Conception de PLA CMOS

Abbas Dandache

THESE

présentée par

DANDACHE Abbas

pour obtenir le titre de **DOCTEUR**

de l'**INSTITUT NATIONAL POLYTECHNIQUE DE GRENOBLE**

(Spécialité : **Microélectronique**)

CONCEPTION DE PLA CMOS

Thèse soutenue le 9 juillet 1986 devant la commission d'examen.

P. GENTIL	**Président**
J. LARDY	
G. SAGNES	**Examinateurs**
G. SAUCIER	
J. TRILHE	

Je tiens à remercier,

- Monsieur P. GENTIL, professeur à l' ENSERG et Directeur du Centre Interuniversitaire de Microélectronique, qui a bien voulu me faire l'honneur de présider le jury de cette thèse ;

- Madame G. SAUCIER, professeur à l' ENSIMAG, m'avoir accueilli dans le Laboratoire Circuits et Systèmes et ainsi que pour ses conseils tout au long de cette étude ;

- Les rapporteurs de cette thèse : Monsieur J. LARDY, responsable de l'Equipe de Conception au CNET, et Monsieur G. SAGNES, professeur à l'Université des Sciences et Techniques de Languedoc (Montpellier), qui ont bien voulu juger le manuscrit de cette thèse et y consacrer une part de leur temps ;

- Monsieur J. TRILHE, Docteur Ingénieur THOMSON Semiconducteurs, pour avoir accepter de participer au jury de cette thèse et pour les discussions constructives que j'ai eu avec lui ;

- tous mes collègues, membres du Laboratoire Circuits et Systèmes, pour les nombreuses discussions et leur sympathie ;

- le Service de Reprographie de l'IMAG dont la compétence et l'efficacité ne sont plus à présenter.

A mes parents

إلى أهلي

إلى بدر

RESUME

Cette thèse concerne plus particulièrement les **PLA CMOS**. Les quatre aspects suivants sont développés :

- performance électrique : spécification d'outil d'évaluation électrique et temporelle de PLA par une technique hybride estimation-simulation, basée sur la recherche du chemin critique d'E/S dans le PLA ;

- distribution des types de pannes en fin de fabrication et leurs manifestations électriques et logiques. Une approche vers le test de PLA CMOS est également présentée ;

- amélioration du rendement de fabrication par la conception de PLA reconfigurable (ajout de lignes supplémentaires) ;

- partitionnement de PLA, en vue de réduire la surface, le temps de réponse, et de faciliter la reconfiguration et l'interconnexion avec les blocs voisins.

MOTS CLES: VLSI - PLA - performance - test - reconfiguration - partitionnement - CAO.

ABSTRACT

This thesis deals with CMOS PLAs. The four following points are developped :

- electrical performance : specification of an electrical and temporal evaluation tool for PLAs with a hybrid estimation-simulation technique based on the search of critical I/O paths in the PLA ;

- distribution of the end of manufacturing fault types, and their electrical and logical manifestation. An appraoch for the test of CMOS PLAs is also presented ;

- improvement of the manufacturing yield by designing reconfigurable PLAs. This is obtained by adding spare unprogrammed lines that can be used to replace faulty ones ;

- PLA partitioning to reduce area, delay time, and to ease the reconfiguration and the connection to neighbouring blocks.

KEYWORDS: VLSI - PLA - performance - test - reconfiguration - partitioning - CAD.

TABLE DES MATIERES

CHAPITRE II
DEFAUTS DE FABRICATION ET TEST DES PLA CMOS

CHAPITRE III

PLA RECONFIGURABLES

CHAPITRE IV
PARTITIONS DE PLA

INTRODUCTION

Notre étude, qui se situe dans le thème **"conception des circuits intégrés"**, porte sur des problèmes concernant la conception de **PLA** en technologie **CMOS**.

Les problèmes de conception des PLA utilisés dans les circuits intégrés complexes sont liés à des contraintes d'ordre électrique, topologique et économique.

En vue de tenir compte de ces contraintes, il est nécessaire d'étudier des méthodes et/ou des outils qui facilitent la tâche du concepteur et qui augmentent le rendement de production. Pour celà, deux voies sont utilisées:

1- Le développement d'outils de **CAO**.

2- La définition de structures régulières de PLA qui pourront alors être traitée efficacement avec de tels outils.

Un **PLA** (Réseau logique programmable) est un circuit combinatoire (ou séquentiel) simple qui permet la réalisation régulière de fonctions booléennes. Il est caractérisé par le nombre de ses entrées, le nombre de

ses monômes, le nombre de ses sorties et par deux matrices : une matrice qui réalise de fonctions **ET** et une matrice qui réalise de fonctions **OU** (figure 1).

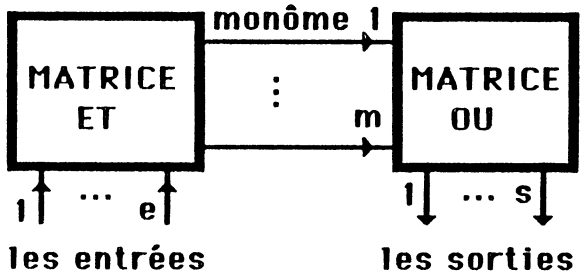


Figure 1- Structure globale d'un PLA.

L'utilisation de PLA, pour l'intégration à grande échelle, réduit les coûts de conception, et permet de bénificier, par rapport à la structure aléatoire, des avantages suivants.

- facilité de conception ;
- facilité d'implantation ;
- facilité de mise en oeuvre des dispositifs de test ;
- implémentation facile d'algorithmes d'interprétation compliqués ;
- facilité de connexion avec les blocs voisins ;
- facilité de simulation électrique ;
- facilité de reconfiguration.

La surface d'implantation d'une structure à base de PLA est supérieure à celle d'une structure en logique aléatoire. Ce problème peut être résolu grâce à des outils d'optimisation de PLA. Une telle

optimisation peut être réalisée à deux niveaux : au niveau logique et au niveau topologique.

Cette thèse concernant la conception des PLA en technologie CMOS, se compose de quatre chapitres présentés selon le plan suivant :

- Dans le premier chapitre, les différentes structures régulières des PLA sont étudiées afin de développer une méthode d'évaluations électriques et temporelles (simulateur).

- Dans le deuxième chapitre, nous allons introduire le phénomène des mécanismes de défaillance affectant les éléments physiques dans le système VLSI, afin de les faire remonter aux niveaux électrique, puis logique.

Nous présenterons, d'une part, une distribution des types de pannes observés afin de faciliter l'étape de reconfiguration des PLA, et d'autre part une méthode de test pour les PLA CMOS.

- Au niveau du chapitre III, en vue d'améliorer le rendement de fabrication de circuits intégrés complexes en VLSI, nous traitons la reconfiguration des PLA par l'ajout de lignes supplémentaires (en tenant compte de la distribution des types de pannes). La méthode utilisée repose sur des techniques de programmation d'interrupteurs

permettant d'échanger une zone partiellement en panne avec une zone saine.

- Enfin, le chapitre IV propose une méthode de partition de PLA afin de réduire la surface totale de Silicium occupée, le temps de propagation et afin de faciliter la reconfiguration et le problème d'interconnexion avec les blocs voisins. Son principe est fondé sur la technique de partition d'un PLA en plusieurs PLA indépendants et fonctionnant en parallèle.

CHAPITRE I

EVALUATIONS ELECTRIQUE ET TEMPORELLE DES PLA CMOS

Introduction

Dans le domaine de conception des circuits intégrés, ce chapitre traite la conception électrique de PLA CMOS, pour la performance en vitesse et en consommation. La conception d'un PLA peut être divisée selon les trois processus suivants, figure I.1:

1- La conception fonctionnelle.

2- La conception topologique.

3- La conception physique.

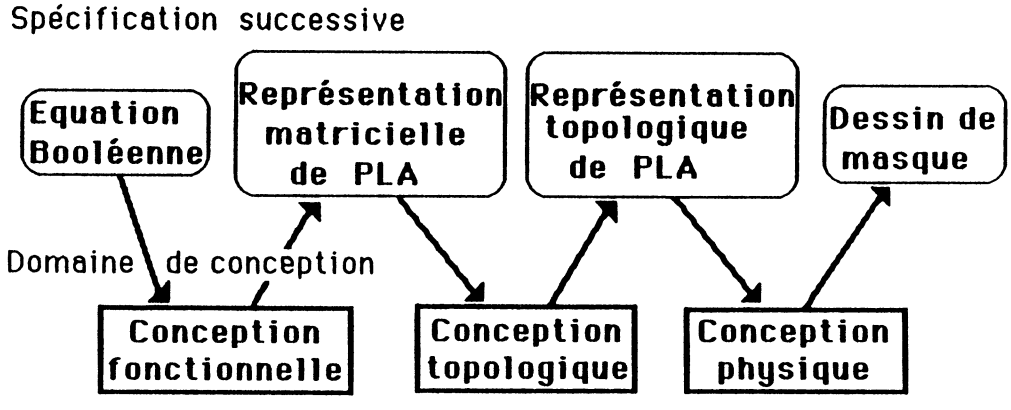


Figure I.1- Processus de conception de PLA.

1- La conception fonctionnelle consiste à transposer des équations Booléenne en deux niveaux logiques et de les mettre sous la forme d'une somme de produits (ou produit de sommes) des variables (PLA) [KAN81]. L'étude de la minimisation logique pour réduire le nombre de monômes et de transistors est effectuée pendant cette étape.
La représentation matricielle de PLA résulte de cette étape. En effet, deux matrices ET-OU correspondent respectivement aux deux niveaux logiques de PLA.

2- La conception topologique consiste à définir la représentation topologique du PLA à partir de sa représentation matricielle. L'étude de l'optimisation topologique peut être effectuée pendant cette étape telle que le pliage du PLA (folding) ou son tassement [HAC82, PAI81].

3- La conception physique comporte la réalisation de dessins de masque pour une technologie donnée. L'étude de performance électrique (vitesse, consommation, fiabilité, ...) peut être effectuée pendant cette étape.
Différentes études de performances électriques sont présentées dans différentes technologies telles que le Bipolaire, NMOS et le CMOS [CAN86, HED85, WES85, DAN83, MAV83].

Notre étude porte sur ce problème de performances électriques de PLA en technologie CMOS.

Les performances, au niveau électrique, nécessitent des outils

particuliers tels que des simulateurs électriques de circuits. Les simulateurs électriques (SPICE, MSINC, ...) constituent ces outils de base. La complexité des algorithmes de simulation est telle qu'elle ne peut s'appliquer que difficilement à des circuits intégrés dotés d'un grand nombre de transistors. Ceci a justifié l'ouverture de recherches vers une nouvelle génération de simulateurs et d'autres outils de conception électrique.

Dans ce chapitre, une méthode d'évaluations électrique et temporelle de PLA CMOS va être developpée. Ce chapitre comporte deux parties:

I- La première partie étudie les différentes structures électriques les plus classiques des PLA CMOS: dynamiques et semi-statiques.

II- La deuxième partie: présente des méthodes d'évaluation des performances électriques et temporelles des PLA CMOS. Ceci consiste à calculer des paramètres électriques et topologiques et à estimer le temps de propagation dans les PLA pour une technologie donnée et pour une programmation spécifique, figure I.2.

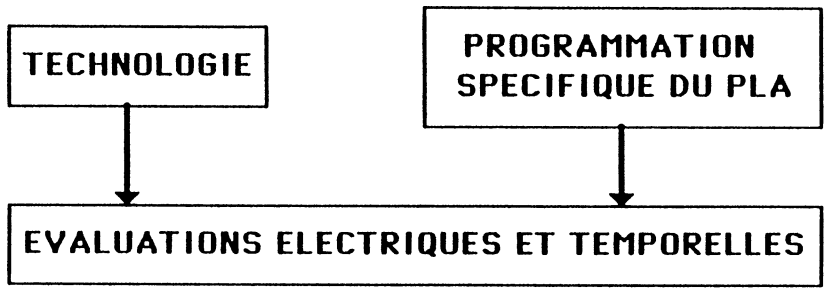


Figure I.2- Evaluations électriques et temporelles du PLA.

La variation des valeurs de ces paramètres donne les diverses possibilités de réalisation d'un PLA. Un système interactif est réalisé à partir de ces parmètres pour permettre à l'utilisateur ou au concepteur de choisir l'organisation d'un PLA en fonction de sa consommation, de sa surface et de son temps de réponse.

I- Réalisation de PLA en CMOS

I.1- Choix de la technologie CMOS

La technologie d'implantation dominante actuellement dans les systèmes **VLSI** est la technologie **NMOS** (transistor canal N) dans la réalisation des microprocesseurs tels que (6800, 68000, INS8080, Z8000, ...). Cette technologie a remplacé la technologie **PMOS** (transistor canal P), à partir des années 67, pour différents avantages dont la rapidité, due à la mobilité ($\mu n > \mu p$), par rapport à celle du PMOS.

L'apparition de la technologie **CMOS** (transistors MOS complémentaires, canal N et canal P) présente des avantages dans son développement sur différents points par rapport à la technologie NMOS et principalement dans la structure des circuits dynamiques et des circuits à précharges pour augmenter la densité de portes par unité de surface:

- La technologie CMOS [HAR84] représente une faible consommation statique.

- Une haute immunité au bruit.

- le bloquage alterné des réseaux de transistors P/N en régime statique implique Une haute impédance entre l'alimentation et la masse.

- Elle est aussi rapide que la technologie NMOS.

- Elle est aussi dense que la structure NMOS dans les circuits dynamiques et à précharge.

Les avantages de la technologie CMOS vont conduire à réaliser des microprocesseurs dans les systèmes **VLSI**.

Pour cette raison, nous allons utiliser la technologie CMOS, qui devient une technologie dominante, pour implémenter des structures régulières à précharge (PLA,ROM).

I.2- Structures CMOS

Les circuits CMOS sont utilisés en structures dynamiques ou à précharge et en particulier dans les structures régulières. La structure à précharge a pour but de réduire la surface occupée et d'augmenter la vitesse de fonctionnement.

La structure régulière qui va être traitée, pour les PLA et les ROM, est à base de portes **NOR** ou **NAND**. Ces portes peuvent être en cascade à deux niveaux logiques:

- Le premier niveau de portes peut être soit un décodeur d'entrées pour la ROM, soit une matrice ET pour le PLA.
- Le deuxième niveau de portes peut être soit la ROM, soit la matrice OU pour le PLA.

Dans la suite nous pouvons assimiler la structure régulière à la réalisation de PLAs, la ROM étant un PLA particulier, et plus particulièrement nous nous intéresserons à la structure de PLA à base de portes NOR (transistors N en parallèles), figure I.3, car sa vitesse de fonctionnement est plus grande que celle du type NAND [CHO81] (transistors en séries), et La génération des dessins de masques pour de PLA en porte NOR (transistors N en parallèles) est plus simple que celle de porte NAND. On peut ajouter aussi que le dimensionnement de transistors d'une porte NAND est plus difficile que celui d'une porte NOR.

Nous allons introduire les différentes structures de fonctionnement de PLA.

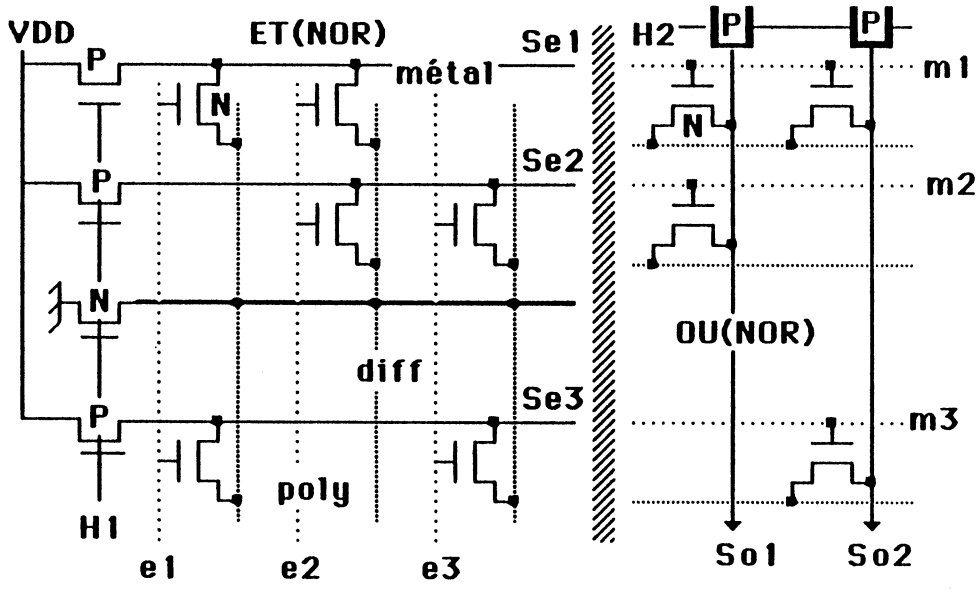


Figure I.3- PLA CMOS NOR-NOR.

Deux différentes structures de fonctionnement suivantes vont être présentées:

1- Structure dynamique.

2- Structure semi-statique.

Chaque structure comporte deux niveaux de portes, en vue de réaliser des PLA (une somme de produits des variables ou un produit de sommes) tels que les portes qui réalisent les fonctions logiques sont les suivantes:

NOR-NOR, NAND-NAND, NOR-NOT-NAND, NAND-NOT-NOR.

Nous reviendrons sur ces structures par la suite.

I.2.1- Structure dynamique

En vue de réduire la surface occupée et d'augmenter la vitesse de fonctionnement, la structure dynamique est souvent utilisée en technologie CMOS. Des méthodes descriptives pour concevoir des circuits dynamiques ont été décrites [PEN72].

Il existe un mode de fonctionnement commun pour tous les circuits dynamiques. Celui-ci comprend une phase de précharge de la sortie à un niveau logique. Il comprend également une phase d'évaluation qui consiste à décharger la sortie d'une porte active ou à la maintenir à son niveau logique (celui de la précharge).

Pendant cette phase d'évaluation, un problème essentiel se pose dans la structure dynamique. Ce problème est le temps de conservation de l'information dans les capacités parasites. Ce temps impose la fréquence minimale de fonctionnement.

Pendant l'activation d'une porte logique, des niveaux logiques stables à ses entrées sont nécessaires. Ceci pose un sérieux problème d'horlogerie dans la réalisation des circuits dynamiques.

Différentes manières de concevoir ce problème d'horlogerie sont proposées par [PEN72, HEB78, KRA82]. Nous allons présenter deux structures de logiques dynamique de type "**DOMINO**" [KRA82] et "**NP-CMOS**" [GON82].

Pour chaque type, nous allons étudier la structure régulière qui réalise des PLA à deux niveaux de porte.

a- Structure DOMINO

Dans la structure DOMINO, figure I.4, une seule horloge peut syncrhoniser le système. Une opération logique se déroule sur deux phases:

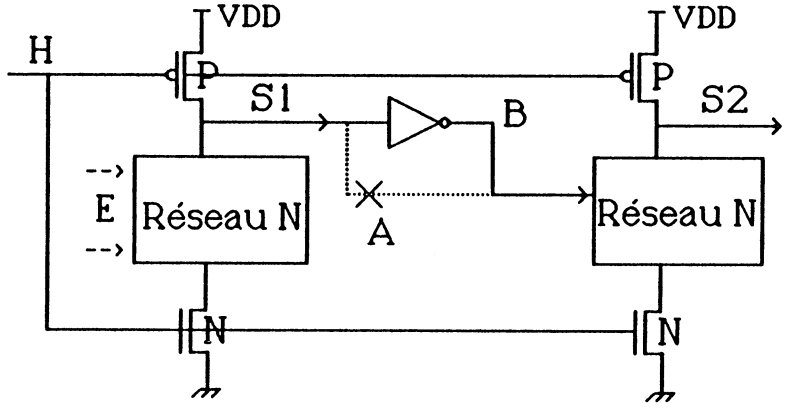


Figure I.4- Structure DOMINO

La première phase est définie pour H="0", les sorties (S1,S2) se préchargent à "1".

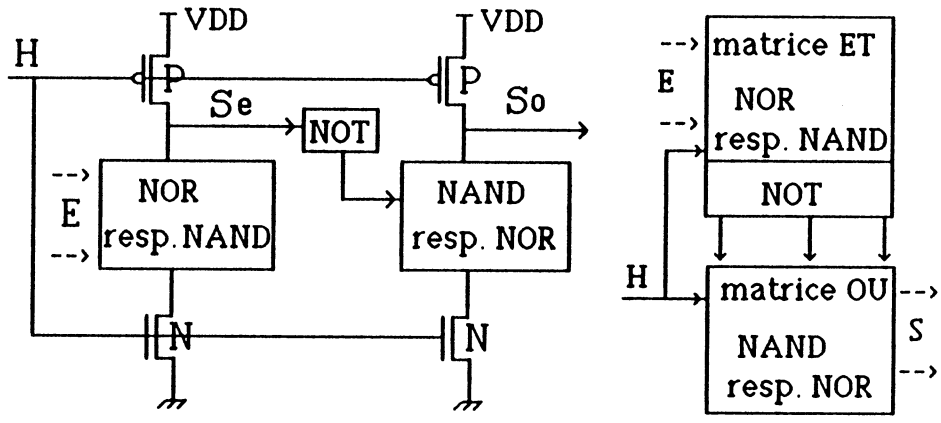
La deuxième phase est définie pour H="1". C'est la phase d'évaluation où les sorties dépendent de l'état des entrées. Une sortie peut avoir deux cas possibles: soit elle se décharge vers "0", soit elle reste à "1".

Un assemblage de blocs identiques ne peut fonctionner seul dynamiquement. En effet, pendant la phase de précharge (H="0"), les sorties (S1,S2) sont à "1".
Au début de la phase d'évaluation (H="1"), les transistors de réseau N du deuxième bloc restent conducteurs pendant un certain temps. Ceci peut provoquer la décharge du deuxième bloc (S2) avant l'évaluation du premier bloc (S1) (voie A). La solution qui est adoptée dans la structure DOMINO consiste à mettre un inverseur entre deux blocs (voie B) pour empêcher la décharge prématurée du deuxième bloc.

En vue de réaliser la structure régulière en produit de sommes ou somme de produits des variables (PLA) à base des portes logiques NOR ou NAND, quatre structures logiques sont possibles:

NOR-NOT-NAND, NAND-NOT-NOR et NOR-NOR, NAND-NAND.

Les deux premières structures logiques (NOR-NOT-NAND, NAND-NOT-NOR) sont compatibles avec la structure DOMINO, les deux blocs étant séparés par un inverseur, figure I.5.



-a- -b-

Figure I.5- PLA compatible avec la structure DOMINO.

Les deux dernières structures logiques NOR-NOR, NAND-NAND ne sont pas directement compatibles avec la structure DOMINO. En effet, on ne peut pas mettre un inverseur entre deux blocs par la contrainte logique (somme de produits des variables). Ceci pose une autre contrainte qui est la possibilité de décharge prématurée du deuxième bloc. Une solution à cette contrainte consiste à utiliser deux horloges (H1,H2), figure I.6.

Pendant la première phase (Ø1) les deux matrices sont en état de précharge (H1=H2=0).

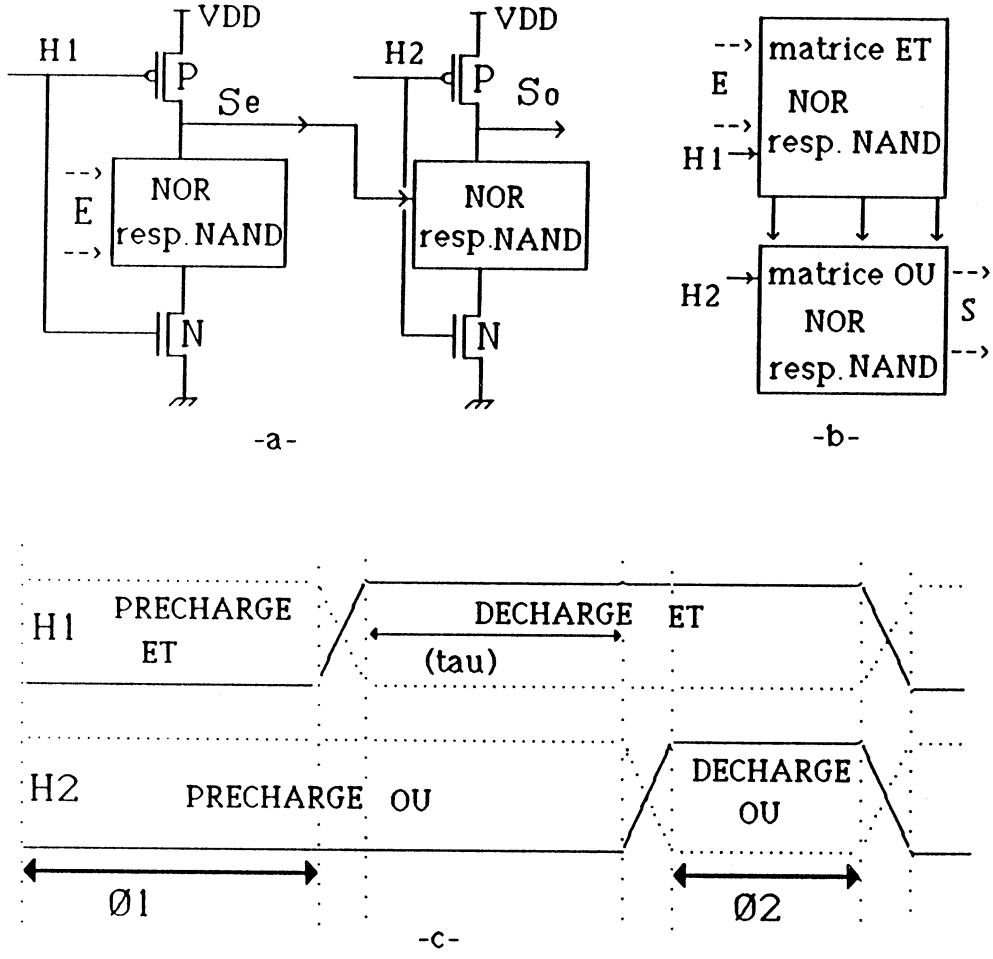


Figure I.6- PLA à deux horloges en deux phases.

Pendant la phase d'évaluation (Ø2), les entrées de chaque bloc doivent d'être à des niveaux logiques stables. Ceci nécessite l'évaluation du premier bloc avant l'activation du deuxième. Autrement l'horloge H2 doit être en retard par rapport à H1 un certain temps (tau) (H1=1,H2=0), jusqu'à ce que la sortie Se (sortie de la matrice ET) du premier bloc se stabilise. Ce temps correspond au temps de décharge du premier bloc. Ceci empêche la décharge prématurée du deuxième bloc. Différentes structures électriques ont été données par [DAN83] en respectant le principe de garder le

deuxième bloc en état de précharge pendant la décharge du premier.

La structure logique NOR-NOR (transistors N en parallèles) est la structure dominante dans la réalisation des PLAs; en effet sa vitesse de fonctionnement est plus grande que celle du type NAND [CHO81]. Dans cet objectif, nous allons présenter différentes structures électriques qui réalisent la logique NOR-NOR.

- La première structure électrique, figure I.6, comporte deux blocs DOMINO. Les réseaux N réalisent des portes NOR entre un transistor P qui tire au niveau d'alimentation (VDD) et un transistor N qui tire à la masse, qu'on appelle transistor de couplage à la masse.

Les inconvénients de cette structure sont:

a- l'interconnexion nécessaire entre les sources des transistors du réseau N et le drain du transistor N du couplage à la masse. Ceci présente une charge résistive et capacitive importante pendant la phase de décharge pour une grande ligne. Ce qui est souvent le cas pour de grands PLA. Cet effet augmente le temps de propagation dans le PLA. La réduction de ce temps nécessite l'ajout de plusieurs transistors de couplage à la masse.

b- Les lignes d'entrées des matrices telle que la matrice OU. Ceci présente une charge d'entrée résistive et capacitive pendant la phase

de décharge. Cet effet augmente le temps de descente de la matrice OU et il devient important pour de grands PLAs où le nombre des sorties est important.

Une solution à cet effet peut être prise: soit par l'utilisation d'un montage sur les entrées de la matrice OU à base de boot-strap [DAN83] pour réduire le temps de montée des lignes d'entrées, soit par l'utilisation d'une deuxième couche métal (en parallèle avec la ligne Poly) qui assure la connection des lignes d'entrée afin de réduire la charge d'entrée pour les grands PLA.

Une comparaison par simulation (SPICE) de deux méthodes montre l'efficacité du domaine d'utilisation de chacune en fonction de la charge d'entrée (ou nombre de sorties), figure I.7.

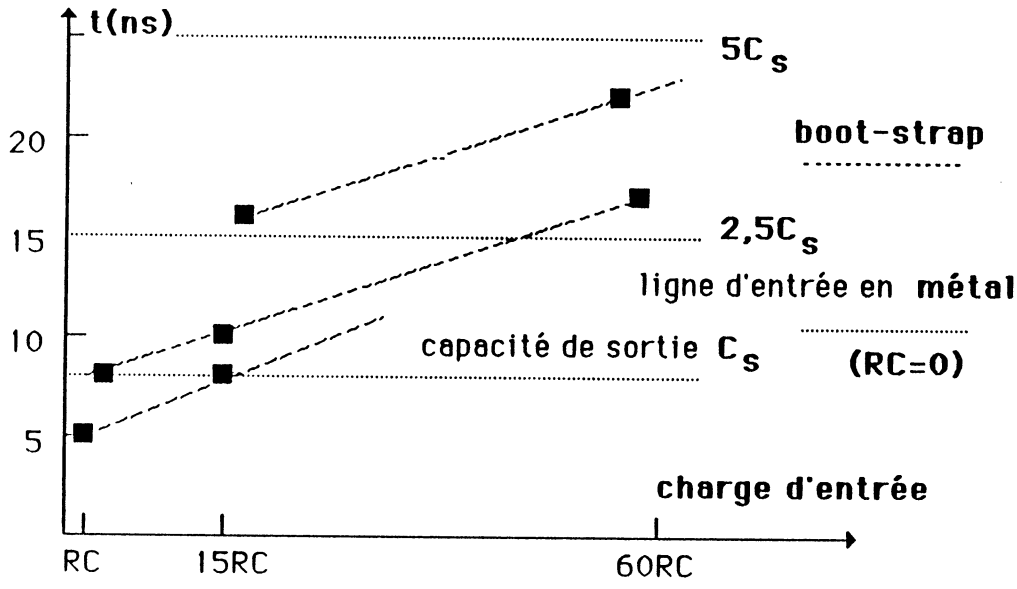


Figure I.7- temps de descente de la matrice OU en structure Boot-Strap ou en ligne d'entrée métal.

- La deuxième structure électrique propose de remplacer les transistors de couplage à la masse par des portes NOR aux entrées du réseau [COH84] afin de bloquer ces transistors pendant la phase de précharge, figure I.8.

L'inconvénient de cette structure est la surface importante occupée par la circuiterie sur les entrées et le retard sur ces lignes.

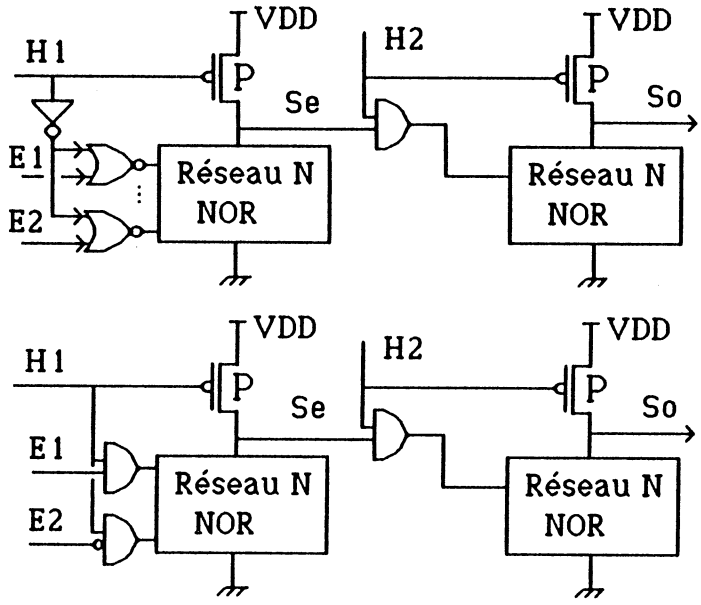


Figure I.8- PLA CMOS NOR-NOR.

- La troisième structure mixte les deux précedentes, figure I.9. Le premier bloc garde le transistor de couplage à la masse pour réduire sa surface ainsi que le retard sur les lignes d'entrées.
En effet, dans ce cas les niveaux d'entrées sont toujours positionnées.

Tandis que sur les entrées du second bloc, un montage à base de boot-strap [DAN83] peut être envisagé pour réduire le temps de montée des lignes d'entrées, ceci réduit le temps de décharge du second bloc.

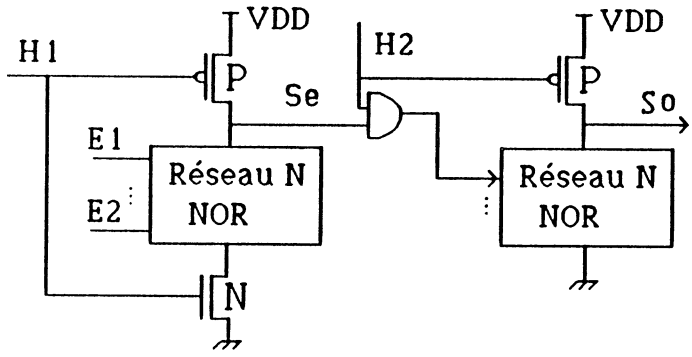


Figure I.9- Structure mixte de PLA CMOS NOR-NOR.

Cette structure est efficace dans le cas où la capacité se sortie de la matrice OU est importante par rapport à celle d'entrée (c'est-à-dire, le nombre des monômes est important par rapport au nombre des sorties nm/ns >1).

En effet la justification de cette efficacité est le fait que le nombre des monômes représente une charge capacitive importante à la sortie. Pendant la phase de décharge, cette capacité se décharge au travers d'un seul transistor en lieu et place de deux transistors en séries.

