

Introduction

L'objectif du cours *Circuits Microelectroniques* est de présenter les techniques de conception des circuits intégrés CMOS VLSI. La conception de circuits électroniques complexes a débuté dans les années 50 avec le développement du RADAR et des premiers ordinateurs. Elle s'est ensuite développée, surtout aux Etats Unis, dans l'industrie et dans les départements d'Electrical Engineering des grandes universités américaines, au cours des années 70 et 80 avec l'apparition des microprocesseurs VLSI nMOS puis CMOS.

La conception des circuits intégrés se fait principalement via deux approches. La première consiste à considérer la technologie VLSI comme une simple évolution technologique de la conception des cartes électroniques. Ces techniques de conception des circuits intégrés peuvent être vues comme une transposition de celles utilisées pour la conception des cartes électroniques. De telles approches sont utilisées par les techniques de conception "rapide" telle que celle des circuits prédiffusés et des circuits programmables (FPGA, EPLD,...). La seconde approche consiste à se rendre compte que la technologie VLSI possède ses propres spécificités qui peuvent être exploitées pour réaliser des circuits beaucoup plus optimisés et plus performants. La circuiterie des circuits VLSI devient alors une nouvelle discipline qui s'appuie sur de nouveaux concepts (logique d'interrupteurs, rétention dynamique de l'information, transparence des bascules, circuits polyphasés,...) et qui intègre les aspects topologiques qui conduisent au dessin des masques. Cette relation entre la fonctionnalité des blocs et leur topologie ouvre une dimension tout à fait nouvelle dans la conception des circuits électroniques.

La maîtrise de la conception des circuits intégrés VLSI est une condition nécessaire au développement d'une industrie électronique performante. Une chaîne de fabrication permet en effet la réalisation de beaucoup de circuits différents. Le développement de nombreux circuits spécialisés par l'industrie européenne nécessite de nombreux concepteurs, ce qui ouvre de larges perspectives professionnelles dans cette discipline.

L'étude de la circuiterie VLSI fait appel à tout un ensemble de notions qui ne font pas partie des cursus habituels de physique et d'informatique. Certains ne sont jamais enseignés. Ils font partie du savoir-faire des concepteurs de circuits intégrés complexes, comme par exemple la technique des circuits polyphasés.

Une étude approfondie de l'ensemble des sujets abordés nécessiterait un volume horaire beaucoup plus important. Nous avons donc choisi de ne traiter que les éléments importants de chaque domaine pour pouvoir disposer d'un point de vue global sur cette discipline et permettre son approfondissement ultérieur.

Chapitre 1 :

Du silicium aux portes de base

(par Yvan Bonnassieux, Ecole Polytechnique)

1 Du silicium aux portes de base

1.1 Introduction aux systèmes numériques

Les Sciences et Technologies de l'Information et de la Communication (STIC) ont pour objet de transmettre, traiter et stocker de grandes quantités d'information.

Ces opérations doivent être exécutées de façon fiable, et donc être robustes en présence de bruit. Il est pour cela fait appel aux systèmes numériques dans lesquels l'information est représentée par des valeurs discrètes d'un signal continu qui varie en fonction du temps, de l'espace ou d'une autre variable.

1.1.1 Logique binaire de niveaux de tension

Le plus souvent, pour maximiser l'immunité au bruit, on n'utilise que deux symboles notés "0" et "1" pour représenter les états du signal, il s'agit d'une logique binaire :

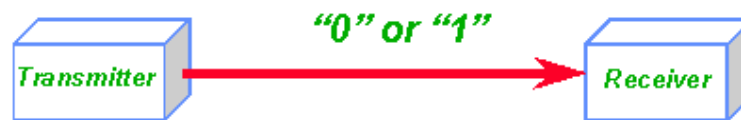


Fig. 1.1 Canal de transmission binaire

La grandeur physique la plus fréquemment utilisée pour représenter ces deux symboles est une tension électrique pour laquelle il faut définir les valeurs nominales des tensions correspondant aux états "0" et "1", ainsi que les tolérances associées.



Fig. 1.2 Etats logiques transmis

On se place le plus souvent dans le cas de la logique "positive" où l'état "1" est associé au niveau "haut" de tension, aussi appelé niveau "actif".

Il est essentiel qu'une chaîne de plusieurs circuits, où chaque sortie alimente l'entrée suivante, et considérée dans un état statique, soit également robuste au bruit : il ne faut pas qu'il y ait de dégradation progressive du signal. Cette condition, appelée discipline statique, impose que chaque élément de la chaîne restaure des signaux reçus à la marge de la zone de validité. Chaque élément doit ainsi avoir des contraintes de validité sur ses sorties plus sévères que celles sur ses entrées. Il doit donc avoir une caractéristique de transfert non linéaire.

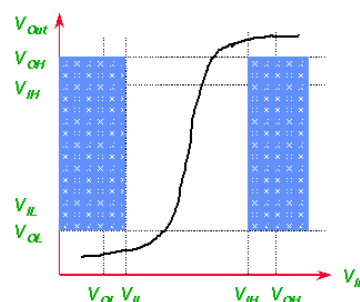


Fig. 1.3 Caractéristique de transfert d'un élément

On définit ainsi :

- Les niveaux d'entrée limites admissibles V_{il} et V_{ih}
- Les niveaux de sortie garantis V_{ol} et V_{oh} (avec $V_{ol} < V_{il}$ et $V_{oh} > V_{ih}$)
- Les marges de bruit $NM_l = V_{il} - V_{ol}$ et $NM_h = V_{oh} - V_{ih}$

L'obtention d'une telle courbe de transfert, avec un gain supérieur à 1, est le plus souvent obtenue en utilisant des transistors comme éléments amplificateurs. Chacun d'eux se comporte essentiellement comme une source de courant commandée par une tension (transconductance). Elle implique un apport d'énergie fournie par les alimentations électriques du système.

1.1.2 Le rôle du Temps

Le plus souvent, les signaux logiques considérés sont fonction du temps t et dépendent les uns des autres par des relations de causalité. Dans un monde idéal, on pourrait imaginer que des actions puissent avoir un effet immédiat, mais dans le monde réel, il y a toujours, pour des raisons physiques, un certain délai (ou retard), fini et positif, entre une action et sa conséquence.

Plusieurs approches sont possibles pour prendre en compte ce problème :

Dans les **circuits asynchrones**, on traite le temps comme une variable continue et on modélise chaque élément du circuit par un élément parfait (instantané) suivi d'un retard pur dont la valeur est soit connue, soit encadrée par deux limites. Il faut alors analyser le fonctionnement du circuit en tenant compte de toute la complexité des combinaisons possibles des effets des retards.

Dans les **circuits synchrones**, on traite au contraire le temps de manière discontinue au moyen d'une horloge (temps discret) qui déclenche le fonctionnement de chaque composant, et on s'arrange pour que la période de l'horloge soit supérieure au plus grand retard possible dans le circuit. Dans un tel système échantillonné, et en s'imposant cette contrainte de fréquence maximale de l'horloge, on masque complètement le problème des retards.

On distingue également les circuits combinatoires des circuits séquentiels :

Dans un **circuit combinatoire**, les signaux de sortie à l'instant t ne dépendent que de l'état des signaux présents sur les entrées du circuit au même instant t (à l'effet des retards près). Ils ne dépendent pas de l'évolution antérieure des entrées : le circuit n'a pas de mémoire. Cette caractéristique importante implique l'absence de boucle de rétroaction dans le système. Dans de nombreux cas, on ne considérera que les états stationnaires ou quasi stationnaires du circuit.

Cette situation s'oppose à celle des **circuits séquentiels** dans lesquels les signaux de sortie d'un circuit à l'instant t dépendent non seulement des entrées au même instant, mais aussi de leur évolution antérieure (de leur historique). Un tel effet de mémoire est obtenu en introduisant des boucles de rétroaction dans le système.

Lorsque le signal électrique réel doit être traité comme une fonction continue d'un temps lui-même continu, il est commode de le caractériser par des paramètres t_r et t_f décrivant ses temps de montée et de descente (de 10% à 90% de son amplitude), et de caractériser le système par les temps de retard t_{phl} et t_{plh} à mi-amplitude d'une sortie par rapport à une entrée pour chacune des transitions étudiées.

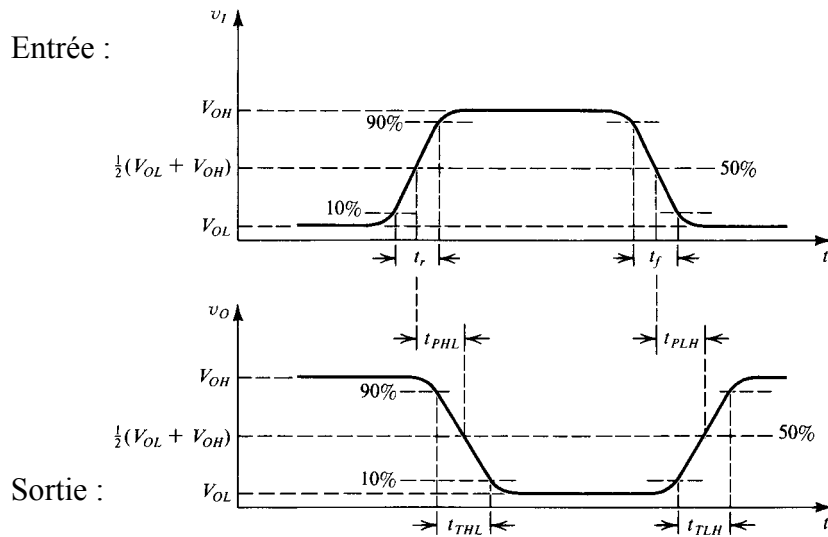
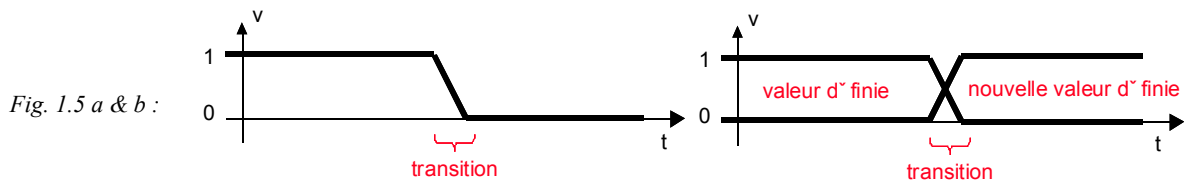
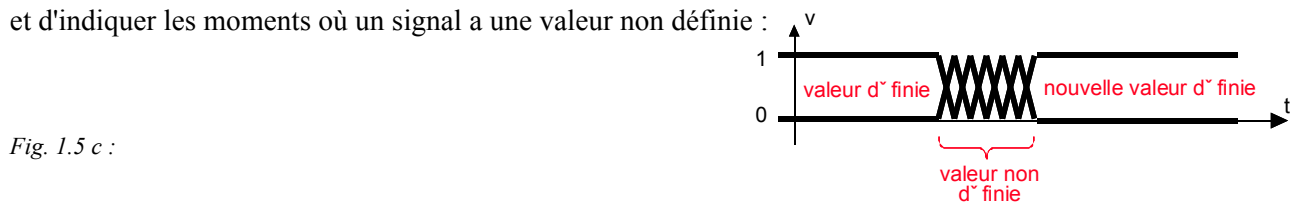


Fig. 1.4 Caractéristiques temporelles des signaux

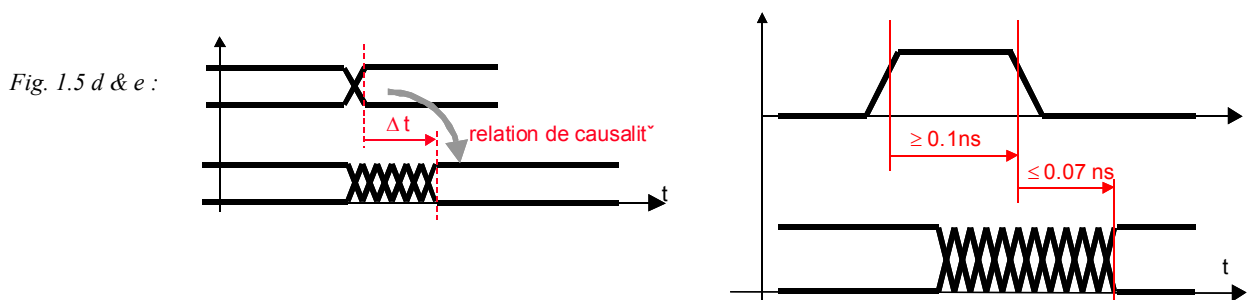
Cette description des signaux réels est habituellement faite de façon stylisée en utilisant le **langage des chronogrammes** qui permet une représentation simplifiée de l'évolution d'un signal. Il est ainsi possible de représenter une transition entre deux états connus ou non précisés :



et d'indiquer les moments où un signal a une valeur non définie :



Ainsi que les relations de causalité et les contraintes temporelles :



1.1.3 Phénomènes physiques

Plusieurs phénomènes physiques placent des contraintes sur les circuits logiques, et nous avons vu le rôle du temps dans la section précédente avec les limitations sur la vitesse de fonctionnement qui en découlent.

Les échanges de signaux électriques se traduisent également par la circulation de courants et une dissipation de puissance thermique.

Pour chaque connexion au sein d'un système de modules interconnectés, les courants consommés par les entrées des circuits placés en aval doivent être fournis par la sortie du circuit placé en amont qui leur délivre le signal. Des règles de charge doivent être respectées pour que la somme des courants consommés n'excède pas la capacité de fourniture de la source de signal.

Enfin la puissance dissipée par effet Joule doit pouvoir être évacuée en respectant les températures maximales de fonctionnement des circuits.

1.1.4 Couches conceptuelles

Un système complet ne s'étudie pas en le réduisant à l'ensemble des équations de son modèle électrique ou électronique. Le niveau de complexité de l'ensemble nécessite de passer par une ou plusieurs couches conceptuelles intermédiaires, de façon à avoir une approche modulaire et hiérarchique, dans laquelle la complexité de chaque sous-ensemble reste maîtrisable.

Comme indiqué dans l'introduction, il est possible d'effectuer la conception et l'étude d'un module, soit en assemblant des fonctions standard (plus ou moins complexes), soit en concevant un réseau ad hoc de transistors (traités comme des interrupteurs).

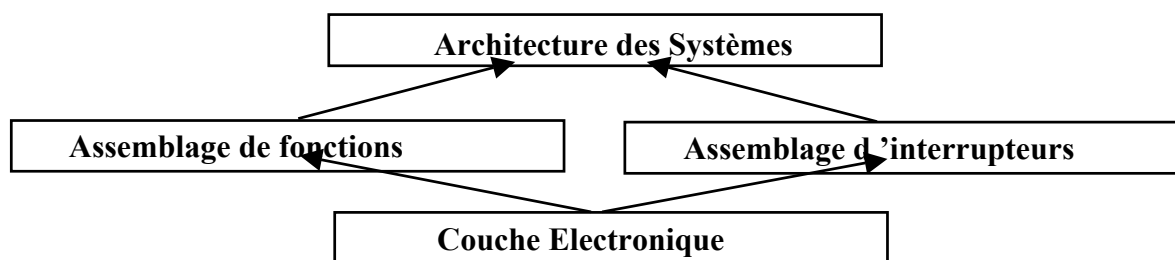


Fig. 1.6 Couches conceptuelles pour l'étude d'un système

1.2 Du matériau au transistor MOS

Les semi-conducteurs sont, comme leur nom l'indique, des éléments ou des composés dont la conductance (ou la résistivité) est intermédiaire entre celle des conducteurs et celle des isolants. Leurs propriétés électroniques, et celles des composants qu'ils permettent de réaliser, découlent de la structure quantique en bandes d'énergies permises ou interdites pour les électrons du matériau dans la structure périodique d'un cristal. Elles sont décrites de façon rigoureuse et détaillée dans le cours de physique statistique et de physique des solides, et dans le cours "Semi-conducteurs et composants" commun à la Majeure de Physique et à la Majeure ECS. L'approche choisie ici est d'admettre ces propriétés, en en expliquant les grandes lignes sans équations, de façon à pouvoir les utiliser pour comprendre le comportement des composants. Il est ainsi possible de modéliser ces derniers et de les utiliser pour bâtir des circuits.

1.2.1 Conductivité et dopage

La figure suivante montre la conductivité électrique σ (et la résistivité associée $\rho \equiv 1/\sigma$) pour un certain nombre de solides isolants, semi-conducteurs et conducteurs. Le quartz fondu est extrêmement isolant ($\sigma \approx 10^{-18}$ Siemens/cm), et l'aluminium, le cuivre et l'argent sont de très bons conducteurs ($\sigma \approx 4$ à $6 \cdot 10^5$ Siemens/cm). L'intervalle couvert est donc immense puisqu'il couvre une dynamique de 10^{24} .

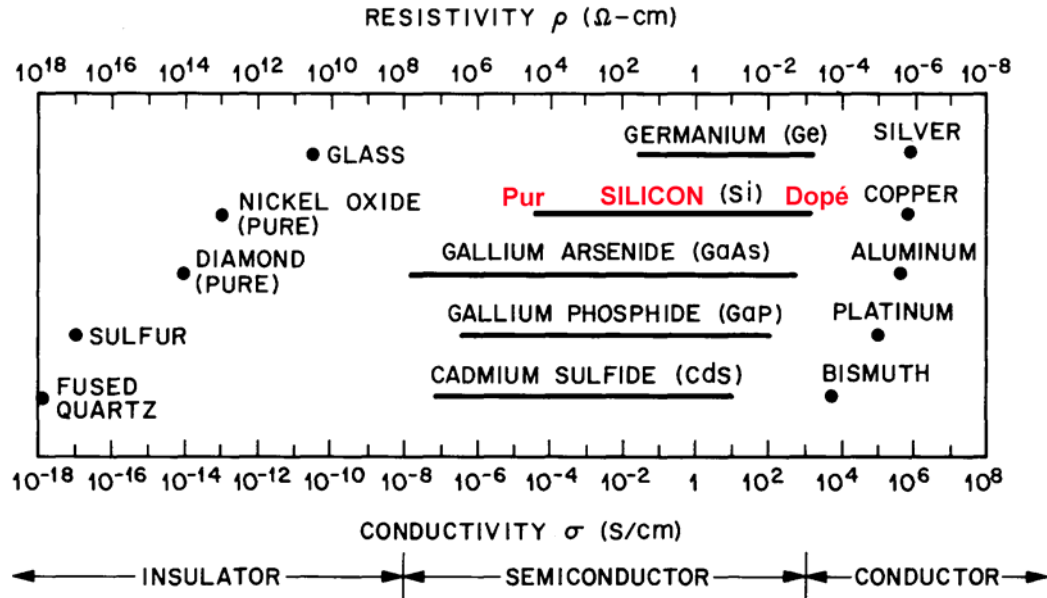


Fig. 1.7 Conductivité de divers matériaux

Les semi-conducteurs y occupent un grand intervalle car leur conductivité dépend sensiblement de la température, mais elle varie surtout très fortement en fonction de la présence de petites quantités d'impuretés bien choisies.

Dans un monocristal de silicium "intrinsèque" (Si ultra pur), élément de la colonne IV du tableau périodique, toutes les liaisons covalentes (4 par atome) sont satisfaites à $T = 0$ K. A la température ambiante, quelques électrons ($1,5 \cdot 10^{10}$ électrons pour $5 \cdot 10^{22}$ atomes dans un cm^3) sont excités thermiquement depuis la bande de valence vers la bande de conduction, et les "électrons libres" et les "trous" positifs ainsi créés contribuent ensemble à la conductivité du matériau, mais celle-ci reste assez faible.

Si on ajoute au Si ultra pur quelques atomes de la colonne V (par exemple du phosphore P) qui possèdent chacun 5 électrons de valence, quatre de ces électrons vont participer aux liaisons covalentes du cristal, et un électron va rester quasiment libre (avec une énergie de liaison très faible). A la température ambiante, ces électrons libres vont tous passer dans la bande de conduction. On dit que l'on a dopé le Si avec un donneur (le phosphore donne un électron à la bande de conduction) et le semi-conducteur est dit de **type n**. Même si le taux d'impuretés est très faible, de l'ordre de quelques parties par million (ppm), il en résulte une densité de donneurs et donc d'électrons libres de 10^{16} atomes/ cm^3 , qui donne une conductivité bien supérieure à celle du Si intrinsèque (elle est multipliée par 10^6). Cette conductivité est dominée par les électrons qui constituent dans ce cas les porteurs majoritaires (de charge).

Si, à l'inverse, on ajoute au Si ultra pur quelques atomes de la colonne III (par exemple du bore B) qui possèdent chacun 3 électrons de valence, la 4^{ème} liaison covalente ne pourra être satisfaite qu'en empruntant un électron du cristal, et un trou faiblement lié va se former et contribuer à la conductivité du matériau ainsi dopé. Un tel élément est appelé un accepteur (d'électron), et le semi-conducteur dopé est dit de **type p**. De façon similaire, la conductivité y est assez importante et est dominée par les trous qui constituent alors les porteurs majoritaires.

Le courant dans un semi-conducteur étant en général la somme de celui dû au déplacement des électrons libres, et de celui dû au déplacement des trous, la conductivité σ est reliée à la densité des porteurs n et p , et à leur mobilité μ par la relation :

$$\sigma \equiv 1/\rho = e (n \mu_n + p \mu_p)$$

où e est la charge de l'électron, n et p sont les densités d'électrons et de trous, et μ est le coefficient qui, à une température donnée, relie la vitesse de dérive moyenne d'un porteur au champ électrique appliqué.

Dans le cas de Si dopé (n ou p) la contribution des porteurs minoritaires est en général négligeable devant celle des porteurs majoritaires.

Des valeurs typiques de μ_n et μ_p à température ambiante sont données dans le tableau :

| μ ($\text{cm}^2 \cdot \text{V}^{-1} \cdot \text{s}^{-1}$) | Si pur | Si fortement dopé |
|---|--------|-------------------|
| Electrons libres | 1350 | 100 |
| Trous | 480 | 60 |

Table. 1.1 Mobilité des électrons et des trous dans divers semi-conducteurs

On constate que les porteurs sont un peu moins mobiles dans le Si dopé que dans le Si pur, mais comme le dopage augmente la densité des porteurs majoritaires par un facteur de l'ordre de 10^7 , il en résulte le fort accroissement de conductivité mentionné plus haut.

