


Fig1: Se all'interno del reticolo cristallino del silicio puro(chiamato silicio intrinseco) inserisco degli atomi pentavalenti di Fosforo(P) o Arsenico, si libera un elettrone che è in grado di aumentare la conduttività del silicio drogato. Questo accade perchè il fosforo ha 5 elettroni di valenza, ovvero 5 cariche esterne che possono legarsi con gli atomi intorno. Quattro di queste cinque cariche si legano ad altrettanti atomi di silicio, la quinta non può formare legami ed è libera di muoversi dietro l'applicazione di un leggero campo elettrico.

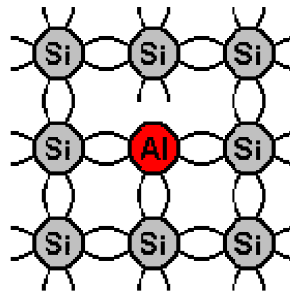


Fig.2: Se invece inserisco un atomo trivalente come il Gallio o l'Alluminio, quest'ultimo, offrendo tre soli elettroni di valenza, può formare solo tre legami anzichè quattro. Si libera così un 'buco', chiamato lacuna, che può essere occupato da un elettrone. Se un elettrone occupa una lacuna, si può formare il quarto legame, ma detto elettrone lascerebbe dietro di sé a sua volta un legame scoperto, e quindi una lacuna. Ripetendo il di scorso, dall'esterno come se una carica positiva fluisse attraverso il materiale se venisse applicato un campo elettrico .

A seconda del tipo di drogante, il silicio modifica le proprie qualità conduttive. Infatti il silicio di suo è come il carbonio, un isolante. Drogandolo, acquista una maggiore propensione a condurre elettricità. Certo, non ai livelli del rame o dell'alluminio, ma certamente più del silicio puro altrimenti detto intrinseco. Insomma, se drogato, il silicio offre proprietà di **semiconduttore**, a metà strada fra l'isolante e il conduttore.

Di per se, il Si drogato di tipo P o N non serve a nulla. E' più economico infatti costruire delle resistenze in carbonio. Le cose cambiano, e cambiano MOLTO, se i due tipi di silicio li unisco, cioè se aggancio una regione drogata P ad un'altra drogata N. Questa giunzione è chiamata anche **diodo**. Solo che noi segusteremo a chiamarla per chiarezza giunzione p -n.

Le basi: la giunzione p-n

Quando unisco due metà drogate diversamente, l'interfaccia fra le due è tale che gli elettroni occupano le lacune. Di conseguenza, se pensiamo al silicio N, è come dire che una carica positiva ha annullato la carica negativa dell'elettrone. Poiché però il silicio drogato era elettricamente neutro, ciò significa che le lacune hanno lasciato delle cariche positive scoperte. Discorso diametralmente opposto vale per la regione P. Guardando la figura, capiamo come stanno messe le cose.

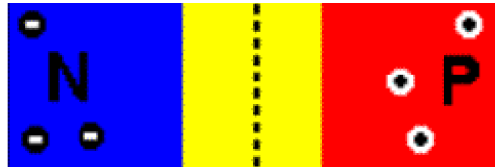


Fig.3: nella regione centrale, in giallo, elettroni liberi e lacune si annichilano a vicenda, lasciando delle cariche fisse scoperte, positive nel lato N(zona gialla sinistra) e negative nel lato P(zona gialla destra). La regione centrale detta Regione di svuotamento(depletion region). Gli elettroni e le lacune create dai droganti permangono lontano dalla regione di svuotamento.

Le cariche scoperte generano un campo elettrico interno, il che **non** vuol dire che scorre corrente nel cristallo così ottenuto, altrimenti violeremmo il primo principio della termodinamica! Infatti appena applichiamo dei contatti ohmici alle estremità si genera una differenza di potenziale opposta che bilancia quella di giunzione. E' da rimarcare il fatto che il mondo dei dispositivi a semiconduttore è vastissimo. Esistono anche delle giunzioni metallo-semiconduttore, o fra semiconduttori diversi(vedi articolo strained-silicon, ad esempio)e, a seconda delle proprietà chimico fisiche dei materiali impiegati, della temperatura in gioco e della scala di integrazione il diagramma corrente tensione varia in modo significativo. Il diodo, al pari del transistor MOSFET, che vedremo fra poco, offre un comportamento estremamente non lineare della corrente se l'escursione della tensione è sufficientemente ampia (nell'ordine del Volt). Per questo e per altri aspetti che esulano da questo articolo, la giunzione p-n è molto versatile, e può essere usata come un rettificatore di tensione, una struttura isolante, come circuito di protezione dai picchi di tensione, o come una capacità dipendente dalla tensione. Inoltre, è utilizzabile all'interno di dispositivi optoelettronici come **celle solari** (costa poco), **fotodiodi** (trasduttori dell'impulso luminoso in impulso elettrico, quindi a valle di una fibra ottica), **LED**(anche se a bassissima efficienza) e persino **diodi laser**.

