Corso di Microelettronica

Introduzione al Processo Tecnologico

INTRODUZIONE AL PROCESSO TECNOLOGICO

Il processo di fabbricazione di circuiti integrati ha inizio con la produzione di silicio policristallino ottenuto da biossido di silicio allo "stato naturale" (cioè da minerali come quarzo o quarzite); il successivo passaggio è la generazione di silicio monocristallino. E' necessario l'uso del silicio monocristallino perché la struttura dei suoi orbitali, le sue bande di valenza e di conduzione ed anche i suoi gap energetici sono tali da permettere di ottenere le funzioni logiche desiderate.

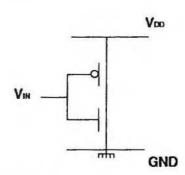
Il silicio policristallino è disponibile in fette con spessore dell'ordine di 1mm e diametro dell'ordine di qualche centimetro. Tipicamente già nella fase relativa alla produzione di silicio policristallino si effettua il drogaggio della fetta con una opportuna specie drogante: se p.e. si vogliono ottenere dei transistori a canale n, essendo necessario un substrato di tipo p, si usa come drogante il boro (B) altrimenti si usa dell'antimonio (Sb), dell'arsenico (As) o del fosforo (P).

Per formare sul silicio la struttura del transistore sono necessari i seguenti strati:

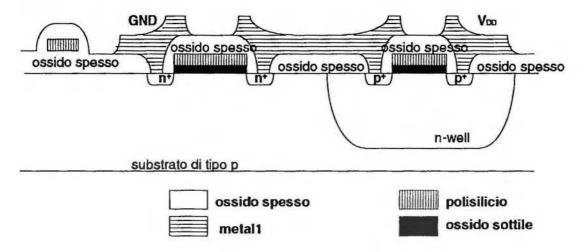
- zone diffuse;
- zona di ossido sottile (cioè lo strato di ossido tra gate e substrato);
- zona di polisilicio (che nelle attuali tecnologie costituisce l'elettrodo di gate).

Per collegare il transistore a strutture analoghe sono inoltre necessarie piste di collegamento, per esempio in alluminio (Al); ed isolamenti, ottenuti attraverso deposizione di ossido (ossido spesso), per separare i collegamenti dal dispositivo.

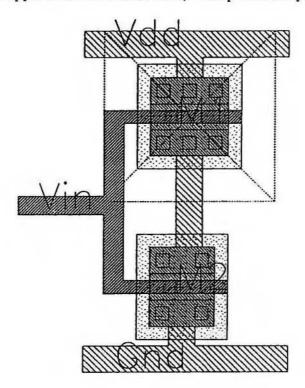
Dato uno schema del tipo



si può considerare la sua sezione verticale in cui sono messi in evidenza gli strati sovrapposti che lo caratterizzano:



In alternativa si può analizzare il suo layout, ottenuto da Cadence cioè uno strumento software di tipo CAD (Computer Assisted Design) che facilità lo sviluppo, il disegno e la verifica della correttezza (attraverso opportuni check automatici) dell'operato del progettista:



Dall'analisi delle due rappresentazioni relative alla realizzazione del circuito si possono fare le seguenti osservazioni:

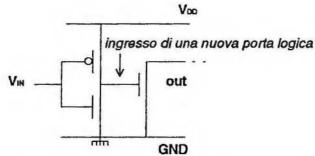
- il substrato della struttura è di tipo p;
- il transistore a canale p è realizzato attraverso la creazione di un n-well sul bulk di tipo p;
- la diffusione è di tipo p+ per un dispositivo a canale p mentre è di tipo n+ per un dispositivo a canale n;
- il segnale di ingresso è portato al gate dei due transistori attraverso piste di polisilicio;
- il gate dei transistori è caratterizzato dalle regioni in cui si ha l'AND tra polisilicio e strato di ossido sottile (od in altre parole in cui sono contemporaneamente presenti polisilicio ed ossido sottile);
- gli strati consecutivi sono tra loro isolati da uno strato di ossido spesso che però non viene evidenziato sul layout (o più precisamente vengono rappresentate le arec dove detto ossido non è presente cioè le zone in cui si hanno i contatti);
- il contatto è rappresentato da un buco nell'ossido spesso effettuato per collegare la metal con la zona diffusa oppure la metal con il polisilicio;
- sia GND che V_{DD} sono rappresentate sul layout con una pista di metal;
- l' elettrodo di bulk è sottinteso e si suppone cortocircuitato a V_{DD} nei transistori a canale p ed a GND nei transistori a canale n;

 nel circuito esaminato il source del transistore a canale p è collegato a V_{DD} quello del transistore a canale n a GND.

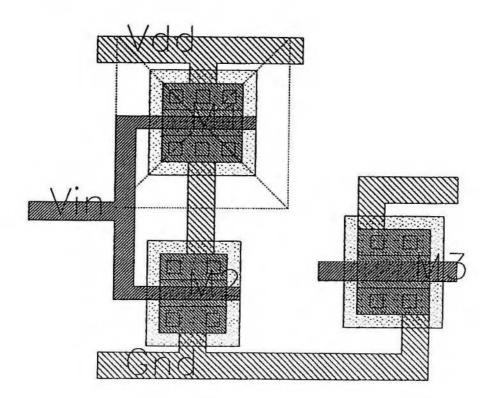
Quello appena visto è un processo standard n-well (esiste anche il processo duale detto p-well) mentre per quanto riguarda il processo C-MOS si ha un substrato intrinseco in cui vanno creati sia i p-wells che gli n-wells. Ricapitolando nell'esempio si hanno:

- 1. due tipi di dispositivi cioè transistore a canale n e transistore a canale p;
- 2. due tipi di collegamenti cioè metal e polisilicio;
- 3. un tipo di contatto (che nell'esempio è utilizzato per collegare la metal alle diffusioni ma che, come si è già anticipato, serve anche per collegare polisilicio e metal).

Un ulteriore passo è quello di complicare un pò il circuito; si suppone di voler collegare i drain dei due transistori al gate di un terzo dispositivo che rappresenta l'ingresso di una successiva porta logica

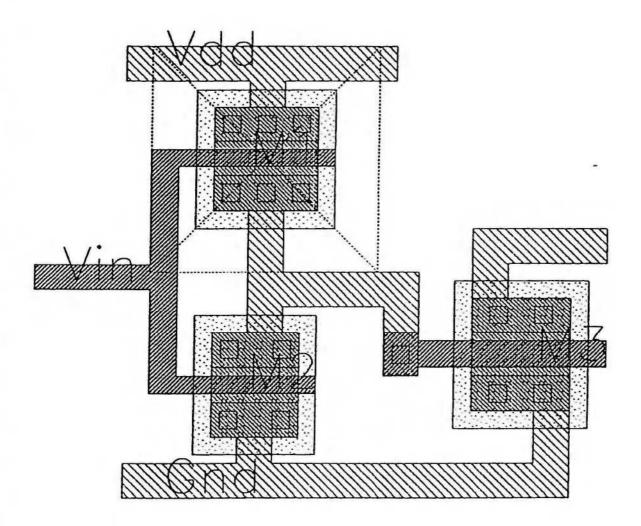


Al layout viene perciò aggiunto un transistore a canale n (in figura il gate del terzo dispositivo è momentaneamente non collegato agli altri due):

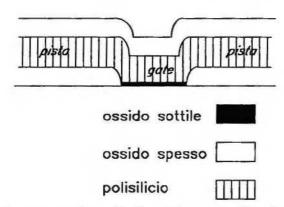


Per come si presenta il layout del circuito di partenza per effettuare il collegamento è sufficiente prolungare opportunamente l'alluminio che congiunge i drain dei due dispositivi e contattarlo al gate del nuovo transistore.

Il corrispondente layout è:



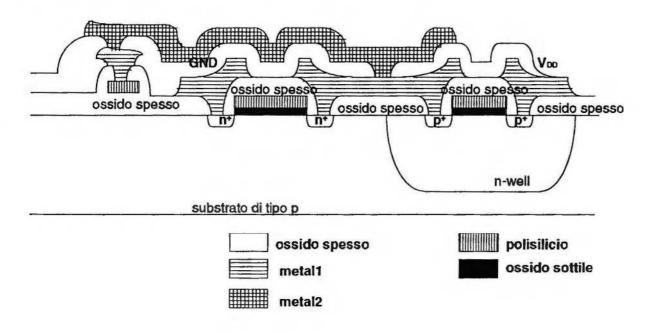
E' utile ribadire che il polisilicio diventa gate quando è sovrapposto all'ossido sottile altrimenti può essere utilizzato nei collegamenti come pista di trasmissione dei segnali; in quest'ultimo caso è isolato dal substrato da uno strato di ossido spesso. Visualizzando quanto appena esposto



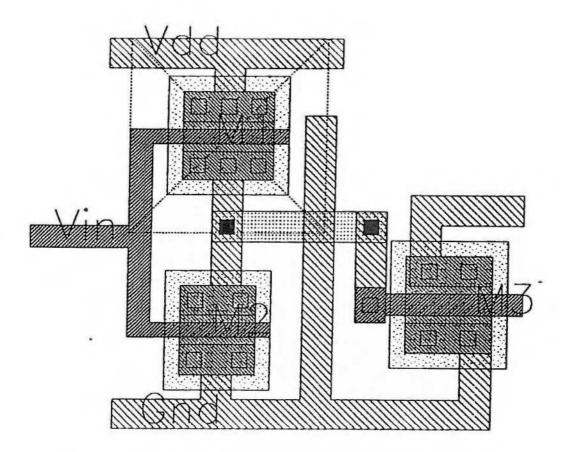
Se sul layout originario ci fosse una pista di alluminio tra quella che congiunge i due drain del circuito e il polisilicio non si potrebbe prolungare, al fine di connettere la nuova porta logica, la metal perché altrimenti le due piste sarebbero indesideratamente cortocircuitate.

Per ovviare a questi inconvenienti i processi tecnologici hanno più strati di metal: se per esempio gli strati di metallizzazione del processo sono 2 (si definisce metal 1 lo strato più basso e metal 2 quello più alto che è isolato dalla metal 1 da una deposizione di ossido spesso) allora il problema è risolubile prolungando la pista di metal 1 che congiunge i due drain con un ponticello di metal 2, superando così la "barriera" di metal 1, per poi continuare il percorso fino al contatto con il polisilicio con una pista di metal 1. Il buco dell'ossido che permette di collegare metal 1 e metal 2 è detto via.