On constate également que les électrons sont beaucoup plus mobiles que les trous (par un facteur 1,7 à 2,8) à cause de leur masse effective plus faible. Il en résulte que le silicium de type n , dans les mêmes conditions de dopage et de géométrie, offre une plus faible résistance au passage du courant, et ceci a des conséquences importantes dans la conception des circuits intégrés.

1.2.2 La structure MOS

Une structure MOS (Métal-Oxyde-Semiconducteur) est constituée d'un substrat semi-conducteur (monocristal de Silicium) dopé recouvert d'une couche d'oxyde isolante (SiO_2) sur laquelle est déposée une couche métallique appelée "grille" G :

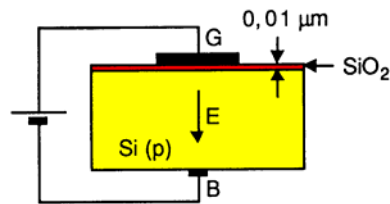


Fig. 1.8 Structure MOS

Ce système est analogue à un condensateur plan et, quand une tension est appliquée entre la grille G et le substrat B , il apparaît une charge sur les deux armatures avec formation d'une zone de charge d'espace. Selon le signe et la grandeur de la tension appliquée entre la grille et le substrat, on peut obtenir différentes situations dans le semi-conducteur : **accumulation** ou **désertion** des porteurs libres à l'interface isolant-semi-conducteur ou encore **inversion**, c'est-à-dire création à cette interface d'une couche de porteurs minoritaires. Par conséquent, en faisant varier la tension appliquée, et donc le champ électrique perpendiculairement au plan des interfaces, on peut fortement modifier la conductivité électrique dans le semi-conducteur, au voisinage de l'interface et parallèlement à ce plan.

Par exemple pour un substrat de type p où les porteurs majoritaires sont les trous, on obtient le comportement suivant en fonction de la tension V_G appliquée entre la grille et le substrat. Le

raisonnement est ici imagé, et un traitement rigoureux est donné dans le cours "Semi-conducteurs et composants" cité précédemment :

- Si $V_G < 0$: le potentiel négatif de la grille attire les trous, porteurs majoritaires du substrat de type p, près de l'interface isolant-semiconducteur où ils sont ainsi accumulés (régime d'**accumulation**).
- Si $V_G > 0$: Le potentiel positif de la grille repousse les trous et attire les électrons. La densité des trous près de l'interface diminue, c'est le régime de **déplétion**.
- Si $V_G \gg 0$: la diminution de la densité de trous au voisinage de l'interface est telle qu'elle devient inférieure à la densité des électrons. Ceux-ci qui étaient minoritaires deviennent majoritaires, et le semi-conducteur devient localement de type n au voisinage de l'interface avec l'isolant. C'est le régime d'**inversion**.

Notes : • La transition entre le régime d'accumulation et celui de déplétion n'a pas forcément lieu exactement à $V_G = 0$, et le potentiel correspondant à cette transition dépend du dopage du substrat utilisé par le biais de son potentiel chimique.

• Le potentiel de transition entre le régime de déplétion et le régime d'inversion est un paramètre essentiel de la structure et sera noté V_T .

• La grille était initialement réalisée en aluminium. Actuellement, pour des raisons de fiabilité, en particulier pour les faibles épaisseurs d'oxyde, on réalise la grille en silicium polycristallin fortement dopé et appelé **polysilicium**.

1.2.3 Le transistor N-MOS

Un transistor MOS est une structure hétérogène, à 3 ou 4 connections, dans laquelle la conductivité dans un "canal" entre deux électrodes (appelées « Source » et « Drain ») est commandée par un champ électrique créé au moyen d'une troisième électrode (appelée Grille).

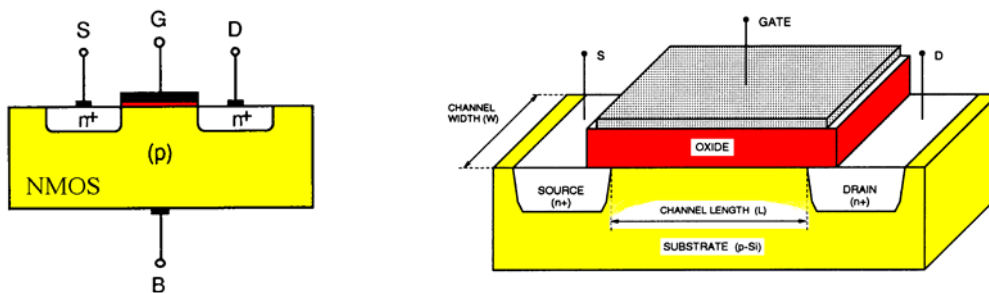


Fig. 1.9 Structure d'un transistor N-MOS

Nous allons dans un premier temps décrire un transistor MOS de type N (N-MOS), constitué d'un substrat de type p à la surface duquel a été formée une structure MOS décrite ci-dessus, encadrée de deux zones de type n+ (fortement dopées) et constituant la Source et le Drain. Le choix d'un fort dopage pour ces zones permet d'avoir un contact métal-semiconducteur de type ohmique pour les fils de connexion. On observera que la structure est géométriquement symétrique par échange de la Source et du Drain, mais on conviendra d'appeler « Source » la région n+ de potentiel le plus bas, et « Drain » la région n+ de potentiel le plus élevé de telle sorte que $V_{DS} \geq 0$.

Les dimensions de la structure dépendent des propriétés électriques recherchées. Des valeurs typiques sont les suivantes :

- Longueur $L = 0,15$ à $1,2 \mu\text{m}$

- Largeur $W = 1 \text{ à } 10 \text{ }\mu\text{m}$
- Epaisseur d'oxyde = $0,005 \text{ à } 0,1 \text{ }\mu\text{m}$ (50 à 1000 Angströms).

1.2.4 Le fonctionnement du transistor N-MOS

Examinons le comportement du transistor N-MOS en fonction des tensions appliquées, en prenant comme référence le potentiel de la source V_S , et en supposant que le substrat est soit connecté à la source, soit à un potentiel plus négatif ($V_{BS} \leq 0$).

Nous supposons que le dopage du substrat de type p est tel que pour $V_G = 0$ la structure MOS est en régime d'accumulation. Cette situation est la plus courante, et le N-MOS est alors dit à enrichissement. Nous ne traiterons pas ici des transistors MOS à déplétion qui sont plus rarement utilisés.

Lorsque $V_G \leq 0$, la couche de type p proche de l'isolant est en régime d'enrichissement en trous (porteurs majoritaires), et le trajet Source-Drain est constitué de deux jonctions p – n (diodes) tête-bêche. Il est donc non conducteur (aux courants de fuite près qui sont très faibles) et l'on dit alors que le transistor est bloqué.

Pour $0 \leq V_G < V_T$, où V_T est une tension de seuil (positive pour un N-MOS), et pour V_{DS} nul ou faible, la couche proche de l'isolant passe progressivement en mode de déplétion, le trajet Source-Drain reste bloqué mais se rapproche de la conduction.

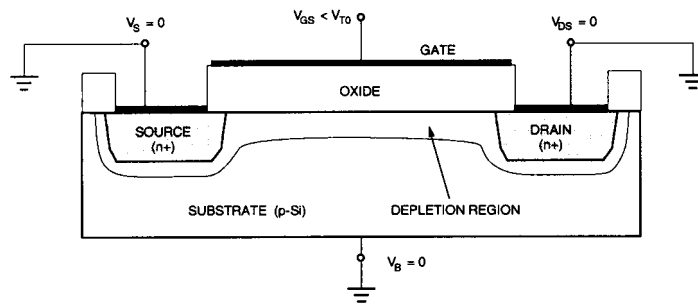


Fig. 1.10 Transistor N-MOS avec une zone de déplétion

Pour $V_G > V_T$ et V_{DS} nul ou faible, la structure MOS est en régime d'inversion, un canal de type n se forme au voisinage de l'interface avec l'isolant et constitue un circuit conducteur entre les deux zones n^+ . Un courant électronique peut alors circuler de la Source vers le Drain. Le transistor est alors dit conducteur ou passant. Il faut noter que $I_S = I_D$ puisque la Grille est isolée, et que le substrat est par hypothèse hors circuit. La valeur de ce courant commun à la Source et au Drain dépend des potentiels V_{GS} et V_{DS} .

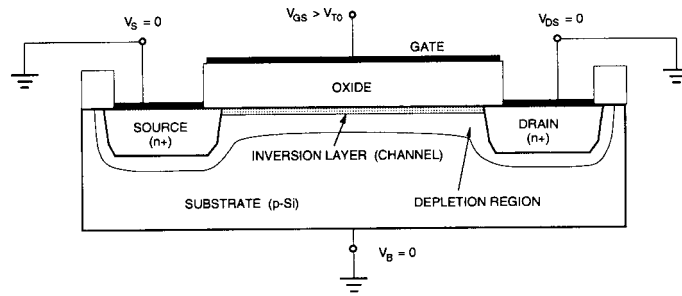


Fig. 1.11 Transistor N-MOS avec un canal d'inversion

Tant que $V_{GS} > V_T$ et $V_{GD} > V_T$ (et donc pour $V_{DS} < V_{GS} - V_T$), le canal s'étend sur toute la longueur entre la Source et le Drain et se comporte comme une résistance ($I_S = I_D \approx V_{DS}/R$) dont la valeur R est indépendante de V_{DS} , mais varie avec la tension de commande V_{GS} . La densité électronique dans le canal augmente lorsque la tension de grille augmente, et donc la conductivité du canal augmente elle aussi. On dit que le transistor est en mode résistif.

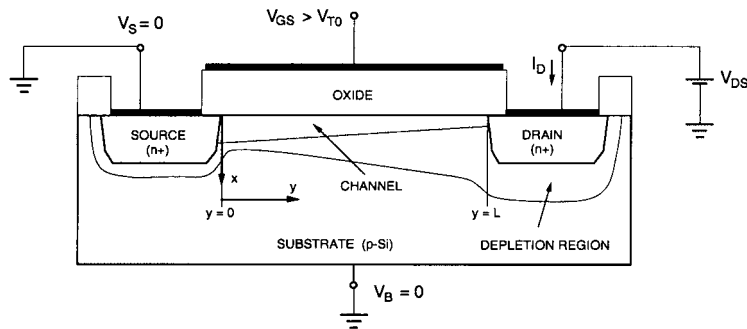


Fig. 1.12 Transistor N-MOS en mode résistif

Fortes valeurs de V_{DS} : on maintient V_G fixé ($V_{GS} > V_T$), et on augmente V_D qui se rapproche de V_G . A partir d'un certain potentiel V_{DSat} appelé tension de pincement, la condition $V_{GD} > V_T$ ne sera plus satisfaite, il n'y aura plus d'inversion de population au voisinage du Drain, et l'épaisseur du canal sera localement réduite à zéro. Au-delà de cette tension de pincement, le point de pincement se déplace vers la Source de sorte que la tension V_{DSat} apparaît appliquée aux extrémités d'un canal dont la longueur est diminuée de ΔL , l'excédent de tension ($V_D - V_{DSat}$) étant chuté dans une zone désertée (diode en inverse de forte résistivité) de longueur ΔL .

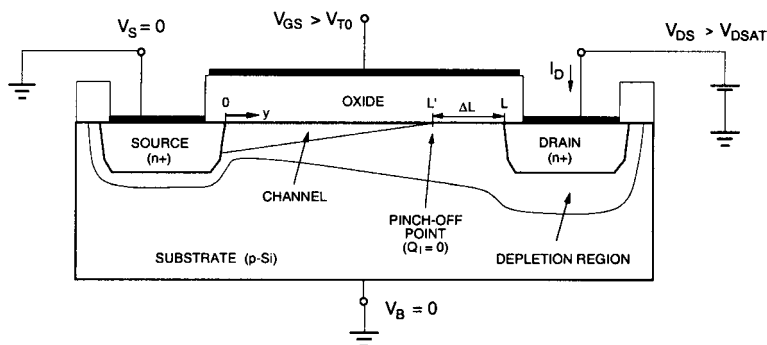


Fig. 1.13 Transistor N-MOS en mode saturé

La présence de cette zone désertée en série avec le canal conducteur ne diminue pas le courant car les

électrons qui parviennent au point de pincement sont aspirés par le fort champ électrique pour être injectés dans le Drain. Dans la mesure où $\Delta L \ll L$, le courant I_D est principalement déterminé par la conductivité du canal et la différence de potentiel V_{Dsat} , et ce courant reste approximativement constant lorsque V_D excède V_{Dsat} . Le transistor est en mode saturé.

1.2.5 Les caractéristiques courant-tension du transistor N-MOS

Comme indiqué précédemment, le transistor MOS est principalement utilisé pour ses propriétés de transconductance (source de courant commandée par une tension) : on s'intéressera donc surtout aux caractéristiques du dipôle Drain-Source et à sa dépendance en fonction de la tension de Grille. La description quantitative du transistor N-MOS peut être plus ou moins précise en fonction de la complexité de la modélisation utilisée, et pour les besoins de ce cours, une modélisation simple (quadratique) est suffisante. Elle fournit le courant de Drain $I_D = I_S$ dans les divers modes de fonctionnement que nous venons de voir par les relations suivantes :

a) $V_{GS} < V_T$ **Transistor Bloqué** : $I_D = 0$

b) $V_{GS} > V_T$ et $V_{GD} > V_T \iff V_{DS} < V_{GS} - V_T$ **Mode « résistif »** :

$$I_D = \mu_n C_{ox} \frac{W}{L} \left[(V_{GS} - V_T) V_{DS} - \frac{1}{2} V_{DS}^2 \right]$$

c) $V_{GS} > V_T$ et $V_{GD} < V_T \iff V_{DS} > V_{GS} - V_T$ **Mode saturé** :

$$I_D = \frac{1}{2} \mu_n C_{ox} \frac{W}{L} (V_{GS} - V_T)^2$$

\implies La frontière entre le mode résistif et le mode saturé est donnée par la relation :

$$V_{GD} = V_T \iff I_D = \frac{1}{2} \mu_n C_{ox} \frac{W}{L} V_{DS}^2$$

où : W et L sont la largeur et la longueur du canal;
 μ_n est la mobilité des électrons (porteurs majoritaires du canal n);
 C_{ox} est la capacité par unité de surface de la capacité MOS;
 $C_{ox} = \epsilon_{ox} / t_{ox}$ où ϵ_{ox} est la permittivité de l'oxyde et t_{ox} son épaisseur.

On définit ainsi les paramètres de transconductance :

$$k'_n = \mu_n C_{ox}$$

$$k_n = \mu_n C_{ox} \frac{W}{L}$$

Les courbes caractéristiques courant-tension ont ainsi l'allure suivante :

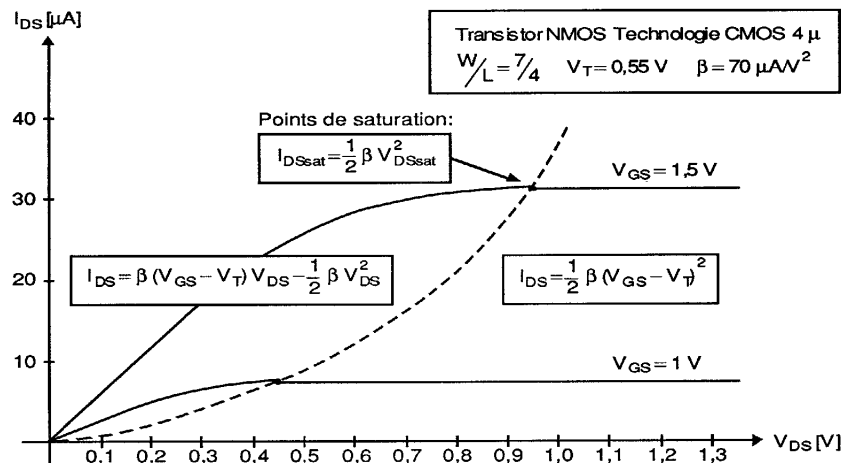


Fig. 1.14 Caractéristique courant-tension d'un transistor N-MOS

Des valeurs typiques des paramètres sont les suivantes :

- Mobilité $\mu_n \approx 580 \text{ cm}^2/\text{V.s}$ ($\mu_p \approx 230 \text{ cm}^2/\text{V.s}$)
- Oxyde
 - Epaisseur : $t_{\text{ox}} \approx 0,02 \text{ à } 0,1 \text{ }\mu\text{m}$
 - Permittivité : $\epsilon_{\text{ox}} = 3,97 \epsilon_0 = 3,5 \cdot 10^{-13} \text{ F/cm}$
 - Capacité caract: $C_{\text{ox}} = 1,75 \text{ fF}/\mu\text{m}^2$ pour $t_{\text{ox}} = 0,02 \text{ }\mu\text{m}$
 - $C_{\text{ox}} = 0,35 \text{ fF}/\mu\text{m}^2$ pour $t_{\text{ox}} = 0,1 \text{ }\mu\text{m}$
- Transconductance
 - $k'_n = \mu_n C_{\text{ox}} \approx 100 \text{ }\mu\text{A}/\text{V}^2$ pour $t_{\text{ox}} = 0,02 \text{ }\mu\text{m}$
 - $\approx 20 \text{ }\mu\text{A}/\text{V}^2$ pour $t_{\text{ox}} = 0,1 \text{ }\mu\text{m}$

Les limites de ce modèle sont les suivantes :

- La réduction ΔL de la longueur de canal en mode saturé entraîne en réalité une légère augmentation du courant I_D lorsque V_{DS} augmente au delà du seuil de saturation. Cet effet peut être pris en compte au moyen d'un terme multiplicatif $(1 + \lambda V_{DS})$, où $1 / \lambda$ est un paramètre homogène à une tension (tension d'Early V_A) qui vaut entre 30 et 200 volts. Il est particulièrement gênant pour les transistors de petite taille.
- Les effets liés au substrat d'un transistor unique ont été négligés en supposant le substrat relié à la Source. Dans ce cas en effet, la "jonction p-n" entre le substrat et le canal est bloquée car polarisée en inverse, et le substrat ne joue effectivement aucun rôle. Dans un circuit complexe, le substrat est commun à de nombreux transistors, et ne peut donc pas être relié à la Source de chacun d'eux. On obtient néanmoins un résultat similaire en maintenant le substrat à un potentiel constant et plus négatif que l'ensemble des électrodes. Cependant, la polarisation inverse V_{SB} variable a pour effet de réduire la profondeur du canal, et donc d'augmenter légèrement la tension de seuil V_T en fonction de V_{SB} .
- Les effets de température ont été négligés (k' décroît avec la température).
- Au delà de certaines limites de tension, le champ électrique est trop élevé et des avalanches ou des claquages peuvent se produire.
- C'est un modèle statique, valable à basse fréquence. Il est possible de l'étendre aux hautes fréquences en y ajoutant les capacités des éléments du circuit (capacités entre conducteurs séparés par un isolant, mais aussi capacités dynamiques des jonctions polarisées en inverse).

1.2.6 Le transistor P-MOS

D'une façon similaire au transistor N-MOS, on peut construire un transistor P-MOS en remplaçant toutes les zones "p" par des zones "n" et vice-versa.

On nommera Source la zone p+ de potentiel le plus élevé, ce qui donnera $V_{DS} \leq 0$.

La conduction se fera par un canal de type p où les porteurs majoritaires sont les trous, et le courant de trous ira de la Source au Drain.

La **tension de seuil V_T est négative**, la zone intéressante de fonctionnement sera donc pour $V_{GS} \leq 0$.

On obtient également 3 modes de fonctionnement pour le transistor P-MOS :

a) $V_{GS} > V_T$ **Transistor Bloqué** : $I_D = 0$

b) $V_{GS} < V_T$ et $V_{GD} < V_T \iff V_{DS} > V_{GS} - V_T$ **Mode « résistif »** :

$$I_D = \mu_p C_{\text{ox}} \frac{W}{L} \left[(V_{GS} - V_T) V_{DS} - \frac{1}{2} V_{DS}^2 \right]$$

c) $V_{GS} < V_T$ et $V_{GD} > V_T \iff V_{DS} < V_{GS} - V_T$ **Mode saturé** :

$$I_D = \frac{1}{2} \mu_p C_{ox} \frac{W}{L} (V_{GS} - V_T)^2$$

\implies La frontière entre le mode résistif et le mode saturé est donnée par la relation :

$$V_{GD} = V_T \iff I_D = \frac{1}{2} \mu_p C_{ox} \frac{W}{L} V_{DS}^2$$

où : W et L sont la largeur et la longueur du canal;
 μ_p est la mobilité des trous (porteurs majoritaires du canal p);
 C_{ox} est la capacité par unité de surface de la capacité MOS;
 $C_{ox} = \epsilon_{ox} / t_{ox}$ où ϵ_{ox} est la permittivité de l'oxyde et t_{ox} son épaisseur.

On définit ainsi les paramètres de transconductance :

$$k'_p = \mu_p C_{ox}$$

$$k_p = \mu_p C_{ox} \frac{W}{L}$$

A cause de la mobilité plus faible des trous, la transconductance des P-MOS est, à géométrie égale, plus faible que celle des N-MOS, ils sont moins bons conducteurs du courant.

Pour mieux visualiser le comportement respectif des N-MOS et des P-MOS, il est possible de représenter leur mode de fonctionnement dans un diagramme où les potentiels des 3 électrodes sont placés sur une échelle commune :

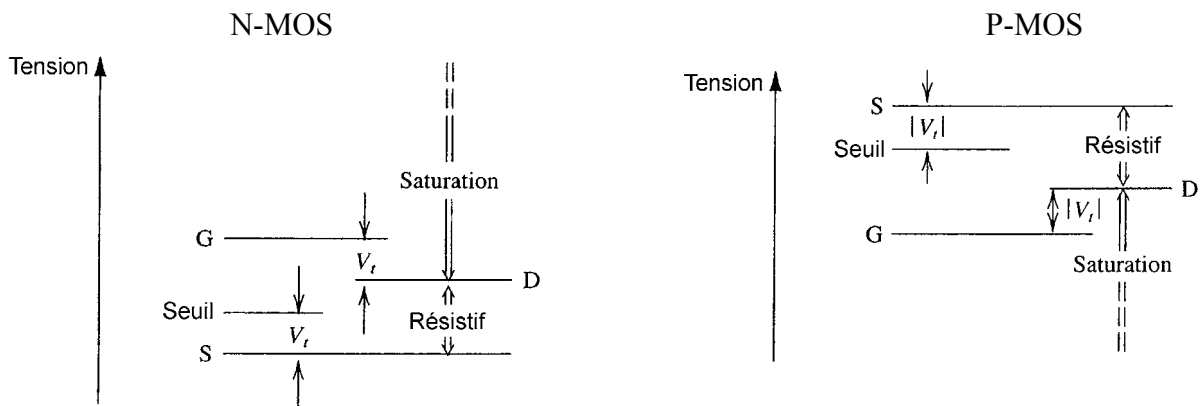


Fig. 1.15 Modes de fonctionnement des transistors N-MOS et P-MOS

Dans de nombreux cas, il est possible d'approximer encore plus simplement le comportement des transistors en les considérant comme des interrupteurs commandés par la tension de grille V_G et possédant deux états :

- Etat **ouvert** : Résistance infinie et courant nul;
- Etat **fermé** : Résistance finie (faible, mais non nulle, souvent de l'ordre de 1000 Ω).