Une comparaison du temps de descente (par simulation) de la matrice OU pour les trois cas différents (ligne d'entrée en poly, en métal et boot-strap en structure mixte) montre le domaine d'utilisation de chaque structure, figure I.10:

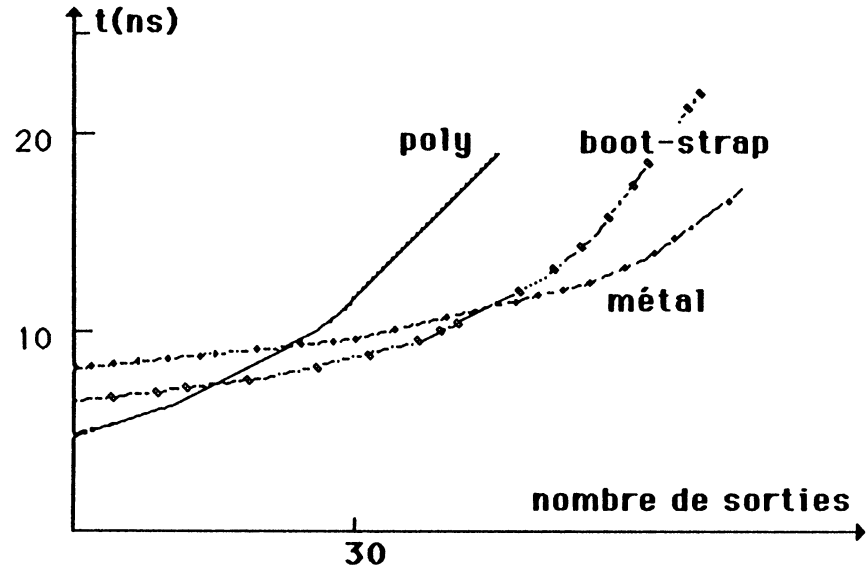


Figure I.10- temps de descente de la matrice OU en fonction du nombre des sorties (le nombre des monôme est constant) pour les trois structures (ligne d'entrée poly, métal ou avec boot-strap).

L'efficacité d'utilisation de chaque structure dépend de la taille du PLA:

- Le PLA avec des lignes d'entrée poly est efficace pour de petits PLA.
- Le PLA avec le boot-strap entre les deux matrices est efficace pour des PLA moyens et en particulier pour des PLA ou nm/ns > 1 (nm, ns: nombre des monômes et des sorties de PLA).
- Autrement pour un PLA large, l'utilisation d'une deuxième couche de métal peut devenir nécessaire.

b- Structure NP-CMOS (Nora)

La structure NP-CMOS, figure I.11, utilise des blocs en cascade qui sont constitués alternativement des transistors nMOS (Tn) et des transistors pMOS (Tp). Ces blocs sont connectés directement.

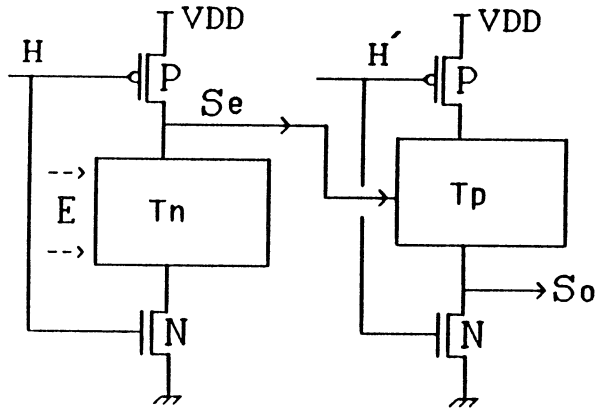


Figure I.11- Structure NP-CMOS.

Une seule horloge H avec son complément H' peut syncroniser le système. Une opération logique se déroule sur deux phases:

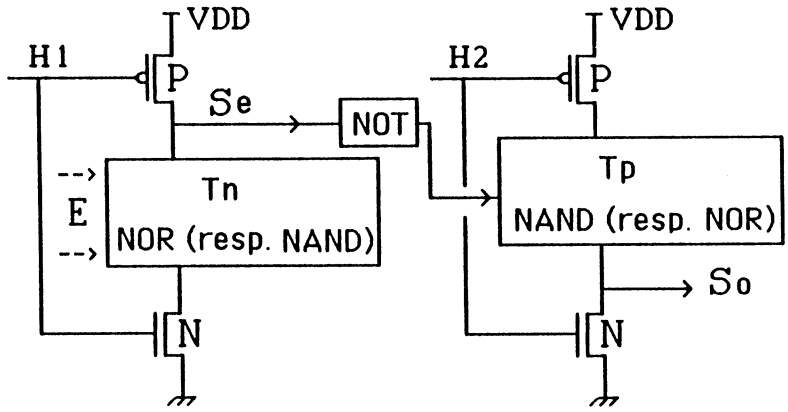
- Pendant la première phase (H=0, H'=1), les deux blocs sont en état de précharge (Se="1", So="0").
- Pendant la phase d'évaluation (H=1, H'=0), les deux sorties (Se,So) des deux blocs s'évaluent successivement par rapport aux entrées.

En vue de réaliser la structure régulière des PLAs en NP-CMOS, les quatre structures logiques vont être réalisées:

NOR-NOR, NAND-NAND et NOR-NOT-NAND, NAND-NOT-NOR.

Les deux premières structures logiques (NOR-NOR, NAND-NAND) sont compatibles avec la structure NP-CMOS. Car les interconnexions entre les deux blocs sont directes.

Le problème se pose pour les deux dernières structures logiques (NOR-NOT-NAND, NAND-NOT-NOR). Elles ne sont pas directement réalisables en NP-CMOS du fait que les sorties entre deux blocs sont inversées. Ceci provoque la décharge prématurée du deuxième bloc, figure I.12.



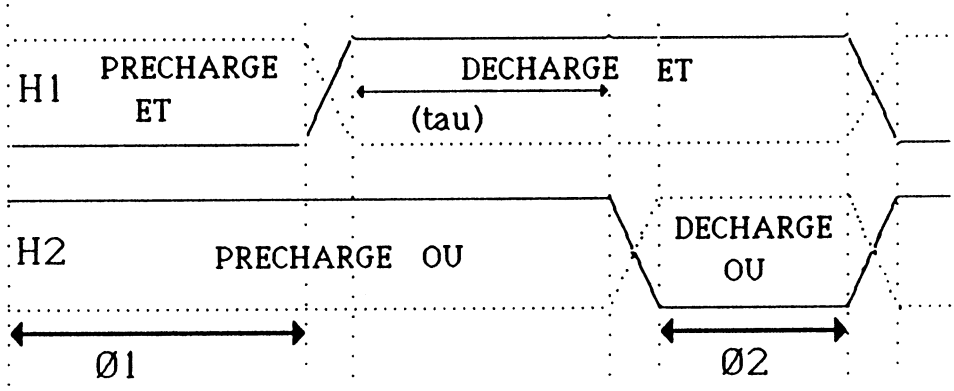


Figure I.12- PLA en NP-CMOS.

Comme dans le cas de la structure DOMINO, la solution consiste à utiliser deux horloges (H1, H2) pour assurer l'évaluation successive des deux blocs.

1.2.2- Structure semi-statique

La structure semi-statique est réalisée à partir de la structure dynamique en ajoutant d'autres composants électriques pour rendre l'opération totalement statique. Ceci consiste à ajouter un maintien de niveau (ou maintien de la précharge) sur chaque sortie qu'on appelle "accrocheur".

L'accrocheur est capable de maintenir une ligne isolée au niveau "1". Donc il maintient une sortie au niveau haut pendant qu'elle doit être au niveau haut.

Une structure d'accrocheur qui est donnée par [PRI82], comporte un inverseur et un transistor pMOS, figure I.13a.

Une autre structure d'accrocheur peut être simplement un transistor déplété nMOS, figure I.13b. Dans cette structure, le temps de montée peut augmenter si le transistor nMOS est mal dimensionné (ce transistor est dimensioné de façon à compenser le courant de fuite).

Dans les deux cas, la résistance de chaque transistor (p ou n) doit

être importante afin de permettre seulement de passer un courant qui compense le courant dû au courant de fuite et de la distribution de charge dans le cas le plus défavorable.

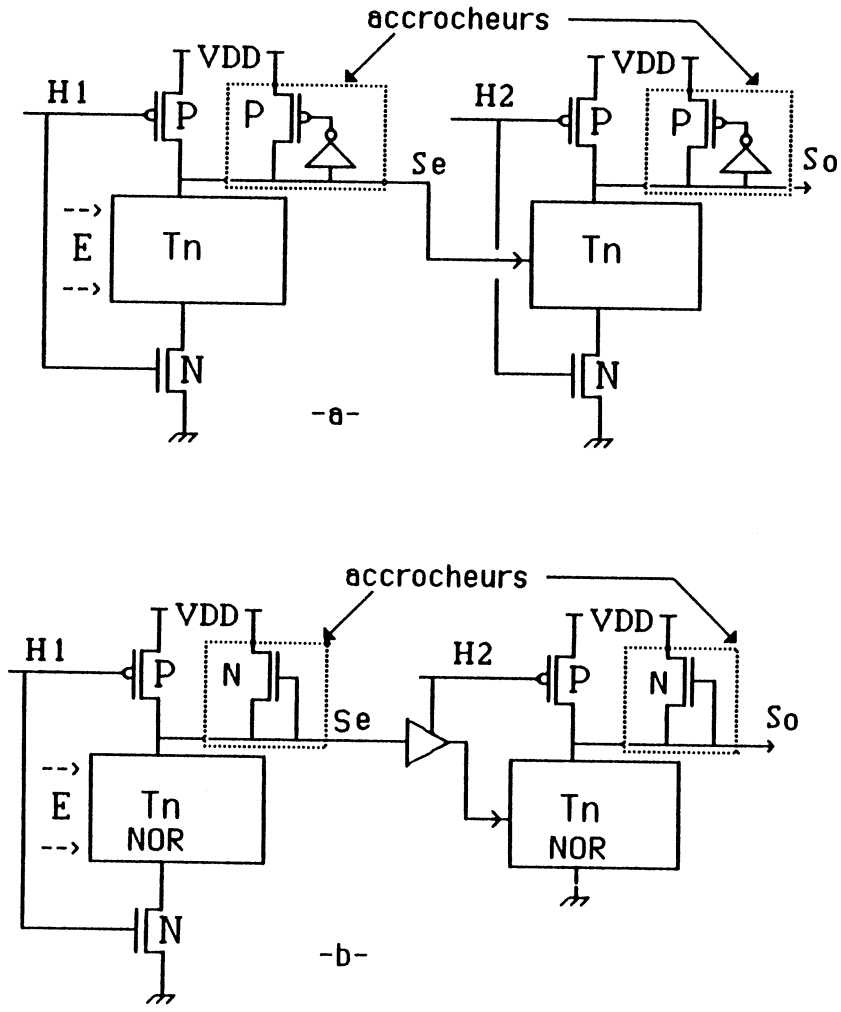


Figure I.13- PLA semi-statique.

II- Evaluations électrique et temporelle des PLA CMOS

Cette partie des évaluations électrique et temporelle de la structure régulière consiste à définir un évaluateur électrique qui permet d'estimer le temps de propagation dans les PLA CMOS.

Cette méthode d'évaluation des performances électriques et temporelles consiste à normaliser la structure d'un PLA et la ramener à l'étude d'une porte CMOS.

Ensuite nous présenterons la méthode de la normalisation d'une structure régulière et ses performances pour estimer le temps de propagation.

La structure qui nous intéresse à étudier est celle qui concerne les PLA du type des portes NOR-NOR à réseau N, pour leurs rapidités par rapport aux autres structures logiques citées auparavant, et pour la facilité de générer leurs dessins de masques.

Tout d'abord nous allons présenter des rappels théoriques qui introduisent des formules de base de l'évaluateur électrique concernant le calcul des capacités d'un inverseur pendant son temps de montée et de descente, puis le calcul de retard d'un inverseur chargé soit par une capacité, soit par une capacité et une résistance.

2.1- Formules de base

2.1.1- Calcul des capacités

Les capacités à calculer se situent à la sortie d'un inverseur CMOS pendant son temps de commutation. Pour rendre constante chaque capacité qui varie en fonction de la tension, figure I.14. On lui donne une valeur maximale ou une valeur moyenne.

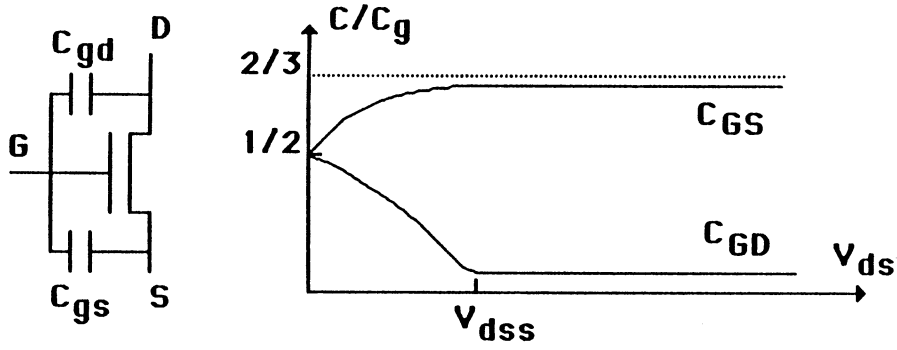


Figure I.14- La variation des capacités de transistor MOS en fonction de la tension drain source (Vds).

On suppose que l'inverseur considéré attaque un autre semblable, figure I.15.

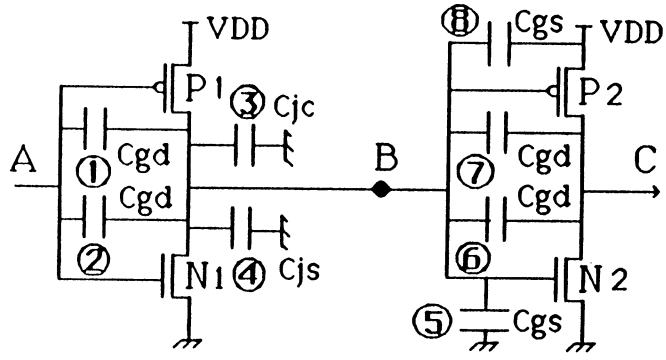


Figure I.15- Les capacités de deux portes CMOS.

1- Temps de descente:

La tension aux bornes de la capacité ramenée au point B évolue du niveau haut au niveau bas pendant que l'entrée au point A est attaquée par un échelon de 0 à 1.

Premier inverseur:

Le transistor TP1, figure I.18, est bloqué, sa capacité grille drain (Cgd) (1) est nulle. Celle de sa jonction drain-substrat (3) est Cjc.

Le transistor Tn1 conduit, sa capacité grille-drain (Cgd) (2) agit en effet Miller; elle sera multipliée par deux (variations opposées et égales à Vg et Vd) et ramenée à la masse. Elle devient Cgd=2.Cgs/2.
Celle de sa jonction drain substrat (4) est Cjs.

Second inverseur:

Le transistor Tn2 conduit. La capacité grille-drain (5) est Cgs/2. La capacité grille-drain (6) agit en effet Miller; elle devient Cgd=2.Cgs/2.
Le transistor Tp2 est bloqué. Sa capacité grille-drain (7) est Cgs=0. Sa capacité grille-source (8) est maximale 2/3.Cgc (Cgc: capacité grille-source du transistor Tp).

Pendant le temps de descente (td), la capacité (Cs) de sortie du premier inverseur est la somme des capacités vues du point B.

$Cs(td) = Cjc+Cjs+Cgs+Cgs/2+Cgs+2/3.Cgc = Cjc+Cjs+5/3.Cgs+2/3.Cgc.$

2- Temps de montée:

La tension aux bornes de la capacité ramenée au point B évolue du niveau bas au niveau haut pendant que l'entrée au point A est attaquée par un échelon descendant de 1 à 0.

Premier inverseur:

Tp1 conduit. Sa capacité Cgd agit en effet Miller; elle devient Cgd=2.Cgc/2. Celle de sa jonction drain-substrat (3) est Cjc.
Tn1 est bloqué. Sa capacité Cgd (2) est nulle Cgd=0. Celle de sa jonction (4) est Cjs.

Second inverseur:

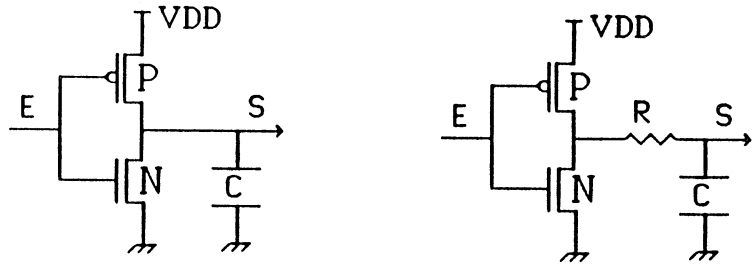
Les rôles des capacités (5, 6, 7, 8) sont inversés par rapport au cas du front de descente. Leur valeurs sont: 2/3.Cgs, 0, Cgc, Cgc/2.

Pendant le temps de montée (tm), la capacité (Cs) de sortie du premier inverseur est la somme des capacités suivantes (vues du point B):
Cs(tm) = Cgc+Cjc+Cjs+2/3.Cgs+Cgc+Cgc/2 = Cjc+Cjs+5/2.Cgc+2/3.Cgs.

2.1.2- Calcul du retard

Un inverseur CMOS, figure I.16, est caractérisé par ses transistors complémentaires et sa capacité de charge.

Nous allons calculer le temps de descente et le temps de montée d'un inverseur, qui est chargé soit par une capacité, figure I.16a, soit par une résistance et une capacité (R.C), figure I.16b, dans le cas où la capacité se charge et se décharge à travers la résistance.



-a- -b-

Figure I.16- Inverseurs CMOS.

La capacité pendant le temps de montée est (Cm(tm)). La capacité pendant le temps de descente est (Cd(td)).

Dans tous les calculs qui suivent, la mobilité (μ) dépend de la tension grille-source (Vgs):

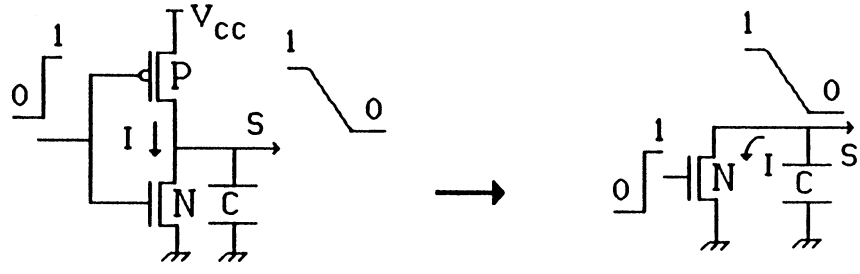
$$\mu\ (n, p) = \mu 0\ (n, p)/(1+\theta.(Vgs-Vt))$$

Vt(s, c): tension de seuil pour le transistor N et P respectivement.

a- **Porte chargée par C**:

Temps de descente:

Nous supposons qu'on attaque l'entrée de l'inverseur CMOS par un échelon d'amplitude Vcc, figure I.17a. Le transistor P est bloqué instantanément. Nous obtenons le schéma électrique équivalent pendant le temps de descente, figure I.17b.



-a- -b-

Figure I.17- Porte CMOS chargée par C pendant le temps de descente.

Le transistor N traverse deux modes de fonctionnement. Il fonctionne d'abord en saturation, puis en mode ohmique. Le temps de descente de l'inverseur est la somme des temps de descente dans ces deux modes.

- Zone saturée: En zone saturée, on a la relation suivante pour le transistor N:

$$Vds = Vs > Vdss$$

Vds: tension drain-source.

Vdss: tension de saturation.

Vs: tension à la sortie de l'inverseur.

Dans cette zone de saturation: $Vgs-Vts < Vds <= 0,9.Vcc$, où $Vgs=Vcc$. Le courant qui traverse le transistor N est I(sat):

$$I(sat) = \frac{\mu n \,.\, Cox \,.\, \gamma s}{2 \,.\, (\beta s+1)} \,.\, (Vcc-Vts)^2$$

Cox: capacité unitaire de l'oxyde mince.

γ (s,c) : rapport de dimension des transistor N et P.

β (s, c): coeffiçients d'effet du substrat pour les transistors N et P.

I(sat) décharge la capacité de la sortie (Cs=Cd(td)) de l'inverseur. On obtient l'égalité de courant:

$$I(sat) = -Cs. \, d(Vds)/dt.$$

Le temps de descente en zone saturée (td(sat)) s'obtient en intégrant cette égalité entre les deux niveaux de tension Vcc-Vts et 0,9.Vcc (Vts>0.5 volts):

$$td(sat) = \frac{Cd \,.\, 2.(\beta s+1)\ (Vts-0,1.Vcc)}{\mu n \,.\, Cox \,.\, \gamma s \,.\, (Vcc-Vts)^2}$$

- Zone ohmique: En zone ohmique la relation de Vs devient:

$0,1.Vcc <= (Vds=Vs) < Vcc-Vts$ (Vts<4 volts).

Le courant qui traverse le transistor N devient I(ohm):

$$I(ohm) = \mu n \,.\, Cox \,.\, \gamma s \,.\, (Vcc - Vts - \frac{\mu s + 1}{2} . Vds) \,.\, Vds$$

Alors l'égalité du courant: $I(ohm) = - Cs \,.\, d(Vds)/dt.$

L'expression I(ohm) contient des produits de Vds. Elle peut être développée en facteurs simples. Un relation logarithmique de td(ohm) s'obtient en intégrant les facteurs simples entre les deux niveaux de tension 0,1.Vcc et Vcc-Vts:

$$td(ohm) = \frac{Cd \cdot \ln\left(((19-\beta s)Vcc - 20.Vts/(1-\beta s))/Vcc \right)}{\mu n \cdot Cox \cdot \gamma s \cdot (Vcc-Vts)}$$

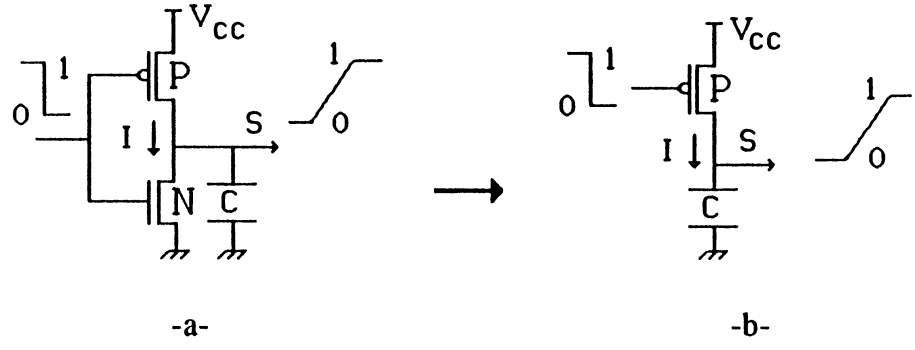
Un résultat similaire est obtenu dans [NEI85] sans tenir compte de l'effet de substrat β.

Comme le temps de descente est la somme de deux: td= td(sat)+td(ohm):

$$td \quad \frac{Cd.(2.(\beta s+1)(Vts-0,1.Vcc)/(Vcc-Vts) + \ln((19-\beta s)-20.Vts/((1-\beta s)Vcc))}{\mu n \cdot Cox \cdot \gamma s \cdot (Vcc-Vts)}$$

Temps de montée:

Pour le temps de montée on suppose qu'on attaque l'entrée de l'inverseur par un échelon descendant d'amplitude Vcc, figure I.18a. Le transistor N est bloqué instantanément. Nous obtenons le schéma électrique équivalent pendant le temps de montée, figure I.18b. Le transistor P traverse deux modes de fonctionnement. Il fonctionne d'abord en saturation puis en mode ohmique.



-a- -b-

Figure I.18- Porte CMOS chargée par C pendant le temps de montée.

Le temps de montée de l'inverseur est la somme du temps de montée dans les deux modes.

On obtient une expression du temps de montée (tm) identique à celle du temps de descente:

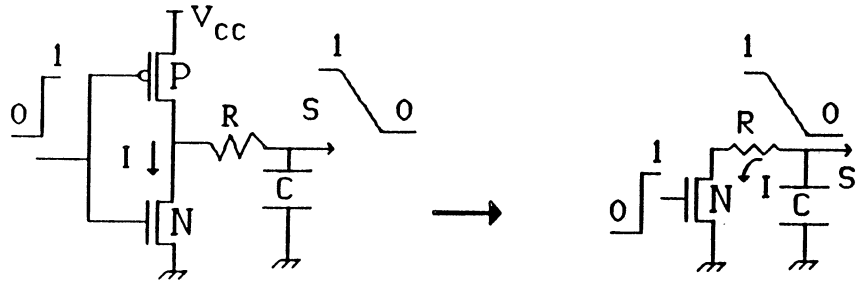
$$tm = \frac{Cm.(2.(\beta c+1)(Vtc-0,1.Vcc)/(Vcc-Vtc) + \ln((19-\beta c)-20.Vtc/((1-\beta c)Vcc))}{\mu p.Cox.\gamma c.(Vcc-Vtc)}$$

Une comparaison des résultats numériques (annexe 1.1) des temps de montée et de descente sont obtenues par calcul d'une part et par simulation d'autre part (simulation MSINC), avec une présision de l'ordre de 15%.

b- Porte chargée par R.C:

Temps de descente:

Nous allons calculer le temps de descente et le temps de montée d'une porte chargée par une résistance et une capacité (R.C). Les mêmes conditions pour le signal d'entrée sont gardées, figure I.19a. Nous obtenons le schéma électrique équivalent pendant le temps de desente de l'inverseur, figure I.19b. Le transistor N traverse deux zones de fonctionnement: saturée et ohmique.



-a- -b-

Figure I.19- Porte CMOS chargée par R.C pendant le temps de descente.

La résistance (R) modifie la relation de la tension de sortie pour l'inverseur chargée par R.C. En effet cette relation devient :

$$Vs = Vds + R.I$$

Nous allons calculer le temps de descente dans la zone saturée (tdr(sat)) et dans la zone ohmique (tdr(ohm)) en tenant compte de R.

- Zone saturée: En zone saturée, on a la relation suivante pour le transistor N:

$$Vgs-Vts < Vds <= 0,9.Vcc \quad \text{où } Vgs=Vcc$$

$$I(sat) = -Cs \cdot \frac{d(Vs)}{dt}$$

Mais Vs devient: $Vs = Vds + R.I(sat)$, on obtient donc:

$$I(sat) = -Cs \cdot \frac{d(Vds + R.I(sat))}{dt}$$

Comme le courant I(sat) du transistor N est indépendant du temps (t) et de la tension drain source (Vds):

$$I(sat) = \frac{\mu n \cdot Cox \cdot \gamma s}{2 \cdot (\beta s+1)} \cdot (Vcc-Vts)^2$$

La dérivée du courant I(sat) par rapport au temps est nulle. Cette égalité du courant devient:

$$I(sat) = - Cs \cdot \frac{d(Vs)}{dt}$$

Le temps de descente en zone saturée (tdr(sat)) s'obtient de la même manière que celui de l'inverseur chargée par C et avec la même expression:

$$tdr(sat) = td(sat) = \frac{Cd \,.\, 2.(\beta s+1)\ (Vts-0,1.Vcc)}{\mu n \,.\, Cox \,.\, \gamma s \,.\, (Vcc-Vts)^2}$$

- Zone ohmique: En zone ohmique la relation de la tension de sortie (Vs) est:

$$Vs = Vds + R.I(ohm). \quad Avec \quad 0,1.Vcc <= Vds <= Vcc-Vts.$$

L'expression du courant I(ohm) dépend de la tension Vds:

$$I(ohm) = \mu n \,.\, Cox \,.\, \gamma s \,.\, (Vcc - Vts - \frac{\beta s + 1}{2} \,.\, Vds) \,.\, Vds$$

L'égalité du courant devient:

$$I(sat) = - Cs \,.\, \frac{d(Vds + R.I(ohm)}{dt}$$

$$I(sat) = - Cs \,.\, \frac{d(Vds)}{dt} - Cs.R. \frac{d(I(ohm))}{dt}$$

On remplace I(ohm) par sa valeur en fonction de Vds on obtient:

$$\frac{d(I(ohm))}{dt} = \mu n \,.\, Cox \,.\, \gamma s \,.\, (Vcc - Vts - (\beta s + 1) \,.\, Vds) \,.\, \frac{d(Vds)}{dt}$$

On remplace I(ohm) et d(I(ohm))/dt par leurs valeurs en fonction de (Vds) dans l'égalité du courant, on obtient:

$$\mu n.Cox.\gamma s.(Vcc-Vts-\frac{\beta s+1}{2}.Vds).Vds = -Cs(1+\mu n.Cox.\gamma s.(Vcc-Vts-(\beta s+1).Vds)) \frac{d(Vds)}{dt}$$

On pose:

$$a = \mu n \,.\, Cox \,.\, \gamma s$$

$$b = Vcc - Vts$$

$$c = \beta s + 1$$

On obtient:

$$-Cs.(1 + R.a.b - R.c.Vds) \,.\, \frac{d(Vds)}{dt} = \frac{a.c}{2} \,.\, (\frac{2.b}{c} - Vds).Vds$$

puis:

$$dt = \left(\frac{-Cs \,.\, (1+a.b.R)}{a.c/2 \,.(2.\, b/c - Vds) \,.Vds} + \frac{Cs \,.\, a.c.R}{a.\, c/2 \,.(2.\, b/c - Vds)} \right) \,.\, d(Vds)$$

Une relation logarithmique du temps de descente (tdr(ohm)) en zone ohmique s'obtient en intégrant cette égalité entre les deux niveaux de

tension (Vcc-Vts) et (0,1.Vcc):

$$tdr = td(ohm) + R.a.b \cdot \left(td(ohm) - \frac{Cs.c}{2.a.b} \cdot \ln\left(\frac{(19-\beta s).Vcc - 20Vts}{5.(1-\beta s)(Vcc-Vts)} \right) \right)$$

Pour R=0, le temps de descente (tdr(ohm)) devient identique à celui de td(ohm). Pratiquement la résistance R augmente le temps de descente de valeur suivante (tr(ohm)):

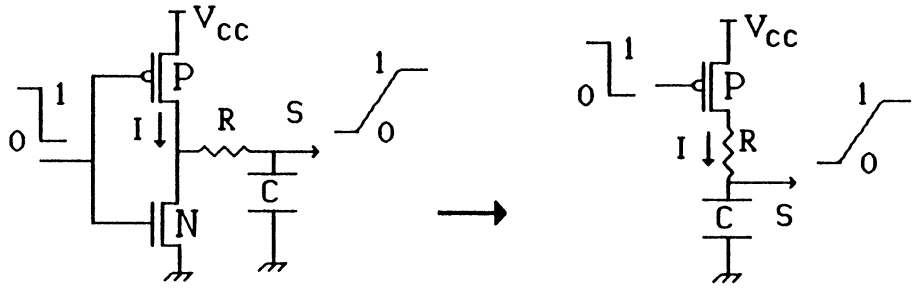
$$tr(ohm) = R.a.b \cdot \left(td(ohm) - \frac{Cs.c}{2.a.b} \cdot \ln\left(\frac{(19-\beta s).Vcc - 20Vts}{5.(1-\beta s)(Vcc-Vts)} \right) \right)$$

Le temps de descente d'un inverseur chargé par R.C (tdr) est la somme de temps en deux zones: saturée et ohmique (pour 0,5<Vts<4). On remplace tdr par sa valeur en fonction de td:

$$tdr = tdr(sat) + tdr(ohm) = td(sat) + td(ohm) + tr(ohm) = td + tr(ohm).$$

Temps de montée:

Avec les mêmes conditions que celles utilisées pour calculer le temps de montée d'un inverseur chargé par C, on calcule le temps de montée (tmr) d'un inverseur chargé par R.C, figure I.20:



-a- -b-

Figure I.20- Porte CMOS chargée par R.C pendant le temps de montée.

On obtient une expression de temps de montée tmr identique à celle du temps de descente tdr à condition de remplacer:

- Vts par la valeur absolue de Vtc.
- γs par γc.
- μn par μp.

Donc: tmr = tm + tr(ohm).

L'annexe 1.2 montre également une comparaison des résultats numériques (pendant le temps de montée et le temps de descente) du temps tr(ohm), par calcul d'une part et par simulation d'autre part (simulation SPICE), avec une présision de l'ordre de 15%.

2.2- Estimation du temps de propagation

Cette méthode d'estimation du temps de propagation d'une structure régulière ou d'un PLA, CMOS NOR-NOR, consiste à normaliser un bloc et à le ramener au cas d'un inverseur pendant une phase de commutation à partir des données suivantes:

- Une technologie donnée
- la programmation spécifique d'un PLA

A partir de ces données, la stratégie de calcul est:

- Génération de la géométrie d'un point de PLA à partir des règles de dessin et des paramètres électriques de la technologie.

- Calcul des capacités de chaque entrée/sortie dans chaque bloc à partir de la géométrie d'un point PLA et de la spécification du contenu d'un PLA.

- Ensuite, nous estimons le temps de propagation du PLA en estimant le temps de propagation dans chaque bloc en cherchant le chemin critique. En effet, on calcule les quatre temps suivants:

- temps de précharge de la matrice ET.
- temps de décharge de la matrice ET.
- temps entre la fin de la décharge de la matrice ET et le début de la

décharge de la matrice OU.

- temps de décharge de la matrice OU.

Le temps de précharge et de décharge peuvent être calculer en calculant le temps de montée et de descente de la matrice correspondante.

Tout d'abord, pour chaque calcul de temps on calcule les capacités correspondantes.

2.2.1- Calcul des capacités

a- Point de PLA

A partir de la géométrie, figure I.21, d'un point de PLA, on peut calculer les paramètres électriques et topologiques correspondants afin de calculer les paramètres électriques d'un PLA en fonction du nombre des points de PLA considéré. Les capacités qui varient avec la tension sont majorées pour qu'elles deviennent constantes [DAN83].

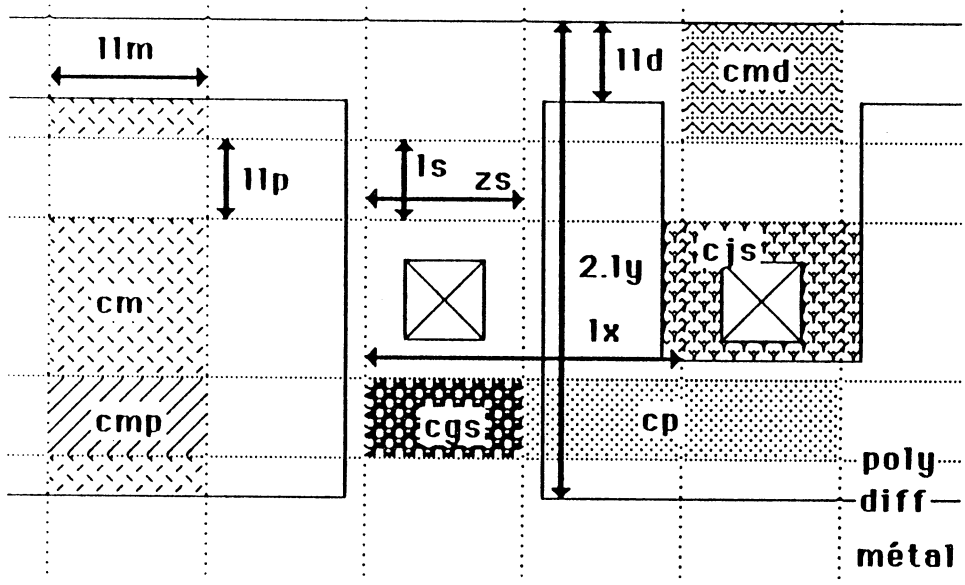


Figure I.21- Les paramètres dans un point de PLA.