In sezione la struttura è la seguente



Il corrispondente layout è:



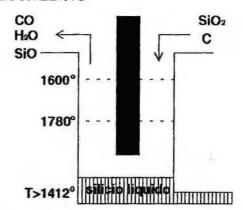
Le figure costituenti il layout sono approssimazioni del circuito effettivo: i contorni non sono nella realtà così regolari ed inoltre gli spessori presentano una certa tolleranza rispetto al valore che nominalmente dovrebbero avere. Queste non idealità rendono necessario il rispetto, in fase di progetto, di un insieme di regole o di vincoli geometrici che sono classificabili in tre categorie:

- minime dimensioni, al di sotto delle quali si rischia di avere interruzioni delle piste (per esempio, se l'incertezza per i bordi della metal è di 0.5μ e la metal è spessa 1μ, si rischia di avere interruzioni della pista);
- minime distanze (non rispettare le distanze può provocare cortocircuiti indesiderati a causa delle incertezze sui bordi):
- minima sovrapposizione (per avere la sicurezza di ottenere contatti perfetti è necessario che gli strati da collegare siano sovrapposti almeno per una certa superficie).

Il rispetto delle regole di progetto garantisce di avere una resa accettabile, dove per resa si intende la percentuale di circuiti funzionanti sul totale dei circuiti prodotti. Il numero di circuiti funzionanti decresce con il crescere della complessità del progetto ovvero tipicamente si ha una resa:

- dell'ordine del 70-80% per circuiti semplici;
- dell'ordine del 50% per circuiti di media complessità;
- dell'ordine 5-10% per circuiti molto complessi.

PRODUZIONE DI SILICIO POLICRISTALLINO



In un crogiolo, schematizzabile come in figura e dotato di un elettrodo di grafite che fornisce l'energia necessaria alla trasformazione chimica, vengono immessi silicio minerario (tipicamente quarzite cioè SiO₂) e carbonio (per esempio carbon-coke cioè carbonio liberato da idrocarburi leggeri). La temperatura massima che si raggiunge nel crogiolo è pari a 1800°C e dalla reazione si liberano ossido di carbonio (CO), acqua (H₂O), ossido di silicio (SiO) e silicio liquido; quest'ultimo si deposita sul fondo e che può essere estratto(si ricorda che la temperatura di fusione del silicio è pari a 1412°C).

La reazione chimica che si sviluppa è la seguente

Il silicio fuso che si ottiene da questa trasformazione è di tipo MGS (Metallurgical Grade Sylicon) ed ha un grado di purezza pari al 95%; le impurezze droganti hanno una concentrazione troppo elevata (44ppm(B)) per la produzione di circuiti integrati per cui è necessario silicio di tipo EGS (Electronical Grade Sylicon) in cui le impurità sono inferiori ad una parte su un miliardo (1ppb(B)). Chemical Vapor Deposition (CVD).

E' il processo che permette la produzione di EGS e necessitando di silicio aereiforme usa il triclorosilano (SiHCl₃); l'SiHCl₃ si ottiene sostituendo al silano tre atomi di H con tre atomi di Cl. La struttura del silano è la seguente

mentre la struttura del triclorosilano è

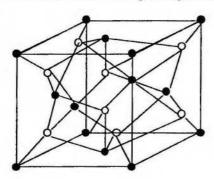
La reazione da cui si ottiene il triclorosilano è

 $Si+3HCl\rightarrow SiHCl_3+H_2\uparrow+calore$

Dalla trasformazione

si ottiene silicio puro a struttura policristallina ma per la produzione di circuiti integrati è necessaria una struttura monocristallina.

Per capire la differenza tra monocristallo e policristallo bisogna far riferimento al sistema di cristallizzazione del silicio che è cubico a facce centrate cioè ha una struttura cristallina cubica che ha un atomo su ogni vertice ed al centro di ogni faccia inoltre ogni vertice è collegato ad un atomo interno alla struttura che è anche collegato agli atomi delle facce adiacenti:



- O atomi interni alla struttura
- atomi dei vertici o atomi al centro delle facce

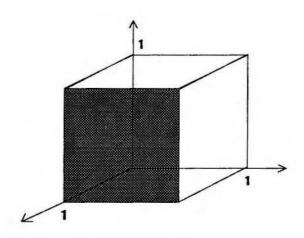
Il cristallo si definisce:

- monocristallo se la cella elementare è ripetuta nella struttura sempre con lo stesso orientamento;
- policristallo se la cella elementare viene ripetuta con diverso orientamento.

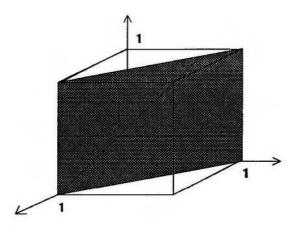
Si chiamano piani cristallografici le superfici notevoli con cui è possibile sezionare il cristallo.

I tagli possibili sono:

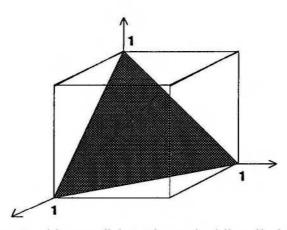
- taglio [1,0,0]



- taglio [1,1,0]



- taglio [1,1,1]



Il piano [1,1,1] ha la massima densità superficiale di atomi ed il taglio lungo questo piano è usato in tecnologia bipolare mentre il taglio [1,0,0] è usato dalla tecnologia mos (vale la pena puntualizzare che nella realtà vengono utilizzati tagli che differiscono di qualche grado da quelli visualizzati).

DIFETTI DI UN CRISTALLO

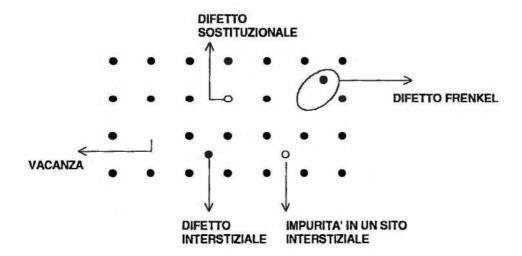
I difetti di un cristallo si distinguono nelle seguenti categorie spaziali:

- difetti puntuali
- difetti lineari
- difetti planari (o superficiali)
- · difetti di volume

Difetti del primo ordine o difetti puntuali

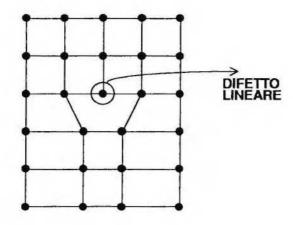
Il più semplice difetto puntuale è la vacanza o difetto schottky che è caratterizzato dalla mancanza di un atomo nel reticolo cristallino. Si parla di difetto sostituzionale, quando un atomo di silicio è sostituito da un atomo per esempio di boro (B), di difetto interstiziale, quando un atomo di silicio si posiziona in un minimo energetico (ma ovviamente ad energia superiore a quella del reticolo) circondato dagli atomi del reticolo stesso, di impurità in un sito interstiziale, quando l'atomo nel minimo energetico non è silicio ma per esempio boro.

La combinazione tra vacanza e difetto interstiziale è detta difetto Frenkel.



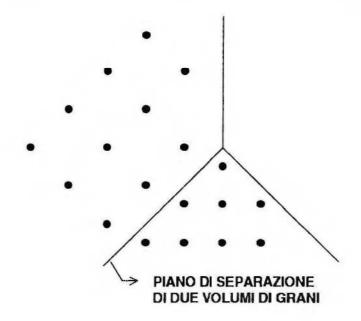
Difetti del secondo ordine o difetti lineari

Le distorsioni lineari, definite dislocazioni, sono causate da un atomo che è connesso reticolarmente solo a tre atomi a lui complanari; questi difetti si propagano in linea retta e si tratta di difetti molto gravi che provocano il non funzionamento del dispositivo.



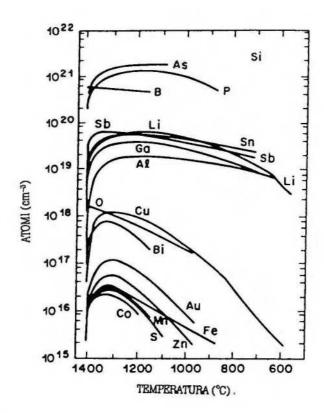
Difetti planari

Il difetto planare è valutato rispetto ai piani che delimitano i volumi dei grani, infatti per un piano che delimita due grani non si può parlare di struttura monocristallina bensì di struttura policristallina.



Difetti di volume

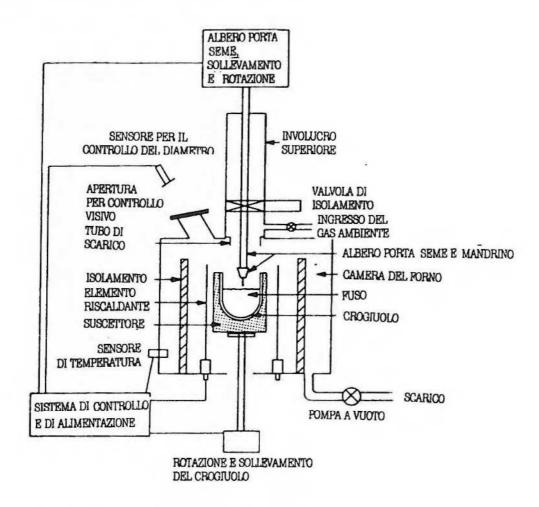
Per introdurre il discorso sui difetti di volume è utile riportare il grafico delle solubilità di alcuni tipi di impurità del silicio:



Supponiamo di avere del silicio alla temperatura T=1300°C in cui ci è presente del rame (Cu) con una concentrazione pari a 10¹⁷ cm⁻³; se il silicio viene raffreddato p.e. alla T=600°C la quantità di rame in esso presente supera la soglia di solubilità (vedi grafico) condensando e formando così un precipitato, cioè delle vere e proprie gocce di rame, all'interno del cristallo. Nei punti in cui si condensa il rame il reticolo si distorce rendendo il silicio inutilizzabile.

PRODUZIONE DI SILICIO MONOCRISTALLINO

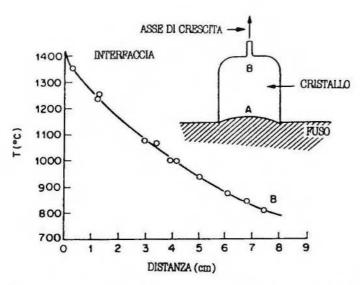
Tra i vari metodi esistenti per la produzione di silicio monocristallino il più diffuso è il metodo Tchochralski, questo processo utilizza il seguente impianto:



La struttura sopra riportata è costituita dai seguenti componenti:

- crogiolo;
- camicia del crogiolo;
- motore del crogiolo;
- asta;
- sistema di controllo;
- protezioni.