Dans une convention logique "positive" la plus usuelle, où l'état "1" est associé au niveau "haut" de tension, le comportement des interrupteurs N et P est donné par la figure suivante :



Fig. 1.16 Approximation des transistors N-MOS et P-MOS par des interrupteurs

Cette figure montre également la représentation symbolique des transistors. Noter la différence entre le symbole du N-MOS et celui du P-MOS, ainsi que la flèche optionnelle indiquant la Source.

1.2.7 Les circuits C-MOS

La technologie C-MOS (Complementary MOS) fait appel aux deux types de transistors N-MOS et P-MOS réalisés sur le même support de façon à profiter de la dualité de leurs propriétés. Cette dualité permet des constructions élégantes et puissantes et elle offre plusieurs avantages décisifs :

- Faible puissance dissipée par l'absence de courants statiques (dans les états stables, un transistor est toujours bloqué lorsque l'autre est conducteur);
- Rapidité de commutation en éliminant les résistances de charge.

Il existe cependant une difficulté importante dans leur fabrication puisque les N-MOS sont construits sur un substrat de type "p", et les P-MOS sur un substrat de type "n". On contourne ce problème en créant par implantation ionique des zones spécifiques (appelées puits ou caissons) de type "n" à la surface d'un substrat unique de type "p".

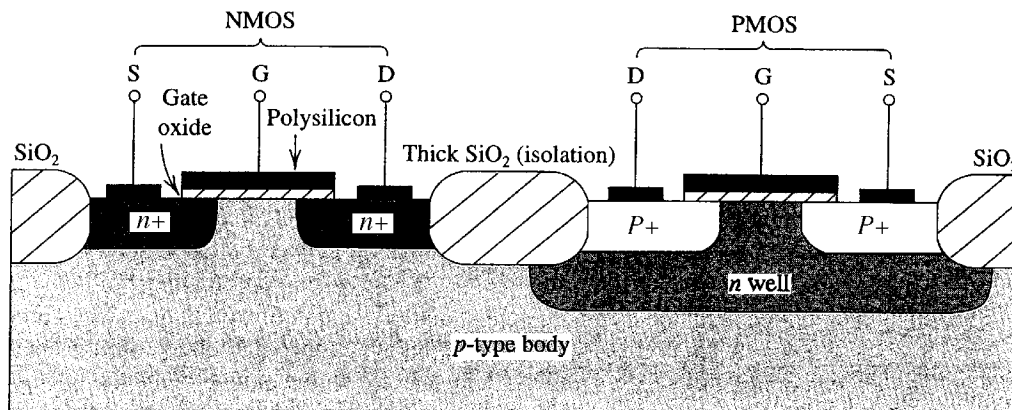


Fig. 1.17 Structure d'un circuit intégré C-MOS

Il est ainsi possible de faire coexister les deux types de transistors au sein d'un même circuit intégré monolithique.

La technologie C-MOS a actuellement pratiquement supplanté les technologies antérieures, en particulier pour les circuits numériques très rapides et à très haute intégration tels que les microprocesseurs et les mémoires.

La suite de ce cours se concentrera donc exclusivement sur les circuits C-MOS.

Les éléments de base seront des transistors N-MOS et P-MOS (ces derniers réalisés dans des caissons) qui seront en général dessinés à la taille minimale permise (pour avoir la densité d'intégration maximale). Pour améliorer la symétrie électrique entre les N-MOS et les P-MOS, et compenser la différence de mobilité entre les électrons et les trous, les transistors P-MOS seront le plus souvent dessinés deux fois plus larges que les N-MOS.

Dans toute la mesure du possible, les transistors seront traités comme des contacts résistifs.

Chapitre2 :

Fabrication des Transistors CMOS

(par Yvan Bonnassieux, Ecole Polytechnique)

1. Introduction

Depuis le début de la microélectronique la dimension des transistors c'est réduite aux taux de 13% par an. Et comme le montre les prévisions « roadmap » cela vas perdurer encore au moins pour cette décennie.

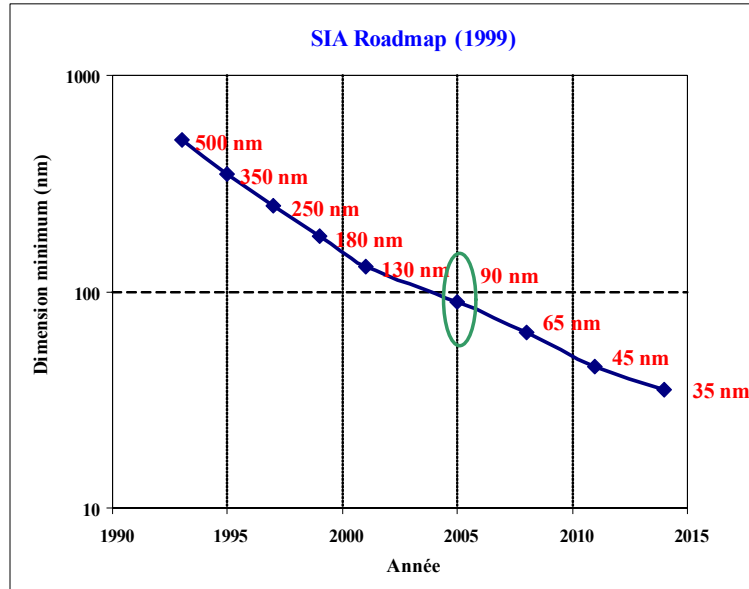


Figure 1 : Taille minimale des transistors (Roadmap SIA 1999)

Cette miniaturisation a engendrée par fonction élémentaire une forte diminution des coûts, ainsi par exemple pour les DRAM le coût par points mémoire est divisé par 2 tous les 2ans. De manière similaire pour les portes logiques élémentaires le temps de commutation a réduit de 4 magnitudes depuis 1959 et la consommation énergétique divisée par 1 millions.

Cette évolution fantastique à permit de concevoir des circuits dont la puissance de calcul à crue de manière exponentielle (un facteur 2 tous les 18 mois). Cette évolution avait été prévue de manière empirique par Gordon MOORE en 1965 (cofondateur d'Intel).

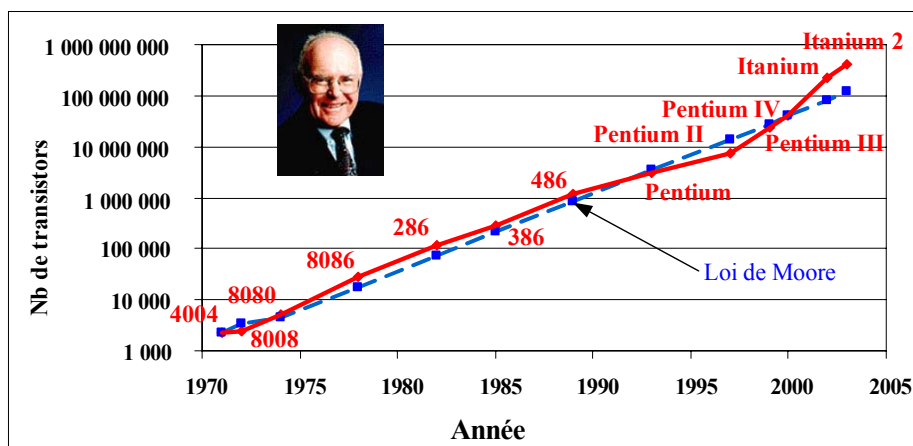


Figure 2 : Loi de Moore

Si les voitures avaient Evolué à la même vitesse que les circuits intégrés, une voiture pourrait aujourd'hui :

- Ce déplacer à la vitesse de la lumière
- Rouler des années avec un seul plein d'essence
- Transporter une ville entière dans une voiture

1.1. Un bref historique

Nombre des technologies de fabrication des semi-conducteurs dérive de procédés inventés il y a de nombreux siècles. Ainsi, par exemple, la lithographie à été inventé en 1798. Initialement l'image était transférée depuis une pierre (ce qui justifie l'étymologie).

En ce qui concerne plus particulièrement la fabrication des composants semi-conducteurs on peut citer les étapes suivantes (nous reviendrons sur ces technologies dans le corps de ce chapitre):

En **1918**, *Czochralski* développe la technique de croissance liquide solide toujours utilisée pour la fabrication des wafers de silicium.

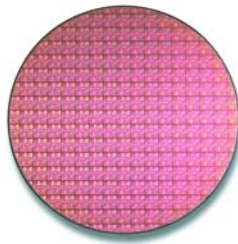


Figure 3 : Wafer

1947, Invention du transistor bipolaire à point par *Bardeen, Brattain et Schockey*

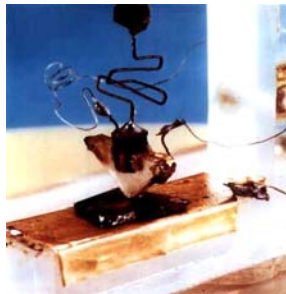


Figure 4 : premier transistor

L'idée d'utiliser la diffusion d'impuretés, pour changer la conductivité du silicium a défini par un brevet de *Pfann* en **1955**.

En **1957**, *Andrus* applique la lithographie pour la fabrication des composants semi-conducteurs. Il a utilisé les caractéristiques photosensibles de polymères résistants à la gravure pour le transfert des dessins.

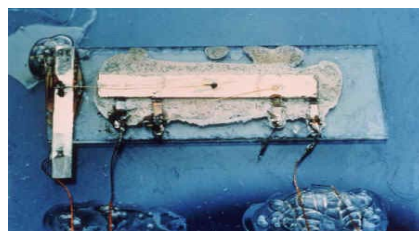


Figure 5 : lithographie

Le masquage par oxyde a été développé par *Frosch et Derrick* en **1957**. Ils ont constaté qu'une couche d'oxyde peut empêcher la diffusion de la plupart des atomes d'impureté.

Toujours en **1957**, l'épitaxie, c'est-à-dire la croissance du cristal de silicium par déposition de vapeur chimique a été inventée par *Sheftal et al.*

En **1959**, *Noyce* propose le premier circuit intégré monolithique. C'est-à-dire la construction sur un seul substrat de semi-conducteurs d'un dispositif complet. La figure 6 montre ce premier circuit, il s'agit d'une bascule comportant 6 composants avec leurs interconnexions en aluminium obtenu par gravure.

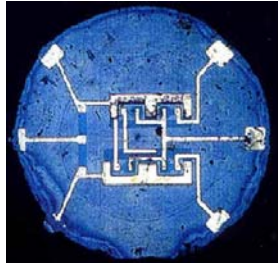


Figure 6 : premier circuit intégré

Le procédé Planar a été conçu par *Hoerni* en **1960**. On commence par créer une couche d'oxyde sur le semi-conducteur. Puis par lithographie des fenêtres sont créées dans cet oxyde. Les atomes d'impuretés diffusent seulement dans le semi-conducteur non protégé par l'oxyde.

Les premiers circuits intégrés étaient NMOS, les composants CMOS (NMOS et PMOS) sont proposés en **1963** par *Sah*.

1965, première formulation de la loi de *Moore*

En **1967**, *Dennard* invente la mémoire DRAM (Dynamic Random Access Memory) par association d'un transistor MOSFET et d'une capacité de stockage.

1971 est l'année de la conception du premier microprocesseur par *Hoof et al* de la société Intel. Il s'agit d'un processeur 4 bits (Intel 4004) que nous pouvons voir sur la figure 7. C'est un circuit de 3mm par 4mm comportant 2300 transistors ayant une longueur de 8 μ m.

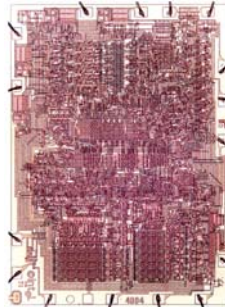


Figure 7 : Processeur Intel 4004

Depuis lors de nombreuses autres nouvelles technologies ont permis, sans changer vraiment les méthodes de fabrication de réduire la taille des transistors et ainsi la densité des mémoires et la puissance des microprocesseurs. On peut citer plus particulièrement les interconnexions par cuivre proposées par *Rung et al.* en 1993. La meilleure conductivité de ce métal par rapport à l'aluminium a permis de réduire de manière drastique la taille des interconnexions.

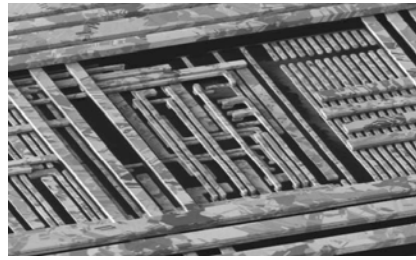


Figure 8 : Interconnexions en cuivre

Cette course à la miniaturisation permet aujourd'hui de concevoir des transistors d'une longueur de 130nm tel que celui présenté sur la figure 9. Ce qui permet de concevoir des processeurs tels que le Pentium IV (Intel) qui comporte environs 40 millions de transistors.



Figure 9 : processeur Intel Pentium IV

1.2. Les principales étapes de fabrication des semi-conducteurs

Aujourd'hui la technologie PLANAR est celle utilisée de manière quasiment exclusive pour la fabrication des circuits intégrés.

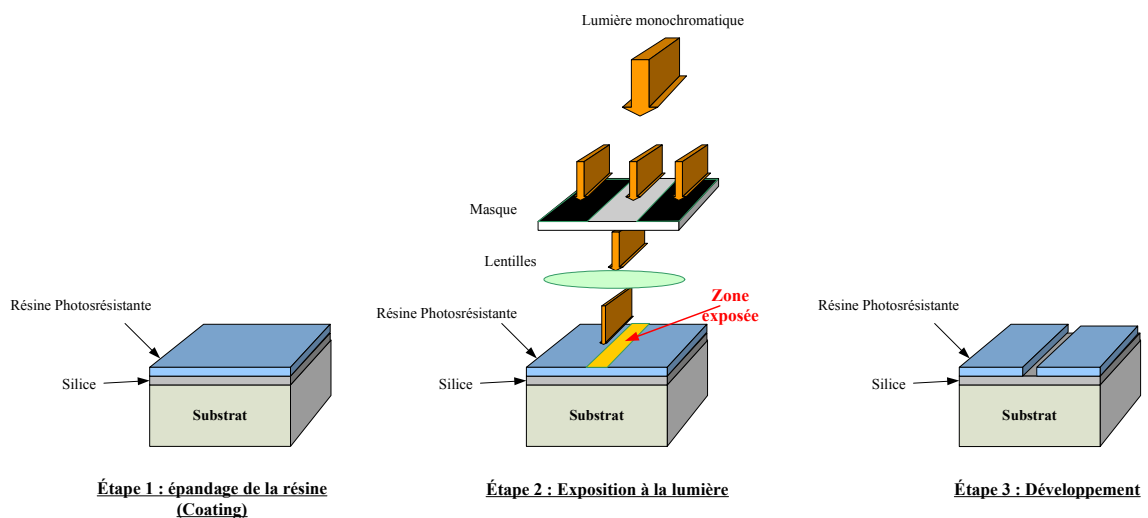


Figure 10 : Technologie Planar (1^{ère} partie)

Les figures 10 & 11 montre de manière rapide les différentes étapes du process PLANAR, étapes qui seront plus complètement explicité dans les chapitres suivants.

Partant d'un cristal de silicium : wafer.

Une première étape d'oxydation permet d'obtenir une couche de silice : SiO_2 .

Puis la photolithographie permet de définir la géométrie de du transistor. Elle consiste en un dépôt de résine photorésistante (photoresist coating), de l'exposition de cette résine à un rayonnement ultraviolet à travers un masque (Exposition), puis développement de cette résine par attaque chimique.

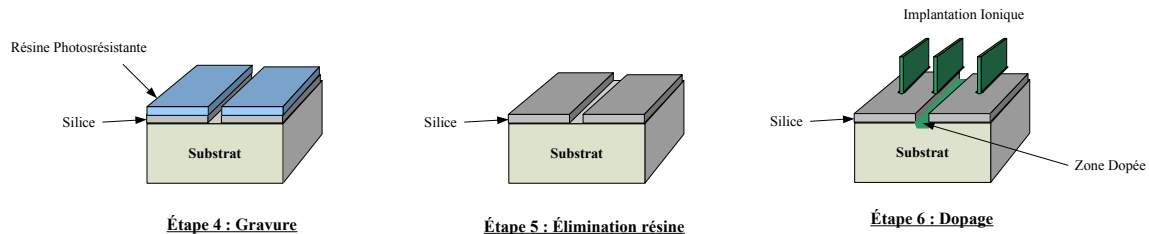


Figure 11 : Technologie Planar (2^{ème} partie)

Une seconde attaque chimique ou physique permet d'enlever la silice la ou elle n'est pas protégée par la résine (Gravure ou Etching).

Une nouvelle attaque chimique permet d'enlever la résine restante.

Le dopage ce fait par implantation ionique ou par diffusion d'impuretés. La silice joue un rôle de barrière et empêche l'implantation de ces impuretés.

Un dépôt de métal est ensuite réalisé par déposition de vapeur pour réaliser les contacts ohmiques et les interconnexions.

2. La Photolithographie

La photolithographie est le processus de transfert de formes géométrique d'un masque sur une fine couche de matériaux photosensibles (résines photorésistantes) qui recouvrent d'un wafer de semi-conducteur. Ces formes définissent les différentes régions d'un circuit intégré tels que

Les zones de dopage, les connections métalliques, les points de contacts. Les dessins de résines définies par la photolithographie ne sont pas des éléments permanents d'un composant mais seulement la réplification des dispositifs des circuits. Comme nous l'avons déjà brièvement d'écrit cette réplification est réalisée par une gravure (voir paragraphe suivant) des couches qui ne sont pas protégés par la résine photorésistante.

2.1. Lithographie Optique

La plupart des équipements de lithogravure utilisé lors de la fabrication des circuits intégrés utilise des lampes à ultraviolet ($\lambda=0,2-0,4\mu\text{m}$). Avant de présenter les différents outils de la photolithographie optique, il nous semble nécessaire de faire une rapide présentation des salles blanche car ce processus de fabrication se doit d'être réalisé dans des conditions de très grande propreté.

2.1.1. Salle blanche

La photolithographie utilise un processus de masquage pour le transfert de forme sur les différentes couches constituant un composant. En présence de poussières plusieurs défauts peuvent apparaître :

- Si les poussières adhérentes sur le masque il modifie les zones masquées et ainsi les formes obtenues après gravure comme le montre la figure 12.

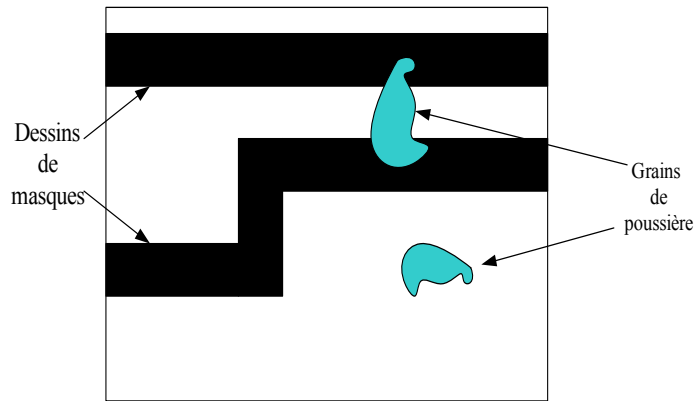


Figure 12 : Effet des poussières sur le masquage

- Si elles sont présentes dans l'oxyde de grille, elles peuvent modifier la conductivité et ainsi provoqué un affaiblissement nuisible de la tension de claquage.
- De grosses particules peuvent aussi avoir des tailles supérieures à l'épaisseur de certaines couches et ainsi réaliser de véritables contacts parasites (« pinhole »).

Pour limiter la présence de ces poussières la fabrication des circuits intégrés se doit d'être réalisée dans des locaux les plus propres possibles : Les salles blanches. Dans ces dernières la quantité de particules et leur taille par unité de volume est parfaitement maîtrisée ainsi que la température et l'humidité. La figure 13 présente le nombre et la taille de particules suivantes la classe de la salle blanche.

| Classe | Nb de particules de 0,5mm Par pied cube | Nb de particules de 5mm Par pied cube |
|---------|--|--|
| 1 | 1 | 0 |
| 10 | 10 | 0 |
| 100 | 100 | 0 |
| 1 000 | 1 000 | 7 |
| 10 000 | 10 000 | 70 |
| 100 000 | 100 000 | 700 |

Figure 13 : classe des salles blanches (norme FS209)

On notera que 2 systèmes de classification coexistent :

- Pour le système anglo-saxon la classe est définie par un chiffre qui correspond au nombre maximal de particules de taille supérieure à 0,5µm par pied cube.
- Pour le système métrique la classe représente le logarithme (en base 10) du nombre maximum de particules de taille supérieure à 0,5µm par m³.

Ainsi une salles blanche de classe 100 (qui est classique pour la fabrication industriel des circuits intégrés actuels) correspond à une classe M3,5 en système métrique.

La course à la réduction de la taille des composants nécessite de manière évidente la réduction de la taille des particules supportables en salle blanche, ainsi les chaînes de production de techno 130nm se font dans des salles blanches de classe '10' voir '1'.

Par comparaison une salle de cours est une salle de classe supérieure à '1 000 000'.

2.1.2. Principe

La photolithographie optique est réalisée par des masqueurs tels que la MJB3 représenté sur la figure suivante. La résine est modifiée chimiquement par exposition pendant un temps donné à une lampe ultraviolet à travers une lentille de collimation.