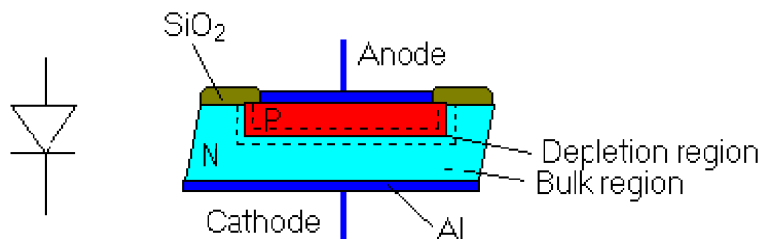


Fig.4: a sinistra il simbolo del diodo. Nella figura a fianco, una sezione di un diodo reale. La depletion region, la regione cioè dove si genera un forte campo elettrico, è quella compresa fra le linee tratteggiate. Per analogia con i vecchi tubi a vuoto, l'anodo è il polo positivo, quello a contatto con la regione di tipo P, il catodo il polo negativo, quello a contatto con la regione N. Da notare come la regione P venga costruita sopra la regione di tipo N. I contatti metallici sono in alluminio, nell'esempio.

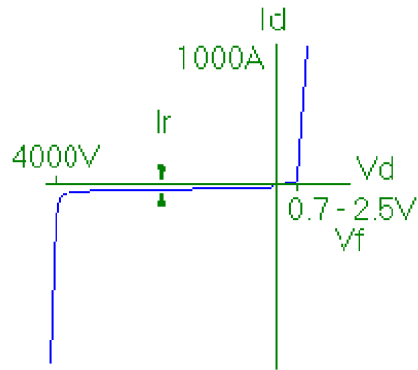


Fig. 5: Caratteristica tensione corrente per un diodo. Se la tensione, sull'asse orizzontale, sale oltre una soglia, tipicamente nell'ordine del Volt, la corrente I_d cresce rapidamente. Parimenti, se la tensione inversa (negativa) scende sotto un'altra determinata soglia(di Zener, in figura 4000V, ma la si può variare in sede di fabbricazione fino a qualche decina di Volt) la corrente in valore assoluto sale rapidamente.

Sulla giunzione **PN** riposano tutti i dispositivi a tre contatti: transistor BJT, MOSFET e JFET. Del BJT ci disinteressiamo poichè, salvo rari casi, con questa famiglia di dispositivi non si realizzano più circuiti digitali. Al solito, quanto segue è trattato comunque in maniera semplice. Il mondo reale è assai più complesso.

Enhancement MOSFET

Il Mosfet ad arricchimento è il classico transistor che viene impiegato in elettronica digitale. E' un tipo particolare di FET (field effect transistor) ed essendo un dispositivo a tre terminali, può essere utilizzato come un interruttore. Agendo sulla tensione del contatto di gate, può scorrere o meno la corrente attraverso gli altri due, cioè fra drain e source

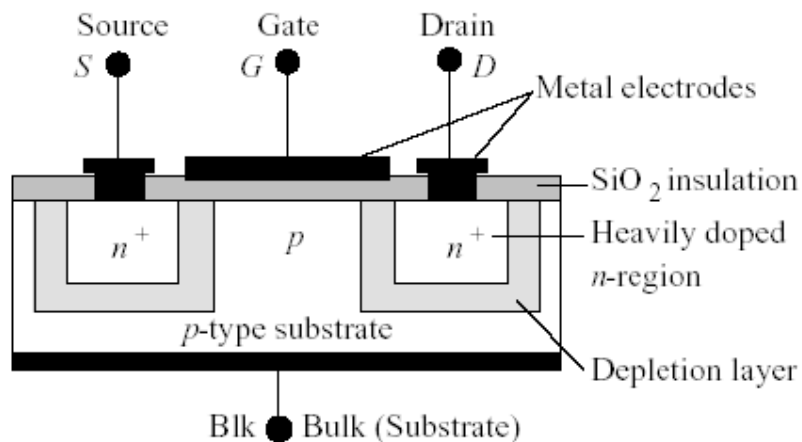
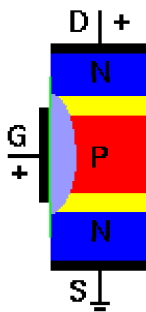


Fig.6: Il gate si trova fra drain e source. Il simbolo n^+ sotto ai contatti di drain e source indica che quella zona di silicio, di tipo n, è drogata pesantemente (alta concentrazione di atomi donatori come fosforo o arsenico). In figura sono mostrate anche le regioni di svuotamento lungo le interfacce fra regioni a drogaggio opposto (il substrato è di tipo p, i pozzi di drain e source di tipo n). Il gate è dato dalla sequenza metallo (contatto), ossido (l'isolante SiO_2) e semiconduttore, da qui il nome MOSFET (metal -oxide-semiconductor field effect transistor).



Per capirne il funzionamento, osserviamo la figura a sinistra. La giunzione del drain è polarizzata inversamente, quella del source direttamente. Quando collego il source S a massa, e applico una tensione positiva sul gate G, ciò che ottengo è attrarre e lettroni (che hanno carica negativa), verso il gate: questi elettroni andranno a costruire un canale che collegherà le regioni di drain e source. Variando la tensione di gate, andrò a modulare la larghezza della regione di canale (in viola).

Poichè è presente un canale laddove prima non c'era (nessuno ha costruito un canale n in sede di fabbricazione del dispositivo) ecco che si parla di MOSFET ad arricchimento(enhancement), ovvero mosfet intensificato.

La spiegazione fin qui data è molto qualitativa,ma serve a farci capire cosa è un canale, un gate e le relative problematiche quando si costruiscono milioni di questi aggeggi su scale dell'ordine dei decimi di micron.

Quando uso questo mosfet in combinazione ad un altro mosfet a canale p, ecco che nasce la logica **C-MOS** (complementary mosfet), con la quale si costruisce la quasi totalità degli integrati moderni.

Quando un transistor è spento, l'altro è acceso e viceversa.

A titolo di esempio, ecco uno schema di inverter, che fa uso di un MOSFET a canale p e di uno a canale n disposti in serie. L'inverter costituisce il circuito base della logica C -MOS.

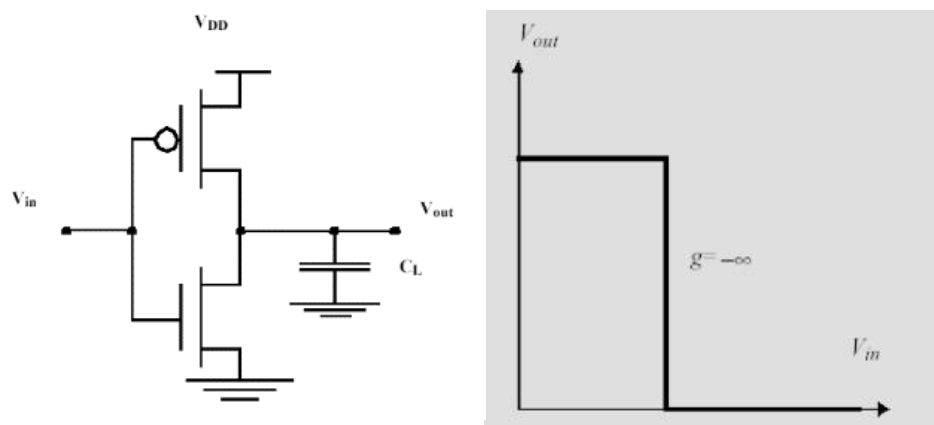


Fig 7: Schema di un inverter in logica C-mos. Il circuito è il più semplice in assoluto: quando la V_{in} , tensione in ingresso, è bassa, il transistor n-mos in basso è spento, mentre il p-mos in alto è acceso. Di conseguenza, l'uscita V_{out} è alta perchè l'uscita è connessa alla V_{DD} . Se invece V_{in} è alta, è il transistor n-mos in basso ad essere acceso, e la V_{out} va a zero (massa). Nella figura di destra il comportamento ideale di un inverter. Appena la V_{in} scavalca la soglia, il transistor commuta. La Capacità C_L di uscita tiene conto delle capacità a valle del dispositivo (capacità di gate del transistor successivo).

Dove stanno i problemi?

1) per commutare, un transistor nmos deve avere una tensione fra gate e source (V_{gs}) maggiore di una tensione detta di soglia. Al diminuire della scala di integrazione la tensione di soglia si abbassa al punto tale che il transistor, anche quando dovrebbe essere spento, in realtà conduce.

2) La curva dell'inverter non è più verticale, ma obliqua. In altri termini la commutazione è lenta.

DOUBLE GATE

Nell'ottica di realizzare dei MOSFET con una lunghezza di gate minore di 0.1 μ m, IBM ha modificato la struttura base del MOSFET, utilizzando una configurazione a doppio gate.

Grazie al fatto che c'è un elettrodo di gate su ogni lato del canale, il campo elettrico generato dall'elettrodo di gate è meglio schermato dagli effetti del source situato alla fine del canale stesso. Non solo, i due gate possono controllare circa il doppio della corrente rispetto alla soluzione a singolo canale, permettendo delle commutazioni on-off molto più rapide. Dualmente, possono abbassare la quantità di corrente richiesta attraverso un dato gate per migliorare il management di potenza, ovvero **raddoppiare** le velocità di commutazione!

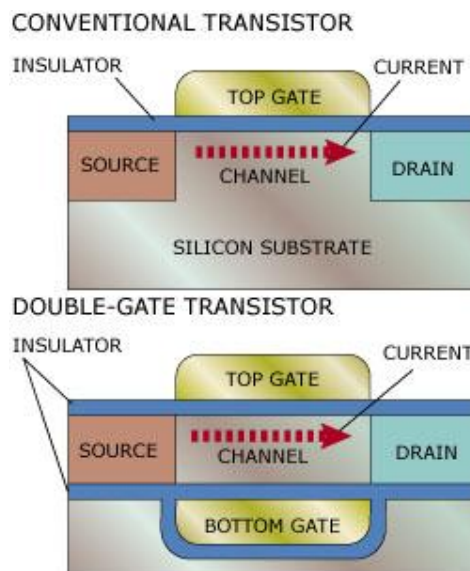


Fig.8: Double gate transistor realizzato in tecnologia planare. Da notare il secondo gate sprofondata nel bulk o substrato del transistor.

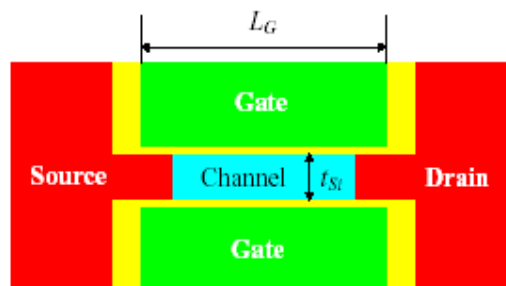


Fig.9: Schema di principio del dual gate. In giallo lo spessore dell'ossido di gate. Questo schema è valido in generale, indipendentemente dal tipo di realizzazione del dual gate(planar o vertical, come vedremo fra poco).

Il problema di fondo è infatti rappresentato dagli effetti di canale corto (**short channel**): più il canale è corto, più è facile che fluiscono cariche fra drain e source anche quando non dovrebbero. Si osserverebbero fluttuazioni della tensione di soglia, capacità incontrollabili, e perdite di corrente.

Oltre ad agire sui profili dei dopanti del drain e nel source per mitigare tali perdite (cosa che in ogni caso richiede procedimenti litografici costosi e raffinati) una soluzione alternativa è rappresentata proprio dal dual gate e dal più intenso campo elettrico di contro llo che questo approccio consente.

Per dirla in parole povere, se il canale rimpicciolisce, il drain e il source si ostentano la sua gestione, mantenendo il transistor in un perenne stato di ON più o meno accentuato. Come dice il proverbio, due è meglio di uno e anche in quest'ambito è stata applicata la stessa filosofia.

Chiaramente ogni medaglia ha il suo rovescio: il dual gate richiede alcuni accorgimenti, infatti non è così semplice impiantare un secondo gate nel substrato di silicio di un transistor esistente. Lo spessore della wafer di silicio preclude infatti questa possibilità. Per di più, il source, il drain e il gate devono essere allineati in modo molto preciso, altrimenti i campi elettrici interni tenderanno a sovrapporsi compromettendo quindi il corretto flusso di corrente nel canale.

IBM, grazie alle conoscenze acquisite con il procedimento **SOI** (per approfondire il quale vi rimando a **quest'altro articolo**), che è un'altra tecnica volta alla riduzione delle correnti di perdita in un transistor, ha sviluppato un metodo all'avanguardia ma molto costoso per produrre circuiti dual gate. Non se ne sa molto (i documenti a tal riguardo non si riescono a reperire) ma pare che su due wafer vengono costruiti i dispositivi a semiconduttore che vengono poi uniti a sandwich garantendo un perfetto allineamento fra i pozzi dei droganti e i gate.

Finora però abbiamo preteso che i due gate fossero costruiti con tecnologia planare ma se potessimo disporre di un grado di libertà in più? se potessimo costruire i transistor anche in verticale cosa accadrebbe?

Transistor 3D

Premetto subito che il transistor 3D, come è stato battezzato dagli addetti ai lavori, altro non è che un transistor in cui gli elementi di base possono giacere non più su piani paralleli ma su piani ortogonali, ammettendo quindi l'utilizzo di una terza dimensione nella costruzione dei dispositivi a semiconduttore.

Un esempio è fornito da questa figura, in cui viene proposta la realizzazione di un transistor verticale.

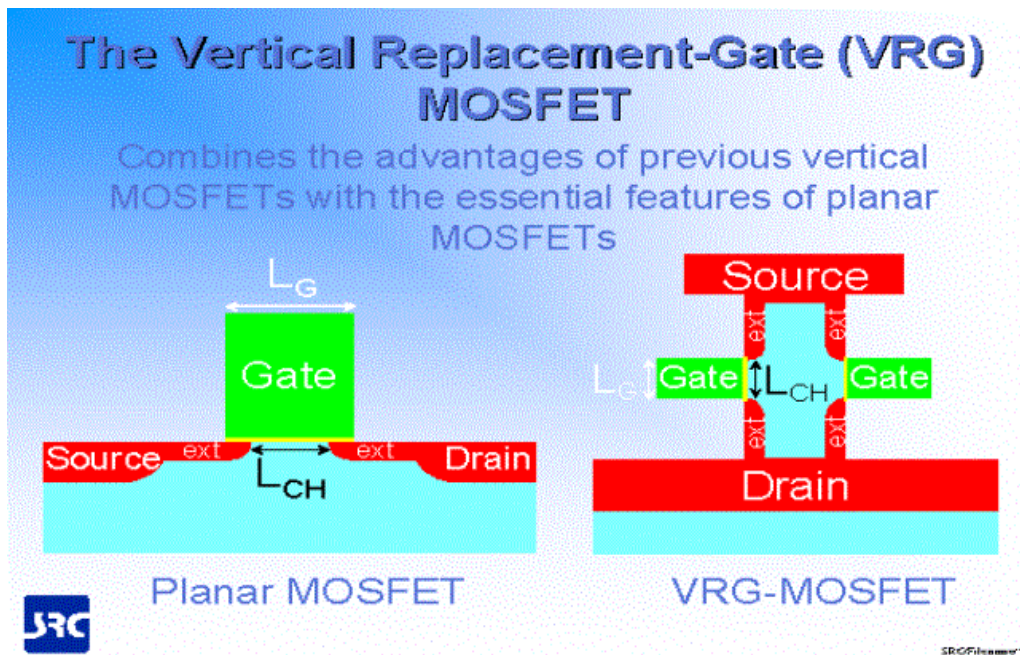


Fig.10: A sinistra un mosfet tradizionale. A destra un Mosfet 3 -D, in cui il canale è ortogonale alla disposizione dei due gate. I mosfet 3D in letteratura sono chiamati mosfet verticali, ed esistono diverse varianti realizzative: la finfet è una di queste varianti.

Un esempio di transistor 3D è dato dal **finfet**. 'Fin' in inglese significa pinna, e il nome dice tutto.

Nel design del **finfet**, una pinna di silicio verticale è costruita tra il source e il drain, dopodichè il materiale dell'elettrodo di gate viene depositato su entrambi i lati del fin per dare origine ad una struttura a doppio gate.

Osserviamo infatti la figura sottostante.

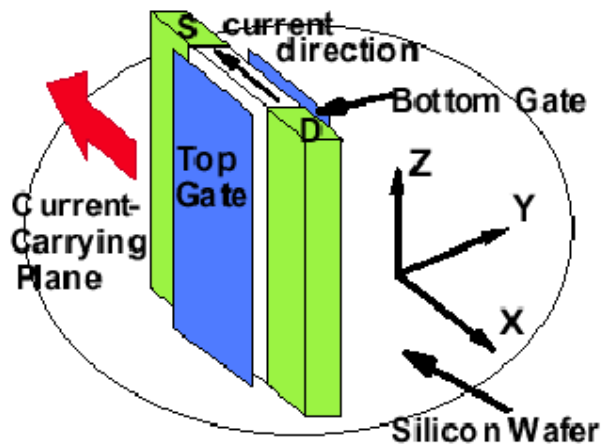


Fig 11 : Da questo schema si vede come il finfet si sviluppi in verticale lungo l'asse z. Nel piano x-y è sistemato il wafer che ospita il transistor. tra il drain e il source viene interposta una pinna di silicio all'interno della quale si genera il canale.

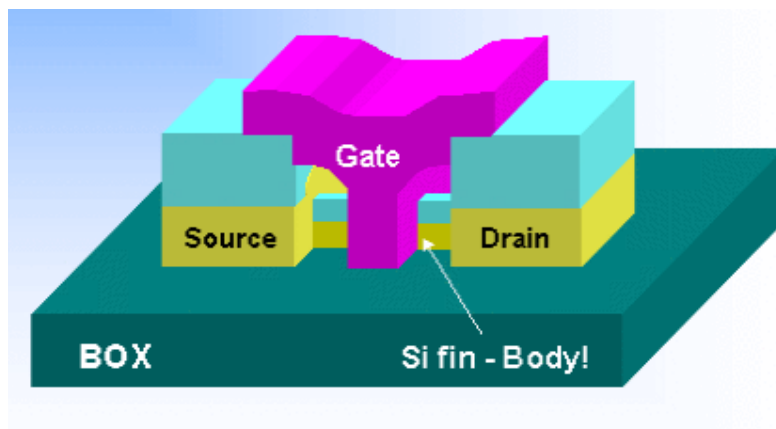


FIG 12: Realizzazione di un finfet. Il materiale di gate avvolge completamente il canale.

Perchè soluzioni verticali? E' presto detto: se lo spessore dell'ossido di gate diventa pari a 5 strati atomici e la lunghezza di canale entra nella soglia dei nanometri, si possono studiare materiali rivoluzionari oppure rivoluzionare il design per impedire che il transistor sia un colabrodo di corrente. Se nei primi due articoli ci siamo occupati del primo aspetto, col transistor verticale ci affacciamo in un nuovo mondo. Nel caso del finfet, il problema della limitata lunghezza del gate che rendeva troppo vicini fra di loro source e drain viene aggirato, in quanto posso agire sull'**altezza**, variandola in misura maggiore rispetto ad una soluzione planare, che impone una lunghezza di gate fissata. Inoltre, è possibile depositare materiale di gate su entrambi i lati del fin in modo da avere un maggiore controllo del canale che si viene a generare nel condotto e limitare il più possibile i danni da canale corto. I transistor 3D necessitano di accorgimenti particolari per la loro costruzione e messa a punto, accorgimenti che esulano dalle tradizionali tecniche litografiche impiegate finora, tecniche che per loro natura sono planari. Chiaramente, si tratta di tecnologie di frontiera, ma nulla vieta che possano essere impiegate in maniera estesa fra qualche anno nella costruzione di processori ad altissime prestazioni e basse perdite.

TEORIA DEI SEMICONDUTTORI

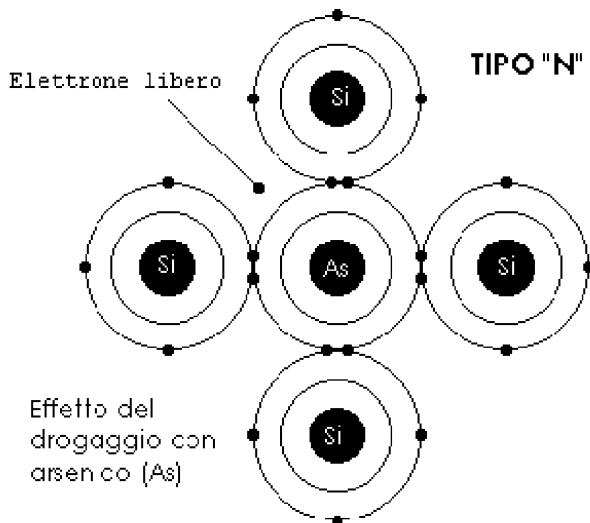
I SEMICONDUTTORI

I semiconduttori sono particolari materiali che non si possono definire nè conduttori nè isolanti. Dispongono di alcuni elettroni liberi, liberi per agitazione termica, molti in più di un normale isolante ma molti meno di un buon conduttore. Da qui il nome di semiconduttori. Il semiconduttore intrinseco (ovvero puro) non ha interessanti virtù elettriche. Al contrario, interessanti capacità si possono creare drogando con opportune impurità il semiconduttore. Il termine drogare indica l'immissione nel materiale semiconduttore di impurezze che lo rendono appunto impuro. I semiconduttori più conosciuti sono il germanio ed il silicio. Il germanio è ormai caduto in disuso, a causa delle sue caratteristiche inferiori rispetto al silicio (maggiore deriva termica, maggiori perdite da correnti parassite inverse, etc...). Ambedue sono tetravalenti, ovvero hanno nell'orbita più esterna 4 elettroni che possono combinarsi con gli elettroni degli atomi adiacenti per formare il classico legame covalente (mutuo interscambio di un elettrone con quello di un atomo vicino), ovvero la presenza di 8 elettroni sulle orbite esterne di cui 4 di un atomo e 4 interscambiati con altri 4 atomi vicini. Il silicio ha 14 elettroni che compensano i suoi 14 protoni presenti nel nucleo. È essenziale che sia chiaro il concetto che elettricamente i materiali semiconduttori sono 'neutri', ovvero le cariche negative (elettroni) sono in egual numero di quelle positive (protoni).



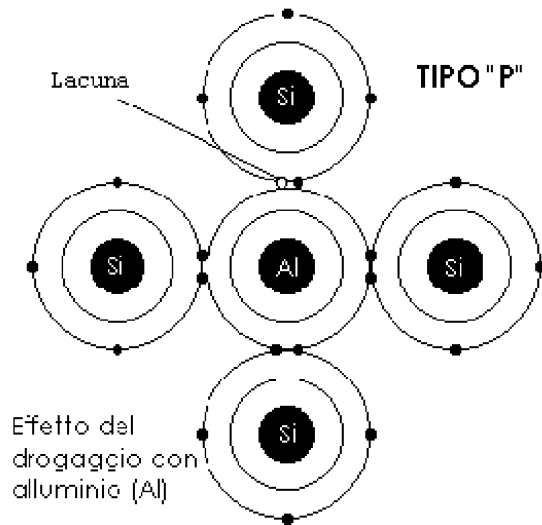
IL DROGAGGIO

Il drogaggio si effettua immettendo, come abbiamo visto, materiale 'drogante' nel semiconduttore. Tale drogaggio può essere effettuato con diversi materiali, secondo il risultato da raggiungere. Usando ad esempio arsenico, che contiene 5 elettroni liberi nell'orbita di valenza ovvero dicasi pentavalente, otterremo che il semiconduttore 'libererà' nella struttura cristallina 1 elettrone per ogni atomo di arsenico immesso. Perché ciò? Dal momento che i materiali conduttori tendono a formare un reticolo cristallino tramite il legame covalente visto prima, (8 elettroni nelle orbite esterne di cui 4 condivisi) si otterrà che all'atomo di arsenico si legheranno 4 atomi di silicio. Siccome l'arsenico ha però 5 elettroni da donare, sulle orbite di valenza rimarrà un elettrone 'libero', dal momento che l'ottetto viene raggiunto con soli 4 elettroni (4 dell'arsenico + 4 donati da 4 atomi adiacenti di silicio). In tal modo nella struttura cristallina rimane questo elettrone libero, pur rimanendo l'intera struttura sempre neutra (infatti l'elettrone libero proviene da un atomo di arsenico che era neutro, ovvero dotato di ugual numero di cariche positive e negative). Questo elettrone 'vaga' dunque nel semiconduttore (ovviamente non è sempre lo stesso, ma si accoppia ad un legame covalente vicino liberando un altro elettrone al suo posto e così via). Siccome il drogaggio ha creato parecchi di questi elettroni liberi avremo un continuo interscambio di elettroni che passano da un atomo ad un altro. La somma però di tutti questi movimenti dà origine ad una corrente nulla, essendo le direzioni degli elettroni sempre diverse e casuali tali da autoannullarsi. Il semiconduttore così trattato si dice di tipo N



Effettuando invece un drogaggio con materiale trivalente (Alluminio) che contiene quindi 3 elettroni liberi nella banda di valenza otterremo questa volta di creare un 'buco' nell'ottetto generato con i 4 atomi di silicio posti nelle immediate vicinanze dell'atomo di alluminio. Tale buco si definisce lacuna. L'alluminio 'cede' i suoi tre elettroni al legame covalente che però non viene raggiunto completamente a causa appunto della mancanza di un elettrone. Tale lacuna viene riempita con un elettrone di un atomo vicino che però farà mancare il legame agli atomi con i quali è legato. In questo modo si può dire che la lacuna si 'sposta' nella struttura cristallina (in realtà sono sempre gli elettroni a spostarsi, ma in pratica si assiste ad uno spostamento del buco da un atomo all'altro). Anche in questo caso il materiale rimane neutro essendo il drogante (Alluminio) neutro.

Il semiconduttore così trattato si dice di tipo P

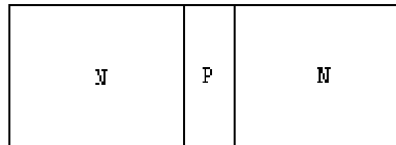


I semiconduttori così trattati sono ancora però in un certo senso 'senza' scopo. Non abbiamo garantito nessuna particolarità elettrica degna di nota al materiale. Le speciali caratteristiche elettriche si creano quando due materiali drogati diversamente (uno tipo N ed uno tipo P) vengono uniti insieme in una struttura chiamata 'giunzione'.

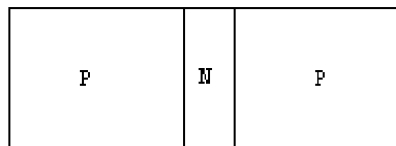
IL TRANSISTOR

IL TRANSISTOR

Come visto nel caso del diodo anche il transistor, ovviamente, sfrutta le particolari caratteristiche della giunzione a semiconduttore. Nel transistor, però, le giunzioni diventano due e sono ottenute accostando materiale N ad due P oppure materiale P a due N. Le tipologie di transistor sono dunque due: N-P-N e P-N-P. Il funzionamento delle due tipologie di transistor è pressochè identico, semplicemente nel primo si parla di circolazione di elettroni e nel secondo di lacune.

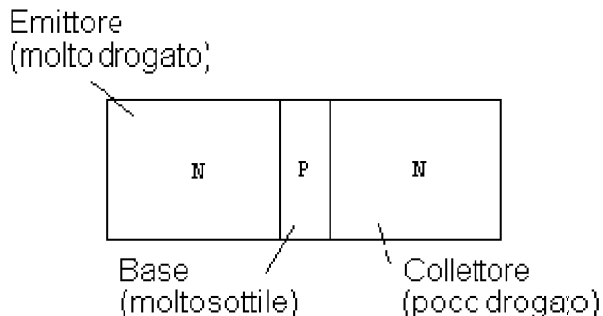


Transistor NPN



Transistor PNP

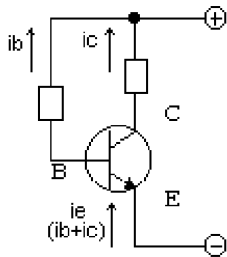
Per spiegare il funzionamento del transistor occorre prima fare qualche precisazione. In particolare è necessario che lo spessore del materiale che sta in mezzo agli altri due (il materiale P nel caso dell'NPN e N nel caso del PNP) abbia particolari caratteristiche. Più precisamente il suo spessore deve essere minore della 'distanza media di diffusione' delle cariche presenti nel semiconduttore, ovvero quella distanza media entro la quale una carica libera si ricombinerà con una di carica opposta. Per semplificare di molto potremmo concepire questa distanza come lo spostamento MINIMO che una carica in movimento nel semiconduttore DEVE ESEGUIRE ogni volta che si sposta. Il materiale posto al centro della giunzione prende il nome di BASE. Uno degli altri due materiali agli estremi verrà drogato in maniera molto più massiccia dell'altro e prenderà il nome di EMITTORE. Il restante materiale prenderà il nome di COLLETTORE.



L'emittore, molto drogato, dispone dunque di parecchi portatori in eccesso che potranno spostarsi nel materiale ed attraverso le giunzioni.

POLARIZZAZIONE

Per poter funzionare il transistor deve venir polarizzato in maniera univoca. Più precisamente dovremo polarizzare direttamente la giunzione Base-Emittore ed inversamente la giunzione Base-Collettore. E' altresì necessario che la giunzione Base-Collettore sia polarizzata ad un livello di tensione molto più elevato (almeno 3..4 volte) di quella Base-Emittore. Analizziamo ora il comportamento del transistor in queste condizioni. Dal momento che la giunzione B -E è polarizzata direttamente avremo che un notevole flusso di cariche cercherà di ricongiungersi in base, partendo dall'emittore che, fortemente drogato, può fornirne in discreto numero. Dal momento che, però, lo spessore della base è inferiore alla distanza media di percorrenza delle cariche, molte di queste cariche tenderanno a 'saltare' letteralmente la base per ricongiungersi direttamente nel collettore, polarizzato oltretutto ad un livello maggiore della base e, dunque, in grado di attrarle con maggior efficacia. In termini pratici otteniamo che solo una piccola percentuale delle cariche 'espulse' dall'emittore riusciranno a ricombinarsi in base, mentre la maggiorparte di esse lo farà nel collettore. Il rapporto tra le due correnti (la corrente tra emittore e collettore e quella tra emittore e base) prende il nome di guadagno in corrente del transistor o beta. Questo numero si mantiene 'abbastanza' lineare entro una discreta gamma di correnti e indica in ultima analisi il fattore di amplificazione di corrente del transistor. Già, perchè la forza del transistor è proprio quella di 'amplificare' la corrente di base x beta volte nel collettore. Con una piccola variazione della corrente di base otteniamo dunque una grande variazione di quella di collettore.



Polarizzazione tipica di un transistor NPN. Nel caso di PNP la freccia interna dell'emittore è rivolta verso l'interno e l'alimentazione ha i poli invertiti .

Pulitura laser di Wafer di Silicio

Questo nuovo sistema di pulitura rimuove rapidamente e con elevata efficienza le minuscole particelle che contaminano i wafer di silicio. I metodi convenzionali si basano sull'utilizzo di melamina, una sostanza chimica tossica. La conservazione di tale sostanza è pericolosa e il metodo di pulitura è estremamente lungo (circa 3 ore) e deve essere eseguito a mano.

Con l'introduzione della tecnologia laser, le sostanze chimiche non sono più necessarie e la pulitura è eseguita in 30 minuti, con ciclo automatico.

Grazie ai **Sistemi Laser di Pulitura Quanta System** è possibile ottenere le migliori performance nella pulitura di wafer di silicio.

Nel campo delle nanotecnologie, attualmente in rapida crescita, è assolutamente necessario che le superfici siano prive di contaminanti. Quindi la pulitura dei wafer è di interesse strategico e negli ultimi anni si stanno sviluppando diverse tecniche di pulizia che sfruttano la luce laser, per rimuovere particelle con dimensioni dell'ordine di 100nm.

Negli ultimi anni le due metodologie seguenti sono quelle che hanno trovato maggiore riscontro:

Pulitura laser a secco (DLC): la superficie da pulire è irradiata con un impulso laser corto. In questo caso svolge un ruolo determinante l'espansione termica del substrato, che accelera le particelle tanto da contrastare le forze di adesione al substrato.

Pulitura laser a vapore: (SLC): un liquido trasparente viene fatto condensare sulla superficie prima di colpirla con il laser. L'assorbimento della radiazione da parte del substrato porta ad un rapido aumento della temperatura e anche il liquido si scalda per conduzione termica. All'interfaccia liquido-substrato si generano delle bolle che poi esplodono evaporando, trascinando via le particelle contaminanti.