Il processo consiste nell'attaccare all'asta, con una particolare orientazione, un seme di silicio monocristallino purissimo, nel posizionare l'asta in modo tale da lambire il silicio fuso contenuto nella fornace ed infine nell'alzarla molto lentamente facendola ruotare: in questo modo all'interfaccia astaaria il silicio fuso si solidifica con una struttura cristallina che ha la stessa orientazione del seme intorno a cui cresce.



Gradiente di temperatura sperimentalmente determinato in un cristallo di silicio riferito al cristallo in crescita rappresentato nell'inserto. (Da deKoch e van de Wijggert, Ref. 14).

La reazione relativa al bilancio dei flussi di calore verso il cristallo all'interfaccia solido-liquido è la seguente:

$$L\frac{dm}{dt} + k_L \frac{dT}{dx_1} A_1 = k_S \frac{dT}{dx_2} A_2$$

dove

- Lè il calore di fusione;
- din dt rappresenta la velocità di solidificazione del silicio;
- k, è la conducibilità termica del liquido;
- k_s è la conducibilità termica del solido;
- dT/dx, rappresenta il gradiente di temperatura nel liquido;
- $\frac{dT}{dx_2}$ rappresenta il gradiente di temperatura nel solido;
- A₁ e A₂ sono le superfici isotermiche rispettivamente nel liquido e nel solido.

Chiamata d la densità di volume del liquido si ha che

$$dm = d dV = d A_2 dx$$

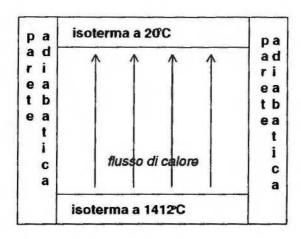
essendo

 $dV = A_2 dx$ (cioè volume = superficie \times innalzamento della quota del cilindro).

Si può dimostrare che il flusso di calore nel solido è costante. La situazione che si analizza è caratterizzata da due flussi:

- un flusso che va dal silicio fuso alla superficie isoterma a 1412°C
- un flusso che va dall'isoterma a 1412°C all'ambiente esterno

ed è cosi schematizzabile



Il flusso di calore φ è esprimibile come

$$\varphi = R_T \Delta T$$

dove $R_T = \frac{k_S H}{A}$ è la resistenza termica della struttura in cui

- ks, conducibilità termica del solido, è costante
- · A, sezione del crogiolo, è costante
- Hè in prima approssimazione costante

⇒ il flusso φ è costante perché prodotto di quantità costanti.

Nel processo di fabbricazione del silicio monocristallino riveste un ruolo fondamentale la velocità di estrazione dell'asta (cioè la quantità $\frac{dx}{dt}$), perché se questa aumenta provoca una diminuzione del gradiente di temperatura del liquido. Se il gradiente di temperatura del liquido, diminuendo, uguaglia quello del solido il silicio liquido non riesce a cedere calore e conseguentemente non si deposita più intorno al seme monocristallino.

Per questo esiste un limite superiore all'aumento di velocità di estrazione dell'asta ricavabile dalla seguente relazione

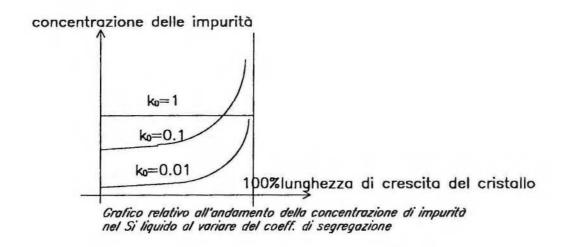
$$A_2 Ld \frac{dx}{dt}\Big|_{MAX} = k_S \frac{dT}{dx_2} A_2 \Rightarrow v_{MAX} = \frac{k_S}{Ld} \frac{dT}{dx_2}$$

la v_{MAX} per un cilindro di diametro pari a 75mm è dell'ordine dei $5\frac{mm}{min\ uto}$ ed in generale v_{MAX} è inversamente proporzionale alle dimensioni del diametro del cilindro (infatti più il cilindro è grande più aumenta la resistenza termica e quindi diminuisce la quantità $\frac{dT}{dx_2}$).

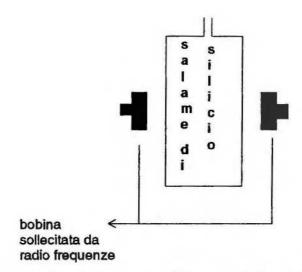
Si definisce coefficiente di segregazione $k_0 = \frac{C_S}{C_L}$ la quantità di specie drogante presente nel silicio solidificato. Il seguente esempio può aiutare a capire il significato di k_0 : facendo solidificare un cm³ di silicio drogato con una quantità di As pari a $10 \frac{\text{at om i}}{\text{cm}^3}$ si ottiene un volume di Si contenente 3 atomi di As (essendo $k_0|_{\text{ARSENICO}} = 0.3$).

elemento drogante	coefficiente di segregazione		
alluminio (Al)	0.002		
arsenico (As)	0.3		
boro (B)	0.8		
carbonio (C)	0.07		
rame (Cu)	0.0004		
ferro (Fe)	0.000008		
ossigeno (O)	1.25		
fosforo (P)	0.35		
antimonio (Sb)	0.023		

Legato al coefficiente di segregazione è l'andamento della concentrazione delle impurità al crescere della lunghezza del cristallo; questo avviene perché il silicio solidificandosi porta via solo una parte delle impurità del corrispondente volume di liquido per cui con l'avanzare del processo la concentrazione di impurità nel liquido cresce e conseguentemente aumenta anche la presenza di impurità nel silicio solido che successivamente si forma.



Il salame di silicio monocristallino che si ottiene alla conclusione del processo viene ripulito dalle impurità facendolo scorrere attraverso una bobina sollecitata con radiofrequenze; le radiofrequenze inducono nel salame di silicio delle correnti che via via fondono fettine di silicio permettendo la deposizione delle impurità sul fondo.



Il cristallo viene spostato molto lentamente all'interno della bobina per permettere una risolidificazione che mantenga la struttura monocristallina.

Alla fine della procedura la parte alta del pezzo di silicio è molto più pura della parte sottostante in cui si sono convogliate le impurità.

Vale la pena puntualizzare che la fornace in cui si fonde il silicio è caratterizzata da un crogiolo di grafite ricoperto da biossido di silicio (o, come talune volte sperimentalmente accade, ricoperto da nitruro di silicio); questo accorgimento è indispensabile per evitare che il crogiolo, sottoposto a temperature elevatissime, fonda passando in soluzione con il silicio contaminandolo (questo inevitabilmente accadrebbe se fossero utilizzati semplicemente crogioli di grafite o di carburo di tantalio).

La copertura di SiO₂ durante la fusione del silicio si consuma in parte (cioè si trasforma in SiO gassoso che può essere estratto) ed alla fine la concentrazione di ossigeno nel liquido è molto bassa cioè dell'ordine dei 10-17 cm⁻³.

Un ulteriore accorgimento volto ad ottenere una orientazione perfetta è quello di posizionare sull'asta il seme intorno a cui il silicio cresce attraverso un sistema a raggi x altamente preciso.

Una volta prodotto il salame di silicio monocristallino questo viene privato della testa, perché con diametro irregolare, e della coda, perché ricca di impurità, per poi venire sottoposto ad una azione di tornitura allo scopo di ottenere un cilindro perfetto.

Successivamente si effettuano dei tagli flat sul cilindro monocristallino ottenuto (cioè si tolgono dal cristallo delle lunette) per permettere di identificare automaticamente sia il tipo di drogaggio sia il tipo di taglio cristallografico.



drogaggio n taglio cristallografico 100



drogaggio p taglio cristallografico 100



drogaggio n taglio cristaliografico 111



drogaggio p taglio cristallografico 111

I tagli flat si distinguono in primari, cioè più profondi, e secondari, cioè meno profondi.

In un secondo tempo il silicio viene tagliato in fette dello spessore di 1mm con una sega a diametro interno (ovvero con una sega circolare a lama interna tagliente) per rovinare il meno possibile il cristallo.

La fetta è poi sottoposta a particolari accorgimenti fisico-meccanici quali:

- l'eliminazione degli spigoli vivi essendo questi i punti più fragili cioè quelli più soggetti a fratture;
- la lucidatura.

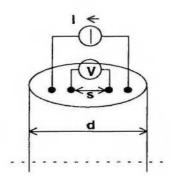
Per lucidare si usa un feltro di SiO₂ (biossido di silicio) e NaOH (idrossido di sodio) umidi su cui viene fatta ruotare la fetta. La lucidatura si svolge in due fasi, una più grossolana ed una più fine, ed ha come obiettivo quello di migliorare il più possibile le qualità della superficie di silicio: se infatti il piano cristallografico fosse perfetto i dispositivi integrati su di esso sarebbero di altissima qualità (uno dei fattori che influiscono sulla resa del processo, che varia a seconda della complessità del circuito tra 0.35÷0.8 è infatti legata ai difetti superficiali).

Un aspetto non trascurabile è il costo del processo produttivo: il diametro delle fette € attualmente intorno ai 100÷125 mm ma la tendenza è quella di aumentarlo in modo da produrre un numero più elevato di chips per passata di processo con conseguente diminuzione del costo del singolo chip.

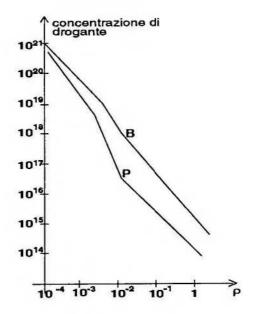
Lo spessore della fetta a conclusione dei trattamenti analizzati varia da 0.5 a 0.7mm e la diminuzione rispetto al valore originario è principalmente imputabile alla lucidatura.

Poiché un'alta concentrazione di metalli di tipo pesante (p.e. Fe) determina un elevato tasso di ricombinazione e perciò alte correnti di perdita, vengono creati, per mezzo di un laser, dei difetti sul fondo della fetta in modo tale che questi, essendo centri di cattura, forzino i residui di metallo pesante a concentrarsi nel bulk, dove non creano problemi.

Ultimo passo è quello di verificare se la concentrazione di drogante ottenuta è quella voluta, per far ciò si effettuano della misure sulla resistività del cristallo attraverso il seguente circuito:



Ricavata la resistività ρ dalla relazione $\rho = \frac{V}{I} 2\pi s$ (d>10s) si risale al valore della concentrazione della specie drogante attraverso tabelle del tipo:



PROCESSO DI FABBRICAZIONE DEI CHIPS

Chemical Vapor Deposition (CVD) e Plasma-assisted Chemical Vapor Deposition (PCVD).