Figure 14 : Masqueurs MJB3 pour lithographie

Il existe 2 méthodes de masquage :

- Le masquage par contact qui nécessite un contact direct entre le wafer recouvert de résine photosensible et le masque comme le montre la figure 15(a). Cette méthode peut engendrer des dégradations du masque surtout en présence de poussières.
- Le masquage de proximité, ou une distance faible ($50\mu\text{m}$) est laissée entre le masque et la résine comme le montre la figure 15(b). Cette méthode engendre un phénomène de diffraction qui dégrade la résolution (minimum $2\mu\text{m}$)

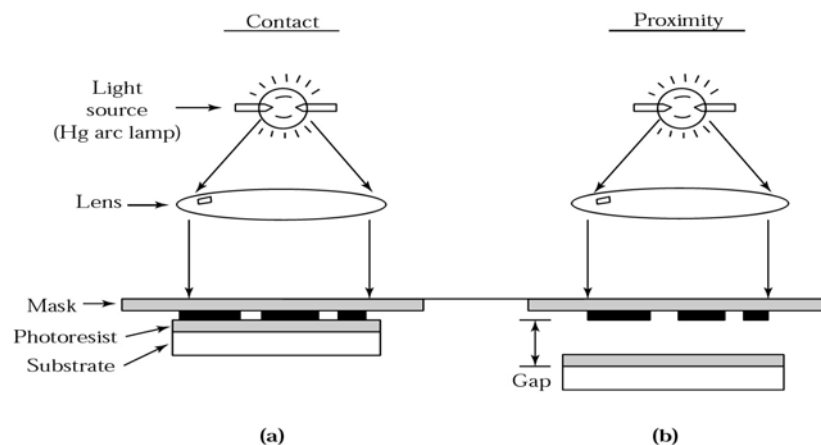


Figure 15 : méthodes d'impression par masquage

Avec la méthode précédente la résolution des masques est identique à celle des dessins sur la résine cela devient quasiment impossible de fabriquer des masques pour une résolution submicronique. Pour réduire la résolution sans engendrer de dégradation des masques, un système de masquage par projection du masque a été développé : « le Stepper ». Le principe décrit sur la figure 16 utilise la projection du masque directement sur la résine via un système de lentille. Cette approche permet de

travaillé avec des maques de grande dimension sont la projection sera réduite par l'optique en outre elle permet de faire une structure de masque simple qui sera répéter pas à pas sur le wafer.

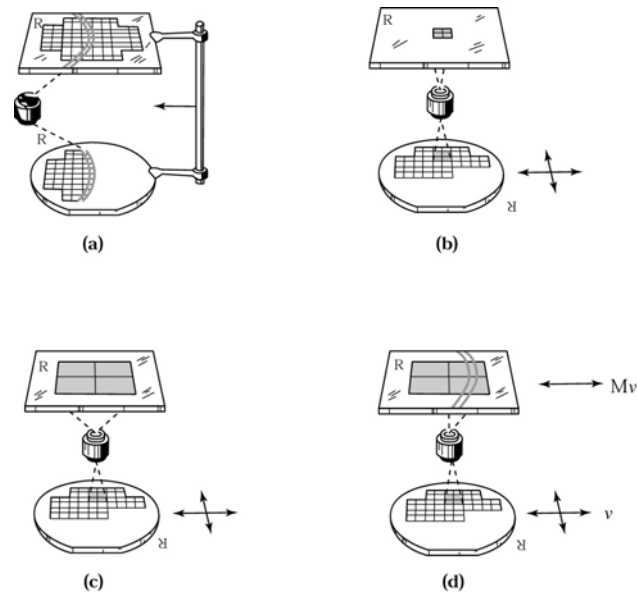


Figure 16 : masquage par les Stepper

Les lampes utilisées par les masqueurs sont en générale des lampes à arc à vapeur de mercure car elle possède une grande intensité avec une grande stabilité, elles permettent d'atteindre une résolution maximale de 300nm. Pour les résolutions supérieure on développe des lasers solides de type EXIMER qui permettent avec des longueurs d'onde de $\lambda=157\text{nm}$ d'atteindre une résolution de 70nm pour la suite la solution reste à inventer.

2.1.3. Les Masques

La première étape de fabrication des masques utilise un système de CAO qui permet une description complète du circuit via des modèles électrique. La chaîne de CAO traduit de manière automatique (ou semi-automatique) sous forme de dessin de masques. Ces dessins représente un niveau de désigne du circuit imprimé. Chaque séquence du processus de fabrication comme les zones d'isolation les régions de grilles,... sont donc traduites par un masque différent. Ces dessins sont directement traduits par un système de lithographie par faisceau d'électrons sur un masque physique. Ce masque est constitué d'un substrat de silice recouvert d'une couche de chrome

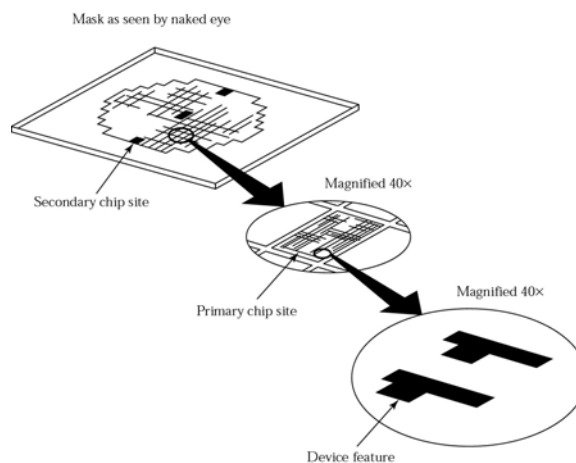


Figure 17 : Dessin de masque

2.1.4. Résines photorésistantes

Il existe deux types de résines photosensibles. Elles se caractérisent par leurs différences de réponses aux radiations comme le montre la figure 18.

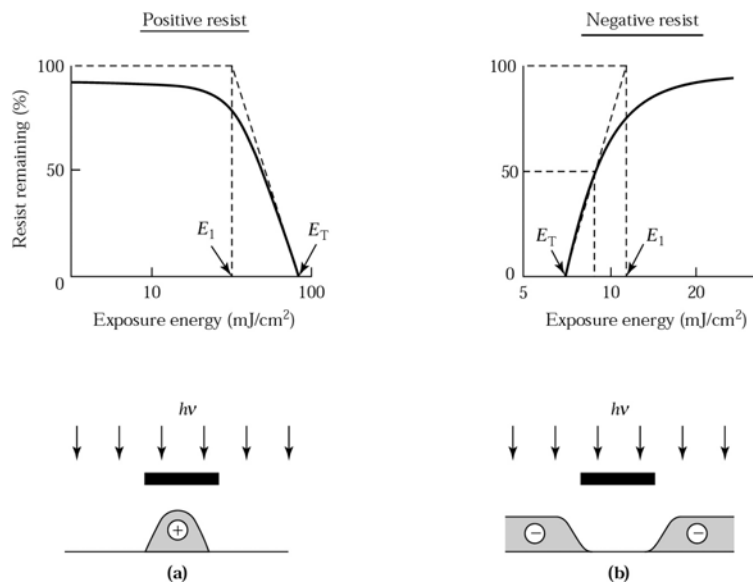


Figure 18 : Les 2 types de résines photosensibles

L'étalement des résines sur les wafer se fait par un système de rotation : « les tournettes » tels que celui présenter sur la figure 19.

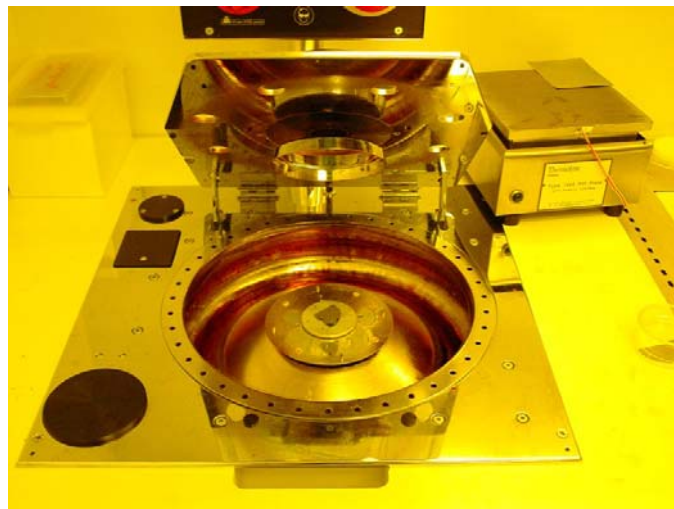


Figure 19 : Tournette d'étalement de la résine photosensible

2.2. Prochaines générations de lithographie

Afin d'améliorer encore la définition et ainsi réduire la taille des dessins sur le wafer donc celle des composants de nouvelles méthodes de lithographie ont été développées.

2.2.1. Lithographie e-beam

On utilise ici un faisceau d'électron.

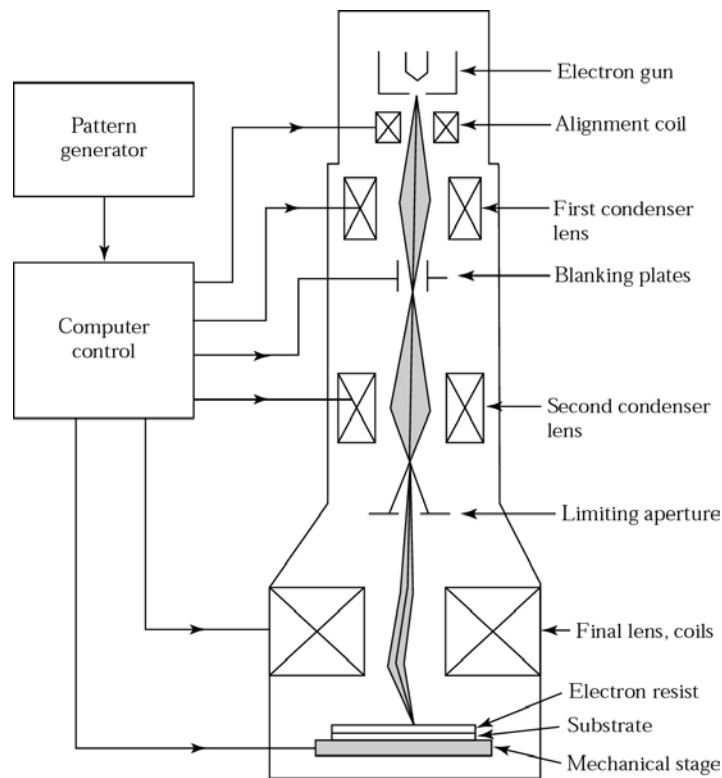


Figure 20 : Principe d'un système de lithographie électronique

Dans cette méthode pas de masques physique mais un système de gravure par pixel comme le montre la figure suivante.

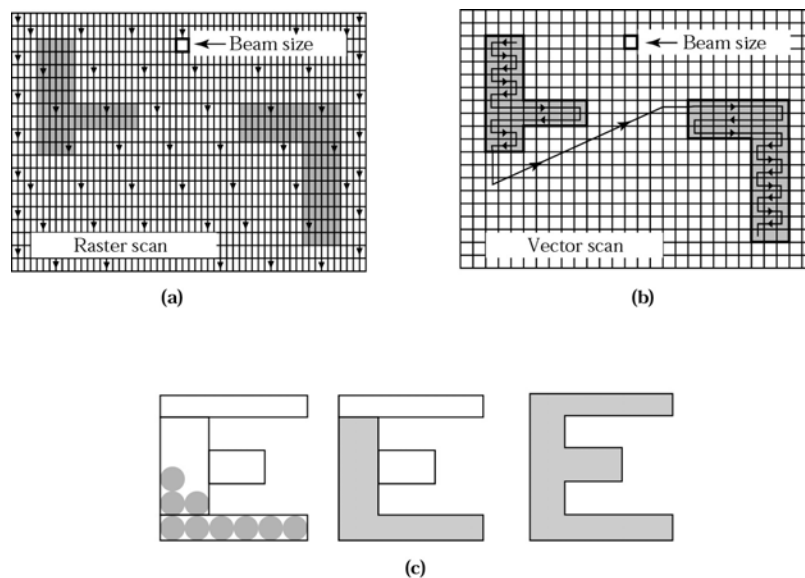


Figure 21 : dessin avec une e-beam lithographie

2.3. Quelques comparaisons

Pour terminer ce chapitre sur la photographie, nous nous attacherons à quelques chiffres et comparaisons fournis par Intel.

| <i>Pixels</i> | |
|---|---------------------------------------|
| Sur un masque en technologie 90nm | 1 000 000 000 000 |
| Sur une photo numérique | 5 000 000 |
| <i>Défauts</i> | |
| Tailles des défauts devant être détectés et corrigés | 0,1 μm |
| Nb de ces défauts supportés | 0 |
| Rapports de tailles entre ces défauts et le un masque | Terrain de football/Californie |
| <i>Données</i> | |
| Nombre typique de masques 90nm pou un process | 22-25 |
| Taille des données nécessaires pour ces masques | 200 GOctets |
| <i>Coûts</i> | |
| Coûts d'un masque 90nm | 1 M\$ |

Figure 22 : Chiffre et données sur les masques (Intel)

3. Séquences de fabrication CMOS

3.1. Introduction

Nous nous attacherons dans ce chapitre à décrire les principales étapes de la réalisation pratique d'un inverseur CMOS. Nous présenterons en parallèle les éléments clefs des technologies utilisées. Nous commencerons tout d'abord par rappels brièvement la structure des transistors NMOS et PMOS puis celle de l'inverseur CMOS. Des éléments plus précis ont déjà été donnés dans le chapitre précédent de ce documents et surtout dans le cours de semi-conducteur de la majeure MNO.

3.1.1. Le transistor NMOS

Sans revenir sur la physique des transistors CMOS de type N nous nous contenterons ici de rappeler sa structure comme le montre la figure 23.

- Dans un substrat de type P on été générées deux zones fortement dopées N. Ces dernières connectées au métal 1 correspondent au drain et à la source.
- La grille réalisée par une couche de poly-silicium est isolée du canal par de la silice.
- Une autre couche de silice isole enfin le transistor de ces congénères et protège aussi la grille.

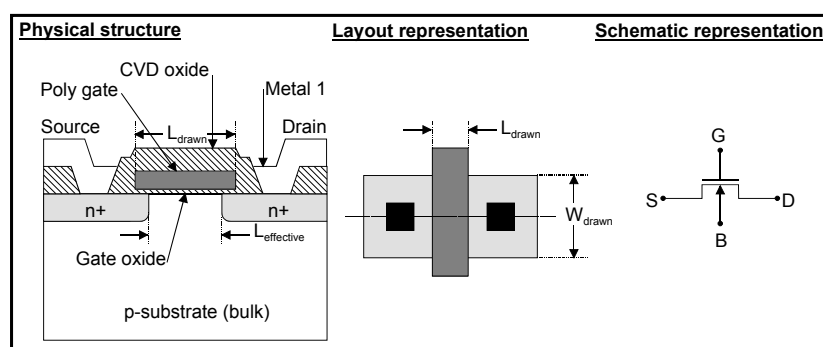


Figure 23 : Transistor NMOS

Sur la figure 23 est aussi représentée le Layout du transistor c'est-à-dire la représentation en 2 dimensions des différents masques (ou zones) du transistor NMOS.

3.1.2. Le transistor PMOS

La figure 24 donne de manière similaire la représentation physique et le Layout du transistor PMOS

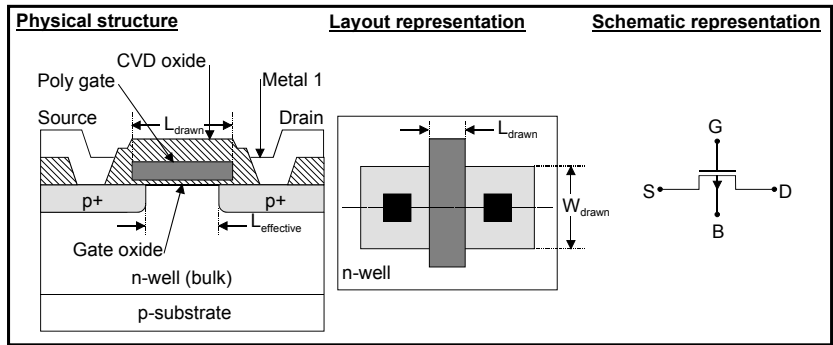


Figure 24 : Transistor PMOS

3.1.3. Inverseur CMOS

Pour décrire au mieux les techniques de fabrication CMOS nous nous baserons sur la conception d'un inverseur CMOS, tels que celui représenté sur la figure suivante :

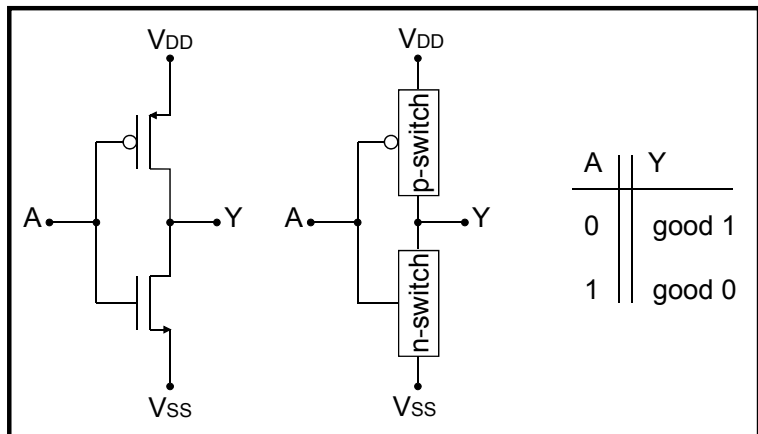


Figure 25 : Inverseur CMOS

Nous décrivons pour des raisons pédagogiques ici le process de fabrication dit LOCOS «**LOC**al **O**xidation of **S**ilicon » qui se caractérise principalement par le fait que l'oxyde de champ FOX « **Thick Field OXide** » est positionné sur toutes les régions non actives.

3.2. Etape 0 : le Wafer

3.2.1. Introduction

Avant de nous occuper de la conception de l'inverseur à proprement parlé il est nécessaire de définir au mieux le matériau de départ la tranche de monocristal de silicium (Wafer)

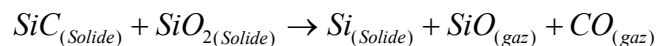
3.2.2. Croissance du cristal de silicium

Les 2 semi-conducteurs utilisés pour la conception des circuits intégrés sont le Silicium et l'Arséniure de Gallium. Ce dernier étant actuellement peu utilisé nous nous limiterons dans ce chapitre à la description de la croissance des monocristaux de silicium.

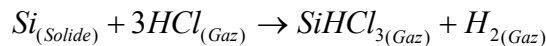
3.2.2.1. Matériau initial

Le matériau de départ est un type particulier très pure de sable (SiO₂) nommé « Quartzite ».

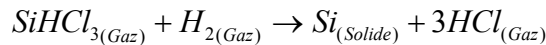
Il est placé dans un four avec différente forme de carbone (Charbon, Coke, bois), la réaction principale obtenue est :



On obtient ainsi un Silicium métallique avec une pureté de l'ordre de 95%. L'étape suivante consiste en une pulvérisation du Silicium suivie d'une attaque à l'acide chlorhydrique (HCl) pour former du Trichlorosilane :



Le Trichlorosilane étant liquide à température ambiante une simple distillation fractionnée permet d'éliminer les impuretés. La dernière étape consiste alors à une réduction du Trichlorosilane purifié sous hydrogène.



On obtient ainsi du Silicium de qualité électronique (Electronic-Grade Silicon), il se caractérise par une concentration d'impureté inférieure à une partie par milliard.

3.2.2.2. La technique dit « Czochralski »

La technique dit « Czochralski » utilise un appareil nommé extracteur de cristal. Une présentation simplifiée est donnée sur la figure 26.

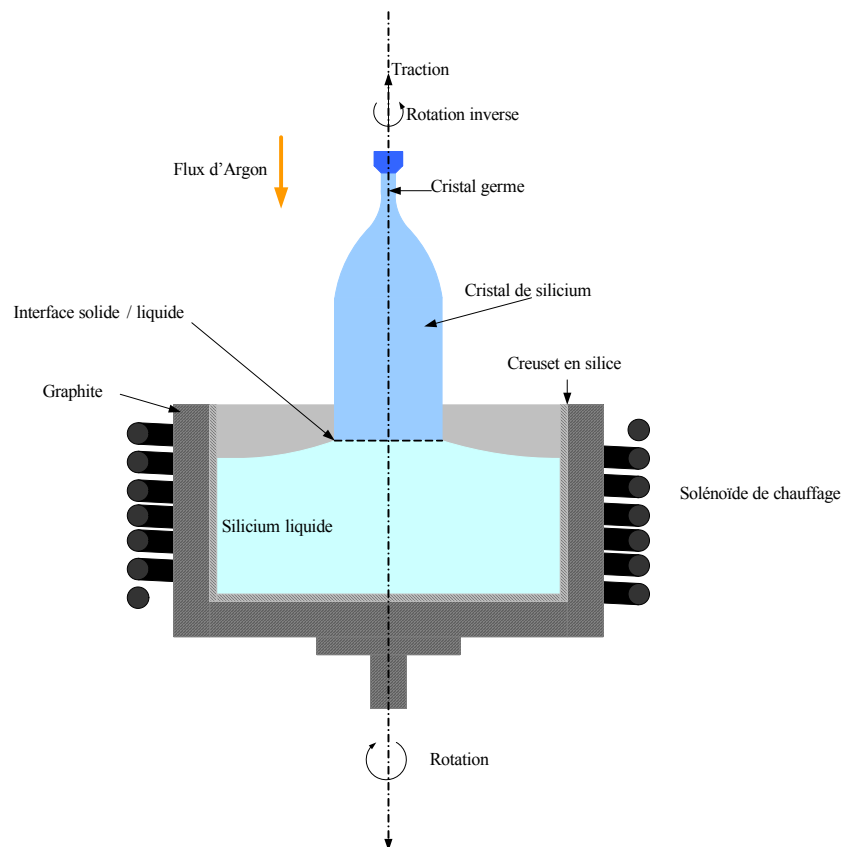


Figure 26 : Méthode Czochralski

L'extracteur se décompose en 3 parties :

- Un four comportant un creuset en silice contenant le silicium en fusion. Ce creuset est dans une structure en graphite qui est entraînée en rotation dans le sens des aiguilles d'une montre. Le chauffage de l'ensemble se fait par induction grâce à des électroaimants.
- Le mécanisme d'extraction du cristal qui se compose d'un germe c'est à dire d'un monocristal de silicium qui sert de précurseur, d'un support de germe qui est entraînée en

rotation dans le sens contraire des aiguilles d'une montre, mais aussi en translation vers le haut au fur et à mesure de la croissance du cristal.

- Un système de contrôle qui inclut le contrôle de l'atmosphère (argon), de la température, des vitesses de rotation et de translation.