- lld: largeur de la ligne de diffusion.
- llp: largeur de la ligne de poly.
- ls, zs: dimensions du canal du transistor N dans les matrices.
- lx, ly: dimensions du point de PLA.

- cgs: capacité grille-source.
- cp: capacité poly-substrat sur oxyde épais.
- cmp: capacité métal-poly. S'il y a deux couches métals, il faut calculer les deux capacités correspondantes.
- cj: capacité de jonction entre la diffusion et le substrat.
- cmd: capacité diff-métal.
- cm: capacité du métal-substrat.

b- **Capacités de la matrice ET**

Vu la structure électrique de la matrice ET (portes NOR), les transistors N sont en parallèle entre une ligne de métal et une ligne de diffusion. Le calcul des capacités de ET consiste à calculer les trois capacités suivantes:

1°- La capacité de la ligne d'entrée (Ceet).
2°- la capacité de sortie pendant son temps de montée (Cmet).
3°- La capacité de sortie pendant son temps de descente (Cdet).

1°- Capacité de la ligne d'entrée (Ceet):

La capacité d'une ligne d'entrée de ET correspond à la capacité de la ligne d'entrée polysilicium. Elle dépend de celles du point de PLA (cmp, cp), de la longueur de la ligne d'entrée en nombre de pas (npe) et du nombre de grilles sur cette ligne (nte):

$$Ceet = npe \,.\, (2.cmp + cp) + nte \,.\, cgs$$

2°- Capacité de sortie de ET pendant le temps de montée (Cmet):

Pendant le temps de montée de la matrice ET la ligne métal (monôme) de sortie évolue de 0 à 1 logique ; ainsi que la ligne de diffusion qui connecte les sources des transistors N entre eux. Le calcul

de Cmet consiste à effectuer la somme des capacités de la ligne métal et de celles de la ligne de diffusion.

Les capacités sur la ligne de métal par pas de PLA sont:

- cm: capacité métal-substrat.
- cmd: capacité métal-diffusion.
- 2.cmp: capacité métal-poly qui agit en effet Miller.
- cjs: capacité jonction-substrat des transistors
- ct = 5/2cgc+cjc+cjss+2/3cgs: capacité, due aux deux inverseurs, qui a déjà été calculée. cjss est la capacité de jonction du transistor de connexion à la masse.
- cld: capacité de la ligne de diffusion en fonction de la surface du fond et du bord de la ligne par pas de PLA.

Cmet s'obtient alors en fonction du nombre des pas de la ligne métal (npm) et du nombre des drains de transistor (ntm) sur la ligne:

$$Cmet = ntm.cjs + npm.(cm+cmd+2.cmp) + ct + cld.$$

3 - Capacité de sortie de ET pendant le temps de descente (Cdet):

La capacité de sortie de la matrice ET pendant le temps de descente (Cdet) a la même expression que celle de Cmet à condition de remplacer l'expression de la capacité entre deux inverseurs par sa valeur. Cette

valeur a déjà été calculée :

ct = cjc + cjss + 5/3.cgs + 2/3.cgc.

c- Capacité de la matrice OU

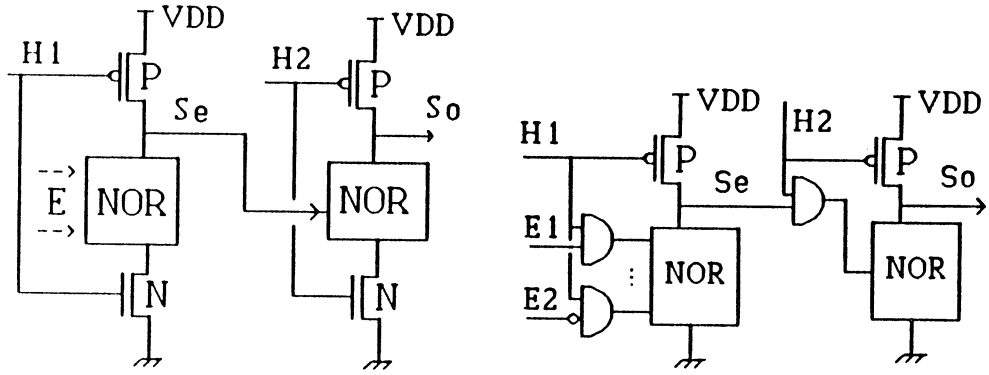
La structure de la matrice OU est identique à celle de ET (dans le cas où les transistors N sont directement connectés à la masse, structure mixte, on n'ajoute pas la capacité de la ligne de diffusion cld). Ceci consiste à calculer:

- La capacité de la ligne d'entrée (cl) est calculée comme Ceet à condition de remplacer (npe, nte) par (ns, s: longueur de monôme en poly dans OU par nombre de pas de PLA et nombre de grilles respectivement).

- Les capacités de sortie de la matrice OU, pendant le temps de montée (Cmou) et de descente (Cdou), ont les mêmes expressions que celles de Cmet et Cdet, en remplaçant (npm, ntm) par (nps, nts: longueur de la ligne de sortie en nombre de pas et nombre de drains sur cette ligne respectivement). On ajoute aux deux capacités, une capacité de charge sur la sortie (cch).

2.2.2- Calcul du temps de propagation

Nous allons traiter les deux structures électriques NOR-NOR, figure I.22. chaque structure correspond à deux blocs qui vont être traités à part, ainsi que le temps de commutation.



-a- -b-

Figure I.22- Deux structures NOR-NOR de PLA CMOS.

a- La structure (a) comporte des blocs NOR (transistors N, Tn) qui réalisent la matrice ET ou la matrice OU, avec des transistors N de couplage à la masse (Tnc) et des transistors P de commutation à l'alimentation (Tp). Nous allons traiter le temps de propagation dans chaque matrice pendant les deux phases de précharge et d'évaluation.

La phase de précharge d'une matrice produit:

- Un niveau logique haut sur les sorties
- Des niveaux logiques stables sur les entrées.

Le temps de précharge d'une matrice est le temps maximum nécessaire pour charger les sorties (temps de montée) en suivant le chemin le plus long parmi les différentes sorties.

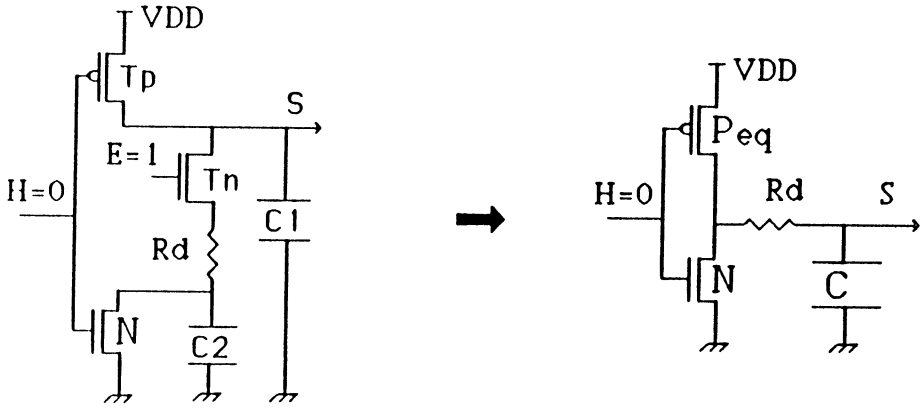
Le temps de montée d'une porte NOR est calculé à partir d'un schéma électrique équivalent dans le cas le plus défavorable, figure I.26a:

- H=0, le transistor de couplage à la masse Tnc est bloqué, le Tp conduit.
- Un seul transistor Tn conduit et les autres représentent des capacités parasites.

Le temps de montée dépend des paramètres suivants:

- Les dimensions du transistor Tn ($\gamma s=ws/ls$) sont données par le dessin du point de PLA.
- Les dimensions du transistor Tp ($\gamma c=wc/lc$) qui donnent un degré de liberté en donnant γc.
- La capacité de sortie (Cm) est calculée à partir de la technologie et du contenu de la matrice.
- La résistance de la ligne de diffusion (cette ligne est parallèle aux lignes d'entrées, figure I.3) qui connecte les sources des Tn en tenant compte des autres courants qui débitent dans la ligne. Ceci donne une résistance équivalente $Rd=n(n+1)R/8$.

n: nombre des pas entre deux Tnc. Ceci donne aussi un degré de liberté en donnant n.



-a- -b-

Figure I.23- Schémas électriques équivalents d'une matrice pendant son temps de montée.

A partir du schéma électrique équivalent, figure I.23a, le temps de montée (Tm) est majoré par le schéma électrique équivalent, figure I.23b, en remplaçant les deux transistors (Tp et Tn) par un transistor équivalent Tpeq et en tenant compte des mobilités:

$$\gamma ceq = \gamma c.\gamma s / (\mu_p/\mu_n \, . \, \gamma c+\gamma s).$$

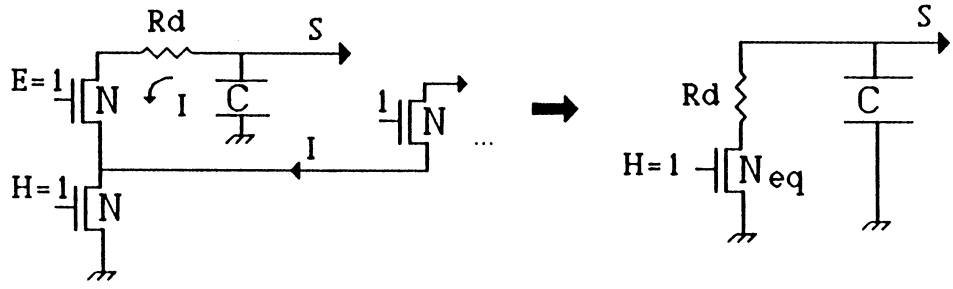
L'annexe 1.3 montre la validitée de cette majoration.

On obtient alors le schéma électrique d'un inverseur qui charge une capacité Cm au travers d'une résistance Rd, et dont on peut calculer le temps de montée.

Pendant la phase d'évaluation (H=1), le schéma électrique équivalent, figure I.24a, dans le cas le plus défavorable correspond à un seul transistor TN qui conduit et les autres sont bloqués. Ils représentent des capacités parasites.

La sortie se décharge au travers du transistor Tnc (γss). En tenant compte des autres sorties (n sorties) qui peuvent se décharger au travers du même transistor Tnc. On obtient le schéma I.24b en remplaçant les transistors Tn et Tnc par un transistor équivalent Tnceq:

$$\gamma ssseq = (\gamma s.\gamma ss/n) / (\gamma s+\gamma ss/n).$$



-a- -b-

Figure I.24- Schémas électriques équivalents d'une matrice pendant le temps de descente.

Ceci donne un schéma électrique équivalent à un inverseur où sa

capacité de sortie se décharge au travers d'une résistance dont on peut calculer le temps de descente.

Le temps de commutation entre les deux blocs intervient dans le temps de montée et de descente du premier bloc, en tenant compte de la capacité (Cl) et de la résistance (Rl) de la ligne d'entrée correspondante dans le deuxième bloc. Cette capacité (Cl) et cette résistance (Rl) de la ligne d'entrée poly s'ajoutent à la capacité et à la résistance de la sortie du premier. Ceci augmente le temps de propagation.

b. Deuxièmement, nous allons traiter la structure NOR-NOR de la figure I.22b. Les transistors Tn sont directement connectés à la masse.

Pendant la phase de précharge (H1=H2=0), tous les transistors Tn sont bloqués. Ils représentent des capacités parasites. Le transistor Tp conduit. Ceci peut donner un schéma électrique équivalent à un inverseur qui est chargé par une capacité, et dont on peut calculer le temps de montée.

On ajoute à ce temps de montée (pour calculer le temps de précharge d'une matrice) le temps nécessaire pour décharger toutes les lignes d'entrées à "0".
Ce temps est équivalent au temps de descente d'un inverseur chargé par R.C (où R est la résistance et C est la capacité de la ligne d'entrée la plus

longue) ; car chaque entrée est attaquée par une porte.

Pendant la phase d'évaluation (H1=H2=1), Les sorties peuvent se décharger au travers des transistors Tn, (toujours dans le cas le plus défavorable où un seul transistor Tn conduit). Ceci donne un schéma électrique équivalent, à un inverseur qui décharge sa capacité de sortie au travers d'une résistance Rd équivalente à la ligne de diffusion qui connecte les sources des transistors Tn entre elles.

Il faut ajouter au temps de descente (pour calculer le temps de décharge d'une matrice), le temps nécessaire pour charger la ligne d'entrée la plus longue à "1" (ou du transistor correspondant). Ce temps détermine aussi le temps de commutation entre deux blocs.

En vue de réduire ce temps, nous allons utiliser des boot-strap entre les deux matrices, figure I.25, qui donnent une tension grille-source du transistor de charge de l'ordre de 3/2VDD due à sa capacité grille-drain (cgd). En effet, quand H=0 et la sortie de ET est à 1, la capacité cgd se charge à VDD ; lorsque H évolue à 1 la tension grille-source du transistor de charge augumente par l'effet capacitif). Ceci augmente le courant disponible pendant le temps de montée et réduit le temps de commutation.

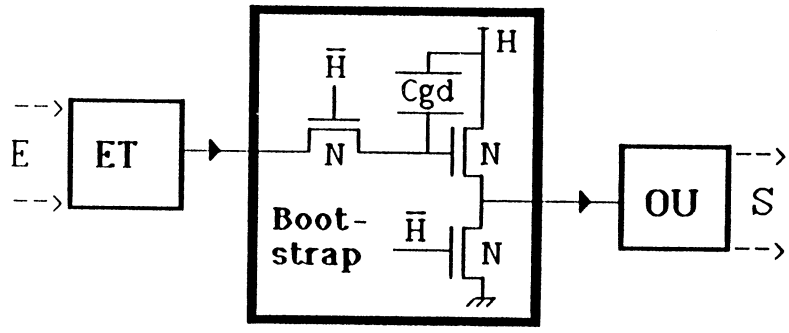


Figure I.25- Boot-strap entre deux blocs.

En vue de réduire le temps de propagation du PLA pendant la phase d'évaluation, une solution peut être envisagée. Elle s'agit d'une structure mixte, figure I.26.

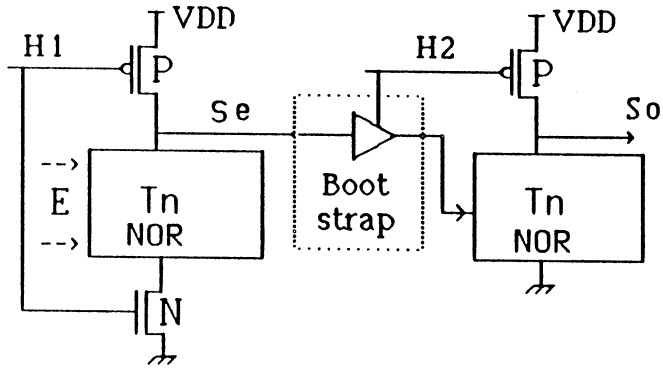


Figure I.26- Structure mixte.

Les avantages de cette structure par rapport aux deux précédentes sont:

- Le temps de descente de la matrice OU peut être réduit, en supprimant le transistor de couplage à la masse Tnc.

- Le boot-strap réduit le temps de descente du deuxième bloc par l'effet de chargement rapide de la ligne d'entrée. Il sépare la résistance et la capacité des deux blocs (ligne poly d'entrée du deuxième bloc, ligne sortie correspondante au premier bloc.

- Le premier bloc garde la première structure. Car pour chaque entrée (E), on a une entrée complémentaire (E'). Ceci demande une surface importante de la circuitrie des Boot-straps d'entrées.

L'autre raison de garder la première structure est la suivante: les entrées de la matrice ET sont activées pendant la phase de précharge qui est un temps mort après l'exécution.

En résumé, le temps de propagation du PLA de la structure mixte est la somme de trois temps suivants:

- Temps de descente du premier bloc qui correspond à sa sortie la plus lente.

- Temps du retard entre les deux blocs par l'effet de boot-strap qui correspond à la ligne la plus longue. Ce temps est multiplié par le facteur Vts/VDD (le transistor conduit à partir de la tension de seuil).

- Temps de descente du deuxième bloc qui correspond à sa sortie la plus lente.

Conclusion

Ce premier chapitre concernant l'évaluation électrique et temporelle de PLA CMOS, est basé sur la recherche d'un chemin critique dans chaque matrice pour le ramener à une porte simple. Ceci donne le temps de précharge et le temps de décharge de chaque matrice et aussi le temps du retard entre les deux matrices.

Le temps de propagation dans un PLA est la somme des temps correspondant aux chronogrammes pour chaque structure, figure I.6.

Les étapes successives nécessaires à cette recherche sont:

- Etude des portes simples pour une technologie donnée. Ceci définie les formules de base pour calculer le temps de réponse.

- Etude d'un point du PLA. Ceci définie les dimensions de transistors N des deux matrice et les dimensions des interconnexions.

- Calcul des paramètres électriques (capacité et résistance de chaque ligne, ...) à partir de la spécification d'un PLA donné et dimensionnement des transistors de précharge P (respectivement de décharge N).

- Recherche d'un chemin critique dans chaque matrice pour donner un schéma électrique équivalent à une porte, où on peut calculer le temps de propagation (formules de base §2.1). Ceci définie le temps de

précharge et le temps de décharge de chaque matrice. En effet, un système interactif est réalisé entre le temps de réponse et les paramètres suivants :

- γc: dimension du transistor P de couplage à l'alimentation
- γss: dimensions du transistor N de couplage à la masse
- n: nombre de monômes entre deux transistors de couplage à la masse ou entre deux rappels de masse).

Ceci permet au concepteur de choisir la réalisation d'un PLA en fonction des paramètres qui agissent sur le temps de propagation, la surface et la consommation.

CHAPITRE II

DEFAUTS DE FABRICATION ET TEST DES PLA CMOS

Introduction

La complexité des circuits VLSI actuels, rend difficile leur test: cependant ce test est crucial, car plus les dimensions des éléments de base diminuent, plus les modes de pannes se compliquent [ELZ83]. Ce chapitre étudie ce problème dans le cas des PLA.

Les défauts qui peuvent se produire dans un circuit intégré, outre les défauts de conception, sont: soit des défauts en fin de fabrication, soit des défauts pouvant survenir au cours de la vie d'utilisation des circuits.

La phase de test est alors un besoin vital pour un PLA : soit pour vérifier l'absence de pannes afin de vérifier qu'il fonctionne correctement, soit pour détecter et localiser des pannes afin de corriger un circuit partiellement en panne s'il s'agit d'une structure PLA reconfigurable.

En vue de détecter les pannes, trois différentes méthodes de test sont introduites par [SMI79]:

- Dans la première, on ajoute de la circuiterie spéciale pour le test sur les matrices [SIG76]. Cette circuiterie est activée en plaçant sur les E/S un niveau électrique qui est en dehors du niveau normal. Cependant, cette méthode ne teste pas le PLA dans des conditions normales de fonctionnement. De plus, elle est difficile à appliquer, lorsque le PLA est placé dans un système.

- Dans la deuxième, la méthode de test est exhaustive: tous les vecteurs d'entrées possibles sont appliqués sur les matrices, afin de vérifier si les réponses sont correctes. Cette méthode ne s'applique qu'à des PLA isolés (non intégrés dans un circuit) et nécessite des équipements de test à grande vitesse et plus sophistiqués. Elle devient impraticable pour un PLA ayant un nombre important d'entrées.

- Dans la troisième, la méthode de test consiste à sélectionner un certain nombre de configurations spéciales de test et à ne tester que ces configurations. Ceci permet d'effectuer un test avec un nombre de vecteurs de test relativement faible.

- D'autres voies de test sont aussi ouvertes, il s'agit d'ajouter de la circuiterie (registres de signature, ...) en vue de rendre le PLA facilement testable ou autotestable [WAN79].

Ce chapitre comporte deux parties:

I- La première partie étudie les défauts physiques de fabrication des circuits intégrés, afin d'obtenir la distribution des types de pannes dans les PLA. Ceci simplifie la définition d'une structure reconfigurable, où

seuls les types de pannes les plus probables seront corrigés (chapitre III).

II- La deuxième partie est une approche vers le test des PLA intégrés dans des circuits VLSI, en particulier pour la technologie CMOS: définition d'un modèle de panne et d'une méthode de test.

I- Les défauts physiques de fabrication

1.1- Introduction

Nous allons classifier les défauts physiques de fabrication [CAS76, ELZ83, MAL84, MAN80] des circuits intégrés et en particulier des PLA en technologie CMOS. Ceux-ci comportent les mécanismes de défaillance ainsi que leur manifestations logique et électrique dans cette technologie. On étudie la distribution des types de pannes pour déterminer les plus fréquents et les éléments les plus souvent défectueux. En effet, cette étude présente pour les structures reconfigurables l'intérêt suivant : seuls les types de pannes les plus probables seront envisagés. Ceci permet:

- de reconfigurer dans la majorité des cas.
- de réduire la surface de circuiterie ajoutée par la reconfiguration, et donc de diminuer le nombre d'interrupteurs à programmer. Par conséquent, le surcoût de fabrication peut être diminué (le temps pour griller les fusibles) avec une surface minimum à ajouter.

Deux méthodes peuvent être utilisées conjointement à partir de l'étude de la distribution des types de pannes:

a- La première méthode consiste à déterminer les défauts de fabrication les plus probables au niveau des composants, et à ajouter la circuiterie nécessaire dans la structure reconfigurable, afin de pouvoir corriger uniquement ces pannes.

b- La deuxième méthode consiste à réduire la probabilité d'occurrence des défauts physiques les moins probables. Dans ce cas, il n'y a pas d'utilisation de circuiterie de reconfiguration mais par contre nous allons prendre en compte de certaines règles de dessin.

Une défaillance due aux défauts physiques, dans un circuit intégré, se manifeste par un changement physique local. Ces derniers se traduisent par des défauts des composants qui engendrent (ou non) de mauvais fonctionnements (changement des caractéristiques électriques, erreurs logiques, ...).

Ces mécanismes de défaillance présentent une difficulté de classification due au mélange des causes de défaillance et des effets [ELZ83] (e.g: Les phénomènes de transport peuvent être considérés comme cause, mais le claquage diélectrique est considéré comme effet). Une même cause peut donner des effets différents ou bien plusieurs causes peuvent produire le même effet.

Une défaillance (cause ou effet) peut provenir soit des défauts

résultant de la fabrication des circuits, soit de l'usure au cours de la vie utile des circuits.

L'étude des défauts de fabrication est un problème vital en vue d'étudier les structures reconfigurables dans les circuits intégrés ; ceci nécessite de pouvoir détecter et localiser une panne.

1.2- Mécanismes de défaillance

Une technologie est définie comme une séquence d'étapes technologiques élémentaires. En effet, une tranche de circuit intégré (MOS) se compose de trois types de couches, autre que le substrat, figure II.1:

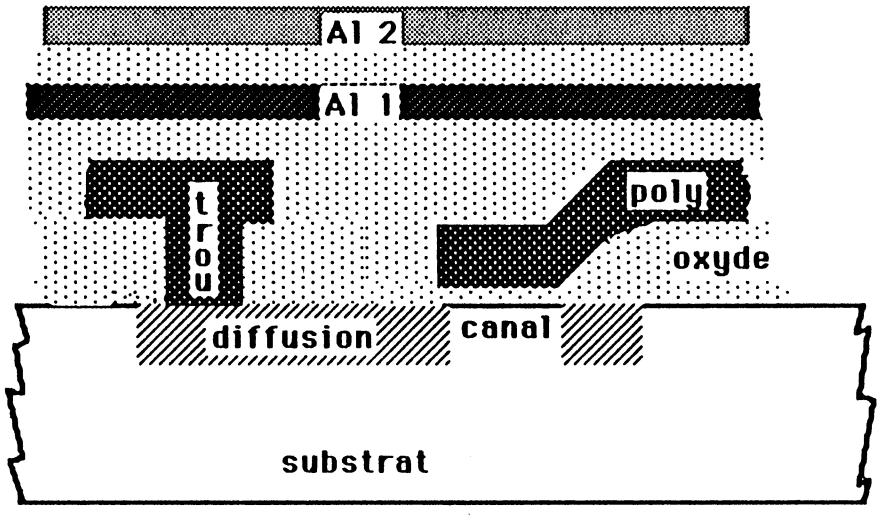


Figure II.1- Les trois couches d'une technologie MOS

- Une couche inférieure de connexions réalisées sous forme de diffusions tracées dans un substrat en Silicum.

- Des couches supérieures de connexions réalisées sous forme de métallisations (Si-Poly, Al et éventuellement deuxième Al).

- Des couches d'oxyde épais (SiO2) qui isolent deux couches conductrices: cette couche comporte des ouvertures pour permettre le contact entre les deux couches inférieure et supérieure. Et des couches d'oxyde mince entre Silicium polycristallin et diffusion pour former des grilles.

L'analyse du processus de fabrication de la tranche (les couches) peut permettre de déterminer les défauts en fin de fabrication les plus probables, en effet, chaque couche subit plusieurs procédés technologiques. Ceci produit des défauts physiques dont la probabilité d'occurence augmente avec le nombre des étapes pour chaque technologie, toutes choses égales par ailleurs.

Muroga [MUR82] présente par exemple une séquence de conception et de fabrication de circuits intégrés, où on ajoute une étape de reconfiguration, figure II.2.

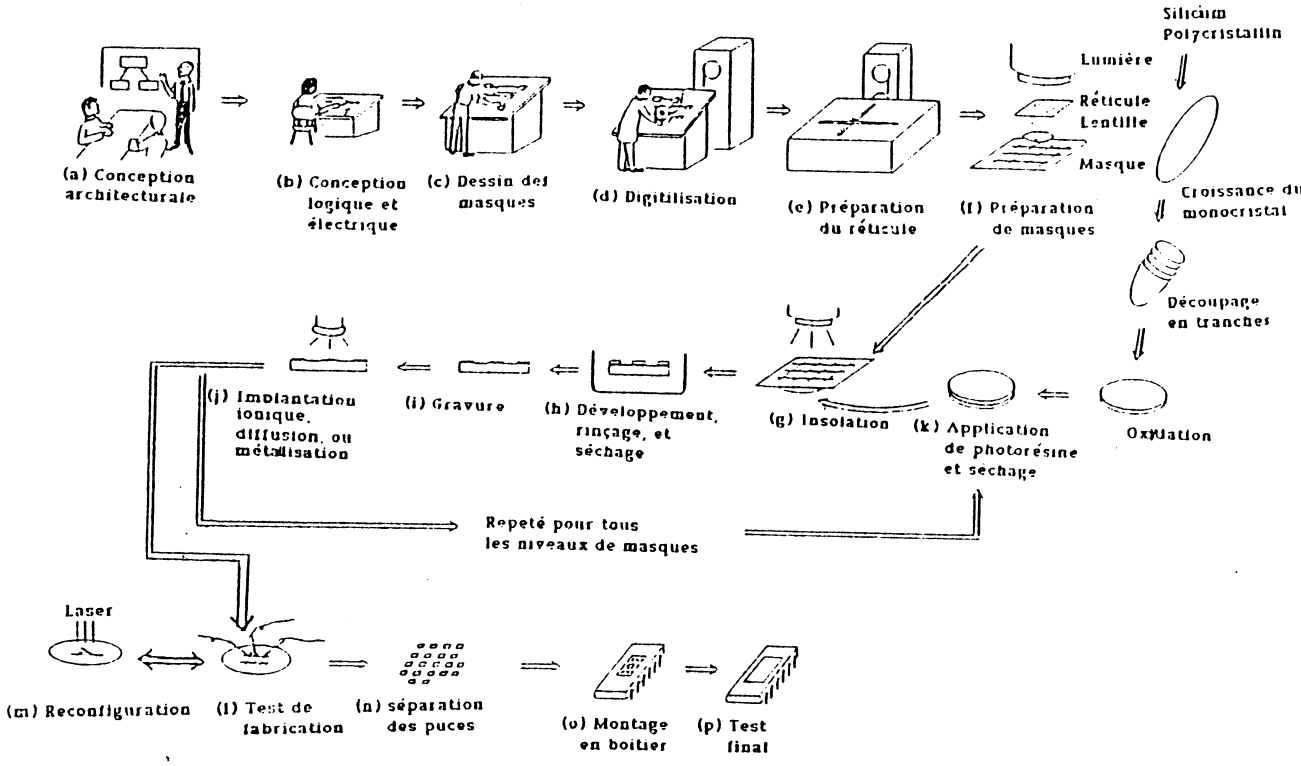


Figure II.2- Séquence de conception et de fabrication de circuits intégrés.

Citons quelques exemples des défauts de fin de fabrication dans chaque couche :

- La réalisation des couches diffusées et des interconnexions exige chacune un procédé de masquage sélectif de zones de la surface. Ceci peut produire des défauts en masquant partiellement une zone qui ne doit pas être masquée ou par le non-masquage d'une zone qui doit être masquée. Ces défauts peuvent se manifester, soit par des courts-circuits dans chaque couche, soit par des coupures.

- Les processus de formation de la couche d'oxyde (épais ou mince) affectent de manière uniforme toute la surface de la tranche et permettent par conséquent d'obtenir des couches homogènes avec des

possibilités de défauts limités.

Du fait que l'oxyde épais constitue un très bon isolant, on peut admettre que la probabilité d'apparition de courts-circuits entre deux couches (métallisation-diffusion) est négligeable (sauf Al1/Al2).
En ce qui concerne les trous dans la couche d'oxyde pour les contacts, une mauvaise prise de contact entre les deux couches conductrices est possible. Ce défaut est équivalent à la coupure de l'équipotentielle qui traverse l'oxyde au niveau du trou. Ceci produit des défaillances du transistor par une coupure du drain ou de la source.

Par contre, au niveau de la couche d'oxyde mince, des défauts peuvent survenir, soit par amincissement local qui favorise le claquage par décharge électrostatique, soit par la présence de charge due à une pollution ponctuelle qui provoque une dérive de la tension de seuil pour le transistor concerné.
Ces défauts peuvent se manifester respectivement, par des courts-circuits résistifs entre la grille et le drain (ou la grille et la source), et par une dérive de la tension de seuil, qui rend un transistor nMOS conducteur en permanence (stuck-on (s-on): court-circuit entre les diffusions ou canal en permanence), et qui rend un transistor pMOS ouvert en permanence (stuck-open (s-op): pas de canal).

En dehors des pannes causées par des outillages des processus technologiques tels que les masques, il existe d'autres mécanismes de défaillance, tels que :

- Les phénomènes de transport dans les solides résultent de la diffusion de particules dans un milieu. Ceci dépend de leur concentration et de la température ainsi que d'autres facteurs, tels que la force électrostatique et la force due au vent électronique, en présence d'un champ électrique (ces phénomènes dans ce cas s'appellent l'électromigration). Les conséquences de la diffusion de particules sont:

  * Une création de déplétions (pits)-accumulations (hillocks ou whiskers) dans les métallisations, qui peuvent produire des coupures ou casser la passivation.

  * Les phénomènes d'interdiffusion qui peuvent intervenir à l'interface de deux métaux (ou deux conducteurs). Ceci résulte de la diffusion, en surface, des atomes de l'un à l'autre (normalement on cherche à avoir d'interdiffusion profonde pour obtenir une bonne cohésion entre deux surfaces). Un déplacement de l'interface entre les métaux peut provoquer la dislocation.

- Les croissances cristallines peuvent être dues à des phénomènes électrochimiques, en présence d'humidité et d'un champ électrique où se développer en dehors de toute action électrochimique. Dans tous les cas elles risquent de créer des courts-circuits dans les métallisations.

- Les phénomènes des claquages sont soit de diélectrique, soit de conducteurs. Le claquage diélectrique se traduit par un déplacement d'ions positifs qui peuvent produire des courts-circuits dans la jonction NP. Le claquage de conducteurs se produit pendant la fusion de métallisation et donne une coupure.

- Les effets thermomécaniques sont dûs aux coefficients de dilatation. Ils peuvent produire des coupures des fils de liaison avec l'extérieur ou dans la ligne de métallisation. De plus, sans qu'il y ait de conséquence mécaniques, une variation de la tension de seuil peut apparaître par l'effet de la température.

- La contamination ionique (mécanisme de défaillance importante) sur la surface de la puce, causée par les procédés technologiques ou autres, produit des charges sur la surface ou dans l'oxyde. Ceci modifie les caractéristiques électriques des composants la dérive de tension de seuil qui rend le transistor toujours conducteur (s-on) ou ouvert (s-open).

- L'énergie due à des radiations et des vibrations peut provoquer des défauts au niveau supérieur (Aluminum). Des éléments radioactifs qui créent des paires électrons-trous, sont capables de casser des liaisons atomiques. Ainsi des électrons chauds qui entrent dans l'oxyde mince, peuvent changer la tension de seuil.

- La corrosion chimique causée principalement par l'humidité du milieu est un mécanisme important (la passivation est une solution contre la corrosion). Ceci peut produire des coupures dans les métallisations.

D'autres auteurs [CAS76, ELZ83, MAL84, SIE82] ont étudié les mécanismes de défaillance de fabrication :

* [MAL84] distingue les procédés de fabrication du niveau de couche et la structure de la couche. Il classe les défauts physiques en trois catégories:

- non-homogénéité du substrat.
- contamination locale de la surface.
- défauts ponctuels dans la lithographie.

1°- Le non-homogénéité du substrat provient d'un désordre du réseau du substrat ; une dislocation peut être causée soit par des impuretés, soit par des chocs thermiques ou mécaniques.

2°- La contamination locale sur la surface est causée par différentes origines, telles que les impuretés introduites au cours des procédés chimiques (l'eau utilisée pendant les procédés). Ceci, par exemple, peut changer la mobilité du canal si la contamination se situe entre Si et SiO2. Des "pinholes" dans l'oxyde de grille engendrent un court-circuit grille/canal dans le transistor.

3°- Les défauts ponctuels peuvent survenir pendant les procédés de photolithographie à cause de la poussière ou des trous sur les masques. Des courts-circuits drain-source, pinholes dans l'oxyde, de manque de traits (missing features) ou de manque de transistors peuvent se produire.

* Par ailleurs, [CAS76] propose également trois types de défaillance de fin de fabrication ; mais au lieu des non-homogénité du substrat propose les défauts dans les niveaux de métallisation (Al, coupures ou

courts-circuits dans les lignes de connexion) comme type de défaillance majeure.

Quoique chaque défaut est présumé causer une seule faute, ceci peut produire un effet électrique ou logique sur plus d'un transistor ou plus d'une ligne d'interconnexion.

I.3- Distribution des types de pannes

L'étude de la distribution des types de pannes résultant de processus de fabrication présente un intérêt important pour les deux domaines suivants :

- Amélioration des facteurs technologiques (sources de pannes).
- Testabilité des pannes.

a- La connaissance de la distribution des pannes peut permettre dans le premier domaine une amélioration technologie de fabrication. Ceci consiste à étudier des facteurs technologiques en vue de réduire les défaillances qu'ils engendrent.