La deposizione di films di opportuno spessore è utilizzata nella produzione di circuiti VLSI per ottenere per esempio isolamento elettrico tra le metals o per proteggere il chip dalle contaminazioni chimiche dell'ambiente esterno.

I metodi per la deposizione dei materiali devono essere sicuri, riproducibili, facilmente automatizzabili ed economici. Tra i processi più diffusi si ricordano:

- · CVD (atmospheric-pressure, chemical vapor deposition);
- · LPCVD (low-pressure, chemical vapor deposition);
- PCVD (plasma-assisted, chemical vapor deposition).

La reazione chimica che regola la deposizione dell'SiO, è la seguente:

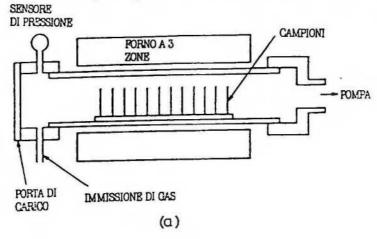
$$SiCl_2H_2 + 2N_2O \rightarrow SiO_2 + 2N_2 + 2HCl$$
 $T = 900^{\circ}C$

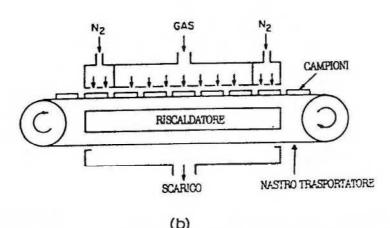
dove

- SiCl₂H₂ è la formula chimica del diclorosilano;
- N₂O è la formula chimica dell'ossido d'azoto;
- HCl è la formula chimica dell'acido cloridrico.

La deposizione del polisilicio si ottiene per pirolisi con una reazione che avviene alla temperatura di $600/650^{\circ}$ C e che scinde il silano in silicio policristallino, che si deposita, ed H_2 .

Per attuare il processo di chemical vapor deposition sono comunemente usati i seguenti reattori:





Lo schema a) rappresenta un reattore a parete riscaldata e pressione ridotta (hot-wall, reduced-pressure reactor); esso è costituito da un tubo di quarzo riscaldato in cui il gas è immesso da una estremità ed è estratto dall'estremità opposta attraverso una pompa. La pressione durante la reazione varia tra i 30 e i 250 Pascal mentre la temperatura appartiene ad un intervallo che va dai 300°C ai 900°C.

Nel reattore i wafers sono tenuti in posizione verticale da sostegni di quarzo affinché il flusso di gas circolante nel tubo incida perpendicolarmente sulle fette. Durante ciascuna passata di processo sono trattati dai 50 ai 200 wafers.

Lo schema b) mostra un reattore a gettito continuo e a pressione atmosferica (continuous-troughtput, atmosferic-pressure reactor), in esso le fette sono spostate da un nastro trasportatore e sono scaldate per convezione. Il gas necessario per la reazione è immesso nel reattore attraverso un'apertura centrale ed è arginato da due barriere gassose ciascuna costituita da N₂.

I reattori a parete riscaldata e a pressione ridotta garantiscono una eccellente uniformità, grandi carichi ma hanno lo svantaggio di ottenere basse percentuali di deposizione e di utilizzare gas tossici, corrosivi e altamente infiammabile.

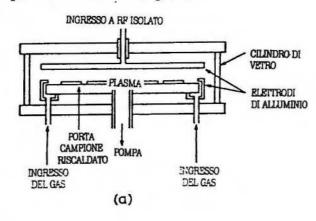
I reattori a gettito continuo ed a pressione atmosferica risolvono i problemi causati dai reattori "hotwall" ma hanno la svantaggio di richiedere flussi di gas molto veloci e frequenti puliture.

Poiché i processi appena analizzati sono fortemente dipendenti dalla temperatura, si possono manifestare effetti secondari di disturbo tra passi contigui di uno stesso processo. Si supponga per esempio di effettuare una diffusione che porti ad un esatto posizionamento degli atomi: se per passare alla fase successiva del processo è necessario alzare la temperatura accade che gli atomi appena diffusi vengono indesideratamente allontanati dalle posizioni raggiunte.

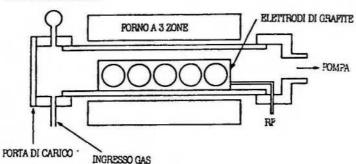
Per limitare gli effetti collaterali legati alla crescita della temperatura si utilizzano processi che sfruttano plasma invece di gas (PCVD – Plasma Chemical Vapor Deposition) e che lavorano a temperature più basse (la scissione del silano con un processo CVD avviene a T=600°C mentre con un processo PCVD a T=200°C).

PCVD

Gli schemi dei reattori utilizzati per la PCVD sono i seguenti:







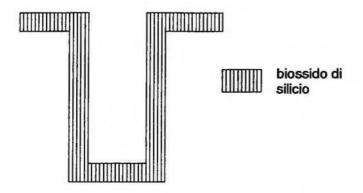
(b)

La figura a) rappresenta un reattore radiale ad armature parallele (radial-flow, parallel-plate, plasma-assisted CVD reactor); la camera di reazione è un cilindro di vetro o di alluminio fornito di due armature. Le fette sono adagiate sull'elettrodo inferiore che è cortocircuito a massa ed è riscaldato alla temperatura di 100/400°C mentre all'elettrodo superiore è applicata una tensione a radiofrequenza. Nel cilindro il gas è immesso attraverso due ingressi laterali e viene fatto scorrere verso il centro dell'armatura dove è estratto da una pompa. Questo tipo di reattore è utilizzato per la deposizione di biossido di silicio e di nitruro di silicio.

La figura b) rappresenta un reattore PCVD a parete riscaldata (hot-wall, plasma-assisted CVD reactor) in cui la reazione avviene in un tubo di quarzo scaldato da una fornace. Le fette sono tenute verticalmente e sono parallele al flusso del gas; il montaggio dell'elettrodo, che sorregge le fette, contiene lunghe piastre di grafite o di alluminio che sono collegate alla tensione di alimentazione in modo alternato.

Step coverage.

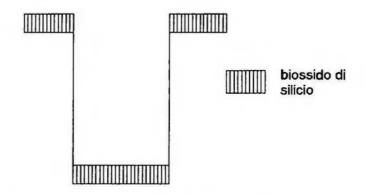
L'alternanza di processi di deposizione e di attacco chimico crea sulla fetta di silicio, originariamente piana, dei gradini che devono essere ricoperti da uno strato di SiO₂.



coverage del gradino con uno strato uniforme di SiO

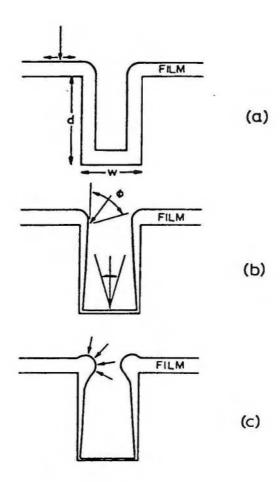
Alla conclusione della deposizione dell'SiO₂ si vorrebbe una copertura uniforme (conformal step coverage) ma questa è difficile da ottenere.

Attuando una deposizione completamente verticale l'SiO₂ non riuscirebbe a depositarsi sulle pareti verticali e si otterrebbe come risultato finale:



risultato di un coverage verticale dell'SiQ

Le possibili configurazioni della copertura dei gradini sono le seguenti:

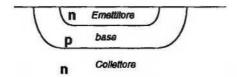


La configurazione a) si ottiene se la deposizione avviene a temperatura elevata perché il materiale è fluido e la copertura diventa uniforme per migrazione superficiale delle particelle. Se la migrazione delle particelle non è significativa lo spessore di deposizione è proporzionale all'angolo di incidenza delle molecole di gas:

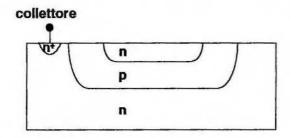
in figura b) si ha un esempio di deposizione in assenza di migrazione ed in cui le dimensioni del gradino da coprire sono più grandi del cammino libero medio delle molecole, mentre in figura c) si ha un esempio di deposizione in assenza di migrazione e con breve cammino libero medio delle molecole.

Epitassia.

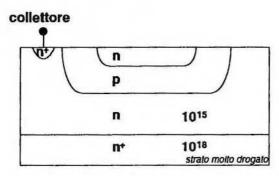
Con l'epitassia si va a crescere il substrato stesso, questo processo è nato per produrre transistori bipolari di potenza la cui struttura è grossomodo la seguente:



Per ottenere un transistore bipolare di potenza bisogna avere un dispositivo con alta tensione base-collettore che sostenga anche alte correnti: queste due esigenze sono in contrasto perché per avere alte tensioni di rottura base-collettore è necessario disporre di un collettore poco drogato ma questo significa avere un collettore altamente resistivo. Il problema nasce dal fatto che nei transistori bipolari, usati in circuiti integrati come stadi di potenza, il contatto di collettore sta in basso e l'alta resistenza distribuita che lo caratterizza è rappresentata dallo strato inferiore drogato n:



La soluzione è quella di diminuire tale resistenza aggiungendo sotto la struttura del transistore bipolare uno strato molto drogato:



Le resistenze nei circuiti integrati sono resistenze distribuite e perciò risulta laborioso calcolare il loro valore: per determinare la resistenza tra due elettrodi di un dispositivo è necessario rilevare i percorsi della corrente che vanno da un elettrodo all'altro (cioè valutare in modo approssimato le linee di forza del campo elettrico nella struttura) ed integrare la resistività rispetto a detti percorsi (questo metodo vale anche se l'andamento della resistività non è continuo).

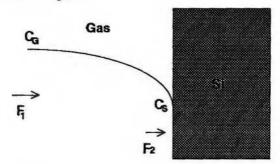


Per ottenere una struttura con uno strato inferiore più drogato ed uno superiore meno drogato, senza ricorrere all'epitassia, sarebbe necessario utilizzare un substrato con una concentrazione di As (arsenico) pari a 10^{18} e fare una diffusione con una densità di boro (che è di polarità opposta all'arsenico) pari a 10^3 ($10^{15} = 10^{18} - 10^3$). In questo modo la parte superficiale avrebbe una concentrazione di drogante pari a 10^{15} per la ricombinazione degli atomi di boro con quelli di arsenico.

Il processo appena descritto non è però utilizzabile perché la quantità di boro diffusa non è controllabile (cioè è praticamente impossibile ottenere le concentrazioni volute nelle zone volute) per cui è necessario ricorrere all'epitassia.