De manière synthétique, on peut décrire le processus de croissance du cristal de la manière suivante :

- On place dans le creuset le silicium EGS précédemment préparé on le chauffe jusqu'à sa température de fusion.
- Un cristal de silicium orienté « 111 » ou « 100 » est suspendu sur le support de germe. Ce germe est partiellement plongé dans le silicium en fusion, la fusion d'une partie de ce dernier engendre une interface liquide/solide.
- On commence alors à retirer lentement le germe du silicium en fusion et le refroidissement progressif permet la croissance d'un grand cristal ayant la même orientation que le germe.

La vitesse de croissance classique est de l'ordre de quelques mm/min. pour obtenir les cristaux de silicium de grand diamètre on rajoute à la méthode initial un champ magnétique, ce dernier permet le contrôle de la concentration de défauts, d'impuretés et d'Oxygène.

La figure 27 présente 2 cristaux obtenus par la méthode Czochralski le premier à un diamètre de 300mm (12''), le second à un diamètre de 400mm (16'').

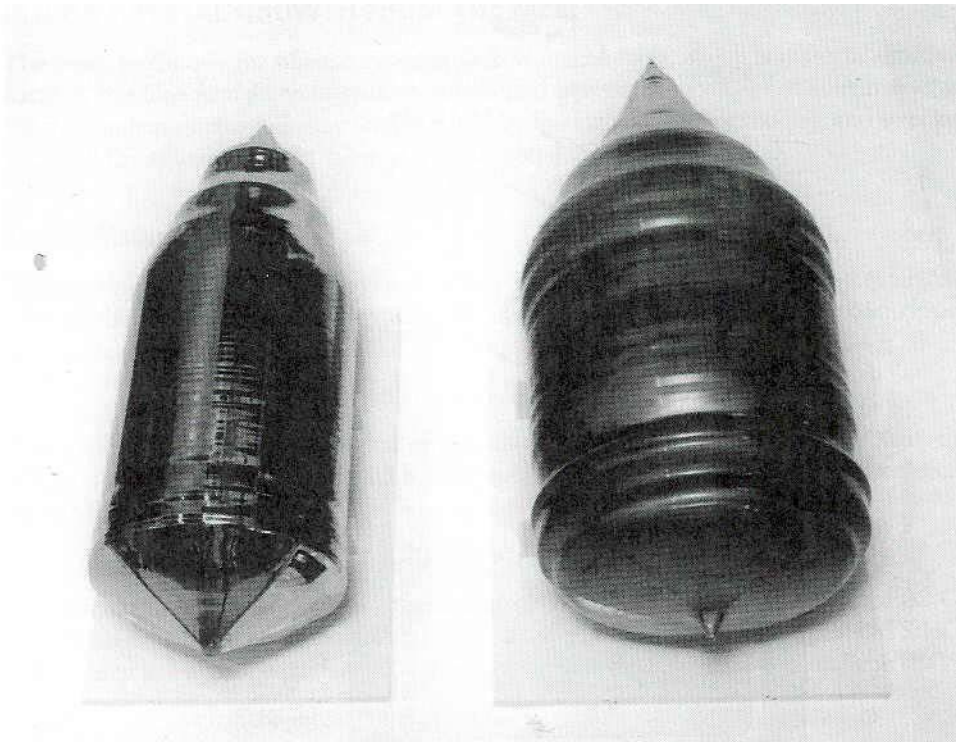


Figure 27 : lingots de silicium monocristallin

Lors de la croissance du cristal une quantité connue de dopant est mélangé au silicium en fusion pour obtenir un cristal ayant le dopage désiré. Ainsi avec le silicium, pour un dopage de type « P » on utilise le Bohr et pour un dopage de type « N » du phosphore.

3.3. Etape 1 : Caisson dopé N

3.3.1. Introduction

On commence par un substrat de type P C'est-à-dire un wafer de silicium contenant des impuretés constituées d'atomes de Bore. La réalisation de transistors de types PMOS nécessite au préalable la réalisation de zone dopé N que l'on nomme caisson N. Il est à noter que dans les process submicroniques profonds il est d'usage de faire aussi des caissons P pour les transistors NMOS (on contrôle ainsi mieux le dopage du canal de ces transistors).

Les figures suivantes montrent les 6 tâches nécessaires à la réalisation du caisson N

- La croissance de l'oxyde de protection
- Le dépôt de la résine photorésistante
- Le patterning par masquage de cette résine
- La gravure de l'oxyde de silicium
- L'implantation ionique de la zone N

3.3.2. Principe

Le principe de réalisation de cette première étape est décrit par les quatre figures suivantes (elles donnent la représentation en coupe et sous forme de Layout du futur inverseur)

3.3.2.1. Silice et résine photorésistante

On réalise tout d'abord une oxydation du silicium, cet oxyde servira de masque pour la future implantation ionique des atomes de Bore. Puis on dépose par tournette une couche de résine photorésistante, que l'on insole à travers le masque du caisson N (lithographie)

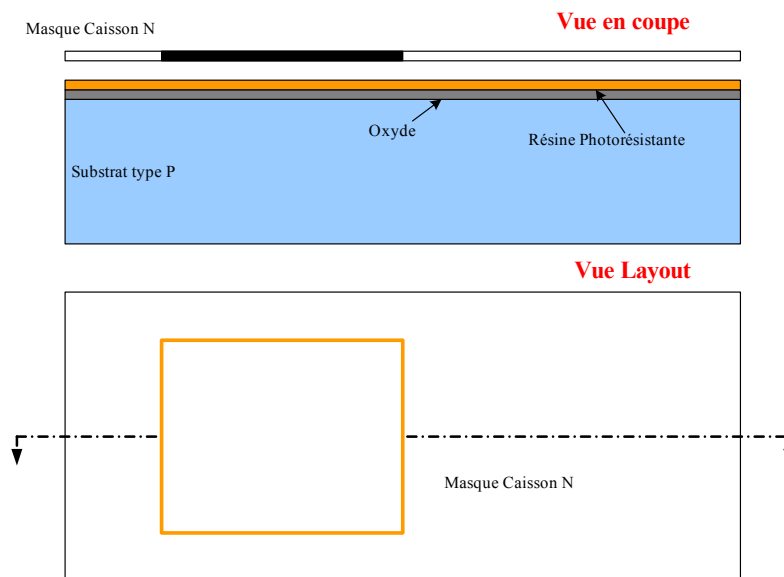


Figure 28 : Etape 1 (Silice et Résine photorésistante)

3.3.2.2. Développement de la résine

Le développement enlève la résine qui n'a pas été insolée. Seule la zone correspondant au futur caisson N laisse apparaître la silice.

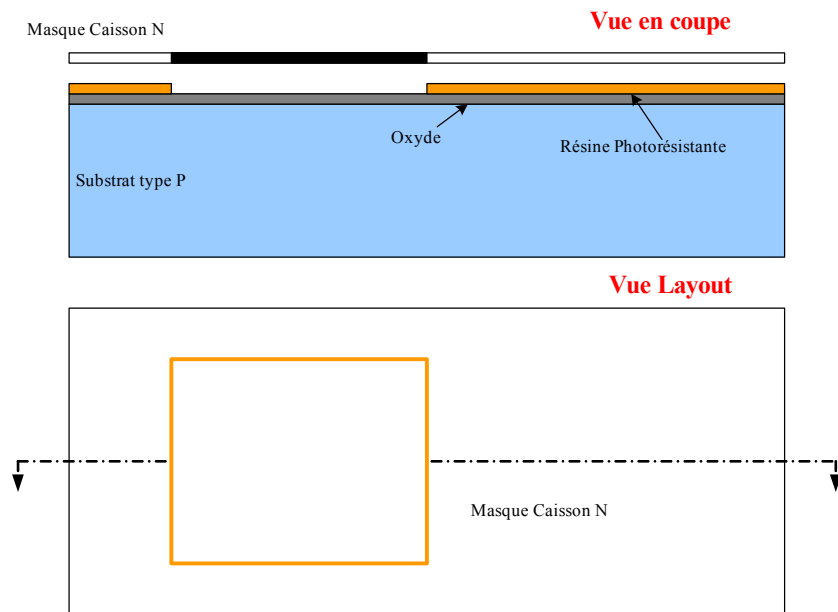


Figure 29 : Etape 1 (développement de la résine)

3.3.2.3. Gravure de la silice

On grave ensuite la silice par gravure chimique ou par plasma. On obtient comme le montre la figure 30 la zone correspondant au caisson N nue directement le silicium) le reste du wafer est recouvert donc protégé par l'oxyde de silicium.

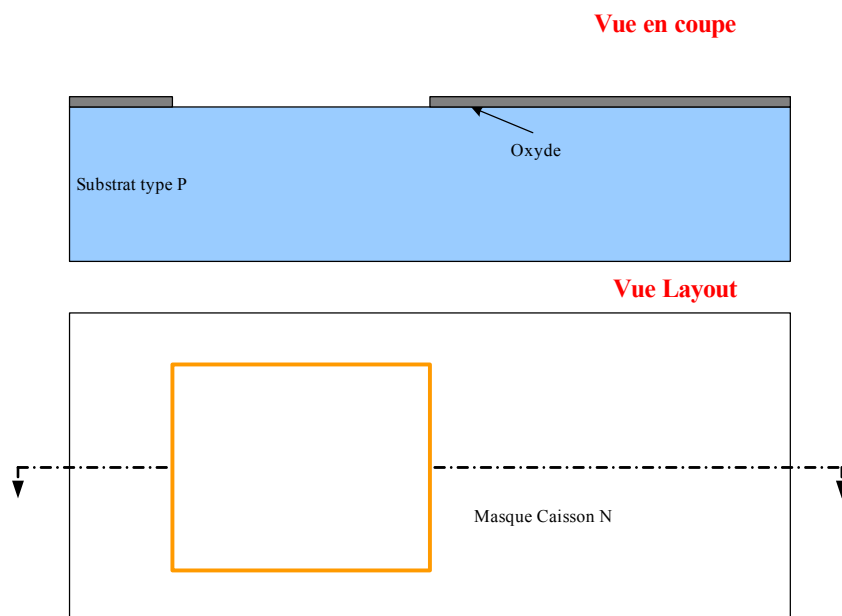


Figure 30 : Etape 1 (gravure de la silice)

La résine est ensuite enlevée par dissolution chimique

3.3.2.4. Implantation ionique du Bore

On réalise enfin l'implantation ionique du bore pour réaliser la zone dopé N. Seule la zone non protégée par la silice est dopée cette dernière réalise donc un masque qui définit proprement la zone du caisson N

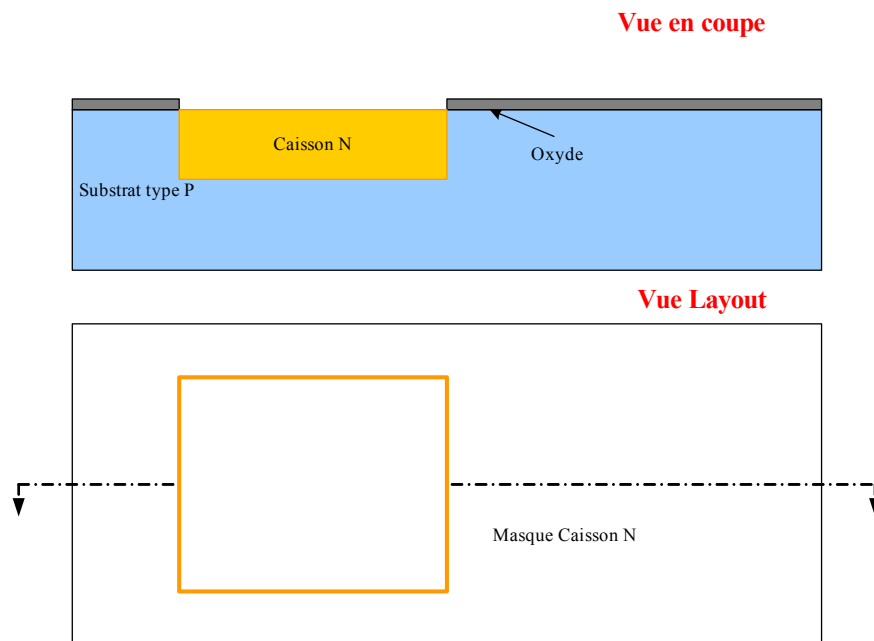


Figure 31 : Etape 1 (Implantation du Bore)

3.3.3. *Technologies associées à l'étape 1*

3.3.3.1. Oxydation du silicium

Différents types de couches minces sont utilisées pour la fabrication des composants discrets et des circuits intégrés : Oxydes Thermiques, couches diélectriques, silicium polycristallin et couches métalliques.

Oxydation thermique

Les semi-conducteurs peuvent être oxydés par différentes méthodes, tels que l'oxydation thermique et le dépôt par plasma PECVD (Plasma-Enhanced Chemical Vapor Deposition). La silice (SiO_2) obtenue par ces méthodes est de très bonnes qualités d'isolation électrique et de protection de surface mais aussi par sa facilité de fabrication, est une des principales raisons de l'utilisation du silicium dans les composants actuels. Nous nous concentrerons dans ce chapitre sur l'oxydation thermique la plus utilisée le chapitre « Dépôt de couche mince fera un rappel sur les croissance par PECVD ».

Le principe de base de l'oxydation thermique est décrit par la figure 32.

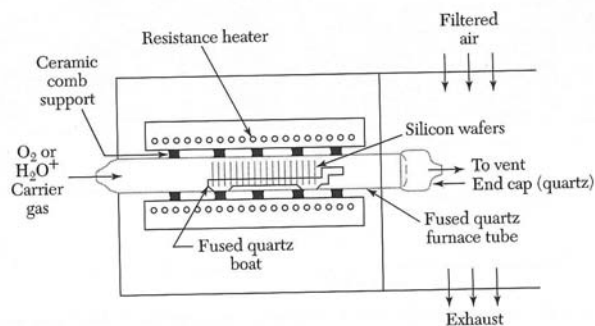


Figure 32 : réacteur d'oxydation thermique du silicium

Le réacteur se compose d'un système de chauffage par effet joule. Au centre se trouve un tube de quartz qui contient les wafers de silicium placés verticalement. Dans ce tube on crée un flux continu d'oxygène pur gazeux ou de vapeur d'eau (pour les couches épaisses). Ce flux est contrôlé via un écoulement d'air purifié.

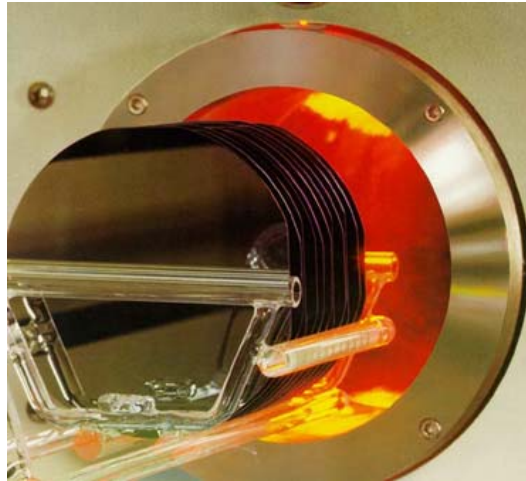


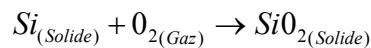
Figure 33 : Introduction de wafer dans un four d'oxydation

La température se situe de manière classique dans une plage de 900°C à 1200°C avec un débit de gaz d'environ 1l/min.

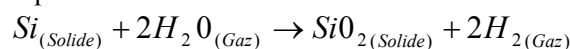
Vitesse de croissance

Les réactions thermiques mise en œuvre dans le cadre de l'oxydation thermique du silicium sont très simples :

Pour une oxydation par oxygène.



Pour une oxydation par vapeur d'eau.



La croissance de la silice se fait bien sûr aux dépens du silicium mais du fait de la différence de densité et de poids moléculaire de ces 2 éléments l'obtention d'une couche de silice nécessite la consommation d'une couche de silicium 0,44 fois plus fine.

Volume d'une mole de silicium

$$\frac{\text{Poids moléculaire du Silicium}}{\text{Densité du Silicium}} = \frac{28,9 \text{ g/mol}}{2,33 \text{ g/cm}^3} = 12,06 \text{ cm}^3 / \text{mol}$$

Volume d'une mole de silice

$$\frac{\text{Poids moléculaire du Silice}}{\text{Densité du Silice}} = \frac{60,08 \text{ g/mol}}{2,21 \text{ g/cm}^3} = 27,18 \text{ cm}^3 / \text{mol}$$

Si on suppose qu'une mole de silicium se transforme en une mole de silice

$$\frac{\text{Epaisseur de Si} * \text{surface}}{\text{Epaisseur de SiO}_2 * \text{surface}} = \frac{\text{Volume d'une mole de Si}}{\text{Volume d'une mole de SiO}_2} = \frac{12,06}{27,18} = 0,44$$

Ainsi par exemple la croissance d'une couche de 100nm de silice consomme une couche de Silicium de 44nm.

Croissance d'oxyde mince

Dans les technologies actuelles la couche de silice de grille est très mince comme le montre la figure suivante (Intel)

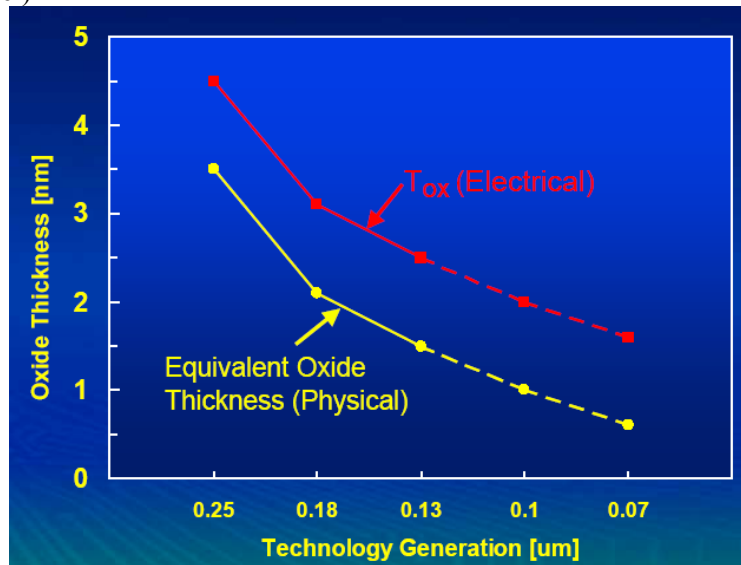


Figure 34 : Epaisseur d'oxyde de grille suivant les Technologies

La croissance de cette silice se fait à pression atmosphérique et à relativement basse température (800°C à 900°C). Afin de mieux contrôler la vitesse et d'assurer une parfaite reproductibilité des épaisseurs on réduit la pression partielle d'O₂ par une dilution par des gaz inerte comme N₂, Ar ou He.

Propriété de masquage de l'oxyde de silicium

Une autre propriété intéressante de la silice est sa capacité d'être utilisée comme masque sélectif contre la diffusion des dopants à haute température. En effet la vitesse de diffusion est de l'ordre de 2 à 3 ordres de grandeur plus faible dans la silice que dans le silicium. Cette caractéristique est très utilisée dans le processus de fabrication des circuits intégrés.

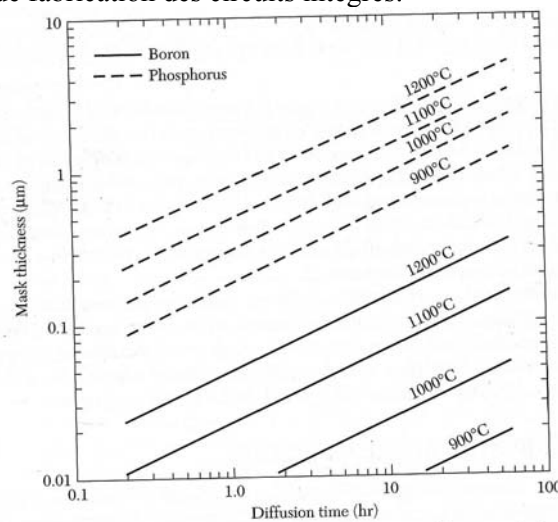


Figure 35 : Effet de masquage des dopants de la silice

Il est à noter que la silice utilisée pour le masquage est classiquement obtenue par oxydation humide. L'oxydation sèche qui donne une silice de plus grande densité et donc de plus grande tension de claquage (5-10 MV/cm).

3.3.3.2. Gravures

L'étape technologique précédente la photolithographie permet de transférer les dessins des masques sur la résine. La gravure va permettre de reproduire ces dessins sur les couches minces situées en dessous. En fonction des matériaux à graver et des dessins à réaliser on trouve deux types de gravure :

- Une gravure humique qui utilise un produit chimique.
- Une gravure sèche qui utilise les plasmas

Gravure humide

La gravure chimique (ou humide) est très utilisée dans l'ensemble des processus de fabrication des circuits intégrés. Outre pour la reproduction des structures des masques, elle est utilisée pour le polissage des wafers, On l'utilise aussi pour éliminer les traces d'impuretés et de contaminations préliminairement à une oxydation ou à une croissance par épitaxie.

- La gravure chimique, comme le montre la figure 33 peut se décomposer en 3 étapes :
- Le produit chimique (réactant) se dépose par diffusion sur la surface à graver.
- La réaction chimique se produit à la surface.
- Le produit de cette réaction est éliminé par diffusion.

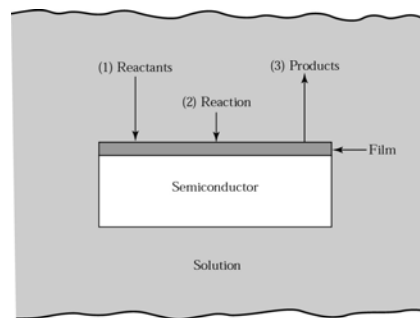


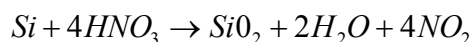
Figure 36 : Mécanismes basiques d'une gravure humide

Il est à noter que la vitesse de gravure peut être accrue et contrôlée par la maîtrise de la température et une agitation thermique. De manière générale une gravure humide se fait par immersion du wafer dans la solution de gravure mais afin d'accroître la vitesse et améliorer l'uniformité on utilise de plus en plus une gravure par pulvérisation.