En effet, l'avancement et la complexité de la technologie de circuits intégrés, tels que les LSI/VLSI, résultent de ces facteurs que nous nous limitons à citer :

- Les techniques de dépôt (isolants et conducteurs)
- Les techniques d'oxydation et de traitement thermique
- Les techniques de la photolithographie
- Les techniques de dopage
- Les techniques de gravure
- Les techniques de fabrication de masques (précision et propreté)
- La conception assistée par ordinateur CAO.

b- Par contre, le deuxième domaine (qui nous concerne ici) concerne la testabilité. Dans ce domaine, l'étude de la distribution des types de pannes a pour but de réduire le temps de test et de localisation des pannes en cherchant les plus probables.

Dans le cas des circuits reconfigurables on ne cherchera à corriger que les plus probables. Les circuits reconfigurables auquels nous nous intéressons, sont des structures régulières et principalement des PLA reconfigurables en technologie CMOS.

Ce problème de distribution des types de pannes (pourcentage de chaque type des pannes) est traité par différents auteurs [CAS76, SCH78, RAC] [UCL82]. Les données sur les pourcentages de pannes extraites par chaque auteur ne sont pas toujours cohérentes. Ceci peut être dû aux différences de :

- classification utilisé
- méthode de test
- circuit d'étude
- technologie de fabrication.

Les données extraites par les trois premiers auteurs sont relativements cohérentes:

- [CAS76] présente la distribution des types de pannes de circuits défectueux, en technologie CMOS, par densité de surface. Il classe ces données en trois ensembles suivants:

a- Les défauts de photolithographie

1- les pinholes dans l'oxyde	$2,5/cm^2$
2- le manque des traits (missing features)	$2,5/cm^2$
3- les courts-circuits drain-source	$2,5/cm^2$

b- Les défauts de métallisation

1- les courts-circuits entre lignes de métal	$2/cm^2$
2- les coupures dans les lignes de métal	$2/cm^2$

c- Les défauts de la diffusion $2/cm^2$

On remarque, à partir des données de [CAS76], que les défauts de photolithographie sont les plus importants. Ceux-ci engendrent principalement de mauvais fonctionnements de transistors (pinholes, courts-circuits drain source ...).

- [SCH78] présente ces données en pourcentage des différents types de pannes en technologie MOS. Ces données sont étudiées par rapport aux composants électriques (transistors, lignes), pour cinq catégories de mécanismes de défaillance:

1- transistor conducteur par contamination de Na	35%
2- court-circuit dans l'oxyde de grille	25%
3- défauts de métallisation	20%
4- dégradation mécanique (bonding, schathed metal...)	15%
5- courant de fuite dû à l'humidité (moisture leakage)	5%

On remarque que les deux premières catégories sont dominantes. Celles-ci engendrent principalement, aussi, de mauvais fonctionnements de transistors (transistor conducteur par contamination et court circuit dans l'oxyde de grille).

- Le tableau II.1 résume des données extraites de [RAC] Reliability Analysis Centre of Rome Air Developpement Centre (RADC). Il présente des pourcentages des types de pannes de circuits CMOS en mauvais fonctionnement, pour différents degrés d'intégration (MSI, SSI, LSI).

mécanismes de défaillance	nombre des pannes de ce type	pourcentage relatif
surface	65	38%
oxyde	54	32
métallisation	14	8
diffusion	10	6
substrat	12	7
E/S de circuit	15	9

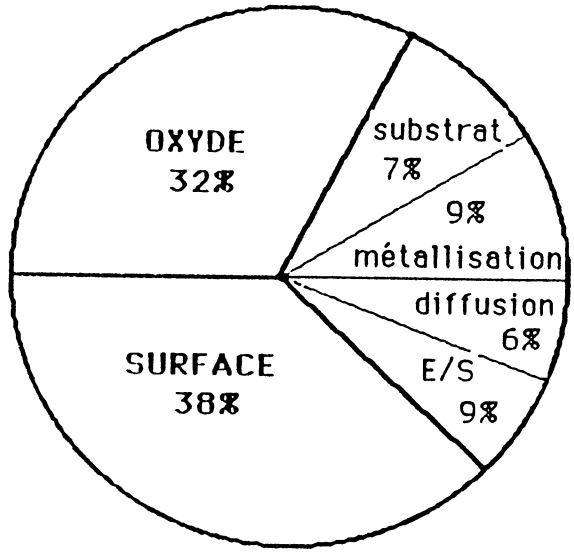


Tableau II.1

On remarque que les types de pannes de surface et d'oxyde sont dominantes. Ceux-ci engendrent encore des mauvais fonctionnements de transistors. Nous allons étudier leur manifestation électrique et logique (tableau II.2).

On admettra que ces pourcentages relatifs des types de pannes

peuvent rester valables pour les systèmes VLSI.

- D'autres données en pourcentage, extraites de [UCL82], présentent les mécanismes de défaillance dominant comme le court-circuit entre les lignes métal adjacentes (resp. diffusion) 53%. Ceci concerne à la fois les lignes d'interconnexion et les transistors.

En ce qui concerne les PLA, nous pouvons nous baser sur les données extraites par les trois premiers auteurs cités précédemment, où les défaillances de transistors sont importantes, pour les raisons suivantes:

1°- Les lignes de diffusions sont intercalées avec des lignes de poly (à cause de la régularité des PLA). Donc le court-circuit entre deux lignes de diffusion est moins probable.

2°- Les lignes de métal dans les PLA sont régulières. On peut imposer alors des contraintes dans les régles de dessin qui peuvent réduire la probabilité de court-circuit en augmentant la distance de garde métal/métal (en gardant les même dimensions du point de PLA).

1.4- Manifestations électriques et logiques des pannes

Chaque défaillance physique peut causer une panne. Celle-ci peut affecter plus d'un transistor ou d'une interconnexion dans le circuit.

Les manifestations électriques et logiques des pannes dépendent, non

seulement, des mécanismes de défaillance mais aussi de la circuiterie concernée. Exemple : un défaut physique (coupure d'une ligne) causé par la contamination et par la migration dans le métal, peut se manifester dans un circuit comme une panne transitoire, due à un couplage capacitif; par contre, dans un autre circuit il peut se manifester comme une panne de collage.

Différents auteurs ont déjà classé ces types de pannes électriques et logiques, sous trois catégories principales :

- circuit ouvert
- court circuit
- transistor MOS s-on/s-open).

Nous allons classer ces types de pannes électriques et logiques par ordre de fréquence, en vue de corriger dans la structure reconfigurable les composants défaillants les plus probables. Ces composants du circuit intégré (ou du PLA) sont les transistors et les lignes d'interconnexions.

- Les défaillances dans le transistor sont :

les courts-circuits entre (grille/source, grille/drain, grille/substrat, drain/source), grille flottante ou grille ouverte, drain ouvert et source ouverte.

- Les défaillances des interconnexions sont :

les coupures des lignes ou des circuits ouverts, et les courts-circuits entre deux lignes qui sont soit du même niveau (Al/Al, Poly/Poly, Diff/Diff) soit à des niveaux différents (Al/Poly, Al/Diff, Diff/Poly).

* [MAN84] présente un pourcentage d'occurrence des pannes en fin de fabrication dans ces composants (transistors et interconnexions):
En effet, l'occurrence des pannes dans un circuit intégré VLSI/WSI concerne pour 3O% à 40% les interconnexions, le reste affectant les transistors.

Le tableau II.2 montre cette relation des défaillances physique/électrique (ou physique/composant) pour chaque type de panne.

Défaut général	Panne physique	Mode de défaillance électrique (composant)
oxyde	pinhole claquage cassure court-circuit charge piégée électron chaud manque d'oxyde mince	MOS (s-on, s-open) interconnexion (coupure et court-circuit entre les deux couches conductrices)
surface	contamination ionique courant de surface charge piégée électromigration défaut photolithographie	MOS (s-on,s-open) interconnexion (circuit ouvert, court-circuit)
métallisation	corrosion masque électromigration contamination ionique vibration défaut photolithographie	interconnexion (Al circuit-ouvert, court-circuit)
diffusion, poly	claquage contact résistif coupure masque défaut photolithographie	MOS (s-on,s-open) interconnexion (circuit-ouvert, court-circuit)

Tableau II.2

A partir de l'étude de la distribution des types de panne (surface et oxyde dominants), et particulièrement dans une structure régulière, la défaillance des transistors intervient pour un pourcentage important dans la défaillance des composants. Ceci implique (pour un choix des technologies de fabrication) le résultat suivant:

- Les défaillances dans les transistors sont les plus probables.
- Les défaillances dans les interconnexions sont les moins probables. D'autant que dans les PLA, elles peuvent être minimisées.

Les données extraites du rapport de [BAN82] justifient et confirment ce résultat:

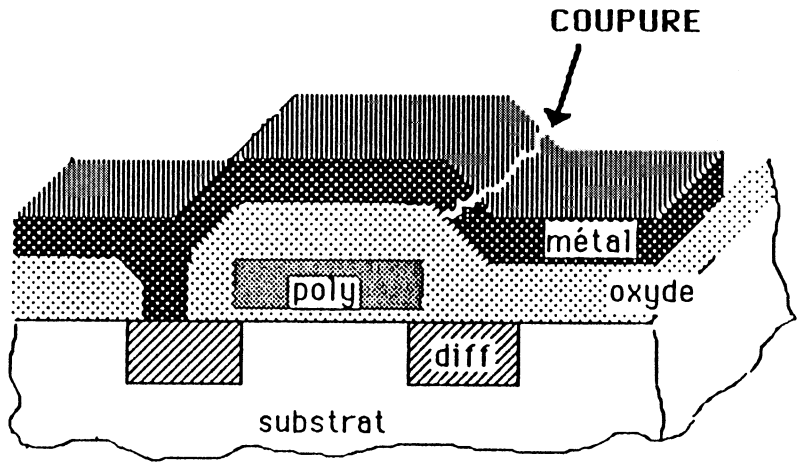


Figure II.3- Coupure de la ligne Aluminum sur une marche de Poly.

Le rapport de [BAN82] classifie les défaillances des composants (transistors et interconnexions) résultant des défauts physiques, en

trois classes par ordre décroissant de probabilité (les probabilités des défaillances dans la classe i sont plus importantes que les probabilités des défaillances dans la classe i+1):

Classe 1 (très probable)

a- transistor

court-circuit (grille/drain et grille/source)

b- interconnexion

court-circuit entre lignes de diffusion

Classe 2 (moyennement probable)

a- transistor

drain ouvert - source ouverte

b- interconnexion

coupure des lignes Aluminum sur une marche de poly, figure II.3.

Classe 3 (peu probable)

a- transistor

court-circuit grille/substrat - grille flottante

b- interconnexion

court-circuit entre lignes d'Aluminum

En fait, dans le cas des PLA, nous pouvons négliger la classe 1b par rapport à la classe 1a (court-circuit entre les lignes de diffusion), car entre ces lignes sont intercalées des lignes de Poly.

Résultat:

En résumé, à partir de l'étude des mécanismes de défaillance et de la distribution des types de pannes résultants de [CAS76] [SCH78] [RAC], nous avons montré le lien entre les défauts physiques et leurs manifestations électriques et logiques, et nous obtenons un résultat qui converge avec le résultat de [MAN84]. Ce résultat est le suivant:

- Les défaillances dans les transistors sont les plus probables.
- Les défaillances dans les interconnexions sont les moins probables.

Ce résultat, qui va être pris en compte comme facteur important dans la structure reconfigurable des PLA, ouvre la voie à la conception de circuits à base de PLA reconfigurables ; en remplaçant les transistors défaillants par des transistors sains d'une part, et d'autre part, en étudiant les géométries des lignes des interconnexions (par contraintes sur les régles de dessin) afin de réduire la probabilité des pannes d'interconnexions.

II- Test de PLA CMOS

Différents auteurs [DEC85, LOT85, ABR82, OST79, SMI79, CHA78] ont déjà proposé des algorithmes de test qui utilisent pratiquement le même modèle de panne, basé sur une analyse des PLA NMOS. Ce modèle de panne comporte trois classes de pannes:

a- Stuck-at-fault: collage à 0 ou à 1 d'une ligne (ou plus). Cette panne résulte physiquement d'une coupure d'une ligne ou d'un court-circuit avec la masse ou bien, de transistors manquants dans les matrices (ET-OU).

b- Short-faults: court-circuit entre deux lignes voisines dans le PLA tel que (métal/métal, diffusion/diffusion, ou métal/diffusion ...).

c- Extra/missing crosspoints defects: présence ou absence d'un, ou plusieurs transistors dans les matrices.

Ce modèle de panne est applicable pour la structure classique de PLA NMOS. Il devient insuffisant pour couvrir les pannes qui peuvent exister dans un PLA optimisé topologiquement [CHU83]. Cette technique d'optimisation modifie la structure topologique de PLA pour avoir plusieurs lignes d'entrées/sorties sur la même colonne. Le modèle de pannes doit suivre cette évolution topologique, afin de couvrir d'autres pannes résultant, par exemple, du problème de routage dans le PLA optimisé qui assure la connexion entre les entrées (sorties) et les blocs voisins. En effet, des entrées/sorties normalement non voisines,

peuvent devenir voisines à cause du type de routage. Il en résulte une possibilité de court-circuit entre n'importe quelles lignes d'entrées/sorties en métal, polysilicium ou diffusion. Ces pannes doivent être ajoutées dans la classe "short-fault" pour assurer la détection des pannes dans les PLA optimisés.

Aux trois classes de pannes citées auparavant, s'ajoutent d'autres classes de pannes qui se produisent dans la technologie CMOS. Le modèle classique de pannes ne peut pas couvrir ces classes. Récemment WADSACK [WAD78], TIMOC [TIM83] et EL-ZIQ [ELZ81] ont développé des modèles de pannes qu'on appelle stuck-open (SOP) et stuck-on (SON).

Ce modèle de pannes SOP/SON ne peut pas être testé par le test combinatoire conventionnel (ou logique). Ceci est dû à l'effet de mémorisation qui nécessite un test séquentiel d'un circuit combinatoire réalisé par des portes CMOS. Des procédures de test sont proposées pour tester le SOP par [ELZ81, CHA83] et pour tester le SOP/SON par [CHI83, OKL84] dans les portes CMOS en général. L'étude de ce problème de test de panne SOP/SON nécessite d'examiner ce modèle.

Dans notre cas, nous avons un PLA CMOS à précharge qui est un circuit séquentiel ou pseudo-séquentiel. Alors le modèle de test séquentiel d'un circuit CMOS combinatoire n'est pas applicable pour tester un circuit séquentiel en CMOS.

Vu la structure régulière de PLA CMOS (NOR-NOR à précharge), figure

II.5, le problème se décompose en deux parties:

Premièrement: les matrices, qui sont des réseaux en NMOS, peuvent être toujours testées par les algorithmes cités auparavant, en couvrant le modèle de panne de PLA NMOS.

Deuxièmement: nous allons étudier le modèle de pannes SOP/SON. Ceci concerne le test des transistors P qui tirent à l'alimentation et des transistors N qui tirent à la masse (figure II.5), afin de déterminer une méthode de test qui complète le test pour les PLA CMOS.

a- **Stuck-open SOP**

Un transistor SOP correspond à un circuit ouvert entre son drain et sa source. La panne SOP se manifeste dans le circuit logique CMOS comme panne séquentielle due à l'effet de la mémorisation de niveau.

* Chandramouli [CHA83] a analysé l'effet SOP, dans un circuit combinatoire, et développé une procédure de test, puisée sur la méthode classique de chemin sensible "path sensitizing".
Ceci nécessite une étape d'initialisation du circuit dans un état connu, suivie par une autre étape d'activation du chemin sensible pour manifester l'erreur [HEN64].
La procédure de test de Chandramouli aboutit à un test séquentiel d'un circuit combinatoire.

* Dans notre cas de PLA CMOS NOR-NOR à précharge, cette méthode ne peut pas être appliquée pour tester le SOP. En effet les PLA CMOS à précharge sont des circuits séquentiels. Par contre, nous allons garder ce principe (initialisation puis activation du chemin sensible) pour tester les pannes séquentielles en testant les pannes de collage en séquence. Ceci nous amène à tester le SOP dans les PLA CMOS à précharge.

* Les PLA à précharge comportent deux phases de fonctionnement plus une inter-phase, figure II.5:

La première est la phase de précharge ($\varphi 1$) qui va être définie comme l'état d'initialisation ou connu (niveau logique 1).
E: représente un ensemble d'entrées du PLA.
Se: est une sortie de la matrice ET qui représente un mônome en fonction de I.

La deuxième est la phase d'activation ou de décharge ($\varphi 2$) qui va être définie comme l'état d'activation pour détecter la panne (SOP).
L'inter-phase (φ) assure la décharge de la matrice ET avant le début de décharge de la matrice OU, pendant laquelle H1 passe à 1 et H2=0.

Pendant la phase de précharge (H1=0, H2=0) les transistors P sont conducteurs et les transistors N sont bloqués. Les monômes (sorties de la matrice ET) et les sorties (sorties de la matrice OU) sont à 1.
Pendant la phase de décharge, on a (H1=1, H2=0) pendant l'inter-phase (φ) puis (H1=1, H2=1), les transistors N sont conducteurs

et les transistors P deviennent bloqués.
Des monômes restent à 1, si toutes les entrées correspondantes sont à 0; et des monômes passent à 0, si au moins une parmi toutes les entrées correspondantes est à 1.
Des sorties restent à 1, si tous les monômes correspondants sont à 0; et des sorties passent à 0, si au moins un parmi tous les monômes correspondants est à 1.

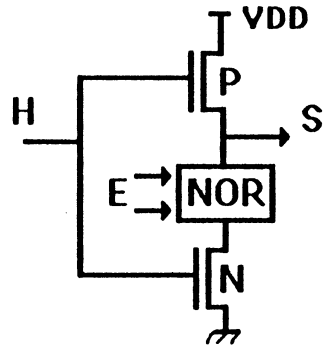
Notre méthode de test est basée sur l'initialisation pendant la phase de précharge et sur l'activation d'un chemin sensible pendant la phase de décharge du PLA, car les sorties de PLA sont uniquement observables pendant la phase de décharge.
La figure II.5 montre un exemple de test de SOP (et stuck-on fault SON) pour un PLA CMOS NOR-NOR à précharge.

Tout d'abord, nous allons traiter le cas d'une simple porte NOR (figure II.4):

- Pendant la phase de précharge, la sortie de la porte NOR est au niveau 1 (le transistor P étant conducteur). Si le transistor P est SOP, la sortie de la porte NOR est à 0 (car l'état initial à la mise sous tension est 0).
La porte NOR représente soit la matrice ET d'un PLA, soit la matrice OU. Les entrées d'une porte NOR dans la matrice ET sont un ensemble accessible des entrées du PLA. Les entrées d'une porte NOR dans la matrice OU sont un ensemble de monômes du PLA.

- Pendant la phase d'évaluation, le transistor N doit être conducteur et le transistor P devient bloqué.

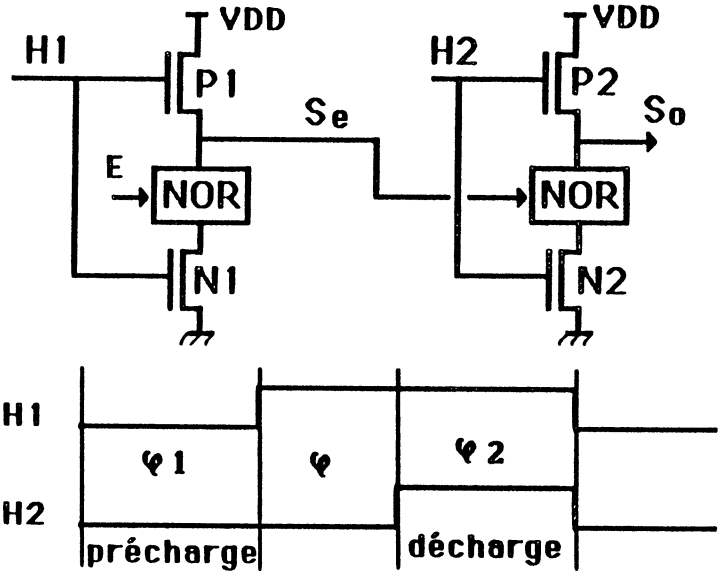
Si la porte NOR ou la matrice est opaque (c'est-à-dire toutes ses entrées E sont à 0), sa sortie doit rester à 1. Si le transistor P est SOP, la sortie reste à 0 (sortie fausse Sf). Donc, le SOP du transistor P d'une porte NOR implique un collage à 0 de la sortie. Par conséquent, on peut tester le SOP du transistor P d'une porte NOR en testant le collage à 0 (s-a-0) de sa sortie (à condition que E=0).



H	E	S	Sf	SOP	s-a-f
0	0	1	0	(précharge ou initialisation)	
1	0	1	0	P S	s-a-"0"
0	1	1	1	(précharge ou initialisation)	
1	1	0	1	N S	s-à-"1"

Figure II.4- Porte NOR d'un PLA CMOS à précharge.

Si la porte NOR ou la matrice est transparente (c'est-à-dire une au moins parmi toutes ses entrées E est à 1), sa sortie doit passer à 0. Si le transistor N est SOP, la sortie reste à 1. Donc, le SOP du transistor P d'une porte NOR implique un collage à 1 de sa sortie. Par conséquent, on peut tester le SOP du transistor P d'une porte NOR en testant le collage à 1 (s-a-1) de sa sortie.



H1	H2	E	Se	Sef	So	Sof	SOP	s-a-f	SON
0	0	0	1	0	1	1	(précharge ou initialisation)		
1	1	0	1	0	0	1	P1, N2	Se"0", So"1"	N1, P2
0	0	1	1	1	1	0	(précharge ou initialisation)		
1	1	1	0	1	1	0	N1, P2	Se"1", So"0"	P1, N2

Figure II.5- Test de SOP/SON, d'un PLA CMOS NOR-NOR à précharge, par un test de s-a-f en séquence.

Dans le cas de PLA CMOS à précharge, deux portes NOR sont en séquence. Notre méthode de test de SOP qui est appliquée sur une porte NOR, va être prolongée pour le PLA. En effet, les SOP des transistors P et N vont être testés en testant les collages à 0 et à 1 des sorties et des monômes du PLA (étant donné que seules les sorties de la matrice OU sont observables). Pour tester le collage d'un monôme, il faut activer le chemin sensible qui manifeste l'état du monôme à la sortie du PLA (sorties observables). Nous verrons plus tard les conditions à prendre en compte pour détecter le collage d'un monôme (figure II.5):

Pendant la phase de précharge (H1=0, H2=0) les états des sorties sont connus (Se=So=1).

Pendant la phase d'évaluation, tout d'abord (H1=1, H2=1).

Si E=0: (c'est-à-dire toutes les entrées correspondantes au monôme Se sont à 0), la sortie Se doit rester à 1 et So s'évalue de 1--> 0.
Si le transistor N2 est SOP, alors la sortie So est collée à 1. Donc le SOP du N2 peut être testé par le test de collage à 1 de So (So s-à-"1").

Si P1 est SOP, alors la sortie Se (monôme) est collée à 0. Donc le SOP du P1 peut être détecté par le test de collage à 0 de Se (Se s-à-"0"). Pour manifester le SOP de P1 à la sortie So, il faut activer le chemin sensible qui manifeste le test de collage à 0 de Se (Se s-à-"0") à la sortie So. Ceci implique la condition suivante: tous les monômes correspondants à So, autre que Se, doivent être à 0. Alors si Se est s-à-0, la sortie So reste à 1 .

(c'est-à-dire que le SOP de P1 et N2 sont détectés en testant le s-à-1 de So).

(Sef et Sof sont les sorties fausses)

Si E=1: (c'est-à-dire une au moins parmi toutes les entrées correspondantes au monôme Se est à 1), la sortie So doit rester à 1 et Se s'évalue de 1--> 0. En utilisant la même méthodologie que le cas précédant (E=0), le SOP de P2 peut être détecté par le test de collage à 1 de So ; et le SOP de N2 peut être détecté par le test de collage à 1 de Se. Ce dernier peut être observé à la sortie So, à condition que tous les monômes correspondants à So, autre que Se, doivent être à 0. Alors si Se est s-à-1, la sortie So passe à 0.

b- **Stuck-on fault SON**

Un transistor est SON s'il est dans un état de conduction permanent à cause :

- d'un court-circuit entre le drain et la source.
- d'un court-circuit entre la grille et le drain pour le transistor canal N ou entre la grille et la source pour le transistor canal P.
- d'une variation de la tension de seuil qui rend le transistor toujours conducteur par la présence du canal.

Dans le cas d'un inverseur figure II.6, cet état de conduction

permanent d'un transistor SON dépend du rapport des résistances du transistor canal P (Rp) et celle du transistor canal N (Rn).

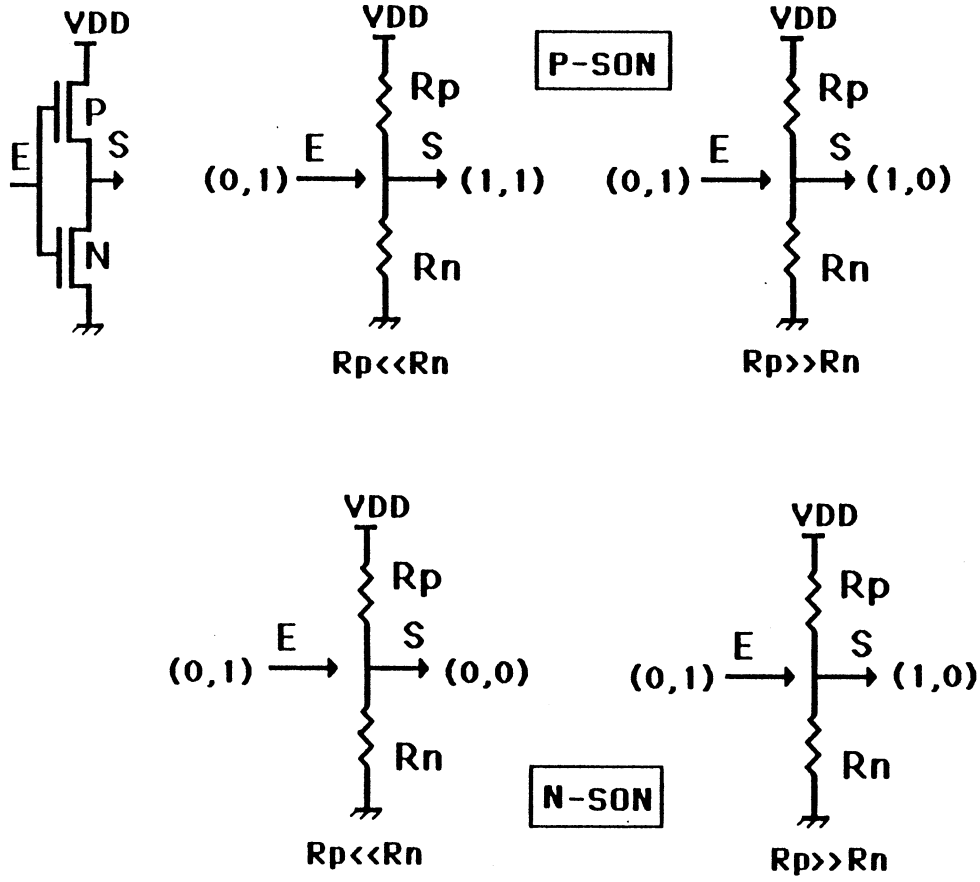


Figure II.6- Les quatre états de SON possibles dans un inverseur CMOS.

Résis-tance	Tran-sistor	Entrée	Sortie en panne	Sortie correcte	Commentaire
Rp << Rn	P SON	0	1	1	SON détectable par s-a-1
		1	1	0	
	N SON	0	1	1	SON non détectable
		1	0	0	
Rp >> Rn	P SON	0	1	1	SON non détectable
		1	0	0	
	N SON	0	0	1	SON détectable par s-a-0
		1	0	0	

Tableau II.3- Les quatre états de SON.

1°- Premièrement, si le transistor P est SON. Deux cas sont possibles :

- si Rp<<Rn, dans ce cas la sortie est en permanence collée à 1 (s-a-1) quelque soit l'entrée. Cet état peut être détectable en testant le s-a-1 de la sortie.

- si Rp>>Rn, dans ce cas la sortie dépend de l'entrée. Cet état est indétectable par un test logique.

2°- Deuxièmement, si le transistor N est SON. Deux états sont aussi possibles:

- si Rp<<Rn, cet état est indétectable par un test logique.

- si Rp>>Rn, la sortie est en permanence collée à 0 (s-a-0). Cet état peut être détectable en testant le s-a-0 de la sortie.

En résumé, le tableau II.3 illustre les quatre états SON d'un inverseur.

Deux états sont détectables en testant les pannes de collage. Les deux autres sont indétectables par un test logique. Mais, ils peuvent être détectés par d'autres méthodes analogique, telleque les méthodes de test temporelle (un transistor SON, par son effet résistif, peut augmenter le temps de réponse d'un circuit). En effet, les transistors SON modifient le temps de propagation du circuit. Ce temps peut devenir important si la résistance équivalente du transistor SON est grande. Ceci est détectable en étudiant le signal de sortie en fonction du temps.

Pour de raisons de rapidité de test, seuls les deux états détectables de SON sont considérés. Le test de panne de collage à 1 (s-a-1) de la sortie d'une porte détecte le SON du transistor P. Le test de panne de collage à 0 (s-a-0) détecte le SON du transistor N.

Nous allons appliquer cette méthode pour tester le SON dans les PLA CMOS à précharge. On considére la même structure du PLA NOR-NOR, figure II.5:

Pendant la phase de précharge (H1=0, H2=0) les états des sorties sont connus (Se=So=1).

Pendant la phase d'évaluation (H1=1, H2=1) :

Si E=0, le SON du transistors P2 peut être détecté en testant le s-a-1 de

So. Le SON du transistor N1 peut être détecté en testant le s-a-0 de Se.

Si E=1, le SON des transistors P1 et N2 peut être détecté en testant respectivement le s-a-1 de Se et le s-a-0 de So.

Conclusion

Notre étude du test de PLA concerne les deux domaines suivants: défauts de fabrication et test.

* En fabrication, l'étude des mécanismes de défaillance, de la distribution des types de pannes, du lien entre les pannes physiques et leurs manifestations électrique et logique, a permis d'obtenir le résultat suivant (pour un choix des technologies de fabrication), qui va être pris en compte pour la structure des PLA reconfigurables:

- Les défaillances dans les transistors sont les plus probables.
- Les défaillances dans les interconnexions sont les moins probables.

* Pour le test, le PLA CMOS à précharge est un circuit séquentiel. Ceci nécessite un test séquentiel dû à l'effet de mémorisation (SOP, SON). Ce problème peut être traité en deux parties: l'une sert à tester les matrices (ET, OU), l'autre sert à tester les transistors qui tirent à l'alimentation (P1, P2) et les transistors qui tirent à la masse (N1, N2).

En effet, les pannes de PLA en technologie CMOS peuvent être détectées de la manière suivante:

- Les pannes dans les matrices ET-OU peuvent être détectées de la même manière que dans le PLA NMOS (étant donné que l'observabilité est limité à la sortie de la matrice OU).

- Les pannes SOP/SON, pour les transistors (P1, P2 et N1, N2), peuvent être détectées par une séquence de test de panne de collage (s-a-f) en initialisant pendant la phase de précharge et en activant le chemin sensible pendant la phase de décharge de PLA CMOS à précharge (en signalant que la difficulté d'un algorithme de test est un problème à étudier).

CHAPITRE III

PLA RECONFIGURABLES

Introduction

Nous nous interessons ici aux PLA tolérant des défauts de fabrication, afin d'accroitre le rendement de fabrication de tels dispositifs. Cette tolérance est obtenue par des techniques de "reconfiguration". La reconfiguration consiste à remplacer des éléments défectueux (lignes ou colonnes) par des élements sains en additionnant des lignes supplémentaires.

Différentes méthodes ont été utilisées dans ce domaine telles que la duplication et la redondance dans les RAM [KOK81, SMI81, MIN82].

Une reconfiguration efficace doit prendre en compte la distribution des types de pannes et donc s'appuyer sur des techniques de test et de diagnostic efficace ainsi que sur des éléments de commutation (interrupteurs) adéquats.

Nous allons, tout d'abord, étudier les caractéristiques des éléments et leurs regroupements dans les PLA. Ces derniers dépendent de la régularité du PLA et de ses particularités. Ensuite, nous allons étudier les éléments de la circuiterie nécessaires au circuit.

Puis, nous verrons que le résultat de la distribution des types de défaillance (les défaillances dans les transistors sont dominantes par rapport à celles dans les interconnexions pour un choix des technologies de fabrication) va influer sur la structure des unités en réserve.

Enfin, nous présenterons les interrupteurs (switches), tels que les fusibles à laser [KEN80, MET83, SMI81, STE85] et les transistors à grille flottante [SHA83], et leurs compatibilités avec le coût de fabrication des circuits intégrés.

Nous étudierons enfin une autre possibilité de reconfiguration de PLA par commande de transistors.

I- Méthode

La structure d'un PLA est un facteur important pour sa reconfiguration. En effet, sa régularité donne la possibilité de regrouper des éléments semblables dans un ensemble. Ceci facilite le problème d'échange d'un élément défectueu avec un élément sain en échangeant des ensembles en lieu et place de l'échange des éléments. Cela facilite aussi le test pour localiser une défaillance dans un ensemble par rapport à un élément.

La structure considérée, dans notre cas, est le PLA CMOS [DAN85] reconfigurable. Cette structure régulière de PLA comporte trois groupes différents qui sont:

a- les entrées

b- les monômes

c- les sorties

Chaque groupe est constitué d'éléments semblables (mais pas identiques par leurs répartitions), ce qui rend le problème de la reconfiguration à priorie simple à réaliser.

Par contre l'hétérogéneité de répartition des éléments sur les lignes pose de problème d'échange des lignes. En effet, chaque groupe devra être traité indépendamment des autres. La reconfiguration de chaque groupe nécessite la définition des lignes à reconfigurer. Nous appellerons ces lignes : lignes services.

A partir de ces lignes services, notre méthode consiste à ajouter des lignes réserves, afin de remplacer une ligne contenant des défauts (ligne service) par une ligne saine (ligne réserve).
Autrement dit, trois sortes des lignes réserves correspondent aux trois groupes des lignes services (entrées, monômes et sorties) de PLA, figure III.1.