Il processo di epitassia prevede l'utilizzo di un monocristallo di antimonio (solitamente si sceglie questo elemento perché diffonde poco) con una concentrazione di drogante pari a 10¹⁸ su cui si fa una deposizione di silicio per pirolisi (cioè cottura) del tetraclorosilano (SiCl₄). Il silicio si deposita sul substrato monocristallino e replica, sotto certe condizioni, la struttura sottostante facendo perciò crescere il cristallo. Se durante il processo di deposizione al tetraclorosilano è aggiunta una opportuna quantità di gas drogante il cristallo crescerà con la voluta concentrazione di impurità. La reazione che regola questo processo è la seguente:

L'andamento della concentrazione di silicio nel gas è:



In riferimento alle quantità visualizzate in figura:

- C_G è la concentrazione di silicio nel gas lontano dall'interfaccia;
- C_s è la concentrazione superficiale all'interfaccia di silicio;
- F, è il flusso di particelle di tetraclorosilano nel gas per unità di superficie;
- F₂ è il flusso attraverso la superficie all'interfaccia per unità di superficie.

I flussi F_1 e F_2 sono espressi dalle seguenti espressioni:

$$F_1 = h_G(C_G - C_S) \tag{1}$$

$$F_2 = k_s C_s \tag{2}$$

Poiché si è in condizioni stazionarie $F_1 = F_2 = F$ dalla relazione (1) = (2) si può ricavare una formula per la concentrazione superficiale all'interfaccia di silicio cioè

$$C_{s} = \frac{C_{g}}{1 + \frac{k_{s}}{h_{g}}}$$

dove

- k_s è una costante relativa alla velocità della reazione chimica che avviene all'interfaccia nell'epitassia (dipendente dalla pressione e dalla temperatura);
- h_G è la costante di trasferimento di massa (riguarda la diffusione dell'SiCl₄ nell'atmosfera).

La velocità di crescita, cioè il rapporto tra lo spessore di silicio che si deposita sul substrato ed il tempo necessario per la formazione di detto strato, è esprimibile con la seguente relazione

$$v = \frac{F}{N_1} = \frac{k_s h_G C_G}{(k_s + h_G) N_1}$$

Il valore N_1 rappresenta la quantità di silicio per unità di volume e si riferisce al silicio solido (mentre C_G si riferisce al silicio gassoso).

Poiché $C_G = YC_T$ (essendo C_T la concentrazione di molecole della miscela gassosa costituita da H_2 e tetraclorosilano e Y la frazione molare del tetraclorosilano) si può esprimere la velocità di crescita con la seguente espressione:

$$v = \frac{k_S h_G Y C_T}{(k_S + h_G) N_1}$$

Se $k_s \ll h_G$ si dice che il processo è a controllo da reazione superficiale e $v = \frac{k_s Y C_T}{N_1}$.

Si parla di processo a controllo da trasferimento di massa se $h_G \ll k_S$ cioè se $v \equiv \frac{h_G Y C_T}{N_s}$.

Con l'epitassia si ottiene il silicio opportunamente drogato utilizzando una miscela di tetraclorosilano (SiCl₄) e gas drogante, se a tal scopo si usa l'arsenico il gas impiegato è AsH₃ e la reazione risulta essere:

$$AsH_3 + SiCl_4 + 2H_2 \rightarrow Si + 4HCl + AsH_3$$

in cui l'AsH3 viene assorbito dalla superficie e migra per diffusione.

Il silicio che si ricava con l'epitassia è molto più puro di quello che si ottiene con il metodo Tchochralski perché non contiene ossigeno. Tuttavia, per dispositivi mos, questa tecnica è poco usata in quanto, essendo un processo abbastanza costoso, è utilizzato solo per alcuni tipi pregiati di fet quali VMOS.

Ossidazione.

Un metodo per ottenere strati isolanti è l'ossidazione che si distingue in:

ossidazione secca, regolata dalla reazione

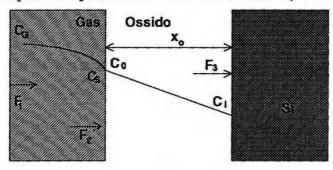
$$Si + O_2 \rightarrow SiO_2$$

ossidazione termica umida, regolata dalla reazione

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

Attraverso l'ossidazione secca si ottiene uno strato di biossido di silicio che, per le sue caratteristiche perfette, può essere reso molto sottile ed utilizzato sotto il gate. L'SiO₂ che si ottiene da una reazione umida è invece più grossolano e viene quindi utilizzato come ossido di campo.

Lo schema base relativo all'ossidazione termica del silicio (che solitamente avviene a temperature tra i 700°C ed i 1300°C e con pressioni parziali che vanno da 0.2 a 1 atm) è il seguente



La figura rappresenta il substrato di silicio ricoperto da uno strato d'ossido, che è a diretto contatto con la fase gassosa, ed in essa sono evidenziabili tre flussi:

- F₁ è il flusso che trasporta la specie ossidante verso l'interfaccia gas-ossido;
- F₂ è il flusso che trasporta la specie ossidante verso il silicio, attraversando lo strato di ossido esistente;
- F₃ è il flusso con cui, all'interfaccia ossido-silicio, la specie drogante reagisce formando l'SiO₂.

 F_1 può essere approssimato linearmente, assumendo che il flusso di ossidante che dal gas si sposta verso l'interfaccia gas-ossido sia proporzionale alla differenza tra la concentrazione di ossidante nel gas (C_G) e la concentrazione di ossidante adiacente alla superficie dell'ossido (C_S) , cioè

$$F_1 = h_G(C_G - C_S)$$

dove h_G è il coefficiente relativo al trasferimento di massa della fase gassosa.

Per trovare la relazione che lega la concentrazione di specie drogante all'equilibrio nell'ossido (C_o) a quella in fase gassosa (C*) è necessario ricorrere alla legge di Henry per cui:

$$C_o = H p_s$$

 $C^* = H p_g$

dove

- H è la costante di Henry;
- p_s è la pressione parziale della specie ossidante all'interfaccia;
- p_G è la pressione parziale della specie ossidante nel gas.

$$C_i \rightarrow 0 \text{ e } C_o \rightarrow C^* \text{ mentre se } D \rightarrow \infty \quad C_i = C_o = \frac{C^*}{1 + \frac{k_s}{h}}.$$

Definita N₁ la concentrazione di specie ossidante nel gas, si può esprimere F₃ come

$$F_3 = N_1 \frac{dx_0}{dt} \qquad (1)$$

poiché vale anche la relazione

$$F_3 = k_s C_i = k_s \frac{C^*}{1 + \frac{k_s}{h} + \frac{k_s x_o}{D}}$$
 (2)

Uguagliando l'equazione (1) con la (2), risolvendo l'equazione differenziale che ne risulta e ponendo come condizione iniziale che $x_0 = x_i$ per t=0 si ricava che

$$x_{o}(t) = \frac{A}{2} \left(\sqrt{1 + \frac{t + \tau}{\frac{A^{2}}{4B}}} - 1 \right)$$

in cui

A = 2D
$$\left(\frac{1}{k_s} + \frac{1}{h}\right)$$
, B = $\frac{2D C^*}{N_1}$, $\tau = \frac{x_i^2 + A x_i}{B}$.

A commento del processo di ossidazione analizzato si può evidenziare il fatto che il biossido di silicio si forma consumando il silicio del substrato; gli atomi di ossigeno che si depositano si combinano con atomi di silicio che diffondono verso l'alto.

Diffusione

La diffusione di atomi di impurità nel silicio è importante nel processo VLSI infatti viene utilizzata, nella tecnologia bipolare, per creare basi, emettitori e resistenze, mentre, nella tecnologia MOS, per ottenere regioni di sorgente e di collettore ed inoltre per drogare il polisilicio.

I metodi per introdurre atomi di drogante nel silicio sono i seguenti:

- 1. per diffusione da una sorgente chimica in forma di vapore a temperature elevate;
- 2. per diffusione da una sorgente d'ossido drogato;
- 3. per diffusione e successiva ricottura da uno strato di ioni impiantati.

Il processo di diffusione in un solido è caratterizzato da movimenti atomici, attraverso lacune o posizioni interstiziali, della specie che diffonde nel reticolo cristallino.

Un modello monodimensionale per le equazioni di diffusione è stato introdotto da Fick nel 1855: la sua teoria si fonda sull'analogia tra il trasporto di materiale in una soluzione e la trasmissione di calore per conduzione.

Il flusso monodimensionale che rappresenta il trasferimento di atomi in soluzione è rappresentato dalla seguente equazione:

$$J = -D \frac{\partial C(x, t)}{\partial x}$$

dove

- Jè il flusso di diffusione:
- C è la concentrazione del soluto;
- x è l'asse coordinato nella direzione del flusso;
- tè il tempo di diffusione;
- Dè il coefficiente di diffusione.

Il segno negativo della relazione che esprime J, cioè la velocità di diffusione del soluto per unità di area, indica che la diffusione avviene nella direzione di concentrazione di soluto decrescente (ovvero dove il gradiente $\frac{\partial C}{\partial x}$ è negativo).

Per la legge di conservazione della materia vale la relazione $\frac{\partial C(x,t)}{\partial t} = -\frac{\partial J(x,t)}{\partial x}$

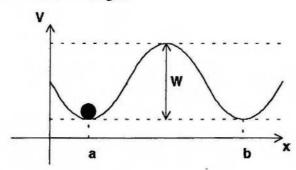
$$\frac{\partial C(x,t)}{\partial t} = -\frac{\partial J(x,t)}{\partial x}$$

che, sostituendo l'equazione di J, diventa

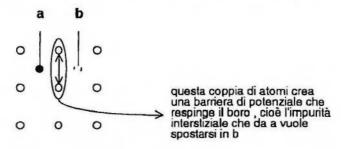
$$\frac{\partial C(x,t)}{\partial t} = D \frac{\partial^2 C(x,t)}{\partial x^2}$$

Si vuole a questo punto trovare un'espressione per il coefficiente di diffusione e a tale scopo è utile fare un esempio.

Si consideri il seguente andamento dell'energia:



La rappresentazione periodica dell'andamento dell'energia è ragionevole essendo il reticolo cristallino periodico: la particella che si trova in x = a sta su un minimo energetico e per spostarsi in x = b deve superare una barriera di potenziale di ampiezza pari a W. La situazione energetica appena analizzata potrebbe rappresentare il reticolo di silicio con impurità interstiziale (p.e un atomo di boro) in x = a:



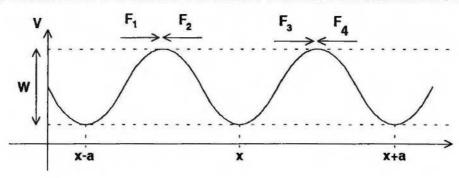
Allo zero assoluto le particelle del reticolo sono immobili e l'impurità di boro rimane fissa in x = a, ma, in presenza di agitazione termica, questa particella può "schizzare via", ad esempio colpita da un fotone, e portarsi in x = b.