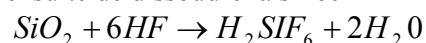
Gravure du silicium

Pour le silicium le mécanisme de gravure le plus courant est basé sur une oxydation préliminaire puis par une dissolution de la silice. Ainsi on utilise un mélange d'acide nitrique (HNO_3) et d'acide hydrofluorhydrique (HF) :

Oxydation à l'acide nitrique



L'acide fluorhydrique permet ensuite de dissoudre la silice



L'eau peut être utilisée comme diluant mais on préfère classiquement de l'acide acétique.

On remarque que certaines solutions de gravure ont des vitesses de gravure différentes suivant les plans du cristal de silicium. Ainsi le plan (111) a une bien plus grande résistance à un mélange de KOH d'eau et d'isopropanol. Ainsi une solution de 19% de KOH dans de l'eau désionisée à $80^\circ C$ élimine le plan (100) plus rapidement que le plan (110) et (111). Le ratio de vitesse est 100:16:1.

Dans ce cas, on obtient des formes gravure telles que celle représenté sur la figure 34 (noté l'angle de $54,7^\circ$ du plan (111)).

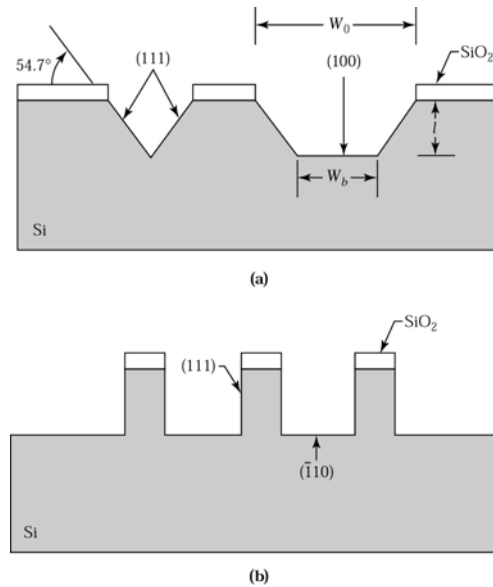


Figure 37 : Différence de gravure en fonction de l'orientation du cristal de silicium

La gravure humide du polysilicium se fait de manière similaire à celle du silicium monocristallin.

Gravure de la silice

La gravure humide de la silice se fait généralement par une solution de HF de manière similaire à ce que nous avons vu pour la seconde étape de la gravure du silicium. Une gravure par HF en phase vapeur est aussi possible.

Un des problèmes de la gravure est la sélectivité en effet, on remarque que les solutions chimiques utilisées ne gravent pas qu'un seul matériau mais attaque (pas à la même vitesse tous les éléments. La sélectivité qui quantifie la propension d'une solution de graver un matériau plutôt qu'un autre ce mesure par le ratio des vitesses de gravure.

Quelques autres gravures

De manière succincte nous donnons dans le tableau suivant les produits classiquement utilisés pour graver les principaux matériaux utilisés dans la conception des circuits intégrés.

| Matériaux | Solution | Vitesse (nm/min) |
|--|--|------------------|
| SiO ₂ (silice) | HF NH ₄ F | 100 |
| | HF HNO ₃ H ₂ O | 12 |
| Si ₃ N ₄ (Nitrure de silicium) | HF NH ₄ F | 0,5 |
| Al (aluminium) | H ₃ PO ₄ | 10 |
| | HNO ₃ H ₂ O H ₃ PO ₄ CH ₃ COOH | 30 |
| Au (Or) | KI | 1000 |
| Mo (Molybdène) | I ₂ HNO ₃ H ₂ O H ₃ PO ₄ CH ₃ COOH | 500 |
| Pt (platine) | HNO ₃ H ₂ O HCl | 50 |
| W (tungstène) | KH ₂ PO ₄ KOH K ₃ Fe(CN) ₄ H ₂ O | 160 |

Figure 38 : Gravure humique des isolants et des conducteurs

Gravure sèche

Dans le cadre du transfert des dessins sur les couches minces d'un circuit intégré la résine sert de masque protégeant les zones que l'on ne désire pas graver. La plupart de ces couches minces sont amorphes ou polycristalline ; dans ce cas la gravure humide a une vitesse de gravure isotrope (vitesse latérale et verticale de gravure sont identiques) ce qui engendre des phénomènes de sous gravure comme le montre la figure 39b.

Pour obtenir un parfait transferts des dessins de la résine nécessaire dans le cadre des circuits VLSI/ULSI, on utilise une méthode de gravure anisotrope (verticale) ; C'est la gravure sèche ou gravure par plasma, figure 39c.

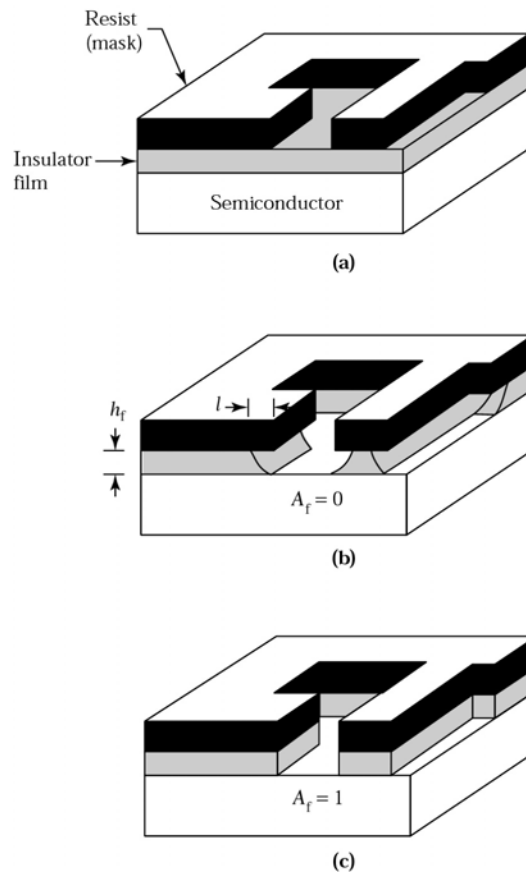


Figure 39 : Comparaison gravure sèche/ gravure humide

Rappels sur les plasmas

Un plasma est un gaz totalement ou partiellement ionisé composé d'un nombre égal de charges positives et négatives et de molécules non ionisées. Un plasma est généré par un champ électrique qui arrache des électrons libres à une électrode négative et les accélèrent. Lors de leurs déplacements ils percutent des molécules de gaz créant ainsi des ions.

Gravure par plasma

La gravure par plasma est basée sur la génération de plasma dans un gaz à base pression. Elle peut se décomposer en 5 étapes principales, comme le montre la figure 40 :

- Les composés de gravure (ions) sont générés dans le plasma.
- Ils sont transportés par diffusion à travers une couche de gaz stagnante sur la surface du wafer.
- Ils sont ensuite absorbés par la surface
- La réaction chimique (ajouté à l'effet physique du bombardement ionique) se déroule et forme un composé volatil.
- Finalement, ce dernier est désorbé de la surface, diffuse dans le gaz et est pompé par le système de vide.

On distingue deux types de gravure par plasma :

- Une gravure physique ou ce sont les ions positifs qui bombardent à grande vitesse (généralisé par le champ électrique) la surface on a un effet de gravure. Cette approche engendre une faible vitesse de gravure mais un très fort effet anisotropie.
- Une gravure chimique ou les ions engendrent une réaction chimique de gravure. Cette approche engendre une grande vitesse de gravure une bonne sélectivité mais une aussi une grave isotrope.

Dans la pratique les deux effets sont présents et conjugués.

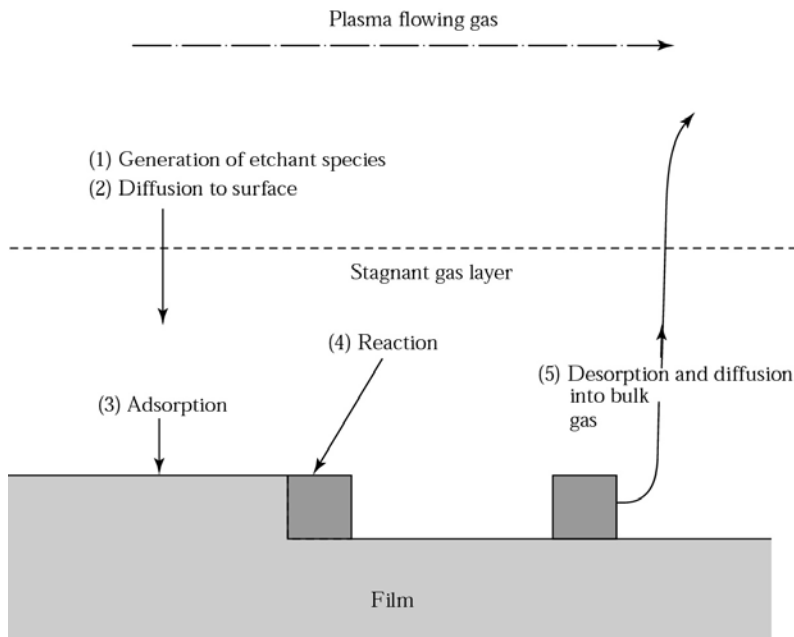


Figure 40 : principe de gravure par plasma

3.3.3.3. Méthodes d'implantation ionique

Principe de l'implantation ionique

Le dopant est ici implanté grâce à flux d'ions. On peut noter que la distribution de dopant à un pic situé sous la surface du wafer.

La concentration est ici fonction de la masse des ions mais aussi de l'énergie d'implantation. Classiquement on travaille avec des énergies comprises entre 1 keV et 1 MeV ce qui donne des profondeurs de distribution comprise entre 10 nm et 10µm avec une densité surfacique de dopant comprise entre 10^{12} ions/cm² et 10^{18} ions/cm².

Le principal intérêt de cette méthode est sa précision dans la distribution obtenue mais aussi sa répétitivité en outre elle peut se faire à température ambiante.

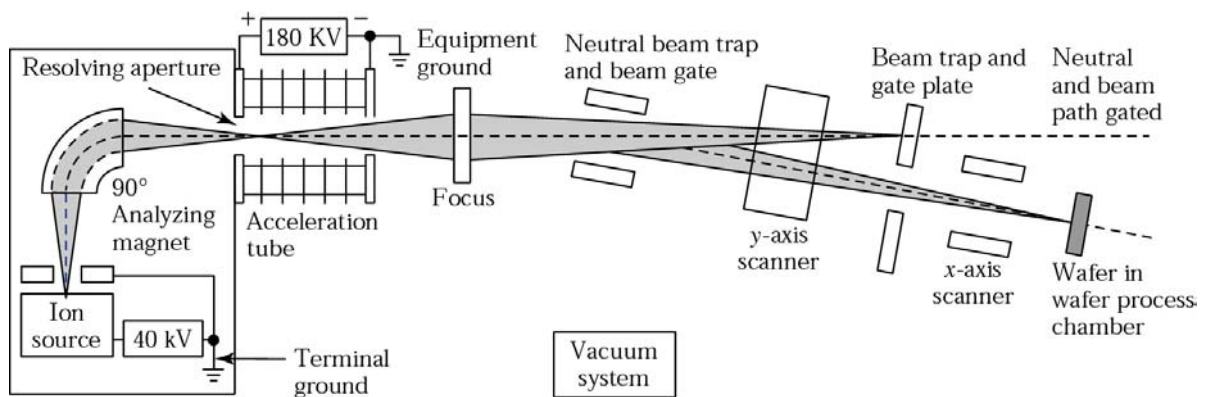


Figure 41 : Principe de l'implantation ionique

La figure 41 donne un schéma de description simplifié d'un implanteur ionique : La source contient un filament permettant de casser les molécules de gaz de dopant (PH_3 , BF_3 ,...). Un potentiel de 40 kV appliqué par des électrodes permet d'accélérer ces ions et de créer le flux. Un champ magnétique permet de choisir (au niveau de la sortie de la source) uniquement les ions ayant la bonne masse. Puis un tube d'accélération avec un champ électrique de 180kV (réglable) permet de choisir le niveau d'énergie voulue. Des plaques de déflexion ainsi qu'un focus permettent enfin de diriger le flux sur le wafer.

Afin de minimiser le risque de collision parasite de dépôt de fait dans des conditions de vide secondaire 10^{-4} Pa.

Domage engendré par l'implantation ionique

La collision des ions avec le wafer engendre des dégradations qu'un recuit permet de compenser en grande partie. Lorsqu'ils pénètrent dans le semi-conducteur, ils perdent leur énergie par une série de collisions avec les électrons et les noyaux des atomes de silicium. Dans le cas des noyaux, elles peuvent engendrer un déplacement des atomes (un désordre), déplacement qui ensuite se propage dans la structure du réseau cristallin comme le montre la figure 42. Cela peut aller jusqu'à créer en surface du wafer une couche de silicium amorphe (complètement désordonné)

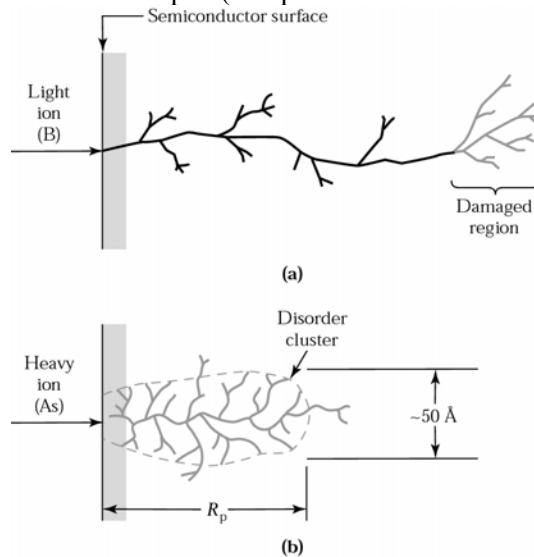


Figure 42 : dommages générés par l'implantation ionique

Recuit thermique rapide

Les dégradations du réseau cristallin engendrées par le dopage diminuent drastiquement les caractéristiques du semi-conducteur telles que la mobilité et la durée de vie. En outre les ions de dopant ne se positionnent pas toujours au bon endroit. Une étape de recuit est donc nécessaire pour activer le dopage et ainsi restaurer les caractéristiques du matériau.

Pour ce faire on utilise classiquement des machines de recuit thermique rapides telles que décrites sur la figure 40. Dans une enceinte basse pression (gaz inerte) une série de lampes (tungstène ou arc) permet d'appliquer sur le wafer dopé, avec un très fort gradient (100°C/s) une température de l'ordre de 600 à 1000°C .

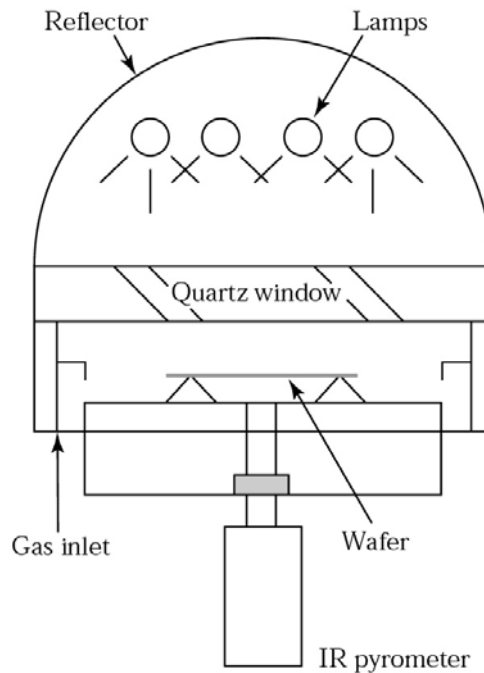


Figure 43 : Principe de recuit thermique rapide

3.4. Etape 2 : Définition des Zones actives

3.4.1. Introduction

La seconde étape consiste en la définition des zones actives c'est-à-dire des zones dopés N ou P qui correspondront soit aux drains et sources des transistors respectivement NMOS et PMOS soit aux contacts de caisson.

Les figures suivantes montrent les 7 tâches nécessaires à la définition des zones actives

- Dépôt du nitrure de silicium SiN
- Dépôt de la résine photorésistante
- Patterning de la résine photorésistante
- Gravure du SiN
- Élimination de la résine
- Croissance de l'oxyde de champ FOX
- Élimination du SiN

3.4.2. Principe

Le principe de réalisation de cette seconde étape est décrit par les cinq figures suivantes (elles donnent la représentation en coupe et sous forme de Layout du futur inverseur)

3.4.2.1. Nitrure de silicium et résine photorésistante

On réalise tout d'abord un dépôt par plasma de nitrure de Silicium Si_3N_4 , ce dernier servira de masque pour la future croissance d'oxyde de champ FOX. Puis on dépose par tournette une couche de résine photorésistante, que l'on insole à travers le masque des zones actives (lithographie)

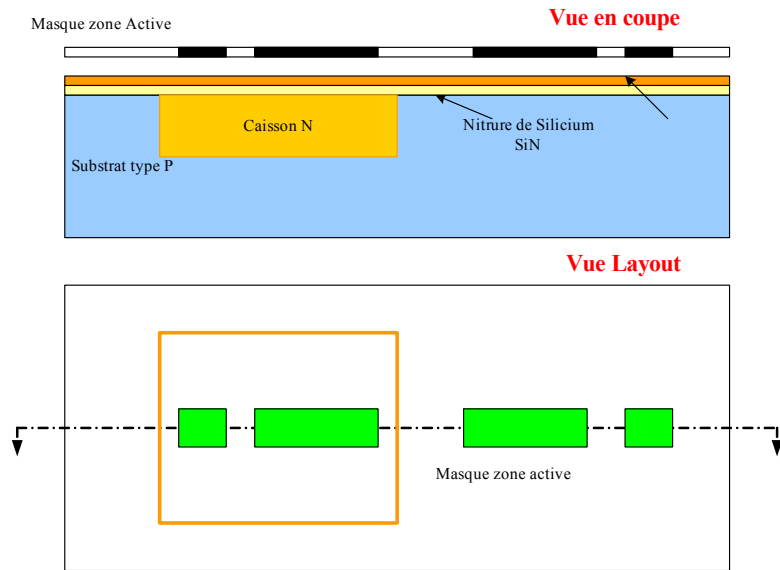


Figure 44 : Etape 2 (SiN et résine Photo résistante)

3.4.2.2. Développement de la résine

Le développement enlève la résine qui n'a pas été insolée. Seule la zone correspondant aux futures zones actives laisse apparaître le SiN.

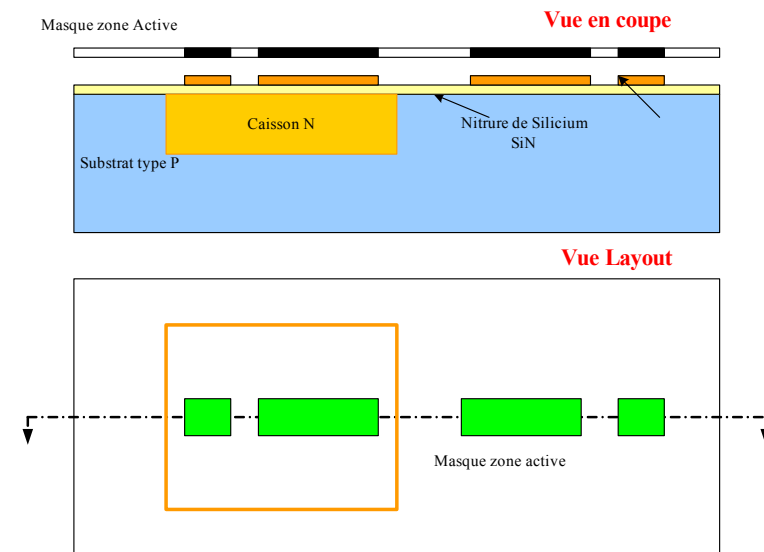


Figure 45 : Etape 2 (développement de la résine)

3.4.2.3. Gravure du nitrure de Silicium

On grave ensuite l'oxyde de silicium par plasma. On obtient comme le montre la figure 43 les zones ne correspondant pas aux régions actives nues directement le silicium). les futures zones sont recouvertes donc protégées par le nitrure de silicium.

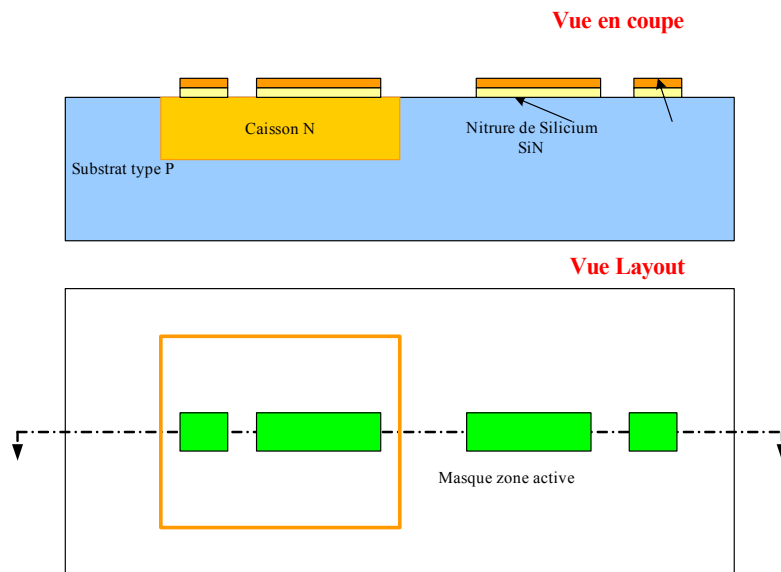


Figure 46 : Etape 2 (gravure du SiN)

La résine est ensuite enlevée par dissolution chimique

3.4.2.4. Croissance de l'oxyde de champ

Les zones non protégées par le SiN sont oxydées par oxydation thermique. La silice ainsi obtenue isole électriquement entre eux les différents éléments du wafer. On parle d'oxyde de champ « Field Oxyde (FOX) ».

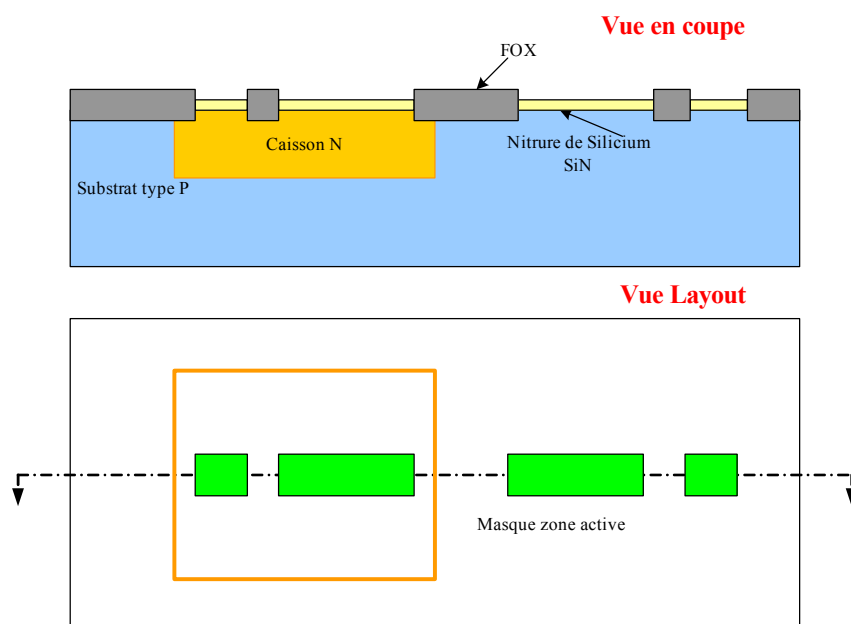


Figure 47 : Etape 2 (croissance de l'oxyde de champ)

3.4.2.5. Elimination du niture de silicium

On enlève enfin le niture de silicium. On a ainsi défini les zones actives qui seront ensuite dopées P ou N. Ces zones sont isolées par un oxyde de silicium épais dit oxyde de champ.

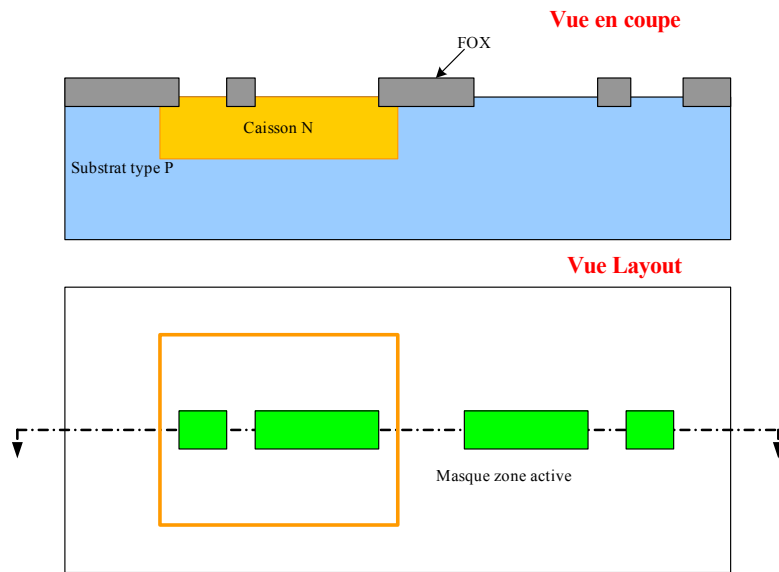


Figure 48 : Etape 2 (Elimination du SiN)

3.4.3. Technologies associées à l'étape 2

3.4.3.1. Déposition de diélectriques

Le dépôt de couches mince de diélectrique est utilisé principalement pour la passivation des composants discrets (isolation électrique). Ils existent trois méthodes classiques de déposition :

- Déposition chimique en phase vapeur à pression atmosphérique 'APVCD'.
- Déposition chimique en phase vapeur à basse pression 'LPVCD'.
- Déposition en phase vapeur avec assistance par plasma 'PECVD'.

Pour les 2 premières méthodes le schéma de principe et proposé sur la figure 49(a). Les wafers positionnés verticalement dans un four en quartz sont soumis à un flux de gaz.

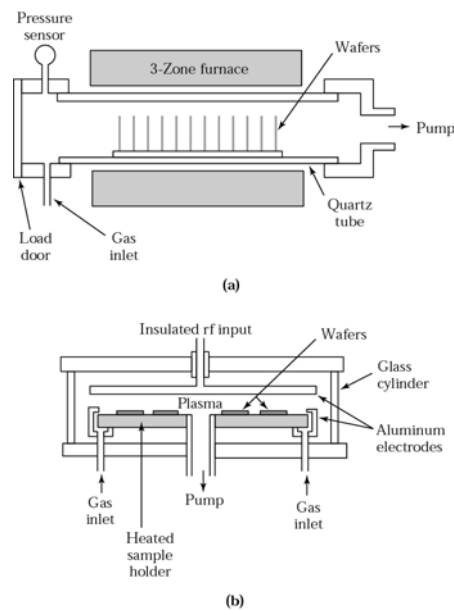


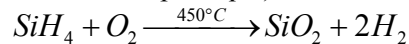
Figure 49 : principe de déposition de silice

Pour la PECVD (principe sur la figure 49(b)) l'énergie d'un plasma est ajouté à l'énergie thermique d'une CVD classique Ce qui permet de travailler à plus basse température (100 à 400°C). Par contre la capacité est limitée particulièrement au niveau de la taille des wafers. En outre une contamination est possible avec les dépôts précédents qui se sont fixé sur les parois de la chambre à plasma.

3.4.3.2. Déposition de silice

Une Silice obtenu par CVD est de qualité inférieure à celle obtenu par oxydation thermique. Ainsi cette silice est utilisée de manière complémentaire. Non dopé elle est utilisé comme isolant entre les différentes couches de métallisation ou comme masque dans une opération d'implantation ionique

A basse température (300 à 500°C) la réaction chimique de la CVD m'est en œuvre le silane et l'oxygène (risque d'explosion à pression atmosphérique)



3.4.3.3. Déposition de nitrure de silicium

La croissance thermique du nitrure de silicium est difficile du faite de la faible vitesse de croissance et de la haute température nécessaire. Il peut être obtenu soit par une LPCVD ver 750°C ou encore par une PECVD à 350°C.

La LPCVD permet d'obtenir un nitrure stochiométrique (Si_3N_4) et de grande densité (2,9 à 3,1 g/cm^3). Ce dernier est utilisé comme couche barrière vis-à-vis de l'eau et du sodium.

Le nitrure obtenu par PECV n'est pas stochiométrique et à une plus faible densité (2,4 à 2,8 g/cm^3). Du faite de l'utilisation d'une faible température ce nitrure est utilisé comme passivation finale des circuits intégrés et comme protection contre les rayures et contre l'humidité.

3.5. Etape 3 : réalisation de la grille

3.5.1. *Introduction*

La troisième étape consiste en la réalisation de la grille des 2 transistors en polysilicium.

Les figures suivantes montrent les 6 taches nécessaires à la définition des grilles

- Croissance de l'oxyde de grille
- Déposition du polysilicium
- Déposition de la résine photorésistante
- Développement de la résine
- Gravure du polysilicium
- Gravure de l'oxyde de grille

3.5.2. *Principe*

Le principe de réalisation de cette troisième étape est décrit par les trois figures suivantes (elles donnent la représentation en coupe et sous forme de Layout du futur inverseur)

3.5.2.1. Oxyde de grille polysilicium et résine photorésistante

On dépose tout d'abord sur l'ensemble du wafer l'oxyde de grille. Sa présence ne modifie par bien sur la partie FOX. La silice choisie se doit d'être de très bonne qualité diélectrique tout en étant extrêmement mince (>4 nm pour la technologie 90 nm). Elle est obtenue classiquement par oxydation thermique comme décrit dans §3.3.3.1.

On dépôt ensuite toujours sure l'ensemble du wafer le polysilicium avec les méthodes que nous décrirons dans le §3.5.3.

On recouvre enfin le tous par une résine photorésistante

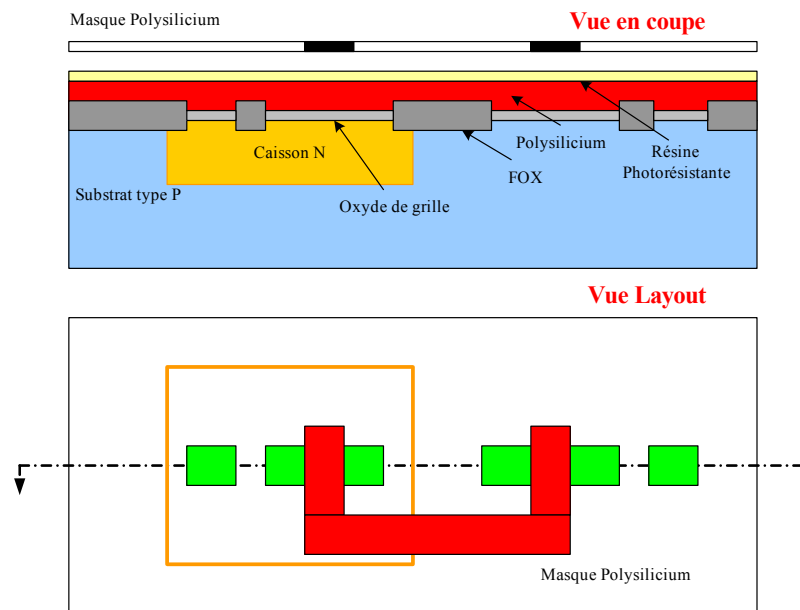


Figure 50 : Etape 3 (silice de grille, polysilicium & résine)

3.5.2.2. Gravure résine et polysilicium

La résine est insolée à travers le masque Polysilicium.

Le développement enlève la résine qui n'a pas été insolée. Elle protège ainsi le polysilicium correspondant aux grilles des 2 transistors. Le reste du polysilicium est éliminé par une gravure physique par plasma « Etching ».

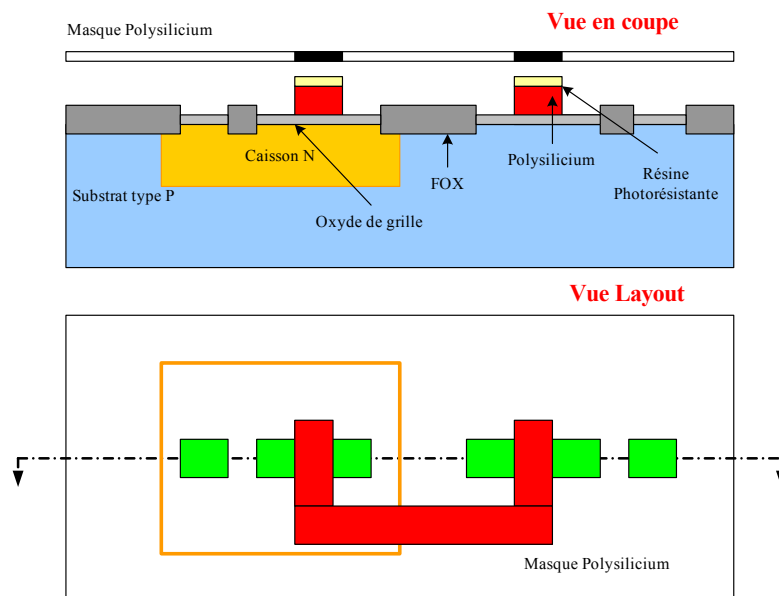


Figure 51 : Etape 3 (gravure résine & polysilicium)

On élimine ensuite la résine photorésistante restante sur le polysilicium.

3.5.2.3. Gravure de l'oxyde de grille

On élimine enfin l'oxyde de grille sauf sous les zones protégées par le polysilicium (concept d'auto alignement). On a ainsi généré les deux grilles des deux transistors MOS. Cette gravure est réalisée de manière physique par plasma.

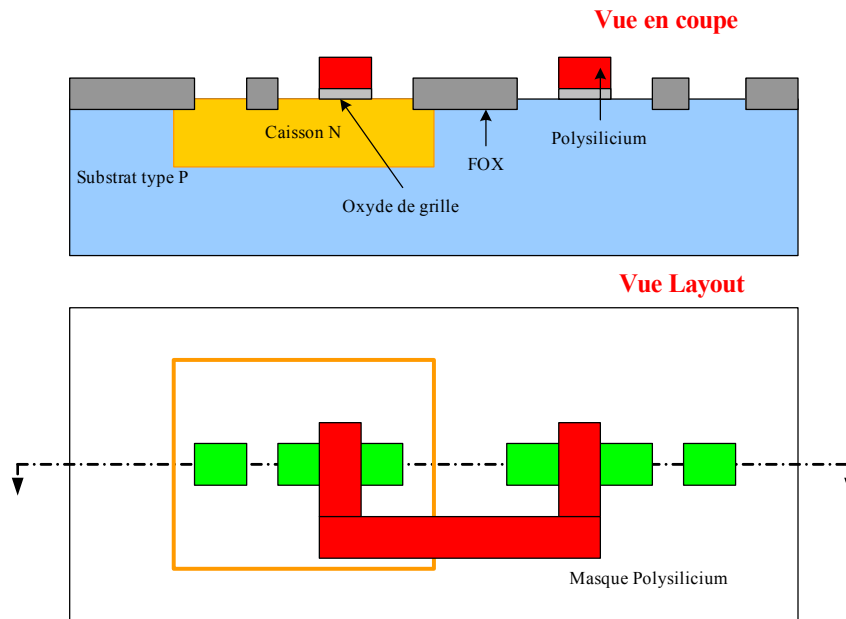


Figure 52 : Etape 3 (gravure oxyde de grille)

3.5.3. Technologie associée à l'étape 3 : Déposition de polysilicium

Le polysilicium pour la grille des transistors MOS c'est généralisé car il surpasse le silicium d'un point de vue de la fiabilité. La figure 53 montre une comparaison de la durée de vie des capacités de grilles. La plus faible durée pour l'aluminium est du à la migration des atomes de Al dans la silice sous l'action du champ électrique.

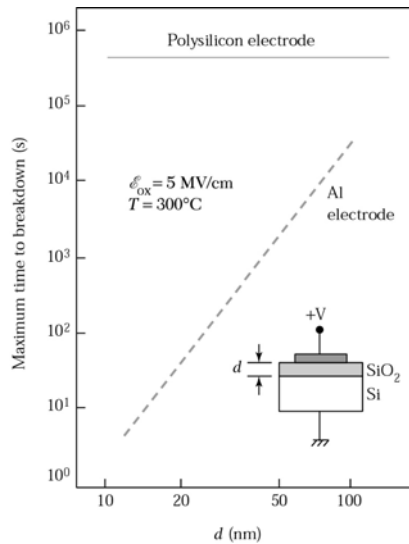
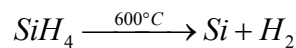


Figure 53 : intérêt du polysilicium face à aluminium

Le polysilicium est obtenu par LPCVD à 600/650°C par pyrolyse du silane sous une pression de 25 à 13à Pa.



3.6. Etape 4 : Zones actives dopage

3.6.1. Introduction

La quatrième étape consiste en la réalisation des zones dopées des 2 transistors ; dopée respectivement N+ pour le NMOS et P+ pour le PMOS.

3.6.2. Dopage P (transistor PMOS)

Le principe de réalisation de cette quatrième partie de la quatrième étape est décrit par les trois figures suivantes (elles donnent la représentation en coupe et sous forme de Layout du futur inverseur)

3.6.2.1. Résine photorésistante

On dépose sur l'ensemble du wafer la résine photorésistante

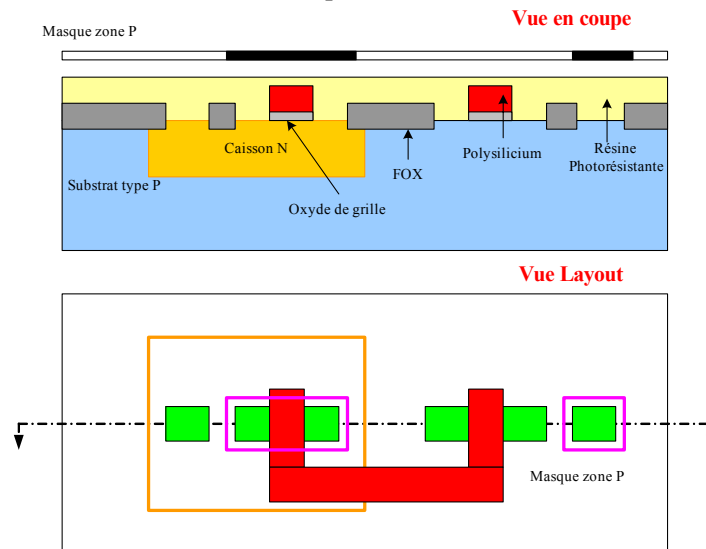


Figure 54 : Etape 4 (dopage P, dépôt résine)

3.6.2.2. Développement résine

La résine est insolée à travers le masque zone P. Le développement enlève la résine qui n'a pas été insolée. On laisse ainsi à découvert les zones qui doivent être dopées P.

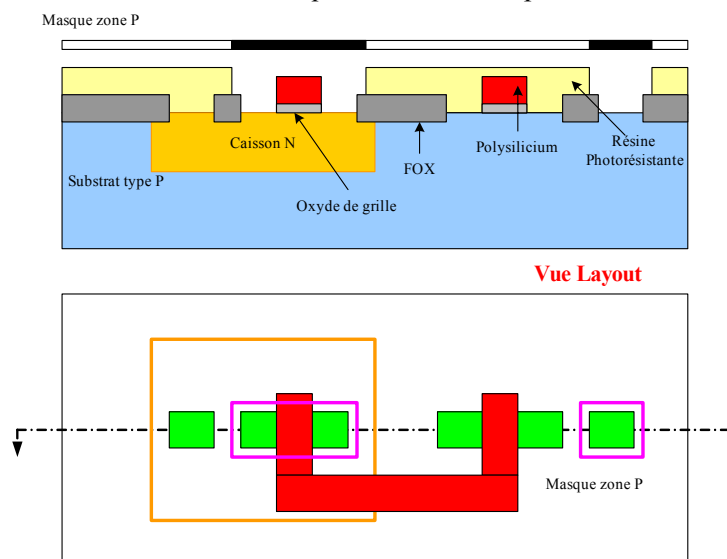


Figure 55 : Etape 4 (Dopage P, développement résine)

3.6.2.3. Implantation ionique zone P+

L'implantation ionique que nous avons décrite dans le paragraphe §3.3.3.3 permet de générer les zones dopé P+ c'est-à-dire la source et le drain du transistor PMOS mais aussi le contact du substrat dopé P.

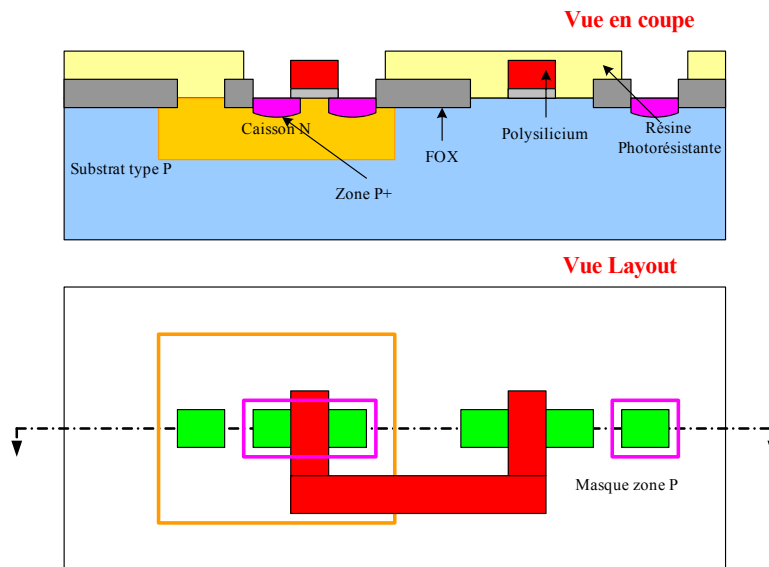


Figure 56 : Etape 4 (Dopage P, implantation ionique)

3.6.3. Dopage N (transistor NMOS)

Le principe de réalisation de cette seconde partie de la quatrième étape est décrit par les deux figures suivantes (elles donnent la représentation en coupe et sous forme de Layout du futur inverseur)

3.6.3.1. Résine photorésistante

On dépose sur l'ensemble du wafer la résine photorésistante

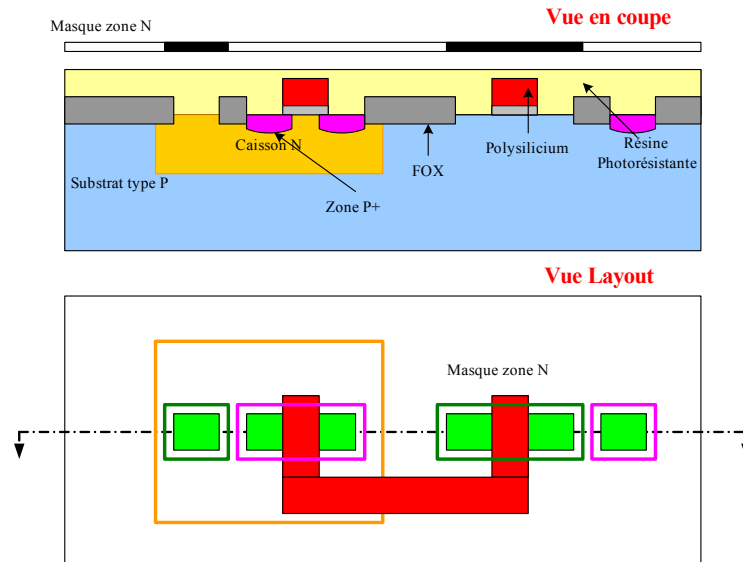


Figure 57 : Etape 4 (dopage N, dépôt résine)

3.6.3.2. Développement résine & implantation ionique zone N+

La résine est insolée à travers le masque zone N. Le développement enlève la résine qui n'a pas été insolée. On laisse ainsi à découvert les zones qui doivent être dopées N.

L'implantation ionique que nous avons décrite dans le paragraphe §3.3.3.3 permet de générer les zones dopé N+ c'est-à-dire la source et le drain du transistor NMOS mais aussi le contact du caisson N.

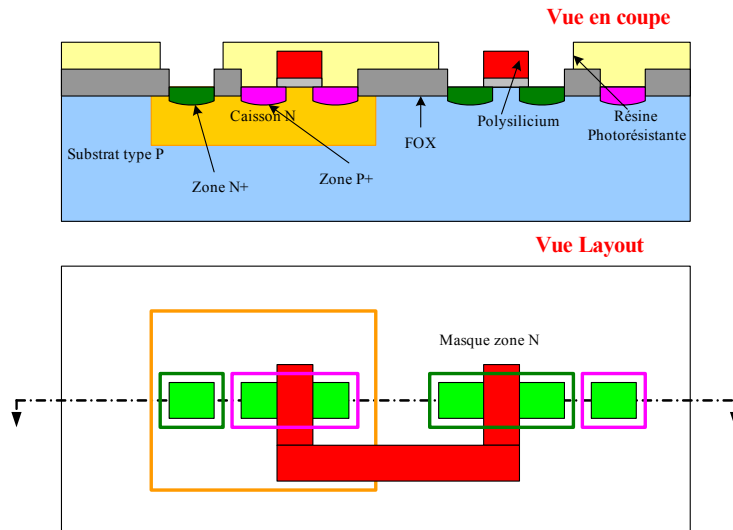


Figure 58 : Etape 4 (Dopage N, développement résine & implantation ionique)

3.7. Etape 5 : Via de contact

3.7.1. Introduction

La cinquième étape consiste en la réalisation des via de contact, c'est-à-dire les zones de contact entre les régions dopées et le métal 1 mais aussi le contact entre la liaison de grille en polysilicium et le métal 1.

Les figures suivantes montrent les 4 tâches nécessaires à la définition des grilles

- Croissance de la silice
- Dépôt de la résine photorésistante
- Développement de la résine
- Gravure de la silice

3.7.2. Principe

Le principe de réalisation de cette cinquième étape est décrit par les trois figures suivantes (elles donnent la représentation en coupe et sous forme de Layout du futur inverseur)

3.7.2.1. Dépôt silice et résine photorésistante

On dépose sur l'ensemble du wafer d'abord de la silice par oxydation thermique du silicium. Puis la résine photorésistante

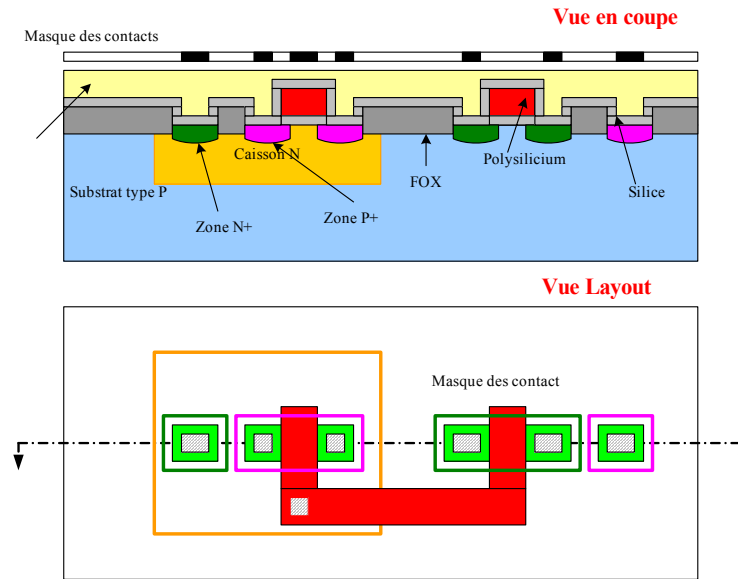


Figure 59 : Etape 5 (dépôt silice et résine photorésistante)

3.7.2.2. Développement de la résine

La résine est insolée à travers le masque zone N. Le développement enlève la résine qui n'a pas été insolée. On laisse ainsi à découvert la silice pour les zones qui vont correspondre au via.

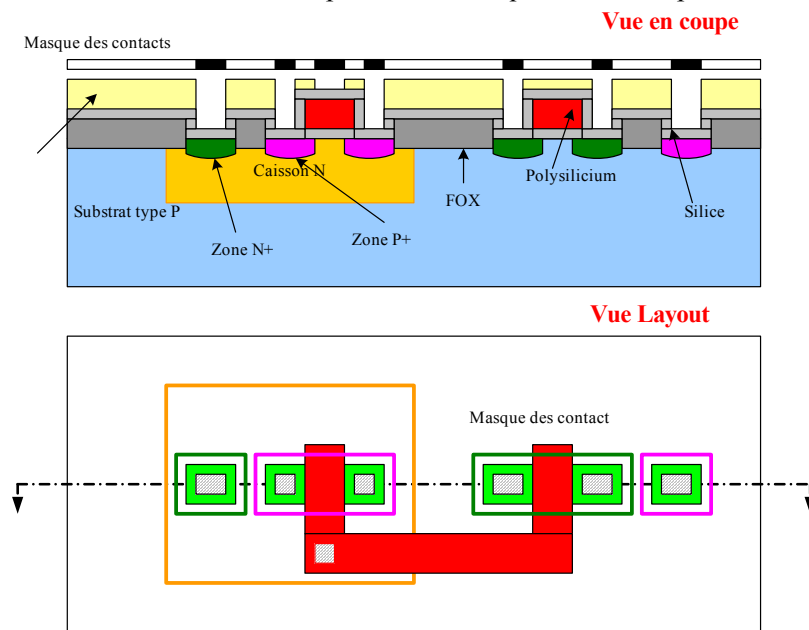


Figure 60 : Etape 5 (développement résine)

3.7.2.3. Gravure de la résine

On élimine enfin la silice dans les zones non protégées par la résine. On a ainsi généré les via ou la métal 1 sera directement en contact ohmique avec les zones dopées. Cette gravure est réalisée de manière physique par plasma.

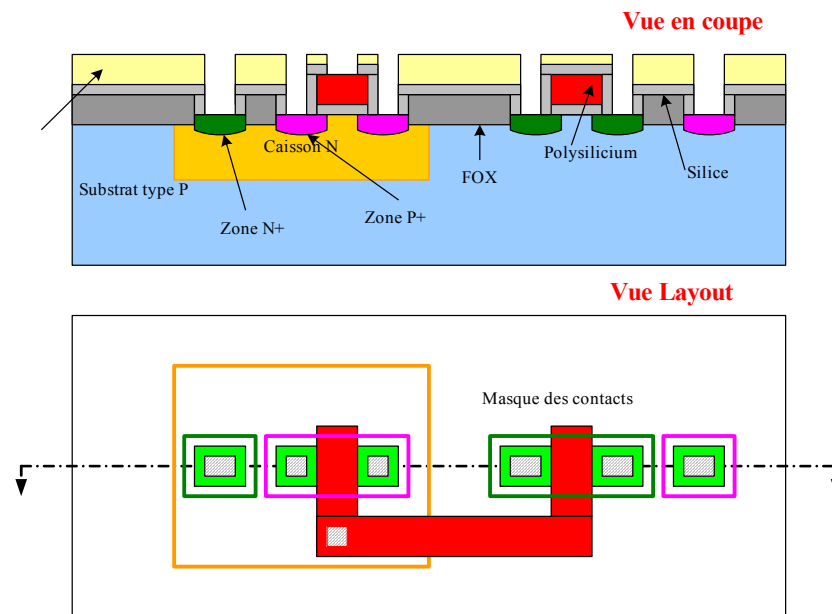


Figure 61: Etape 5 (gravure de la silice)

On élimine ensuite la résine restante par dissolution chimique

3.8. Etape 6 : Couche de métal 1

3.8.1. Introduction

La sixième étape consiste en la réalisation des connexions métal 1.

Les figures suivantes montrent les 6 tâches nécessaires à la définition des grilles

- Dépôt métal 1
- Planarization
- Dépôt de résine photorésistante
- Développement de la résine
- Gravure du métal 1
- Elimination de la résine photorésistante

3.8.2. Principe

Le principe de réalisation de cette sixième étape est décrit par les deux figures suivantes (elles donnent la représentation en coupe et sous forme de Layout du futur inverseur)

3.8.2.1. Dépôt métal 1 et résine

On dépose classiquement par pulvérisation le métal 1 sur l'ensemble du wafer. Le métal est soit de l'aluminium soit du cuivre (plus difficile mais meilleur conducteur).

Il est important pour la maîtrise des couplages capacitifs entre les couches métal que la couche de métal 1 soit parfaitement plate. Cette étape de polissage réalisé mécaniquement se nomme la « **planétarisation** ».

On dépose la résine photorésistante sur l'ensemble du wafer.

La résine est ensuite insolée à travers le masque du métal 1. Il est à noter que la prise du contact d'entrée de l'inverseur (sur le polysilicium) est décalé du plan de coupe de la figure suivante.

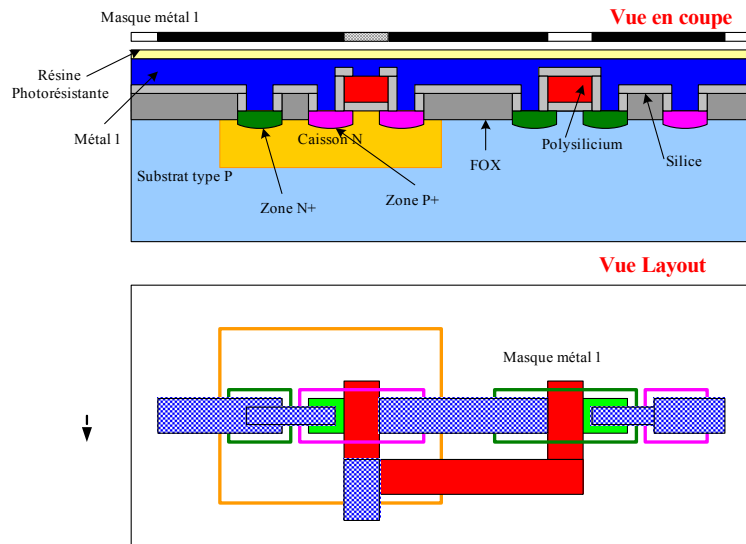


Figure 62 : Etape 6 (dépôt métal 1 et résine)

La résine est ensuite développée. Elle laisse ainsi découverte les zones du métal 1 qui doivent être retirées.

3.8.2.2. Gravure du métal 1

Le métal 1 non protégé par la résine est gravé par méthode physique de type plasma.

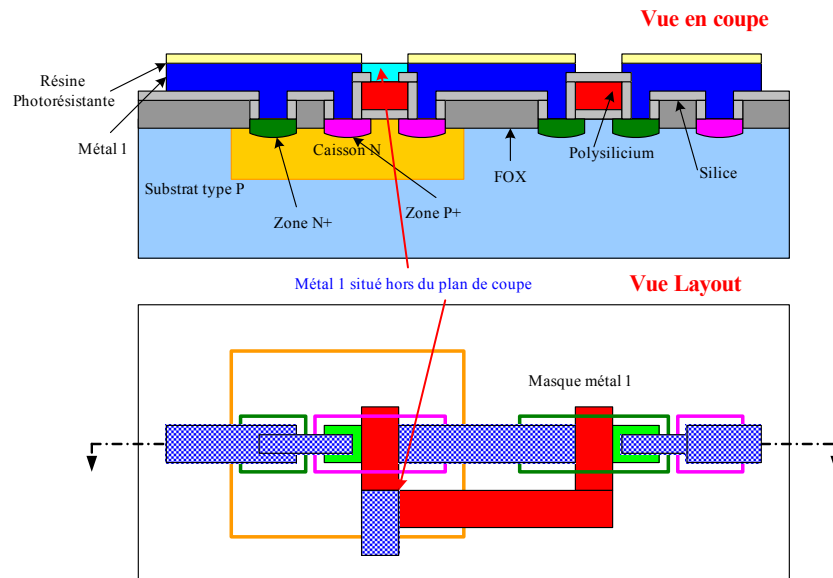


Figure 63 : Etape 6 (gravure métal 1)

On élimine enfin par dissolution chimique la résine photorésistante restante.

3.8.3. **Technologie associée : Méthode de métallisation**

3.8.3.1. Déposition par pulvérisation cathodique

La déposition par pulvérisation cathodique ou 'sputtering' utilise, dans une chambre à vide, un flux d'ions qui bombarde une cible de métal (Ti, Al, Cu, TiN) libérant des atomes qui viennent se déposer sur le substrat (voir figure 61).

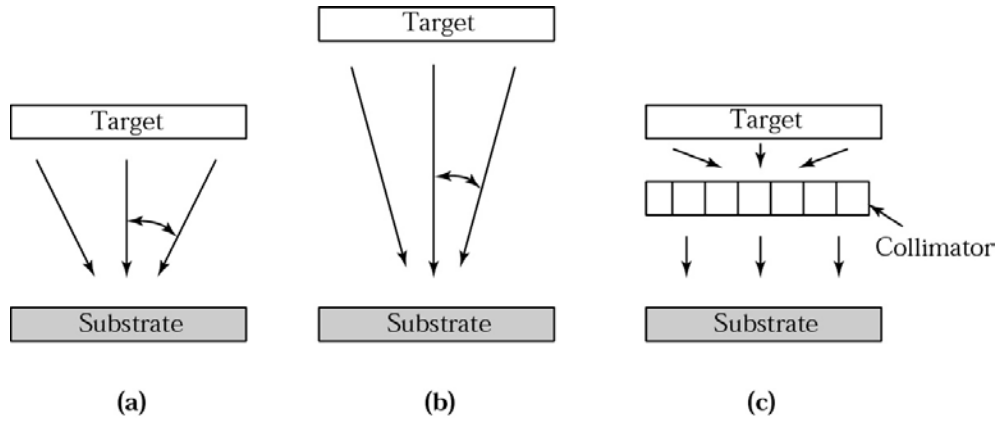


Figure 64 : Principe de sputtering

La figure 65 montre une machine industrielle de sputtering.



Figure 65 : Bâti MRC de dépôt de métaux et diélectriques par pulvérisation cathodique

3.8.3.2. Déposition par évaporateur

La déposition par évaporation se fait, dans une chambre à vide, par sublimation du métal (sous un très fort courant dans un creuset de tungstène ou grâce à un flux d'électron). La figure 66 montre un évaporateur par effet joule.



Figure 66 : Evaporateur par effet Joule RIBER

3.9. Etape 7 : Via du métal 1

3.9.1. Introduction

La septième étape consiste en la réalisation des via entre le métal 1 et le métal 2. Les figures suivantes montrent les 6 tâches nécessaires à la définition des grilles

- Dépôt de la silice
- Planarization de la silice
- Dépôt de la résine photorésistante
- Développement de la résine
- Gravure de la silice
- Elimination de la résine restante

3.9.2. Principe

Le principe de réalisation de cette septième étape est décrit par les deux figures suivantes (elles donnent la représentation en coupe et sous forme de Layout du futur inverseur)

3.9.2.1. Dépôt silice et résine photorésistante

On dépose sur l'ensemble du wafer une couche d'oxyde de silicium par dépôt plasma. Toujours pour maîtriser les couplages capacitifs cette silice est planérisée

On recouvre l'ensemble du wafer par une couche de résine photorésistante

Cette résine est insolée à travers le masque via métal 1.

Après dissolution chimique la résine restante protège la silice entre les via du métal 1.

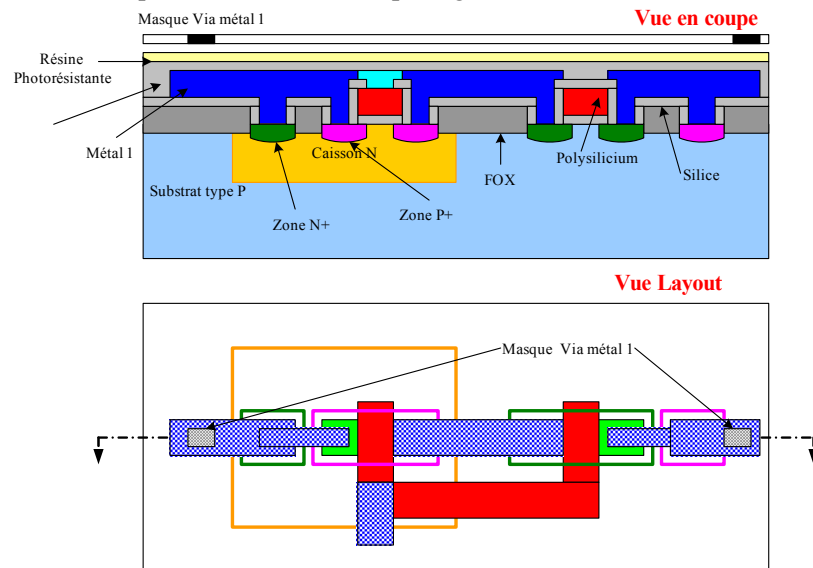


Figure 67 : Etape 7 (dépôt silice et résine)

3.9.2.2. Gravure silice

La silice non protégée par la résine est gravé pour méthode physique de type plasma.

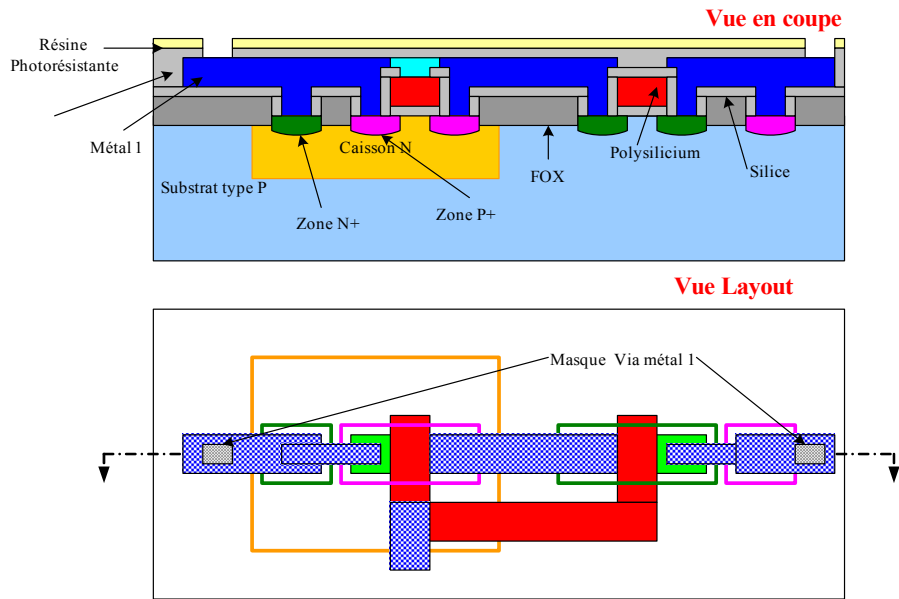


Figure 68 : Etape 7 (gravure silice)

On élimine par attaque chimique la résine photorésistante restante

3.10. Etape 8 : Couche de métal 2

3.10.1. Introduction

La huitième étape consiste en la réalisation des connexions en métal 2.

Les figures suivantes montrent les 5 tâches nécessaires à la définition des grilles

- Dépôt du métal 2
- Dépôt de la résine photorésistante
- Développement de la résine
- Gravure du métal 2
- Elimination de la résine restante

3.10.2. Principe

Le principe de réalisation de cette huitième étape est décrit par les deux figures suivantes (elles donnent la représentation en coupe et sous forme de Layout du futur inverseur)

3.10.2.1. Dépôt métal 2 et résine photorésistante

On dépose par pulvérisation cathodique le métal 2 (aluminium ou cuivre).

On recouvre ensuite l'ensemble du wafer par de la résine photorésistante.

La résine est insolée à travers le masque métal 2.

On découvre ainsi le métal 2 hors des futures connexions

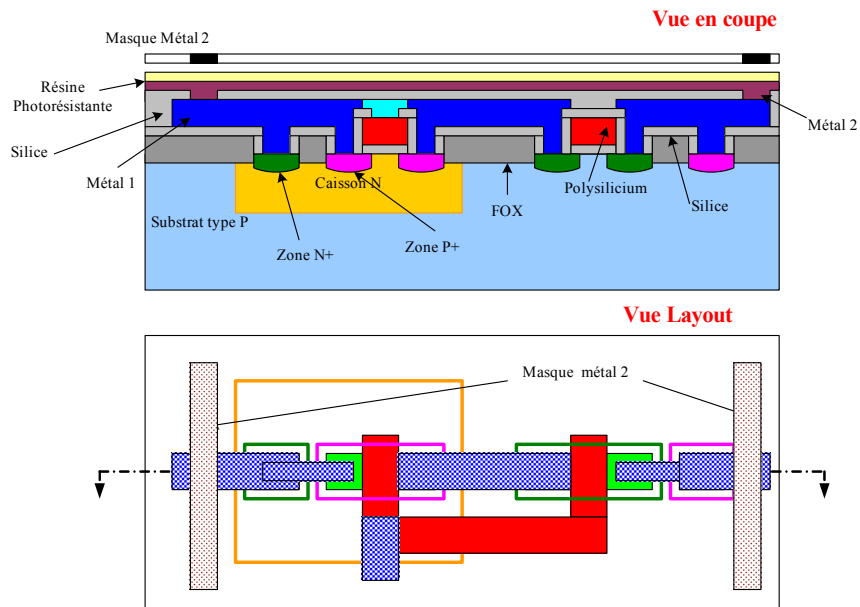


Figure 69 : Etape 8 (dépôt métal 2 et résine)

3.10.2.2. Gravure du métal 2

Le métal 2 est gravé par attaque physique de type plasma.

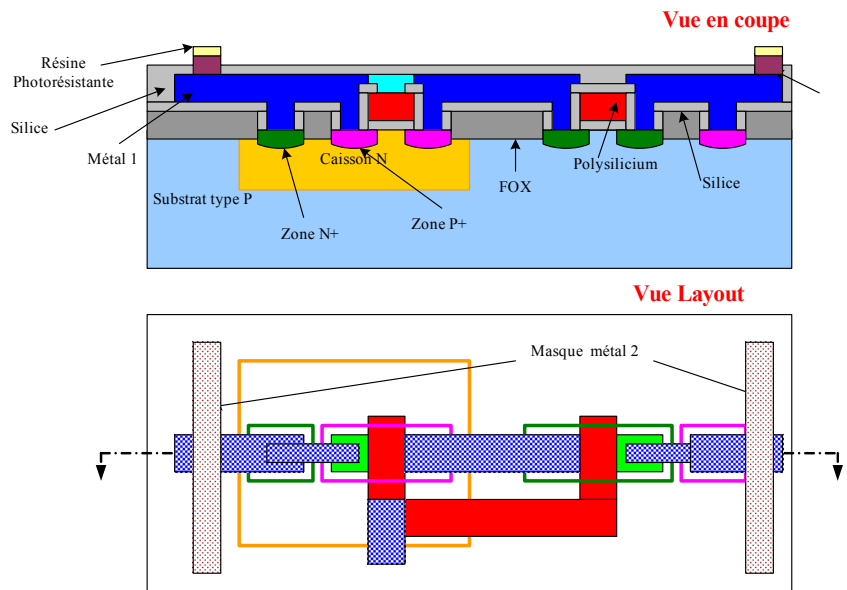


Figure 70 : Etape 8 (gravure métal 2)

On élimine par attaque chimique la résine photorésistante restante