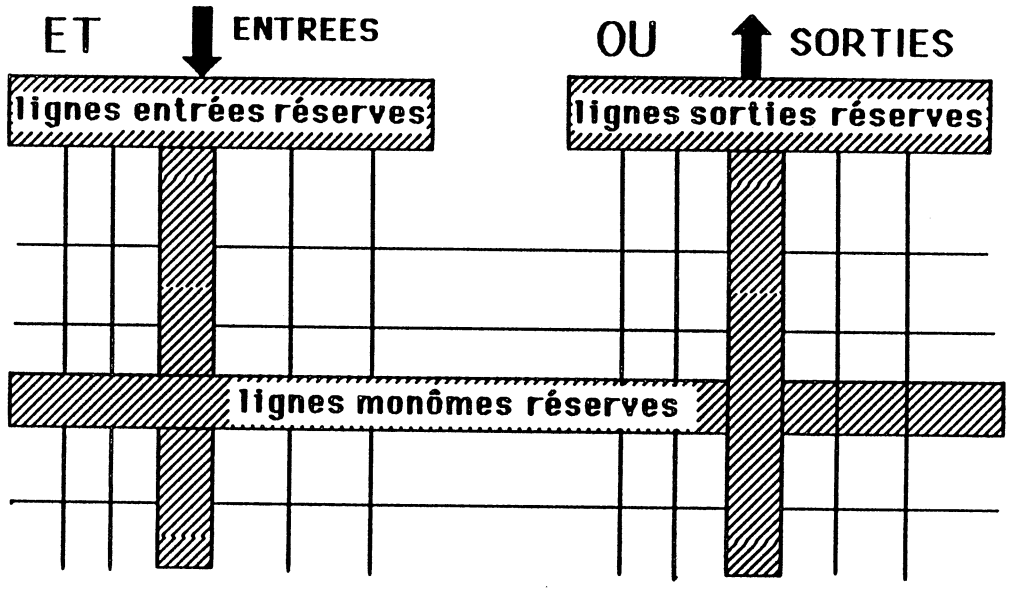


Figure III.1- Structure de PLA reconfigurable.

Les entrées et les sorties du PLA restent inchangées réellement, vues de l'extérieur ou de l'interconnexion avec les autres blocs, avant et après la phase de la reconfiguration.

Cette phase de reconfiguration est réalisée par des étapes technologiques supplémentaires au niveau interne du PLA après avoir détecté et localisé des défauts dans les lignes services pendant la phase de test.

Pour une ligne service contenant des défauts, on remplace cette ligne par une autre en réserve, figure III.2, en changeant le chemin de la ligne considérée.

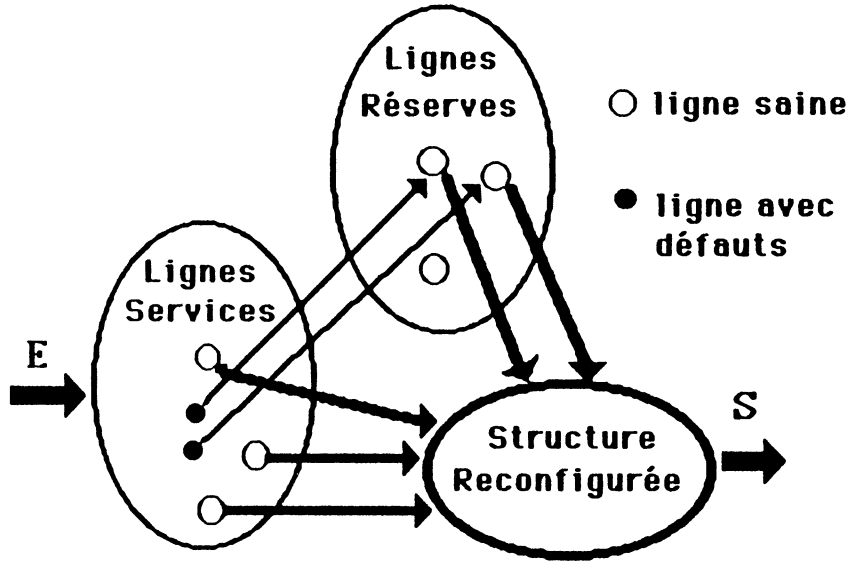


Figure III.2- Méthode de la reconfiguration.

Ce changement de chemin est basé sur la programmation des interrupteurs qui consiste à :

- supprimer des éléments
- couper des lignes
- activer des éléments
- désactiver des éléments

Les moyens de réaliser cette structure (programmation des interrupteurs) sont donnés par des circuiteries ajoutées à base d'éléments programmables, tels que les fusibles, les antifusibles et les transistors bloqués ou conducteurs (comme par exemple les transistors à grille flottante).

Cette technique de changement du chemin dépend de la technologie utilisée et des interrupteurs qui peuvent être réalisés. Aussi, une étude de comparaison des interrupteurs est nécessaire par rapport à d'autres

paramètres tels que :

le surcoût de fabrication, la compatibilité avec la technologie de fabrication et la surface occupée.

Nous allons, tout d'abord, étudier la structure de chaque groupe (entrées, monômes, sorties) en vue de définir les lignes (services et réserves) correspondantes.

Ensuite, nous allons étudier les circuiteries qui réalisent l'échange des lignes services contenant des défauts avec des lignes réserves.

1.1- Reconfiguration des lignes d'entrées

Chaque entrée d'un PLA ou chaque ligne d'entrée (généralement en silicium polycristallin) comporte soit des grilles de transistors soit un inverseur suivi par des grilles qu'on appelle respectivement l'entrée et l'entrée complémentée d'un PLA.

Si on tient compte des inverseurs dans les lignes, une ligne service peut être définie soit comme ligne d'entrée soit comme ligne d'entrée complémentée, figure III.3. Pour le cas où on ne prend pas en compte les inverseurs, une ligne réserve représente une simple ligne d'entrée.

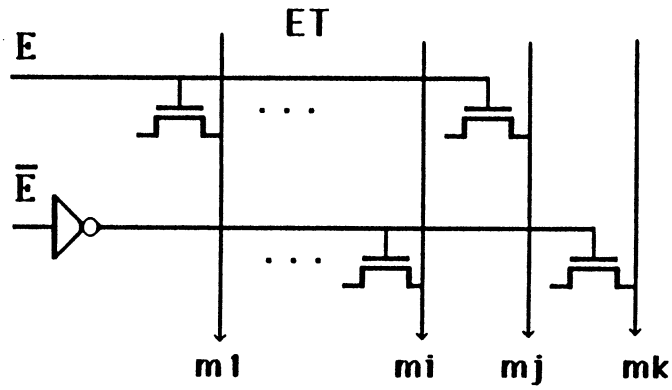


Figure III.3- lignes des entrées services.

Donc dans le cas où on intégre un inverseur dans la ligne, nous définissons deux types de lignes d'entrées services:

- ligne d'entrée service LES
- ligne d'entrée service complémentée LESC

Cette définition des lignes d'entrées services permet de réaliser des lignes d'entrées réserves. La ligne réserve (LER) doit couvrir toutes les lignes services, afin de pouvoir échanger les lignes contenant des défauts avec des lignes saines. Donc, le nombre des transistors d'une ligne réserve est plus grand ou égal au nombre des transistors d'une ligne service. Ainsi à chaque niveau où se trouve un transistor dans la LES, il doit se trouver un transistor dans la LER. Les lignes d'entrées réserves sont aussi deux types :

- ligne d'entrée réserve LER
- ligne d'entrée réserve complémentée LERC

L'échange des lignes s'effectue avec le même type de lignes.

La méthode d'échange entre deux lignes d'entrées (service et réserve) peut être réalisée par la programmation des interrupteurs, figure III.4.

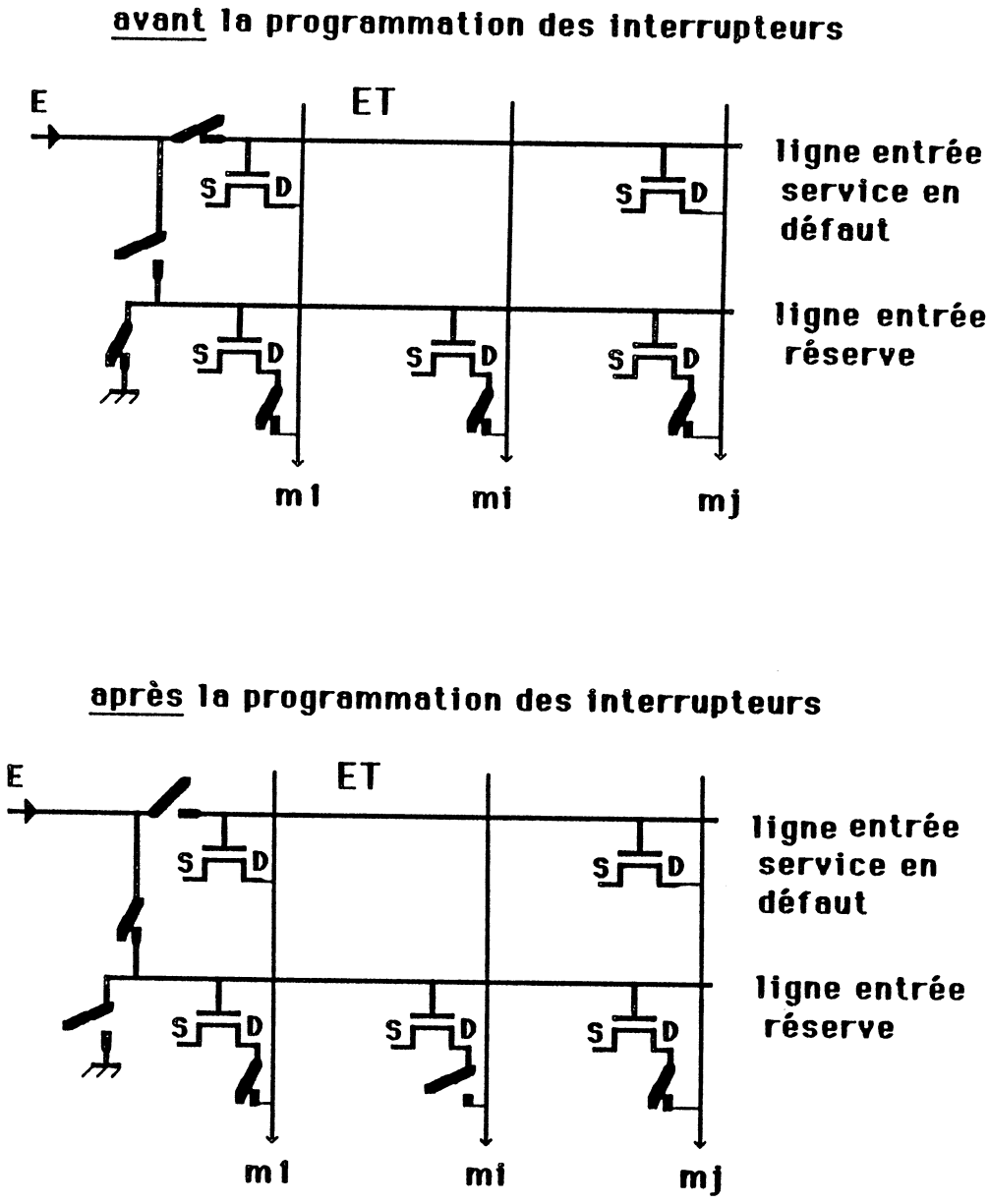


Figure III.4- Structure d'entrées reconfigurables

En vue de réaliser l'échange s'il y a des défauts, deux ensembles d'interrupteurs sont utilisés: les interrupteurs d'échange entre lignes et ceux d'identification des lignes.

Les interrupteurs d'échange sont :

- l'interrupteur qui se trouve entre les deux lignes (LES, LER). Il doit être ouvert pendant la phase de test et fermé pendant la phase de la reconfiguration pour assurer l'échange.
- L'interrupteur qui se trouve entre la ligne LER et la masse, est fermé pendant la phase de test, et ouvert pendant la phase de la reconfiguration. Ceci assure la neutralité de la ligne (au niveau logique 0) de réserve (LER), par rapport aux évaluations des monômes, pendant la phase de test.
- Pendant la phase de la reconfiguration, l'ouverture de l'interrupteur (fermé) qui se trouve à l'entrée de la ligne LES assure sa neutralité.

Les interrupteurs d'identification des lignes (LES, LER) assurent le placement des transistors dans la ligne LER aux mêmes niveaux que dans la ligne LES à reconfigurer. Ils sont placés entre les lignes de sorties (monômes) et les drains de transistors, pour réduire la capacité de sortie après la reconfiguration. Ces interrupteurs peuvent être pendant la phase de test: soit ouverts, soit fermés. S'ils sont ouverts, on ferme ceux qui correspondent aux mêmes niveaux de placement des transistors; sinon on ouvre les autres, figure III.4.

Circuiterie de LER

La circuiterie des lignes d'entrées réserves réalise la reconfiguration des lignes d'entrées du PLA. Elle dépend des interrupteurs utilisés et réalisables en tenant compte de deux facteurs essentiels: la compatibilité avec le coût de fabrication et la surface de silicium utilisée.

Entre les lignes LES et LER se trouvent des interrupteurs ouverts qui

sont programmables pendant la phase de la reconfiguration.

Un exemple de la réalisation électrique de la structure d'entrées reconfigurables est donnée sur la figure III.5 pour un groupe des 4 entrées.

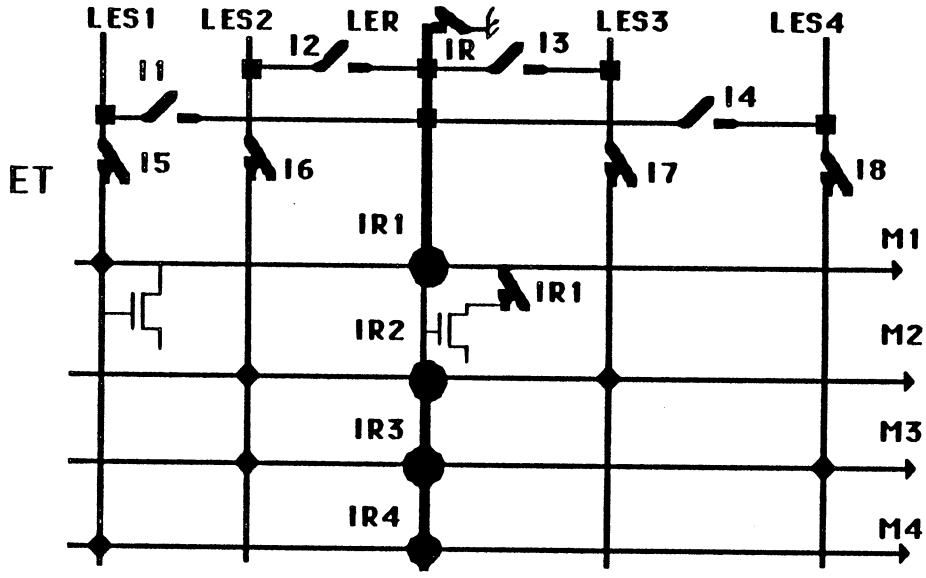


Figure III.5- Circuiterie des entrées de PLA reconfigurables.

La figure III.5 montre la structure des entrées reconfigurables pendant la phase de test. Si la ligne LES1 contient des défauts, la reconfiguration s'établit par l'échange de la ligne LES1 avec LER. Ceci nécessite la programmation d'interrupteurs suivante:

- (I5, IR) vont être ouverts et I1 fermé pour assurer l'échange.
- (IR2, IR3) vont être ouverts pour identifier les lignes (LES1, LER)

I.2- Reconfiguration des lignes de monômes

Les monômes sont du même type de structure électrique. Chaque ligne monôme (généralement en aluminium dans la matrice ET et en silicium polycristallin dans la matrice OU) comporte des drains de transistors dans la matrice ET et des grilles dans la matrice OU, figure III.6.

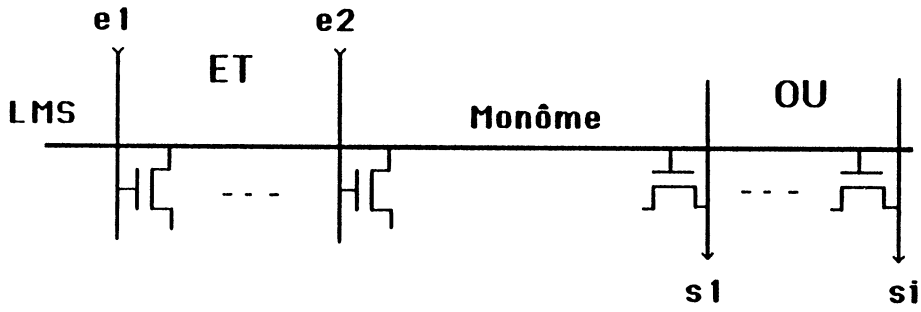


Figure III.6- Structure d'une ligne monôme.

Les monômes peuvent comporter des amplificateurs entre les deux matrices ou un inverseur. Alors nous pouvons définir:

- soit un seul type de ligne de monôme, si chaque ligne monôme (dans les deux matrices ET-OU) représente une ligne monôme service (LMS).
- soit deux types de lignes de monômes, si on distingue entre la partie de la matrice ET et celle de la matrice OU.

Nous allons traiter le cas d'un seul type de ligne de monôme service (LMS) qui simplifie la circuiterie d'échange entre lignes de monômes. A partir des LMS données, on peut définir des lignes monômes réserves (LMR).

Une ligne LMR doit couvrir une LMS à reconfigurer. Le cas le plus

défavorable est lorsque la LMR est pleine, c'est-à-dire lorsqu'on met un transistor à chaque pas dans les deux matrices.

Pendant la phase de test les LMR sont inactives, car elles sont à un niveau logique "0". Ceci est dû au fait que sur chaque LMR il y a au moins une entrée avec son complément ($(x+x')' = 0$).

L'échange d'une LMS contenant des défauts avec une LMR s'obtient par la programmation de deux groupes d'interrupteurs: les interrupteurs d'échange entre lignes de monômes et ceux d'identification, figure III.7.

Les interrupteurs d'échange sont uniquement sur les LMS. Ils sont fermés normalement. Ils servent à neutraliser les lignes monômes contenant des défauts pendant la phase de la reconfiguration.

Les interrupteurs d'identification se trouvent entre les lignes de sorties et les drains de transistors. Ils assurent l'identification entre les monômes à reconfigurer. Ils peuvent être au départ soit ouverts soit fermés.

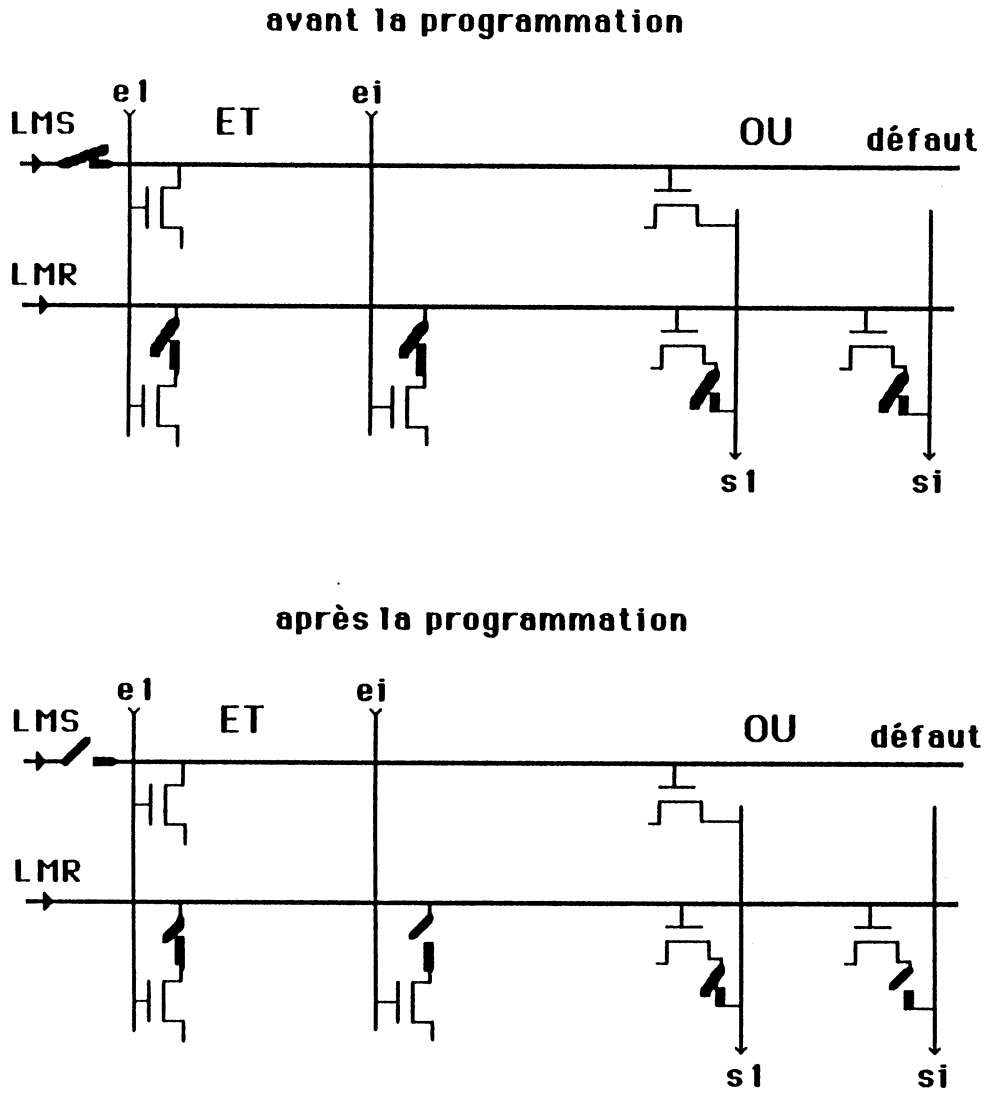


Figure III.7- Structure des monômes reconfigurables.

Circuiterie de LMR

A partir des lignes réserves de monômes (LMR). la circuiterie de reconfiguration assure l'échange d'une LMS contenant des défauts avec une LMR. Le particularité d'une ligne LMR vient du fait qu'elle peut être échangée avec n'importe quelle LMS, sans programmation d'interrupteurs entre les lignes de monômes. Le seul problème est celui de l'identification de LMR avec LMS et la neutralité de LMS.

Un exemple de la réalisation électrique de la structure des lignes de monômes reconfigurables est donné sur la figure III.8 pour un groupe des quatre monômes.

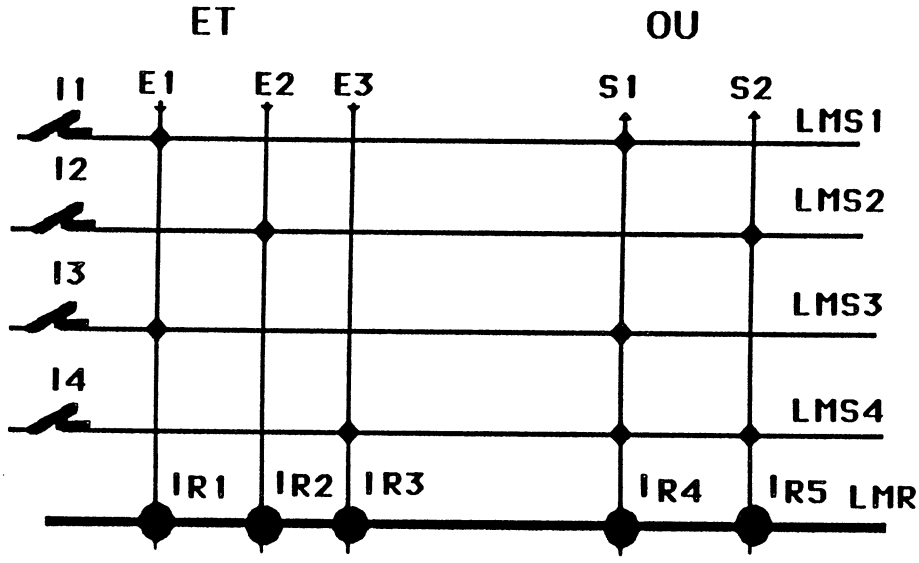


Figure III.8- Circuiterie des monômes reconfigurables.

Nous supposons que la phase de test localise des pannes dans LMS3. L'échange de LMS3 avec LMR peut être réalisé par la programmation d'interrupteurs suivants :

- On ouvre I3 pour rendre le monôme PMS3 neutre.
- On identifie les deux lignes LMR et LMS3 en ouvrant les interrupteurs (IR2, IR3, IR5).

I.3- Reconfiguration des lignes de sorties

La structure des lignes de sorties de PLA semble à la structure des lignes de monômes du fait qu'un seul type de ligne peut être défini.

Chaque ligne de sortie (généralement en aluminium) comporte des drains de transistors avec un inverseur à la sortie de la matrice OU, figure III.9. Alors une ligne de sortie peut être définie comme ligne de service (LSS).

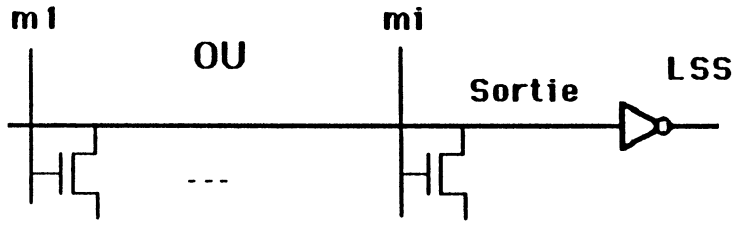


Figure III.9- Structure d'une ligne de sortie.

A partir de ceci, une ligne de sortie réserve (LSR) est définie en ajoutant une ligne de sortie réserve qui couvre les LSS.

Pour obtenir les mêmes lignes des sorties services après la reconfiguration, la même circuiterie utilisée pour les entrées peut être utilisée pour les sorties, en ajoutant les deux groupes d'interrupteurs, figure III.10 :

- Les interrupteurs d'échange
- Les interrupteurs d'identification des lignes de sorties.

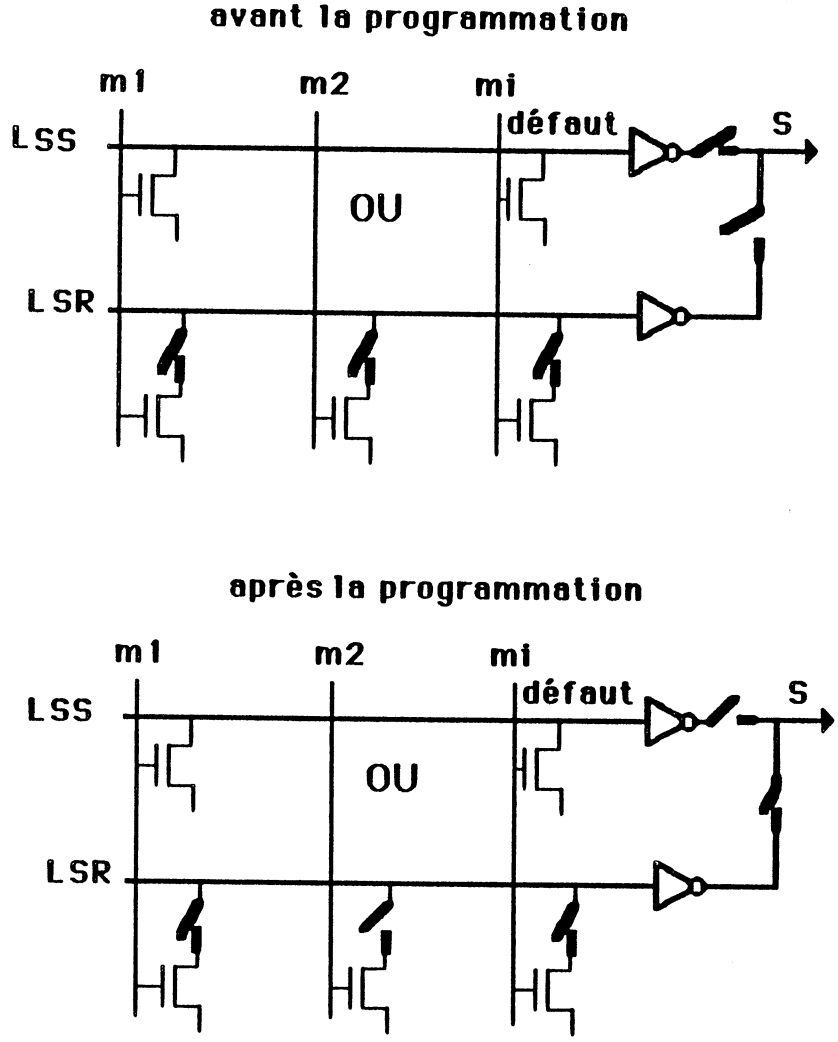


Figure III.10- Structure de sorties reconfigurables.

Circuiterie de LSR

La circuiterie de reconfiguration assure l'échange d'une LSS contenant des défauts, avec une LSR. Un interrupteur ouvert entre les deux lignes (LSS, LSR) est nécessaire pendant la phase de test. Cet interrupteur sera fermé pendant la phase de la reconfiguration.

Un exemple de la réalisation électrique de la structure des sorties

reconfigurables est donné sur la figure III.11 pour un groupe de 4 lignes de sorties. Nous supposons que la phase de test localise des pannes dans LSS3. L'échange de LSS3 avec LSR peut être réalisé par la programmation des interrupteurs suivants:

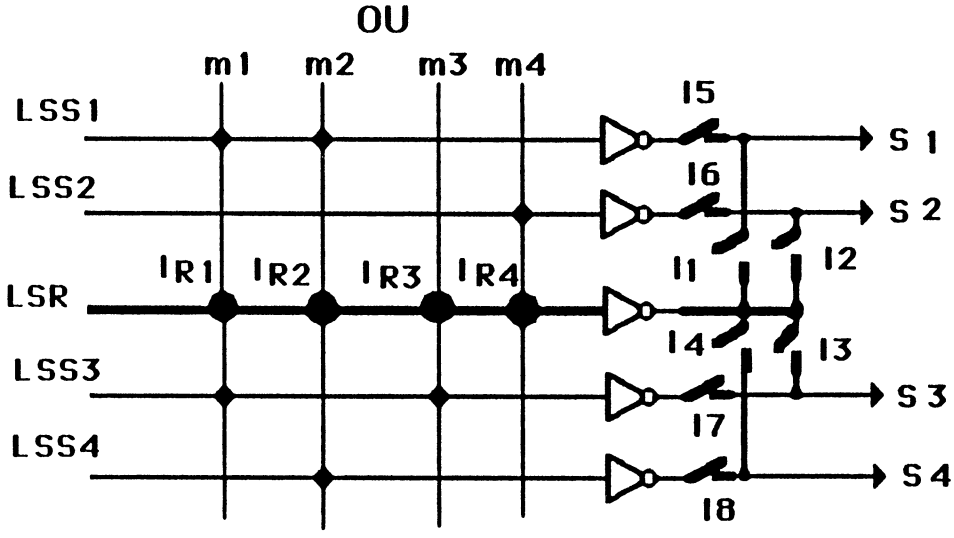


Figure III.11- Circuiterie des sorties reconfigurables.

- On ferme I3 et on ouvre I7 pour assurer l'échange.
- On ouvre les interrupteurs (IR2 , IR4) pour assurer l'identification de deux lignes LSS3 et LSR.

II- Performance de la structure reconfigurable

Le coût des étapes technologiques ajoutées pour cette méthode de reconfiguration (programmation des interrupteurs) nécessite une étude approfondie de la réalisation des interrupteurs, afin de tenir compte du surcoût de fabrication des C.I.

L'importance du nombre d'interrupteurs utilisés dans cette méthode pose le problème d'incompatibilité avec le coût de fabrication, et le problème de la grande surface de silicium utilisée.

En vue de réduire le nombre d'interrupteurs, nous allons tenir compte du premier facteur (les défauts en fin de fabrication) dans la structure reconfigurable. Ceci est résumé par le résultat suivant (chapitre II):

les défaillances dans les transistors sont dominantes par rapport à celles dans les interconnexions.

A partir de ce résultat, la performance de la structure reconfigurable est basée sur la correction des pannes de transistors. Cette méthode va évaluer les ensembles de transistors dans les PLA à reconfigurer. Nous allons étudier la relation entre la reconfiguration de transistors et celle des lignes définies précédemment.

- Les défaillances des transistors de la matrice ET agissent sur l'évaluation de l'état logique des monômes. Alors, un transistor défaillant (stuck-open: SOP, stuck-on: SON) dans la matrice ET implique que la ligne monôme correspondante est en panne.
Donc, la correction des transistors dans la matrice ET peut être réalisée par la reconfiguration des lignes monômes.

- Les défaillances des transistors dans la matrice OU agissent sur l'évaluation de l'état logique des sorties. Ceci implique qu'une ligne sortie est en panne.
Donc, la correction des transistors dans la matrice OU peut être réalisée par la reconfiguration des lignes sorties. Mais, si le transistor est SOP

dans la matrice OU, la correction peut être réalisée par la reconfiguration des lignes monômes.

A partir de ces deux résultats, nous allons envisager uniquement la structure reconfigurable des monômes et des sorties. Ceci couvre la correction des types de pannes les plus probables.
Alors, nous pouvons éliminer la circuiterie de la reconfiguration des lignes d'entrées. Il en résulte une réduction importante du nombre des interrupteurs.

Les pannes couvertes ou corrigées dans le PLA CMOS, par cette évaluation de la structure reconfigurable, sont les pannes des composants suivants:

- Dans la matrice ET:

  - Les transistors de la matrice.
  - Les transistors de pull-up.
  - La circuiterie du maintien de niveau (accrocheurs).
  - Les lignes d'aluminium de monômes.
  - Les amplificateurs (ou boot-strap) entre les deux matrices.

- Dans la matrice OU:

  - Les transistors de la matrice.
  - Les transistors de pull-up.
  - La circuiterie du maintien de niveau (accrocheurs).
  - Les lignes d'aluminium de sorties.
  - Les inverseurs de sorties

- Les lignes d'entrées silicium polycristallin de la matrice OU.

- Les composants qui ne sont pas reconfigurables sont:

- Les lignes d'entrées polycristallin de la matrice ET. Ceci est dû à la faible probabilité des pannes. Cette probabilité peut être aussi réduite avec l'élargissement des lignes d'entrées de polycristallin.

- Les lignes d'interconnexion (en diffusion) entre les sources des transistors dans les deux matrices. La probabilité des pannes dans ces lignes est faible. Elle diminue avec l'élargissement de la ligne de diffusion.

- Les transistors de couplage à la masse. La probabilité des défauts dans ces transitors est faible à cause de leur grande dimension.
Ces transistors peuvent être reconfigurables en mettant plusieurs transistors sur chaque ligne de diffusion avec des interrupteurs fermés sur les drains. Si un transistor est stuck-on, on ouvre l'interrupteur correspondant.

III- Réalisation des interrupteurs programmables

Les interrupteurs programmables (switches) doivent être capables de réaliser, après les étapes technologiques nécessaires: soit la déconnexion, soit la connexion entre deux lignes.

Différentes techniques de réalisation de ces interrupteurs ont été proposées. Trois catégories sont réalisables, tableau III.1 [SMI82]:

- Des interrupteurs réalisant uniquement la déconnexion.
- Des interrupteurs réalisant uniquement la connexion.
- Des interrupteurs réalisant à la fois: soit la déconnexion, soit la connexion.

	INTERRUPTEURS	AVANT	APRES
déconnexion	- Programmation électrique par transistor de commande d'un fusibles poly.	R = 200Ω	R > 10^7Ω
	- Programmation par laser de fusibles poly.	R =200Ω	R > 10^9Ω
	- Programmation électrique+ laser de fusible poly	R = 200Ω	R > 10^9Ω
connexion	- Programmation électrique oxyde	Diélec. Iso	R=5OO-2KΩ
	- Antifusible (niveau Al)	R > 10^7Ω	R = 10Ω
	- poly - diodes		
	- redistribution de dopage	R > 10^7Ω	R=1O^2-10^3Ω
	- migration Al	R > 10^7Ω	R=10-30Ω
	- Programmation par laser vias		
	niveaux métaux	Diélec. Iso	R = 0.3Ω
	métal --> diffusion	Diélec. Iso	R=10-15Ω
	- métal gap	R > 10^9Ω	R=1-10Ω
	- résistance poly-redistribution de dopage	R > 10^9Ω	R > 10^3Ω
déconnexion + connexion	Grille flottante	VT=-1.2V	VT=-6V

Tableau III.1- Caractéristique des interrupteurs avant et après la programmation.

3.1- Interrupteurs utilisés

La réalisation des PLA CMOS reconfigurables, dans notre cas, à base d'interrupteurs programmables, nécessite l'utilisation d'interrupteurs de connexion et de déconnexion. Nous allons étudier l'utilisation des deux techniques d'interrupteurs suivantes:

- La programmation par laser de fusible poly.
- Le transistor à grille flottante.

a- Fusible poly

La technique de programmation par laser de fusible poly [KEN80, SMI81, MET83, STE85] offre simplement la déconnexion entre deux lignes de poly.

La réalisation d'interrupteurs de connexion (antifusible), est nécessaire pour notre méthode de reconfiguration. Ce problème peut être résolu en additionnant des composants, commandés par un fusible, qui réalise la connexion entre deux lignes.

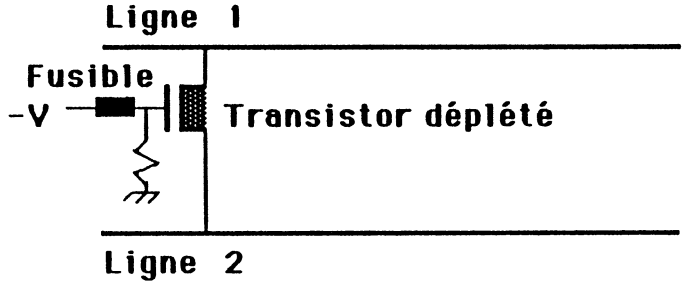


Figure III.12- Connexion entre deux lignes avec un fusible.

L'utilisation d'un transistor déplété peut être une solution. Avant la programmation, le fusible connecte la grille du transistor à une tension négative, figure III.12. Le transistor est bloqué. Après la programmation, le transistor déplété devient conducteur en grillant le fusible (tension de grille est à la masse). Alors une connexion entre le drain et la source (ou entre deux lignes) est réalisée.

Les inconvénients de cette technique sont:

- la grande surface de silicium occupée par un fusible poly électrique-laser (une dizaine de fois la surface d'un transistor) ;
- le temps de programmation pour griller un fusible par laser (temps d'alignement du Laser par rapport au fusible, plus le temps pour le griller (5J)). Ceci présente un coût élévé de programmation ;
- la possibilité d'une nouvelle création du contact après la programmation.

Une coupure de ligne poly (ou métal) et une connection entre deux lignes métal 2, par Laser, sont également possibles [NIC86] ; elles permettent de réaliser la déconnexion et la connexion entre deux lignes. Cependant, cette méthode nécessite de définir et de préparer des zones en poly et en métal. Ceux-ci demande moins de surface de Silicium en respectant les règles de dessin correspondantes.
Les figures III.12 et III.13 montrent l'intégration d'une ligne réserve à base de coupure des lignes de poly dans un PLA reconfigurable. Par contre, les figures III.14 et III.15 montrent l'intégration d'une ligne réserve à base de connexion des lignes de métal 2.

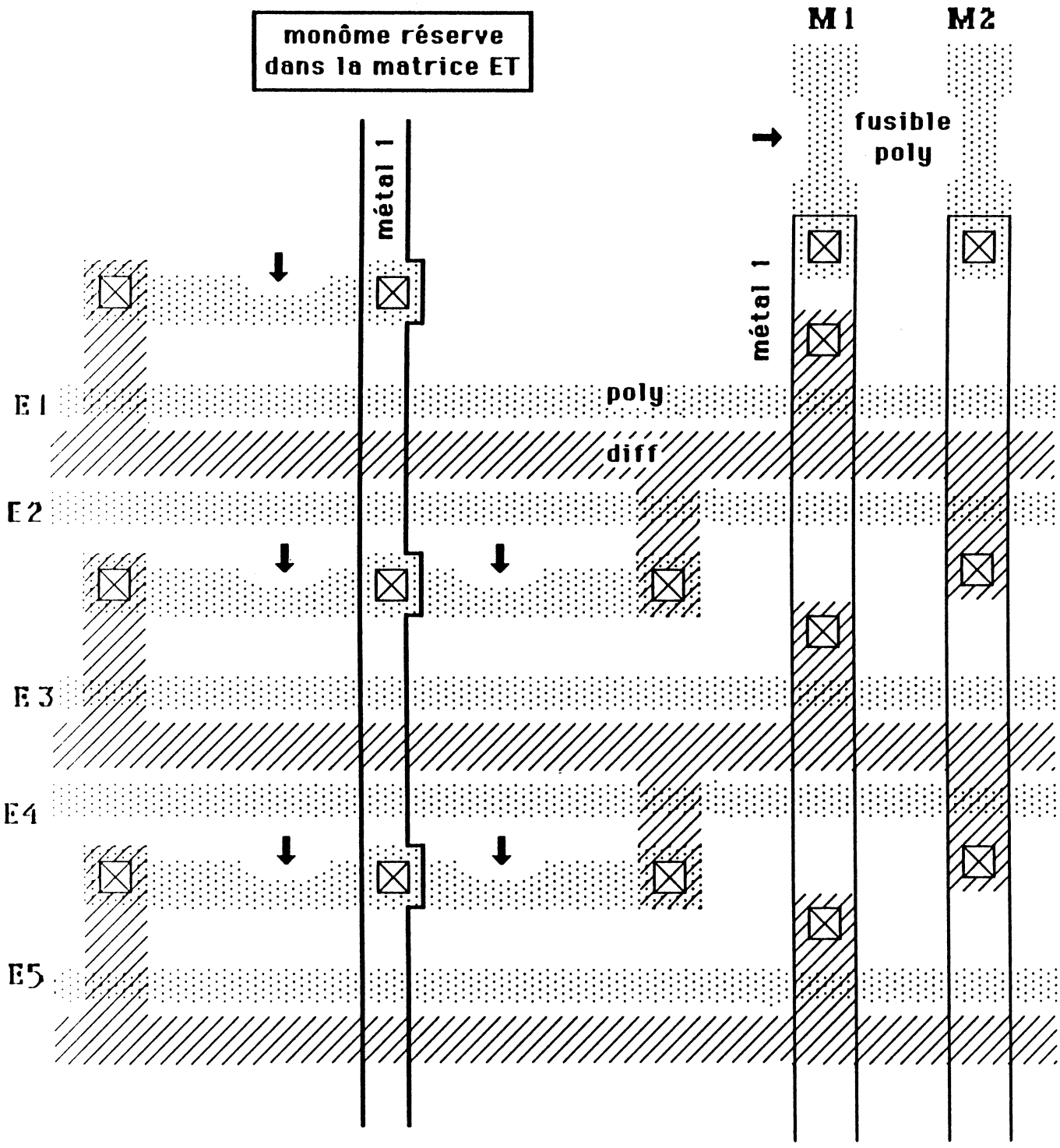


Figure III.13- Ligne réserve de monôme dans la matrice ET à base de coupure de ligne de poly par laser.

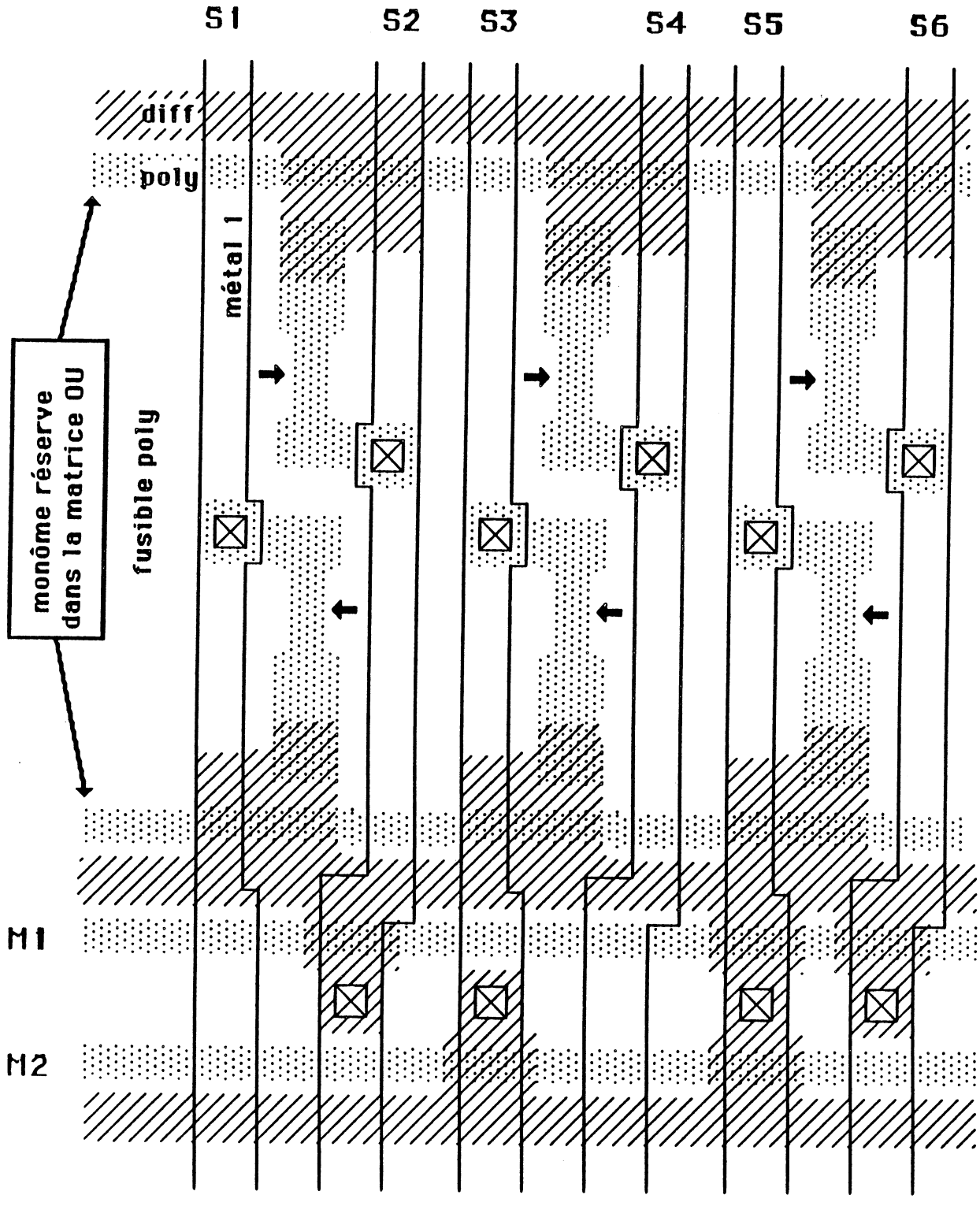


Figure III.14- Ligne réserve de monôme dans la matrice OU à base de coupure de ligne de poly par laser.

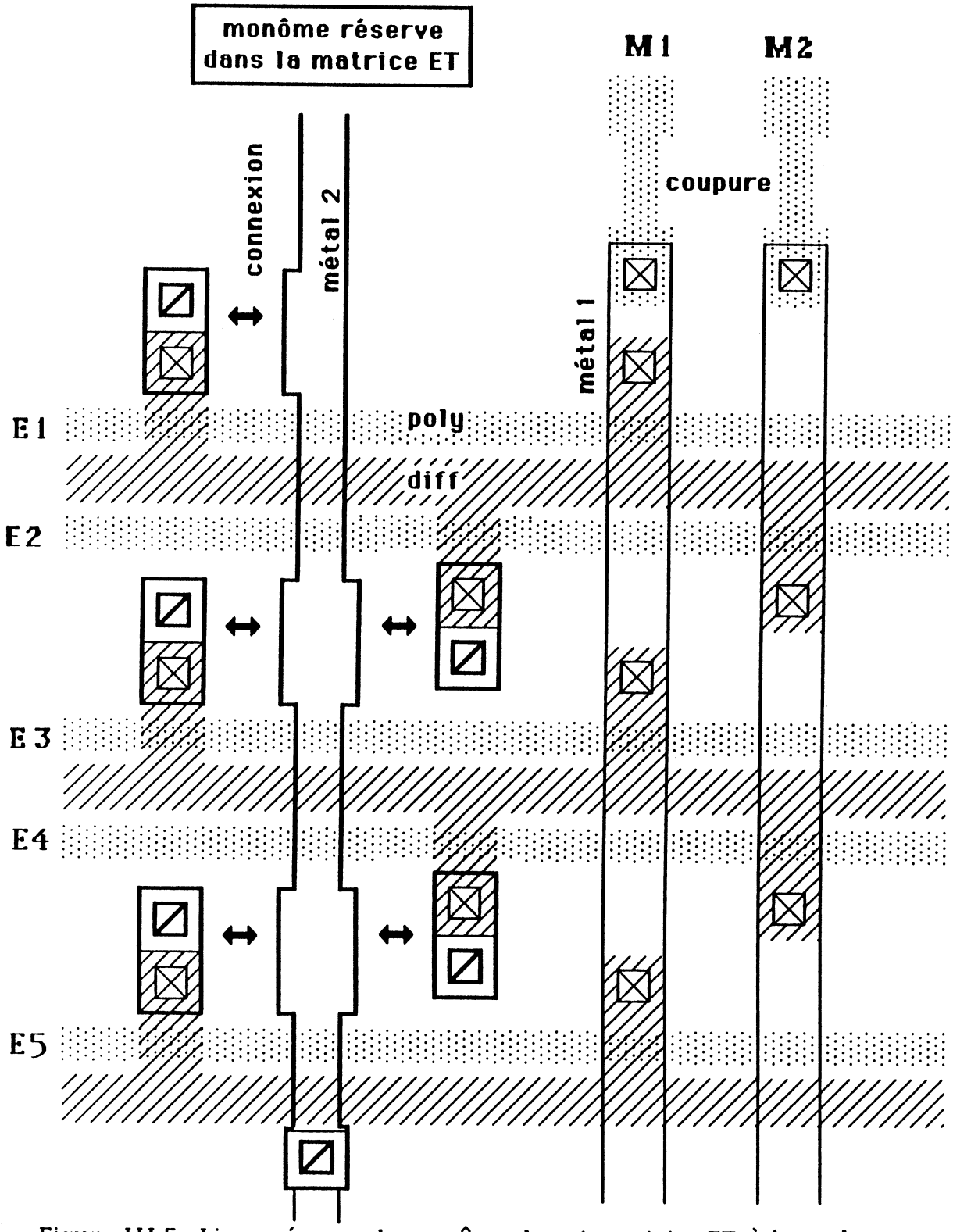


Figure III.5- Ligne réserve de monôme dans la matrice ET à base de connexion entre deux lignes métal 2 par laser.

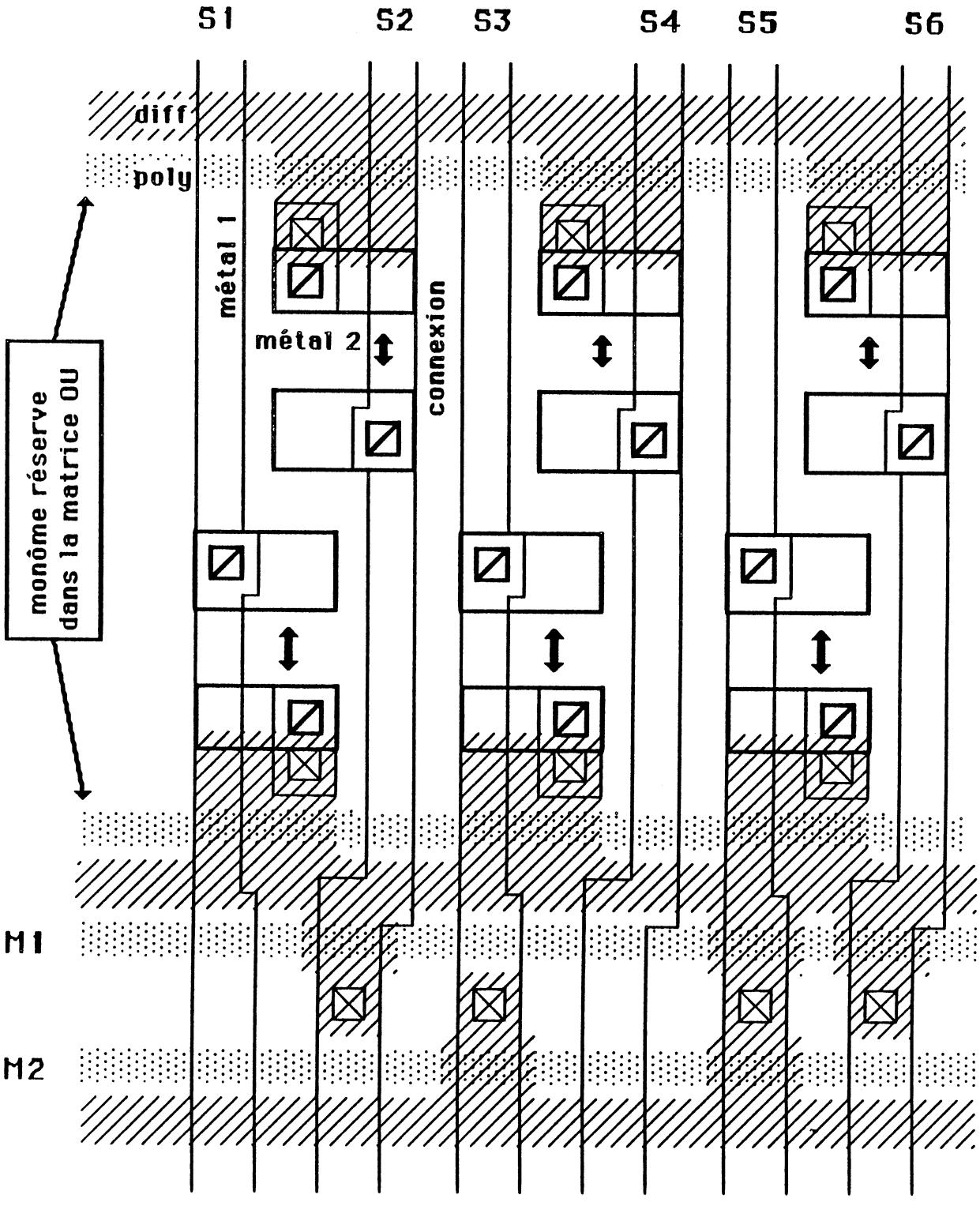


Figure III.16- Ligne réserve de monôme dans la matrice OU à base de connexion entre deux lignes métal 2 par laser.

b- Grille flottante

Le transistor FET à grille flottante [SHA83] offre à la fois la possibilité de la connexion (transistor pMOS enrichi) et la possibilité de la déconnexion (transistor nMOS déplété).

Les avantages de cette technique d'interrupteurs sont:

- La possibilité de programmation à tout moment par un faisceau d'électrons.
- Cette technique de programmation est réversible par la technique d'effaçage sous UV par exemple.
- Le transistor FET à grille flottante occupe une surface comparable à celle d'un transistor normal et présente un faible surcoût de fabrication.

Les seules limites de cette technique sont:

- La possibilité d'effet capacitif sur la grille flottante avec les lignes voisines telles que la ligne d'horloge. Ceci peut être résolu par une garde topologique.
- Le temps de rétention décroît (10 ans à 50°C, 50 heures à 150°C) avec la température, ce qui pose une problème en norme militaire.

La figure III.17 montre l'intégration d'une ligne réserve, à base des transistors à grille flottante, dans un PLA reconfigurable.

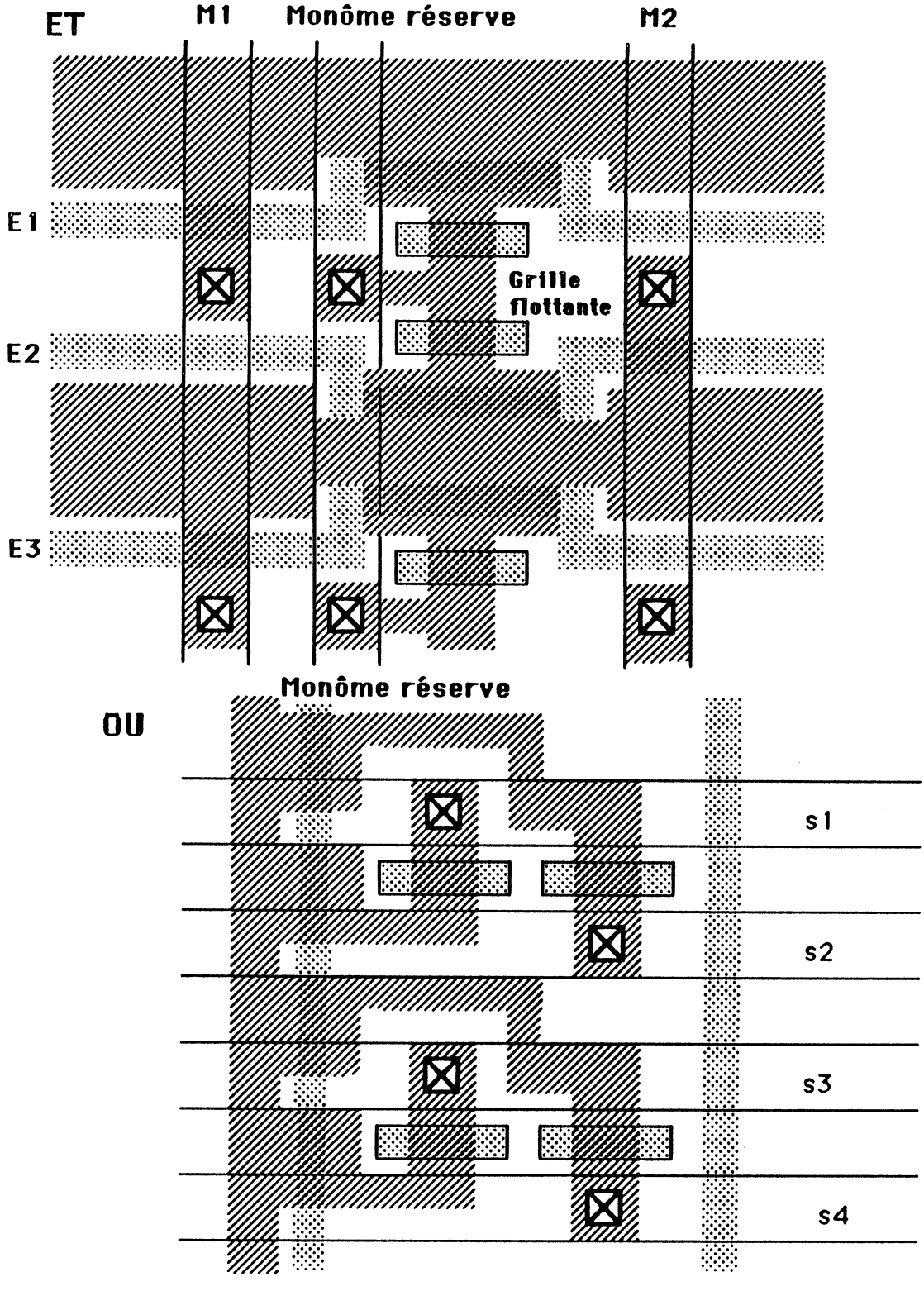


Figure III.17- Ligne réserve de monôme à base des transistors à grille flottante.

IV- Reconfiguration par commande de transistors

4.1- Introduction

Malgré la limitation de la reconfiguration aux monômes et aux sorties, un nombre important d'interrupteurs est toujours nécessaire pour reconfigurer un PLA. Ceci représente, en utilisant des fusibles, une croissance de la surface occupée par un PLA.

En vue de réduire le nombre de fusibles à utiliser, nous allons introduire une méthode de reconfiguration en commandant des transistors plutôt que des fusibles. Cependant, nous montrerons que cette méthode peut présenter une difficulté de conception pour qu'elle soit compatible avec les outils de génération de dessin de masques de PLA et par conséquent, elle est moins facilement industrialisable.

Les même principes, d'échange des lignes contenant des défauts avec des lignes saines, sont utilisés. Les entrées et les sorties d'un PLA restent également inchangées, vues de l'extérieur ou de l'interconnexion avec les autre blocs, avant et après la phase de la reconfiguration.

Par contre le changement de chemin entre deux lignes est basé sur la programmation de commandes de transistors en lieu et place de la programmation de fusibles.
En effet, la ligne service de sortie reste définie de la même manière qu'auparavant, par contre la ligne service de monôme représente uniquement la partie de la ligne monôme dans la matrice ET. C'est-à-dire, une ligne service monôme comporte des drains de transistors. Nous allons voir que ce choix facilite la contrainte d'échange entre lignes.

Cette méthode de reconfiguration par commande de transistors, est

appliquée pour les lignes de monômes et pour les lignes de sorties.

4.2- Méthode d'échange de lignes

Cette méthode d'échange de lignes comporte les deux ensembles de commandes suivants:

a- des commandes sont utilisées pour programmer l'état des transistors, afin d'assurer l'échange des lignes contenant des défauts par les lignes saines.

b- des commandes sont utilisées pour programmer l'état des transistors, afin d'assurer l'identification des lignes échangées.

a- commandes d'échange

Les commandes d'échange des lignes doivent assurer à la fois l'isolement de la ligne contenant de défauts et la remplacer par la ligne saine.

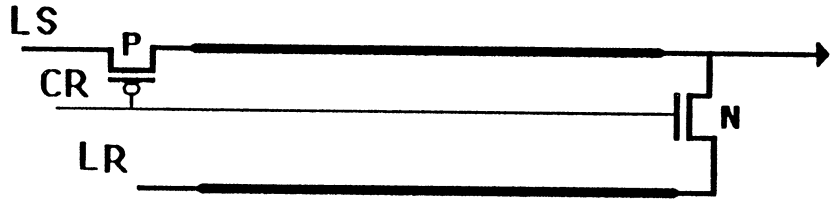


Figure III.18- Echange des lignes.

Ceci peut être réalisé par une commande de remplacement (CR), figure III.18, qui active le transistor pMOS (CR=0) pendant la phase de test et qui active le transistor nMOS pendant la phase de reconfiguration, s'il y a des défauts. Pour chaque ligne correspond une commande CR. Les commandes CR seront traitées ultérieurement.

b- commandes d'identification

Les commandes d'identification (CI) des lignes doivent assurer la commande des transistors d'une ligne réserve, afin de l'identifier à la ligne d'origine de service à corriger. Ceci implique la structure de lignes réserve à base de transistors en lieu et place de fusibles, figure III.19.

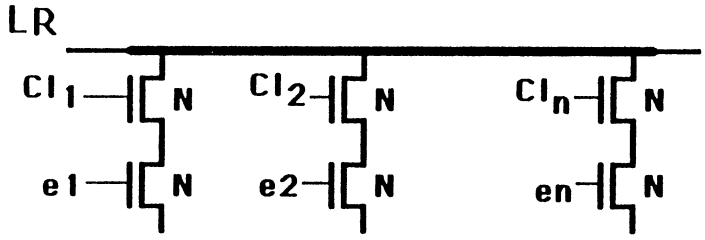


Figure III.19- ligne réserve à base de transistors.

Les transistors (de commande) correspondant à ceux qui appartiennent à la ligne de service (contenant des défauts), doivent être passants (CI=1) aprés la phase de reconfiguration (CR=1, C1=1, C2=1). Les autres transistors restent bloqués (CIn=0), figure III.20.

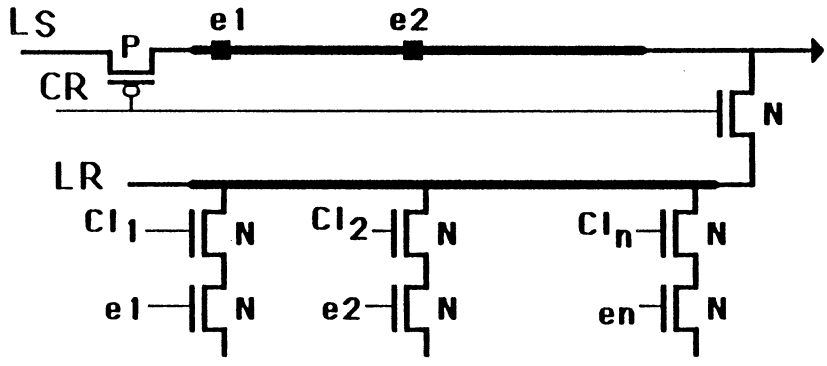


Figure III.20- Echange et identification des lignes.

Les commandes (CR, CI) peuvent être remplacées par des commandes identiques (C), figure III.20 ; car elles ont la même caractéristique. Elles

sont à 0 pendant la phase de test, et à 1 pendant la phase de reconfiguration pour les lignes contenant des défauts.

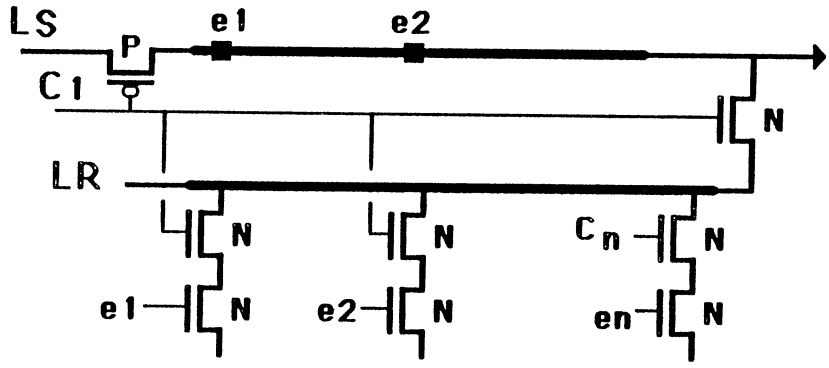


Figure III.21- Identification des commandes CR et CI par C.

Comme une ligne réserve doit assurer la possibilité d'identification avec une ligne service parmi plusieurs lignes. Ceci pose la contrainte suivante:

- soit deux lignes services (LS1, LS2), avec deux commandes (C1, C2), et qui contiennent des transistors au même niveau, figure III.22. Une ligne réserve qui couvre LS1 et LS2 ne peut être identifiée ni à LS1 ni à LS2.

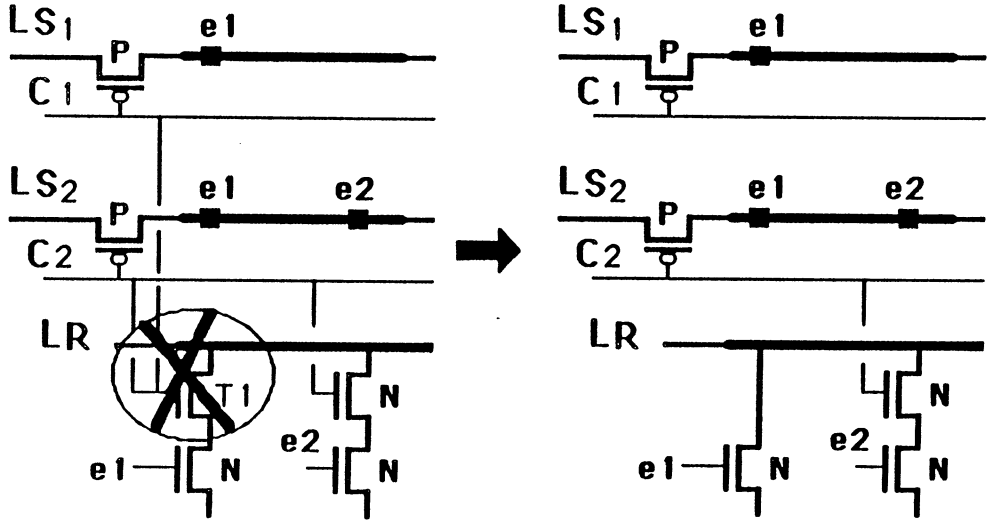


Figure III.22- Contrainte d'identification des lignes.

Car les deux commandes (C1, C2) sont connectées au même transistor T1. Une solution consiste à supprimer T1; car T1 doit être conducteur si l'échange est effectué soit avec LS1 soit avec LS2. A partir de cette contrainte nous pouvons introduire les conditions d'identification de lignes.

c- conditions d'identification de lignes

Une ligne réserve peut être identifiée à une ligne service, qui appartient à un ensemble [LS], si les lignes services de cet ensemble ont les caractéristiques suivantes:

- les éléments de [LS] sont disjoints entre eux. Deux lignes sont disjointes si elles ne comportent pas de transistors au même niveau, figure III.23.

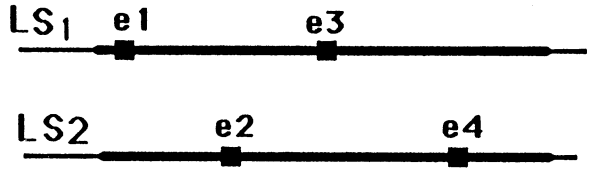


Figure III.23- lignes disjointes.

Si deux lignes ne sont pas disjointes, elles peuvent le devenir par substitution de lignes, pour éliminer les transistors au même niveau, figure III.24.

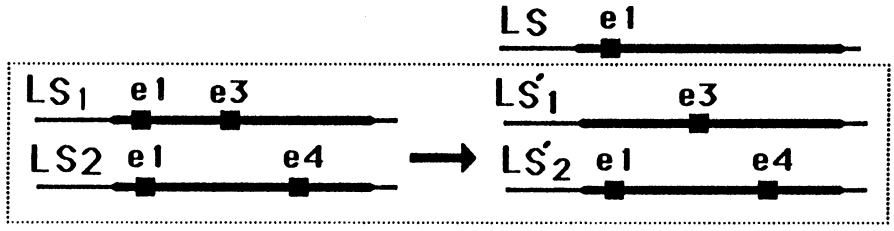


Figure III.24- lignes disjointes par substitution.

- nous venons de montrer que nous savons reconfigurer des lignes dans les deux cas suivants : des éléments ou des transistors sont distribués au même niveau sur toutes les lignes de l'ensemble [LS] et d'autres sont distribués d'une manière disjointe sur les autre niveaux, figure III.25. Ceux-ci définient les conditions d'identification de lignes.