La probabilità di transizione, cioè la probabilità che l'atomo passi da x = a ad x = b, è espressa in questo modo

$$v = v_0 e^{\frac{qW}{kT}}$$

in cui v_0 è la frequenza di oscillazione, ovvero la frequenza dei tentativi effettuati per superare la barriera W e portarsi in x = b, solitamente $v_0 \approx 10^{14}$ e W è pari a qualche frazione di elettronvolt.

Si vede che per $T\to 0$ (zero assoluto, essendo T espressa in gradi Kelvin) $e^{\frac{qW}{kT}}\to 0$ e quindi la probabilità di transizione v è nulla mentre se T si alza anche $e^{\frac{qW}{kT}}$ cresce con conseguente aumento di v. Per calcolare il coefficiente di diffusione è necessario analizzare una successione di barriere:



Le espressioni dei flussi F₁, F₂, F₃ ed F₄ sono le seguenti:

- il flusso da x-a ad x è $F_1 = a C(x-a) v_0 e^{\frac{qW}{kT}}$;
- il flusso da x ad x-a è $F_2 = a C(x) v_0 e^{\frac{qW}{kT}}$;
- il flusso da x ad x+a è $F_3 = a C(x) v_0 e^{\frac{-qW}{kT}}$;
- il flusso da x+a ad x è $F_4 = a C(x+a) v_0 e^{\frac{qW}{kT}}$.

Il flusso \vec{F} all'ascissa x può essere espresso come media tra i flussi \vec{F}_1 ed \vec{F}_4 , in quanto \vec{F}_2 ed \vec{F}_3 si compensano avendo stessa intensità e direzione ma verso opposto:

$$\vec{F} = \frac{\vec{F}_1 + \vec{F}_4}{2} = \frac{1}{2} \left(\vec{F} \left(x + \frac{a}{2} \right) + \vec{F} \left(x - \frac{a}{2} \right) \right)$$

L'intensità di F è data dalla seguente relazione

$$F = \frac{F_1 - F_4}{2} = a v_0 e^{\frac{-qW}{kT}} (C(x-a) - C(x+a))$$

approssimando poi la concentrazione con una serie di Taylor del 1° ordine si ha che

$$C(x \pm a) = C(x) \pm a \frac{\partial C}{\partial x}$$

infine, utilizzando quest'ultima equazione, il modulo del flusso diventa

$$F = a^2 v_0 e^{-\frac{qW}{kT}} \left(-\frac{1}{2} \left(\frac{\partial C}{\partial x} + \frac{\partial C}{\partial x} \right) \right) = -a^2 v_0 e^{-\frac{qW}{kT}} \frac{\partial C}{\partial x}$$

L'espressione di F va equiparata con $J = -D \frac{\partial C(x,t)}{\partial x}$ e dal confronto risulta che

$$D = a^2 v_0 e^{\frac{qW}{kT}}$$

Si riportano a seguito i valori di D, per T=1100°C, di alcuni elementi:

- $\square \qquad D=5.3 \ 10^{-11} \ \frac{\text{cm}^2}{\text{sec}} \text{ per il gallio (Ga);}$
- $\square = 9.9 \ 10^{-17} \frac{\text{cm}^2}{\text{sec}} \text{ per l'antimonio (Sb)};$
- D= 2.9 $10^{-16} \div 2 \cdot 10^{-13} \frac{\text{cm}^2}{\text{sec}}$ per il fosforo(P);
- D= 3.4 $10^{-17} \div 2 \cdot 10^{-14} \cdot \frac{\text{cm}^2}{\text{sec}}$ per il boro (B);
- D= 1.2 $10^{-16} \div 3.5 \ 10^{-15} \frac{\text{cm}^2}{\text{sec}}$ per l'arsenico (As).

Dai dati sopra riportati si può notare che, per alcune specie droganti, il valore di D varia all'interno di un certo range perché dipende dalla concentrazione: quanto più è alta la concentrazione, a parità di gradiente, tanto più basso è il coefficiente di diffusione o, in altre parole, il cammino libero medio della particella di specie drogante.

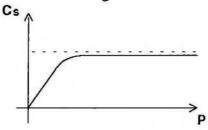
La diffusione si articola in due fasi distinte:

- la fase di predeposition, consiste nella deposizione CVD della specie drogante sul silicio;
- la fase di annealing, consiste in una ricottura della fetta di silicio che,in atmosfera inerte, determina la diffusione in profondità del sottile strato di specie drogante precedentemente deposto.

I controlli da effettuare durante la diffusione riguardano:

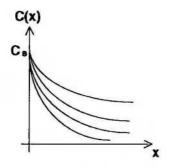
- 1. la quantità di specie drogante nella fase di deposizione;
- la profondità di diffusione nella fase di ricottura.

In fase di deposizione l'andamento della concentrazione superficiale in funzione della pressione parziale di specie drogante è qualitativamente il seguente:

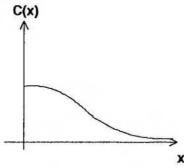


L'andamento di C_s , oltre un certo valore di P, tende al limite di solubilità solida rappresentata in figura con una linea tratteggiata.

Alla fine della fase di predeposizione la concentrazione è la seguente



mentre dopo la fase di annealing le particelle di specie drogante diffondono anche piu in profondità ed il profilo della concentrazione diventa:



La concentrazione di specie drogante che si ha all'inizio della fase di annealing è dipendente sia dal tempo di predeposizione t sia dalla profondità dello strato di specie drogante deposto, in particolare essendo

$$\begin{cases} C(x,0) = 0 \\ C(x = 0,t) = C_s \end{cases}$$
 condizioni iniziali
$$C(\infty,t) = 0 \ \forall t$$

la soluzione dell'equazione $\frac{\partial C(x,t)}{\partial t} = D \frac{\partial^2 C(x,t)}{\partial x^2}$ che le soddisfa è la seguente:

$$C(x,t) = C_s \operatorname{erfc}\left(\frac{x}{2\sqrt{Dt_{pd}}}\right)$$

dove

- t_{pd} è il tempo di predeposizione che si approssima con un impulso di pari area;
- $\operatorname{erfc}(x) = 1 \frac{2}{\sqrt{\pi}} \int_{0}^{x} e^{-a^{2}} da$

Una approssimazione della funzione C(x, t), che suppone la quantità di drogante deposta per unità di superficie costante, è la seguente

$$C(x,t) = \frac{Q}{\sqrt{\pi Dt}} e^{-\frac{x^2}{4Dt}}$$

in cui ponendo x=0 si ottiene che

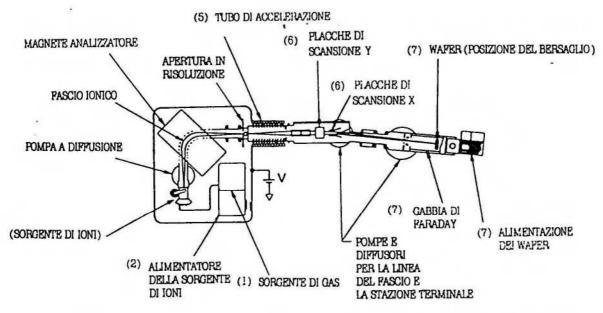
$$C(0,t) = \frac{Q}{\sqrt{\pi Dt}}$$

Impiantazione ionica

Nel processo di diffusione appena analizzato si è visto che le grandezze da controllare, per ottenere il profilo di drogaggio desiderato, sono sia il tempo di deposizione sia il tempo di ricottura. Durante la fase di predeposizione è difficile controllare in modo preciso il numero di atomi di drogante impiantati mentre questo è attuabile se si ricorre all'impiantazione ionica.

Il processo di impiantazione ionica spara atomi ionizzati sul wafer, con energia sufficiente a farli penetrare sotto la regione superficiale; con questa tecnica si possono ottenere precise concentrazioni di drogante che vanno da 10^{14} a 10^{21} $\frac{\text{at om i}}{\text{cm}^3}$.

Un sistema commerciale di impiantazione ionica è il seguente



Schema di un tipico sistema commerciale per impiantazione ionica. (Da Varian-Extrion, DF-3000 Brochure).

dove si identifica con

- (1) la sorgente gassosa, solitamente di BF₃ o di AsH₃, ad alto potenziale accelerante V ed in cui il flusso di ioni è regolato da una valvola;
- (2) l'alimentatore ad alta tensione che ha il compito di "energizzare" la sorgente di ioni;
- (3) la sorgente di ioni contenente il plasma desiderato, per esempio costituito da +As⁷⁵ oppure da +B¹¹ e ad una pressione di 10⁻³ torr;
- (4) il magnete analizzatore che seleziona solo le specie desiderate e respinge le altre;
- (5) il tubo di accelerazione, attraverso cui viene fatto passare il fascio;
- (6) le placche di deflessione, soggette a tensioni con andamento a "dente di sega", che, deviando le particelle sia lungo l'asse x sia lungo l'asse y, permettono agli ioni la scansione uniforme della superficie di silicio;

(7) la camera di bersaglio, costituita da un'apertura che definisce la zona bersaglio, da *una gabbia* di Faraday e da un meccanismo per l'alimentazione del wafer.

La quantità di ioni (Q) deposta sul wafer attraverso l'impiantazione ionica è calcolabile effettuando l'integrazione nel tempo della corrente di ioni che viene accelerata contro il bersaglio di silicio, per questo motivo Q è perfettamente regolabile.

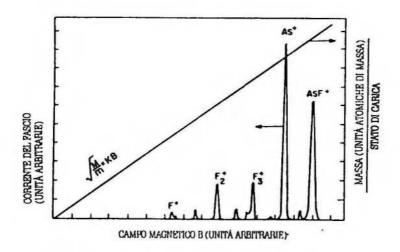
Il magnete deflette le particelle secondo la legge

$$R B = \sqrt{\frac{2VM}{mq}}$$

dove

- R è il raggio dell'orbita;
- B è il campo magnetico;
- V è la tensione di accelerazione;
- M è la massa dello ione;
- m è la valenza;
- q e la carica dell'elettrone.

Regolando il campo magnetico B si seleziona il passaggio delle specie ioniche desiderate attraverso la fenditura da cui si accede al tubo di accelerazione. In particolare aumentando B, ovvero l'intensità del campo, cresce il peso specifico degli ioni a cui è consentito il varco:

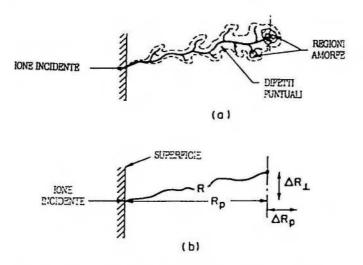


Uno spettro di massa tipico di una sorgente di gas di AsF₃. La corrente del bersaglio ed il campo magnetico sono espresse in unità arbitrarie. Sono anche indicate le specie principali ionizzate. Il rapporto fra massa e stato di carica è determinato dalla linea retta

L'energia di accelerazione fornita agli ioni varia da 3+500 KeV ed è in grado di rompere fino ad un migliaio di legami reticolari (il silicio ha un legame reticolare di energia pari a 4 eV ma, poiché ha una struttura tetravalente, per spostarlo sono necessari almeno 16 eV).

Uno ione impiantato subisce un processo di diffusione a causa degli urti con gli atomi del wafer e la sua energia diminuisce fino a che esso si arresta.

La penetrazione della particella nel silicio provoca la creazione di difetti puntuali e di zone amorfe disordinate:



a) Disordini "ad albero" di uno ione impiantato. b) Schema del "range" ionico R, del "range" proiettato R_p, incertezze di R_p o dispersione proiettata ΔR_p e dispersione laterale ΔR_⊥.

La lunghezza totale del percorso ionico viene definita ionic range (R). Tipicamente uno ione si ferma ad una distanza, perpendicolare alla superficie del wafer, pari ad R_p , definita range proiettato e rappresentante la profondità di penetrazione: gli ioni che incontrano meno ostacoli, cioè che sono coinvolti in un basso numero di urti, si spingono oltre R_p mentre ioni coinvolti in un numero elevato di urti si fermano prima di R_p . La fluttuazione rispetto al range proiettato viene definita ΔR_p , mentre la variazione del punto di arresto dello ione nella direzione ortogonale ad R_p è pari alla dispersione laterale ed è definita ΔR_1 . La distribuzione della concentrazione degli ioni impiantati in funzione della posizione è una gaussiana centrata in R_p e così espressa

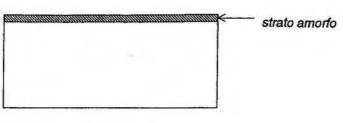
$$n(x) = n(R_p) e^{\frac{(x-R_p)^2}{2\Delta R_p^2}}$$

dove ΔR_p è la deviazione standard o dispersione della distribuzione ed in cui la concentrazione massima si ha per $x = R_p$. La massima concentrazione $n(R_p)$ è esprimibile come:

$$n(R_p) = \frac{\phi}{\sqrt{2\pi} \Delta R_p}$$

Per T elevate si manifesta un fenomeno detto back scattering, questo termine si riferisce alla diffusione all'indietro degli ioni di drogante causata dall'urto di questi ultimi contro atomi oscillanti del reticolo cristallino: in presenza di back scattering la distribuzione di carica non è più di tipo gaussiano ma è meglio approssimata da una distribuzione di tipo Pearson IV. Il valore del range proiettato e della deviazione della distribuzione gaussiana forniscono, infatti, una buona rappresentazione del primo ordine per ioni impiantati in substrati amorfi o policristallini a grani di piccole dimensioni. Alcune distribuzioni, nonostante l'approssimazione gaussiana sia abbastanza fedele in prossimità del picco, presentano accentuate assimmetrie e code della distribuzione per cui è necessario utilizzare formulazioni che si servono di momenti di ordine più elevato.

Poiché l'impiantazione ionica fa diventare il primo strato di silicio amorfo, a causa della rottura dei legami reticolari dovuta alla penetrazione degli ioni, è necessario ricuocere la fetta per ottenere una crescita epitassiale del

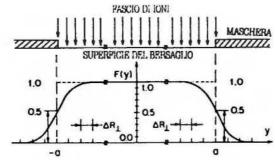


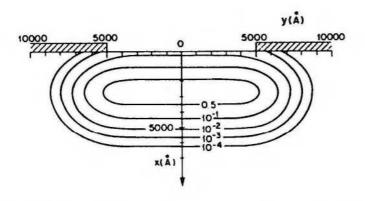
cristallo amorfo da fase solida, che avviene alla T per esempio di 1200°C.

La crescita epitassiale può avvenire da fase solida poiché lo strato amorfo, molto drogato, è assimilabile ad un gas molto concentrato. L'aumento di T determina perciò una ricostruzione dei legami reticolari del silicio con conseguente diffusione verso il basso del drogante: a conclusione della ricottura il silicio ha qualche imperfezione ma buona distribuzione della concentrazione di drogante.

Il costo finale di questa tecnica è, nel complesso, basso perché la resa è buona ed inoltre Q è perfettamente controllabile, perché calcolabile dall'integrazione nel tempo della corrente.

Il processo di impiantazione ionica utilizza, per effettuare un'impiantazione selettiva, opportune maschere ma, nonostante questo accorgimento, non si riesce ad evitare che i profili delle impiantazioni debordino oltre il limite che teoricamente dovrebbero imporre le maschere. Questo effetto è dovuto alla dispersione laterale degli ioni ed una sua possibile rappresentazione è riportata in figura:





Rappresentazione dei profili laterali. a) Concentrazione di ioni lungo la direzione laterale (y) per una maschera di gate con a > $> \Delta$ R \perp e direzione infinita nella direzione z. b) Contorni di equiconcentrazioni ioniche per B * 70 keV (R_p = 2710 Å, Δ R_p = 824 Å e Δ R \perp = 1006 Å) incidente in silicio attraverso una fessura di Iµm. (Da Furukawa, Matsumura e Ishiwara, Ref. 22).

La relazione che quantifica il profilo effettivo della densità di distribuzione ionica è la seguente

$$n(x,y) = \frac{n(x)}{2} \left[erfc \left(\frac{y-a}{\sqrt{2} \Delta R_{\perp}} \right) - erfc \left(\frac{y+a}{\sqrt{2} \Delta R_{\perp}} \right) \right]$$

dove

- n(x) è la densità di distribuzione in profondità, lontano dal bordo della maschera;
- 2a è l'ampiezza della finestra della maschera;
- y è l'asse diretto lungo la direzione laterale.

Dall'espressione di n(x, y) si può constatare sia che la concentrazione di drogante tende a zero via via che ci si allontana dal bordo della finestra, sia che, se a>> R_{\perp} , n(x,a) = $\frac{n(x)}{2}$.

Gli ioni incidenti sulla fetta di silicio possono subire due tipi di urti, ovvero possono scontrarsi contro gli elettroni oppure contro i nuclei del reticolo cristallino, ed a seguito di ogni urto la loro energia diminuisce. Secondo la teoria di Lindhard, Scharff e Schiott i meccanismi di perdita di energia sono tra loro indipendenti ed additivi per cui la diminuzione di energia per unità di lunghezza può essere espressa come

$$\frac{dE}{dx}\Big|_{\text{totale}} = \frac{dE}{dx}\Big|_{\text{nuclears}} + \frac{dE}{dx}\Big|_{\text{elettronico}}$$

Il range espresso in funzione dell'energia dello ione è dato dalla relazione:

$$R(E) = \int_{0}^{R(E)} dx$$

Considerando che lo ione si arresta quando la sua energia è nulla ed effettuando un cambiamento di variabile si ha che

$$R(E) = \int_{0}^{R(E)} dx = \int_{E}^{0} \frac{dx}{dE} dE = -\int_{0}^{E} \frac{dE}{\left| dE/dx \right|_{\text{totale}}} = \frac{1}{N} \int_{0}^{E} \frac{dE}{S(E)}$$

in cui

- N è la densità del materiale ovvero rappresenta il numero di "atomi bersaglio" per cm³;
- S(E) è definito stopping power e rappresenta l'energia di rallentamento;
- E è l'energia iniziale degli ioni incidenti.

Tra gli effetti collaterali, che disturbano l'impiantazione ionica, si ha il channelling o canalizzazione degli ioni: questo fenomeno si manifesta quando una particella di drogante incide sul silicio in corrispondenza di uno spazio aperto tra due file di atomi. Una volta che lo ione si è inserito, le forze direttrici dei potenziali atomici delle file di elementi del reticolo lo spostano verso il centro dello spazio vuoto e lo guidano stabilmente lungo il canale per una considerevole distanza. Per questo motivo le distanze coperte dagli ioni canalizzati sono molto maggiori di quelle percorse da ioni non canalizzati, che, come tali, sono coinvolti in un elevato numero di urti. Per diminuire l'effetto di channeling si sparano gli ioni lungo orientazioni dense di atomi ma ciò non toglie la possibilità di entrare, casualmente, in uno di questi canali anche in una fase successiva della penetrazione nella fetta.

La tabella in cui sono rappresentati, per differenti tensioni di accelerazione, lo ionic range e le fluttuazioni ΔR_p e ΔR_{\perp} del punto di arresto degli ioni di alcune specie droganti è la seguente:

Tabella 2 Valore dei quattro momenti e △R⊥.

loni		E (keV)			
	Parametro	10	30	100	300
В	R,	382	1065	3070	6620
	R_p ΔR_p	190	390	690	1050
	γ1	-0.32	-0.85	~1.12	-1.59
	B ₂	3.2	4.49	5.49	8.35
	ARL	190	465	871	1523
P	R _a	150	420	1350	4060
	R_p ΔR_p	78	195	535	1150
	γ, ΄	0.45	0.20	-0.37	-0.91
	γι Β ₂	3.4	3.1	3.26	4.89
	ARI	61	168	471	1097
As	R.	110	233	678	1946
	R _p ΔR _p	40	90	261	667
		0.57	0.46	0.45	0.30
	Υı B ₂	3.6	3.4	3.4	3.16
	ARL	33	64	187	481
Sb	ΔRp	100	208	507	1303
	ΔR _p	30	62	158	390
		0.54	0.51	0.40	0.18
	Υı B ₂	3.5	3.5	3.3	3.1
	ARL	23	46	108	266

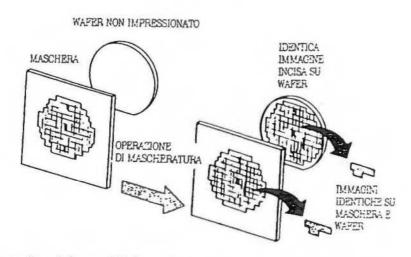
Nota: AR L Furukawa? riportato in scala a AR, di Fichtner25.