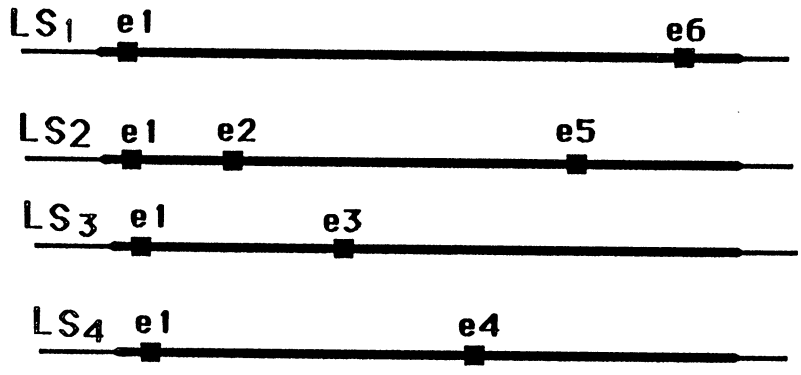


Figure III.25- distribution des transistors d'un ensemble reconfigurable.

Ces conditions nécessitent une étape de conception pour retrouver des ensembles de lignes reconfigurables par cette méthode. Nous allons présenter l'algorithme qui cherche les ensembles reconfigurables.

d- <u>algorithme de reconfiguration</u>

La réalisation de cette méthode de reconfiguration par commande de transistors nécessite les étapes suivantes :

a- classer les lignes suivantes dans des ensembles qu'on appelle des ensembles reconfigurables:

- les lignes qui sont disjointes normalement ou après substitution ;
- les lignes qui sont identiques à un ou plusieurs niveaux et qui sont disjointes dans les autres niveaux.

b- ajouter à chaque ensemble une (ou plusieurs) ligne réserve qui couvre tous les éléments de l'ensemble correspondant. Une seule commande (C à la fois pour l'échange et pour l'identification de lignes) est utilisée pour chaque élément, car une ligne réserve peut corriger uniquement un seul élément de l'ensemble.

c- toutes les lignes commandes sont à 0 pendant la phase de test.

d- des lignes commandes doivent passer à 1 pendant la phase de reconfiguration, afin de corriger les lignes défectueuses.

La programmation de l'état final des lignes de commande peut être réalisée par différentes méthodes. Ceci dépend de différents facteurs qui sont choisis par le concepteur ou par l'utilisateur, tel que la surface, la fiabilité et le surcoût de conception et de fabrication. Par exemple, l'utilisation de fusibles nécessite une surface importante de silicium; par contre les transistors à grille flottante manquent de fiabilité à une température élevée. L'utilisation des points mémoires sensibles à la lumière nécessite des masques ou des caches avec un boîtier transparent. Cette méthode ne peut pas être industrialisée.

Finalement la solution retenue est la coupure par laser des lignes de poly et de métal, ou la connexion par laser des lignes de métal 2.

CONCLUSION

La réalisation de circuits toujours plus complexes (VLSI) pose des problèmes de rendement aigus. Ceux-ci ne peuvent être résolus qu'en reconfigurant les circuits partiellement en panne à la fin de fabrication.

Cette technique consiste à corriger un circuit partiellement en panne et à lui rendre un mode de fonctionnement normal, en remplaçant la zone contenant des défaults par une zone saine. Ceci nécessite des architectures plus élaborées (interrupteurs), ainsi une phase préalable de test pour détecter et localiser la panne ou la zone à reconfigurer.

Cette technique de reconfiguration est applicable dans la structure de PLA CMOS. La stratégie suivie, afin de corriger un PLA partiellement en panne, est la suivante:

- Un PLA est divisé en trois zones à corriger qu'on appelle des lignes:
  - ligne d'entrée service LES
  - ligne monôme service LMS
  - ligne sortie service LSS

- On ajoute dans chaque zone des lignes supllémentaire que l'on appelle des lignes de réserve.

- L'échange, entre une ligne service contenant des pannes avec une ligne réserve, est réalisé à l'aide de la technique de programmation des interrupteurs (fusible à laser, grille flottante et coupure/connexion de lignes par laser). Cette technique peut devenir compatible avec les outils CAO de génération de dessin de masques de PLA. Le nombre

d'interrupteurs utilisés entraîne un surcoût de fabrication élevé.

L'étude de la distribution des types de panne permet de localiser les pannes les plus probables dans les transistors. La technique de reconfiguration peut être alors améliorée pour corriger uniquemment les types de pannes les plus probables. Il suffit de reconfigurer les monômes et les sorties, et de supprimer la reconfiguration des entrées.

Cette technique peut être aussi améliorée par le choix d'interrupteurs présentant à la fois, un faible surcoût de fabrication, une bonne souplesse de programmation, une faible surface de silicium occupée et une stabilité dans le temps et en fonction de la température après la programmation.

Une autre solution, pour réduire le nombre des interrupteurs, est possible (mais moins facilement industrialisable) en commandant des transistors en lieu et place de la programmation des interrupteurs. En effet, cette technique introduit des contraintes sur la conception pour qu'elle soit compatible avec les outils CAO de génération de dessin de masques de PLA.

CHAPITRE IV

PARTITIONS DE PLA

Introduction

La surface de Silicium occupée par un PLA implanté directement, par rapport à celle de la structure en logique aléatoire équivalente est souvent pénalisante en raison de sa structure régulière. Il est donc nécessaire de mettre en oeuvre de techniques d'optimisation pour les PLA.

Ces techniques d'optimisation consistent à réduire la surface occupée par un PLA en agissant sur les différents facteurs suivants:

- Les matrices d'un PLA sont généralement creuses et à faibles taux de remplissage (en particulier pour la matrice OU).

- Les lignes d'entrées (sorties) occupent toute une colonne sur une extrémité du PLA au pas minimal. Par conséquent, la surface occupée par un PLA augmente ainsi que celle occupée par les interconnexions avec d'autres blocs.

- Le nombre des monômes ou de termes augmente la surface de PLA.

- Un gros PLA présente des capacités et des résistances parasites importantes. Ceci augmente le temps de réponse et réduit les performances.

A partir de ces facteurs, l'optimisation de PLA est une solution qui réduit la surface pour améliorer ses performances. Ceci peut être effectué par les deux techniques d'optimisation logique et d'optimisation topologique.

* Les techniques d'optimisation logique [AUG78, KAN81] agissent sur les fonctions logiques afin de réduire le nombre de monômes qui est, à la fois, un facteur dans la réduction de surface et dans la réduction du nombre de transistors dans les matrices de PLA. Ce dernier critère de minimisation facilite la mise en oeuvre de techniques d'optimisation topologique.

* Les techniques d'optimisation topologique réorganisent la structure interne des matrices (sans modifier la structure logique) de manière à trouver un compactage convenable.
Cette réorganisation doit permettre de placer plusieurs entrées ou sorties dans une même colonne et plusieurs monômes dans une même ligne, dans le but de diminuer au maximum le nombre de colonnes et de lignes.

Les techniques d'optimisation topologique les plus importantes à citer sont:

- La technique de triangularisation qui consiste à ordonner les monômes et les entrées (sorties) de manière à pouvoir distribuer les transistors d'une matrice (particulièrement la matrice OU) sous une forme triangulaire, dont la partie ne contenant aucun transistor, peut être utilisée pour y placer d'autres blocs.

- La technique du "folding PLA" qui consiste à placer:
  deux ou plusieurs entrées (sorties) dans une même colonne.
  deux ou plusieurs monômes dans un même niveau ou ligne.

Cette technique a été introduite par WOOD [WOO79] et concerne le folding simple. Elle a été suivie par HACHTEL [HAC82, HAC80] pour placer deux ou plusieurs entrées (sorties) dans la même colonne. Une autre technique semblable à celle de folding est explorée par [LEP82, PAI81, GRE76].

La technique de partionnement de PLA que nous allons exposer consiste à morceler un PLA original en plusieurs petits PLA, en vue de réduire sa surface et son temps de réponse, et de faciliter sa reconfiguration et son interconnexion avec les blocs voisins.

I- Technique de partition de PLA

Le partitionnement de PLA consiste à morceler un PLA original en plusieurs petits PLA en vue de réduire principalement sa surface. Ceci peut être réalisée par différentes techniques et sous les différentes formes suivantes:

- Partitionnement de la matrice ET (entrées): cette technique consiste à morceler la matrice ET en plusieurs blocs en cherchant les "folding inputs", figure IV.1.

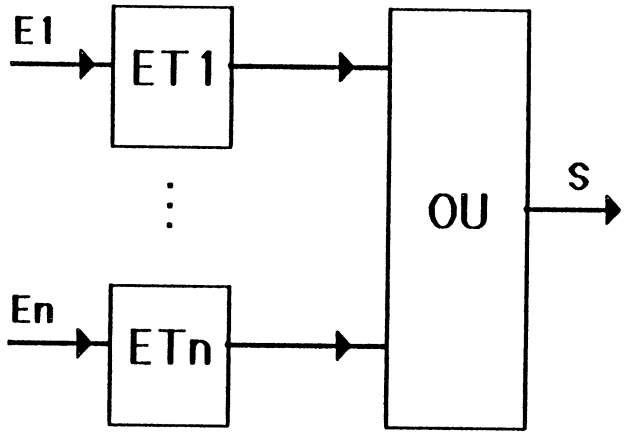


Figure IV.1- Partition de la matrice ET.

- Partitionnement de la matrice OU (sorties): cette technique consiste à morceler la matrice OU en plusieurs blocs en cherchant les "folding outputs", figure IV.2.

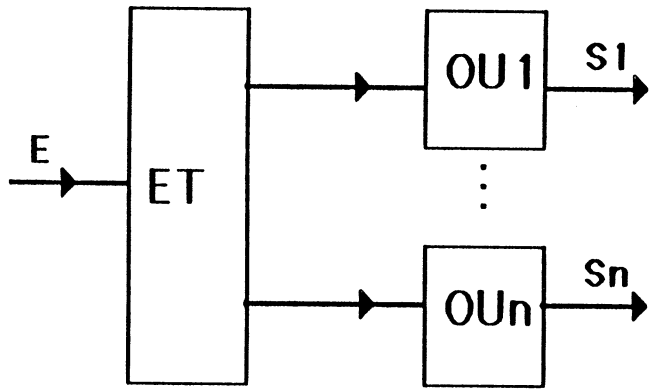


Figure IV.2- Partition de la matrice OU.

- Partitionnement en série: cette technique consiste à morceler un PLA original en plusieurs pętits PLA en série, figure IV.3.

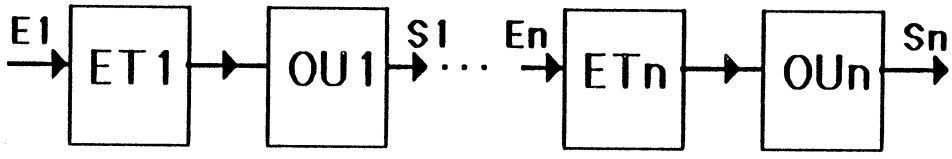


Figure IV.3- Partition de PLA en série.

- Partitionnement en parallèle: cette technique consiste à morceler un PLA original en plusieurs petits PLA fonctionnant en parallèle, figure IV.4.

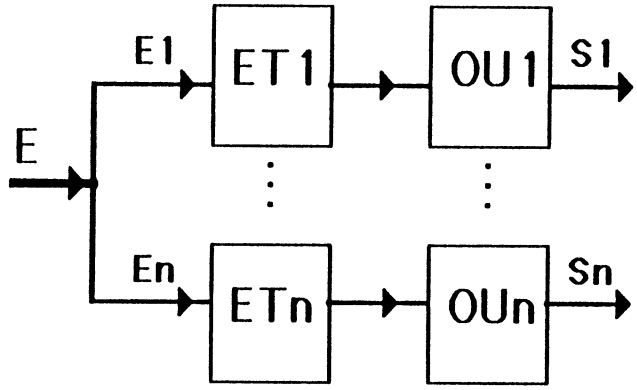


Figure IV.4- Partitions de PLA en parallèle.

- Partitionnement en parallèle/série: cette technique consiste à morceler un PLA original en plusieurs petits PLA en parallèle/série, figure IV.5.

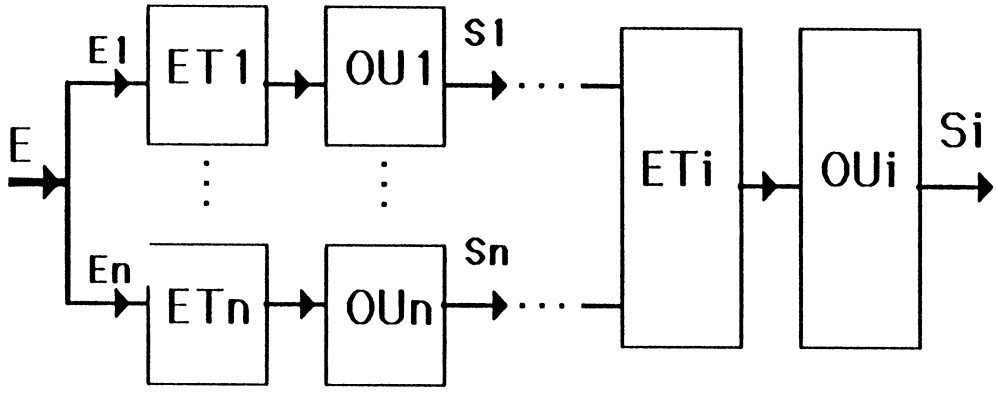


Figure IV.5- Partition de PLA en parallèle/série.

Différentes techniques heuristiques de partitionnement de PLA sont proposées.

- [SUW81] présente une méthode de partition basée sur la segmentation de monômes et sur le pliage des entrées/sorties (segmented-folded PLA). Les PLA partitionnés par cette technique ne sont pas indépendants.

- [SMI83] présente une méthode heuristique de partition en cherchant les matrices équivalentes avec une représentation graphique. Cette méthode, qui optimise principalement la surface, reste limitée à l'utilisation de PLA de taille moyenne.

- [HEN84] a suivi la méthode général de partitionnement de PLA qui était présentée par [KAN81]. En effet [HEN84] a étudié le placement des PLA après partitionnement.

Nous allons présenter une méthode de partitionnement qui diffère des méthodes présentées ci-dessus, ceci en raison de nos objectifs, qui se formulent de la façon suivante:

1- Le premier objectif est de réduire la surface occupée par le PLA original. Ceci est général, du fait que la technique de partitionnement est liée à l'optimisation topologique.

2- Le deuxième objectif vise à améliorer les performances électriques du PLA. Ceci exige de placer les PLA partitionnés en parallèle afin de réduire le temps de réponse.

3- Le troisième objectif consiste à morceler le PLA original en plusieurs petits PLA, qui sont placés de manière à faciliter les connexions (connexions directes) avec le bloc voisin. Ceci exige un placement rectangulaire des PLA de longueur égale à celle du bloc voisin (figure IV.6). Ce critère a tendance à augmenter le nombre de PLA partitionnés et ainsi la surface totale, à cause de la duplication des monômes et des entrées.

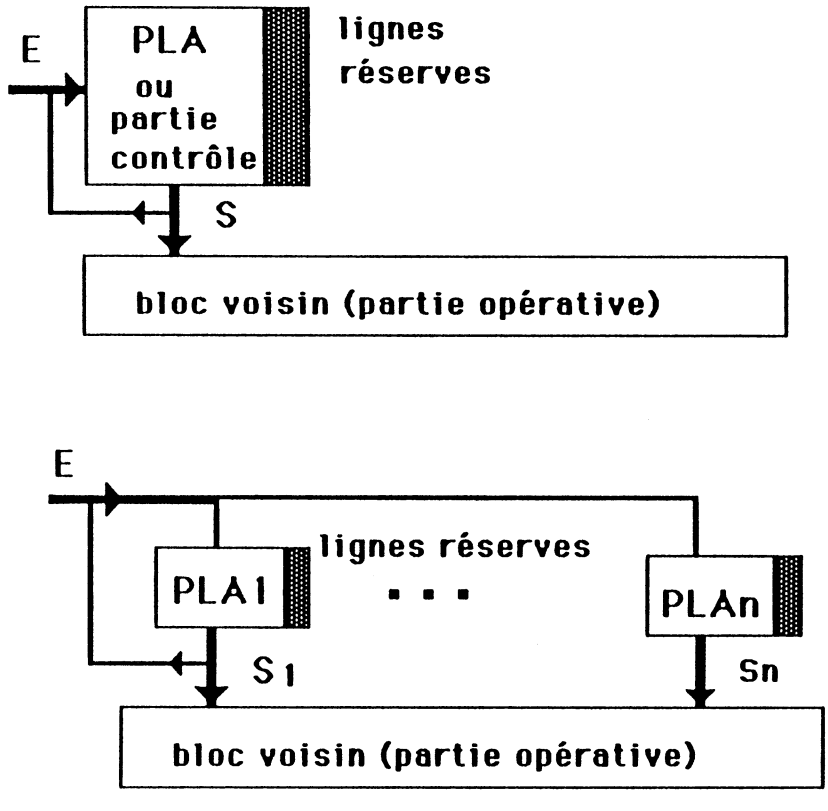


Figure IV.6- Partition et reconfiguration de PLA par rapport au bloc voisin.

4- Le quatrième objectif vise à trouver des PLA indépendants après partitionnement (sorties indépendantes) afin d'obtenir des PLA adaptables à la méthode de reconfiguration [DAN85].

II- Méthode de partition

Après la présentation des quatre objectifs précédents, nous allons introduire une nouvelle méthode "paramétrée" de partitionnement qui doit respecter les deux facteurs suivants:

- avoir des PLA partitionnés indépendants et pouvoir les placer en parallèle.
- pouvoir morceler un PLA original en un nombre déterminé (donné par le concepteur) de petits PLA.

Cette méthode de partition ne vise pas uniquement l'optimisation en surface, par contre, elle vise à faciliter le placement des PLA partitionnés. En effet, nous allons voir que deux degrés de liberté sont possibles par le choix des deux facteurs suivants:

1- nombre de monômes à dupliquer. Ceci représente la limite de la surface à augmenter.
2- les monômes qui peuvent être dupliqués.

Nous allons partir de l'hypothèse topologique que nous voulons un nombre n de petits PLA en parallèles à partir d'un PLA original (nous verons dans la suite les critères pour choisir n). Ceci implique de construire n ensembles de sorties indépendants (PLA indépendants) dans le PLA original.

Plusieurs méthodes sont possibles pour trouver n ensembles de sorties indépendantes. La méthode la plus triviale est de prendre n ensembles quelconques de sorties et de les implanter directement par leur fonctions logiques. Ceci peut représenter un très fort taux de duplication de monômes, et par conséquent, une surface non optimisée. Autrement, nous pouvons chercher la matrice de Jordan à partir de la matrice OU du PLA, et en déduire les ensembles de sorties indépendantes. Cette méthode ne prend pas en compte le choix du concepteur, quant au nombre de PLA partitionnés.

Enfin, nous allons suivre une méthode qui morcelle un PLA en n petits PLA qui sont relativement équivalents en surface. Cette méthode consiste à chercher n ensembles de sorties indépendants en retrouvant les sorties disjointes, ou en cherchant les sorties qui peuvent devenir disjointes par duplication de monômes. Chaque élément d'un ensemble doit être disjoint avec les autres éléments des autres ensembles.
Avant de présenter les étapes de cette méthode, nous allons donner une série de définitions qui concernent la personnalisation d'un PLA.

définition 1

Une colonne (entrée ou sortie) est adjacente avec un monôme si et seulement si il existe un transistor sur le point de croisement.

définition 2

deux colonnes (entrées ou sorties) sont disjointes si et seulement si il n'existe aucun monôme adjacent avec ces deux colonnes.

définition 3

le poids ou le cardinal d'une sortie (card(s)) est égal au nombre de transistors correspondant à cette sortie dans la matrice OU.

Exemple: s'il y a seulement 5 monômes adjacents à la sortie sj, alors card(sj)=5.

définition 4

delta (si,sj) (Δ(si,sj)) est le nombre de monômes adjacents à la fois aux deux sorties (si,sj).

Exemples: s1 et s2 sont disjoints, alors Δ(s1,s2)=0. Si s3 est adjacent aux monômes (m1, m2 et m3) et si s4 est adjacent aux monômes (m1, m3 et m4), alors Δ(s3,s4)=2.

définition 5

deux sorties (si,sj) non disjointes, peuvent devenir disjointes si on duplique les monômes qui sont adjacents à la fois à ces deux sorties.

Exemples: Si s1 est adjacent aux monômes (m1, m3) et si s2 est adjacent aux monômes (m1, m4), alors Δ(s3,s4)=1. Si on duplique le monôme m1, alors les deux sorties (s1, s2) peuvent devenir disjointes (figure III.27).

définition 6

Les sorties disjointes par rapport à une sortie d'origine sont classées par ordre de poids de la manière suivante : soit deux sorties, l'une est disjointe (Si) et l'autre (Sj) peut devenir disjointe (en dupliquant Δ termes) par rapport à une sortie d'origine. L'ordre du poids entre Si et Sj est défini par la comparaison de card(Si) avec (card(Sj)-Δ).

2.1- Algorithme de partitionnement de PLA:

Deux paramètres sont à définir par l'utilisateur:

- $\Delta 0$: taux maximum de duplication de termes.
- **nbclasse** : nombre maximal de classes. Ceci représente le nombre de PLA partitionnés.

1- Constitution d'un ensemble de classes :

Chacune des classes constituées, est formée d'un ensemble de fonctions les plus disjointes possibles, c'est-à-dire comportant le moins possible de monômes en commun.

algoritme 1:

Initialiser l'ensemble des classes EC à vide;

tantque il y a des sorties non selectionnées faire
début

selectionner la sortie S de poids le plus fort;
créer une classe C constituée de la sortie S;
chercher une sortie Sd disjointe à S
telle que le nombre de termes communs à Sd et à S soit
inférieur ou égal à $\Delta 0$;

tantque Sd existe et le nombre d'éléments de C est inférieur ou
égal à nbclasse faire
début

ajouter Sd à C;
chercher une sortie Sd disjointe avec $\cup_{i \in C} S_i$
telle que le nombre de termes communs à Sd et à $\cup_{i \in C} S_i$
soit inférieur ou égal à $\Delta 0$;

fin;

ajouter C dan l'ensemble des classes EC

fin;

2- Génération d'un partitionnement initial:

On selectionne dans l'ensemble des classes EC, la classe C qui contient le plus grand nombre de fonctions, et avec chacune de ces fonctions, on crée un PLA initial; on supprime C de EC.

3- Constitution du partitionnement final :

Il s'agit de prendre chaque fonction, pour chacune des classes de EC, et de l'ajouter dans l'un des PLA crées, de façon à minimiser le nombre de termes à dupliquer et de façon à équilibrer les tailles des PLA partitionnés.

algorithme 2:

Pour chaque classe C de EC faire
  pour chaque fonction S de C faire
    début
      si S est disjointe avec tous les PLA,
        alors ajouter S au PLA qui comporte le moins de monômes
        sinon ajouter S au PLA qui possède le plus de monômes en commun avec S
    fin.

L'application de cet algorithme donne un bonne résultat sur les PLA dotés d'une distribution homogène de transistors dans la matrice OU.

La contrainte de cet algorithme concerne les sorties qui dépendent de plus de 1/2 du nombre total des monômes. Dans ce cas, la ligne de sortie (S) est divisée en deux lignes (S1, S2) et dans deux PLA, où une porte OU qui rélie les deux lignes pour obtenir la veritable sortie (S=S1+S2).

L'annexe 2 montre un exemple de partitionnement (en 3 puis en 4 PLA) d'un PLA de taille moyenne, avec une réduction de la surface occupée de l'ordre de 30%.

2.2- Choix du nombre des PLA à partitionner (n)

Le partitionnement est réalisé, d'une part, pour obtenir des petits PLA reconfigurables et d'autre part , pour obtenir une largeur totale des PLA partitionnés comparable à celle du bloc voisin.

Pour trouver le nombre des PLA à partitionner, nous allons estimer leur largeur afin de la comparer à celle du bloc voisin.
Vu la structure d'un PLA (figure IV.7 et annexe 3), sa largeur dépend du nombre des lignes d'entrées (E), du nombre des lignes de sorties (S), de la largeur du circuit de précharge des lignes de monômes (lcp) et de l'espacement entre blocs (esp) (espacement entre les deux matrices et espacement entre deux PLA). Alors, la largeur d'un PLA (L_{pla}) peut être estimé de la manière suivante:

$$L_{pla} = E.L1 + S.L2 + lcp + esp$$

avec: L_1 : largeur du pas dans la matrice ET.

L_2 : largeur du pas dans la matrice OU.

D'après notre méthode de partition (partitionnement des PLA en parallèle), le nombre des lignes de sorties reste constant, par contre, le nombre des lignes d'entrées augumente (lignes d'entrées communes à plusieurs PLA), et la largeur du circuit de précharge des lignes de monômes (lcp) et de l'espacement entre blocs (esp) augumente en fonction du nombre de PLA partitionnés.

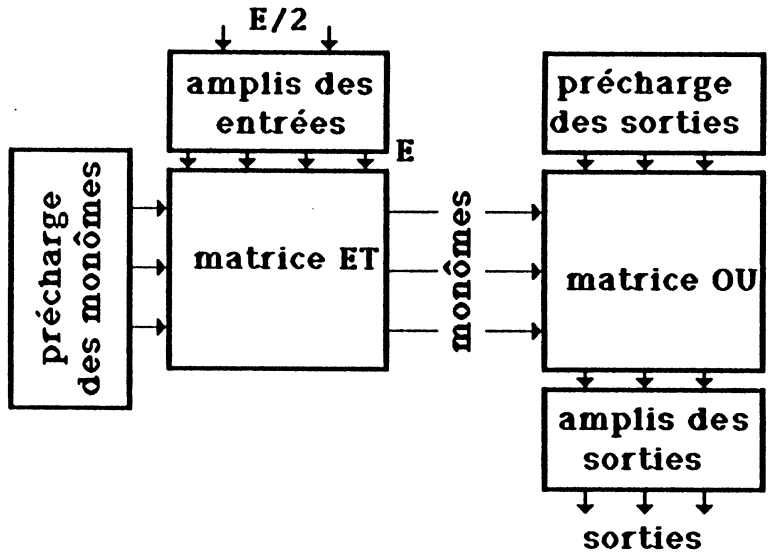


Figure IV.7- Structure générale de la topologie d'un PLA.

En effet, la largeur essentielle à estimer, dans les PLA partitionnés, est celle des lignes d'entrées. Dans la plupart des cas, on estime que les 3/4 des lignes d'entrées (comme limite supérieure) du PLA d'orgine restent en commun dans les deux PLA partitionnés. A partir de cette estimation, le nombre des lignes d'entrées (N_{le}) dans deux PLA devient:

$$N_{le} = (3/4 + 1/8)E + (3/4 + 1/8)E = 2\ (7/8)E$$

D'une manière générale: $$N_{le} = 2^x\ (7/8)^x\ E = n\ (7/8)^x\ E$$

avec $n = 2^x$ = nombre des PLA après partitionnement ($x = \log_2(n)$).

Cette loi d'acroissement du nombre des lignes d'entrées avec le nombre des PLA après partitionnement (n), va permettre d'estimer la largeur de l'ensemble des PLA. Nous aurons:

$$L_{pla} = N_{le} \cdot L1 + S \cdot L2 + n \cdot (lcp+esp)$$

où, $$L_{pla} = n\ (7/8)^x\ E \cdot L1 + S \cdot L2 + n \cdot (lcp+esp)$$

L'estimation de n (nombre de PLA à obtenir après partitionnement) résulte de cette loi. En effet, la largeur des PLA doit être comparable à celle du bloc voisin (L_{bv}), une comparaison des largeurs (L_{pla} et L_{bv}) détermine alors n de façon théorique.
Après le partitionnement, on compare la largeur des PLA du résultat pratique à L_{bv}. La comparaison de ces résultats permettra, si nécessaire, de rectifier la valeur de n.

Conclusion

Notre méthode de partitionnement de PLA atteint les quatre objectifs suivants :

1- l'optimisation topologique ;
2- la performance électrique par le placement en parallèle des PLA partitionnés ;
3- la répartition (en surface équilibrée) des PLA partitionnés sur une surface qui facilite les connexions avec d'autre blocs ;
4- des pétits PLA indépendants pour faciliter la reconfiguration.

En effet, les sorties disjointes (ou qui peuvent devenir disjointes) déterminent le partitionnement d'un PLA d'origine afin d'obtenir des PLA indépendants.

Cette méthode est paramétrée par le nombre des PLA à partitionner et par le taux maximum de duplications de monômes, ces paramètres étant laissés au choix du concepteur.

CONCLUSION

La complexité des circuits intégrés, actuels, rend nécessaire l'utilisation d'outils de conception à la fois plus simples, plus rapides et plus sûrs.

Dans ce domaine, cette thèse traite quatre aspects de conception (électrique, test, reconfiguration, topologique) interdépendants concernants les PLA en technologie CMOS :

1- Le premier aspect concerne les évaluations électriques et temporelles des PLA CMOS afin de développer un outil prototype de simulation électrique pour une structure de PLA donnée et pour une technologie donnée. Ceci est fondé sur la recherche du chemin critique (E/S) le plus long dans le PLA afin de le ramener à l'étude d'un inverseur. En effet, le résultat de cette étude est fonction des paramètres topologiques (dimensionnement de transistors).

2- Le deuxième aspect concerne le test des PLA. La distribution des types de pannes en fin de fabrication et leurs manifestations électriques et logiques sont étudiés afin de déterminer les composants défectueux

les plus probables. En effet, l'étude des pannes en fin de fabrication facilite la reconfiguration des PLA en permettant de ne corriger que les types de pannes les plus probables.

Une méthode de test est également étudiée afin de détecter les pannes SOP/SON dans les PLA CMOS.

3- Le troisième aspect concerne la reconfiguration des PLA afin d'améliorer le rendement de fabrication en corrigeant des zones défectueuses. Celle-ci est fondée sur l'ajout de lignes supplémentaires en tenant compte de la distribution des types de pannes, et sur la programmation des interrupteurs.

4- Le quatrième aspect concerne le partitionnement d'un PLA. Ceci est fondé sur la recherche des sorties disjointes.
Les PLA partitionnés (fonctionnant en parallèle) présentent une facilité de reconfiguration et une amélioration de la performance électrique.

REFERENCES

[ABR82] G. P. MEK, J. A. ABRAHAM and E. S. DAVIDSON, "The design of PLAs with concurrent Error Detection", 12th Int. Symp. of Fault-Tolerant Computing, California, June 22/24, 1982.

[AUG78] M. AIGUIN, "Conception des systèmes de commande à l'aide des Réseaux Logiques Programmables", Thèse de 3ème cycle , Université de Nice, 17 nov. 1978.

[BAN82] P. BANERJEE, "A model for simulating physical failures in MOS VLSI circuits" 1982 repport CSG13, university of Illinois at Urbana USA.

[CAN86] M. CAND, E. DEMOULIN, J. L. LARDY, P. SENN, "Conception des Circuits Intégrés MOS", Editions EYROLLES et CNET-ENST, 1986.

[CAS76] G. R. CASE, "Analysis of fault mecanisms in CMOS gates", 13th Design Automation Conference Proceedings, 1976.

[CHA78] C. W. CHA, "A testing strategy for PLAs", 15th Design Aut. Conf. Proc., pp. 326/334, June 1978.

[CHA83] R. CHANDRAMOULI, "On testing stuck-open faults", FTCS 13th, pp. 1/8, mai 1983.

[CHI83] K. W. CHIANG and G. VRANSIC, "On fault detection in CMOS logic networks", 20th DAC 1983.

[CHO81] CHONG Ming LIN, "A 4 micron NMOS NAND structure PLA", IEEE, J. Solid state Circuits, Vol. SC-16, n°2, April 1981.

[CHU83] S. CHUQUILLANQUI, "Internal connection problem in large optimized PLAs", IEEE-ACM, 20th Design Aut. Conf. Miami, JUNE 27/29, 1983.

[COH84] P. B. COHEN, "An introduction to CMOS design styles", VLSI Design, pp. 88, sept. 84.

[DAN83] A. DANDACHE, "Evaluations électriques et temporelles des PLAs complexes"' Thèse 3ème cycle, INPG, 21 nov. 1983.

[DAN85a] A. DANDACHE, "Etude de structures régulières PLA-ROM", IMAG/LCS R.R. n°516 Avril 85.

[DAN85b] A. DANDACHE, "PLA reconfigurables", IMAG/LCS R.R. n°547 Juillet 85.

[DEC85] P. De CHOUDENS, "Test intégré de processeur facilement testable", thèse 3ème cycle informatique, Grenoble 14 novembre 1985.

[ELZ81] Y. M. EL-ZIQ, "Automatic test generation for stuck-open faults in CMOS VLI", 18th DAC, PP. 347/352, june 1981.

[ELZ83] Y. M. EL-ZIQ, "Modeling and testing of MOS physical faillures",IEEE Aut. test Prog. Gen, workschop, mars 15/16, SANFRANCISCO 83.

[GRE76] D. L. GREER, " An associative logic matrix", IEEE J. Solid-State Circuit, vol.SC-11, pp.679-691, 1976.

[GON82] N. F. GONCALAVES and H. DEMAN., "NP-CMOS: Racefree-Dynamic CMOS technique for pipelined logic structures", ESSIRC Digest. of Technical Papers, pp. 141/144, Sept. 82.

[HAC80] G. D. HACHTEL, A. L. SANGIOVANNI-VICENTELLI and A. R. NEWTON, " Some results in optimal PLA folding", dans ICCC80, pp. 1023-1027.

[HAC82a] G. D. HACHTEL, A. L. SANGIOVANNI-VICENTELLI and A. R. NEWTON, "Techniques for programmable logique array folding", 19th DAC-82, pp.147-155.

[HAC82b] G.D. HACHTEL, A.R. NEWTON, A.L. SANGIOVANNI VINCENTELLI, "An algorithm for optimal PLA folding", IEEE Trans. Comp. Aided Design of ICAS, Vol. 1, n°2, Apr. 82.

[HAR84] HARRY J. M. VEENDRICK, "Short-circuit dissipation of static CMOS circuitry and its impact on the design of buffer circuits", IEEE, Solid State Circuits, Vol. SC-19, n°4, Aug. 1984.

[HEB79] E. HEBENSTREIT and K. HORNINGER, "High-speed programmable logic arrays in ESFI SOS technology", IEEE, J. Solid State Circuits, Vol. SC-11, pp. 370/374, june 1979.

[HED85] K. S. HEDLUND, "Electrical Optimization of PLAs", 22th DAC-85, PP. 681-687.

[HEN64] F. C. HENNIE, "Fault detection experiments for sequential circuits", Proc. 5th ann. symp. swit. circ. th. and logic Des. pp.95, nov.64.

[HEN84] J. HENNESSY, "Partitioning programmable logic array summary", ICCAD84, pp.180-181.

[KAN81] S. KANG et W. M. van CLEEMPUT, "Automatic PLA Synthesis from DDL-P Description", dans DAC-81, pp. 391-397.

[KEN80] V. G. McKenny, "A 5V 64K EPROM utilising redundant circuitry", IEEE Int. Sol. State Cir. Conf. PP 146/147, 1980.

[KOK81] K. KOKKONEN, P. O. SHARP, R. ALBERS, J. P. DISHAW, F. L. LOUIE and R. J. SMITH, "Redundancy techniques for fast static RAMs", IEEE ISSCC, pp 80/81, 1981.

[KRA82] R. H. KRAMBECK, C. M. LEE and H. S. LAW, "High-speed compact circuits with CMOS", IEEE, J. solid state circuits, Vol. sc-17, pp. 614/619, june 1982.

[LEP82] P. LEPINAY, "GPLAM", thèse 3ème cycle, Université des Sciences et Technique de Lanquedoc, Montpellier 1982.

[LOT85] Z. LOTFI, "Etude et analyse des schémas logiques combinatoires a l'aide de représentations numérique et graphique", thèse de Docteur ES SCIENCES, Nancy le 8 février 1985.

[MAL84] W. MALY, F. J. FERGUSON and J. P. SHEN, "Systematic characterization of physical defects for fault analysis of MOS IC cells", IEEE Int. Test Conf. pp 390/399, 1984.

[MAN80] T. E. MANGIR and A. AVVIZIENIS, "Failure modes for VLSI and their effect on chip design", IEEE Int. Conf. Circ. Comp. pp 685/688, 1980.

[MAN84] T. E. MANGIR, "Sources of failures and yield improvement for VLSI and restructurable interconnects for RVLSI and WSI: part 1 - source of failures and yield improvement for VLSI", Proceedings of tne IEE, Vol. 72, n°6, june 84.

[MEA80] C. MEAD, L. CONWAY, "Introduction to VLSI systems", Reading, MA: Addisson-Wesley, 1980.

[MET83] L. R. METZGER, "A 16K CMOS PROM with polysilicon fusible links", IEEE J. solid state circuits Vol. sc 18, n°5, oct. 83.

[MIC83] G. De MICHELI and A. L. SANGIOVANNI VINCENTELLI, "PLEASURE: acomputer program for simple and multiple constrained folding of Programmable Logic Arrays" Proc. DAC. Miami Beach, jun 83.

[MIN82] 0. MINATO, T. MASUHARA, T. SASAKI, Y. SAKAI and T. YASUI, "A Hi-CMOSII 8K X 8 bit static RAM", IEEE JSSC, Vol. 17, n°5, oct 1982.

[MUR82] S. MUROGA, "VLSI System Design", Jonn Wiley and Sons, 1982.

[NIC86] G. NICOLAS, "Technical and economical aspect of laser repair for WSI memory", Workshop of Wafer Scale Integration, Grenoble University, 17-19 march 1986.

[OKL84] G. OKLOBDZIJA, "On testability of CMOS-DOMINO logic", 14th FTCS, ORLANDO, USA 84.

[OST79] D. L. OSTAPKO and S. J. HONG, "Fault analysis and test generation for programmable logic arrays (PLA)", IEEE Trans. On Comp. Vol sc 28, n°9, PP. 617/627, sept 1979.

[RAC] Reliability Analysis Center of the Rome Air Developemment Center (RADC), 1976.

[PAI81] J. F. PAILLOTIN, "Optimization of PLA area", 18th DAC-81, pp.406-410.

[PEN72] W. M. PENSEY and L. LAU, "MOS integrated circuits", NEWYORK: Van Nostrand, 1972 pp. 260/282.

[PRI82] W. D. PRITHARD, "A CMOS PLA generator", ESSIRC Digest of Technical Papers, pp. 94/97 sept.82.

[PET85] J. M. PETER van Laarhoven and E. H. L. Aarts, "PHIPLA- A new Algorithm for Logic minimization", 22th DAC, pp. 739-743.

[SCH78] SCHNABLE, G. L. , L. J. GALLACE and H. L. PRYOL, "Reliability of CMOS Integrated Circuits", Computer, Vol.11, n°10, Oct. 78, PP. 6-17.

[SHA84] D. C. SHAVER, "Electron-Beam customization, repair, and testing of Wafer-Scale Circuits", Sol. State Tech. pp. 135/139, Feb. 1984.

[SIE82] D. P. SIEWIOREK, R. S. SWARZ, "The theory and practice of reliable system design", Digital Equipement Coporation 1982.

[SIG76] Signetics Field Programmable Logic Arrays, Sunnyvale CA:cignatics, Mars 1976.

[SMI79] J. E. SMITH, "Detction of fault in programmable logic arrays", IEEE Trans. On Comp. Vol c-28, n°11, pp. 845/853, nov. 79.

[SMI82] R. J. SMITH, "Redundancy - the new device technologie", IEDM'82 Dig. papers, pp. 608/611 San Francisco, CA, Dec 1982.

[SMI81] R. T. SMITH, J. D. CHLIPALA, J. F. M. BINDELS, R. G. NELSON and T. F. MANTZ, "Laser programmable redundancy and yield improvement in 64K DRAM", IEEE JSSC, Vol.16, n°5, Oct. 81.

[SMI83] G. DeMICHELI & M. SANTOMAURO, "SMILE: a computer program for partitioning of programmed logic arrays", Computer Aided Design Vol.15,n°2, March 1983, pp.89-97.

[STE85] D. M. STEWART, "Lasers fix dynamic RAMs", ElectronicsWeeK Feb. 4, 1985.

[SUW81] I. SUWA & W. J. KUBITZ, "A computer aided design system for segment-folded PLA macro cells", Proc. 18th Design Automation Conf. june 81, pp.398-405.

[TIM83] C. TIMOC, M.BUEHLER, T.GRISWALD, C. PINA, F.STOTT and L.HESS, "Logical models of physical faillures", IEEE int. test conf. USA 83.

[UCL82] V. G. OKLODZIJA, "Design for testability of VLSI structures through the use of circuit techniques" UCLA Computer Science Departement, Univ. California, R. n° CSD 820820, Aug. 82.

[WAD78] R. L. WADSACK, "Fault modeling and simulation of CMOS and MOS integrated circuits", the Bell syst. tech. J. Vol.57, n°5, pp. 1449/1474, june 78.

[WOO79] K. H. WOOD, "High density programmable logic array chip", IEEE Trans. on Computers, vol. C-28, n°9, sept 79 pp.602-608.

ANNEXE 1.1

Nous allons montrer les résultats concernant le temps de réponse d'un inverseur CMOS au point B (figure 1.1), par calcul et par simulation (MSINC) pour la technologie suivante :

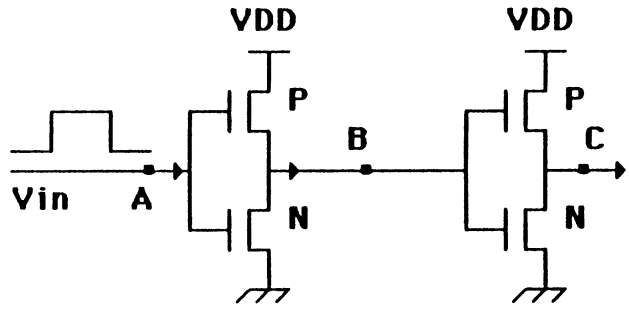


Figure 1.1- Deux portes CMOS identiques.

μ_n = 600 cm^2/v.s : mobilité du transistor enrichi (canal N) ;

μ_p = 200 cm^2/v.s : mobilité du transistor enrichi (canal P) ;

VDD = 5 volts : tension d'alimentation ;

Vtc = 0.75 volts : tension de seuil du tansistor P-canal ;

Vts = 0.75 volts : tension de seuil du tansistor N-canal ;

$\beta_s = \beta_c = 0$: coéfficient d'effet de substrat linéarisé ;

θ = 0.03 : constant pour calculer la mobilité ;

xn = 0.3 μm ;

xw = 0.5 μm ;

xjn = 0.45 μm ;

xjp = 0.8 μm ;

cox = 7.7E-4 pF/μm^2 : capacité d'oxyde mince ;

cjunn = 2.5E-4 pF/μm^2 : capacité de la jonction (canal N) ;

cjunp = 1.7E-4 pF/μm^2 : capacité de la jonction (canal P) ;

wc, lc , γc : dimensions des transistors (canal P) ;

ws, ls , γs : dimensions des transistors (canal N) ;

sd = 60 μm^2 : surface de la jonction drain (source) ;

cch = 0.1E-3 nF : capacité de charge à la sortie.

Les formules utilisées pour le calcul du temps de réponse sont :

cgc = cox . wc . lc : capacité de grille du Tp ;

cgs = cox . ws . ls : capacité de grille du Tn ;

cjs = cjunn.(sd+2.(ws+10).xjn) : capacité de la jonction du Tn ;

cjc = cjunp.(sd+2.(wc+10).xjp) : capacité de la jonction du Tp ;

cm = cjs+cjc+(5/3).cgs+(3/2).cgc+cch : capacité de sortie (au point B) pendant le temps de montée.

cd = cjs+cjc+(5/3).cgc+(3/2).cgs+cch : capacité de sortie (au point B) pendant le temps de descente.

tm : temps de montée de l'inverseur CMOS au point B

$$tm = \frac{Cm.(2.(\beta c+1)(Vtc-0.1.Vcc)/(Vcc-Vtc)+\ln((19-\beta c)-20.Vtc/((1-\beta c)Vcc)))}{\mu p.Cox.\gamma c.(Vcc-Vtc)}$$

td : temps de descente de l'inverseur CMOS au point B .

$$td = \frac{Cd.(2.(\beta s+1)(Vts-0.1.Vcc)/(Vcc-Vts)+\ln((19-\beta s)-20.Vts/((1-\beta s)Vcc)))}{\mu n . Cox . \gamma s . (Vcc-Vts)}$$

Pour différentes valeurs des dimensions des transistors (Tn, Tp), le tableau 1.1 montre les résultats numériques du temps de réponse, d'un inverseur CMOS (entre deux inverseur), par calcul et par simulation avec une précision de l'ordre de 15%.

						SIMULATION		CALCUL	
ws	ls	γs	wc	lc	γc	tm	td	tm	td
(µm)	(µm)		(µm)	(µm)		(ns)	(ns)	(ns)	(ns)
12	6	2	12	6	2	9	3.5	11	3.5
9	6	1.5	9	6	1.5	11	4	12	4
6	6	1	6	6	1	14	5	15	5
9	9	1	9	9	1	20	7	23	9
12	12	1	12	12	1	26	8	34	12
6	9	2/3	6	9	2/3	23	8	28	9

Tableau 1.1.

ANNEXE 1.2

L'annexe 1.2 montre une comparaison des résultats numériques de tr(ohm) (chapitre I, § 2.1.2.b) d'un inverseur chargé par RC :

- par calcul : tr(ohm) = 6ns pendant le temps de descente.
  tr(ohm) = 6ns pendant le temps de montée.

- par simulation: tr(ohm) = 5ns pendant le temps de descente.
  tr(ohm) = 6ns pendant le temps de montée.

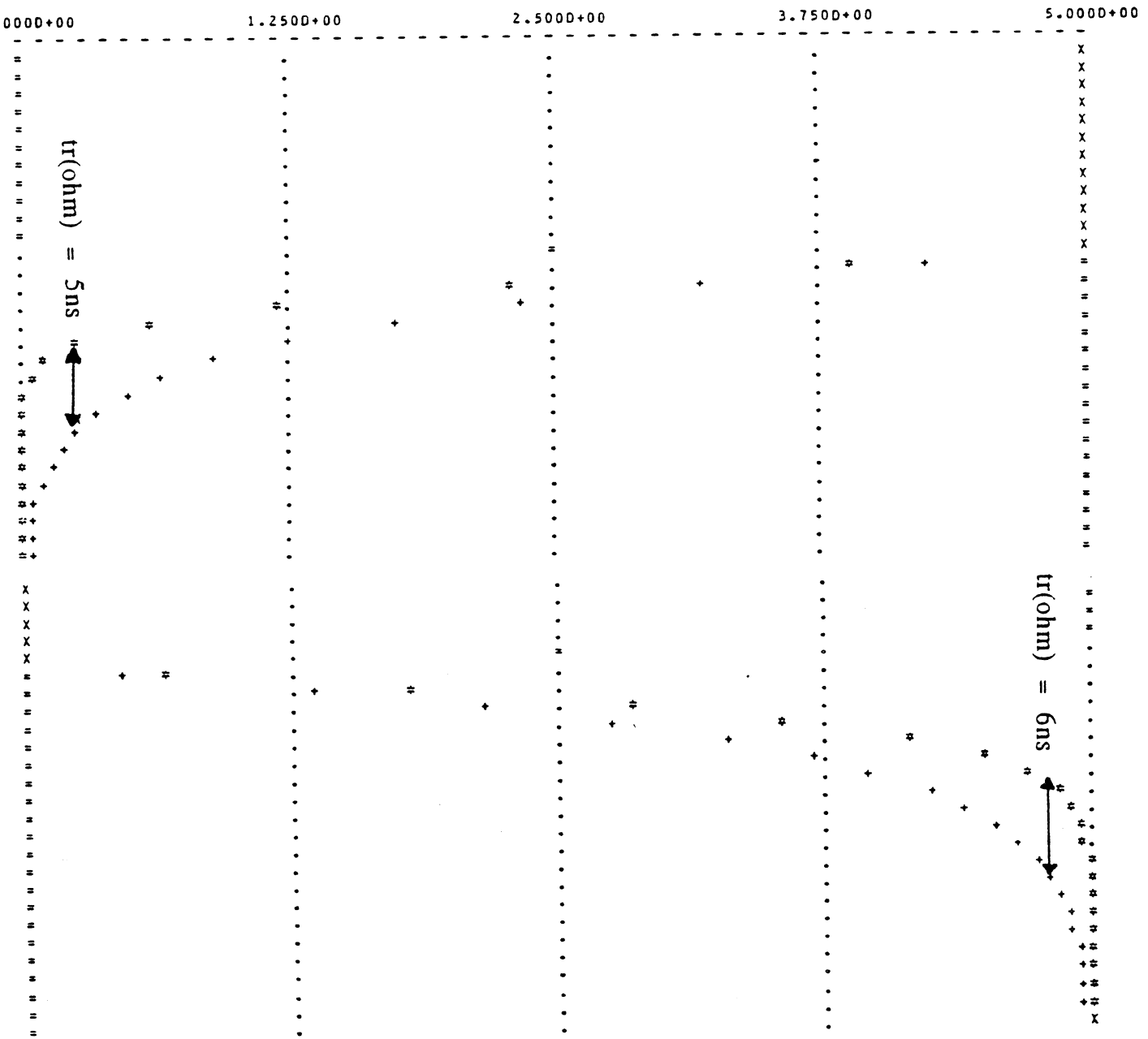
```
**********************************************************************


.MODEL MODN NMOS LEVEL=2 VTO=.6 NSUB=9.4E14 UO=800
+CGSO=4E-11 CGDO=4E-11 CGBO=2E-10 CBD=10FF CBS=10FF
+TOX=.6E-7 UCRIT=1E4 UTRA=.3 XJ=1U
+LD=0.5U UEXP=0.1 CJ=4FF VMAX=5E4 DELTA=1
.MODEL MODP PMOS LEVEL=2 VTO=-.9 NSUB=9.4E14 UO=200
+CGSO=4E-11 CGDO=4E-11 CGBO=2E-10 CBD=10FF CBS=10FF
+TOX=.6E-7 UCRIT=1E4 UTRA=.3 XJ=1U
+LD=0.5U UEXP=0.1 CJ=4FF VMAX=5E4 DELTA=1
M1   1   3   2   1   MODP   5U   15U   10P   10P
M2   2   3   0   0   MODN   5U   5U   10P   10P

C1   2   0   0.2P
M3   1   3   4   1   MODP   5U   15U   10P   10P
M4   4   3   0   0   MODN   5U    5U   10P   10P
R1   4   5   10K
C2   5   0   0.2P
VDD  1   0   5
VIN  3   0   PULSE(0,5,10N,2N,2N,46N,100N)
.TRANS 1N 100N
.PLOT TRANS V(2) V(5) V(3) (0,5)
.END
```

#: V(2)
+: V(5)
=: V(3)

TIME	V(2)
0.0000D+00	5.000D+00
1.000D-09	5.000D+00
2.000D-09	5.000D+00
3.000D-09	5.000D+00
4.000D-09	5.000D+00
5.000D-09	5.000D+00
6.000D-09	5.000D+00
7.000D-09	5.000D+00
8.000D-09	5.000D+00
9.000D-09	5.000D+00
1.000D-08	5.000D+00
1.100D-08	4.988D+00
1.200D-08	3.918D+00
1.300D-08	2.310D+00
1.400D-08	1.211D+00
1.500D-08	5.751D-01
1.600D-08	2.353D-01
1.700D-08	9.873D-02
1.800D-08	4.078D-02
1.900D-08	1.710D-02
2.000D-08	7.138D-03
2.100D-08	2.799D-03
2.200D-08	1.363D-03
2.300D-08	4.081D-04
2.400D-08	1.917D-04
2.500D-08	3.570D-05
2.600D-08	1.678D-05
2.700D-08	3.145D-06
2.800D-08	1.496D-06
5.500D-08	-4.316D-07
5.600D-08	-1.764D-07
5.700D-08	1.281D-07
5.800D-08	5.447D-07
5.900D-08	3.539D-04
6.000D-08	6.504D-01
6.100D-08	1.823D+00
6.200D-08	2.828D+00
6.300D-08	3.564D+00
6.400D-08	4.137D+00
6.500D-08	4.490D+00
6.600D-08	4.688D+00
6.700D-08	4.837D+00
6.800D-08	4.907D+00
6.900D-08	4.953D+00
7.000D-08	4.974D+00
7.100D-08	4.987D+00
7.200D-08	4.993D+00
7.300D-08	4.996D+00
7.400D-08	4.998D+00
7.500D-08	4.999D+00
7.600D-08	4.999D+00
7.700D-08	5.000D+00
7.800D-08	5.000D+00
7.900D-08	5.000D+00
8.000D-08	5.000D+00



ANNEXE 1.3

l'annexe 1.3 montre par simulation (SPICE) la validitée des schémas électriques équivalents d'une matrice pendant son temps de montée (chapitre I, § 2.2.2, figure I.23).

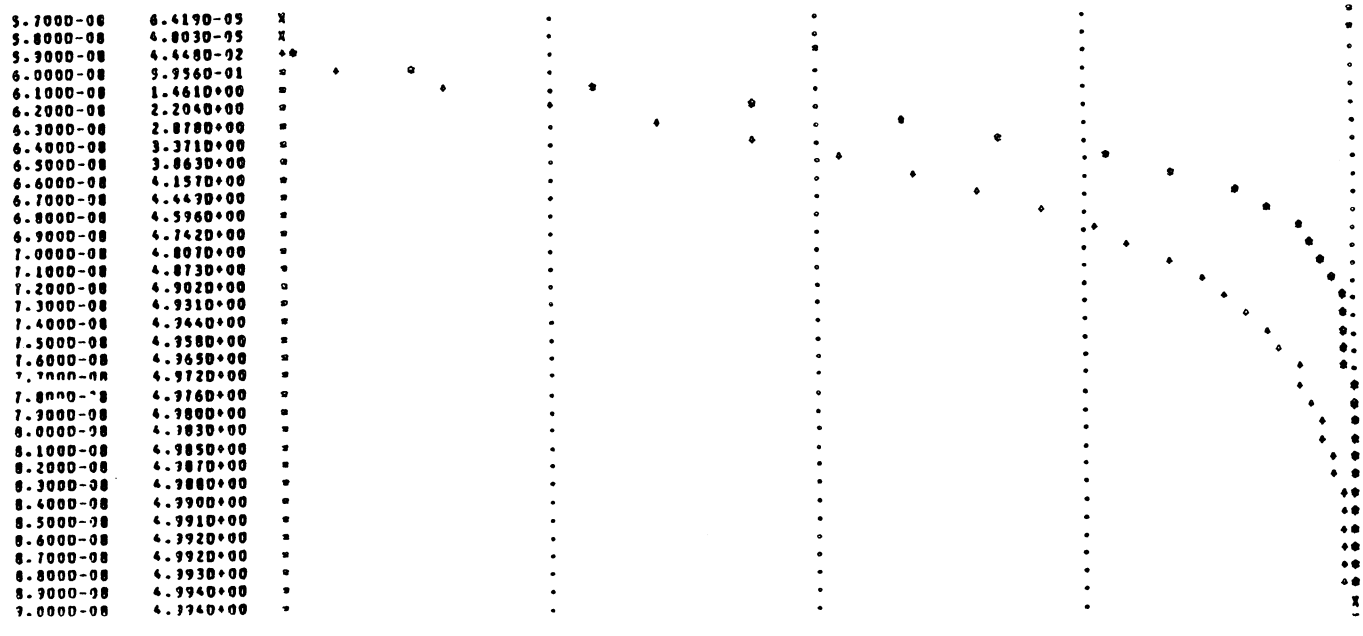
En effet,

* représente l'evolution de sortie ,de la figure I.23a ;

+ représente l'evolution de sortie de la figure I.23b.

```
.MODEL MODN NMOS LEVEL=2 VTO=.6 NSUB=9.4E14 UO=800
+CGSO=4E-11 CGDO=4E-11 CGBO=2E-10 CBD=10FF CBS=10FF
+TOX=.6E-7 UCRIT=1E4 UTRA=.3 XJ=1U
+LD=0.5U UEXP=0.1 CJ=4FF VMAX=5E4 DELTA=1
.MODEL MODP PMOS LEVEL=2 VTO=-.9 NSUB=9.4E14 UO=200
+CGSO=4E-11 CGDO=4E-11 CGBO=2E-10 CBD=10FF CBS=10FF
+TOX=.6E-7 UCRIT=1E4 UTRA=.3 XJ=1U
+LD=0.5U UEXP=0.1 CJ=4FF VMAX=5E4 DELTA=1
M1   1   3   2   1   MODP   5U   15U   10P   10P
M2   2   1   4   0   MODN   5U   5U    10P   10P
M3   5   3   0   0   MODN   5U   5U    10P   10P

C1   2   0   0.2P
R1   4   5   5K
C2   5   0   0.1P
M4   1   3   6   1   MODP   5U   10U   10P   10P
M5   6   3   0   0   MODN   5U    5U   10P   10P
R2   6   7   5K
C3   7   0   0.3P
VDD  1   0   5
VIN  3   0   PULSE(0,5,10N,2N,2N,46N,100N)
.TRANS 1N 100N
.PLOT TRANS V(2) V(7) V(3) (0,5)
.END
```



ANNEXE 2

- PLA d'origine (10 entrées, 59 monômes et 42 sorties)

```
MATRICE
.0.....00. 00000000001100000000000000000000000000000000
...0...00. 00000000001100000000000000000000000000000000
00000..... 00000000000000001000000000010000001000000000
.0...0...0 00000001000000000000000000000000000000000000
..0....00. 00000000001100000000000000000000000000000000
..0..0...0 00000001000000000000000000000000000000000000
.......00. 01000000000000000000000000000000000000000000
0...0....0 00000001000000000000000000000000000000000000
0....0...0 00000001000000000000000000000000000000000000
..0......1 00000010000000000000000000000000000000000000
0........1 00000010000000000000000000000000000000000000
.0.......1 00000010000000000000000000000000000000000000
.......01. 00001000001100000000000000000000000000000000
........1. 00000000100000000000000000000000000000000000
..0....10. 00000000000100000000000000000000000000000000
.......1.. 00000000100000000000000000000000000000000000
.0.....10. 00000000001000000000000000000000000000000000
.......10. 00000000010000000000000000000000000000000000
.......11. 00110000011000000000000000000000000000000000
..0.1....0 00000001000000000000000000000000000000000000
.0..1....0 00000001000000000000000000000000000000000000
00001..... 00000000000000010000000000100000000000000000
.......10. 00000000000100000000000000000000000000000000
0001...... 00000000000000010000000000100000001000000000
...1...10. 00000000000100000000000000000000000000000000
..1....... 00000000000000000000000000001000000000000000
.01......0 00000001000000000000000000000000000000000000
00.0.1.... 00000000000000000111000000000000000000000000
0010.1.... 00000000000000000110000000000000000000000000
001100.... 00000000000000001100000000100000000000000000
001101.... 00000000000011000000000000100000000000000000
00111..... 10000000000000001100000000000000000000000000
.10......0 00000001000000000000000000000000000000000000
010000.... 10000000000000000100011101000001000000000000
.1........ 00000000000000000000000000010000000000000000
0100010... 00000000000000000100010011000001000000000000
0100011... 00000000000000000100001101000001000000000000
01001..... 00000000000000000100000011000001000000000000
0101...... 00000000000000000100001001000001000000000000
011....... 00000000000001000001100000000001000000000000
01101..00. 00000000010000000000000000000000000000000000
100000.... 00000000000000000000000000000000000000001000
1......00. 00000000001100000000000000000000000000000000
1......... 00000000000000000100000000000000000000000000
10000..... 00000000000000010000000000000000000000000000
1......10. 00000000000100000000000000000000000000000000
100001.... 00000010000000000000000000001000000000001101
100100.... 00000000000000010000000000000000000000110000
100101.... 00000010000000010000000000001000000000110101
10011..... 00000000000000000000000000000000000000001010
101000.... 00000000000000010000000000001010000000000000
101001.... 00000000000000010000000000001100000000000000
101010.... 00000000000000010000000000001010000000000000
101011.... 00000000000000010000000000001001000000000000
10110..... 00000000000000010000000000000000000000000000
10111..... 00000000000000010000000000000001000000000000
1100...... 00000000000001100000000000000000000000100000
1101...... 00000000000000000000000000000000000000000010
111....... 00000100000000000100000000000000001000001
```

- partitionnement en 3 PLA : PLA1 (7, 16, 8) PLA2 (7, 19, 9)

PLA3 (8, 26, 25)

MATRICE

```
.0.....00. 00000011
...0...00. 00000011
..0....00. 00000011
.......00. 00010000
.......01. 01000011
........1. 00001000
..0....10. 00000001
.......1.. 00001000
.0.....10. 00000001
.......10. 00000100
.......11. 10100110
.......10. 00000001
...1...10. 00000001
01101..00. 00000100
1......00. 00000011
1......10. 00000001
```

MATRICE

```
00000..... 010000101
.0...0...0 000000010
..0..0...0 000000010
0...0....0 000000010
0....0...0 000000010
..0.1....0 000000010
.0..1....0 000000010
00001..... 000000101
0001...... 100000101
.01......0 000000010
001100.... 000000100
001101.... 000000100
.10......0 000000010
101000.... 001010001
101001.... 000110001
101010.... 001001001
101011.... 000101001
10110..... 000000001
10111..... 000001001
```

MATRICE

```
..0......1 00000000000000000000000100
0........1 00000000000000000000000100
.0.......1 00000000000000000000000100
..1....... 00000000000000000001000000
00.0.1.... 00000000001000000000100001
0010.1.... 00000000001000000000100000
001100.... 00000000000100000000000001
001101.... 00100000000000000010000000
00111..... 00000000000110000000000001
010000.... 00000000110010001000010011
.1........ 00000000000000000001000000
0100010... 00000001010000000000001011
0100011... 00000000100000001000010011
01001..... 00000001000000000000001011
0101...... 00000000000000001000010011
011....... 01000000000000000010100010
100000.... 00000000000000010000000000
1........ 00000000000000000000000001
10000..... 00000000000000000000010000
100001.... 00000100000001100010001000
100100.... 00000010000000010000010000
100101.... 00000110000001010010101000
10011..... 00001000000000010000000000
1100...... 00000000000000001010010000
1101...... 00001000000000000000000000
111....... 10010000000001000000000001
```

- partitionnement en 4 PLA : PLA1 (7, 13, 6) PLA2 (7, 14, 16)
PLA3 (7, 14, 10) PLA4 (7, 22, 10)

MATRICE

```
.0.....00. 000011
...0...00. 000011
..0....00. 000011
.......01. 010011
..0....10. 000001
.0.....10. 000001
.......10. 000100
.......11. 101110
.......10. 000001
...1...10. 000001
01101..00. 000100
1......00. 000011
1......10. 000001
```

MATRICE

```
00.0.1.... 0000000100001001
0010.1.... 0000000100001000
001100.... 0000000010000001
001101.... 0010000000010000
00111..... 0000000011000001
010000.... 0000011001100111
0100010... 0000101000000111
0100011... 0000010000100111
01001..... 0000100000000111
0101...... 0000000000100111
011....... 0100000000011010
1......... 0000000000000001
1100...... 0000000000010000
111....... 1001000000000001
```

MATRICE

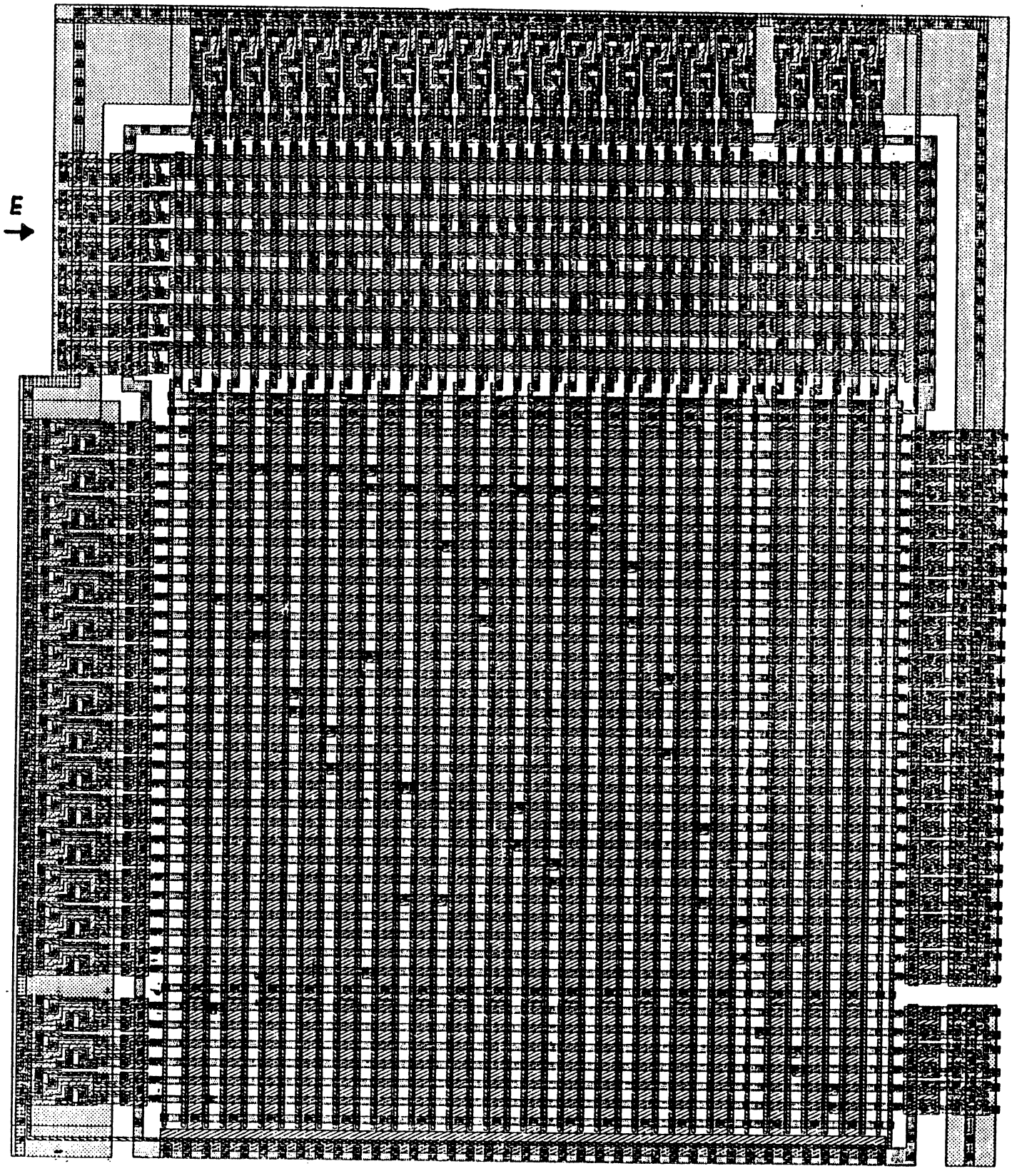
```
00000..... 0100000011
.......00. 0010000000
........1. 0000001000
.......1.. 0000001000
00001..... 0000000011
0001...... 1000000011
001100.... 0000000010
001101.... 0000000010
101000.... 0001010001
101001.... 0000110001
101010.... 0001000101
101011.... 0000100101
10110..... 0000000001
10111..... 0000000101
```

MATRICE

```
.0...0...0 0000000001
..0..0...0 0000000001
0...0....0 0000000001
0....0...0 0000000001
..0......1 0000000010
0........1 0000000010
.0.......1 0000000010
..0.1....0 0000000001
.0..1....0 0000000001
..1....... 0000001000
.01......0 0000000001
.10......0 0000000001
.1........ 0000001000
100000.... 0000100000
10000..... 0000000100
100001.... 0101101010
100100.... 0010010100
100101.... 0111011110
10011..... 1000100000
1100...... 0000010100
1101...... 1000000000
111....... 0001000000
```

ANNEXE 3

- dessin de masques d'un PLA CMOS.



INPG département des études doctorales

AUTORISATION de SOUTENANCE

VU les dispositions de l'article 15 Titre III de l'arrêté du du 5 juillet 1984 relatif aux études doctorales

VU les rapports de présentation de Messieurs
. J. LARDY, Ingénieur
. G. SAGNES, Professeur

Monsieur DANDACHE Abbas

est autorisé à présenter une thèse en soutenance en vue de l'obtention du diplôme de DOCTEUR de L'INSTITUT NATIONAL POLYTECHNIQUE DE GRENOBLE, spécialité "Microélectronique".

Fait à Grenoble, le 26 juin 1986

D. BLOCH
Président
de l'Institut National Polytechnique
de Grenoble

P.O. le Vice-Président,

RESUME

Cette thèse concerne plus particulièrement les **PLA CMOS**. Les quatre aspects suivants sont développés :

- performance électrique : spécification d'outil d'évaluation électrique et temporelle de PLA par une technique hybride estimation-simulation, basée sur la recherche du chemin critique d'E/S dans le PLA ;

- distribution des types de pannes en fin de fabrication et leurs manifestations électriques et logiques. Une approche vers le test de PLA CMOS est également présentée ;

- amélioration du rendement de fabrication par la conception de PLA reconfigurable (ajout de lignes supplémentaires) ;

- partitionnement de PLA, en vue de réduire la surface, le temps de réponse, et de faciliter la reconfiguration et l'interconnexion avec les blocs voisins.

MOTS CLES: VLSI - PLA - performance - test - reconfiguration - partitionnement - CAO.