Nella fabbricazione di dispositivi si fa spesso uso di processi in cui sulla superficie dei bersagli sono presenti rivestimenti protettivi; per esempio l'impiantazione può essere effettuata in ambiente ossidante in modo tale che gli ioni, incidendo sulla superficie, cioè al momento che hanno energia massima, incontrino un primo sottile strato di biossido di silicio che evita che la fetta sottostante subisca gravi danni. La presenza dell'ossido permette inoltre che il silicio danneggiato possa trasformarsi in biossido di silicio, combinandosi con atomi di ossigeno, che, grazie all'energia acquisita penetrano dentro al silicio, e possa essere poi lavato via. Nell'impiantazione in ambiente ossidante bisogna però calibrare il processo di ossidazione al fine di evitare di lavare via anche il silicio profondo, oltre a quello effettivamente danneggiato.

Litografia

La litografia è il processo che trasferisce le forme geometriche delle maschere, fabbricate per realizzare un dato circuito, sul wafer di silicio. Il primo livello nella fabbricazione di un circuito integrato è la generazione dei modelli delle forme geometriche, che si ottengono suddividendo in differenti livelli il disegno composito del circuito. Sistemi di visualizzazione grafica interattiva e digitalizzatori convertono il disegno geometrico in dati digitali che servono a pilotare un generatore di disegni controllato al calcolatore. Il generatore di disegni, che solitamente è una macchina a fascio elettronico, trasferisce le figure geometriche su una piastra di vetro fotosensibile chiamata fotomaschera o maschera.

Il circuito integrato finale viene prodotto trasferendo sequenzialmente il disegno da ogni maschera, livello per livello, alla superficie del wafer di silicio come rappresentato in figura:



Processo di trasferimento delle immagini da una maschera ad un wafer di silicio

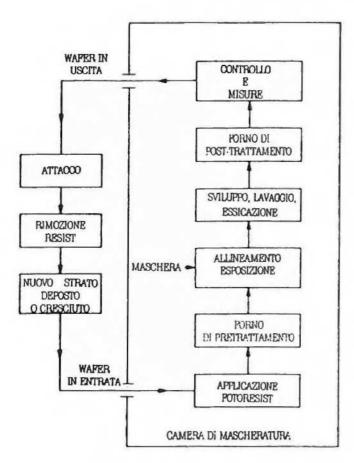
Un tempo il processo di fabbricazione delle maschere consisteva nel disegnare prima un insieme di forme geometriche su grande scala, tipicamente di dimensioni dalle 100 alle 2000 volte più grandi delle dimensioni reali, che veniva poi suddiviso in modo tale da avere, per ogni livello di mascheratura, un disegno. Questo insieme di figure veniva poi ridotto fotograficamente di 10 volte su di un reticolo di vetro ed infine ulteriormente diminuito per essere portato alle dimensioni effettive.

Per evitare il disegno su larga scala, sono stati sviluppati sistemi CAD le cui uscite, che sono rappresentazioni digitali dei layout dei circuiti, vengono memorizzate su un nastro magnetico ed utilizzate per pilotare il generatore di disegni.

In commercio si trovano maschere realizzate sia con piastre di vetro ad emulsione, quali le piastre ad alta risoluzione della Kodak, sia con piastre di vetro ricoperte da materiali a superficie dura:

 le prime sono meno costose ma sono utilizzabili solo in applicazioni circuitali che si servono di tecnologie a 5μm; le seconde sono più care, perché realizzate con macchine a fascio elettronico e con materiali
quali cromo, ossido di cromo, ossido di ferro o di silicio ma possono essere utilizzate con
tecnologie a 1μm.

Il flusso seguito dai wafer di silicio all'interno dell'area di processo litografico si può cosi schematizzare:



Flusso dei waser di silicio attraverso l'area di processo litografico di una produzione di circuiti integrati.

Il wafer pulito, su cui, per prima cosa, devono essere realizzate le diffusioni profonde o wells, accede al diagramma di flusso relativo al processo litografico in corrispondenza del blocco "nuovo strato deposto o cresciuto".

Il wafer entra poi nella camera di mascheratura in cui:

- gli viene applicato uno strato di fotoresist dello spessore di c.a. 1μm;
- passa in un forno di pretrattamento in cui il resist si essica;
- viene esposto a luce ultravioletta attraverso una fotomaschera;
- viene sviluppato in una soluzione che fa sciogliere il resist;
- viene immesso in un ambiente che attacca l'ossido spesso ma non il resist, un solvente adatto a questo scopo è l'acido fluoridrico tamponato (BHF);
- il resist viene infine asportato ma lascia la sua impronta sull'ossido di silicio, quest'ultima è usata come maschera per la deposizione o la crescita dello strato seguente (infatti un

processo di impiantazione ionica droga solo il silicio esposto e non quello protetto dal biossido di silicio).

La procedura continua allineando la fotomaschera successiva sul disegno presente sul silicio e ripetendo il trasferimento litografico; i circuiti VLSI richiedono dalle 5 alle 11 mascherature separate, e conseguenti fasi di trasferimento litografico, al fine di produrre un dispositivo funzionante.

Andando più nel dettaglio del processo di mascheratura si può distinguere tra:

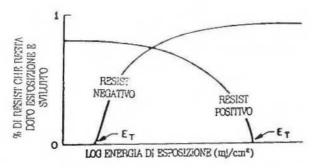
- mascheratura a contatto, che prevede l'appoggio della maschera sulla fetta, si tratta della tecnica migliore perché minimizza la diffrazione, ma si può provocare il danneggiamento della maschera (se per esempio la fetta ha superficialmente un difetto di volume quest'ultimo danneggia la maschera irrimediabilmente rendendola inutilizzabile);
- mascheratura a prossimità, che mantiene la maschera ad una certa distanza dalla superficie del wafer, risolvendo il problema dell'eventuale danneggiamento della maschera, questa tecnica è però affetta da diffrazione;
- mascherature a proiezione, che prevede la proiezione della figura utilizzando un sofisticato, e
 come tale molto costoso, obiettivo fotolitografico monocromatico, cioè sintonizzato su una
 lunghezza d'onda.

Durante la litografia è necessario prestare molta attenzione a limitare i difetti, la resa può, infatti, facilmente cadere a zero, in quanto i difetti possono essere introdotti in ciascuna delle fasi del processo: se, per esempio, diventa difettoso il 10% del chip ad ogni trasferimento, la realizzazione di un dispositivo che richiede 11 mascherature originerà un chip in cui solo il 31% dell'area sarà funzionante.

Precisando meglio le operazioni del processo fotolitografico, si può iniziare col descrivere la tecnica usata per l'applicazione del fotoresist. E' interessante premettere che l'ambiente in cui si deposita il fotoresist è illuminato da luce gialla poiché i fotoresist non sono sensibili a lunghezze d'onda che superano i 5000 Å. Il fotoresist è solitamente centrifugato sulla fetta di silicio, il wafer è perciò montato su una piattaforma ruotante a cui è fissato mediante vuoto; alcune gocce del resist liquido sono depositate sulla superficie del wafer che poi è accelerato sino al raggiungimento di una velocità di rotazione prefissata, a cui è mantenuto per una trentina di secondi.

Il pretrattamento in forno serve a rimuovere il solvente del resist ed ad aumentare l'adesione dello strato al wafer, l'esposizione alla luce ultravioletta, che avviene alla fase successiva, deve essere preceduta da un attento allineamento della maschera alle forme già esistenti sul silicio ed è seguita dallo sviluppo del resist.

E' importante specificare che esistono due tipi di resist cioè si distingue tra resist positivo e resist negativo. I resist negativi diventano meno solubili in soluzioni di sviluppo se sono esposti alla radiazione mentre quelli positivi diventano più solubili dopo l'esposizione, le tipiche risposte all'esposizione sia di resist positivi che negativi sono riportate in figura:



Il comportamento del resist in fase di sviluppo ha una motivazione organica legata al fatto che l'esposizione alla radiazione ultravioletta provoca la variazione della lunghezza delle molecole del resist e conseguentemente della loro solubilità in un determinato solvente.

Le risposte sopra riportate hanno un andamento dipendente da tutte le variabili che regolano l'utilizzazione del resist nel contesto del trasferimento litografico e cioè: spessore iniziale del resist, distribuzione spettrale della radiazione di esposizione, condizioni di pretrattamento, composizione chimica della soluzione di sviluppo, tempo di sviluppo, etc.

Dall'analisi degli andamenti sopra riportati risulta che il resist positivo richiede più energia di esposizione rispetto a quello negativo, cioè deve essere sottoposto ad una esposizione più prolungata, per questo motivo i processi litografici che usano fotoresist positivo sono meno produttivi.

L'obiettivo del processo litografico è quello di riportare fedelmente l'immagine della maschera sul resist e da qui sul wafer di silicio in pratica però questo non avviene, si riporta perciò una figura in cui sono rappresentate le immagini che si ottengono sul resist a conclusione della fase di sviluppo:



Un punto particolarmente delicato del processo è legato all'allineamento delle maschere rispetto alle forme geometriche già trasferite sul silicio, infatti le immagini appartenenti ai vari livelli di mascheratura sono tra loro spazialmente dipendenti: per esempio le geometrie di metallizzazione devono ricoprire completamente le aperture dei contatti. Per garantire che le sovrapposizioni dovute avvengano, è necessario, in fase di progetto, prevedere una certa ampiezza di tolleranza per garantire l'allineamento.

L'ampiezza di tolleranza minima è calcolata considerando che:

- i bordi delle geometrie del dispositivo sul wafer di silicio possono discostarsi da quelli specificati dal progetto originale;
- si ha incertezza nell'allineamento delle immagini presenti su un dato livello di mascheratura rispetto a quelle del livello precedente a causa di errori di allineamento automatico delle maschere.

Il processo litigrafico è ripetuto fino a che si giunge alla deposizione di uno strato di plexiglass che ha lo scopo di proteggere il chip appena fabbricato. Una volta che il wafer è pronto diventa necessario suddividerlo in singoli die che, se elettricamente validi, sono disposti in una configurazione utile per la successiva operazione di assemblaggio. Per effettuare il taglio del wafer solitamente si utilizza una sega diamantata in quanto garantisce un taglio con bordi netti e riduce considerevolmente le possibilità di frammentazione e di rottura sia per incisione parziale sia per separazione totale.

Come già accennato non tutti i chip prodotti vengono montati sul case, solo quelli marcati in nero sono utilizzati in quanto quelli marcati in rosso sono guasti; i circuiti funzionanti sono a loro volta suddivisi in classi di merito a seconda del funzionamento in frequenza (i processori, per esempio, anche se lavorano a frequenze diverse non sono prodotti separatamente).

Per connettere il chip al case sono realizzate delle piazzuole dette *pad*, appartenenti al chip perché realizzate con piste di metal, su cui viene effettuata la saldatura con il filo di connessione ai pin del package:

