



INSTITUT NATIONAL POLYTECHNIQUE DE GRENOBLE

**THÈSE**

pour obtenir le grade de

**DOCTEUR DE L'INPG**

*Spécialité : Micro et Nano Electronique*

préparée au sein du groupe de Microélectronique du  
*Laboratoire Européen pour la Recherche Nucléaire (CERN)*

dans le cadre de l'École Doctorale « **Electronique, Electrotechnique, Automatique et  
Traitement du Signal** »

présentée et soutenue publiquement par

Sandro BONACINI

le 16 novembre 2007

Titre :

**Développement de circuits logiques programmables résistants aux  
aléas logiques en technologie CMOS submicrométrique**

Titre anglais :

**Development of Single-Event Upset hardened programmable logic  
devices in deep submicron CMOS**

Directeur de thèse : M. Raoul VELAZCO

**JURY**

M. Régis LEVEUGLE	Professeur, INP Grenoble	Président
M. Raoul VELAZCO	Directeur de recherche au CNRS	Directeur de thèse
M. Kostas KLOUKINAS	Ingénieur, CERN	Resp. de thèse CERN
M. Laurent DUSSEAU	Professeur, Univ. de Montpellier 2	Rapporteur
M. Sandro CENTRO	Professeur, Univ. de Padova / INFN	Rapporteur
M. Alessandro PACCAGNELLA	Professeur, Univ. de Padova	Examinateur





*In memoria di  
mio padre*



# Contents

<b>Résumé</b>	<b>1</b>
I Introduction . . . . .	1
I-i Le CERN et la physique des hautes énergies . . . . .	1
I-ii Le grand collisionneur d'hadrons . . . . .	1
I-iii L'environnement radioactif d'une expérience typique du LHC . . . . .	2
I-iv Circuits intégrés résistants aux radiations . . . . .	2
I-v Motivations et objectifs de cette thèse . . . . .	3
II Effets des radiations sur les circuits intégrés et durcissement . . . . .	4
II-i Effets de la dose totale ionisante . . . . .	4
II-ii Durcissement à la dose totale ionisante . . . . .	5
II-iii Effets non récurrents des radiations . . . . .	5
II-iv Protection contre les aléas logiques . . . . .	7
III Logiques programmables et environnement radiatif . . . . .	10
III-i Logiques programmables simples . . . . .	10
III-ii Réseau de portes programmables in-situ (FPGA) . . . . .	10
III-iii Effets des radiations sur les dispositifs programmables . . . . .	12
III-iv Techniques de protection contre les SEU pour les logiques programmables du commerce . . . . .	13
IV Un FPGA résistant aux radiations pour la HEP . . . . .	13
IV-i Dessin du bloc logique en CMOS 0.25 micron . . . . .	14
IV-ii Transposition du LB vers une technologie 0.13 micron . . . . .	19
IV-iii Développement des connections programmables . . . . .	20
V Un PLD résistant aux radiations . . . . .	20
V-i Structure . . . . .	21
V-ii Layout de la puce . . . . .	23
VI Conclusions . . . . .	23
<b>1 Introduction</b>	<b>25</b>
1.1 CERN and High Energy Physics . . . . .	25
1.1.1 Accelerators and detectors . . . . .	25
1.1.2 The Large Hadron Collider . . . . .	26
1.1.3 An example of a typical HEP experiment . . . . .	27
1.2 Radiation environment in the LHC . . . . .	30
1.2.1 Radiation environment in the experiments . . . . .	30
1.2.2 Radiation tolerant ICs . . . . .	32
1.3 Motivation and objectives of this work . . . . .	33

<b>2</b>	<b>Radiation Effects and Hardening</b>	<b>35</b>
2.1	Total Ionizing Dose effects . . . . .	35
2.1.1	Radiation effects on matter . . . . .	35
2.1.2	Radiation effects on MOS transistors . . . . .	36
2.2	Hardening against TID . . . . .	39
2.2.1	Layout techniques . . . . .	39
2.2.2	Circuit and system techniques . . . . .	42
2.2.3	Radiation tolerant digital standard cells libraries . . . . .	42
2.3	Single-Event Effects . . . . .	43
2.3.1	Single-Event Latch-up (SEL) . . . . .	43
2.3.2	Single-Event Upset (SEU) . . . . .	43
2.3.3	Critical charge simulations . . . . .	45
2.3.4	Critical LET measurement . . . . .	46
2.3.5	SEUs in finite state machines and SEFIs . . . . .	47
2.3.6	Single-Event Transients . . . . .	48
2.3.7	Multiple bit upset . . . . .	49
2.4	Protection from SEUs . . . . .	49
2.4.1	The Dual Interlocked cell . . . . .	50
2.4.2	The Whitaker cell . . . . .	52
2.4.3	The SERT cell . . . . .	54
2.4.4	Other SEU-hardened memory cells . . . . .	54
2.4.5	Temporal redundancy . . . . .	54
2.4.6	Triple Module Redundancy . . . . .	56
2.4.7	The TREVOTE cell . . . . .	58
2.4.8	Dual-rail logic . . . . .	59
2.4.9	Coding techniques . . . . .	61
2.4.10	High-capacitance signals . . . . .	63
<b>3</b>	<b>Programmable logic and radiation environment</b>	<b>67</b>
3.1	Brief history of programmable logic . . . . .	67
3.1.1	PROM devices . . . . .	67
3.1.2	PLDs . . . . .	67
3.1.3	CPLDs . . . . .	69
3.1.4	MPGAs . . . . .	69
3.2	Field-programmable gate arrays . . . . .	70
3.2.1	Logic block architecture . . . . .	70
3.2.2	Routing architecture . . . . .	71
3.2.3	I/O blocks . . . . .	72
3.2.4	Programming technique . . . . .	72
3.2.5	Special-purpose blocks . . . . .	74
3.3	FPGAs in radiation environment . . . . .	74
3.4	SEU hardening techniques for commercial devices . . . . .	75
3.4.1	Triple module redundancy . . . . .	75
3.4.2	Reconfiguration . . . . .	76
<b>4</b>	<b>A radiation-tolerant FPGA for HEP</b>	<b>77</b>
4.1	Logic block implementation in 0.25 micron CMOS . . . . .	78
4.1.1	The look-up table . . . . .	78
4.1.2	The carry and wide-fanin logic block . . . . .	84
4.1.3	The user register . . . . .	87
4.1.4	The configuration block . . . . .	87

4.1.5	LB pairs and modules . . . . .	87
4.1.6	Test chip in 0.25 micron technology . . . . .	89
4.1.7	SEU hardening of I/O pads and global signals . . . . .	92
4.1.8	Simulation . . . . .	93
4.1.9	Packaging . . . . .	93
4.1.10	Functional testing . . . . .	94
4.1.11	Ion beam testing procedures . . . . .	95
4.1.12	Test board for ion beam testing . . . . .	96
4.1.13	Ion beam test results . . . . .	101
4.2	Migration of the LB design to 0.13 micron . . . . .	102
4.2.1	Single interleaved SEU-robust register . . . . .	103
4.2.2	Double interleaved SEU-robust register . . . . .	105
4.2.3	Test chip for evaluation of SEU-robust structures . . . . .	105
4.2.4	Testing procedures . . . . .	106
4.2.5	Ion-beam test results . . . . .	106
4.3	Development of the FPGA interconnectivity . . . . .	108
4.3.1	Switch matrix architecture . . . . .	108
<b>5</b>	<b>A radiation-tolerant PLD</b>	<b>113</b>
5.1	Structure . . . . .	113
5.1.1	The logic block . . . . .	114
5.1.2	The fuse storage cell . . . . .	117
5.1.3	The AND matrix . . . . .	120
5.1.4	The transition detector . . . . .	121
5.1.5	Tri-state I/O pad design . . . . .	121
5.1.6	Chip layout . . . . .	124
<b>6</b>	<b>Conclusions</b>	<b>125</b>
<b>A</b>	<b>Memory cell layout for SEU-robustness</b>	<b>127</b>
	<b>Bibliography</b>	<b>129</b>
	<b>List of publications</b>	<b>137</b>





# Résumé

## I Introduction

### I-i Le CERN et la physique des hautes énergies

La physique des hautes énergies (HEP) explore les constituants de base de la matière et de leurs interactions mutuelles. Le CERN, le Laboratoire Européen pour la Physique des Particules, a été fondé en 1954 à Genève (Suisse) dans un effort européen commun de fournir un service scientifique important pour les physiciens des particules.

Les études de physique des particules sont basées sur des collisions de particules à énergie cinétique élevée, ce qui signifie que les particules utilisées dans les expériences doivent avoir une haute vitesse. Les accélérateurs de particules, comme les synchrotrons, sont employés pour atteindre la vitesse requise.

Les résultats d'une collision sont observés par un détecteur qui est disposé tout autour de l'expérience. Un détecteur se compose habituellement de plusieurs sous-détecteurs avec différentes capacités et différents buts, et tous sont reliés à un système informatique pour la reconstitution et l'analyse des événements. Le but est d'identifier, compter et tracer, le plus grand nombre possible de particules générées par la collision.

### I-ii Le grand collisionneur d'hadrons

Le grand collisionneur de hadrons (Large Hadron Collider, LHC) est un accélérateur, actuellement en construction au CERN, conçu pour les collisions de protons atteignant des énergies de collision jusqu'à 14 TeV. Il est nécessaire d'atteindre des énergies élevées pour recréer les conditions primordiales de l'univers pendant le Big Bang. Ainsi, plus l'énergie des collisions que nous parvenons à créer est élevée, plus petite est la dimension que nous étudions, et plus loin dans le temps nous pouvons observer. Le LHC est construit dans un tunnel souterrain de 27 km de longueur. Une si grande circonférence est nécessaire en raison de la déperdition d'énergie par rayonnement continu de freinage (Bremsstrahlung). Les deux faisceaux de proton voyageront en directions opposées, mais se heurteront seulement en quatre points, où les expériences ont été construites.

Les faisceaux seront segmentés en 2835 groupes et chaque groupe aura  $1.1 \cdot 10^{11}$  particules. Deux groupes allant en directions opposées se croiseront dans les points d'interaction chaque 24.95 ns à la vitesse nominale. En d'autres termes la fréquence de collision sera de 40.08 MHz.

Une des quatre expériences, le Compact Muon Solenoid (CMS), sera traité plus en détail dans la section suivante comme exemple typique d'une expérience.

### I–iii L’environnement radioactif d’une expérience typique du LHC

Le CMS a la forme d’un cylindre avec un diamètre de 14.6 m et une longueur de 21.6 m. Son poids total est d’environ 14500 tonnes. La fig. 1.4 page 29 en montre une représentation. Les faisceaux entrent des deux côtés et se heurtent au centre du détecteur, le point également appelé vertex.

Seulement une petite fraction des collisions sera intéressante du point de vue de la physique, donc un filtrage des données doit être effectué. Il est également nécessaire de faire ce filtrage en temps réel. Toutes ces opérations sont traitées par les systèmes d’acquisition de donnée et de trigger de l’expérience, qui choisit les événements utiles en évaluant un sous-ensemble des données.

Afin de maximiser le nombre d’événements intéressants obtenus avec les expériences, l’accélérateur LHC est conçu pour atteindre une luminosité<sup>1</sup> maximale très élevée qui apportera, dans le cas des protons, une production moyenne de  $8 \cdot 10^8$  collisions inélastiques proton-proton par seconde, créant un environnement extrêmement hostile en terme de radiations.

De plus, pour le LHC, l’énergie élevée du faisceau combinée avec la luminosité très élevée créent de nombreuses cascades intenses de particules, qui finiront en un immense nombre de particules à basses énergie. En réalité, les particules excédant 10 GeV devraient être très rares dans les détecteurs. Par conséquent les études de radiation ont été concentrées sur la gamme d’énergie autour de 1 GeV et en dessous.

Approximativement 30% des interactions hadroniques inélastiques créent des radionucléides longévitaux qui contribuent au taux de dose par radioactivité induite dans le secteur expérimental. L’activation peut également se produire par des interactions de neutrons, particulièrement dans le régime thermique.

Comme résumé dans la table 1.1 page 32, les valeurs de dose totale ionisante (TID) dans l’expérience CMS peuvent être hautes, dans les conditions le plus défavorables, jusqu’à 50 Mrad après les 10 années de vie prévue de l’expérience. L’électronique de front-end des détecteurs doit alors supporter cette énorme quantité de radiation, particulièrement dans la partie intérieure.

### I–iv Circuits intégrés résistants aux radiations

En conséquence les circuits intégrés utilisés pour l’électronique de front-end des détecteurs doivent être très résistants aux radiations. La nécessité de ce genre de circuits pour les diverses applications mentionnées a mené, dans le passé, au développement de technologies spéciales, mais la modification des étapes de processus est coûteuse.

Dans un transistor métal-oxyde-semi-conducteur (MOS), la partie la plus sensible aux effets des radiations est l’oxyde de la grille. Une manière de réduire ces effets est de réduire l’épaisseur de la grille, qui est la tendance normale en technologie moderne. Les dispositifs d’aujourd’hui, bien au dessus du micron de dimension, ont une épaisseur d’oxyde de grille de moins de 2 nm. Cela suggère la possibilité d’employer des technologies modernes CMOS dans l’environnement de radiations sans ajouter ou modifier aucune étape du processus de fabrication.

Pour cette raison en 1996 le groupe de microélectronique du CERN a commencé à étudier la possibilité d’employer une technologie commerciale CMOS pour intégrer les circuits à employer dans les détecteurs. À ce moment-là la technologie  $0.7 \mu\text{m}$

---

<sup>1</sup>La luminosité est assimilable au nombre de particules par unité de superficie dans le point d’interaction des deux faisceaux.

était utilisée, mais depuis l'évolution de ces technologies a été suivie en caractérisant les technologies 0.5, 0.35, 0.25 et actuellement 0.13  $\mu\text{m}$ .

## I–v Motivations et objectifs de cette thèse

Les progrès en technologies microélectroniques appliquées aux circuits logiques programmables ont diminué le coût et le temps de développement de l'électronique numérique dans le secteur industriel aussi bien dans les secteurs de l'espace que de l'aéronautique. L'utilisation de tels dispositifs est également intéressant pour les détecteurs de l'HEP placés à proximité des accélérateurs de particules tels que le LHC. Comme mentionné précédemment, la présence de radiations dans ces détecteurs rend les composants commerciaux disponibles inutilisables et exige la conception de circuits spéciaux. Le chapitre 2 introduit les effets des radiations sur les circuits intégrés et des solutions de durcissement contre ces effets.

Les circuits programmables les plus avancés sont les FPGAs, qui seront présentés au chapitre 3. Les FPGAs basés sur des SRAM sont flexibles et peuvent répondre à des exigences multiples. Ils peuvent être modifiés après la réalisation des systèmes pour corriger des erreurs de dessin ou pour améliorer les performances. Les FPGAs basés sur SRAM peuvent être fabriqués dans des processus standard CMOS.

De nombreuses études ont été faites à propos des effets de rayonnement sur les FPGAs commerciaux, qui ont montré leur sensibilité à la dose totale ionisante et aux aléas logiques. Les résultats de ces études seront présentés dans la section 3.3.

La sensibilité aux SEUs des FPGAs est due à la grande quantité d'éléments de mémoire situés dans ces dispositifs. Ceux-là doivent être fortement protégés pour éviter des erreurs pendant l'exécution. Il y a deux techniques principales pour atténuer ces effets de radiations : introduire une redondance dans le programme en langage de description d'hardware (HDL) ou durcir au niveau architectural les cellules. La première technique réduit beaucoup les ressources disponibles du FPGA et exige des circuits complexes de reconfiguration pour éviter les changements dans la configuration. A la différence de cette approche, l'objectif de ce travail est le développement des circuits programmables où l'insensibilité aux SEU est intégrée au niveau de cellules de mémoire, n'exigeant pas de la part de l'utilisateur d'utiliser une technique particulière pour la protection contre les aléas logiques.

Les dispositifs logiques programmables (PLDs) sont de petits composants qui peuvent mettre en application des fonctions de logique équivalentes à approximativement 50 portes logiques. Bien que les PLDs soient considérés comme dépassés par les FPGAs, ils sont encore avantageux dans quelques applications pour créer de simple machines à état et pour corriger des systèmes dans les étapes finales de projet. Les PLDs souffrent également des effets de la dose totale ionisante. Les PLDs en général ne sont pas affectés par les aléas logiques dans le stockage de configuration, mais le registre d'utilisateur peut encore être perturbé et il est donc nécessaire de le protéger.

Ce travail s'occupe de la conception d'un FPGA basé sur SRAM et d'un PLD basé sur des fusibles qui sont robustes aux SEUs, résistants aux radiations et compatibles avec des composants industriels, afin de fournir à la communauté de HEP deux dispositifs appropriés pour les expériences de physique de particules.

Afin d'atteindre les caractéristiques désirées, plusieurs techniques durcissantes ont été évaluées, comme présenté dans la section 2.4, et une approche finale a été choisie et mise en application dans plusieurs puces d'essai pour l'évaluation. Des essais ont été effectués avec un faisceau d'ions lourds et les résultats sont présentés dans les chapitres 4 et 5.

## II Effets des radiations sur les circuits intégrés et durcissement

### II-i Effets de la dose totale ionisante

Les transistors MOS ne sont presque pas sensibles aux dommages de déplacement dus aux radiations, puisque leur conduction est basée sur les porteurs majoritaires au-dessous de l'interface silicium-oxyde, une région qui ne se prolonge pas en profondeur. L'ionisation crée des paires électron-trou. Le nombre de paires créées est directement proportionnel à toute la dose absorbée. Pour cette raison, les études sur les effets de l'ionisation se réfèrent seulement à la dose totale et non au type de particule et à son énergie.

*Charge positive  
emprisonnée dans  
l'oxyde*

En cas de polarisation positive appliquée à la grille, les électrons créés par une radiation ionisante dérivent vers l'électrode de grille dans un temps très court, tandis que les trous se déplacent vers l'interface Si-SiO<sub>2</sub> par un phénomène de transport différent beaucoup plus lent. Puis, près de l'interface mais toujours dans l'oxyde, quelques trous peuvent être emprisonnés, créant ainsi une charge positive fixe d'oxyde  $Q_{ox}$ . La quantité de charge emprisonnée est proportionnelle au nombre de défauts dans le dioxyde de silicium. Les électrons peuvent passer depuis la surface du silicium dans l'oxyde par effet tunnel et se recombiner avec les trous emprisonnés (donnant l'origine à un recuit, annealing). Cet effet fait changer la quantité emprisonnée de charge avec le taux de dose absorbée et son historique.

La charge positive d'oxyde abaisse la tension de seuil  $V_T$  dans les transistors à canal N, puisqu'elle attire plus d'électrons pour provoquer l'inversion du silicium. Dans les transistors à canal P la valeur absolue de tension de seuil est augmentée, ou, en d'autres termes,  $V_T$  est plus négatif.

*Pièges à  
l'interface induite  
par les radiations*

La radiation ionisante induit aussi la création d'états d'interface. Ces pièges ont un niveau d'énergie compris dans l'intervalle d'énergie interdit du silicium. Remplir ces états provoque une charge  $Q_{it}$  emprisonnée à l'interface. Pour cette raison, dans les transistors PMOS et NMOS, le seuil augmente (en valeur absolue) avec l'irradiation.

*Variation de la  
tension de seuil*

Les deux phénomènes décrits ci-dessus font changer la tension de seuil avec l'irradiation. Tandis que les PMOS subissent seulement une augmentation de  $V_T$ , celui-ci peut diminuer, augmenter, ou même être stable dans les NMOS.

Pour une technologie 0.25  $\mu\text{m}$ , la valeur absolue d'augmentation de  $V_T$ , montrée en fig. 2.3 page 38, est de toute façon inférieure à 80 mV après une dose d'irradiation de 30 Mrad. Dans la technologie 0.13  $\mu\text{m}$ , la charge  $Q_{ox}$  donnée par les bords du STI domine à faibles doses ( $\approx 40$  Mrad) tandis que  $Q_{it}$  domine à plus haute dose, donnant une courbe  $V_T$ , montrée dans la fig. 2.4 page 38, qu'initialement diminuait et après recuits. Cette figure montre qu'il est préférable d'utiliser des transistors larges qui souffrent moins des effets du bord STI.

*Augmentation du  
courant de fuite*

Puisque l'oxyde de champ est beaucoup plus épais que l'oxyde de grille, il est plus sensible à la charge positive emprisonnée induite par une radiation ionisante. Un chemin parasite peut ainsi être formé près des côtés de la grille des transistors à canal N qui relie le drain à la source, et augmente le courant de fuite. Une technologie 0.25  $\mu\text{m}$  standard peut être employée sans aucune technique spéciale jusqu'à

200 krad [Faccio 98]. Dans une technologie  $0.13 \mu\text{m}$  la courant de fuite recuit à haute dose [Faccio 05] et cela suggère la possibilité d'employer la technologie sans aucune technique spéciale de dessin.

## II–ii Durcissement à la dose totale ionisante

Le choix d'employer une technologie submicrométrique profonde garantit un oxyde de grille résistant aux radiations. Ce qui est donc nécessaire est de résoudre les problèmes liés à la dégradation de l'oxyde de champ des dispositifs à canal N.

Une solution possible est d'utiliser des transistors à structure fermée (enclosed layout transistors, ELTs). Comme montré en fig. 2.7 page 41, dans ce cas on élimine le chemin parasite entre la source et le drain. Les inconvénients principaux de cette structure sont une plus grande surface et une plus grande capacité. D'ailleurs, le choix sur le rapport de  $W/L$  est limité, puisque  $W$  doit être assez grand pour permettre au contact actif intérieur d'être placé.

*Transistors à structure fermée (ELT)*

Le problème du courant de fuite entre différents dispositifs est résolu en entourant chaque dispositif à canal N avec un anneau de garde P+ [Anelli 00]. Cette méthode s'est révélée très efficace mais l'inconvénient est encore la grande surface utilisée. D'ailleurs, les anneaux de garde évitent la génération des SELs (expliquée après) en abaissant le gain du transistor bipolaire parasite NPN.

*Anneaux de garde*

Une librairie de cellules numériques standard résistantes aux radiations a été conçue et examinée dans une technologie  $0.25 \mu\text{m}$  [Marchioro 98, Kloukinas 98] tandis qu'une librairie commerciale est utilisée dans la technologie  $0.13 \mu\text{m}$ . Les librairies contiennent des portes logique combinatoires, comme des NON-ETs (NANDs) et NON-OUs (NORs), aussi bien que des bascules (flip-flops) et des verrous (latches). Un ensemble de ports d'entrée-sortie est également disponible.

*Librairie de cellules numériques résistantes aux radiations*

## II–iii Effets non récurrents des radiations

Les effets non récurrents (Single-Event Effects, SEE) [Kerns 89] sont des phénomènes produits par une seule particule fortement énergétique passant par le dispositif. La particule produit une trace d'ionisation, de longueur dépendant du nombre atomique et de l'énergie initiale, où des porteurs mobiles de charge sont créés.

Le verrouillage maintenu (Single-Event Latch-up, SEL) est un effet destructif qui peut se produire en raison du thyristor parasite constitué par la structure complexe de jonctions établie dans tous les circuits intégrés CMOS (voir la fig. 2.9 page 44). Ce phénomène est habituellement évité avec des techniques de fabrication et de dessin, comme par exemple en plaçant les contacts de substrat (ou well) très près de la source des dispositifs. Il peut se produire qu'une particule énergétique ionisante passant par un dispositif dépose charge une à l'intérieur du thyristor parasite, déclenchant la rétroaction positive.

*Verrouillage maintenu (SEL)*

L'aléa logique (Single-Event Upset, SEU) est un effet réversible (non destructif), qui consiste en la modification de l'état logique d'une cellule de mémoire. Dans les dispositifs modernes, l'information est habituellement emmagasinée comme une quantité de charge. Une particule ionisante croisant la région d'épuisement de drain

*Aléas logique (SEU)*

d'un dispositif crée des paires électron-trou qui sont rassemblées par le champ électrique. La charge rassemblée modifie la tension sur le nœud du circuit de drain, corrompant l'information. Les dispositifs à canal N accumulent seulement les électrons, donc les charges négatives, alors que les dispositifs à canal P ne rassemblent que des trous, donc les charges positives.

### *Charge critique*

La charge déposée change la valeur dans le nœud atteint seulement si elle excède un seuil particulier appelé *charge critique*, qui dépend du type de circuit et de sa capacité à répondre à un courant induit. Par exemple, un nœud à haute impédance n'a aucun composant actif qui peut fournir le courant pour la restauration de la tension correcte, ainsi il est très sensible aux SEUs. Au niveau du circuit, la logique dynamique, où l'information est emmagasinée dans des nœuds à haute impédance, est plus sensible aux SEUs que la logique statique, où l'information est emmagasinée dans des nœuds à basse impédance.

A la différence de nombreux autres effets induits par les radiations, la sensibilité aux SEUs augmente avec la réduction des dimensions des transistors VLSI : en fait, la charge critique est proportionnelle à la capacité du nœud et à la tension d'alimentation, toutes les deux réduites avec les dimensions du dispositif.

### *Transfert linéaire d'énergie (LET) critique*

La quantité d'énergie déposée par une particule par unité de longueur de trace peut être exprimée en termes de *transfert d'énergie linéaire* (Linear Energy Transfer, LET), avec unité  $\text{cm}^2\text{MeV}/\text{mg}$ , qui est la perte d'énergie par unité de longueur  $dE/dx$  divisée par la densité du matériel (en ce cas, Si). Le LET dépend du numéro atomique et de l'énergie de la particule incidente. Fondamentalement, plus haut est le numéro atomique de la particule chargée, plus haut est son LET, alors que la relation avec l'énergie est plus complexe. Les ions légers n'ont généralement pas un LET suffisamment haut pour induire des SEUs directement, mais ils peuvent provoquer des réactions nucléaires produisant des isotopes secondaires avec un numéro atomique plus élevé et donc un LET plus haut. De cette façon, les protons, les neutrons et les particules alpha induisent des erreurs par réaction nucléaire.

Le paramètre expérimental de caractérisation pour les SEUs est le LET critique, qui décrit combien de charge doit être déposée pour produire un aléa logique. Une courbe section transversale / LET montre souvent une étape correspondant au LET critique (un exemple est donné fig. 2.13 page 47), qui peut être défini rigoureusement comme la valeur LET donnant une section transversale de 10% de la section transversale maximale.

### *Aléas logiques dans les machines à états finis*

La logique contenue dans un circuit intégré spécialisé (Application Specific Integrated Circuit, ASIC) peut habituellement être divisée en deux classes : le *chemin de données* (datapath), qui constitue une structure canalisée responsable de faire les calculs sur les données d'entrée et d'apporter les résultats en sortie, et les *machines à états finis* (Finite State Machines, FSMs) qui sont chargées de commander le datapath et traiter par des protocoles spécifiques la logique en dehors de la puce.

La gravité d'un aléa logique dépend souvent du temps de mise en hors service provoqué, qui dépend du genre de logique qui a été frappée. Si l'aléa logique se produit dans un datapath, l'erreur se propage avec les données et est apportée rapidement hors de la puce. Quand un aléa logique se produit dans les machines à état de commande, celles-ci peuvent entrer dans des faux états et exécuter des opérations inattendues qui peuvent durer longtemps, perturbant le datapath et le système entier. Dans les pires cas, une machine à états peut entrer dans un cycle d'où elle ne sortira

jamais jusqu'à une remise à zéro (reset) de la puce. Ce dernier type d'échec induit par la radiation est appelé Single-Event Functional Interrupt (SEFI) [Koga 98].

Les aléas logiques peuvent également se produire dans la logique combinatoire. Tandis que le taux d'erreur dû à des aléas logiques est indépendant de la fréquence pour la logique séquentielle, il augmente linéairement avec la fréquence pour la partie combinatoire [Buchner 97, Reed 96, Wang 04]. Dans le passé, quand les périodes d'horloge étaient beaucoup plus longues que la durée des aléas logiques, ce phénomène a été rarement pris en considération, alors qu'aujourd'hui il commence à ne plus être négligé. Un aléa logique qui se produit dans la logique combinatoire s'appelle un Single-Event Transient (SET), puisque la valeur correcte de tension est immédiatement reconstituée après que l'injection de charge est terminée.

*Aléas logiques  
dans la logique  
combinatoire*

Intuitivement, la dépendance de la fréquence des SETs vient du fait que le profil de temps de collection de charge d'un nœud ne change pas avec la fréquence. En conséquence en augmentant la fréquence il y a plus de fronts montants d'horloge donc la probabilité d'avoir un front montant d'horloge juste pendant l'intervalle de temps de collection de charge est plus haute. Les registres suivant la logique combinatoire atteint peuvent donc emmagasiner la mauvaise valeur.

Une particule voyageant avec un angle incident à près de  $90^\circ$  peut en principe frapper deux drains ou plus de différents dispositifs en proximité et donc peut influencer deux nœuds ou plus. Ce phénomène peut menacer les circuits qui se basent sur la redondance pour protéger ses données. Le dépôt de charge sur nœuds multiples est limité par la longueur des traces d'ionisation, donc les nœuds qui sont suffisamment loin sont peu susceptibles de rassembler la charge de la même trace.

*Aléas logiques  
multiples*

## II-iv Protection contre les aléas logiques

Des techniques de fabrication pour protéger la logique contre les aléas peuvent être employées seulement quand une production en grande quantité est prévue, ce qui n'est pas le cas pour les expériences de physique. Les techniques de circuit et de système sont basées essentiellement sur la redondance de données. Les techniques de circuit consistent en des configurations de cellules mémoire, différentes de la cellule standard SRAM à 6 transistors, qui s'avèrent résistantes à un aléa sur un seul nœud, comme les cellules DICE, Whitaker, SERT, Dooley, Rockett. Les techniques de système exploitent des encodeurs et des décodeurs par codage à correction d'erreurs (Error Correction Coding, ECC) autour des blocs de mémoires standards.

Afin d'obtenir une certaine redondance, il est possible d'emmagasiner l'information dans un nombre double de nœuds par rapport à une cellule normale SRAM [Calin 96]. Une manière intelligente de relier les transistors entre eux et d'éviter la propagation d'erreur est montrée dans la fig. 2.17 page 50, ce qui représente une Dual-Interlocked Cell (DICE). Cette structure est entièrement symétrique et ses nœuds de mémoire sont totalement équivalents entre eux.

*La cellule DICE*

Sur chaque étape de propagation la valeur logique est inversée. Clairement, aucune valeur de logique ne peut se propager pour plus d'une étape dans la même direction. Il en résulte qu'un SEU sur un des nœuds de la mémoire des cellules affecterait seulement un autre nœud. Après qu'un nœud a été frappé, un certain temps est nécessaire pour reconstituer les tensions correctes dans toute la cellule, et ce

retard s'appelle *temps de rétablissement* (recovery time). Un aléa transitoire sur la sortie de la cellule peut être observé pendant le temps de rétablissement.

Il convient de noter que si deux nœuds de la cellule sont frappés en même temps (par la même trace d'ionisation), la cellule est sensible à un SEU. Quand un nœud est frappé, il y a toujours deux autres nœuds plus vulnérables qui gardent l'information sauvée. Par conséquent il est préférable de laisser un certain espace (de layout) entre les nœuds de la même cellule.

La cellule DICE est donc une cellule qui occupe 2 fois plus de surface qu'une cellule de mémoire standard et dissipe presque deux fois plus de puissance. Un verrou peut facilement être construit comme dans la fig. 2.18 en ajoutant les portes d'entrée du signal d'horloge. Cette cellule est appropriée pour substituer les verrous des machines à état de la logique de commande. Le verrou DICE n'est pas approprié pour des applications à haute vitesse.

*Redondance temporelle contre les SETs*

Une technique pour protéger la logique contre les SETs est d'employer la *redondance temporelle*. Il est en fait possible d'enregistrer un signal plus d'une fois pour obtenir plus d'une copie de sa valeur et par conséquent la redondance. L'inconvénient est que, dans la pratique, cela impose une restriction sur la synchronisation du signal qui doit être stable pendant le temps des deux enregistrements, ainsi la fréquence de fonctionnement du circuit doit être abaissée.

La cellule DICE peut être remodelée en dédoublant les entrées et les sorties, comme montré dans la fig. 2.22 page 55 pour un verrou DICE.

*Triple Redondance Modulaire*

À l'origine développé par [Von Neumann 56] en vue d'augmenter la fiabilité de l'électronique en général, la Triple Redondance Modulaire (TMR) a été par la suite appliquée à la microélectronique pour la protection contre les SEUs. Cette technique est basée sur un bloc de base appelé *voteur majoritaire* qui est une porte combinatoire simple avec 3 entrées et une sortie qui donne toujours comme valeur le niveau logique présent sur au moins 2 entrées, ainsi la majorité.

Dans la fig. 2.24(a), trois blocs identiques de logique reçoivent les mêmes entrées et sont reliés à un voteur. Normalement les trois blocs devraient donner les mêmes sorties, mais en cas de défaut ou SEU cela peut être faux. Il est clair qu'un défaut ou un SEU sur un seul des trois blocs sera masqué par le voteur et ne sera pas visible sur les sorties. Naturellement si deux blocs échouent en même temps, les sorties seront corrompues. Le voteur aussi peut être frappé, produisant un SET, et produite un état incorrect dans toute les trois machines à état. Pour cette raison il est plus adéquat d'employer la structure représentée dans la fig. 2.27(a).

Dans le cas où les machines à état sont en cascade ou reliées ensemble, il est possible d'employer le modèle de la fig. 2.27(b) page 58, qui représente une machine à état TMR complète, où le voteur est également triplé. Les I/Os sont triplés de telle manière que le raccordement avec la logique voisine est également redondant. La TMR complète est la protection totale pour la logique, puisqu'il protège la logique entière contre les SEUs et les SETs.

La TMR a une augmentation d'aire de 200%, avec une augmentation proportionnelle de puissance dissipée et de la charge sur l'arbre d'horloge. La TMR n'a pas de la même vulnérabilité accrue aux hautes fréquences que celle observée pour la cellule DICE et n'a pas de limitations de vitesse, ainsi elle convient à des applications à haute fréquence qui ne peuvent pas utiliser cette dernière cellule.

*Logique doublée (dual-rail logic)*

La redondance de la logique combinatoire peut être obtenue par son doublement



(dual-rail) et sa connexion à un registre doublé comme le DICE. Un exemple est montré dans la fig. 2.30, où au lieu de relier un simple bloc combinatoire de logique aux deux entrées d'une bascule DICE, il est possible de placer deux blocs combinatoires identiques pour conduire séparément les deux entrées du registre. Les entrées du bloc combinatoire seront reliées aux sorties séparées d'une bascule DICE.

La configuration présentée crée un double chemin pour les données et un SET affectera seulement une des deux entrées d'une bascule SEU-robuste, ainsi il sera filtré. L'utilisation d'aire est plus grande dans la logique doublée que dans la redondance temporelle, atteignant 100% par rapport à la logique non protégée, avec la même augmentation de puissance dissipée.

Une méthode efficace pour améliorer la fiabilité des mémoires et des communications numériques est le codage. Dans le cas de durcissement aux SEUs, les codes de bloc qui divisent l'information comme les blocs de mémoires sont utiles. Les codes de bloc [Clark 81] transforment des mots d'entrée de  $k$  bit en mots codés de  $n$  bit, en ajoutant  $(n - k)$  bits de parité pour obtenir un certain niveau de redondance.

Il y a plusieurs codes de bloc qui peuvent être employés pour les mémoires, comme [Hamming 50] et les codes de Reed-Solomon. Le codage est habituellement mis en application comme dans la fig. 2.31, où des données sont d'abord codées, puis emmagasinées dans une mémoire et par la suite décodées.

La complexité de l'encodeur est un paramètre important. Généralement une tentative de réduire la complexité par une utilisation plus efficace d'information conduit à l'augmentation de la complexité provoquée par l'introduction des circuits décodeurs. Pour de petits mots codés, la structure peut être concurrentielle avec la TMR pour protéger des registres (mais pas la logique combinatoire). Bien que l'augmentation d'aire pour le codage soit plus petite que pour le TMR, la dernière offre un durcissement plus fort : dans la TMR chaque bit est triplé indépendamment et les aléas logiques multiples sur bits de différents triplets sont masqués, tandis que pour le codage de Hamming un double SEU est sûrement fatal.

Le codage de Hamming n'est pas concurrentiel en ce qui concerne les cellules résistantes aux SEU (comme la DICE) pour la protection des registres, puisque les premières occupent moins d'aire et donnent une meilleure immunité.

Le codage devient avantageux une fois utilisé pour protéger des blocs SRAM. Un bloc SRAM fabriqué à partir de cellules traditionnelles à 6 transistors peut exploiter un encodeur/décodeur de Hamming si une partie de ses bits est réservée pour la parité. Un seul simple bloc encodeur/décodeur seulement est nécessaire pour nombreuses bits de mémoire, donc l'utilisation d'aire devient très efficace.

Les réseaux à haute capacité n'ont habituellement besoin d'aucune redondance pour le durcissement aux aléas logiques, puisque leur capacité les rend intrinsèquement résistants. Le changement de tension donnée par une particule ionisante dépend de la capacité, de la charge déposée et du courant de sortie du circuit de commande. Une LET de seuil nominale doit être choisie afin d'estimer la capacité qui peut être considérée suffisante pour ne pas exiger d'autre protection.

Les lignes de capacité élevées comme l'horloge et les réseaux de remise à zéro (reset) ont en général une haute capacité, donc leurs arbres peuvent fonctionner sans d'autres formes de protection contre les aléas logiques. Toutes les branches peuvent être conçues pour avoir la capacité parasite suffisante.

*Techniques de  
codage*

*Signaux à  
capacité élevée*

### III Logiques programmables et environnement radiatif

Le désir d'avoir du matériel programmable est présent depuis le début de l'électronique numérique, quand la réalisation rapide de prototypes était le but principal pour les dispositifs programmables.

#### III-i Logiques programmables simples

*Dispositifs  
PROM*

Toutes les fonctions logiques de  $n$  entrées peuvent être réalisées sous forme de somme de  $2^n$  minterms, qui sont les combinaisons de produit des signaux d'entrée sous leur forme positive ou négative. C'est l'idée de base des dispositifs PROM, qui consistent en un décodeur de rangée relié à une série de portes OR par une matrice de commutateurs programmables comme dans la fig. 3.1 page 68. Le décodeur est composé de portes AND, qui produisent les signaux de rangée. La structure ressemble donc à la fonction générale AND-OR décrite précédemment.

La programmation est habituellement faite en brûlant de *fusibles* qui neutralisent le chemin vers la masse dans une matrice OR. La matrice AND est par contre fixe.

*PLAs, PALs et  
PLDs*

Plus de généralité peut être offerte par les dispositifs PLAs, dont les deux matrices AND/OR sont complètement configurables. Un exemple de PLA est montré dans la fig. 3.2(a). Souvent, les fonctions logiques réalisées n'ont pas besoin de beaucoup de mintermes, ainsi une manière plus efficace de créer des fonctions universelles est d'avoir une matrice AND configurable ainsi qu'une matrice fixe OR, comme dans la fig. 3.2(b), qui représente les dispositifs PAL.

Une amélioration aux PALs consiste en l'addition de registres programmables aux sorties qui peuvent aussi rétroagir sur les valeurs dans la matrice AND (voir la fig. 3.3(a)). Ces dispositifs s'appellent PLDs. Ce changement a rendu possible la réalisation de machines à état et de logique séquentielle, ce qui a permis une commercialisation très rapide de ce type de dispositifs.

*CPLDs*

Le nombre d'entrées de la matrice AND ne peut pas augmenter indéfiniment puisque la grande capacité la rend inefficace. Une alternative est de mettre plus d'un PLD sur la même puce et de les relier, ainsi que des ressources programmables de connexion : ces dispositifs sont appelés PLD *complexes* (ou CPLDs) ; un exemple est représenté dans la fig. 3.3(b).

#### III-ii Réseau de portes programmables in-situ (FPGA)

L'architecture des réseaux de portes programmables in-situ (Field Programmable Gate Arrays, FPGAs) peut être vue comme un matrice de blocs logiques programmables entourée par une maille de connections configurables (voir la fig. 3.4 page 70). Une distinction importante doit être faite entre la logique de configuration et la logique d'utilisateur : la *logique de configuration* est constituée de l'infrastructure qui écrit, lit et emmagasine le programme dans le FPGA ; la *logique d'utilisateur* se compose de tout le reste des circuits qui, une fois que le FPGA est programmé, sont reliés ensemble pour créer le système désiré.

Les différences principales parmi les architectures de FPGA résident alors dans les constituants et dans la technique de programmation et d'emmagasinement de la configuration.

Le bloc logique de base peut varier d'un inverseur simple à une logique complexe avec des registres. Les blocs logiques sont alors distingués par la leur granularité, qui peut être définie comme le nombre de portes équivalentes (NANDs avec 2 entrées). L'avantage principal d'employer un bloc logique à grain fin est que l'utilisation du bloc est optimisée ; de fait, il est facile d'employer entièrement les portes et les techniques de synthèse logique sont élémentaires. D'autre part, la logique à grain fin exige plus de ressources de connexion qui sont coûteuses en termes de retard et de surface.

*Architecture du bloc logique*

Les tables de correspondance (Look-Up Table, LUT) sont souvent employées dans les FPGAs puisqu'elles offrent une grande polyvalence. Une LUT se compose fondamentalement d'une mémoire représentant la table de vérité de la fonction booléenne désirée. Les lignes d'adresse de la mémoire peuvent être contrôlées par des signaux d'entrée tandis que la sortie fournit la fonction booléenne. Les LUTs deviennent trop grandes pour plus de 6 entrées : c'est-à-dire que les expressions qu'elles peuvent établir ne sont pas souvent employées ; de plus elles sont difficiles à exploiter par les outils de synthèse.

L'architecture de connexion d'un FPGA est la façon dont les commutateurs et les segments programmables de câblage sont placés pour permettre l'interconnexion des blocs logiques. Il y a habituellement un compromis entre la flexibilité et la densité, puisque plus il y a d'interconnexions possibles dans un FPGA, plus il est flexible. En contrepartie une plus grande surface est perdue pour les connexions et la configuration.

*Architecture des connexions*

Les architectures de connexion évoluées incluent une vue hiérarchique des interconnexions, avec des boîtes de commutateurs (switch boxes) où les fils verticaux et horizontaux se croisent et peuvent être reliés ensemble. Les fils sortant d'un bloc logique peuvent entrer dans les boîtes commutateurs pour arriver aux connexions de plus haut niveau. Cette architecture, connue sous le nom de *modèle à îles*, est représentée dans la fig. 3.6(b) ; c'est la plus utilisée dans les dispositifs commerciaux.

Des réseaux spéciaux sont souvent disponibles dans les FPGAs pour la distribution d'horloge, de remise à zéro et d'autres signaux critiques.

Pour permettre une grande flexibilité, les entrées/sorties (I/Os) peuvent être programmées pour supporter différents standards de signalisation. Les blocs d'entrée/sortie contiennent souvent des registres et peuvent exécuter des communications en double rythme de donnée (Double Data Rate, DDR).

*Blocs d'entrée/sortie*

Les possibilités pour la mémorisation de la configuration varient entre dispositifs programmables une seule fois (One Time Programmable, OTP), blocs de mémoires non volatiles ou volatiles.

*Technique de programmation*

Le dispositif OTP le plus commun est l'*antifusible* (fig. 3.7(a) page 73) : composant à deux bornes dont l'état non programmé présente une très haute résistance ( $\approx 1 \text{ G}\Omega$ ). Quand une haute tension est appliquée à l'antifusible, il se déclenche et crée un lien permanent de basse résistance ( $\approx 50 \Omega$ ). Des circuits supplémentaires sont nécessaires pour programmer l'antifusible avec des tensions élevées. Il est nécessaire d'utiliser de grands transistors pour manipuler des courants élevés, ce qui limite le gain de surface. Les antifusibles n'exigent aucune alimentation d'énergie ou aucun stockage externe de configuration quand le système est arrêté. Aucune reprogrammation n'est possible en utilisant des antifusibles.

Les dispositifs à *grille flottante* (floating gate) comme les EPROM, EEPROM ou les Flash (fig. 3.7(b)) sont des mémoires non volatiles. Les transistors à grille flottante ont deux grilles, une grille de contrôle supérieure et une grille flottante inférieure. La grille flottante est isolée de tous les autres nœuds et il est possible d'y injecter et d'y extraire charge. Cette charge représente la valeur mémorisée. La programmation et l'effacement de ce type de dispositif exigent une haute tension, donc la tension normale d'opération n'affecte pas la charge stockée dans la grille flottante. Les dispositifs à grille flottante sont reprogrammables, donnant plus de polyvalence, et dans le cas où une erreur serait faite pendant la conception, le programme peut être corrigé. D'ailleurs, ces dispositifs sont non volatiles donc aucune alimentation d'énergie ou stockage externe ne sont nécessaire pour préserver la configuration. Par contre il est nécessaire d'utiliser des circuits pour la génération des hautes tensions pour la programmation et effacement.

Les *mémoires volatiles* sont des cellules SRAM ou des bascules. Les registres sont plus grands que les cellules de SRAM qui sont à leur tour plus grandes que les cellules non volatiles. En réalité, le stockage statique de configuration de mémoire domine habituellement l'utilisation d'espace sur un FPGA. Puisque la SRAM est volatile, la configuration doit être rechargée après chaque arrêt, donc un stockage externe de configuration est obligatoire. La mémoire statique est reprogrammable, elle n'exige aucune tension élevée et sa production n'a besoin d'aucun processus spécial.

#### *Blocs spéciaux*

Dans beaucoup de FPGAs il est possible de trouver des blocs particuliers comme par exemple des simples mémoires, multiplicateurs, boucles à verrouillage de phase (PLLs), ou des microprocesseurs. Ces blocs sont reliés à la structure de connexion de le FPGA comme les autres blocs logique.

### III–iii Effets des radiations sur les dispositifs programmables

Quelques considérations spéciales doivent être faites au sujet des effets des radiations sur les FPGAs. Tout d'abord, les FPGAs commerciales sont, selon la technologie dans laquelle ils sont conçus, très différemment influencées par les radiations. En outre, la structure interne d'un FPGAs mémorise l'information de configuration et l'information d'utilisateur, qui ont une importance différente pour le comportement du système.

#### *Antifusibles*

La majorité des FPGAs à antifusibles sont résistantes jusqu'à 300 krad de dose totale mais elles se dégradent au-dessus de ce seuil en raison de la pompe de charge interne utilisée pour produire les tensions élevées requises pour la programmation. La pompe de charge et les transistors d'isolation sont fabriqués avec des dispositifs à oxyde de grille épais qui rassemblent donc une charge positive plus grande que les dispositifs normaux à oxyde mince. Les antifusibles sont intrinsèquement immunisés contre les SEUs. Cependant, des erreurs peuvent être observées dans la logique d'utilisateur et dans la logique de commande de le FPGA qui contrôle la programmation et le déclenchement du dispositif.

#### *Flash*

Les dispositifs basés sur mémoire Flash font face aux mêmes problèmes de dose que les dispositifs à antifusibles, puisqu'ils doivent produire des tensions élevées pour programmer et effacer de la même manière. Le transistor à grille flottante souffre lui-même des effets de dose totale : les états d'interface induits par les radiations dans la porte flottante peuvent potentiellement affaiblir la conservation de charge

et par conséquent causer à long terme un problème de fiabilité. Les FPGA à grille flottante sont limités aux applications en-dessous de 100 krad.

Les FPGAs basées sur la mémoire statique sont traitées en technologie standard CMOS. Afin de se conformer à différents standards de signalisation et différents niveaux de tensions, ces dispositifs contiennent des transistors à oxyde de grille épais dans leurs blocs d'entrée/sortie, donc ils ne résistent pas à plus de 200 krad. Quelques FPGAs durcies à la TID sont toujours très sensibles aux SEUs mais un ensemble de techniques au niveau de la programmation a été développé pour mieux protéger la logique.

*SRAM*

### III–iv Techniques de protection contre les SEU pour les logiques programmables du commerce

Les FPGAs non résistants aux SEUs utilisent des approches diverses pour atténuer les aléas logiques dans les mémoires de configuration et dans les registres de l'utilisateur. Ces méthodes sont des techniques au niveau du système et au niveau du programme. Elles incluent la TMR et la reconfiguration.

Puisque les FPGAs sont programmées avec l'aide d'outils de synthèse, il est suffisant d'introduire la TMR dans le langage de description matériel (Hardware Description Language, HDL) pour obtenir une logique résistante aux SEU. Des essais sur FPGAs à antifusibles [Wang 03b] utilisant la TMR ont confirmé la validité de cette approche. Néanmoins, les FPGAs à SRAM sont sensible aux aléas dans la logique de configuration, qui peuvent être bien plus perturbants pour le système.

*TMR*

Certains FPGAs à mémoire statique permettent la reconfiguration pendant l'opération. Cela signifie qu'il n'est pas nécessaire de remettre à zéro le FPGA pour reprogrammer une nouvelle configuration. Cette fonction peut être employée pour reconstituer la configuration d'une puce après un aléa logique [Xilinx 00] : cette technique est appelée *scrubbing*. Des essais expérimentaux ont été effectués sur les dispositifs FPGA à SRAM en utilisant la combinaison des techniques de TMR et de reconfiguration. Ils ont démontré une grande amélioration de la résistance du système aux SEU [Yui 03]. L'inconvénient est la nécessité d'utiliser un contrôleur externe.

*Reconfiguration*

## IV Un FPGA résistant aux radiations pour la HEP

Les progrès en technologies microélectroniques appliquées aux FPGAs ont diminué les coûts et le temps d'élaboration de l'électronique numérique et ce tant dans le secteur industriel que dans le secteur de l'espace et de l'aéronautique. L'utilisation de tels dispositifs présente également un intérêt pour les expériences d'HEP, qui sont maintenant contraintes d'exploiter des ASICs dans leurs détecteurs. Ceux-là sont placés à proximité des collisions de particules comme au LHC du CERN.

Aucun FPGA existant dans le commerce ne peut tolérer toute la dose produite par les expériences d'HEP. Les dispositifs résistants aux radiations sur le marché sont aussi extrêmement chers.

La première partie de cette thèse est une étude de développement d'un FPGA résistant aux radiations (RT-FPGA) pour la physique des hautes énergies. Le but est d'obtenir un FPGA résistante jusqu'à 20 Mrad avec l'immunité aux SEUs pour les

registres d'utilisateur et de configuration. L'insensibilité aux SEU devra être intégrée dans la puce, ce qui n'exige pas de la part de l'utilisateur d'exploiter des techniques spéciales comme la TMR ou la reconfiguration.

#### IV-i Dessin du bloc logique en CMOS 0.25 micron

La majeure partie de la surface du bloc logique sera dédiée à la mémorisation de la configuration. Un bon équilibre entre les quantités de connexion et de logique est obtenu en utilisant des LUT à 4 entrées. La conception du bloc inclus un registre, la logique pour la propagation de la retenue (carry) et les portes pour la génération des fonctions étendues. Le bloc de logique est représenté dans fig. 4.1 page 78 qui montre également que le LB possède un bloc additionnel de 15 bits de configuration.

*La table de correspondance (LUT)*

La LUT est composée de 16 registres qui ressemblent à un registre à décalage au travers de laquelle la configuration peut être chargée. La logique est conçue pour pouvoir utiliser ces registres comme une RAM synchrone de  $16 \times 1$  bit à double accès (dual-port) ou comme un registre à décalage (shift-register). La LUT a donc un bus d'adresse de lecture de 4 bit et un bus d'adresse d'écriture de 4 bit utilisé seulement dans le mode RAM. Un bit de configuration est réservé pour stocker le mode de fonctionnement RAM ou registre à décalage. Un schéma simplifié de la LUT est représenté dans la fig. 4.2 page 79.

La LUT est aussi composée d'un multiplexeur qui choisit une sortie de registres parmi les 16 possibles. Un décodeur choisit quel registre doit recevoir le signal d'horloge en cas de fonctionnement comme RAM. Pendant la phase de configuration, la LUT est mise en mode registre à décalage et le signal d'horloge est donné à tous les registres.

La LUT a donc deux entrées et une sortie auxiliaires pour le début et la continuation des chaînes de registres à décalage. Deux autres entrées et une autre sortie sont utilisées pour la connexion de LBs voisins afin de former des fonctions logiques plus compliquées.

Afin de protéger la LUT contre les SEUs, les registres sont tous remplacés par des bascules DICE. Le circuit utilisé est représenté dans la fig. 4.3 page 81 et il est composé de deux verrous, un maître (master) et un esclave (slave). Le tampon d'horloge local (local clock buffer) est doublé pour redondance.

La disposition de la bascule est représentée dans la fig. 4.4 page 82, où chaque domaine représente un nœud du circuit. Pour rendre une cellule résistante aux SEUs les nœuds de mémoire du même verrou sont placés loin l'un de l'autre afin d'éviter la collection de charge par des nœuds multiples. La distance atteinte est au minimum de  $10 \mu\text{m}$ , ce qui devrait garantir une probabilité assez basse de SEU. Pour ne pas perdre d'espace, les nœuds de l'esclave et les nœuds du maître sont alternés, car ils appartiennent à différents domaines qui n'interagissent pas ensemble. L'augmentation de la distance entre les nœuds qui doivent être reliés augmente également la complexité des connexions locales. Cela devient un facteur limitant dans cette technologie qui ne permet que trois niveaux de métal.

La logique entière dans la puce est protégée par doublement, donc il y a deux copies de chaque bloc logique dans la LUT. La majeure partie de la surface de la puce est de toute façon occupée par les registres. Il est possible de créer une disposition tout à fait symétrique pour la LUT, qui est représenté dans la fig. 4.8 page 84.

*Propagation de la retenue (carry)*

Un bloc logique spécialisé dans la propagation de la retenue facilite la réalisation

des additionneurs, réduisant au minimum le nombre des LBs nécessaires. Sans cette structure, le nombre de LB utilisé pour un additionneur de  $n$  bit serait  $2n$ , avec 2 sorties par bit, alors qu'avec cette architecture seulement  $n$  LBs sont utilisés. L'entrée de retenue peut être initialisée à la valeur d'une entrée auxiliaire pour le premier additionneur dans la chaîne.

Quand le LB fonctionne en mode additionneur, la LUT est configurée comme une fonction XOR à deux entrées. Selon l'expression nécessaire la chaîne de retenue est configurée pour effectuer l'addition, la soustraction ou la comparaison. Le circuit de retenue est montré dans la fig. 4.11 page 86.

Afin de produire des fonctions logiques avec plus de 4 opérandes, l'extendeur relie des LUTs différentes par l'intermédiaire des multiplexeurs. Généralement deux LUTs à  $n$  entrées peuvent alimenter un multiplexeur 2 : 1 formant une fonction booléenne de  $n + 1$  variables où la  $n + 1$ -ième entrée est l'entrée de sélection du MUX. Dans ce dessin, l'extendeur de fonction peut construire un arbre composé jusqu'à 16 LUTs pour former une expression booléenne de au maximum 8 entrées. Une des entrées auxiliaires est utilisée pour la sélection du multiplexeur.

*Extendeur de fonctions*

Le LB a deux sorties ordinaires, l'une avec registre d'utilisateur et l'autre sans. L'entrée auxiliaire peut se connecter directement au registre d'utilisateur et laisser l'autre sortie disponible pour être utilisée par le reste de la logique. De cette manière, le registre d'utilisateur peut être exploité pour une fonction et le reste du LB pour une autre, ce qui augmente l'efficacité.

Le registre d'utilisateur diffère légèrement de la bascule utilisée pour la LUT : il possède des entrées de remise à 0 et à 1 qui peuvent être configurées pour être synchrones (clear/preset) ou asynchrones (reset/set). Le circuit du registre est montré dans la fig. 4.12 page 88.

*Le registre d'utilisateur*

Le registre d'utilisateur participe à la chaîne de configuration en étant chargée d'une valeur initiale. Ainsi, il doit être synchronisé par l'horloge de configuration au début et par l'horloge d'utilisateur ensuite. Cette transition de phase est commandée par un signal global qui reste bas jusqu'à ce que la configuration soit terminée. Le registre d'utilisateur peut être configuré en verrou ou bascule.

Les bits de configuration sont stockés dans une deuxième chaîne de registre à décalage qui inclut le registre d'utilisateur. Les registres utilisés dans cette chaîne sont une version simplifiée de ceux utilisés dans la LUT. De futures améliorations pourraient exploiter des cellules de SRAM pour cette partie, ce qui économiserait de la surface. Il y a 15 bits dans ce bloc additionnel de configuration. Ensemble, le registre d'utilisateur, le bloc de configuration et la LUT forment un total de 32 registres par LB.

*Le bloc de configuration*

Comme montrés par la fig. 4.14 page 90, chaque paire de blocs logique est étroitement couplé et partage la même horloge d'utilisateur, les signaux de set/reset et le bus d'adresse d'écriture. Dans la paire de LBs, la logique pour la génération d'horloge est aussi partagée. La mise en commun des signaux entre LBs permet de réduire le nombre de raccordements à la boîte de commutation. La fig. 4.13 représente la disposition d'une paire de blocs logique avec leur infrastructure commune, qui est physiquement placée au milieu de deux LBs. La paire de LBs représente l'unité qui sera reliée à l'infrastructure de connexion.

*LB couplés et modules*

Une pile de 8 blocs logique (4 paires de LBs) qui partagent des raccordements pour l'extendeur de fonctions forme un sur-ensemble appelé *module*. La fig. 4.14 met en évidence les connexions parmi des blocs logiques. Les raccordements de l'extendeur de fonctions se prolongent aux modules voisins, laissant se joindre jusqu'à 2 modules pour la même fonction logique.

Les signaux de propagation de retenue se prolongent à la logique voisine aussi bien que les signaux du registre à décalage. Ces raccordements organisent les modules dans une chaîne. Sur toute la longueur de la chaîne, l'utilisateur peut programmer un bloc additionneur exploitant la logique de retenue, un registre à décalage ou encore un bloc de RAM de la taille voulue.

Puisque un LB contient 32 registres, une paire de LBs contient 64 cellules de stockage et un module 256 cellules de stockage. Il existe 17 raccordements entre chaque paire de LBs et sa matrice de commutation.

*Puce de test en technologie 0.25 micron*

Une puce de test dans une technologie CMOS 0.25  $\mu\text{m}$  a été développée pour étudier l'opération des blocs logiques et leur comportement en environnement radioactif. La puce de test est un circuit intégré de  $2 \times 2 \text{ mm}^2$  contenant 4 modules, donc 32 blocs logiques ou, en d'autres termes, 1024 registres au total. La puce d'essai n'inclut aucune infrastructure d'interconnexion configurable. La fig. 4.15 page 91 montre une image de la puce prise au microscope.

Le nombre des entrées/sorties de la puce a été limité à 30 pour économiser l'aire et donc le coût de production. Cette limitation a forcé une certaine simplification des raccordements internes car il est impossible dans ces conditions de relier chaque module à approximativement 70 signaux. Par conséquent, dans la puce d'essai, les signaux de set/reset et l'horloge de tous les LBs sont reliés ensemble dans des arbres de tampons, et il en va de même pour les signaux auxiliaires et 3 des 4 entrées de la LUT. La puce ne possède donc que 7 entrées.

Les LBs sont reliées dans une structure vérifiable composée de 4 chaînes, et qui possède le nombre minimum d'I/Os possible. En pratique on pourra étudier indépendamment la fonctionnalité de chaque paire de LB. Il est en effet possible de programmer une configuration inerte dans les LBs inutilisées et une configuration significative dans le LB à l'essai. La disposition de l'entité du module choisie pour cette puce d'essai n'est pas celle qui sera employée dans l'application finale.

La chaîne du registre à décalage est liée à la chaîne de configuration, afin de former une seule unité de 1024 registres. Il y a un total de 22 entrées-sorties, qui doivent être complétées par 8 connections pour l'alimentation et la masse. Une alimentation séparée est employée pour les tampons d'entrée-sortie et pour la logique interne même si la tension d'alimentation reste la même ( $V_{DD} = 2.5 \text{ V}$ ). Quatre bornes fournissent la puissance à l'anneau d'alimentation interne, alors que le reste fournit la puissance pour les tampons d'entrée-sortie. L'aire inutilisée de la puce, au-dessous de la distribution d'alimentation, est remplie de condensateurs de poly-silicium sur N-well pour le découplage.

*Durcissement des signaux globaux*

Pour durcir les signaux globaux, on les relie à un grand nombre de portes, ce qui a pour effet d'élever la capacité des branches. Il n'est plus nécessaire d'utiliser la technique dual-rail pour ces signaux, ce qui simplifie les connexions. Dans ce travail, un réseau est défini pour être de haute capacité s'il est relié à plus de 63 portes, ce qui correspond approximativement à une capacité de  $C_{th} = 1.6 \text{ pF}$ . La LET de seuil correspondante est  $\approx 190 \text{ cm}^2\text{MeV/mg}$  au cas où le nœud est à haute impédance.



C'est plus que ce qui est nécessaire pour l'environnement d'application (au maximum  $17 \text{ cm}^2\text{MeV/mg}$  dans le LHC).

Des entrées spéciales ont été conçues pour tous signaux qui deviennent dual-rail : afin de protéger le signal dès le début, le tampon d'entrée lui-même doit être dual-rail, donc deux tampons sont nécessaires. Aucune sortie spéciale n'a été conçue dans cette puce d'essai ; tous les signaux dual-rail sont convertis en simple-rail en les reliant simplement sur le tampon de sortie.

*Durcissement des entrées-sorties*

Une puce d'essai contient seulement 1024 cellules de stockage, ce qui pourrait être insuffisant pour une caractérisation significative aux SEU et/ou pour produire une statistique satisfaisante, un temps très long d'exposition au faisceau de test serait nécessaire. En considérant aussi que le temps d'exposition au faisceau coûte cher, deux puces ont été incluses dans chaque boîtier, suffisamment proche pour être couvertes entièrement par le faisceau. De cette façon, le double des statistiques est produit. La fig. 4.17 page 94 montre une photo du boîtier.

*Boîtier*

Un essai d'exposition à un faisceau d'ions a été planifié pour caractériser la robustesse aux SEU de la puce et de ses structures internes. Trois essais différents ont été faits : un essai statique de conservation de la configuration, un essai dynamique de configuration et un essai dynamique de données d'utilisateur. Pendant chaque essai la fluence du faisceau a été exactement mesurée.

*Procédures d'essai d'exposition aux ions lourds*

L'essai statique comprends trois étapes : (a) le chargement d'une configuration tandis que le faisceau est éteint ; (b) l'arrêt de l'horloge et le gel des signaux d'entrée de la puce et l'allumage du faisceau à une fluence spécifique ; (c) le redémarrage de l'horloge et la comparaison de la configuration de sortie à l'originale.

L'essai dynamique de configuration est fait en chargeant simplement, alors que le faisceau est allumé, une longue configuration dans le registre à décalage et en comparant sans interruption la sortie à l'original.

L'essai dynamique de données d'utilisateur est par contre réalisé par (a) le chargement d'une configuration tandis que le faisceau est éteint ; (b) le démarrage du faisceau ; (c) le démarrage de la puce en mode d'utilisateur avec des données aléatoires et l'acquisition des données de sortie ; (d) l'extinction du faisceau à une fluence spécifique ; (e) la mise en marche de l'horloge et la comparaison de la configuration de rendement à l'originale. La configuration utilisée pour ce dernier essai a été une configuration XOR à 4 entrées pour toutes les LBs. De cette manière, tous les changements de bits de registres d'utilisateur ou de configuration doivent être vu comme changement aux sorties.

Un circuit imprimé (Printed Circuit Board, PCB) a été conçu pour l'installation de l'essai pendant l'exposition au faisceau. Il comprend un socle pour le dispositif à vérifier (Device Under Test, DUT), un FPGA Xilinx Spartan-3, une interface USB et quelques régulateurs linéaires pour l'alimentation.

*Carte de test*

L'interface USB doit être reliée à un ordinateur, qui exécute un programme de contrôle. Une séquence de test peut être chargée et récupérée par USB dans la mémoire de la Spartan-3. La Spartan-3 peut alors appliquer la séquence de test au DUT. La fig. 4.19 page 97 dépeint un schéma fonctionnel du PCB.

Bien que le faisceau puisse être focalisé sur le DUT en un point de 25 mm de diamètre, les autres composants du PCB sont placés loin pour ne pas être exposés aux radiations car ils n'y sont pas résistants.

Le Spartan-3 a dû être programmé pour exécuter les essais, pour acquérir les données et pour communiquer avec l'interface USB. Le programme consiste en deux automates finis, un pour la communication entre l'USB et la mémoire et un pour les opérations de test. Les deux machines à états finis marchent dans deux domaines différents d'horloge et communiquent par un protocole. Un ensemble des registres de commande, accessible par USB, a été programmé pour fournir des paramètres à la méthode d'essai. La machine à état finis d'essai a deux modes principaux : le mode parallèle et le mode série. Intuitivement, le mode parallèle est utilisé pour l'essai dynamique de données d'utilisateur mais il est également employé pour l'essai statique, alors que le mode série est utilisé pour l'essai dynamique de la configuration.

Pour la protection contre les SEU dans le Spartan-3 les compteurs d'erreur et l'horloge de temps réel, qui sont les registres critiques, sont protégés par TMR.

Un logiciel pour Microsoft Windows XP qui se connecte via USB à la carte de test a été développé. Le logiciel est entièrement écrit en Visual Basic. La fig. 4.22 page 100 montre l'interface du logiciel.

#### *Installation du test*

L'irradiation a été exécutée au Heavy-Ion Facility (HIF) au CYCLONE à Louvain-La-Neuve, Belgique. Ce cyclotron fournit plusieurs ions qui couvrent la gamme LET de 1.7 à 55.9 cm<sup>2</sup>MeV/mg et avec un flux moyen  $2 \cdot 10^4$  cm<sup>-2</sup>s<sup>-1</sup>. La carte d'essai a été montée sur une armature dans la chambre à vide et le boîtier du DUT a été ouvert. Afin d'acquérir des statistiques pour plusieurs valeurs de LET, la carte peut être inclinée de 45 degrés et 60 degrés par rapport au faisceau. Chacun des trois essais a été réalisé avec différents ions et à différents angles d'inclinaison, en couvrant la gamme LET de 15 à 112 cm<sup>2</sup>MeV/mg.

#### *Résultats du test avec ion lourds*

Un résumé des résultats d'essai sous faisceau est montré dans la table 4.1 page 101. Dans toute la gamme LET explorée le nombre d'erreurs recueillies a été nul ou très bas, donc dans la plupart des cas il est possible de donner seulement une limite supérieure pour la section transversale. Cette limite supérieure est donnée avec un niveau de confiance de 95%.

Les données expérimentales montrent bien la robustesse du circuit aux SEUs jusqu'à une LET de 79.6 cm<sup>2</sup>MeV/mg, puisque aucune erreur n'a été observée jusqu'à ce niveau. Pour une valeur de 112 cm<sup>2</sup>MeV/mg le test dynamique a montré une petite sensibilité, avec une section transversale de  $6.2 \cdot 10^{-10}$  cm<sup>2</sup>/bit, ce qui donne une limite supérieure de  $1.2 \cdot 10^{-9}$  cm<sup>2</sup>/bit, alors que les autres essais n'ont pas montré d'erreur.

Avec les statistiques disponibles, la section transversale limite est inférieure ou égale à  $2.9 \cdot 10^{-9}$  cm<sup>2</sup>/bit dans toute la gamme LET observée pour l'essai statique et l'essai dynamique de configuration. À titre de comparaison, un registre fabriqué avec la même technologie et avec les mêmes techniques de durcissement à la TID mais non protégé contre SEUs a montré une LET de seuil de 14.7 cm<sup>2</sup>MeV/mg et une section transversale de saturation de  $2.59 \cdot 10^{-7}$  cm<sup>2</sup>/bit.

Une explication pour les erreurs observées dans le test dynamique peut être la suivante : le registre de la LUT possède un point faible dû à la proximité mutuelle des deux multiplexeurs d'entrées (voir fig. 4.4 page 82). Ceci pourrait causer des phénomènes de collection de charge sur les nœuds multiples. Ces 2 multiplexeurs forment ensemble une aire sensible de  $\approx 44.1$  μm<sup>2</sup> pour les transistors à canal p et de  $24.3$  μm<sup>2</sup> pour les transistors à canal n, qui pourrait très bien être responsable des aléas logiques enregistrés. Il est clair que la perturbation simultanée des deux copies du même signal ait comme conséquence une erreur.

On notera que des erreurs n'ont été observées que lorsque la carte était inclinée de 60 degrés : l'inclinaison augmente la probabilité de frapper des nœuds multiples puisque la particule se déplace le long des dispositifs. Dans de futures versions du LB, ce problème sera corrigé en changeant le placement des multiplexeurs d'entrée du registre de la LUT.

#### IV–ii Transposition du LB vers une technologie 0.13 micron

Dans la perspective d'une production à long terme de la RT-FPGA, l'effort de conception s'est concentré sur la transposition du dessin du bloc logique vers une technologie plus avancée de 0.13  $\mu\text{m}$  qui tient compte d'une logique de densité plus élevée.

En outre, il y a des indications claires que la technologie 0.13  $\mu\text{m}$  est intrinsèquement résistante aux radiations, n'exigeant pas l'utilisation de ELT. Pour les mêmes raisons, les anneaux de garde ne semblent pas être nécessaires. La conception a commencé par la recherche d'une cellule de mémoire résistante aux SEUs dans la nouvelle technologie, qui a été alors employée pour réaliser le LB.

Les résultats obtenus avec la cellule de technologie 0.25  $\mu\text{m}$  ne peuvent pas être facilement transposés en technologie 0.13  $\mu\text{m}$ , puisque la géométrie doit changer. La cellule doit être haute de 3.6  $\mu\text{m}$  pour se conformer à la librairie. Le nombre de traces horizontales pour chacun des 6 niveaux de métal sera donc limité à 9. La difficulté repose dans le fait que l'alternance exige un bon nombre de ressources de connexion, quantitativement 2 fils par nœud de mémoire, donc 16 fils au total pour une bascule simple.

*Registre simple  
alternée résistant  
aux SEUs*

Deux niveaux de métal sont nécessaires et suffisants pour la connexion d'une bascule intercalée simple avec un schéma semblable à celui dans la fig. 4.3 page 81. Une bascule D a été conçue pour le test et sa disposition est décrite dans la fig. 4.24 page 104. La surface de la cellule est  $14.5 \times 3.6 \mu\text{m}^2$ , ce qui est deux fois l'aire de la bascule standard dans la librairie commerciale. La distance minimale entre les nœuds sensibles est 2.4  $\mu\text{m}$  ce qui est 4 fois moins que celle obtenue en 0.25  $\mu\text{m}$ .

Pour protéger la cellule contre la TID, tous les transistors ont une largeur supérieur à 0.3  $\mu\text{m}$ , de manière à limiter la variation de la tension de seuil à 100 mV.

Alterner les nœuds de deux registres au lieu d'un seul peut augmenter la résistance aux SEUs. Un double registre alterné contiendra deux registres indépendants qui seront seulement assemblés dans le layout. Les deux registres peuvent être placés comme dans la fig. 4.25 page 104, en ayant une distance minimum de 9  $\mu\text{m}$  entre les nœuds sensibles, mais 3 niveaux de métal sont obligatoires pour cette stratégie. Une solution de compromis peut être celle représentée dans la fig. 4.26 qui a été développée pour l'essai.

*Registre double  
alternée résistant  
aux SEUs*

Les deux architectures de registre décrites dans les paragraphes précédents ont été assemblées dans une puce d'essai ainsi qu'un registre de librairie standard (non résistante aux SEUs). Chacun des trois registres est répliqué afin de former un long registre à décalage. Il y a 4096 cellules non résistantes, alors que les deux registres résistants se composent de 4608 cellules chacun. La disposition de la puce d'essai apparaît dans la fig. 4.27 page 105.

*Evaluation des  
structures  
résistantes aux  
SEU par puce de  
test*

### Résultats du test avec ions lourds

Les résultats d'essai sont récapitulés dans la Table 4.2 et dans la Figure 4.29. Les essais ont montré que les deux structures durcies aux SEUs ont une grande robustesse en mode statique, alors qu'ils sont sensibles dans le mode dynamique. En mode statique, des erreurs ont été observées seulement dans la cellule alternée simple et à un LET de  $45.8 \text{ cm}^2\text{MeV}/\text{mg}$ , ce qui est bien au-dessus des spécifications.

Dans le mode dynamique les deux cellules robustes aux SEUs ont montré des erreurs avec une section transversale fortement dépendante de l'angle d'incidence du faisceau. C'est une indication claire que la collection de charge par nœuds multiples joue un rôle dans le mécanisme de SEU, puisque plus d'erreurs sont produites quand les ions ont un angle d'incidence élevé. Néanmoins, lors du test dynamique, les cellules durcies aux SEUs ont montrée une résistance  $10\times$  supérieure à la cellule de la librairie standard.

Les deux cellules robustes aux SEUs peuvent être employées pour le stockage de configuration dans l'application puisque leur robustesse en mode statique est suffisante. Un circuit plus résistant devrait être utilisé pour les registres d'utilisateur.

## IV–iii Développement des connexions programmables

Les interconnexions de le FPGA doivent être une combinaison équilibrée de raccordements locaux, qui apportent les signaux entre les cellules voisines, et de raccordements longs, qui apportent les signaux entre des endroits éloignés sur la puce. Une paire de LBs avec ses connexions adjacentes forme un *carreau*, qui est la structure de base répétée dans le deux dimensions pour former une rangée.

### Fils

Pour laisser l'utilisateur mettre en application des connexions non encombrés, le nombre des fils horizontaux et verticaux doit être presque égal au total d'entrées-sorties du LB qui doivent être reliés. Dans cette conception le nombre de fils est de 18 par direction, avec un rapport de 1 : 2 entre les lignes longues et les lignes courtes. Chaque LB sera de préférence reliée aux LBs voisines, afin de réduire la congestion et le retard. La fig. 4.30 page 109 montre l'architecture de câblage conçue. Quatre lignes spécialisées d'arbre d'horloge sont disponibles. Les carreaux adjacents partagent un certain nombre de raccordements directs. Les entrées de chaque paire de LBs sont physiquement réparties sur les quatre côtés du bloc.

### Commutateurs

Les raccordements programmables parmi les fils sont composés par des pass-transistors, des tampons à trois états (tristate buffers) ou des multiplexeurs, selon la longueur des lignes, leur direction et leur utilisation. Les schémas dans la fig. 4.31 explique les types de connexions réalisées.

## V Un PLD résistant aux radiations

Dans l'infrastructure de nombreuses expériences ou accélérateurs, une logique d'interface ou une simple fonction entre des ASICs est souvent nécessaire, par exemple pour l'adaptation ou la réparation. Dans ces cas un PAL/PLD est utile puisqu'il peut être accordé aux besoins de l'utilisateur. Évidemment, dans un environnement de radiations, il est obligatoire d'utiliser un PLD résistant a la TID et aux SEUs, et un travail a été effectuée dans ce sens pour développer un tel dispositif.

## V-i Structure

La contrainte principale pour la conception du PLD résistant aux radiations est le coût de production, qui doit être le plus bas possible, et se traduit directement en une contrainte en termes d'utilisation d'aire. Le but est de construire une puce de  $2 \times 2 \text{ mm}^2$ .

Tout les PLDs sont basés sur des cellules non-volatiles puisqu'ils ne peuvent pas avoir un circuit de démarrage et qu'ils doivent être fonctionnels dès la mise sous tension. Le choix du genre de cellule de mémoire pour la configuration tombe donc sur les cellules à fusibles et dans notre cas sur les fusibles programmables avec laser. La cellule de stockage à fusible ne peut pas être plus petite que  $32 \times 14 \mu\text{m}$  et ceci impose une contrainte dans notre conception.

*Mémorisation du programme*

Le PLD se compose d'une architecture traditionnelle AND/OR. Les entrées du PLD entrent verticalement dans la matrice ainsi que leurs contreparties inversées, suivant les indications de la fig. 5.1 page 114 qui décrit une section du PLD correspondant à une seule sortie.

*Architecture*

Avec une aire de  $1 \text{ mm}^2$  pour le noyau, il est possible de mettre 2048 cellules à fusibles, ce qui correspond à une matrice  $64 \times 32$ , adaptée à un PLD avec 8 entrées, 8 sorties et 8 minterms par sortie. La figure 5.2 représente la structure complète du PLD. Le PLD possède deux entrées additionnelles pour l'horloge et l'output enable qui interagissent directement avec les blocs logiques.

Chaque bloc de logique est relié à deux entrées-sorties de la puce, dont une entrée-sortie alternative. Le comportement de chaque bloc logique dépend de 4 bits de configuration qui choisissent parmi 3 modes de fonctionnement différent : enregistré, simple et complexe. Le mode de fonctionnement décide si le bloc logique utilise l'entrée-sortie primaire ou l'entrée alternative et s'il l'emploie comme entrée, sortie ou entrée-sortie. Le schéma d'un bloc de logique est montré dans la fig. 5.3. Les multiplexeurs dans le schéma sont commandés par des bits de configuration.

*Le bloc logique*

Chaque bloc logique comprend une boucle de retour à la matrice AND, qui peut donc former des fonctions plus complexes. Les bits de configuration du bloc logique indiquent la provenance de la boucle.

En mode enregistré, le bloc de logique exploite sa bascule d'utilisateur, qui est synchronisée par l'horloge d'entrée ; la rétroaction à la matrice AND vient de la sortie de la bascule. En mode complexe, le bloc de logique se comporte de manière asynchrone et il est configuré pour une opération bidirectionnelle ; la rétroaction vient directement de l'entrée-sortie. En mode simple, le bloc de logique est encore asynchrone mais il est configuré pour une opération monodirectionnelle.

La matrice AND est un ensemble de portes AND câblées, étendues en traits horizontaux. Chaque ligne horizontale est une porte AND et possède donc plusieurs transistors de pull-down, un par ligne d'entrée verticale, et un pull-up constant fourni par un transistor toujours actif à canal P avec sa grille reliée à la masse. La fig. 5.7 page 119 représente un trait horizontal et ses raccordements.

*La matrice AND*

Chaque trait horizontal a une capacité d'interconnexion de presque 400 fF. Le retard de la porte AND est directement lié à la valeur de cette capacité et aux forces des transistors de pull-up et pull-down. Pour cette raison d'une part les transistors devraient être forts, mais, d'autre part, le pull-up ne peut pas être fort puisqu'il est

toujours actif et qu'il consomme la puissance statique quand la valeur de logique sur la ligne est basse.

La taille du pull-up est donc un compromis entre la vitesse et la puissance. Comme exemple, avec un pull-up constant de  $100\mu\text{A}$  et des transistors pull-down ELT de la taille minimale, la réponse à une commutation ressemble à celle montrée dans la fig. 5.8(a), avec un retard de propagation de presque 10 ns. Ce résultat est obtenu par simulation, qui prend en considération également la capacité de drain des transistors reliés à la ligne dont le fusible n'est pas sauté. Dans ces conditions, la puissance globale de la puce serait de 16 mW.

Pour cette raison, chaque trait horizontal possède un deuxième pull-up, plus fort que le primaire, qui n'est activé que lorsque les entrées changent. De cette façon le pull-up primaire peut être rendu plus faible, sa puissance statique étant diminuée. En même temps la performance peut être augmentée, puisqu'elle est liée au pull-up secondaire. Le transistor secondaire est alors conçu pour fournir un courant de  $500\mu\text{A}$  une fois activé, alors que le primaire fournit  $10\mu\text{A}$  constants. La fig. 5.8(b) montre la réponse de la porte AND câblée dans ce dernier cas.

Grace à cette configuration, le retard de propagation est diminué à 3.2 ns et la puissance statique à 1.6 mW. La puissance dynamique est augmentée, puisque chaque fois que les entrées changent, le transistor secondaire est activé.

#### *Considérations sur SEUs/SETs*

Les traits horizontaux ont une capacité élevée qui est assez grande pour résister à des SETs venant des particules avec un LET au-dessous de  $25\text{ cm}^2\text{MeV/mg}$ , ce qui est suffisant dans l'application prévue (environnement des neutrons et protons). Chaque piste horizontale alimente deux inverseurs qui produisent deux copies redondantes de la même valeur venant de la porte AND câblée. Les signaux sont doublés depuis ces inverseurs jusqu'aux sorties.

#### *Conception des tampons trois-états d'entrée-sortie*

Les spécifications définissent un tampon trois-états d'entrée-sortie protégé contre les décharges électrostatiques (Electro-Static Discharge, ESD), avec une vitesse de balayage contrôlée (Slew-Rate Controlled, SRC), et un courant de sortie de 20 mA. Dans ce travail les transistors de l'inverseur tampon de sortie sont commandés séparément. Le SRC est obtenu en divisant l'inverseur final en plusieurs inverseurs parallèles. Quand les données d'entrée changent, tous les inverseurs doivent être arrêtés en même temps et allumés chacun après certain retard par rapport au précédent. De cette façon la circulation du courant sur la sortie et sur l'alimentation change lentement, avec une petite chute de tension  $Ldi/dt$  due à l'inductance parasite, évitant le bruit de commutation.

Deux signaux de commande se propagent respectivement par deux chaînes de retard composées par des inverseurs faibles (presque 300 ps de retard chacun). La fig. 5.11 page 122 montre le schéma d'un tampon trois-états SRC d'entrée-sortie divisé en 4 étapes. Le tampon de ce travail est divisé en 5 étapes, chacune capable de délivrer 4 mA. La fig. 5.12 représente le résultat d'une simulation du circuit tampon et prouve que la vitesse de balayage est presque de 10 mA/ns.

Des diodes de limitation (clamping) sont employées pour la protection contre les ESDs. En outre, les secteurs actifs reliés directement à la sortie sont entourés par un double anneau de garde.

## V–ii Layout de la puce

La taille de la puce est  $2 \times 2 \text{ mm}^2$ , alors que la dimension du noyau est approximativement  $950 \times 1150 \text{ }\mu\text{m}^2$ . Le reste de l'aire est utilisée par les tampons d'entrée-sortie et par la distribution de puissance. La fig. 5.13(a) montre la disposition interne de la puce. La puce possède 10 entrées, 8 entrée-sortie et 4 bornes pour l'alimentation et la masse. Les bornes sont distribuées sur le périmètre de la puce. Les deux couples d'alimentation et de masse sont placées sur deux côtés opposés. L'aire inutilisée est exploitée pour la disposition des marqueurs, pour le référencement et la calibration spatiale du laser brûlant les fusibles.

## VI Conclusions

Ce travail démontre la faisabilité de la conception des dispositifs programmables résistants aux radiations et dur aux SEUs. Le dispositif PLD complet a été fabriqué et subira bientôt l'essai fonctionnel et l'essai en environnement radiatif. La conception du bloc logique du dispositif FPGA dans la technologie CMOS  $0.13 \text{ }\mu\text{m}$  est finalisée et le travail continue pour accomplir l'infrastructure d'interconnexion.

Afin d'atteindre les caractéristiques désirées, plusieurs techniques de protection contre les SEUs ont été évaluées, et une approche finale a été choisie et mise en application dans plusieurs puces d'essai pour l'évaluation.

Une structure de registre robuste aux SEUs a été conçue et examinée dans une technologie CMOS  $0.25 \text{ }\mu\text{m}$  aussi bien que dans une technologie CMOS  $0.13 \text{ }\mu\text{m}$ . Le registre robuste aux SEUs a été dessiné afin d'être utilisé comme élément de mémoire dans la conception des circuits logiques programmables.

Les résultats des essais d'irradiation obtenus en technologie CMOS  $0.25 \text{ }\mu\text{m}$  ont démontré la bonne robustesse du circuit jusqu'à une LET de  $79.6 \text{ cm}^2\text{MeV}/\text{mg}$ , ce qui le rend approprié à l'environnement du LHC.

Le circuit en technologie CMOS  $0.13 \text{ }\mu\text{m}$  a montré une bonne robustesse jusqu'à la LET de  $37.4 \text{ cm}^2\text{MeV}/\text{mg}$  dans le mode d'essai statique mais avait une sensibilité augmentée en mode d'essai dynamique. La tolérance aux SEU du registre de technologie  $0.13 \text{ }\mu\text{m}$  est suffisante pour l'application comme registre de configuration mais pas comme registre d'utilisateur. Un travail additionnel de renforcement est nécessaire pour atteindre ce dernier objectif.

Une évaluation de la robustesse à la dose totale ionisante de deux structures logiques programmables est prévue dans un futur proche. Les plans du projet incluent le développement d'un logiciel capable de programmer les composants FPGA et PLD.





# Chapter 1

## Introduction

### 1.1 CERN and High Energy Physics

High Energy Physics (HEP) explores the innermost basic constituents of matter and their mutual interactions. CERN <sup>1</sup>, the European Laboratory for Particle Physics, was founded in 1954 in Geneva (Switzerland) as a joint European effort to provide a major scientific facility for particle physicists. It is today one of the world's largest and most successful scientific laboratories, as well as an outstanding example of international collaboration between its 19 Member States <sup>2</sup>.

#### 1.1.1 Accelerators and detectors

Particle physics studies are based on studying the products of collisions of particles at high energy. As particles of more heavy mass can be created when more energy is available in the center of mass, colliding particles used in the experiments should have very high speed. Particle accelerators, like a synchrotron, are used to accelerate particles to the speed needed.

Inside modern particle accelerators, beams of charged particles traveling in a vacuum pipe are pushed by appropriate electromagnetic fields. The accelerators can be linear or circular: in the second case, such as the LHC, beam bending is performed by dipole magnets accordingly to the Lorentz force law. Quadrupole and higher order magnets are used to focus the beam.

The results of a collision have then to be observed through an experimental apparatus called *detector*. A detector is usually composed by many sub-detectors with different capabilities and goals, and all of them are connected to a computer system for analysis and event reconstruction. The goal is to identify, count and trace, as many particles moving outwards from the collision point as possible. A detector together with its infrastructure provides the means to observe the particles produced in primary collisions by the interacting beams and therefore conduct an *experiment*.

---

<sup>1</sup>Once called Conseil Européen pour la Recherche Nucléaire, now is officially named as Organisation Européen pour la Recherche Nucléaire.

<sup>2</sup>Member states are Austria, Belgium, Bulgaria, Czech Republic, Denmark, Finland, France, Germany, Greece, Hungary, Italy, The Netherlands, Norway, Poland, Portugal, Slovak Republic, Spain, Sweden, Switzerland and the United Kingdom.

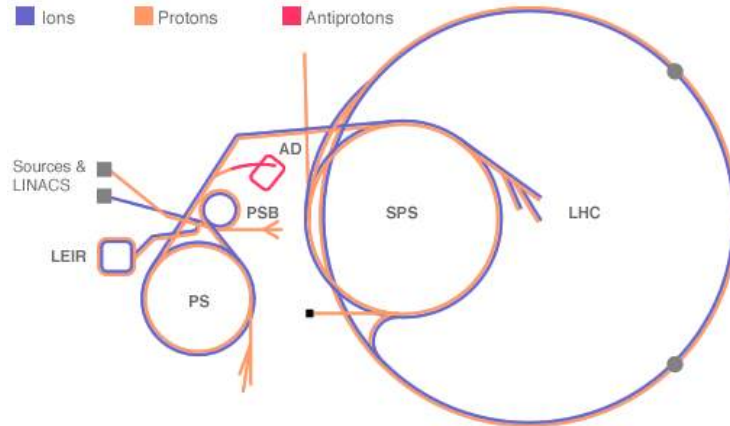


Figure 1.1: Plan of the accelerators at CERN.

### 1.1.2 The Large Hadron Collider

In year 1994 the construction of the world biggest accelerator was approved. Following this decision, the machine existing in the same underground tunnel, the Large Electron Positron collider (LEP), was dismantled in the year 2000 in order to leave place for a new, more powerful, machine: the Large Hadron Collider (LHC). While the LEP was able to reach electron-positron collisions with a centre of mass energy of 200 GeV, the LHC is designed to collide protons, going further up to 14 TeV.

The challenge, in modern particle physics research, is to probe at higher and higher collision energies, either because the basic constituents to be studied are only present at those energies, or because they are normally tied in complex aggregates and need those energies to split apart.

Reaching high energy densities means also recreating the earliest universe conditions present during the big bang. Thus, the higher collision energy physicists manage to create, the smaller dimension they study, and the earlier back in time they can observe.

The LEP was built in a 100 m underground tunnel, with the earth shielding its radiation, following a 27 km long ring. Such a big circumference was necessary because of the energy loss by bremsstrahlung: electrons and positrons emit photons when accelerated, therefore the same happens bending their trajectory; the less the trajectory is bent, the less energy they lose.

In these days the LEP is being replaced with the LHC, which employs the same existing cavern. The LHC is planned to be fully operational from year 2007 onward. LHC will make use of superconducting magnets cooled at 1.9 K installed all along the ring to bend the beams and with a nominal field of 8.33 T, allowing the circulation at the desired energy of 7 TeV of two proton beams. The two beams will run in opposite directions, but colliding only in four points, where the experiments take place. LHC is designed to be able to accelerate also lead ions, much more massive than protons, to attain collision energies of 1148 TeV. However, this will happen only later in the accelerator planned schedule.

The two proton beams will also be segmented in 2835 bunches of up to  $1.1 \cdot 10^{11}$  particles per bunch. This will make two bunches running in opposite directions to meet in the interaction points every 24.95 ns at the nominal speed. In other words the collision frequency will be 40.08 MHz.

The four experiments designed to make use of LHC are:

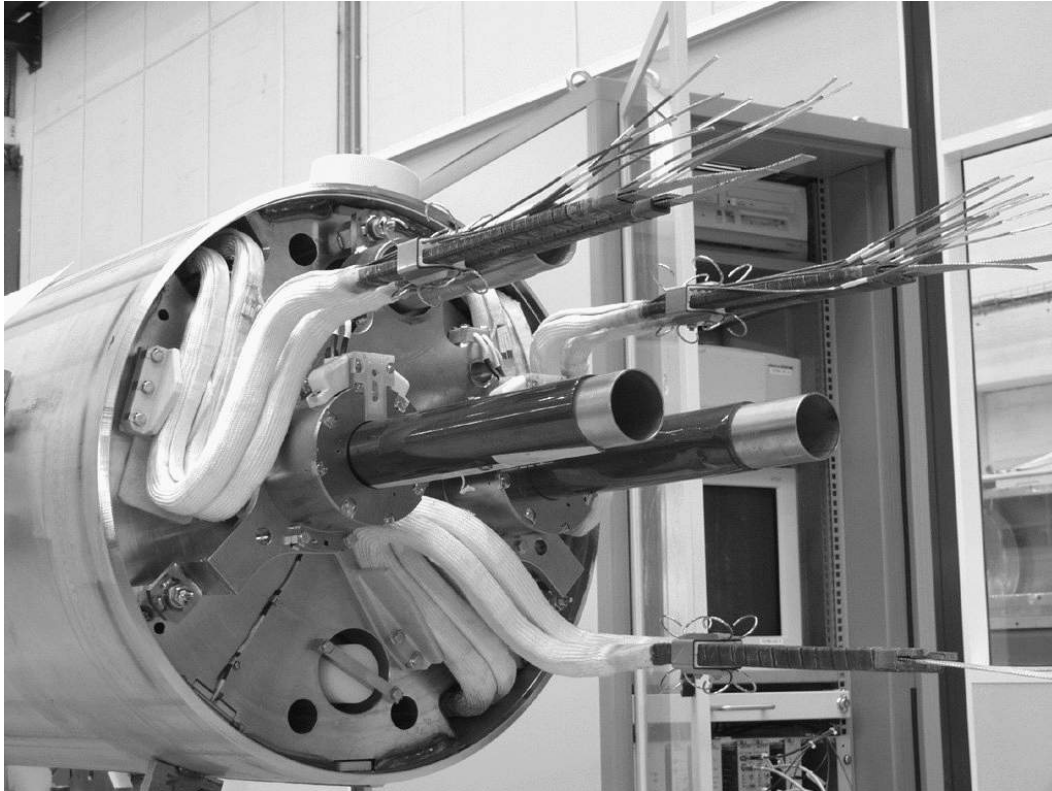


Figure 1.2: LHC accelerator section photograph.

- the Compact Muon Solenoid (CMS);
- A Toroidal Lhc ApparatuS (ATLAS);
- A Large Ion Collider Experiment (ALICE);
- LHCb.

As a typical example of an HEP experiment, in the next section the CMS apparatus will be treated in more detail.

### 1.1.3 An example of a typical HEP experiment

Figures 1.4 and 1.5 show representations of the CMS. As can be seen, it has a cylinder's shape with a diameter of 14.6 m and 21.6 m long, excluding the very forward calorimeter. Its total weight is about 14500 tonnes.

The beams run along the axis entering from the two sides, and collide in the center of the detector, the point also referred as *vertex*. The physics performance is guaranteed by its almost  $4\pi$  solid angle coverage. CMS is optimized for the detection of the expected Higgs boson.

The detector is divided into three main sections: the middle barrel and the two side identical endcaps. A 13 m long superconducting solenoid magnet generates a uniform 4 T field inside the barrel region, which bends the charged particles' trajectory in order to identify them by their momentum and charge. A return path for the magnetic flux is guaranteed by a huge iron structure, covering the whole machine, called return yoke. Inside the return yoke the magnetic field is of about 2 T.

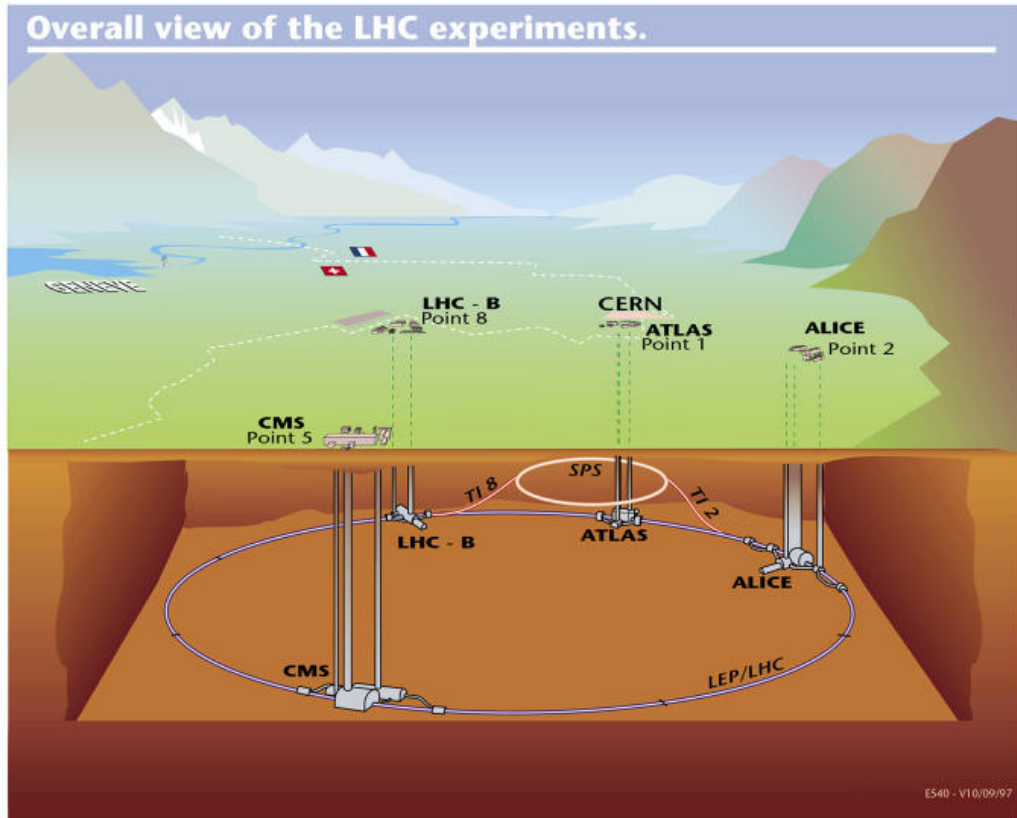


Figure 1.3: Underground view of LHC and its experiments.

The CMS apparatus is composed by several sub-detectors, which, from the inside to the outside are:

**The tracker** composed by silicon pixel detectors in the inner part, and silicon strip detectors in the outer. It traces the trajectory of charged particles with an accuracy of about  $100 \mu\text{m}$ ;

**The electromagnetic calorimeter (ECAL)** which measures the energies of electrons and photons through  $\text{PbWO}_4$  crystals. The ECAL contains also a small silicon strip detector situated in the endcaps' inner part called the *preshower*;

**The hadronic calorimeter (HCAL)** made with thick layers of copper as absorber and thin layers of plastic scintillator, it measures the energies of hadrons;

**The muon chambers** used for detecting muons, which are highly penetrating particles. The muon chambers are interleaved with the iron return yoke and are made with gaseous particle detectors.

**The very forward calorimeter** placed along the axis only in the outer barrel region, it is made with an iron/gas detector.

As mentioned before, the bunch crossing frequency is 40.08 MHz, with an average of 20 inelastic events<sup>3</sup> occurring each time. Given the very large number of electronics

<sup>3</sup>An inelastic event is collision in which particles other than those the two protons participating in the primary collision are found in the products of the collision itself.

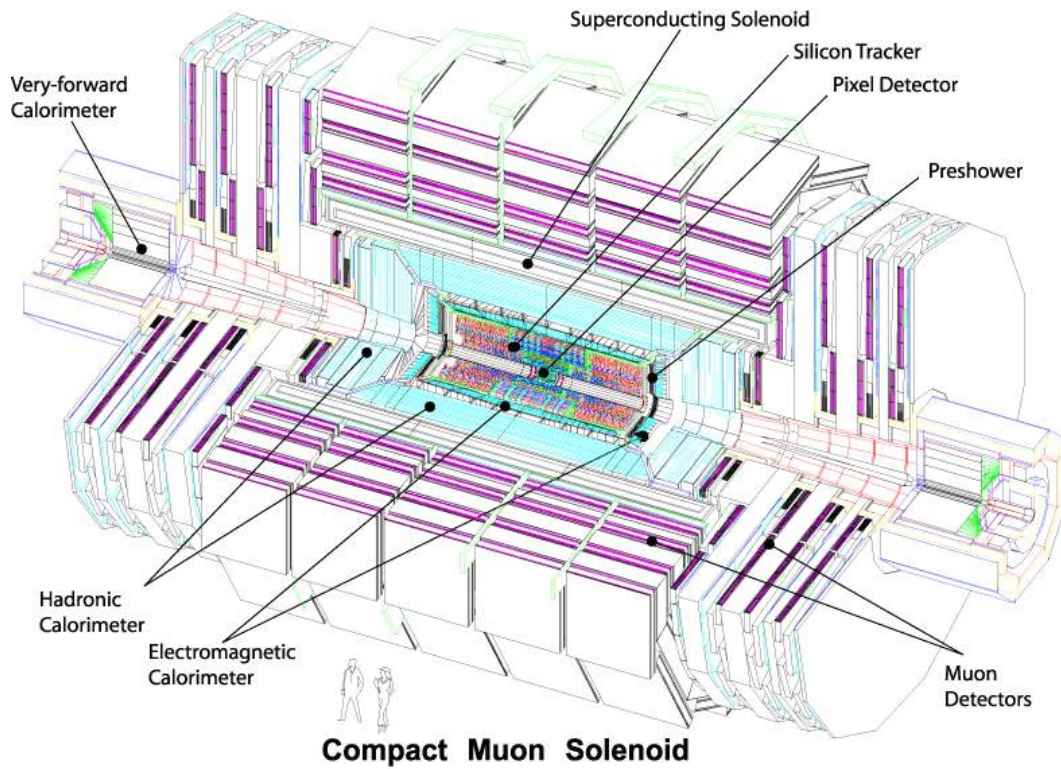


Figure 1.4: View of CMS with its parts and sub-detectors.

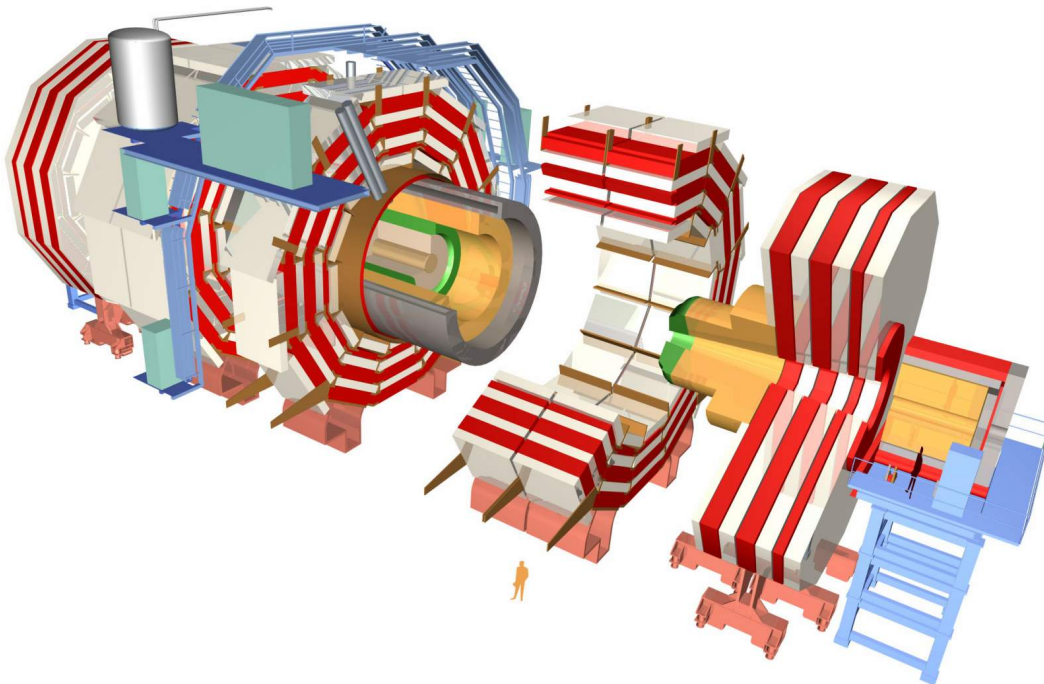


Figure 1.5: 3d split view of the CMS detector.

channels in all sensors in the experiment, the amount of data coming out from the apparatus is enormous. Only a small fraction of the collisions is interesting from the physics point of view, therefore a filtering of the data has to be performed. It is

clearly necessary to do this in real-time, reducing the rate to 100 Hz, which is the maximum rate that should be stored for off-line analysis [CMS 94].

All these jobs are carried out by the trigger and data acquisition system of the experiment, which selects the useful events<sup>4</sup>, rejecting the rest, by evaluating a subset of the data.

## 1.2 Radiation environment in the LHC

In order to maximize the number of interesting events obtained from the experiments, the LHC accelerator is designed to reach a very high peak luminosity<sup>5</sup>:  $10^{34} \text{ cm}^{-2}\text{s}^{-1}$  for protons and  $1.95 \cdot 10^{27} \text{ cm}^{-2}\text{s}^{-1}$  for lead ions. This will bring, in the case of protons, to an average production of  $8 \cdot 10^8$  inelastic proton-proton collisions per second, creating an extremely hostile radiation environment.

### *Studied spectra*

In addition, at LHC the high beam energy combined with the very high luminosity results in numerous intense cascades, which will end up in an immense number of low-energy particles. In fact, particles energies exceeding 10 GeV are expected to be very rare in the detectors' barrel, and also in the major part of the endcap. Therefore the radiation studies focused on the energy range around 1 GeV and below.

### *Induced radioactivity*

While induced radioactivity was negligible in electron-positron colliders (like the LEP), it will be a major concern at LHC. The hadrons produced by the collision will interact with the nuclei composing the detector infrastructure resulting in residual nuclei. Roughly 30% of these inelastic hadronic interactions create long-lived radionuclides which contribute to the dose rate from induced activity in the experimental area. This activity decreases relatively slowly after the end of irradiation, so that even long cooling times do not significantly improve the situation. Activation can also occur through neutron interactions, especially in the thermal regime. However, except for a few special materials, this is usually a minor contribution.

### 1.2.1 Radiation environment in the experiments

As summarized in Table 1.1 Total Ionizing Dose<sup>6</sup> (TID) values in the CMS experiment could be high, in the worst case conditions, up to 50 Mrad in the 10 years of expected lifetime of the experiment<sup>7</sup>. The detectors' front-end electronics has then to stand this enormous amount of radiation, especially in the inner tracker and in the ECAL endcaps, where the levels are higher.

In addition, the silicon layers of the outer tracker and the preshower detector will be exposed to the neutron albedo from the electromagnetic calorimeter. Dose rates drop rapidly when moving from the inner maximum deeper into the calorimeters. Figure 1.6 shows total doses and particle fluences within the experiment. The figure

<sup>4</sup>All the data relative to one bunch crossing is referred as an *event*.

<sup>5</sup>The luminosity can be thought as the number of particles per unit area in the interaction point of the two beams.

<sup>6</sup>Total dose is defined as the total absorbed energy per unit mass. Although the S.I. unit for total dose is the *Gray* (Gy), where  $1 \text{ Gy} = 1 \text{ J/Kg}$ , in the high energy physics community the old unit *rad* is still used:  $1 \text{ rad} = 10^{-2} \text{ Gy} = 100 \text{ erg/g}$ .

<sup>7</sup>For comparison, the average natural background radiation dose rate on earth is about 25 rad/year, while the dose rate absorbed by silicon devices in space on a geostationary orbit is approximately 30 krad/year.

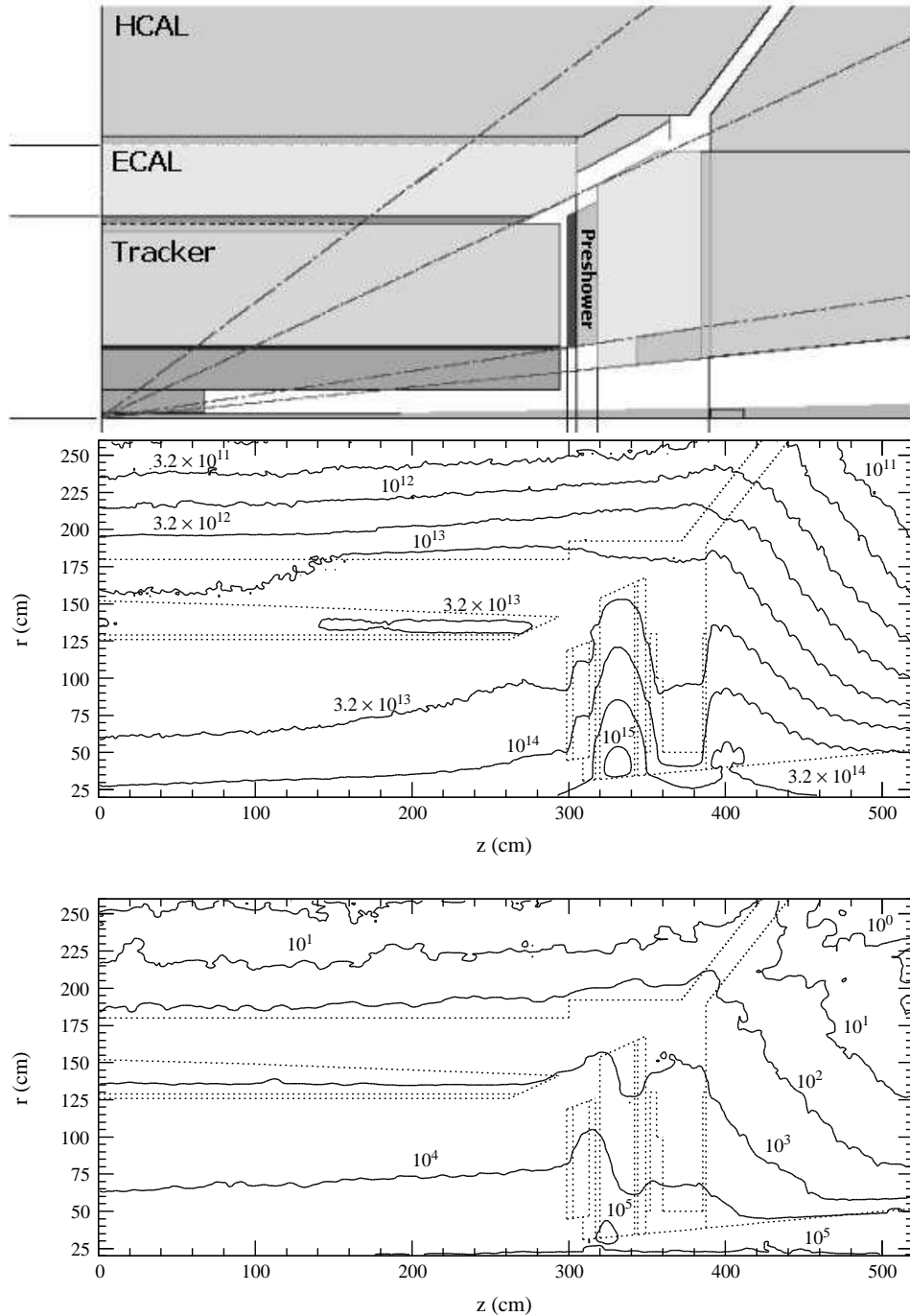


Figure 1.6: Fluence of neutrons (with energy above 100 keV) and charged hadrons in  $\text{cm}^{-2}$  (upper plot) and radiation dose in Gy (lower plot), in the region inside the solenoid [CMS 97]. The dotted lines in the graphs indicate the geometry shown above.

<i>Sub-detector</i>		<i>Total dose</i> [Mrad]	<i>Neutron fluence</i> [ $10^{14}$ cm $^{-2}$ ]	<i>Charged hadron fluence</i> [ $10^{14}$ cm $^{-2}$ ]
Tracker	at 7cm	35.	1.	10.
	at 22cm	6.5	.35	1.5
	at 75cm	.7	.15	.25
ECAL	barrel	.5	.5	.005
	endcaps	20.	10.	.6
HCAL	barrel	.02	.1	-
	endcaps	2.5	5.	-
Muon chambers		.005	.025	-
Forward calorimeter		500.	250.	-
Experimental hall		.0005	.001	-

Table 1.1: CMS sub-detectors' radiation environment in the 10 years experiment lifetime [Giraldo 98], equivalent to  $5 \cdot 10^7$  s. The reported doses and fluences are the maxima inside each sub-detector.

also clearly demonstrates that the ECAL's crystals are the most intense source of fast neutrons inside the CMS.

### 1.2.2 Radiation tolerant ICs

It is then clear that the integrated circuits used for the front-end<sup>8</sup> electronics of the detectors must be resistant to radiation.

The need of these kind of circuits for the various applications mentioned above led, in the past, to the development of special technologies, called *radiation hardened*, where particular processing methods are used in order to improve their radiation tolerance. Modifying the *process* steps is one of the three ways to improve the radiation tolerance of an integrated circuit. The two other possibilities are to use special *layout* techniques or special *circuit and system* architectures.

#### *The gate oxide*

In a metal-oxide-semiconductor (MOS) transistor, the most sensitive part to TID-generated radiation effects is the gate oxide. One way to reduce those effects is to reduce the gate's thickness, which is the natural trend in modern technologies. The market of memories, microprocessors and, in general, digital integrated circuits, has driven a very fast technological evolution in the past 25 years, which has led to today's deep submicron devices with less than 2 nm gate oxide thickness.

#### *Commercial vs rad-hard*

This suggests the possibility of using modern commercial CMOS technologies in radiation environment without introducing or modifying any particular process step. Hardening a technology by introducing special processing steps is generally not affordable for HEP customers since foundries would not modify their processes for such a small market without increasing considerably the prices.

<sup>8</sup>Usually, in the CERN experiments context, the electronics inside the experiments is referred as *front-end* (FE). The detectors and their immediately close analog equipment are instead called *very front-end* (VFE).



Having a radiation tolerant gate oxide does not resolve all the possible problems when irradiating an integrated circuit made in a standard deep submicron technology (increase in leakage current, soft-errors, etc.). To solve these problems one can still adapt the layout and the architecture of the circuits and of the system.

The use of deep submicron CMOS technologies has several beneficial aspects, such as speed, reduced power consumption, high level of integration and high volume inexpensive production. Moreover commercial technologies do not suffer from the problems of radiation hardened technologies, which are more expensive and less advanced (usually a couple of generations behind). Last but not least, the availability of some radiation hardened technologies in the future is not certain, and cases have already been experienced of foundries stopping the production of their radiation hardened processes due to the drop of demand.

For this reason, in 1996, the CERN's microelectronics group started to investigate the possibility of using a commercial CMOS technology to integrate the circuits to be used in the detectors. The very promising results obtained led, at the end of the same year, to the proposal of a Research and Development project<sup>9</sup> which was approved in March 1997. The aim of the project was to assess the improved radiation tolerance of submicron CMOS technologies and to study the use of design and layout techniques to increase it further. At that time, 0.7  $\mu\text{m}$  technology was the state of the art, but since then the evolution has been followed characterizing 0.5, 0.35 and 0.25  $\mu\text{m}$  technologies.

As confirmed in the RD49 status reports [Jarron 99b, Jarron 00], the results were very successful, and allowed the design of integrated circuits which could stand doses of 30 Mrad and beyond [Snoeys 00, Jarron 99a]. At the present time a 0.13  $\mu\text{m}$  technology is being studied, while a rich 0.25  $\mu\text{m}$  digital library is commonly used for design.

*CERN RD49*

### 1.3 Motivation and objectives of this work

The progress in microelectronic technologies applied to programmable logic circuits has allowed to decrease the costs and the development time of digital electronics in the industrial sector as well as in the space and avionics sectors. The use of such devices is also appealing for HEP detectors placed in the vicinity of high-luminosity particle accelerators such as the LHC. As mentioned previously, the harsh radiation environment present in these detectors makes Commercial Off-The-Shelf (COTS) components unsuitable for the application and requires the design of custom-designed circuits. Chapter 2 offers an introduction on the radiation effects on integrated circuits and hardening solutions against these effects.

The most advanced programmable circuits are Field-Programmable Gate Arrays (FPGAs), which will be introduced in chapter 3. SRAM-based FPGAs are inherently flexible to meet multiple requirements and offer significant cost and development time advantages. They can be reconfigured after the commissioning of the systems to correct errors or to improve performance. SRAM-based FPGAs can be implemented in standard CMOS processes while FLASH-based FPGAs require special non-volatile processes.

Many studies have been done on the radiation effects on commercial FPGAs, proving them to be often sensitive to both Total-Ionizing Dose (TID) and Single-

---

<sup>9</sup>“CERN RD49 – Study of the radiation tolerance of ICs for LHC”

Event Upsets (SEUs). Results of these studies will be presented in section 3.3.

FPGAs are critically sensitive to SEUs due to the large amount of memory elements located in these devices. These must be strongly protected to avoid errors during run time. There are two main techniques to mitigate the SEU radiation effects: introducing redundancy in the Hardware Description Language (HDL) program or cell level architectural hardening. Special constructs in the HDL allow introduction of redundancy in the user logic. These techniques reduce drastically the available circuitry resources of the FPGA and require complex reconfiguration schemes to avoid corruption of the configuration data. Unlike this approach, the objective of this work is the development of programmable circuits where SEU insensitivity is built-in at the storage cell level thus, not requiring the user to exploit any special technique for SEU protection.

Programmable Logic Devices (PLDs) are small components which can implement logic functions equivalent to  $\approx 50$  gates. Although PLDs are considered nowadays surpassed by FPGAs, they are still favourable in some applications implementing simple state machines, glue logic circuitry and providing fixes for system design bugs at the late stages of a project. PLDs also suffer from TID. PLDs are in general not affected from SEUs in the configuration storage, but the user register can still be corrupted and therefore needs to be protected.

This work focuses on the design of an SRAM-based FPGA and a fuse-based PLD that are SEU-robust, radiation-tolerant and industry-compatible, in order to provide the HEP community two devices suitable for the construction of particle physics experiments.

In order to reach the desired specifications, several SEU-hardening techniques were evaluated, as presented in section 2.4, and a final approach was chosen and implemented in several test chips for assessment. Tests were conducted in a heavy-ion beam facility and test results are presented in chapters 4 and 5.

## Chapter 2

# Radiation Effects and Hardening

### 2.1 Total Ionizing Dose effects

In this section the effects of total ionizing dose (TID) on matter and on silicon devices are analyzed and solutions for practical applications are proposed.

#### 2.1.1 Radiation effects on matter

The manner in which radiation interacts with solid materials depends on many factors, but the three main criteria of classification are charge, mass and energy of the incident particle. Protons and electrons are charged particles, while neutrons and photons are neutral particles. From the mass point of view, instead, protons and neutrons are heavy particles, while electrons are light particles.

**Charged particles** interact through the Coloumb force with the target material atoms inducing ionization or atomic excitation. Neutral particles instead do not exhibit this kind of behavior.

**Massive particles** can collide with the nuclei of the target material causing displacement, excitation or nuclear reactions if the energy is enough.

**Electrons** also generate Bremsstrahlung (X-rays) when decelerating into the target.

**Photons** have zero mass and no charge, therefore they have a special behavior with respect to other particles. They can interact, ordered by energy of the photon:

- by photoelectric effect, in which an electron of the target atom changes energy state, possibly ionizing the atom, and the photon is completely absorbed;
- by Compton effect, in which an electron of the target atom is set free and a residual photon is emitted;
- by electron-positron pair creation (above 1.024 MeV).

In practice, the effects of radiation on the materials involved in microelectronic devices production can be grouped in two classes: ionization effects and nuclear displacement [Braunig 93].

*Semiconductors  
and insulators*

**Ionization** creates electron-hole pairs. The number of pairs created is directly proportional to the total absorbed dose. For this reason, the studies on the effects of ionization refer only to this quantity and not to the type of particle chosen.

**Displacement** gives origin to crystal defects, most of which are Frenkel pairs. In  $\text{SiO}_2$  at room temperature, 90% of the Frenkel pairs recombine within a minute after the end of irradiation. MOS transistors are almost entirely insensitive to displacement damage, since they are devices whose conduction is based on the flow of majority carriers below the silicon-oxide interface, a region which does not extend deeply in the bulk. This phenomenon has therefore a limited importance.

### 2.1.2 Radiation effects on MOS transistors

*Positive charge trapped in  $\text{SiO}_2$*

As mentioned above, MOS transistors are more sensitive to ionization than to displacement damage. In the gate (metal or polysilicon) and in the substrate the electron-hole pairs generated quickly disappear, since these are materials with small resistance. On the other side, in the oxide, which is an insulator, electrons and holes have a different behavior, as their mobility differ by  $10^5$  to  $10^{10}$  times<sup>1</sup>.

Only a fraction of the induced electron-hole pairs will recombine immediately after being generated, while the rest will be separated by the electric field. In the case of a positive bias applied to the gate, the electrons drift to the gate electrode in a very short time whereas the holes move towards the Si– $\text{SiO}_2$  interface with a very different slower transport phenomenon<sup>2</sup>. Then, close to the interface, but still in the oxide, some holes may be trapped, giving origin to a fixed positive oxide charge  $Q_{ox}$ .

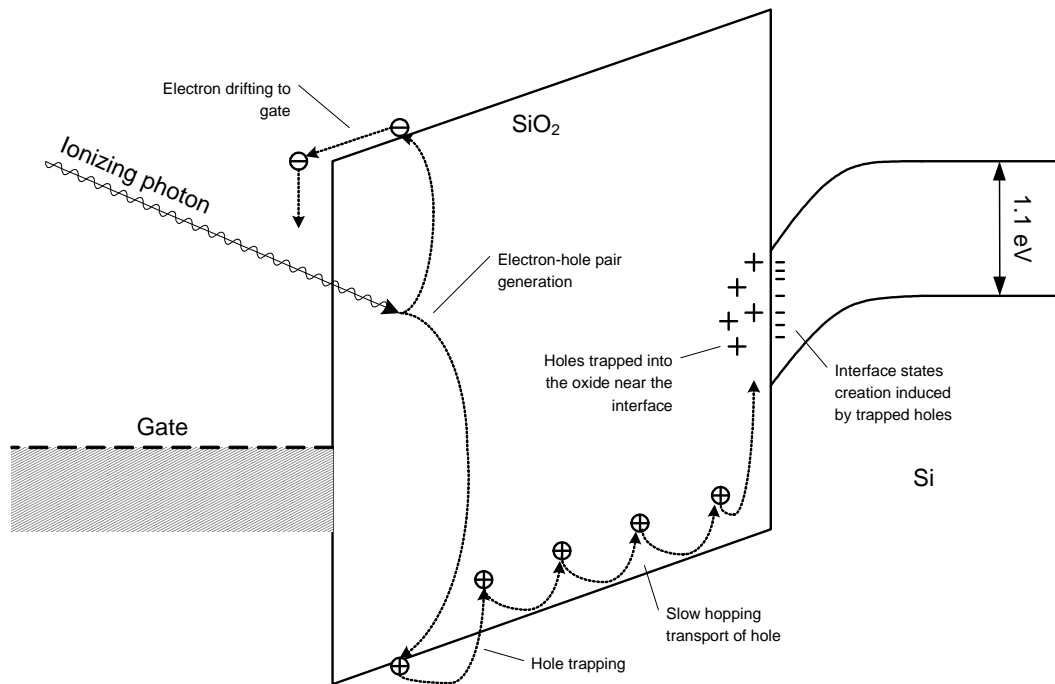


Figure 2.1: Band diagram showing the transport and trapping of holes in the oxide.

The amount of trapped charge is proportional to the number of defects in the silicon dioxide: depending on the oxide quality and on the electric field, the fraction

<sup>1</sup>Typical  $\text{SiO}_2$  electron mobility at room temperature is  $20 \text{ cm}^2\text{V}^{-1}\text{s}^{-1}$ , while for holes it depends strongly on the temperature and on the electric field, and ranges between  $10^{-4}$ – $10^{-11} \text{ cm}^2\text{V}^{-1}\text{s}^{-1}$ .

<sup>2</sup>The transport of holes in  $\text{SiO}_2$  is based on the concept of *small polaron hopping* [Boesch 85, McLean 89], which will not be discussed in this thesis.

of trapped holes varies from 1% to 100% [Boesch 86, Anelli 00]. The non trapped holes which reach the interface, will recombine with electrons coming from the silicon. Moreover these electrons may tunnel from the silicon surface into the oxide and recombine with trapped holes, giving origin to a *tunnel-effect-based annealing* [McWhorter 90]. This effect makes the trapped charge quantity to vary with the absorbed dose rate and its history.

The positive oxide charge lowers the threshold voltage  $V_T$  in n-channel transistors, since it attracts more electrons to form the silicon inversion. In p-channel transistors the threshold voltage absolute value is increased, or, in other words,  $V_T$  is more negative.

Ionizing radiation also induces the creation of interface traps. These traps have an energy laying in the silicon energy gap. Experiments indicate that the major part of the traps present above midgap are acceptors, while traps below are donors<sup>3</sup> [Winokur 89, Anelli 00]. Filling those states gives rise to a interface trapped charge  $Q_{it}$ .

For this reason, in both p- and n-channel MOS transistors, the threshold increases (in absolute value), after irradiation, due to the creation of new interface traps<sup>4</sup>. Again, radiation induced trap generation is strongly dependent on the processing steps of MOS devices. Thus one of the fundamental steps for the fabrication of radiation hardened devices is the control of the gate oxide quality.

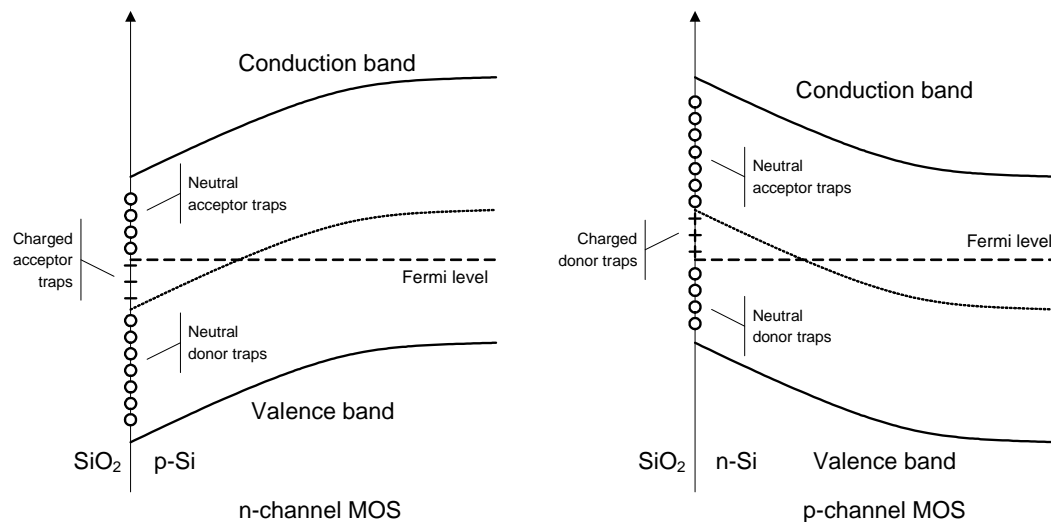


Figure 2.2: Band diagram showing the behavior of interface states for an n-channel and a p-channel transistors. The gate bias is positive for the n-channel while negative for the p-channel.

The two phenomena described above cause the threshold voltage to vary with

*Threshold voltage shift*

<sup>3</sup>A donor trap releases an electron when it passes from below to above the Fermi level. Donor traps are neutral when full and positively charged when empty. An acceptor trap captures an electron when it passes from above to below the Fermi level. Acceptor traps are neutral when empty, negatively charged when full.

<sup>4</sup>Considering an n-channel MOS transistor working in inversion, the acceptor traps in the upper part of the gap, being below the Fermi level, will be filled by electrons and then negatively charged, making necessary an higher gate voltage to have the same channel inversion.

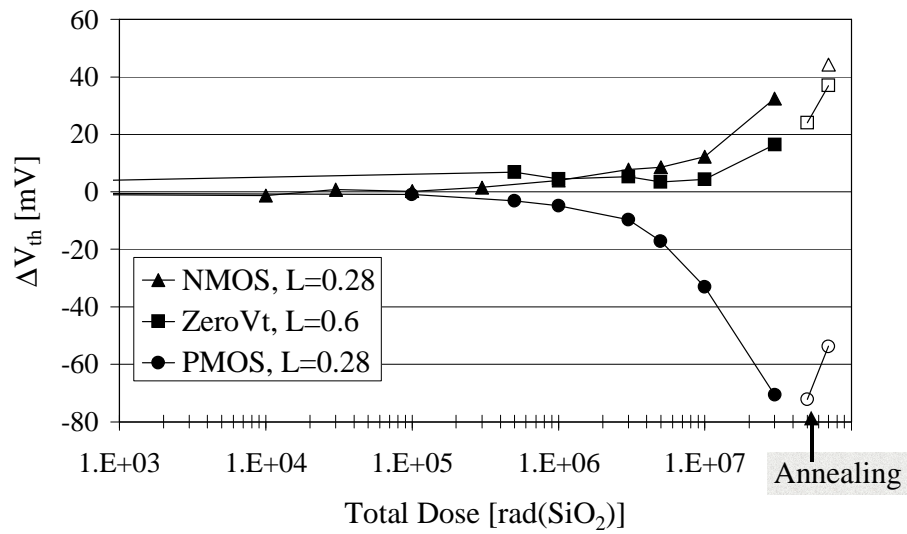


Figure 2.3: Threshold voltage shift of enclosed NMOS, enclosed Zero- $V_T$  NMOS, and normal PMOS transistors in  $0.25 \mu\text{m}$  technology as a function of the total dose [Faccio 98].

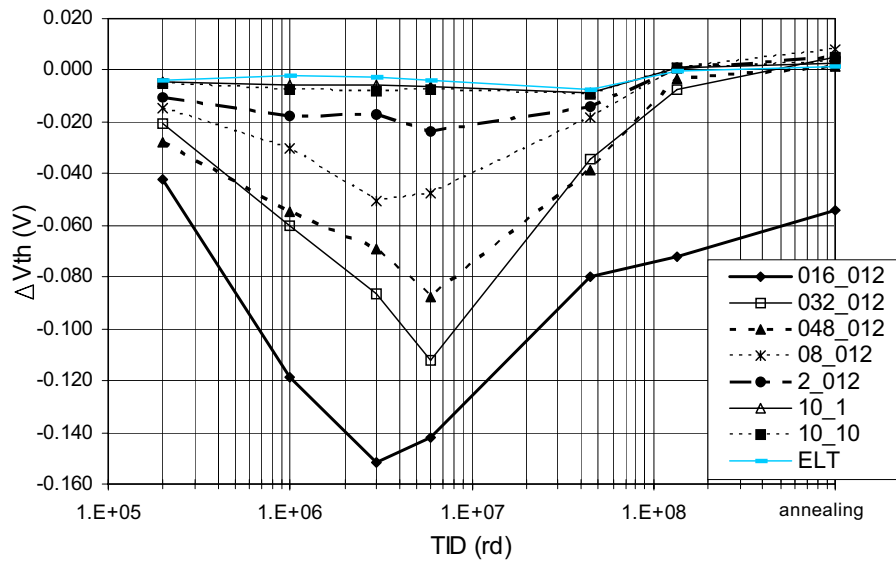


Figure 2.4:  $V_T$  shift with TID for different NMOS transistor size, up to 136 Mrad, in a  $0.13 \mu\text{m}$  technology. The last point refers to full annealing at  $100 \text{ }^\circ\text{C}$  [Faccio 05].

irradiation. While p-channel transistor experience only an increase of  $V_T$ , in n-channel transistors it can decrease, increase, or even be stable, depending of which is the major effect between the positive oxide charge and interface traps. Moreover the  $Q_{ox}$  is influenced by the thickness of the oxide and the dose rate: in oxides thinner than 7 nm, at low dose rates,  $Q_{ox}$  is in general negligible with respect to  $Q_{it}$ . On the other hand, in recent technologies, the Shallow Trench Isolation (STI) oxide present at the sides of the channel contributes to  $Q_{ox}$  and it cannot be neglected.

Thus, in some technologies, like 0.25  $\mu\text{m}$  and below, where the gate oxide is thin and the width of the device is still large, the threshold voltage shows only an increase with irradiation [Anelli 00], while in the other technologies it had more complex behavior related to the balance between  $Q_{it}$  and  $Q_{ox}$ . For a 0.25  $\mu\text{m}$  technology, the absolute value  $V_T$  increase, as Figure 2.3 shows, is anyway less than 80 mV after 30 Mrad irradiation.

In 0.13  $\mu\text{m}$  technology instead, the  $Q_{ox}$  effect given by the STI edges dominates at low TID ( $\approx 40$  Mrad) while  $Q_{it}$  takes over at higher TID, giving a  $V_T$  curve, shown in Figure 2.4, which initially decreases and then self-anneals. From the figure it is clear that it is preferable to use large transistors which suffer less from the edge STI effects.

In MOS devices, a thick oxide is used to isolate between different devices and, within the same device, between the source and the drain [Kuo 99]. Usually the first is referred to as field oxide, while the second as lateral oxide. In many technologies these oxides are made in the same process step, like, for example, the LOCAl Oxidation of Silicon (LOCOS). In deep submicron processes, the thick oxide is often made with the STI technique, which guarantees a better quality than the LOCOS.

Since the lateral oxide is much thicker than the gate oxide, it suffers more of radiation-induced positive trapped charge. This can form a parasitic path near the gate's sides connecting the drain to the source, increasing, in practice, the leakage current. As mentioned before, positive  $Q_{ox}$  lowers the threshold only in n-channel transistors, thus only in those transistors a post-irradiation leakage current is observed. In a 0.25  $\mu\text{m}$  technology this current can grow up to  $\approx 7 \mu\text{A}$  after 10 Mrad irradiation, an unsuitable value for the fabrication of any chip! As Figure 2.5 shows, this technology can be used without any special layout technique up to 200 krad, but not over [Faccio 98].

In 0.13  $\mu\text{m}$  technology instead the leakage current can grow up to  $\approx 200$  nA, with an annealing at high TID, as depicted in Figure 2.6. The effects also depend on the dose rate, having smaller effects at smaller dose rates. This is an indication that fabricating linear transistors instead of ELT could be allowed. The leakage current is independent of the width of the devices, since the edges are the same for any width.

## 2.2 Hardening against TID

The choice of using a deep submicron technology guarantees itself a radiation hardened gate oxide. What is therefore necessary is to solve the problems related to the n-channel devices' field and lateral oxide degradation after irradiation.

### 2.2.1 Layout techniques

The primary problem which has to be addressed is the leakage current inside

*Leakage current  
increase*

*Enclosed Layout  
Transistors  
(ELTs)*

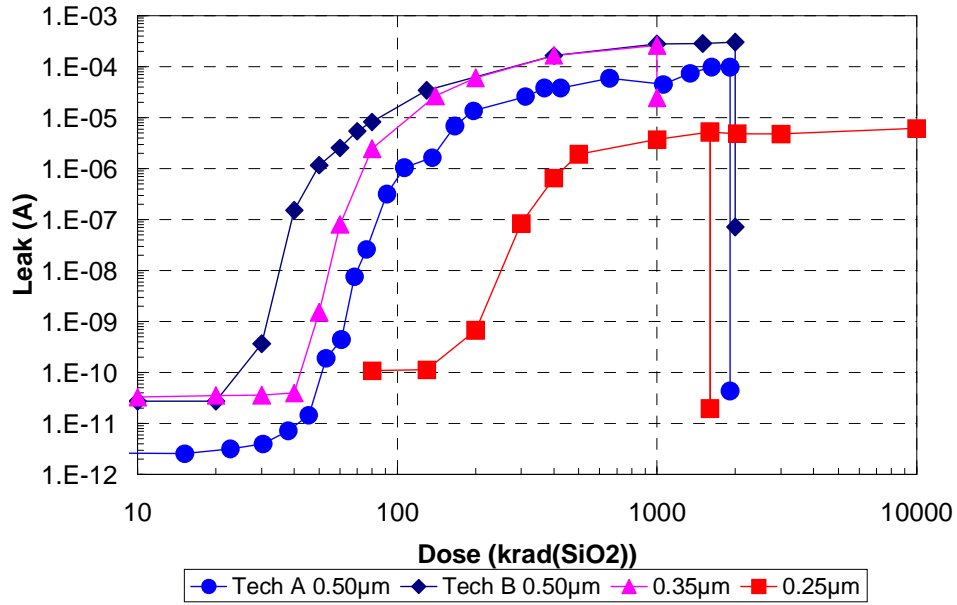


Figure 2.5: Leakage current for normal devices in various technologies (0.50, 0.35 and 0.25  $\mu\text{m}$ ) [Anelli 97]. The measurement was taken with  $V_{DS} = V_{DD}$ . The lower data point on the plot at about 2 Mrad represents the value after annealing in an oven.

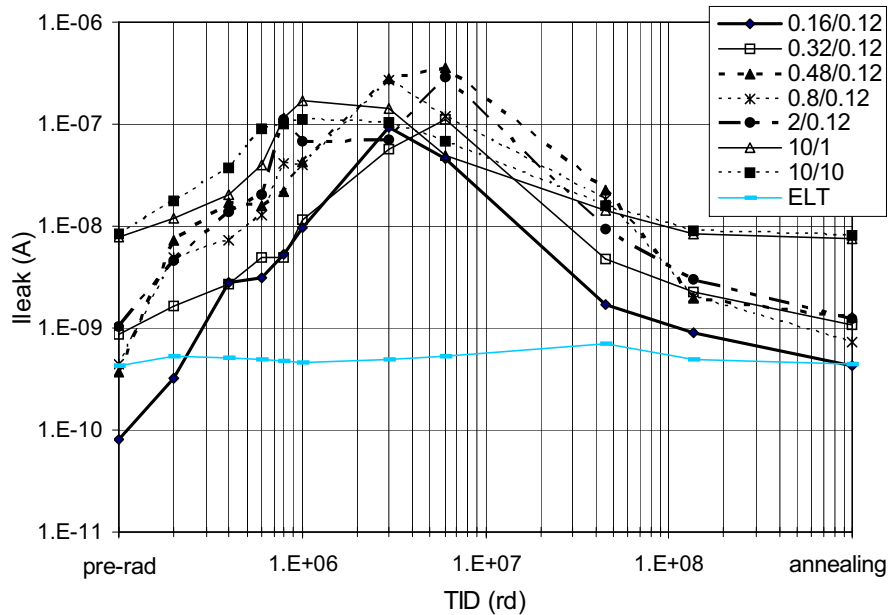


Figure 2.6: Evolution of the leakage current with TID for different NMOS transistor size, up to 136 Mrad, in a 0.13  $\mu\text{m}$ . The last point refers to full annealing at 100  $^{\circ}\text{C}$ . The first point to the left is the pre-rad value [Faccio 05].



n-channel devices. The solution adopted in CERN’s microelectronics group is to use “enclosed layout transistors” (ELTs, also called edgeless). As shown in Figure 2.7, in this case the parasitic path between the source and the drain is eliminated, as well as the lateral oxide.

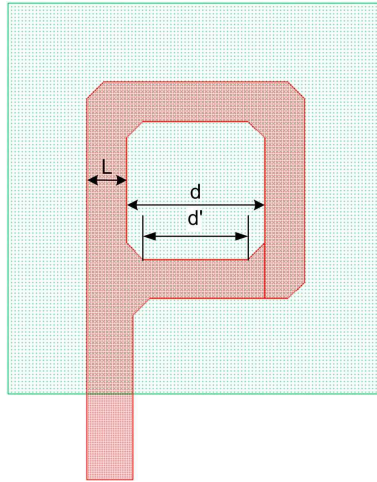


Figure 2.7: Enclosed Layout Transistor. The drain is conventionally in the center while the source is outside the circular gate.

The major disadvantages of this layout style are larger area and increase in capacitances. Moreover, the choice on the  $W/L$  ratio is limited, since  $W$  has to be enough big to allow the inner active contact to be placed.

ELTs have been used in the early days of CMOS [Dingwall 77] and their effectiveness in preventing leakage currents in irradiated integrated circuits is well known. Their intensive use in CERN’s applications lead to the investigation of many issues important for a designer, such as modelling the effective  $W/L$  ratio<sup>5</sup>. There is a wide range of possible enclosed shapes: squared, octagonal, squared with corners cut at 45 degrees and all of them can have a different behavior and require a separate model. To simplify the problem, one specific shape was chosen, compatible with the design rules of the process: square with corners cut at 45 degrees so that the size of the cut is constant for all the gate lengths (see Figure 2.7).

The second problem which can be solved with a layout technique is the leakage between different devices [Anelli 00]. This is done surrounding each n-channel device with a p+ guard ring. This method has been verified to be very effective but the drawback is again the big consumed area. Moreover, guard rings avoid the generation of SEL by lowering the gain of the parasitic NPN bipolar transistor.

*Guard rings*

<sup>5</sup>As described in [Giraldo 98], the model for the effective  $W/L$  of enclosed transistor, if applied to the shape in Figure 2.7, leads to the following expression for the aspect ratio:

$$\left(\frac{W}{L}\right)_{eff} = 4 \frac{2\alpha}{\ln \frac{d'}{d' - 2\alpha L_{eff}}} + 2K \frac{1 - \alpha}{1.13 \cdot \ln \frac{1}{\alpha}} + 3 \frac{\frac{d-d'}{2}}{L_{eff}}$$

where  $\alpha$  is constant usually set to 0.05, while  $K = 7/2$  for short channel transistors ( $L \leq 0.5 \mu\text{m}$ ), otherwise  $K = 4$ . To derive this expression, the enclosed transistor is decomposed into three parts. The first corresponds to the linear edges of the transistor, the second to the corners without the 45 degrees cut, which then is taken into account in the third part. It can be shown that the minimum reachable aspect ratio is around 2.26 with this geometry.

### 2.2.2 Circuit and system techniques

While designing circuits for radiation environment applications, one must take into account and foresee the drift of the circuit's operating point due to absorbed total dose. For digital circuits, the synchronous mode of operation limits the sensitivity to electrical parameters' variation [Anelli 00].

### 2.2.3 Radiation tolerant digital standard cells libraries

In order to help in the design of complex digital ICs, a digital library has been designed and tested in a  $0.25 \mu\text{m}$  technology [Marchioro 98, Kloukinas 98] while a commercial library is employed in  $0.13 \mu\text{m}$ . Only the  $0.25 \mu\text{m}$  library exploits radiation hardening techniques.

The basic features of the technologies are given in Table 2.1.

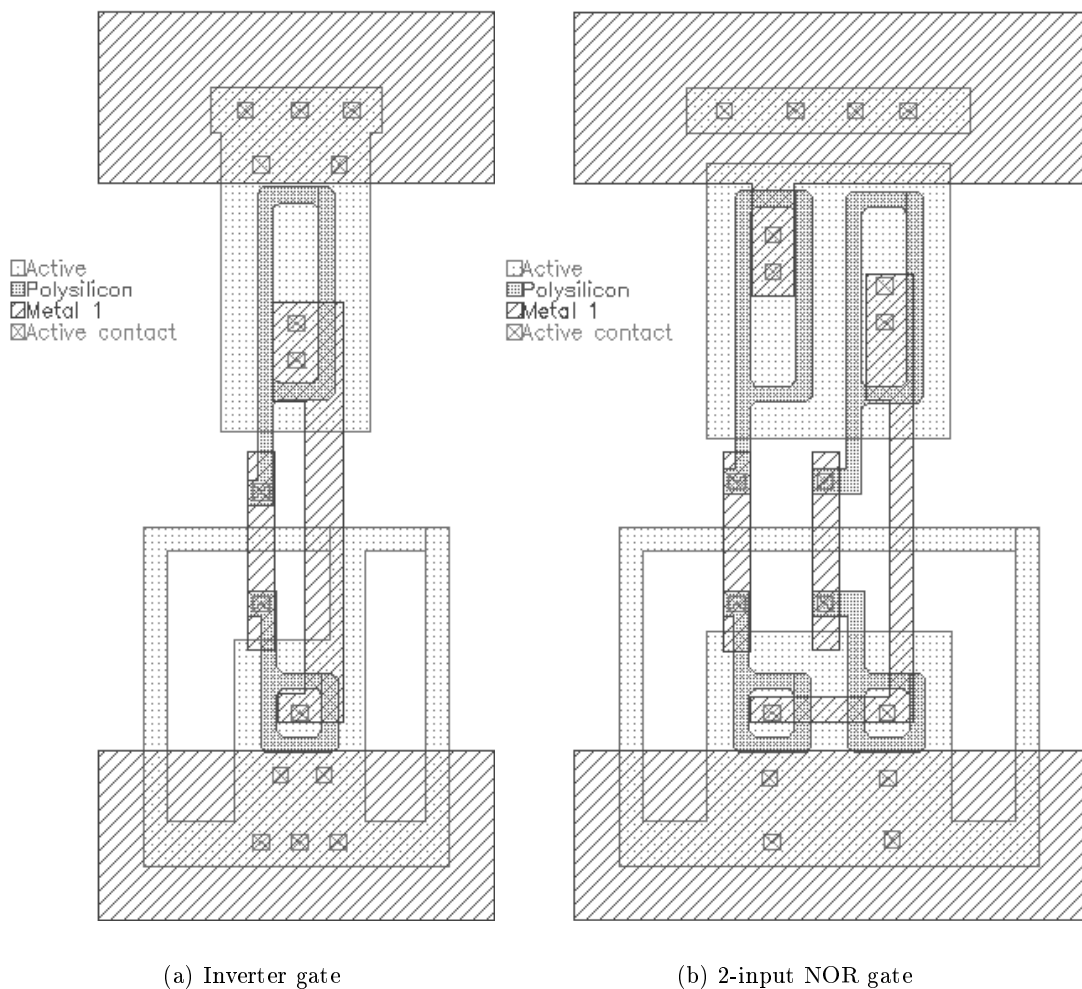


Figure 2.8: Library cells in  $0.25 \mu\text{m}$ .

The standard cells are designed to be abutted one to the other in horizontal rows. Figure 2.8 shows two  $0.25 \mu\text{m}$  library cells.

The power rails are routed in the first metal layer horizontally all along the rows; great effort was spent to keep intracell interconnections on the first metal layer, leaving the rest of the metal layers for global routing. For that purpose the salicided

Minimum lithography	0.240 $\mu\text{m}$	0.120 $\mu\text{m}$
$L_{eff}$	0.180 $\mu\text{m}$	0.098 $\mu\text{m}$
$V_{DD}$	2.5 V	1.2 V
Gate oxide thickness	5.0 nm	3.0 nm
Process	Twin well CMOS	Twin well CMOS
Device isolation	Shallow trench (STI)	Shallow trench (STI)
Salicidation (n+, p+)	Ti	Co
Interconnectivity	2 to 5 metal layers	5 to 8 metal layers
Standard cell pitch	16.0 $\mu\text{m}$	3.6 $\mu\text{m}$
Horizontal M1 tracks	15	9
Typical x1-inverter input capacitance	25 fF	1 fF
M1 wiring capacitance	0.20 fF/ $\mu\text{Fm}$	0.26 fF/ $\mu\text{Fm}$

Table 2.1: Technology features.

polysilicon layer was used as a local intracell interconnect, but since polysilicon cannot be allowed to cross the guardrings, this layer was used only for horizontal routing.

The area penalty paid for ELT style and the guardrings is anyway mitigated by the small feature size of the technology: the only alternative to this approach would be to use process radiation hardened technologies which offer overall a much smaller device density.

The libraries contain combinatorial logic gates, like NANDs and NORs, as well as flip-flops and latches. A set of I/O pads is also available.

## 2.3 Single-Event Effects

Single-event effects [Kerns 89] are phenomena generated by one single highly energetic charged ( $> 1$  MeV) particle passing through a device. The particle produces a track of ionization, whose length depends on the atomic number of the material traversed and the initial energy, and where mobile charge carriers are created (one electron-hole pair per  $E_{eh} = 3.6$  eV in Si).

### 2.3.1 Single-Event Latch-up (SEL)

Latch-up is a destructive effect which can occur because of the parasitic thyristor formed by the parasitic junction structure built in some CMOS IC (shown in fig. 2.9). This phenomenon is usually avoided with process and layout techniques, like for example placing well contacts very close to the devices' source. Even though, it can happen that a ionizing energetic particle passing through the device deposits charge inside the parasitic thyristor, initiating the positive feedback and causing it to turn on. This effect is called "single event latch-up" (SEL). Its importance is limited in deep-submicron technologies since the highly doped substrates and the presence of trench isolation between wells deteriorate the parasitic thyristor.

### 2.3.2 Single-Event Upset (SEU)

Ionizing particles can also change the state of a memory circuit and cause information to be lost in flip-flops: this phenomenon is called "single event upset" (SEU) and

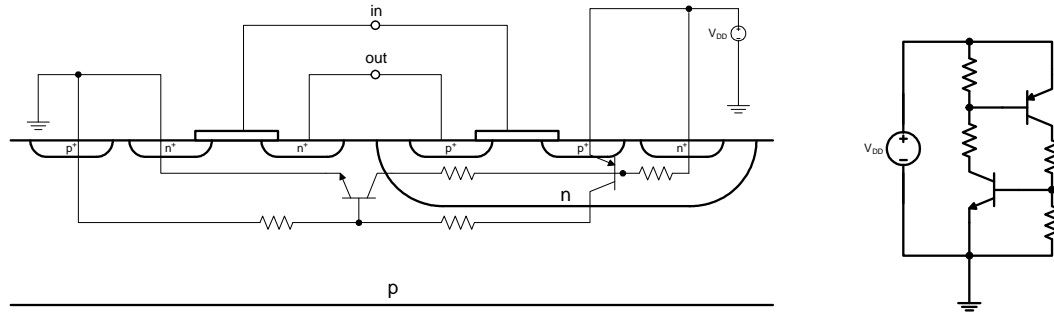


Figure 2.9: Cross-section of a CMOS inverter showing the parasitic thyristor (left) and its circuit (right).

sometimes referred to as *soft-error*.

In modern CMOS devices, information is usually stored as a quantity of charge, either in a single node or in a subcircuit. An ionizing particle crossing the drain (or resp. the source) depletion region of a device creates electron-hole pairs which are collected by the electric field. The collected charge modifies the voltage stored in the drain (resp. source) circuit node, possibly corrupting the stored information. In this case the node will be referred as the hit node or as the stroke node. It is important to point out that n-channel devices only collect electrons, therefore negative charge, while p-channel devices only collect holes, therefore positive charge. The charge

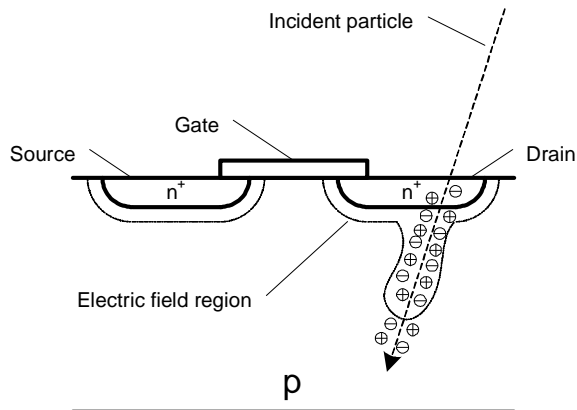


Figure 2.10: Particle strike on a drain node and funnelling.

deposition process also changes the shape of the electric field in such a way that the amount of collected charge is greater than the amount which would be collected in the equilibrium depletion region only: this phenomenon is called *funnelling* and the length of the region involved in the charge collection is called *funnel length* (see fig. 2.10).

### Critical charge

Obviously the deposited charge changes the value stored in the hit node only if it exceeds a particular threshold called *critical charge*, which depends on the circuit details and its ability to respond to a sudden current draw.

For instance, a high-impedance node has no active driver that can provide current to restore the correct voltage, thus it is very sensitive to SEUs. From the circuit

level point of view, dynamic logic, where information is stored in high-impedance nodes, is more sensitive to SEU than static logic, where information is instead stored in driven nodes.

Unlike many other radiation-induced effects, SEU sensitivity increases with the scaling down of VLSI technologies: in fact, the critical charge is proportional to the node capacitance and the supply voltage which both are being scaled down with feature size [Baumann 04]. Critical charge is a very convenient measure of circuit robustness to SEUs and can be obtained by simulation.

The amount of energy deposited by a charged particle per unit of length can be expressed in terms of *linear energy transfer* (LET) in units of  $\text{cm}^2\text{MeV}/\text{mg}$ , which is the energy loss  $dE/dx$  divided by the density of the traversed material (in our case Si, with  $\rho_{\text{Si}} = 2330 \text{ mg}/\text{cm}^3$ ). The LET depends on the atomic number of the material traversed and the energy of the incident particle.

*Critical LET*

Basically, higher atomic number ions have higher LET, while the dependence on the energy is more complex. Light ions generally do not have sufficiently high LET to induce direct SEUs but they can initiate nuclear reactions producing secondary isotopes with higher atomic number and therefore higher LET.

In this way, protons, neutrons and alpha-particles induce errors via nuclear reaction<sup>6</sup>.

LET relates to the deposited charge through the formula

$$Q_{\text{dep}} = \frac{q\rho_{\text{Si}}L_f\text{LET}}{E_{\text{eh}}},$$

where  $q = 1.602 \cdot 10^{-19} \text{ C}$  is the electron charge and  $L_f$  the funnel length which is usually chosen between 1 and 5  $\mu\text{m}$ .

The experimental characterization parameter for SEU threshold is critical LET, which describes how much charge has to be deposited to generate an upset. On the other hand, critical LET does not tell how much charge is collected by the circuit nodes, hence the difference between critical charge and critical LET.

### 2.3.3 Critical charge simulations

Under the approximation that charge collection processes and circuit response dynamics are independent, it is possible to study the two phenomena separately. This assumption is intrinsically valid for dynamic memory cells, where no active devices are involved, while it is basically flawed for static cells where charge-collection and circuit-response overlap in time. In fact, the time profile of charge collection by a node is a strong function of the voltage applied to the junction, which is in turn controlled by the circuit response to the event.

Nevertheless, it is still meaningful to study the circuit response separately to have a qualitative estimation of single-event vulnerability for a cell. This is usually done using a circuit simulator, for instance SPICE or one of its various commercial versions. Single-event interactions are modelled using a one-shot pulse current source with exponential time profile to inject charge into the node of interest. Figure 2.11 shows a possible simulation circuit for a standard SRAM cell, while Figure 2.12 illustrates a typical photocurrent.

---

<sup>6</sup>Alpha particles may induce SEE even without triggering any nuclear reaction, and lately this has been shown by IBM to occur even for protons.

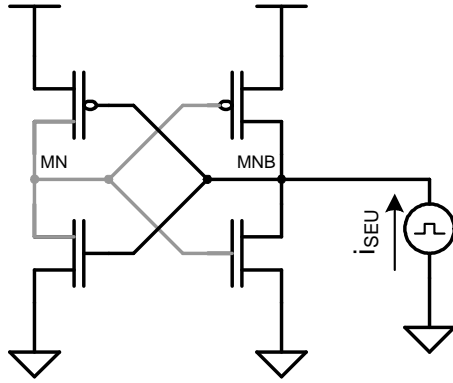


Figure 2.11: Charge collection simulation circuit for a standard SRAM cell.

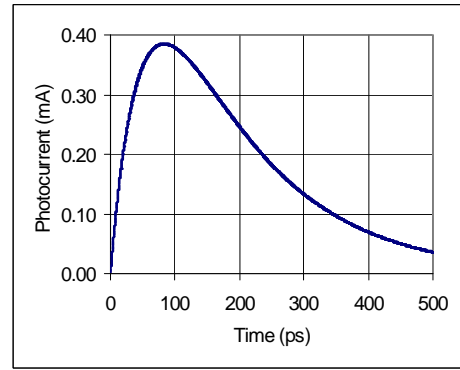


Figure 2.12: Typical exponential time profile photocurrent.

The injected current is also referred as photocurrent and its integral equals the collected charge. By sweeping the photocurrent amplitude it is possible to find the critical charge of the node. Often a triangular pulse is used in place of the exponential pulse, simplifying the simulation. During design time, the critical charge parameter is very useful to compare different architectures and different strategies to fight SEUs.

### 2.3.4 Critical LET measurement

Ion- and proton-beam experimental characterization are useful to determine the critical LET of a device, which can in turn be used to calculate the upset rate in the real application radiation environment. The typical procedure for these characterizations is to direct a monoenergetic (ion or proton) beam of known flux onto a device. The device can be in operation (clock running and I/Os switching) for a dynamic test, or idle (clock stopped and I/Os steady) for a static test. In a dynamic test the outputs of the device under test are checked continuously for errors during the irradiation, while in a static test the device is pre-loaded with a known pattern and checking of the outputs is done only at the end of the irradiation.

The beam is usually characterized in terms of the LET value it has as it enters the device, even though proton beams are an exception since they are characterized by the energy because of their peculiar indirect upset generation mechanism. The value obtained from the experiment is the number of the errors  $N$ . It is possible to compute the cross-section, which represents the sensitive area of the device, for a given LET by the formula

$$\sigma = \frac{N}{Ft \cos \theta} ,$$

where  $\theta$  is the beam angle with respect to the chip normal<sup>7</sup>,  $t$  is the exposure time and  $F$  is the beam flux. The product  $Ft$  gives the total fluence, usually expressed in ions/cm<sup>2</sup>. The cross-section is often referred as the “sensitive area” of the device and it’s usually expressed in cm<sup>2</sup>.

Using different ions and different energies it is possible to change the absorbed energy in order to plot a cross-section versus LET diagram (an example is given in fig. 2.13). Another way to vary the LET [Messenger 97] is modifying the incident angle of the beam: as the angle is increased, the amount of charge deposited in

<sup>7</sup>The validity of the *cos* relation to define the cross-section value is questioned, as many adjacent nodes may be struck by a single particle at grazing incidence, as will be seen later for this work.

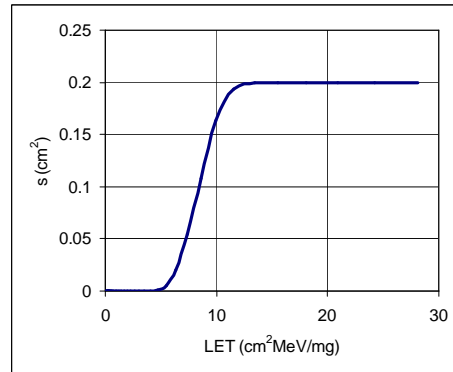


Figure 2.13: Example of a cross-section vs. LET curve (Weibull).

the vicinity of a sensitive device node is also increased since, the ion track in the sensitive volume becomes longer<sup>8</sup> by  $1/\cos\theta$ , assuming that the sensitive volume is small compared to the complete ion path, thus the effective linear energy transfer becomes  $LET_{\text{eff}} = LET/\cos\theta$ .

A cross-section vs. LET diagram can be fit with a Weibull curve [Messenger 97] and often shows a step corresponding to the critical LET, which can be defined rigorously as the LET value giving a cross-section of 10% of the maximum cross-section.

When protons are used instead of ions, a cross-section versus incoming beam energy plot is rather done.

In order to estimate the SEU rate of a device in the actual radiation environment [Faccio 04, Huhtinen 00, Huhtinen 97] it is necessary to know the dependency curve giving the probability per unit fluence to have within the sensitive volume an ionizing deposition greater than or equal to any given energy value. This curve describes the radiation environment and has to be integrated with the cross-section Weibull fit to obtain the error rate.

*Estimation of  
SEU rate*

### 2.3.5 SEUs in finite state machines and SEFIs

The logic present within an application specific integrated circuit (ASIC) can usually be divided in two classes: the datapath, which constitutes a pipelined structure in charge of computing the input data and bringing the results to the output, and the finite state machines which take care of controlling the datapath and dealing via specific protocols with the logic connected outside of the chip.

The severity of an SEU depends often on the out-of-service time caused by the strike, which depends on the kind of logic that has been hit. If the SEU occurs in a datapath, the error follows the data in its journey and it's brought quickly out of the chip. When an SEU occurs in the control state machines or configuration registers, these can enter in wrong states and start unexpected operation sequences which can last a long time, affecting the datapath and the whole system. In the worst case, a state machine can enter a cycle-loop from where it will never come out until a chip reset restores it in a safe state. This last type of radiation-induced failure is also referred as *single-event functional interrupt* (SEFI) and it has been observed for a

<sup>8</sup>See previous note.

long time [Koga 98] in memory devices where the read-write control logic gets upset and prevents from performing several correct I/O cycles. As it will be seen later, SEFs appear also in FPGA devices where the configuration is stored in memory cells.

### 2.3.6 Single-Event Transients

Recent experiments have demonstrated that sensitivity to single-event upsets in logic circuits increases with increasing circuit clock frequency [Buchner 97, Reed 96]. There is some evidence that at high frequencies the dynamic SEU rate is dominated by errors generated in combinatorial logic, composed by gates without memory, rather than in sequential logic, composed by memory devices like flip-flops and latches. Both these two kinds of gates are sensitive to SEU, but while soft error rate is independent of frequency for sequential logic, it increases linearly with frequency for the combinatorial part.

For older technologies where clock periods were far longer than any SEU generated current spike, this phenomenon was seldom taken into account in the design of an IC, while with current deep-submicron technologies errors generated in combinatorial logic start to be not negligible. An SEU occurring in combinatorial logic is also called *single-event transient* (SET), since in static logic the correct voltage value is immediately restored after the charge injection has terminated.

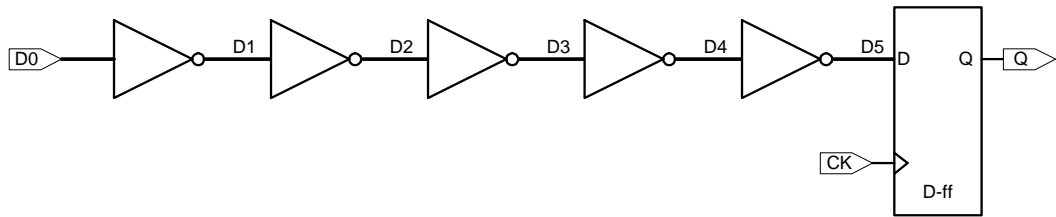


Figure 2.14: Inverter chain and flip-flop which illustrates nodes sensible to SETs.

Intuitively, the dependency of frequency of SETs follows from the fact that the charge collection time profile of a stroke node does not change with frequency, therefore, if the node is connected to the input of a flip-flop (like node D5 in fig. 2.14), the higher the frequency, the higher the number of clock sampling edges per unit time, the higher the probability of having a clock sampling edge just during the charge collection time interval, sampling the wrong value. This also means that experimental cross-section vs. LET curves vary with increasing frequency [Wang 04], where the effect of the combinatorial part gets amplified.

Each node in a combinatorial logic block has its own sensitivity time window. In other words a hit can cause an error only when this happens to be within a specific interval depending on the logic path to the next flip-flop. In fact the SET has to propagate through the logic to be presented at the input of the flip-flop and this takes some delay, thus the sensitivity window of a node is a propagation delay before the clock sampling edge. In Figure 2.14, node D0 has an early sensitivity window in respect to node D4.

SETs can be also masked by the logic function itself, since a variation in the value of one input does not always affect the output. This lowers the probability of a SET causing an error, but it's usually balanced by the fact that a single node often influences more than one output.



### 2.3.7 Multiple bit upset

With the continuous drive toward higher integration levels, the intercell spacing is also decreasing. A particle travelling with an incident angle close to  $90^\circ$  can in principle hit two or more drains of different devices in proximity to each other and therefore upset two or more nodes.

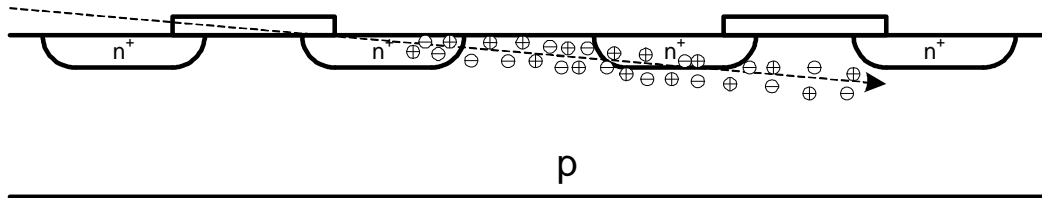


Figure 2.15: Multiple nodes charge collection.

As it will be seen later, this phenomenon can threaten circuits which rely on spatial redundancy to protect data. Charge deposition on multiple nodes is limited by the range of ionization paths, therefore nodes which are widely split apart are more unlikely to collect charge from the same traversing particle.

## 2.4 Protection from SEUs

As pointed out in the previous section, SEUs are a major concern for integrated circuits exposed to radiation environment, especially for modern deep submicron technology products. It is clear that for applications of modern technologies in environments such as the LHC it is necessary to protect the logic from upset and this can be done in several ways which can be divided again in process, circuit and system techniques.

Of course, modified manufacturing process techniques are only available at very high costs. An example of a modified foundry process is mentioned in [Roche 05]: since the critical charge of a sensitive node depends on its capacitance an obvious way for hardening is to increase that capacitance using a special manufacturing process to get efficient area utilization. A slight improvement can also be obtained by using high- $V_T$  transistors [Degalahal 04].

Circuit and system techniques are based essentially on data redundancy. If the data is stored somehow in several circuit nodes (or several bits), it is in some cases possible to reconstruct the correct data even from a fraction of those nodes (or bits). Therefore circuit techniques generally consist in storage cell configurations, different from the standard cross-coupled inverter cells, which prove to be robust to a hit on one node. On the other side, system techniques exploit error correction coding (ECC) methods that make use of special encoders and decoders around standard storage cells.

Due to its lower sensitivity to SEU, static logic, rather than dynamic logic, is used more often in applications where SEU robustness is required, as mentioned in Section 2.3.2.

The choice among the different possibilities for SEU hardening is therefore done taking into account, besides soft error rate, also area overhead, power dissipation and speed penalties.

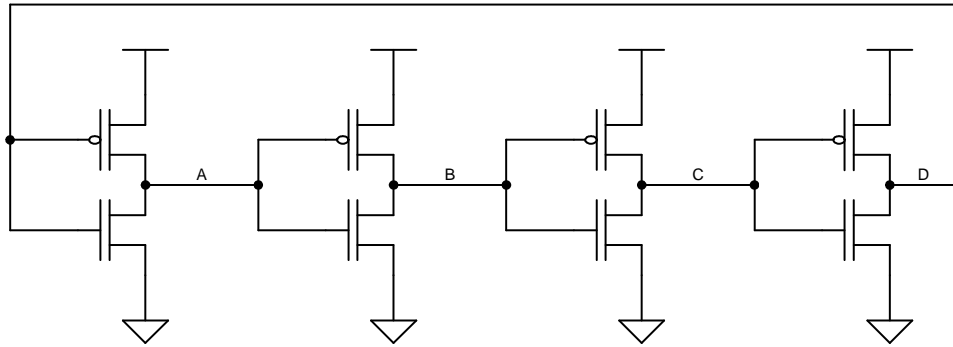


Figure 2.16: Doubled SRAM cell. Pass-gates are not shown.

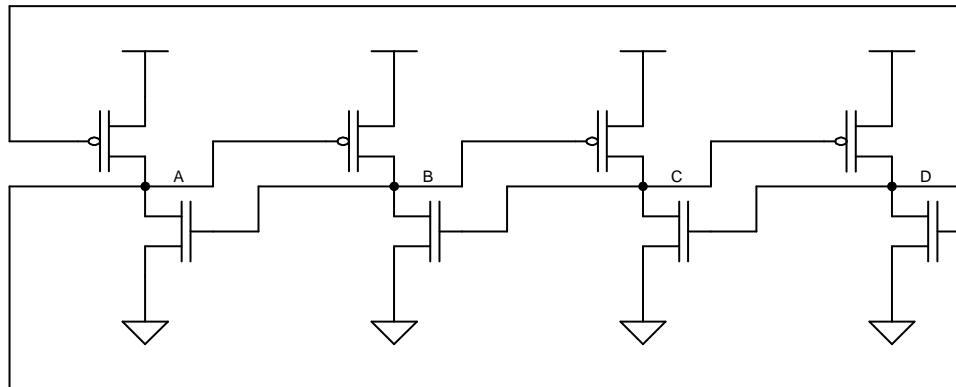


Figure 2.17: Dual Interlocked memory cell. Pass-gates are not shown.

#### 2.4.1 The Dual Interlocked cell

Standard SRAM cells store information in 2 nodes having opposite values. As first described in [Calin 96], to obtain some redundancy, it is possible to think of storing the information in a double number of nodes. The first idea could be to just use four inverters instead of the usual two cross-coupled inverters in the standard SRAM cell and connect them in a loop like in Figure 2.16. But this configuration doesn't give any advantage, since an error in one node can propagate through the loop to the whole cell.

A more powerful way to connect the transistors to each other and avoid error propagation is shown in Figure 2.17 which represents a *dual-interlocked cell* (DICE). This structure is fully symmetric and its memory nodes A, B, C and D are totally equivalent to each other. The cell allows data propagation from each node in two directions, one per logic level, since the gate of every p-channel transistor is connected to the memory node on its left, and the gate of every n-channel transistor is connected to the memory node on its right. In fact, low logic levels propagate towards the right direction, since only a low level can turn on the p-channel transistor, while high logic levels propagate towards the left direction since only a high level can turn on the n-channel transistor.

On each propagation stage the logic value is inverted. Clearly, no logic value can propagate for more than one stage in the same direction. For instance a low level on node B would propagate right through the p-channel transistor which can pull-up

	A	B	C	D	
t0	1	0	1	0	SEU
t1	1	1	1	0	
t2	-	1	1z	0z	
t3	1	0	1	0	

(a)

	A	B	C	D	
t0	0	1	0	1	SEU
t1	0	0	0	1	
t2	0z	0	-	1z	
t3	0	1	0	1	

(b)

Table 2.2: Sequences of states in the case of an SEU on node B. (a) starting with configuration (1,0,1,0), while (b) with configuration (0,1,0,1). High-impedance states are indicated with ‘z’, while contention states are indicated with ‘-’.

node C to a high level, but this doesn’t affect node D; conversely, a high level on node B would propagate left through the n-channel transistor which can pull-down node A, but again this doesn’t affect node D. It follows that an SEU on one of the cell’s memory nodes would affect only another node beside the hit one.

The cell nodes (A,B,C,D) have two stable logic configurations only which are (1,0,1,0) and (0,1,0,1), and each other configuration would settle to one of these two. Two examples of possible state sequences are depicted in table 2.2. In table 2.2(a), the cell is initially in the configuration (1,0,1,0), then a particle hits the p-channel transistor having the drain connected to node B. A high logic level is therefore forced in node B and this causes node A to lose its stored value and enter a contention state, since both its pull-up and pull-down transistor are active. Moreover, nodes C and D enter high-impedance state, since the pull-up from their p-channel transistors is now missing. Still, C and D won’t lose their value, thus maintaining the necessary information to restore the correct value throughout the cell. As soon as the upset ceases, node C tends to restore node B to the low logic state while node D does the same for node A. This marks the end of the state sequence after an upset and no data is lost. The sequence in table 2.2(b) is the converse of what described for table 2.2(a).

After a node has been hit, some time is necessary to restore the correct voltage throughout the cell, and this delay is called *recovery time*. Since a node inside the cell is in contention during the recovery, the cell draws more power than average in this interval, but this is usually negligible since upsets are seldom events. A glitch in the output can be observed during the recovery time.

It should be noted that if two nodes of the cell collect charge at the same time (from the same particle ionization track), the cell is likely to upset. For instance, when node B is hit, nodes C and D, which the cell relies on to restore the information, are more vulnerable because of their high-impedance state. Because of the symmetry of the cell this is a general rule: when a node is hit, there are always two other more vulnerable nodes which keep the saved information. Therefore it is good practice to allow some layout space among nodes of the same DICE [Velazco 96]. Measurement studies [Hazucha 04] over a DICE latch fabricated in a modern 90 nm process demonstrated 10× better reliability of the DICE latch with respect to a normal latch, which clearly shows the presence of charge collection by multiple nodes.

The DICE is therefore a 12-transistor cell which occupies 2× more area than a standard memory cell and burns almost twice more power. A latch can easily be built like in Figure 2.18 by adding clock transmission gates. Again a DICE latch has an area overhead of 100% with respect to a normal latch, and loads twice the clock lines. Since the area overhead is small compared to other SEU-hardening techniques, this cell is suitable for substituting latches of control logic state-machines.

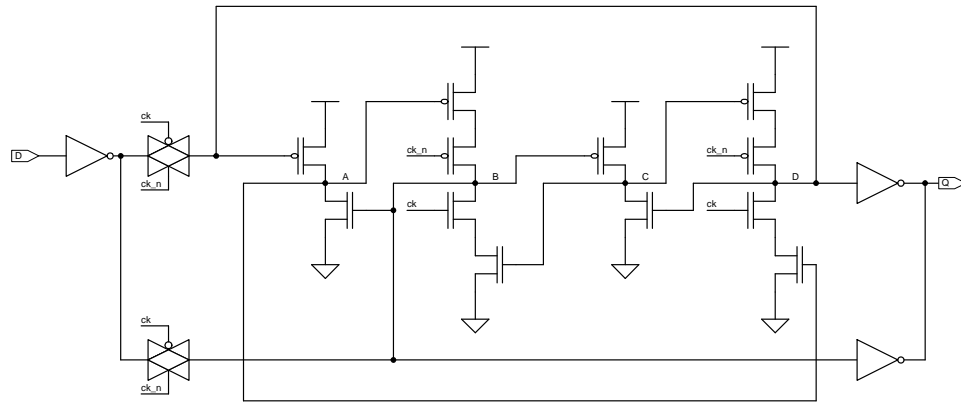


Figure 2.18: DICE latch. Local clock buffer is not shown.

Even though the DICE latch is considered to be the best solution for an SEU robust latch in terms of area overhead and power consumption, it still has speed limitations. In fact, a DICE latch can in principle be upset by a charge injection lasting more than half of the clock period and with the appropriate timing with respect to the clock. This happens because the latch is more vulnerable when open (in other words when it's transparent and the outputs replicates the input) and a hit on one of its nodes capable of injecting charge during the whole open time will prevent the cell to store the correct value. At low frequencies this is an extremely unlikely and negligible condition, since usual upsets last around 200 ps, but at frequencies above 1 GHz this upset mode is possible and becomes dominant at even higher frequencies. For this reason the DICE latch is not suitable for high-speed applications.

#### 2.4.2 The Whitaker cell

A possible cell configuration to obtain SEU robustness was developed by [Whitaker 91]. The cell is represented in Figure 2.19 and has 4 memory nodes (A,B,C,D) like the DICE. The cell is divided in two sections composed by a single type of transistor: on the left lies the p-channel section (or p-section), while on the right lies the n-channel section or (n-section). Hence, only low logic levels in the p-section memory nodes can be upset, since p-channel transistors can only collect holes, and conversely only high logic levels can be upset in the n-section, since n-channel devices only collect electrons.

The p-section exploits the property of p-channel transistors to be strong in pull-up but weak in pull-down. Conversely the n-section exploits the stronger pull-down of n-channel transistors. Proper sizing of the transistors enhances these strength ratios. The two sections are then connected together in such a way that the memory nodes of one section are connected to the weak transistors of the other section, creating a “weak” feedback loop which can restore the correct values after an SEU but cannot propagate errors. In fact, an upset on a node in the p-section will bring it from 0 to 1, but this can only turn off a strong transistor in the p-section, leaving a node in high impedance, and turn on a weak transistor in the n-section which will be balanced by its stronger counterpart. The same reasoning applies to an upset in the n-section.

It's important to note that the valid configurations are (1,0,1,0) and (0,1,1,0), but those are not the only stable ones: also (1,0,0,1) and (0,1,0,1) are possible, but

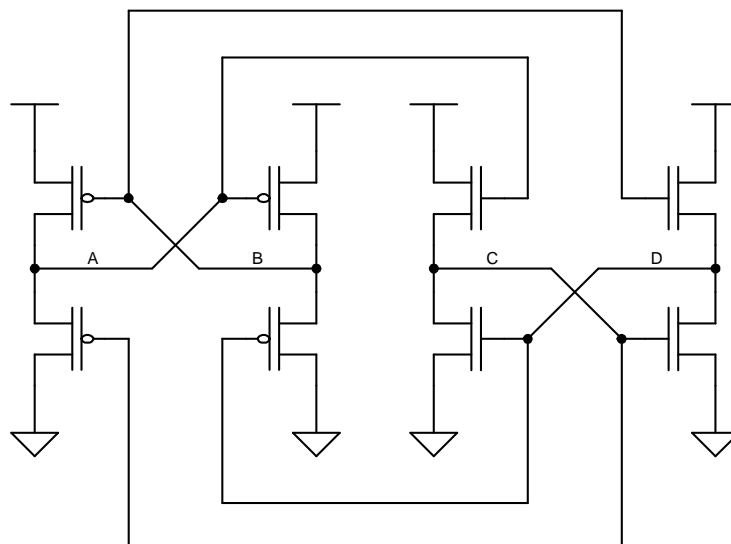


Figure 2.19: Whitaker memory cell. Pass-transistors are not shown.

not reachable by an upset.

The major drawback in the design is that, due to the degraded voltage levels in each section, significant static power is consumed by the cell and this limits the number of cells in a design. On top of that, scaling down the voltage becomes difficult. To solve these problems an improvement has been done by [Liu 92].

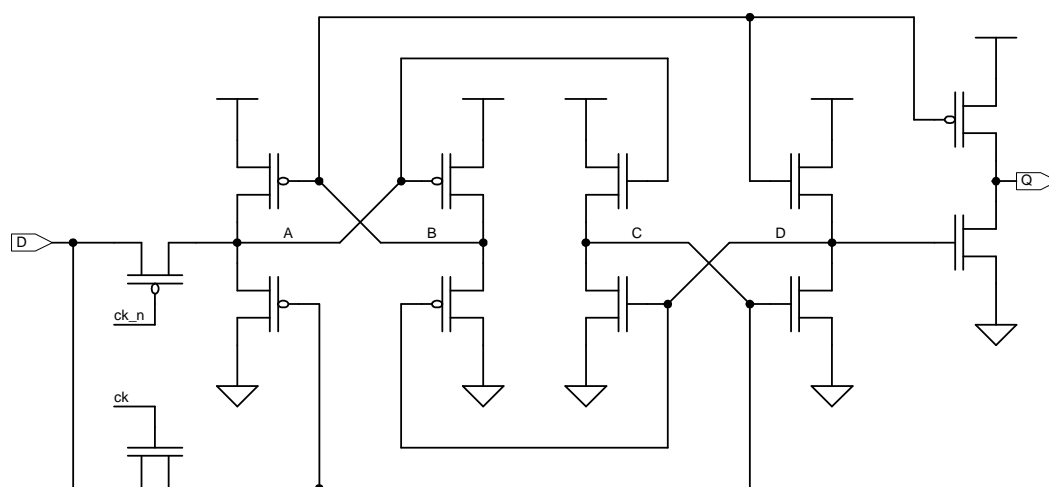


Figure 2.20: Whitaker SEU tolerant latch. Local clock buffer is not shown.

A latch can be made starting from the cell by adding an output buffer and two input pass-transistors like in Figure 2.20. Note that the pass-transistors have to be of the same type of the transistors in the section to preserve the unupsetability of the cell.

The cell is slower than a DICE cell due to the ratioed logic and the degraded logic levels. Besides the two cells are equivalent in terms of number of transistors and area. The cell still suffers from multiple node charge collection and increased vulnerability at frequencies above 1 GHz (see Section 2.4.1).

### 2.4.3 The SERT cell

A variation of the DICE cell is the so called *single-event resistant topology* (SERT) cell [Maki 01] shown in Figure 2.21 which has been used in [Gambles 03] for a space application. By adding a series n-channel transistor per memory node, any con-

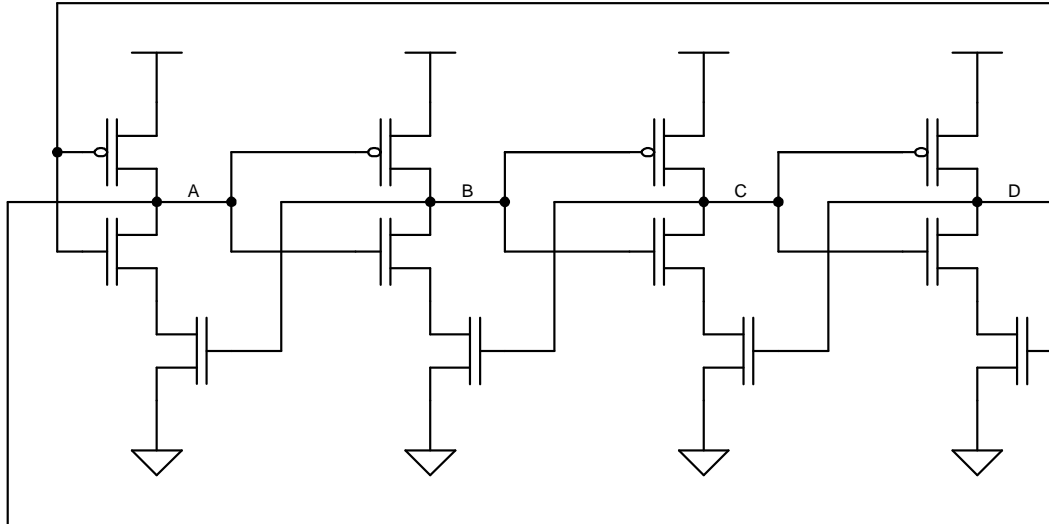


Figure 2.21: SERT memory cell. Pass-transistors are not shown.

tention state is avoided, therefore the cell decreases the recovery time and the power consumption during the recovery. The cost is an increased area.

### 2.4.4 Other SEU-hardened memory cells

Another cell which avoids conflicts in the nodes is the Dooley cell [Dooley 94], similar to the SERT cell, but with the drawback of having a bigger number of transistors.

The Rockett cell [Rockett 88] instead uses a complex ratioing of the transistors and again the upsettability of low logic levels only in p-channel transistors.

### 2.4.5 Temporal redundancy

The cells presented so far are hardened against particle hits on the cell area, but can still store wrong values if the input combinational block is upset generating SETs [Blum 05]. In order to protect the combinational part also some more effort has to be made. One common technique is to use *temporal redundancy* or, in other words, redundancy in the time domain.

Assuming that a signal coming from a combinational block is evaluated to its correct state within the propagation delay and then remains stable for some time, it is possible to sample the signal more than once in this stable time to get more than one copy of the value, hence redundancy. The drawback is that, in practice, this imposes a timing constraint on the signal which has to be stable during the sampling time, thus the operating frequency of the circuit is lowered. Since in many designs the operating frequency is not the most important constraint, it is often possible to use this hardening technique.

In order to sample a signal more than one time it is necessary to have a flip-flop capable of doing that. Fortunately, the upset-hardened cells seen before come to

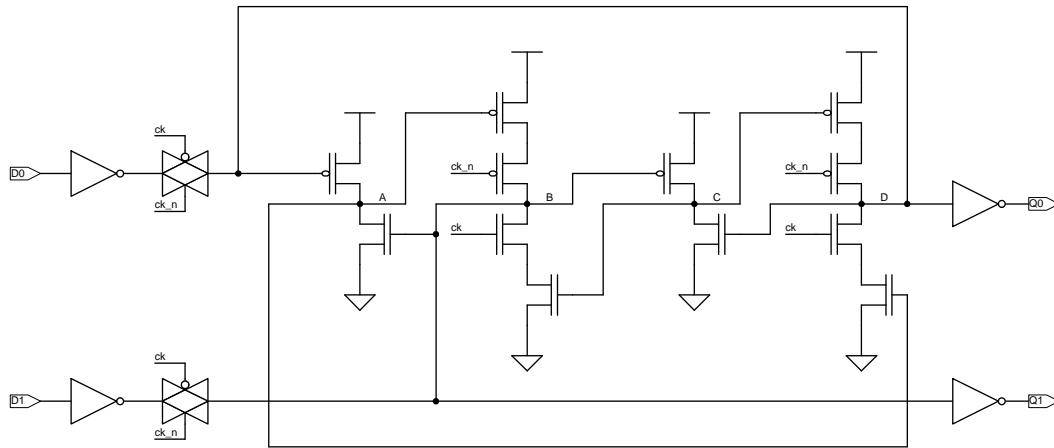


Figure 2.22: SEU-robust latch with split input/outputs based on DICE.

help: all of them can be used splitting the input/outputs, like shown in fig. 2.22 for a DICE-based SEU-robust latch. When the two inputs ( $D_0, D_1$ ) agree, thus they are (0,0) or (1,1), and the clock is low, a corresponding value is loaded into the cell. Instead, if the inputs disagree, the cell will behave like under an upset in the memory nodes, trying to restore the previously stored data.

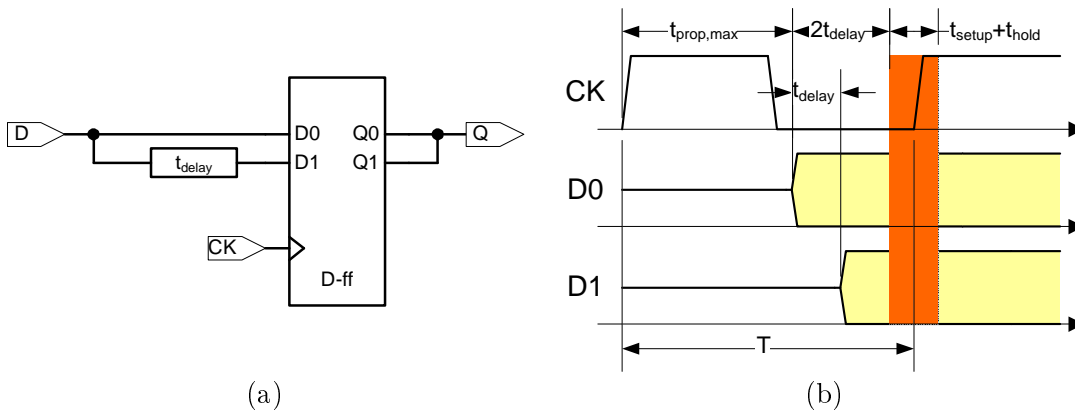


Figure 2.23: (a) Temporal redundancy technique schematic; (b) Temporal redundancy timing example.

Anyhow, since the latch must work with a single clock, the two inputs are sampled together, thus to obtain two samples of the same signal at different times it is necessary to delay one of the input signals with respect to the other. To accomplish this, a simple delay element  $t_{\text{delay}}$  can be added on one of the two inputs like in Figure 2.23(a).

In this way, the structure is immune to SETs lasting at most  $t_{\text{delay}}$ , since while one of the two inputs is upset, the other isn't. If a SET exceeds  $t_{\text{delay}}$  then the data will be corrupted.

By adding the delay element on one input of the flip-flop, its effective setup time is increased by the amount of the delay. Besides, if the cell has to be protected by SETs lasting  $t_{\text{delay}}$ , the effective setup time has to be augmented by another  $t_{\text{delay}}$ , which is then assumed to be equal to the maximum foreseen SET duration  $t_{\text{SET}}$ .

Thus the maximum propagation time of the input logic becomes

$$t_{\text{prop,max}} = T - t_{\text{setup}} - 2t_{\text{SET}} , \quad (2.1)$$

where  $t_{\text{setup}}$  is the intrinsic setup time of the flip-flop and  $T$  is the clock period. Figure 2.23(b) illustrates this timing constraint.

The importance of this timing constraint is limited since the upset-hardened memory cells shown before are not suitable for circuits having a clock period comparable with  $t_{\text{SEU}}$  (see Section 2.4.1).

In order not to have SETs longer than  $t_{\text{delay}}$ , it is possible to design the combinational logic with special techniques which attenuate glitches and upset-induced pulses [Baze 97, Mavis 00].

Temporal redundancy has also been implemented in [Wang 04] by modifying the DICE cell to have a delayed latching edge, in [Hass 98] using the Whitaker cell and in [Hass 03] using the SERT cell.

### 2.4.6 Triple Module Redundancy

Originally developed long ago by [Von Neumann 56] with the purpose of enhancing reliability of electronics in general, this concept was soon applied to microelectronics and ICs for the protection against ionizing particles. This technique is based on a basic block called *majority voter* which is a simple combinational part that has an odd number ( $2n+1$ ) of inputs and 1 output and always gives as output value the value present in at least  $(n+1)$  inputs out of  $(2n+1)$ , thus the majority. The smallest meaningful majority voter has 3 inputs. A majority voter circuit is depicted in Figure 2.24(b).

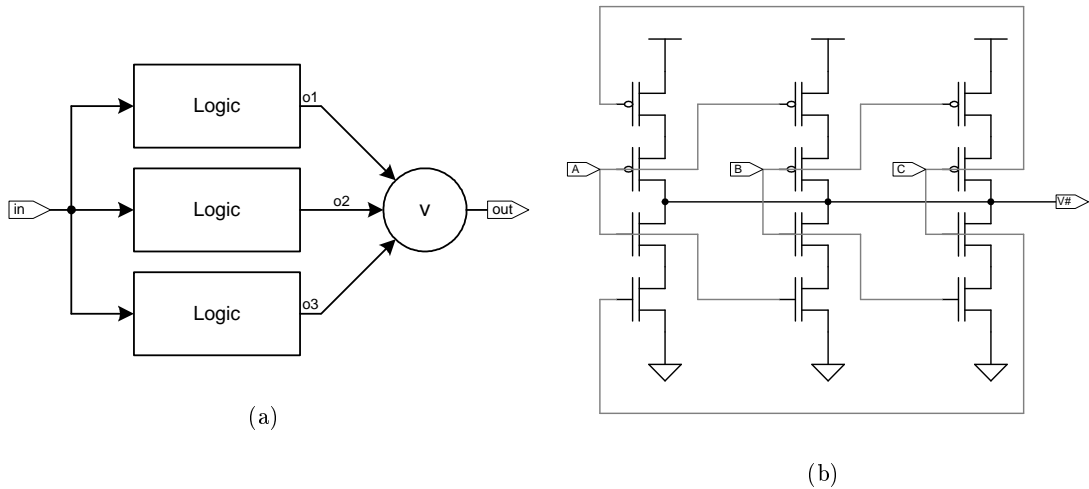


Figure 2.24: (a) Generic triple module redundant logic. The three logic blocks are identical. (b) 3-input majority voter circuit with negated output.

In Figure 2.24(a), three identical logic blocks receive the same inputs and are connected to a voter. Normally the three blocks should give the same output, but in case of fault or upset this can be false. It is clear that a fault or an upset in one out of the three blocks will be masked by the voter and will not be seen in the output. This is the basic principle of *triple module redundancy* (TMR) and has been used for a long time at the circuit board level, where is possible to have redundant chips with voted outputs [Hopkins 71].



Of course if two blocks are failing at the same time, the output will be also corrupted. Nevertheless, after the first failure, the majority voter block can have a diagnostic output telling which of the logic blocks is defective, information that can be used to reset the block or the system before a second failure.

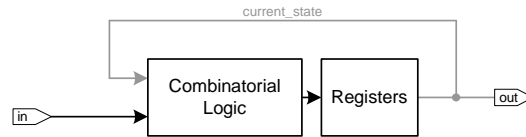


Figure 2.25: Traditional finite state machine.

The TMR technique can be easily applied to a state machine<sup>9</sup>. Figure 2.25 illustrates a traditional state machine, while Figure 2.26 represents a TMR state machine. Traditional, non-SEU-tolerant components are used. The registers keep the current state vector and the output vector in both the state machines. Even though, unlike a traditional state machine, in the TMR structure the current state is fed to the combinatorial logic through the voter, rather than directly from the registers.

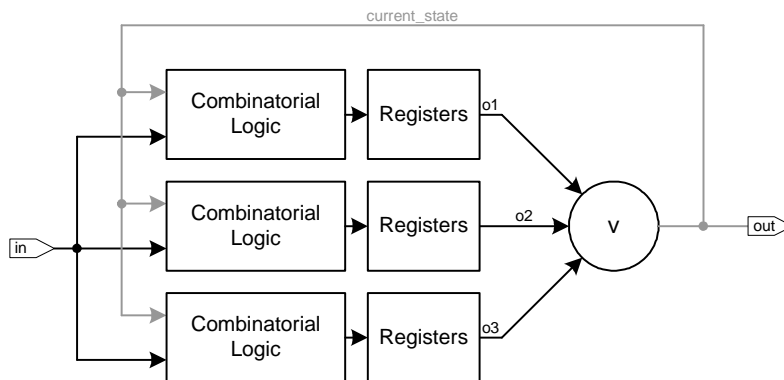


Figure 2.26: Simple scheme of TMR state machine.

It follows that if one of the register blocks loads an incorrect state, this will be masked out by the voter and all three state machines will get a correct input. In the next cycle all three register block will be thus in the right state. The voted feedback is therefore a mean of self-correction after an upset. A remaining weak spot remains the voter, which can be hit, generating a SET, and load an erroneous state into all three state machines. For this reason, with a small cost in area, it is better to use the structure represented in fig. 2.27(a).

In case state machines are cascaded or connected together, it is possible to use the style in fig. 2.27(b), which represents a full TMR state machine where the voter is also triplicated. In addition, the I/Os are triplicated in such a way that the connection with the neighbouring state machine (of the same kind) is protected from SEU: an error in one voter will propagate to only one register block in the following state machine and this will be recovered at the next cycle. Full TMR is the ultimate protection for the logic, since it protects the whole logic from SEUs and SETs.

<sup>9</sup>The more general case of Mealy state machines only will be analyzed, since Moore machines are a subset of the others. In Moore state machines the output depends only on the state vector. In Mealy machines the output depends on the state vector and on the input vector.

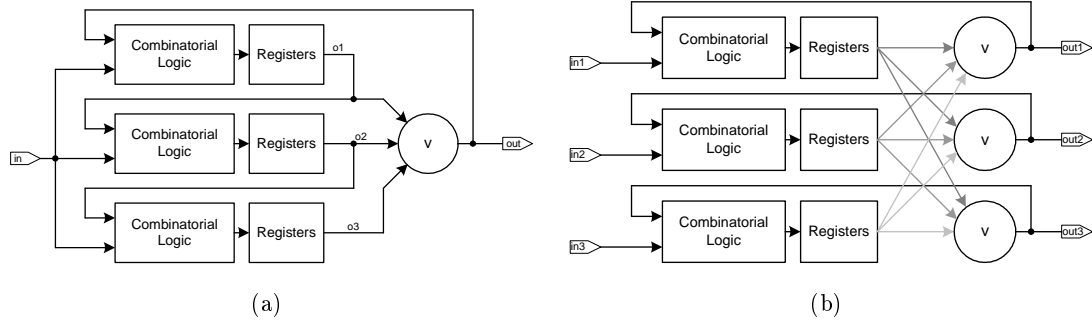


Figure 2.27: (a) TMR-FSM with state propagation; (b) Full TMR state machine.

Clock and reset networks can also be upset and the impact could be very dramatic since they usually feed many registers. On the other hand, the high capacitance present in every branch of the clock and reset networks avoids the generation of large glitches or transients. In case the latter argument doesn't apply, hardening has to be done and triplication might be a solution.

Clearly, TMR has an area overhead of 200%, with a proportional increase in power consumption and clock tree loading. Still, TMR has been used in many successful applications [Kloukinas 03, Bonacini 03] and has proven to be the most effective technique against SEUs. In fact, since this style is often used in combination with automatic place and route tools, multiple bit upset are very unlikely because of the spatial separation of the components.

TMR doesn't suffer of the same increased vulnerability at high frequencies seen for the DICE and neither it has speed limitations like temporal redundancy, thus it is suitable for those high-end applications which cannot employ these latter techniques.

A trade-off between protection and area overhead can be done by not triplicating the combinational logic, losing SET robustness. The latter can be then obtained by temporal redundancy as suggested in [Mavis 00].

#### 2.4.7 The TREVOTE cell

In principle it is possible to apply full triple module redundancy to a single register like in Figure 2.27(b) assuming the combinational part is just composed by an AND gate whose inputs are the enable and the data signals. A more efficient way to implement this circuit is instead to integrate the voters and the registers together. Let's consider the voter circuit shown in Figure 2.24(b), which has a negated output: each one of the three vertical branches in the circuit acts as a 2-way voting block. In fact, the output is left high-impedance by a single branch unless 2 inputs agree on a value. It is possible to build a memory cell modifying the circuit like in Figure 2.28, which exploits the voter's behaviour.

The cell was developed during the present work and it was named *triple-register-voter* (TREVOTE) cell. It has 3 memory nodes (A,B,C) with their respective negated counterpart and each one of the formers is evaluated by the voting of the other two. It follows that if a node disagrees from the others, these two are in high-impedance but retain their values and tend to store the same in the third node. A hit on the inverted nodes is equivalent to a hit on the non-inverted nodes.

Particular care during the design of this cell has to be taken against the charge-sharing mechanism which can degrade the voltage levels in the high-impedance nodes

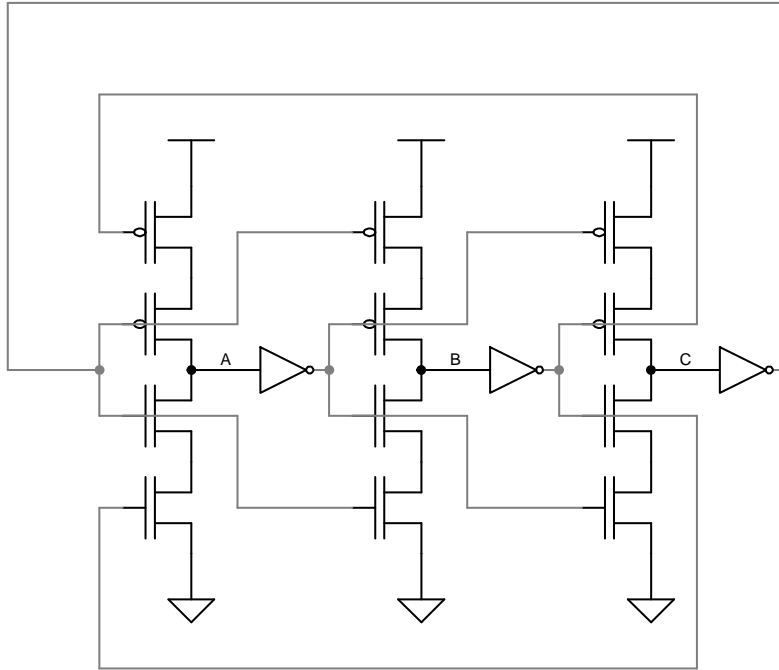


Figure 2.28: TREVOTE memory cell. Pass-transistors are not shown.

during an upset.

It is evident that the TREVOTE cell has more than  $3\times$  more transistors than a normal cell, thus is not suitable for SRAM designs but it can be used as a latch or flip-flop. A latch is illustrated in Figure 2.29.

The latch has again  $3\times$  the transistors needed for a traditional latch, but it has an embedded voting feature. This makes it convenient in case TMR is used (triplicating the combinational logic), since it practically saves the transistors used for the voters. On top of that, wherever TMR is not used, it is in principle possible to connect the three inputs together to the same combinational block and exploit only the cell's intrinsic SEU robustness. Temporal redundancy techniques are also suitable for this cell: two delay elements have to be used for two of the three input lines.

Charge collection by multiple nodes is still a threat for this cell. Moreover, when a node is upset, the other two are in high-impedance, thus more vulnerable. An effort has to be put during layout to space apart sensitive nodes.

### 2.4.8 Dual-rail logic

As seen before, temporal redundancy exploits the property of some storage cells (i.e. DICE, Whitaker and SERT) to be used with split input/outputs and extend their upset tolerance to the preceding gates. The same extension can be obtained by logic duplication. For instance, like in Figure 2.30, instead of connecting a single combinational logic block to both the inputs of an SEU-robust flip-flop, although delaying one, it is possible to place two identical combinational blocks to separately drive those inputs. The inputs of the combinational block will be in turn connected to separate outputs of an SEU-robust flip-flop and so forth.

The presented configuration creates a dual end-to-end path for the data and a SET will affect only one of the two inputs of an SEU-robust flip-flop, thus it will be

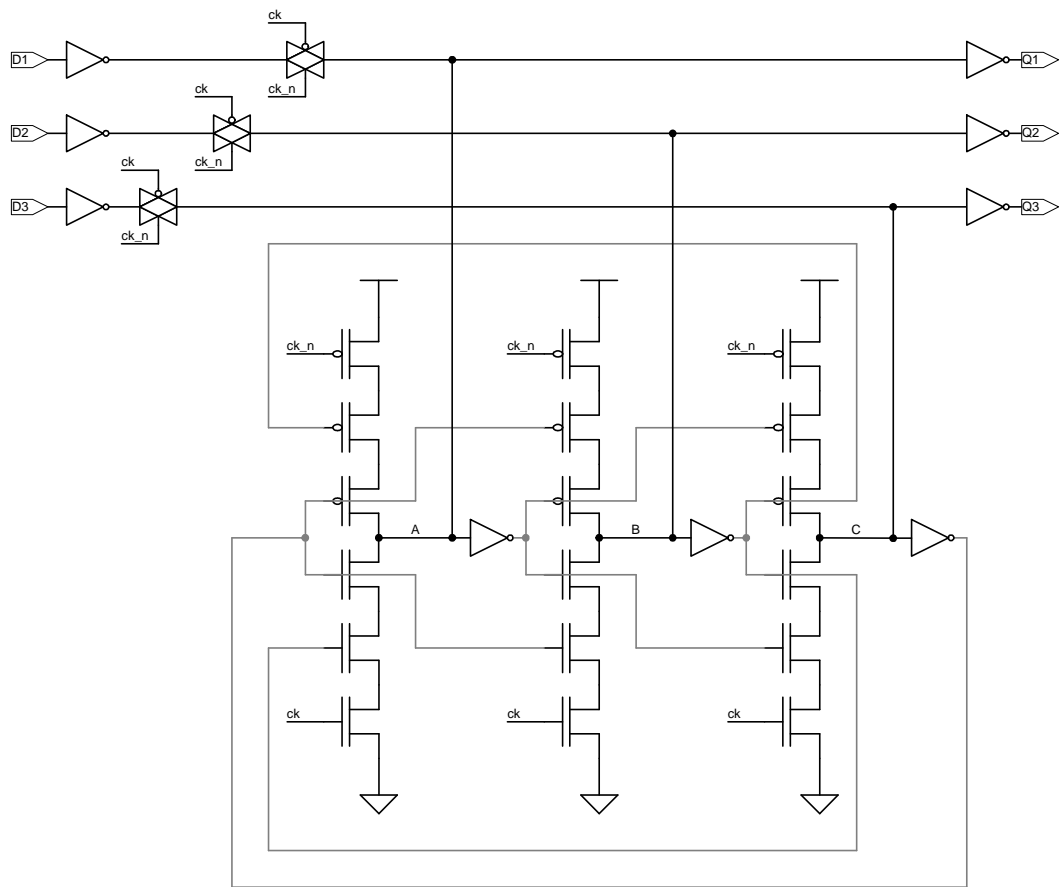


Figure 2.29: TREVOTE latch. Local clock buffer is not shown.

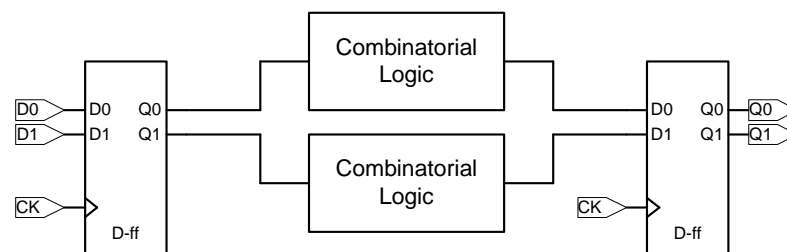


Figure 2.30: Dual-rail logic example.

filtered out. There is only one constraint which the design must respect. Clearly, an SET is harmful only if it arrives on the register's input during the sampling edge of the clock, thus only SETs with a duration interval that overlaps the setup time of the register shall be considered. The register is therefore able to withstand an SET lasting  $t_{\text{SET}}$  if the register inputs had settled already to their correct value before the beginning of the upset. Therefore the constraint is that the maximum propagation delay for a combinatorial logic block is

$$t_{\text{prop,max}} = T - t_{\text{setup}} - t_{\text{SET}} , \quad (2.2)$$

where  $t_{\text{SET}}$  is the expected time duration of an upset. Only SET lasting less than  $t_{\text{SET}}$  will be tolerated. Usually the time duration of an SET is in the order of magnitude of 100 ps.

Comparing the latter equation with (2.1) it is obvious that this time constraint is less tight than the one mentioned for temporal redundancy, it is in fact half, therefore dual-rail logic is suitable for critical paths which cannot respect the requests imposed by temporal redundancy. Nevertheless, the SET robustness of the cell is compromised when  $t_{\text{SET}}$  is comparable to half of the clock period (see Section 2.4.1). In these latter cases the use of TMR for SET protection is necessary.

The area overhead is bigger in dual-rail logic than in temporal redundancy, reaching 100% with respect to unprotected logic, with the same increase in power consumption.

Dual-logic has been successfully exploited in [Hass 03] with the use of the SERT cell.

### 2.4.9 Coding techniques

An efficient method to improve the reliability of data in memories and digital communications is error correction coding. In the case of SEU hardening, block codes which divide the information into blocks like ordinary memory devices are useful. Block coders [Clark 81] transform  $k$ -bit input words into  $n$ -bit codewords, by adding  $(n - k)$  parity bits to obtain some level of redundancy. One of the basic principles of coding is the *Hamming bound*, which, given  $t$  the number of errors the code can correct, states

$$2^{n-k} \geq \sum_{i=0}^t \binom{n}{i} , \quad (2.3)$$

therefore limiting the size of the codeword. The most efficient codes are those which respect the equality in the formula and are called *perfect codes*.

The most interesting perfect codes are Hamming codes [Hamming 50] which have  $t = 1$ , thus they have a single-error-correction (SEC) capability. Equation (2.3), becomes for this kind of codes:

$$2^{n-k} = 1 + n . \quad (2.4)$$

The Hamming codes are therefore limited to have  $(n, k)$  pairs that fit the latter equation, like (7, 4), (15, 11) or, in general,  $(n, n - \log_2(n + 1))$ .

Building an Hamming code is quite easy. Given  $(a_1, \dots, a_k)$  the input word and  $(c_1, \dots, c_n)$  the codeword, it is necessary to copy the input word over the codeword skipping the positions with index equal to a power of two, therefore

$$\forall i \in \mathbb{N}, i < \log_2(n + 1) : \forall q \in \mathbb{N}, 1 \leq q < 2^i : \quad c_{(2^i+q)} = a_{(2^i+q-i-1)} , \quad (2.5)$$

*Hamming  
encoding*

which, expanded becomes

$$\begin{aligned} c_3 &= a_1 \\ c_5 &= a_2 \\ c_6 &= a_3 \\ c_7 &= a_4 \\ c_9 &= a_5 \\ &\vdots \\ c_n &= a_n \end{aligned}$$

The remaining bits of the codeword, like  $c_1, c_2, c_4, c_8, \dots, c_{(n+1)/2}$ , are the parity bits obtained by the XOR operation of some of the other bits of the codeword. The XOR operation will be here indicated as a sum. The parity bits are defined by the set of equations

$$\forall i \in \mathbb{N}, i < \log_2(n+1) : \quad \sum_{p=0}^{\frac{n+1}{2^{i+1}}-1} \left( \sum_{q=0}^{2^i-1} c_{[2^i(2p+1)+q]} \right) = 0, \quad (2.6)$$

which means that, starting from its index, each parity bit alternatively checks a number of bits equal to its index and then skips another number of bits equal to its index. Expanding the latter equation it becomes:

$$\begin{aligned} c_1 &= c_3 + c_5 + c_7 + c_9 + c_{11} + c_{13} + c_{15} + \dots \\ c_2 &= c_3 + c_6 + c_7 + c_{10} + c_{11} + c_{14} + c_{15} + \dots \\ c_4 &= c_5 + c_6 + c_7 + c_{12} + c_{13} + c_{14} + c_{15} + \dots \\ c_8 &= c_9 + c_{10} + c_{11} + c_{12} + c_{13} + c_{14} + c_{15} + \dots \\ &\vdots \end{aligned}$$

The encoded data  $(c_1, \dots, c_n)$  can be stored in a  $n$ -bit wide memory device where one of the bits can get corrupted. In this case the retrieved data  $(r_1, \dots, r_n)$  will be different than  $(c_1, \dots, c_n)$ . In order to correct the error, it is necessary to recompute the parity bits for the retrieved word obtaining a third word  $(s_1, \dots, s_n)$  which differs from the previous only by the parity bits  $s_1, s_2, s_4, s_8, \dots, s_{(n+1)/2}$ . At this point, the location of the error is calculated by summing up the indexes of the parity bits that don't match. Therefore the erroneous bit is at index

$$\epsilon = \sum_{i=0}^{\log_2 \frac{n+1}{2}} 2^i (1 + s_{2^i} + r_{2^i}). \quad (2.7)$$

where the addition inside the parenthesis is a boolean XOR, while the rest of the operators are the ordinary natural number group operators. Although the last equation might look complicated, the intrinsic operation is quite efficiently implemented in logic by a simple row decoder. Knowing the index of the erroneous bit, it sufficient to flip that bit to obtain the corrected word. If more than one bit is corrupted, the logic will not correct the error and might not even recognize that there is an error.

Coding is usually implemented like in Figure 2.31, where data is first encoded, then stored in a memory and eventually decoded. The structure of the decoder is pointed out in the figure. The complexity of the encoder can be estimated easily<sup>10</sup>

<sup>10</sup>by equation (2.6): the inner summation is composed by  $2^i$  operands and then repeated  $(n+1)/2^{i+1}$  times by the outer summation, thus the total number of operands is  $(n+1)/2$ . Since one of the operands is the result and the operation is done  $\log_2(n+1)$  times, the number of 2-input XOR gates necessary to implement the function is  $[(n+1)/2 - 2] \log_2(n+1)$ . The decoder is composed by another encoder, a set of  $n$  2-way XORs and a row decoder. The row decoder is made out of  $[n - \log_2(n+1)]$  NORs having  $\log_2(n+1)$  inputs.

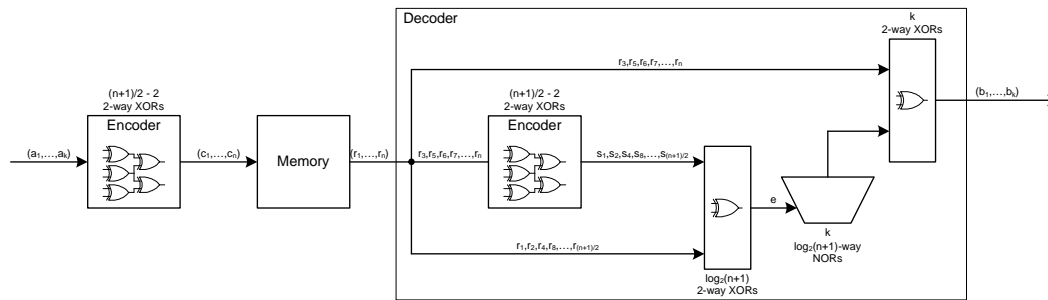


Figure 2.31: Hamming encoding/decoding scheme.

and it grows faster than linearly with respect to  $n$ , being  $O(n \log n)$ . In general, an attempt to reduce complexity in the storage of information competes with the increase in complexity caused by the introduction of decoding circuits.

Codes different from the perfect  $(n, n - \log_2(n+1))$  can be obtained by setting the unused bits to zero or one. The encoding and decoding logic simplifies consequently. For example, a  $(21, 16)$  code, useful for storing 16-bit data, can be obtained from the perfect  $(31, 27)$  by assuming  $a_{17} = \dots = a_{27} = 0$ .

For small values of  $k$ , the Hamming structure can be competitive with TMR to protect registers (but not the combinational logic); comparisons have been done in [Kumar 04, Niranjana 96, Larsen 72]. Although encoding area overhead might be smaller than TMR's, the latter offers stronger hardening: in TMR every bit is triplicated independently and a multiple-hit on bits of different triplets is masked, while in Hamming encoding a double-hit can only be detected, not corrected.

Hamming encoding is not competitive with respect to SEU-robust cells (like DICE) in the protection of registers, since the formers occupy less area and give better upset-immunity. Still, encoding can be exploited to protect state machines in an efficient way using codewords as state vectors, as done in [Meyer 71].

Error coding becomes advantageous when used to protect SRAM blocks. An SRAM block made out of traditional 6-transistors cells can exploit a Hamming encoder/decoder if a small number of parity bits are added. One single encoder/decoder block is necessary for an entire memory block, therefore the area utilization becomes very efficient. Thus, Hamming encoding is not suitable to protect locally registers and state machines, but it is very convenient to protect memories. Researches for an optimal code to maximize speed and minimize area were done in [Hsiao 70, Fuja 88].

To improve in error correction capability, interleaving is often used. For example in SRAMs, where the multiple-bit upset probability is not negligible, interleaving of columns can be used: odd bits are coded separately from even bits, in such a way that two adjacent bits can be upset together without consequences. This technique requires some extra area but is rather efficient in reducing soft error rates.

Error correction blocks based on Hamming coding are often found among IP cores of common logic synthesis tools for ASICs and FPGAs. Hamming coding has been successfully used in many applications, including [Klounkinas 03, Bonacini 03].

#### 2.4.10 High-capacitance signals

High-capacitance networks usually don't need any redundancy for upset hardening, since their capacitance makes them intrinsically hard. A particle hitting those nets would result in a voltage alteration which depends on the capacitance, on the deposited charge and on the strength of the driver. A target LET threshold has to be

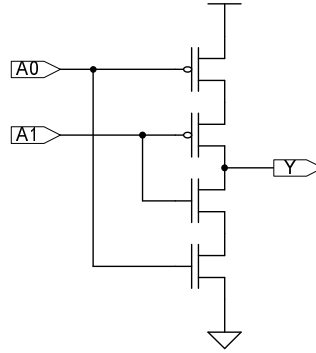


Figure 2.32: Interlocked buffer. When the inputs of the buffer disagree, the output is unmodified and only when the inputs agree the output is driven to the correct value.

chosen in order to estimate the capacitance which can be considered sufficient not to require any further protection.

A network can be defined to be high-capacitance if its fan-out is above or equal a certain number of gates, which corresponds to a known capacitance  $C_{th}$ . Given  $V_{DD}$  the supply voltage,  $E_{eh}$  the energy necessary to create one electron-hole pair,  $q$  the electron charge absolute value,  $\rho_{Si}$  the silicon density and the funnel length  $L_f$ , then the LET threshold corresponding to  $C_{th}$  is

$$LET_{th} = (C_{th}V_{DD}E_{eh})/(2q\rho_{Si}L_f)$$

in case there is no driver.

High fan-out lines like the clock and reset networks have in general high-capacitance, therefore their trees can run without other forms of SEU protection from the pad to the leaf cells. All the branches can be designed to have sufficient parasitic capacitance.

High-capacitance protection can be easily integrated together with other protection styles for low-capacitance signals like duplication or TMR. The distinction between high- and low-capacitance signals creates separate domains for redundant and not redundant logic. A high-fanout signal can easily drive the two respective inputs of a duplicated gate or the three inputs of a TMR gate, therefore the transition between a high-capacitance-protected domain to a TMR or duplication domain is trivial.

Conversely, the transition between a duplication or TMR domain to a high-capacitance-protected domain implies a buffering and requires some special considerations. Using only one copy of the (TMR or duplicated) redundant signal is dangerous since creates a weak spot, therefore all the redundant copies must be involved in the transition.

A simple solution is to buffer all the redundant copies and connect the outputs of the buffers together. This would cause a conflict in case of a propagating upset, therefore the voltage on the output line would have a transition to  $V_{DD}\frac{n-1}{n}$ , where  $n$  is the number of redundant copies (usually 2 or 3). A more elaborate solution is to use a voter in the case of TMR and an interlocked buffer like the one in Fig. 2.32 in the case duplication. When the inputs of the latter disagree, the output is unmodified and only when the inputs agree the output is driven to the correct value [Shuler 05].

Input and output pad buffers can be considered to be borders between a non-redundant domain, the pad and the exterior of the chip, and a redundant domain



inside the chip. The same considerations mentioned before apply to the input/output pad buffers.



## Chapter 3

# Programmable logic and radiation environment

### 3.1 Brief history of programmable logic

The desire to have programmable hardware has been in existence ever since the very beginning of digital electronics, when fast-prototyping was the main goal for programmable devices.

#### 3.1.1 PROM devices

Any  $n$ -input logic function can be realized in the form of sum of  $2^n$  minterms, which are the product combinations of the  $n$  input signals in either their positive or negated form. This is the idea at the base of PROM devices, which can consist in a row decoder connected to a series of OR gates through a matrix of programmable switches like in Figure 3.1. The row decoder is in fact made of  $2^n$   $n$ -way AND gates, which generate the word line signals. These are then driving the inputs of the OR gates or not, depending on the program. The structure resembles therefore the AND-OR general function described before.

The AND/ORs are wired gates, which means that they have a load as pull-up while the pull-down transistor network can be spread all along the input lines. Programming is usually done burning *fuses* which disable the pull-down of the transistors in the OR matrix. The AND matrix is instead fixed, since the nature of PROMs is usually to store data rather than implement logic functions.

More generality can be offered by *programmable logic array* (PLA) devices, which have both matrices with full configurability. An example of PLA is shown in Figure 3.2(a). On the other hand, often the logic functions realized don't need many minterms to be summed, thus a more efficient way to create general-purpose functions is to have a configurable AND matrix and a fixed limited OR matrix, like in Figure 3.2(b). The latter structure is referred as PAL.

*PLA and PAL devices*

#### 3.1.2 PLDs

One refinement to the PALs was the addition of a programmable register at the outputs and feeding back the stored values into the AND matrix (see Figure 3.3(a)): these device are called *programmable logic devices* (PLD). This change made possible

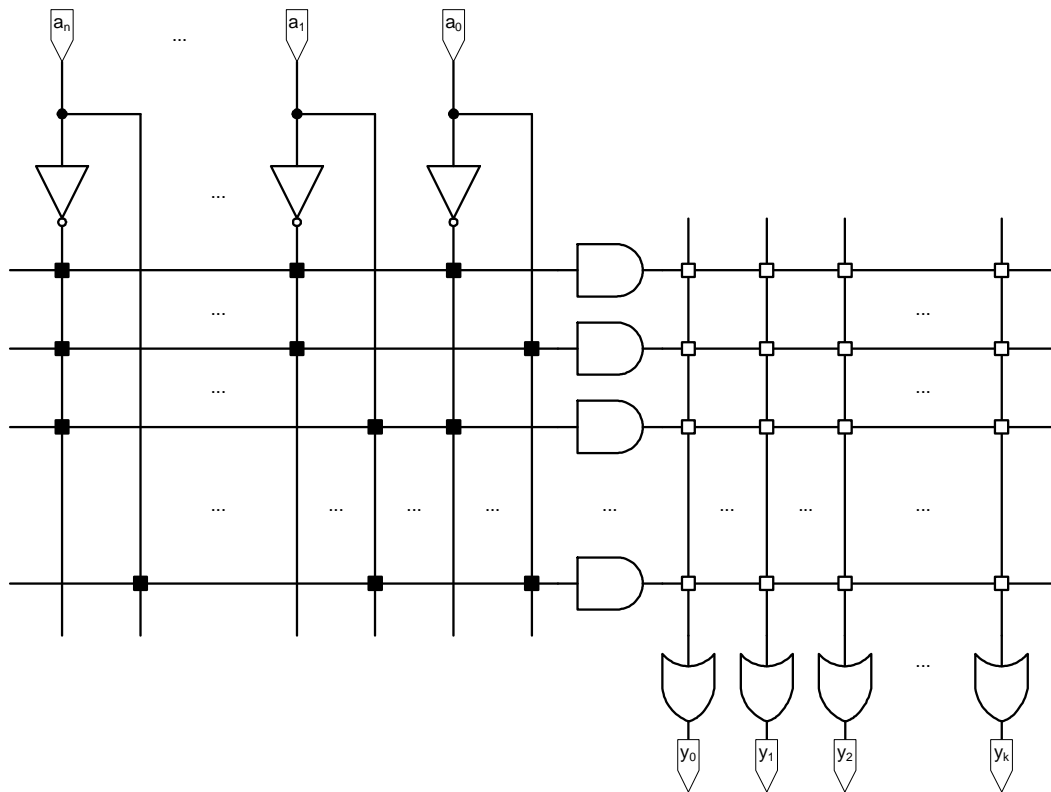


Figure 3.1: PROM structure made out of wired AND/ORs. The squares indicate where a connection to the wired gates is possible; squares are filled where the connection is made.

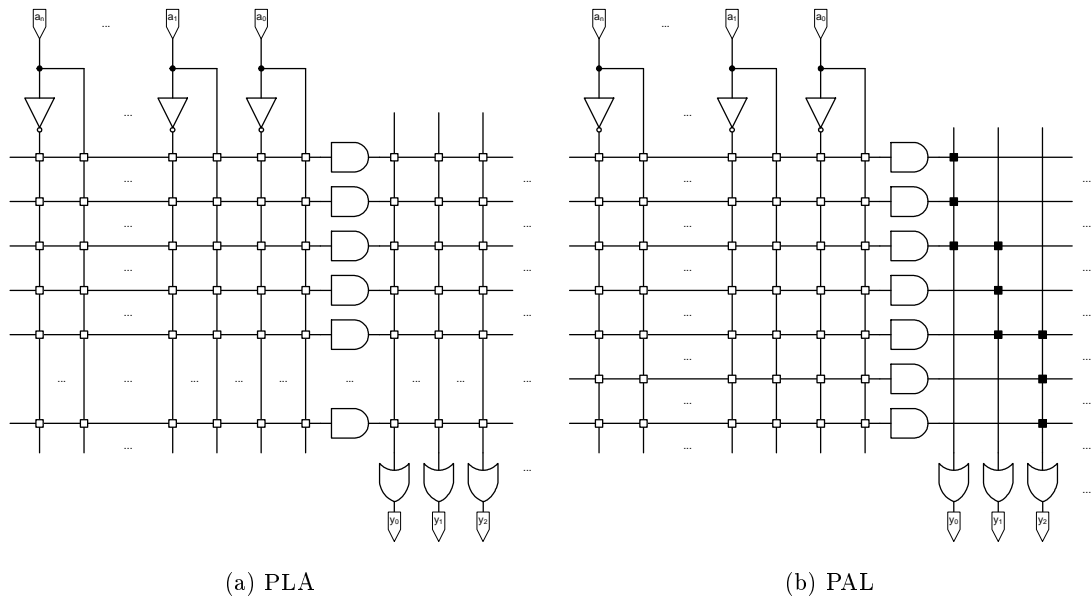


Figure 3.2: Simple programmable logic examples

the realization of state machines and sequential logic, making the devices of this kind

spread commercially very quickly<sup>1</sup>.

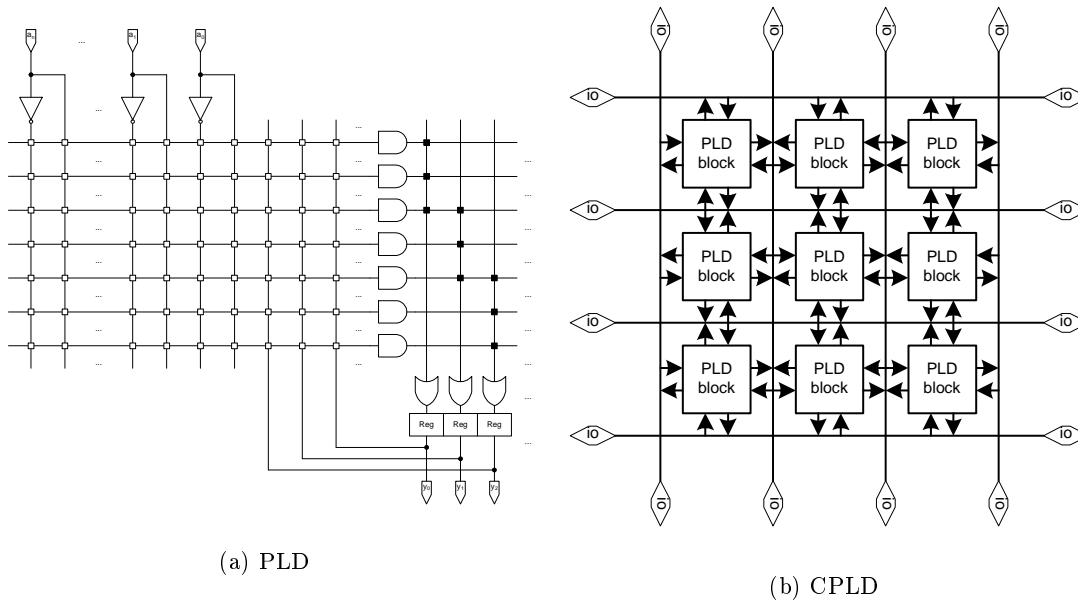


Figure 3.3: Evolved programmable logic devices

### 3.1.3 CPLDs

Since PLDs are typically small devices, often a design contains more than one of them. A solution to integrate all the functionality in one chip could be to increase the size of PLDs. On the other hand the number of inputs in the AND matrix cannot be increased indefinitely since the large fan-in makes it inefficient. An alternative is to put more than one PLD on the same die and connect them together with some programmable routing resources: these devices are named *complex* PLDs (or CPLDs) and an example is represented in Figure 3.3(b). The more PLD blocks a CPLD contains, the more complex functions it is possible to implement.

Programming of PALs, PLAs, PLDs and CPLDs is usually done either by burning fuses or with some non-volatile memory element like EPROM, EEPROM or Flash devices. Clearly, logic programmed through fuses is *one time programmable* (OTP), while the other mentioned devices are in some way rewritable.

### 3.1.4 MPGAs

A mechanism to shape the logic is the one used in *mask-programmed gate arrays* (MPGAs) which are composed by a simple gate, like a 2-way NAND, replicated in a large number of copies (also called “sea of gates”). To implement a design, these silicon gates have to be connected with metal traces, thus the programming is done by shaping those traces, or, in other words, defining the layout mask for the metal. This kind of programming can be done in the late stages of production of the integrated circuit, thus the lead time is short compared to a truly custom ASIC, even though the design is not independent of the fabrication and the masks can be expensive. During the evolution of MPGAs it soon turned out that a more complex basic logic block, like a 4:1 MUX, could utilize the area in a more efficient way.

<sup>1</sup>The 22V10 is the most famous example.

## 3.2 Field-programmable gate arrays

The architecture of *field-programmable gate arrays* (FPGA) is similar to that of a MPGA [Rose 93], but the major difference is that FPGAs are programmed via switches or fuses much the same as traditional CPLDs or even with SRAM elements. Besides, in a CPLD logic is implemented using predominantly two-level AND-OR logic with wide input AND gates and inefficient crossbar-like structures, while in an FPGA logic is realized using multiple levels of lower fan-in gates.

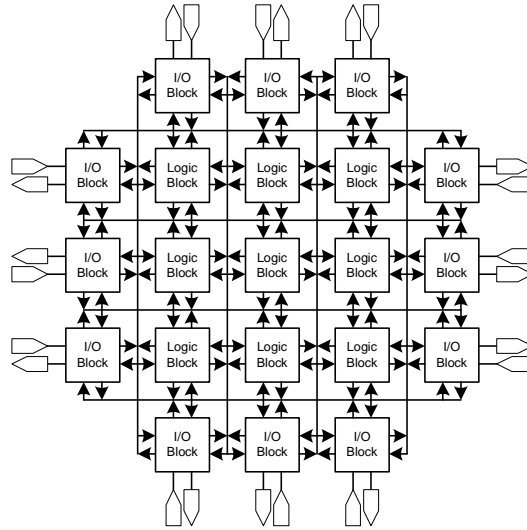


Figure 3.4: Field-programmable gate array.

FPGAs can be visualized as programmable logic embedded in programmable interconnect (see Figure 3.4) [Admrel 02]: an array of programmable logic blocks is surrounded by a mesh of configurable routing. The outer blocks take care of the interface with the external world. One important distinction has to be done between configuration logic and user logic: *configuration logic* is formed by the infrastructure which write, reads and stores the program in the FPGA including all the programmed switches; *user logic* is composed by all the rest of the circuits which, once the FPGA is programmed, are connected together to create the desired system, including registers and gates. Indeed the user logic is composed mainly of the content of logic blocks.

The main differences among FPGA architectures are then found in the former three constituents and in the programming and configuration storage technique.

### 3.2.1 Logic block architecture

Various possibilities exist for the design of the basic logic blocks and the main decision parameter is the implementation capability. This can span from a simple inverter [Marple 92] to a complex logic with registers. Logic blocks are then distinguished by their *granularity*, which can be defined as the number of equivalent gates (2-way NANDs).

The main advantage of using fine grain logic blocks is that the block utilization is optimized, since it is easy to fully use simple gates and logic synthesis techniques are elementary. On the other hand, fine grain logic requires more routing resources which are costly in delay and area.

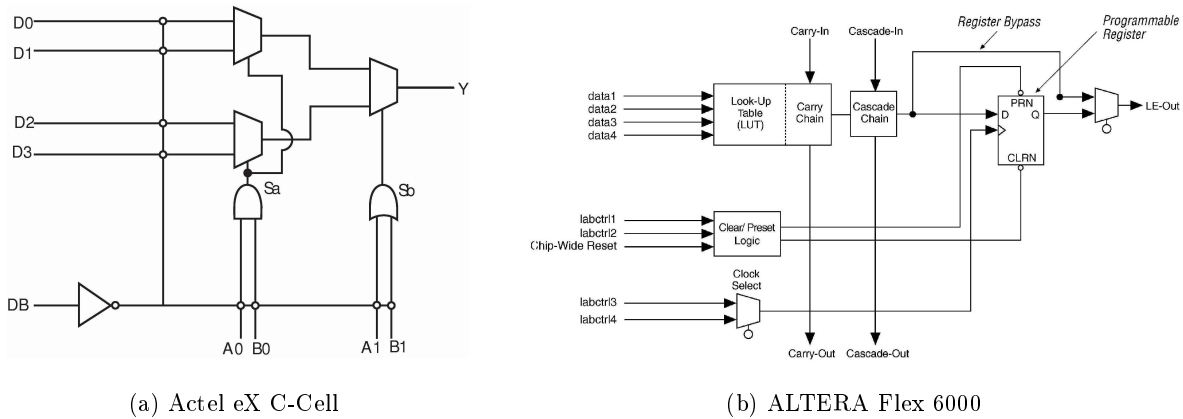


Figure 3.5: Logic block structure examples

For instance, Actel's eX family logic has two possible modules and one of them is represented in Figure 3.5(a): it is simply composed by a few MUXes and gates, being able to implement a range of about 4000 functions. Even though, the cell has 9 inputs and 1 output, forcing a complex routing for a small piece of logic.

Another example is Altera's Flex 6000 basic cell, represented in Figure 3.5(b). It contains a 4-input *look-up table* (LUT), which acts as a function generator that can implement any boolean expression of four variables. Besides, the Flex 6000 cell has a built-in carry chain for addition/subtraction, a cascade chain for the generation wide fan-in logic and, last but not least, a register. This cell has only 4 general purpose inputs and it has a considerable implementation capability, therefore its granularity is coarser than Actel's eX cell, but at the same time it requires less routing effort.

Look-up tables are often used in FPGAs since they offer great versatility. A  $n$ -input LUT is basically composed by a  $2^n \times 1$  memory representing the truth table of the desired boolean function. The address lines of the memory can be driven by the  $n$  input signals while the single-bit output provides the boolean function. A  $n$ -input LUT implements any expression of  $n$  variables, therefore it can realize  $2^{2^n}$  functions. On the other hand, LUTs become unacceptably large for more than 6 inputs and the expressions they can build are not often used in practical designs and difficult to exploit by synthesis tools.

Investigations on the best granularity to obtain best area utilization [Rose 93] showed that, in case LUTs are used, a 4-input LUT is the optimum.

### 3.2.2 Routing architecture

The routing architecture of an FPGA is the manner in which the programmable switches and wiring segments are positioned to allow interconnection of the logic blocks. This is usually a trade off between flexibility and density, since the more interconnections are possible in an FPGA, the more flexible it is, but also the more area is wasted for routing and configuration bits.

Many different possibilities exist to realize the routing, the earliest is the *sea-of-gates* structure represented in Figure 3.6(a), where each logic block is connected only to its neighbours. Direct neighbor-to-neighbor routes are fast and the logic blocks cover virtually the entire floorplan, without area wasted for routing. Actel's SX family is an example of this class of architectures. To realize longer connections,

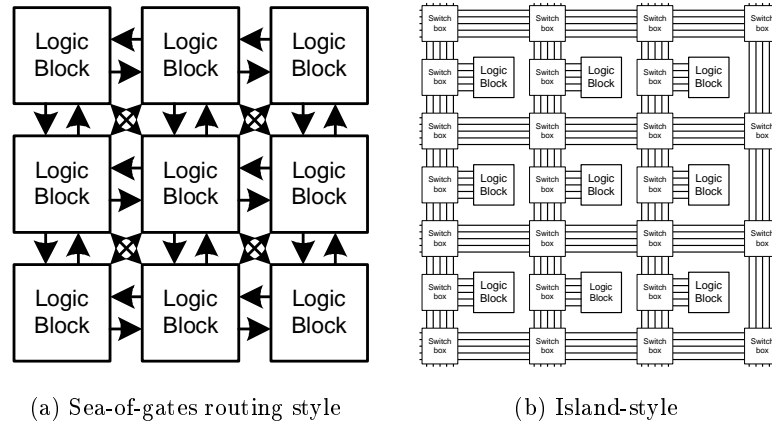


Figure 3.6: Routing structure examples

signals have to go through logic blocks, which means wasting logic resources and slowing down the propagation.

Evolved routing architectures include a more hierarchical view of the interconnections, guaranteeing long distance wires which skip one or more logic blocks. These wire segments usually end-up in switch-boxes, where vertical and horizontal wires intersect and can be programmably connected together, or where the wires coming out of a logic block can enter the higher-level routing. This architecture is known as *island style*, depicted in Figure 3.6(b), and it is the most used in commercial devices.

Switch-boxes are arrays of programmable pass-transistors or tristate buffers which can connect two wires. Often not every wire is connectible with every other wire, in order to save some configuration bits. Each switch slows down a signal by some amount, thus usually some unbroken lines exist which have the goal of bringing a signal bigger distances away without much delay.

The island-style architecture presented is symmetric since it has the same amount of vertical segments as horizontal segments. Some other architectures break this symmetry and give different amount of lines for different directions, like for instance the *row-based* architecture, used in some Actel devices, which has more row wires than column wires. At the limit this technique can result in a *one-dimensional* architecture, although seldom used.

Special tree-networks are often available in FPGAs for distribution of clock, reset and other time-critical signals.

### 3.2.3 I/O blocks

I/O blocks usually connect to the routing in the same way logic blocks do. To allow great flexibility, usually I/Os can be programmed to support many different signalling standards, voltage levels, drive strength, slew rate, open-drain or not, pull-up or pull-down if necessary and, last but not least, differential signalling. I/O blocks often contain registers and can perform *double-data-rate* (DDR) transeiving.

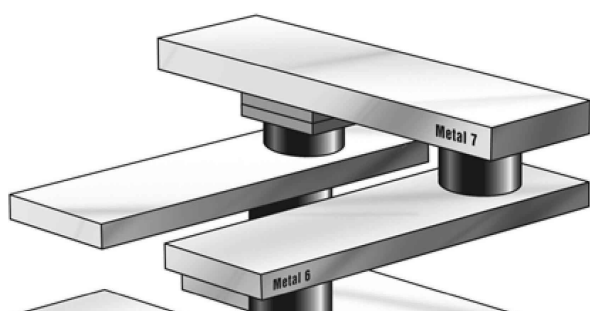
### 3.2.4 Programming technique

Important considerations have to be done on the storage technology used for configuration bits. As mentioned before the possibilities span among one-time programmable devices, non-volatile memory devices and volatile devices.

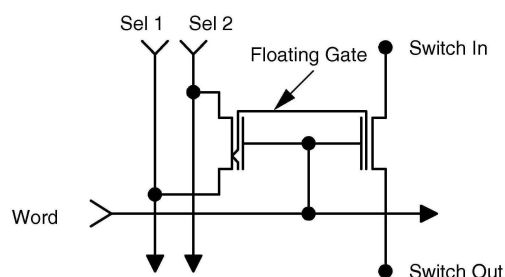


Antifuses are the most common OTP device in use. They are basically a two terminal device with an unprogrammed state presenting a very high resistance ( $\approx 1 \text{ G}\Omega$ ) between the terminals. When a high voltage (in respect to the normal operating supply voltage) is applied across the terminals the antifuse will enter breakdown (or “blow”) and create a permanent low-resistance link ( $\approx 50 \text{ }\Omega$ ). Antifuses are commonly made by a layer of dielectric<sup>2</sup> interposed between two metal layers and this device has about the size of a via, saving a big amount of area compared to other memory cells (see Figure 3.7(a)). Nevertheless, extra circuitry is necessary to program the antifuse with high voltages and programming transistors must be able to handle high currents, thus the area savings are mitigated.

The resistance of a programmed antifuse is lower compared to the resistance of a typical on-state pass-transistor used as a switch in case other memory devices are used. Moreover, antifuse don't require any power supply or any external configuration storage during power down. No reprogrammability is given using antifuses.



(a) Antifuse switch used in Actel's RTAX-S family.



(b) Flash switch used in Actel's ProASIC family.

Figure 3.7: Programmable switches in commercial devices.

Floating-gate devices like EPROM, EEPROM or Flash are non-volatile memories. Floating-gate transistors have two gates, an upper select gate and a lower floating gate. The floating gate is insulated from any other node and, exploiting physical processes that will not be treated here, it is possible to inject electrons into the floating gate, thus negative charge. A negative voltage into the floating gate prevents the transistor to be turned on even by applying a high logic level to the select gate, thus the transistor is programmed to be off. The charge accumulated in the floating gate can be removed by a process called erasing. Programming and erasing require high voltage, therefore normal voltage operation does not affect the floating gate stored charge.

This kind of transistor can be used directly to form a switch or a wired-AND structure like in PLDs, but the terminals of the transistor have to be used for programming also, which requires high voltages, therefore some isolation structures might be necessary. Another possibility is coupling together two floating gate transistors to have common floating gate and common select gate like in Figure 3.7(b): one transistor is used for programming, while the other is used in-logic as a switch.

<sup>2</sup>The dielectric used is often amorphous silicon or silicon oxide-nitride-oxide (ONO).

Floating gate devices are reprogrammable, giving a lot of versatility since in case an error is done during design, the program can be corrected in-circuit. Moreover, these devices are non-volatile therefore no power supply or external storage is needed to preserve the configuration. Some circuitry for high voltage generation, programming and erasing is necessary.

#### *Volatile memories*

Volatile memories are SRAM cells or flip-flops. SRAM cells can be organized in an array, while registers in a long shift-register chain. The output of a register or the content of a memory cell can be used to drive a transistor employed as a switch. The peculiarity of using registers to store the configuration is that this allows a shared user/configuration use: the same registers that contain configuration information in one design can in principle be used as user registers (thus design resources) in another design. On the other hand, registers are bigger than SRAM cells which are in turn bigger than non-volatile cells. In fact, static memory configuration storage usually dominates the area utilization on an FPGA.

Since SRAM is volatile, the configuration must be reloaded after each power down, therefore an external configuration storage is mandatory. Static memory is reprogrammable and programming is much faster than in floating gate devices. On top of that, programming does not require any high voltages and production does not need any special steps.

#### **3.2.5 Special-purpose blocks**

In many FPGAs it is possible to find some special-purpose blocks which range from simple memories to microprocessors. These blocks are then connected with the routing structure of the FPGA like other logic blocks. Usual special-purpose blocks are DLLs, SRAM blocks and multipliers.

### **3.3 FPGAs in radiation environment**

Some special considerations have to be done for radiation effects in FPGAs. First of all, commercial FPGAs span different process techniques which are very differently influenced by radiation. Besides, the internal structure of the FPGAs presents configuration information and user information which have different importance for the system behaviour.

#### *Antifuses*

Antifuse FPGAs have proven to be resistant up to 300 krad total dose [Actel 04] but they fail somewhere above this threshold because of the internal charge-pump used to generate the high voltages needed for programming [Wang 03a]. In fact, even though programming does not need to be done in-circuit, the charge-pump drives the insulation transistors present in the FPGA to separate the high-voltage wires from the rest of the logic, thus a failure on the charge-pump would cause mis-connection of the routing infrastructure. The charge-pump and the insulation transistors are made out of thick-gate devices which therefore collect more positive charge from radiation than normal thin-gate devices. The positive charge tends to keep the n-channel transistors on even when the gate voltage level is low, thus the correct operation of the circuit is affected.

Antifuses are intrinsically immune to SEE<sup>3</sup>, thus the observed soft-errors derive

---

<sup>3</sup>There is evidence of *single-event dielectric rupture* (SEDR) in antifuses caused by an heavy ion hitting the insulator while biased, but it was found to be an extremely low probability effect

from the user logic. The only exception is the FPGA control logic which takes care of programming and starting-up the device (for example a JTAG sequencer). An upset in the control logic can in principle corrupt the whole FPGA behaviour and require a reset. SEU-hardened FPGAs exist on the market [Actel 04].

Flash-based devices face the same total dose problems [Langley 04, Nguyen 99] of antifuse devices in an accentuated way, since they need to generate high voltages for programming and also erasing. FPGAs without charge-pump which require a high-voltage supply from outside are available commercially [Speers 99] but still the floating-gate transistor itself suffers from total dose effects [Wang 03a]. Threshold voltage shift due to holes trapped in the tunnel oxide reduce the long term retention characteristics in floating-gate devices [Cellere 04]. In addition, stored charge loss due to single-event phenomena was observed in floating-gate devices. Floating-gate memory devices are limited to applications below 100 krad.

*Flash*

FPGAs based on static memory are processed in standard CMOS technology, thus they suffer only of total dose and single-event effects in the same way as standard CMOS circuits do. Commercial radiation-hardened FPGAs protected from SEL and total dose are available [Xilinx 04] but they don't usually go beyond 200 krad. In fact, these devices contain thick-gate transistors in their I/O blocks, in order to comply many different signalling standards and voltage levels.

*SRAM*

On top of that, some commercial radiation-hardened FPGAs are still very susceptible to SEUs and a set of techniques was developed to protect the logic by programming.

### 3.4 SEU hardening techniques for commercial devices

Non-SEU-hardened FPGA chips use various approaches to mitigate soft-errors both in configuration registers and user-registers. These methods are mainly system-level and program-level techniques and include TMR and reconfiguration.

#### 3.4.1 Triple module redundancy

Full TMR, which was described earlier, is a classic method to protect user logic [Actel 97, Xilinx 01]. Since FPGAs are programmed with the aid of synthesis tools, it is sufficient to implement full TMR in the HDL code to obtain an SEU-hardened user logic. Tests on antifuse-based FPGAs [Wang 03b] using TMR also confirm the validity of this approach. Nevertheless, SRAM-based FPGAs are prone to upsets in the configuration logic as well, which can be even more catastrophic for the system. A fact that might not be evident is that exploiting TMR for the user logic gives some redundancy also to the configuration logic and protects it, even though weakly, from upsets. The point is that a triplicated user logic uses also more configuration bits and an upset in one of them results in a malfunctioning of one out of three state machines, which is anyway masked by TMR.

The difference between an upset in the user logic and an upset in the configuration bits is that while the former is corrected by TMR and vanishes after a clock cycle, the latter causes the configuration to be corrupted until is re-written to the registers, therefore the user circuit happens to have a faulty block until the FPGA

---

[Swift 95].

is reconfigured. It is important to exploit a full TMR structure because the fault could in fact be in one of the majority voters: if there is only one voter, the user data is lost.

### 3.4.2 Reconfiguration

Some commercial SRAM-based FPGAs allow reconfiguration while in-operation, meaning that it is not necessary to reset the FPGA to store a new configuration. This feature can be used to restore the configuration in the chip after an upset [Xilinx 00]: it is possible to read the configuration of the FPGA from time to time to check whether there were upsets and correct them in that case. Another possibility is to continuously write the configuration in the FPGA, without checking for errors: this technique is referred as *scrubbing*. Experimental tests have been done on SRAM-based devices using the combination of TMR and reconfiguration and demonstrated a big improvement in the system SEU robustness [Yui 03]. The drawback is the necessity of an external controller.

## Chapter 4

# A radiation-tolerant FPGA for HEP

In recent years, the progress in microelectronic technologies applied to Field Programmable Gate Array (FPGA) has allowed to decrease the costs and the development time of digital electronics in the industrial sector as well as in the space and avionics sector. The use of such devices is also appealing for HEP experiments, which are now forced to exploit ASICs in their detectors placed in the vicinity of high-luminosity particle accelerators such as the LHC.

The harsh radiation environment present in these detectors makes Commercial Off-The-Shelf (COTS) components unsuitable for the application and requires the design of custom-designed circuits. All the considerations done in the previous chapters bring to one definite conclusion: no FPGA exists that can stand the total dose of high-energy physics experiments, while commercial products satisfy the space applications' requirements. On top of that, radiation tolerant devices on the market are extremely expensive.

In the present work, a study for the development of a radiation-tolerant FPGA (RT-FPGA) for high-energy physics has been done. The goal is to make a 20-Mrad-resistant FPGA with SEU immunity for the user registers as well as the configuration storage. As mentioned in the previous chapters solutions have been proposed for the SEU sensitivity problem of FPGAs, which involve introduction of redundancy in the user logic. These techniques reduce drastically the FPGA available circuitry resources and require complex reconfiguration schemes to avoid corruption of the configuration data. Contrary to this approach, the final aim of this thesis is the development of an FPGA where SEU insensitivity is built-in, not requiring the user to exploit any special technique for SEU protection.

The 0.25  $\mu\text{m}$  CMOS technology has been used at CERN since some years and 0.13  $\mu\text{m}$  is undergoing characterization and qualification studies. These technologies are standard CMOS, thus the choice of storing the configuration in static cells is forced by the availability of the process. Investigations have been done on the possibility of using antifuses in these processes but it's not yet clear and, on top of that, the reprogrammability of SRAMs is a big advantage in respect to antifuses. The work started from the realization of the logic block (LB), defining its granularity and implementation capability.

## 4.1 Logic block implementation in 0.25 micron CMOS

Since the SRAM cell is quite big, the area of the chip will be dominated by configuration storage. A good balance between routing and logic has to be found and tends to be best with coarse-grain logic blocks. As described in Section 3.2 the optimum is obtained with 4-input LUT blocks therefore the design of the block included a register, some logic for carry propagation and a 4-input LUT. Some gates for generating wide fan-in functions are also included.

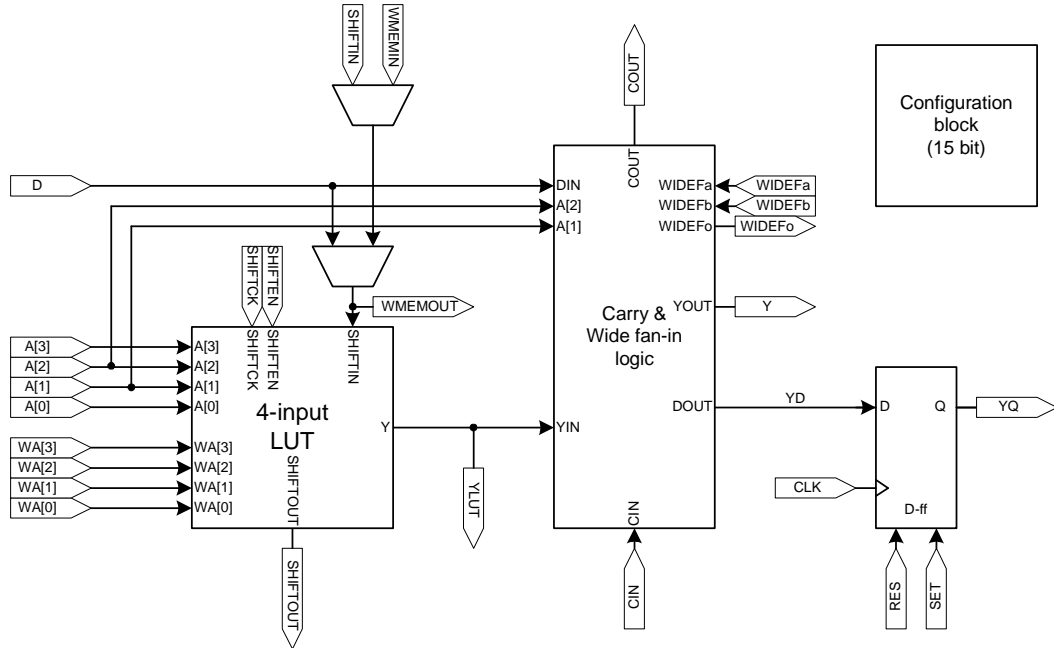


Figure 4.1: Logic block of the RT-FPGA. The connections of the configuration logic are not shown for clarity.

The logic block is represented in Figure 4.1 showing also that a LB has an additional block of 15 configuration bits.

### 4.1.1 The look-up table

The look-up table is basically composed by 16 registers holding the truth table of the generated function. These 16 registers resemble a scan chain which the configuration can be loaded through as a bitstream. In order to fully exploit the hardware, some logic is added to have the possibility of using the registers as a user dual-port synchronous  $16 \times 1$  bit RAM or as a user 16-bit shift-register. The LUT has therefore a 4-bit read address bus  $A[3:0]$  and a 4-bit write address bus  $WA[3:0]$  used only in RAM mode. A configuration bit is dedicated to store the operating mode which can be shift-register or RAM, since the function generator LUT equals to a RAM which is only read.

A simplified schematic of the look-up table is represented in Figure 4.2. A multiplexer selects one among the 16 possible register outputs, implementing the look-up table (or, from another point of view, a  $16 \times 1$  read-only memory). The select inputs of the multiplexer are the 4 inputs  $A[3:0]$  of the LUT which act as read address bus when the block is used as a RAM.

Each register has a couple of 2:1 multiplexers driving its data and clock lines.

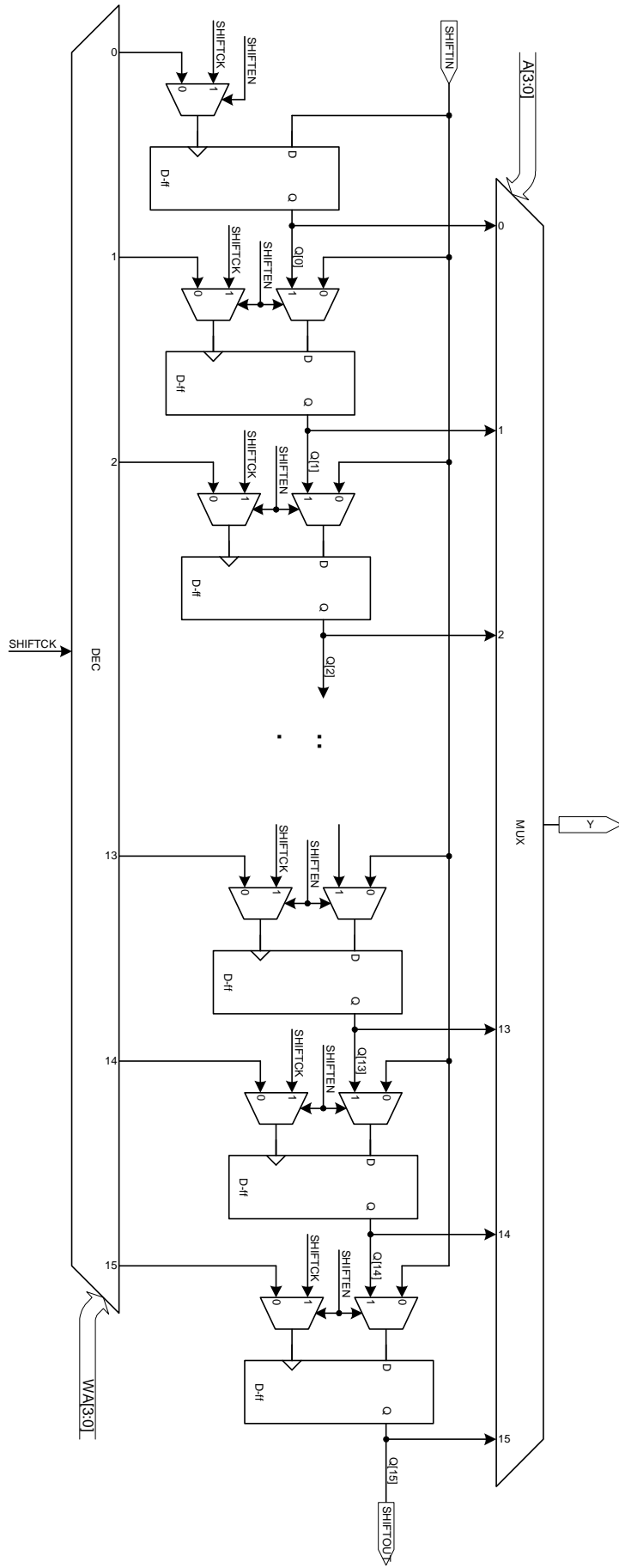


Figure 4.2: Look-up table simplified schematic. The data input multiplexer of each register is in reality included in the registers, forming scan registers, while the clock input multiplexer is included in the decoder. Dual-rail logic is employed for SEU tolerance and this is neglected in the figure.

The selection of these MUXes is done by the `SHIFTEN` signal which decides whether the block behaves as a RAM or as a shift-register. Hence, when `SHIFTEN` is high, the registers are connected in a chain, the first one fed by the `SHIFTIN` data input, and they are all clocked together by `SHIFTCK`. Conversely, when `SHIFTEN` is low, the register receives all the same data input `SHIFTIN` and are clocked selectively by the decoder. The decoder decides which register is to be clocked and operates in principle as the word-line decoder of the  $16 \times 1$  RAM. The decoder select signals are therefore the `WA[3:0]` write address bus.

As will be seen later, the data input multiplexer is in fact included in the registers, forming scan registers, while the clock input multiplexer is included in the decoder: Figure 4.2 is therefore a simplified view. On top of that, dual-rail logic is employed for SEU tolerance, and this is neglected in the figure.

Evidently, as will be seen later for many other signals, the `SHIFTCK` and `SHIFTEN` signals are user programmed but must have a specific value during configuration: to load the configuration into the LUT, this must be in shift-register mode, thus `SHIFTEN` must be high, and it must be clocked by a special configuration clock. The output of the last register in the chain goes out from the block as `SHIFTOUT` to form a chain with another LUT.

Two cascaded multiplexers drive the `SHIFTIN` input of the LUT (see Figure 4.1): depending on the user configuration the input can come from the auxiliary input `D` or from the adjacent LB through the wide-memory input `WMEMIN` or the shift-register chain input `SHIFTIN`. Basically, the auxiliary input `D` is chosen when the user has to begin a shift-register chain, while the `SHIFTIN` input is used to extend a chain beyond the 16-bit limit of a single LUT and use more than one, in fact `SHIFTIN` is connected to the previous LB's `SHIFTOUT`. The wide-memory input `WMEMIN` is instead used to feed the same input to various LUTs when in RAM mode, thus this input is the previous LB's `WMEMOUT`.

#### *The LUT register*

The registers in the LUT are all DICE-based flip-flops for SEU hardening. The circuit used is represented in Figure 4.3 and is a scan flip-flop, as mentioned before. Basically the structure introduced in Section 2.4.1 is used to build a register which is composed by two latches, a master and a slave, therefore about two times the DICE-based SEU-robust latch circuit in Figure 2.18. The difference is that the output inverting stage of the master latch is not present. This is balanced by the different node chose for driving the output buffer of the slave latch. The two latches are still separated by an inverter to avoid a peculiar SEU mode which takes place when a slave node is upset during a negative clock edge: the slave nodes are in a weak condition since the master is changing content and the pass-transistor are half-way on, thus an upset in the slave has more chance to flip the whole content of the cell despite the redundancy.

A local clock buffer is present, since each register can be clocked individually, and it is also duplicated. To each clock buffer is assigned one of the two memory nodes accessed by the transmission gates in each latch (for example `MB` and `MD` in the master). Clocks coming from one buffer drive only the transmission gates connected to the assigned memory node. In this way, an upset in one of the two clock buffers resembles an upset on one of the memory nodes, and is therefore tolerated.

The input multiplexer is just composed by a couple of transmission gates and it is duplicated accordingly to the dual-rail logic style. The scan select signals are driven externally by the `SHIFTEN` signal and its negated counterpart. Transmission gates instead of C<sup>2</sup>MOS gates are used in the memory cell for better performance.



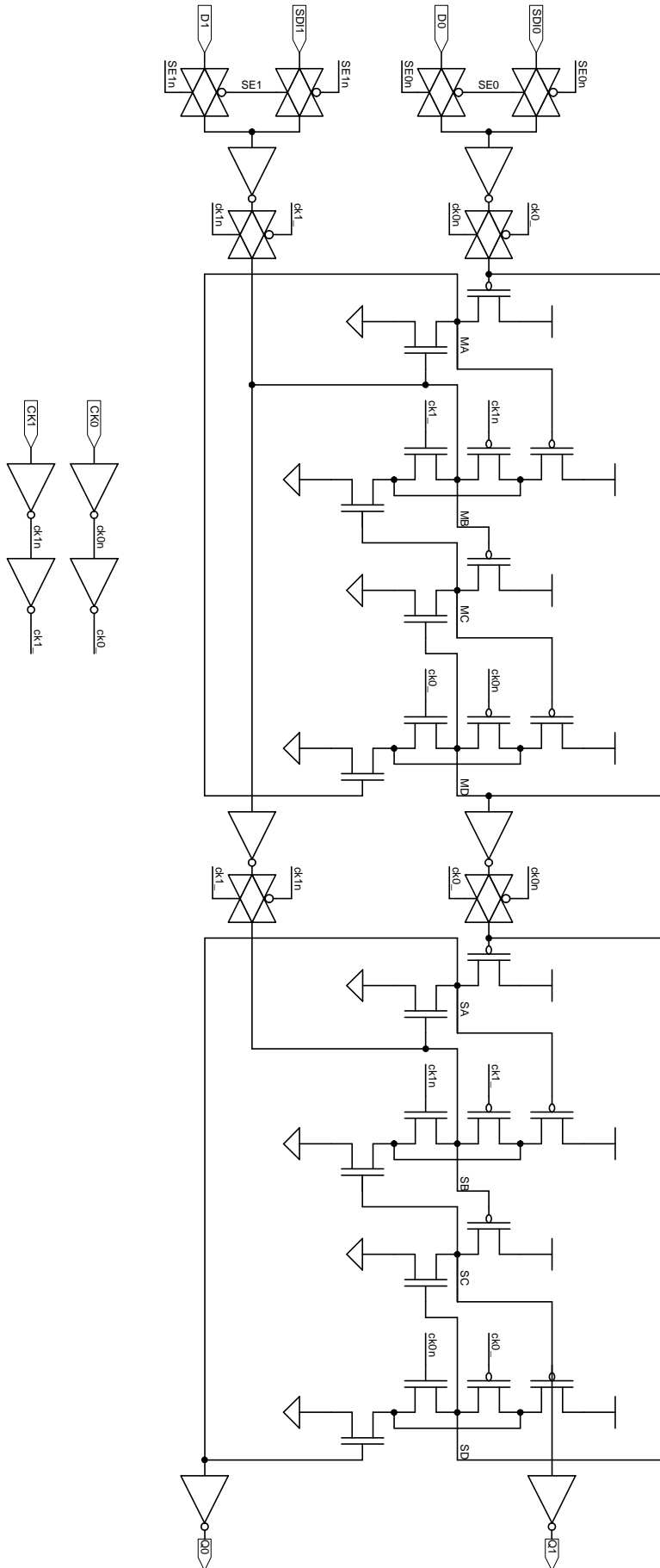


Figure 4.3: DICE-based SEU-robust scan D-flip-flop used in the LUT.

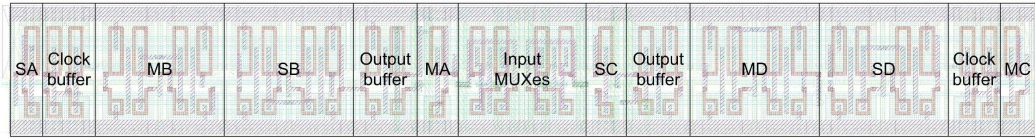


Figure 4.4: LUT register layout.

The layout of the flip-flop is represented in Figure 4.4, where node domains are highlighted. In fact, to make an SEU robust cell, memory nodes of the same latch are spaced apart in order to avoid charge collection on multiple nodes. The distance reached is at minimum  $10 \mu\text{m}$ , which should guarantee a low enough probability of upset<sup>1</sup>. In order not to lose area, the spacing is done by interleaving the slave nodes and the master nodes, since they belong to different domains they don't interact in any upset mechanism.

In Figure 4.4 the nodes are indicated with a rectangle which covers the transistor devices whose drain or source diffusions can collect charge and directly affect the specified node. Therefore series transistors are part of the same box.

Increasing the distance between nodes that have to be connected together increases also the complexity of the routing, which in the used 3-metal layers technology becomes a limiting factor. Further improvements would require a deeper metal stack.

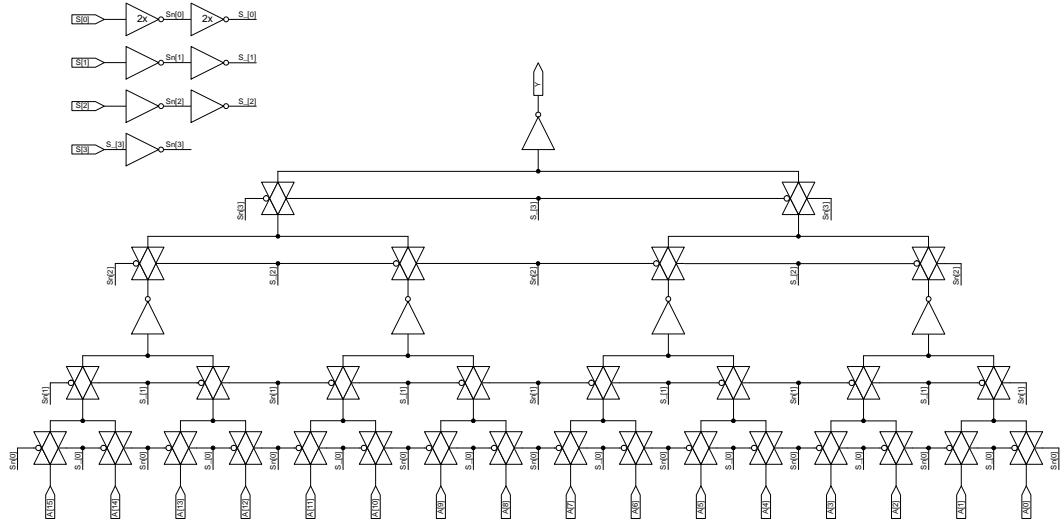


Figure 4.5: 16:1 Multiplexer used in the LUT.

### The multiplexer

The multiplexer circuit used in the LUT block is represented in Figure 4.5 and it is a simple network of transmission gates. An inverter is placed between every two transmission gate stages for buffering.

### The decoder

The decoder in the LUT is composed by a NAND pre-decoder and a dynamic NOR decoder. The circuit is represented in Figure 4.6. The choice of using a dynamic network is suggested by the fact that the output of the circuit is a gated clock thus

<sup>1</sup>Studies on a radiation-tolerant SRAM fabricated in the same CMOS technology [Gagliardi 03] shown very little correlation between upsets in cells separated by more than  $10 \mu\text{m}$ .

it replicates the phase signal. Using dynamic logic is a big advantage since only half of the transistor is needed for the evaluation even though two precharge transistors are necessary. Commonly the drawback is that only half of the clock cycle is left for evaluation, but in this case it does not apply since the output is a clock which is anyway high (resp. low) for only half of its period.

Since the decoder has 4 inputs, it could be realized by stacking 4 transistors per output line in a NAND or NOR fashion and then driving them with the input signals or their negated counterpart. To avoid stacking and improve speed, pre-decoding is possible. Pre-decoding is done with normal static gates and generates all product combinations of 2 variables. The choice of having a NOR p-type dynamic stage was done because in the 0.25  $\mu\text{m}$  CMOS radiation hardened library p-type transistors are smaller than n-channel transistors, since the formers don't need to be enclosed and have a guard-ring. A p-type dynamic 2-way NOR has 3 p-type transistors and only one n-type transistor.

An input for selecting all output lines is provided for shift-register mode LUT operation. The `single` signal allows the selection of one single line when high, while selects all lines when low. This signal is NAnDED to each dynamic stage adding a p-type transistor per output line.

When the input clock is high, the dynamic logic is in pre-discharge state, therefore all outputs are low. When the input clock is low, the dynamic logic is in evaluate state and depending on the input values, one or all outputs go to a high logic level.

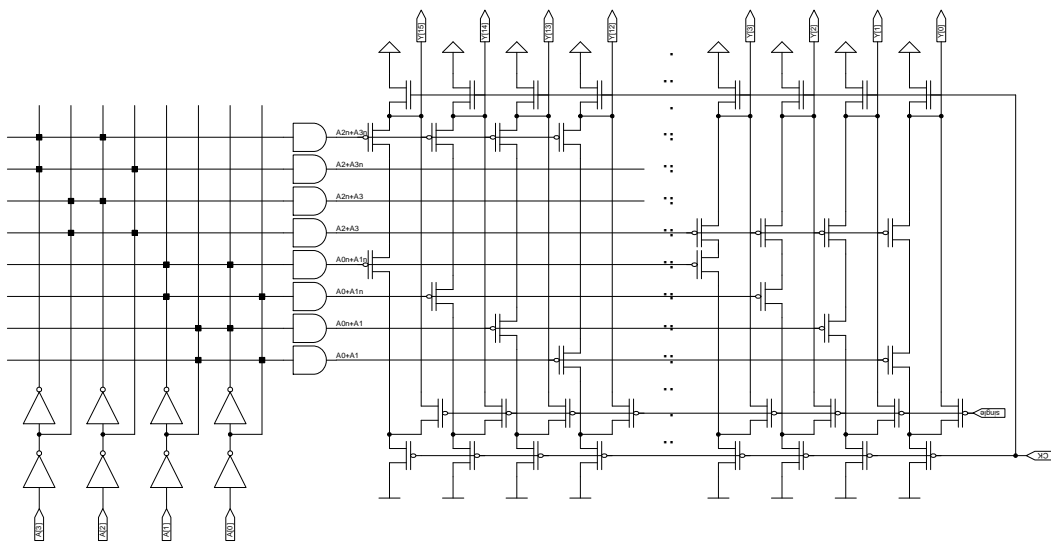


Figure 4.6: LUT decoder.

Although the dynamic logic exploited in the decoder is more sensitive to SEU, dual-rail logic is used to protect the block, thus there are effectively two decoders per LUT. The whole logic in the chip is protected by dual-rail hardening, therefore there are as well two MUXes in the LUT. This means that the chip is fully 2 $\times$  redundant and this can be justified since most of the space is anyway taken up by the SEU-robust registers storing configuration data.

Figure 4.7 clarifies the dual-rail technique for the LUT block.

Since all the logic is duplicated it is possible to create a quite symmetric layout for the LUT, which is represented in Figure 4.8. The two decoders are placed on the sides, with the input clock lines running vertically, while the two MUXes are placed

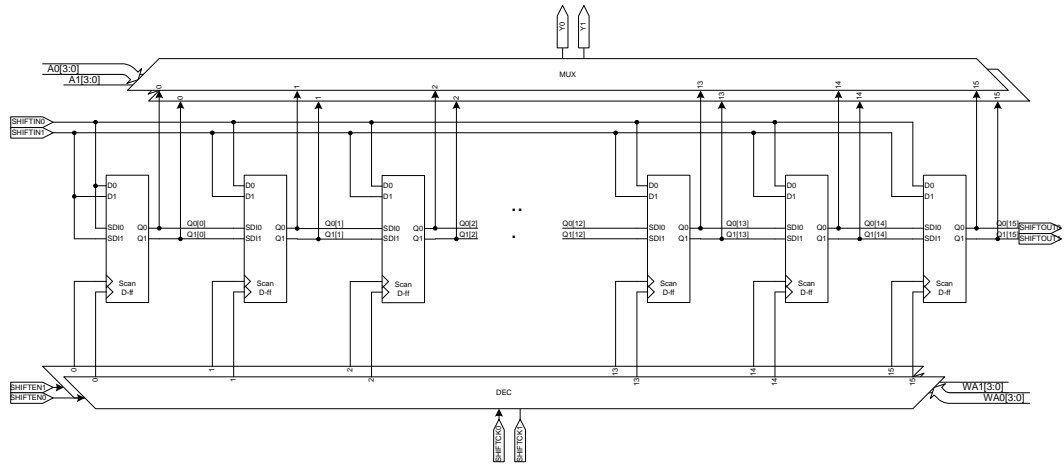


Figure 4.7: LUT schematic showing the dual-rail technique.

on the top and on the bottom. It is evident from the figure that most of the area is taken by the memory devices.

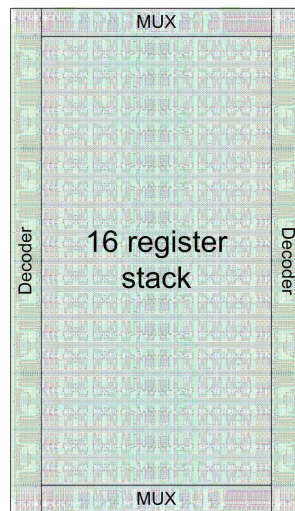


Figure 4.8: LUT layout.

#### 4.1.2 The carry and wide-fanin logic block

The carry and wide-fanin logic block is in fact split into the two homonym functional parts which share some inputs as shown in Figure 4.9. In general the carry input is used to evaluate both the carry output and the sum output, therefore it affects also the wide-fanin component, since there is the place where the decision whether the output of the LB should depend on other LBs is made. The carry input  $CIN$  can be initialized to the value of the auxiliary input  $DIN$  for the first adder in the carry chain. The configuration signal  $confCINIT$  decides whether the carry is initialized or not.

The output of the LUT enters as  $YIN$  and it is as well used both by the carry evaluation and the wide-fanin part. In case the carry chain is used,  $YIN$  is the XOR of 2 of the LUT input signals, that has to be XORed again with the carry input. In

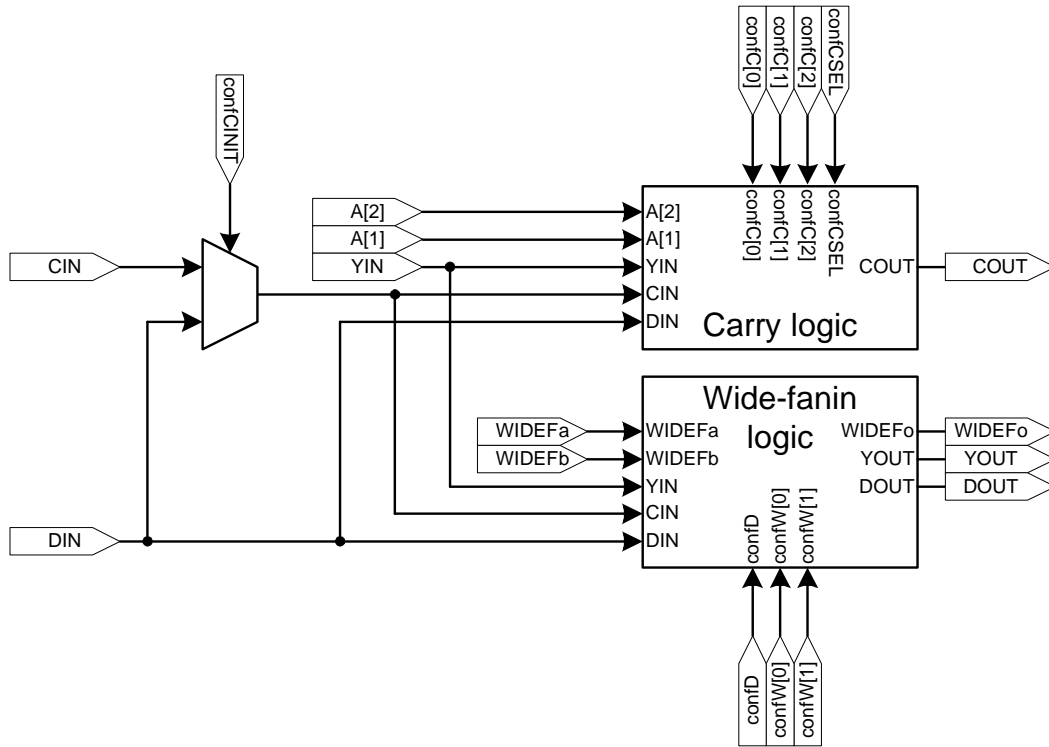


Figure 4.9: Carry & wide-fanin logic block.

Figure 4.10, which represents the wide-fanin logic part, the XOR gate between YIN and CIN is shown.

The last input common to both the carry and the wide-fanin components is the auxiliary input DIN which can load directly the user register and be an operand in the carry evaluation.

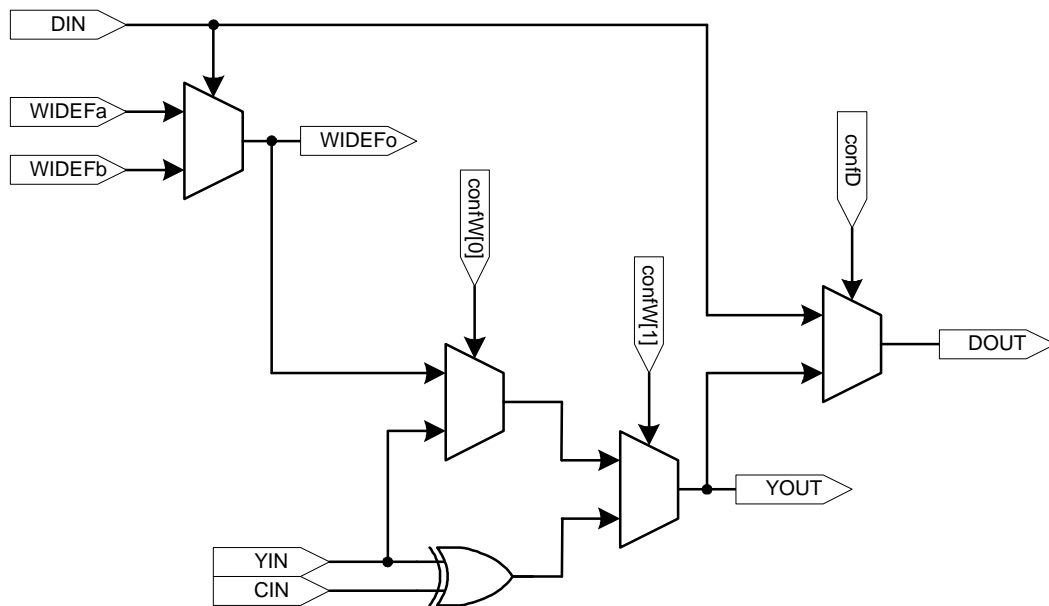


Figure 4.10: Wide-fanin logic.

As depicted in Figure 4.10, the wide-fanin function generator is nothing but a

2:1 multiplexer which selects between the two inputs `WIDEFa` and `WIDEFb`, and the auxiliary input `DIN` acts as select signal. Each one of the two input signal of the MUX are to be connected with the output of a LUT: in fact, in the same way an  $n$ -input LUT can implement any  $n$ -input function, two  $n$ -input LUT can be connected together via a MUX to form any  $(n + 1)$ -input function and the  $(n + 1)$ -th input is the select input of the MUX. Therefore the `DIN` input behaves in this case as the 5th input of a 5-input logic function. Even higher fanin functions can be generated by connecting the `WIDEFo` output of two MUXes to the input of a third MUX and so on.

The output `WIDEFo` can be connected to another wide-fanin MUX to generate even wider functions. Eventually the output will have to exit the LB through the `DOUT` or `YOUT` outputs.

Depending on the configuration bits `confW[1:0]` and `confD`, three MUXes are set to form a specific path that can alternatively bring one of the inputs or the calculated functions to one or both the outputs `DOUT` and `YOUT`. The difference between these two is that `DOUT` is connected to the user register while `YOUT` exits the LB unregistered. The auxiliary input `DIN` can connect straight to `DOUT` to the register and leave the other output `YOUT` available for use by the rest of the logic. In this way, while the user register is exploited directly for some purpose, the rest of the LB can be used for some other function, increasing efficiency.

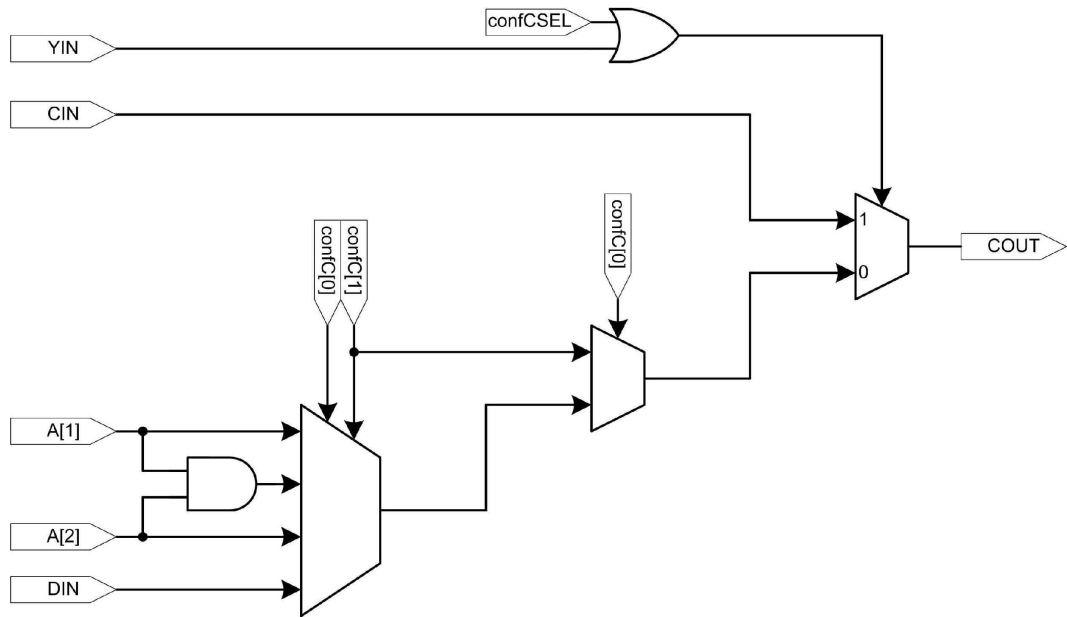


Figure 4.11: Carry-chain logic.

A 1-bit full-adder can be implemented in a LB using only 2 inputs of the LUT, which are usually `A[1]` and `A[2]`. The LUT is configured to be the XOR function of the two inputs. Depending on the expression needed the carry-chain is configured to implement addition, subtraction, comparison and so forth. As can be seen in Figure 4.11, bit `confCSEL` switches between adder and force-propagate mode.

In adder mode, `confCSEL` is high and `YIN` provides the XOR of the two operands, which therefore drives the select input of the MUX. It follows that the carry is propagated when the two operands differ and their XOR is high. Instead, the carry does not propagate and is set to the operands' value when these are equal: the MUXes are in fact configured to let either `A[1]` or `A[2]` pass.

In force-propagate mode, signal **YIN** is ignored and **CIN** is propagated, since **confCSEL** is high. This can be used for skipping one LUT in the carry chain or for initializing the carry with the auxiliary input **D** of the LB.

As results from the previous paragraphs, the carry and wide-fanin block takes overall 8 configuration bits.

### 4.1.3 The user register

The user register slightly differs from the flip-flop used for the look-up table, since it has set and reset inputs which can be configured to be synchronous or asynchronous. As evident from Figure 4.12, a set of NAND gates generates the asynchronous preset and clear signals in case the configuration bit **confASYNC** is high, otherwise they are forced to zero. Low-active asynchronous preset and clear are obtained by substituting the inverters in the memory cell with NANDs. High-active synchronous set and reset are implemented by simply gating the data input with two series NORs.

The clock can be selected between the configuration clock and the user clock. This is because the user register takes part in the configuration scan-chain and it can be loaded with an initial value at start-up, thus it needs to be clocked by the configuration clock before start-up and by the user clock after start-up. The start-up phase transition is ruled by a global signal called general write enable **GWE**, which remains low until the configuration has terminated. Neither the configuration clock nor the global write enable are protected for SEU by duplication since they are global high capacitance networks.

The register behavior can be configured to be either latch or flip-flop. Two taps are in fact available for the choice of the outputs signals **Q0** and **Q1**: one input of the multiplexers is connected to the output of the master latch for latch behavior, while the other input is connected to the output of the slave latch, for flip-flop behavior.

### 4.1.4 The configuration block

The configuration bits besides the LUT are stored in a second shift-register chain which includes the user register. The shift-register is driven by the configuration clock **confCK**. The registers used in the chain are a simplified version of the ones used in the LUT, since they are not scan flip-flops. These registers occupy an area of  $68 \times 11 \mu\text{m}^2$  each. The choice of using registers to store the configuration was done for simplicity. Future improvements might exploit SRAM cells for this purpose, saving area.

The number of bits in this additional configuration block is 15. Together, the user register, the configuration block and the LUT form a total of 32 register per LB.

### 4.1.5 LB pairs and modules

As highlighted by Figure 4.14, each pair of logic blocks is tightly coupled sharing the same user clock **CLK** and asynchronous/synchronous set and reset signals (**SET** and **RES**), besides the same write address **WA[3:0]**. The LB pair has also some common logic for the LUT clock **SHIFTCK** generation. Sharing signals among LBs helps in reducing the number of connections to the switch matrix. Figure 4.13 represents the layout of a pair of logic blocks with their common infrastructure, which is physically placed in the middle of the two LBs.

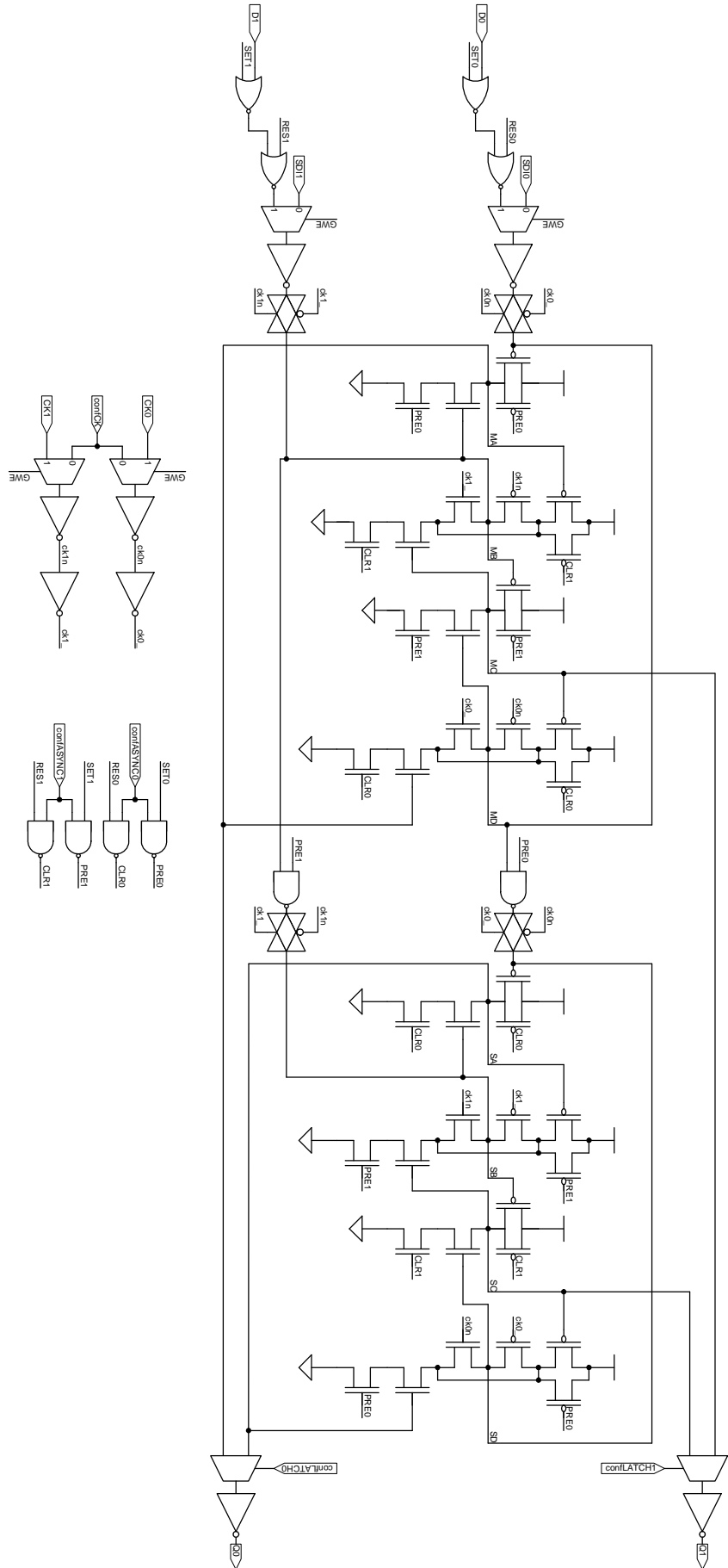


Figure 4.12: User register of the LB, with asynchronous/synchronous set and reset.



The LB pair represents the grain unity which will be connected to the routing infrastructure.

More attention has to be focused on the write addressing for RAM mode: since two LBs share some signals, they are not totally independent. This is especially true for RAM operation: having the same write address means that when both the LBs in a pair are configured as RAM, the user can only write to the same address in both LBs. On top of that the write address WA[3:0] is connected to the read address A[3:0] for one LB in each pair, making it single-port. This limits the implementation capability of a pair to  $1 \times 32$  single-port,  $2 \times 16$  single-port or  $1 \times 16$  dual-port RAM blocks. In dual-port blocks, one port is read/write, while the second is read-only.

A stack of 8 identical logic blocks (4 LB pairs) which share connections for the wide-fanin functionality forms a superset called *module*. Figure 4.14 emphasizes the connections among logic blocks: basically, a binary tree of up to 16 logic blocks acting as nodes can be structured, having 8 leaf cells thus implementing a 8-inputs logic function. In fact, the wide-fanin connections extend to the neighboring modules, allowing joining up to 2 modules for the same logic function.

The carry propagation signals CIN and COUT also extend to the neighboring logic as well as the shift-register chain signals SHIFTIN and SHIFTOUT and the wide-memory signals WMEMIN and WMEMOUT. These connections practically organize the modules in a chain. Everywhere along the chain the user can program a user-defined-width adder block exploiting the carry logic, a user-defined-length shift-register or a user-defined-size RAM block.

These chains would be organized in columns or in a snake-like fashion in the final design. Since a LB contains 32 registers, a LB pair has 64 storage cells and a module has 256 storage cells. A greater or equivalent amount of configuration bits would be needed for the switch matrix adjacent to every LB pair. The number of connections that each LB pair interfaces to its switch matrix is 17.

#### 4.1.6 Test chip in 0.25 micron technology

A test chip in a CMOS 0.25  $\mu\text{m}$  technology was developed to test the operation of the logic blocks and their behavior in radiation environment. The goal of the test chip was to test:

- (a) the functionality of the logic blocks;
- (b) the total dose robustness;
- (c) the SEU sensitivity.

The test chip is a  $2 \times 2 \text{ mm}^2$  integrated circuit containing 4 modules, therefore 32 logic blocks or, in other words, 1024 total registers. The test chip does not include any configurable interconnection infrastructure. Figure 4.15 shows a microscope picture of the chip.

Due to the area constraint, imposed by project costs, the number of pads was limited to 30, in order to have only 2 sides of the chip covered by pads, while, on the other 2 sides, core logic exploits the area which would normally be used for I/Os. This limitation forced some simplification of the connections in the test chip, since each module would need to be connected to  $\approx 70$  signals and this was not of course possible.

*Internal structure*

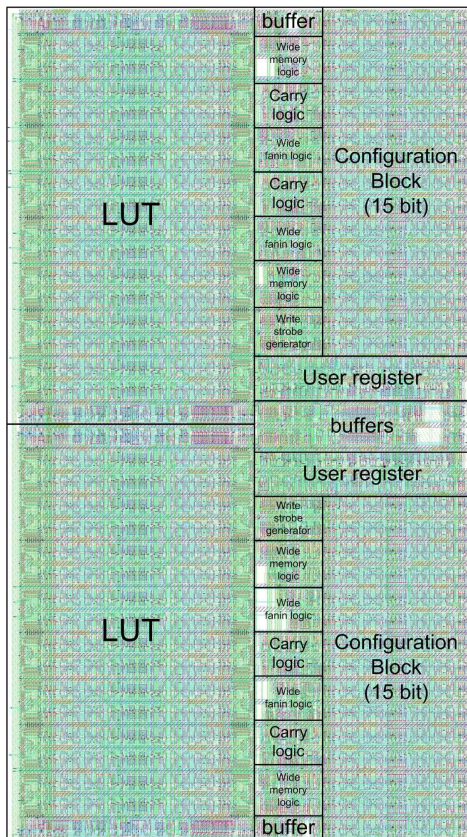


Figure 4.13: Logic block pair layout.

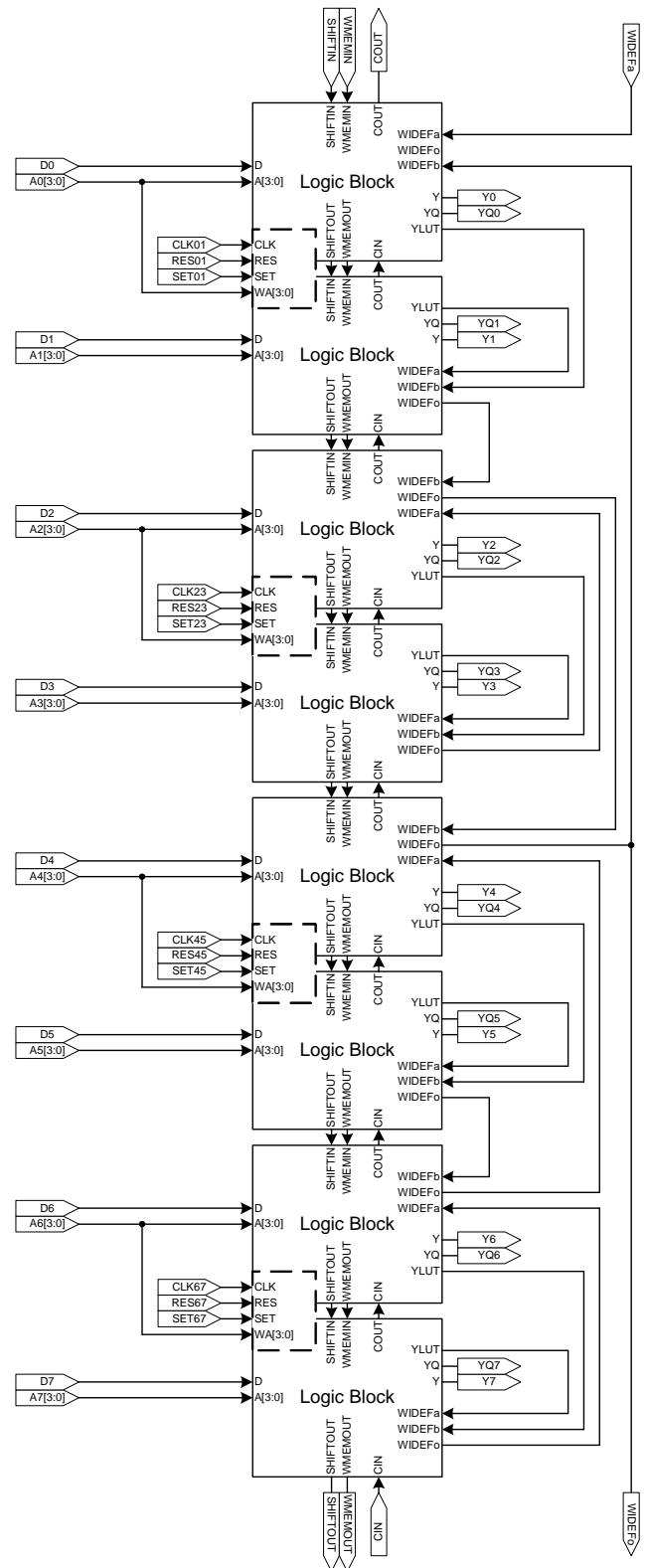


Figure 4.14: Module composed by 8 logic blocks.

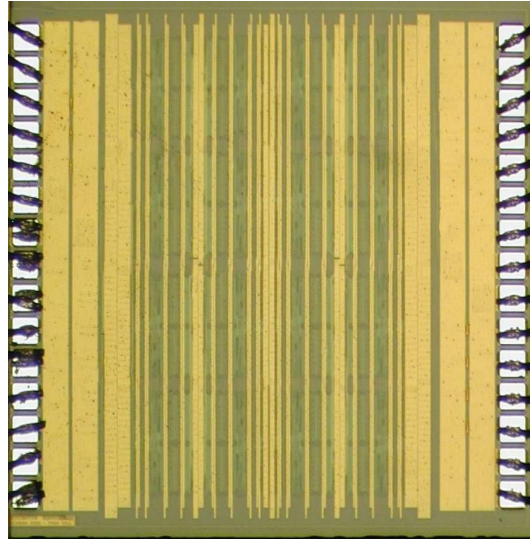


Figure 4.15: Microscope photograph of the test chip.

Therefore, in the test chip, user reset, set and clock signals of all LBs are all connected together in a buffer tree, as well as the auxiliary input signals D and 3 out of 4 of the LUT inputs A[3:1]. These account for 7 of the inputs of the chip.

One of the LUT inputs is nevertheless not connected as a global signal but wired to the outputs of other LBs: since it is necessary to test the whole circuitry, it is mandatory to somehow connect all the LBs in a testable structure which has the minimum number of outputs as possible. The easiest way is to create a chain connecting the output of one LB to the input of the next LB. This is what was done, with the only difference that there are 4 chains and not only one. It is in fact desirable to be able to test a full pair of LBs independently from the others since a first test on the chip is functional. This means that signals inside a pair of LBs cannot be merged. Therefore each pair of LBs has its 2 outputs YQ connected to the 2 inputs A[0] of another pair forming 2 data chains.

In practice it will be possible to test the functionality of each LB pair independently by programming a “feedthrough” configuration in the unused LBs, while a meaningful configuration in the LB under test. During the design, the connection of 4 chains instead of 2 was possible and this allows to test functionally 4 LBs simultaneously. Therefore there are 4 input pads and 4 output pads dedicated to the beginning and to the end of these data chains.

The beginning and the end positions of the data chains were chosen on the basis of easy layout and they don’t correspond to the beginning and end of the carry chain, shift-register chain and wide-memory chain. In the same way, the placement of the LBs favored an easy routing, minimizing the wire length. The result is that the logical path of the chains in the circuitry resembles the shape of a Peano curve. Moreover, it is clear that the module entity layout chosen for this test chip is not the one which will be used in the final application. Figure 4.16(a) depicts a pseudo-layout view of the test chip where the carry chain and one data chain are highlighted. The other three data chains have a similar logical path to the one represented in the figure.

Also the carry-chain and the shift-register chain use an input pad and an output pad each, while the wide-memory chain is terminated with the input set to ground and without an outgoing connection. The shift-register chain is concatenated to the

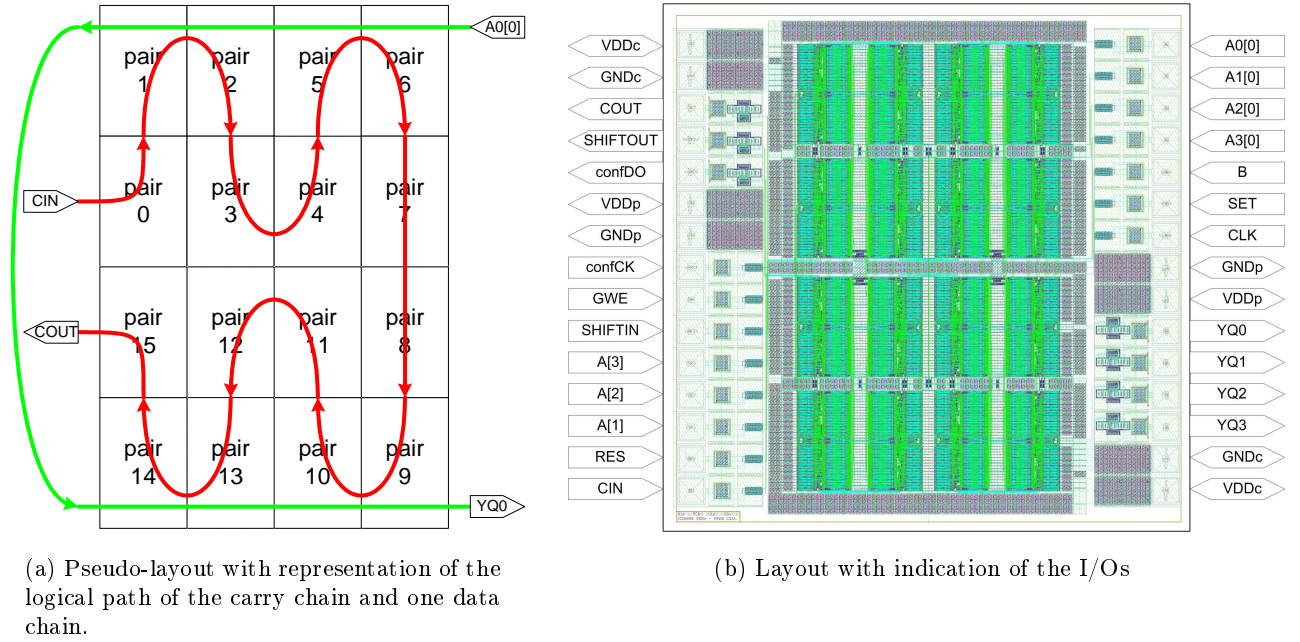


Figure 4.16: Layout of test chip.

configuration chain, in order to form a single 1024-bit-long register chain. Therefore only 3 more pads are necessary for configuration: the configuration clock, the general write enable *GWE* and the output of the configuration chain.

#### *Power distribution*

This makes a total of 22 I/O pads, which have to be completed with 8 power pads. A separate powering scheme is used for the pads and for the core logic even though the supply voltage is the same ( $V_{DD} = 2.5\text{ V}$ ), therefore 4 pads supply power to the core ring, while the rest provide power for the pads. Besides, the two pad strips on the only 2 sides used for pads in the chip are not connected in a ring, thus the powering of these two strips is also separate, in practice one  $V_{DD}$  and one ground pad each. Within the two pad strips, the two adjacent power pads divide the strip between inputs and outputs, in order to minimize the effect of switching noise propagation on the power rails from the output pads to the input pads.

Besides, balancing of the number of outputs on the 2 sides was done, in order not to have more output pads, and consequently bigger voltage drop, power consumption and noise on one side in respect to the other. The chip has a total of 7 output pads, 3 on one side and 4 on the other.

The unused area of the chip, mostly under the power rails, is filled with polysilicon-gate-N-well capacitors for decoupling.

#### 4.1.7 SEU hardening of I/O pads and global signals

As explained before, SEU protection in the chip is done by employing the SEU-robust register for storage and dual-rail logic for the combinational circuitry. Hence, most of the signals run on dual-rail, except some global signals like the configuration clock.

#### *Global signals*

All these global signals are hardened by designing the capacitance of their

branches to be high, or, in other words, to impose a high fan-out on them. Dual-rail is not anymore needed for these nets, simplifying the routing. Obviously a clock signal has a high fan-out, therefore it is necessary to build a tree for its distribution in order to minimize skews. At each stage of the buffering, thus at each node of the tree, the capacitance must be above a certain value in order to guarantee SEU immunity for a given LET.

In this work, a network is defined to be high-capacitance if its fan-out is above or equal 64 gates, which corresponds approximately to a capacitance of  $C_{th} = 1.6$  pF. Given  $V_{DD} = 2.5$  V the supply voltage,  $E_{eh}$  the energy necessary to create one electron-hole pair,  $q$  the electron charge absolute value,  $\rho_{Si}$  the silicon density and the funnel length  $L_f \approx 1 \mu\text{m}$ , then the corresponding LET threshold is

$$\text{LET}_{th} = (C_{th}V_{DD}E_{eh})/(2q\rho_{Si}L_f) \approx 190\text{cm}^2\text{MeV/mg}$$

in case there is no driver. This is more than what is necessary for the target environment of the application.

In the test chip, the high fan-out domain consists also in those nets which have been grouped together for pad limitation like A[3:1], D, SETb, RESb, CLK, GWE. They become global signals and can be distributed with a buffer tree.

The connection of these global signals to the input pad buffers must also respect the high fan-out condition, which is anyway easy since the input pads have usually a high drive capability.

Special pads have instead to be designed for all the input signals which become dual-rail: in order to protect the signal since the beginning, the pad buffer itself must be dual-rail, therefore two input buffers are necessary and both must be connected to the pad.

*Input pads*

Special output pads would also be advised, in order to postpone the conversion of dual-rail signals to single-rail as close as possible to the pad, where high fan-outs are involved. The design of such a pad can be difficult if it has to be slew-rate controlled (SRC). No special output pads were designed for the present test chip; all dual-rail signals are converted to single-rail by simply connecting the two rails on the output pad buffer.

*Output pads*

#### 4.1.8 Simulation

During the design, computer simulation was always extensively performed for functionality, performance and stability verification. The simulation tool HSPICE, a commercial version of SPICE was used for relatively small logic blocks ( $\leq 1000$  transistors). With SPICE it is possible to check accurately timing constraints, driving strengths, noise margins and other digital logic issues, as well as SEU immunity. For SEU immunity also the more specific tool SmartSpice was used.

Bigger blocks can be modeled with Verilog code and simulated logically with an appropriate software. Functional verification at higher level was done therefore with Verilog and a model for the whole chip was created.

#### 4.1.9 Packaging

A test chip contains only 1024 storage cells which could be an insufficient number for a meaningful SEU characterization. In fact, due also to the wanted SEU robustness,

in order to generate a high enough statistics, a very long beam time would be needed. Unfortunately, beam time is expensive. Since the chip has a relatively small number of I/Os, two chips were included in each package, close enough to cover them with a single beam. In this way, double the statistics is generated.

A package which can accommodate two chips and their I/Os, considering the special I/O layout, is the PGA-100. Most of the package pads remain however free. Figure 4.17 shows a picture of the package.

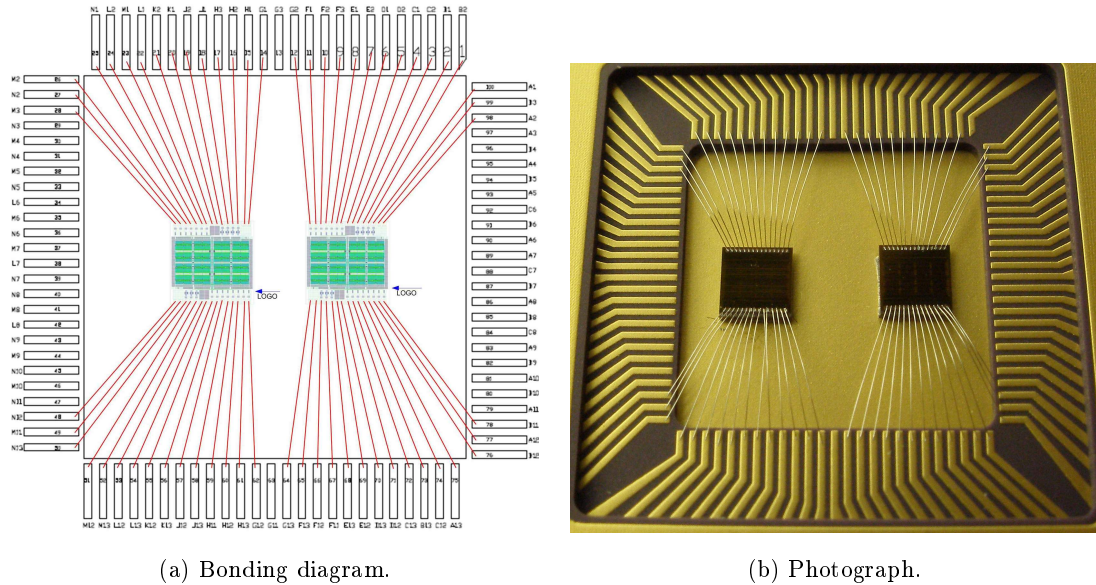


Figure 4.17: Two test chips in a PGA-100 package

A good practice for bonding is to keep the angle between the wire and the pad direction as small as possible and in any case within 45 degrees. The position of the two chips in the cavity had to fulfill this constraint.

The PGA-100 has openable metal lid, allowing the direct exposure of the chip to the beam.

#### 4.1.10 Functional testing

A wire-up board was made for testing on an IMS ATS-200 digital tester. The tester is capable of applying user-defined test patterns of input values to the input pins of the board and reading out the output values from the output pins of the board. The test pattern sequences can be obtained by simulation and for this purpose the Verilog model developed during the design of the chip was used.

The IMS ATS-200 allows very fine adjustments of timing and voltage parameters, and can test components to frequencies up to 100 MHz.

Verification of the test chip was tricky since in order to test all logic areas it is necessary to load different configurations in the LBs. Therefore a number of bitstreams with varying purpose was prepared and tried one by one on the digital tester. Examples are: 4-input XOR operation of all LBs, carry-chain operation of all LBs, 2-bit full-adder on a pair of LBs, user shift-register operation.

The tester connects to the board through 50  $\Omega$  impedance cables, therefore the input lines were terminated on board and series matching resistors were provided for

the output lines. Due to signal integrity problems the tests could be run only up to the clock frequency of 75 MHz.

Figure 4.18 shows a picture of the test setup.

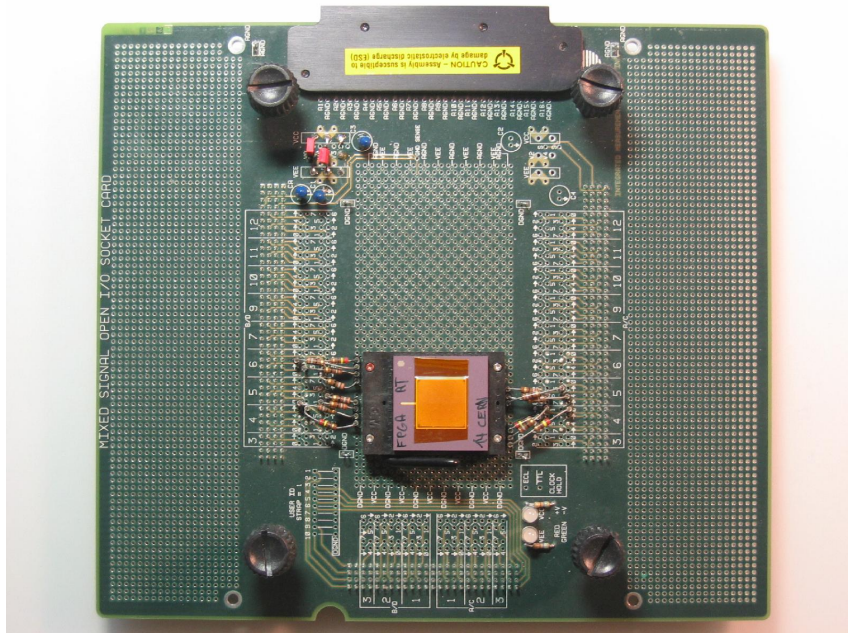


Figure 4.18: Wire-up board for the digital tester (photograph).

#### 4.1.11 Ion beam testing procedures

Ion beam testing was planned to characterize the SEU robustness of the chip and its internal structures. Ions rather than protons were chosen because the cross-section curve produced by an ion beam test contains more information about the sensitivity of the device and can be used to extrapolate this sensitivity to other environments (protons, neutrons or other particles).

Various kinds of tests were done: a static configuration retention test, a dynamic configuration test and a dynamic user data test. During each test the fluence<sup>2</sup> of the beam is accurately measured.

The static test proceeds as follows:

*Static test*

- loading while the beam is off a configuration bitstream through the shift-register chain;
- stopping clock and freezing all DUT input signals while turning on the beam for a specific fluence;
- running again the clock and comparing the output bitstream with the original one.

The dynamic configuration test is done by simply loading, while the beam is on, a long bitstream through the shift-register chain and continuously comparing

*Dynamic configuration test*

<sup>2</sup>The fluence of a beam is the number of particles which cross a unit of area and it is expressed in  $\text{cm}^{-2}$

the output bitstream with the original one. Three kinds of bitstream were used: an alternate 1s and 0s stream, an all-1s stream and an all-0s stream. These simple streams are easy to compare at high-speed on the test board.

#### *Dynamic user data test*

The dynamic user data test is instead performed in this way:

- loading while the beam is off a configuration bitstream through the shift-register chain;
- turning on the beam;
- running the DUT in user mode with random user data and acquiring the output data;
- turning off the beam at a specific fluence;
- running again the clock and comparing the output configuration bitstream with the original one.

The configuration used for this last test was chosen to be a 4-input XOR for all LBs, in such a way that any change in any configuration bit or user register should be seen at the end of the user data chain as an effective change in the output.

#### **4.1.12 Test board for ion beam testing**

A custom made printed circuit board (PCB) was designed for the beam test setup. It basically comprises a socket for the device under test (DUT), a Xilinx Spartan-3 FPGA, a USB interface and some glue logic and linear power regulators.

Practically the USB interface is to be connected with a computer, which runs a user interface program. A test pattern can be loaded and retrieved via the USB connection in the Spartan-3 on-chip memory. The Spartan-3 can then apply the test pattern to the DUT in more or less the same way as a digital tester would do; the main difference is that timings, voltages and transmission lines are not calibrated like in a digital tester. Figure 4.19 depicts a block diagram of the test board.

The USB interface comes with a 5 V supply voltage from the host computer. A regulator generates 3.3 V necessary for the USB chip; the communication between the USB chip and the Spartan-3 uses CMOS 3.3 V signaling and it comprises a 16-bit data bus, an 8-bit address bus, some control signals and a 48 MHz clock. The latter is used as a master clock for the Spartan-3.

The Spartan-3 chip has its own 3.3 V supply which is generated by another regulator from a dedicated 5 V input. From the same 3.3 V, some other regulators generate the 2.5 V supply voltage for the DUT. The communication between the Spartan-3 and the DUT takes place through level translators with dual power supply.

The board is a 2-layers PCB even though there are 4 different power supply networks and one ground network: the choice was dictated by cost and production time. The top layer is divided into signaling and supply voltage distribution, while the bottom layer is exploited for signaling and ground. An efficient grounding is essential for the operation of the board.

Since each DUT contains in fact two chips, the same board signals are applied to the inputs of both the test chips, saving some routing resources. Therefore there are 15 DUT inputs and 14 DUT outputs, since the outputs cannot share a wire.



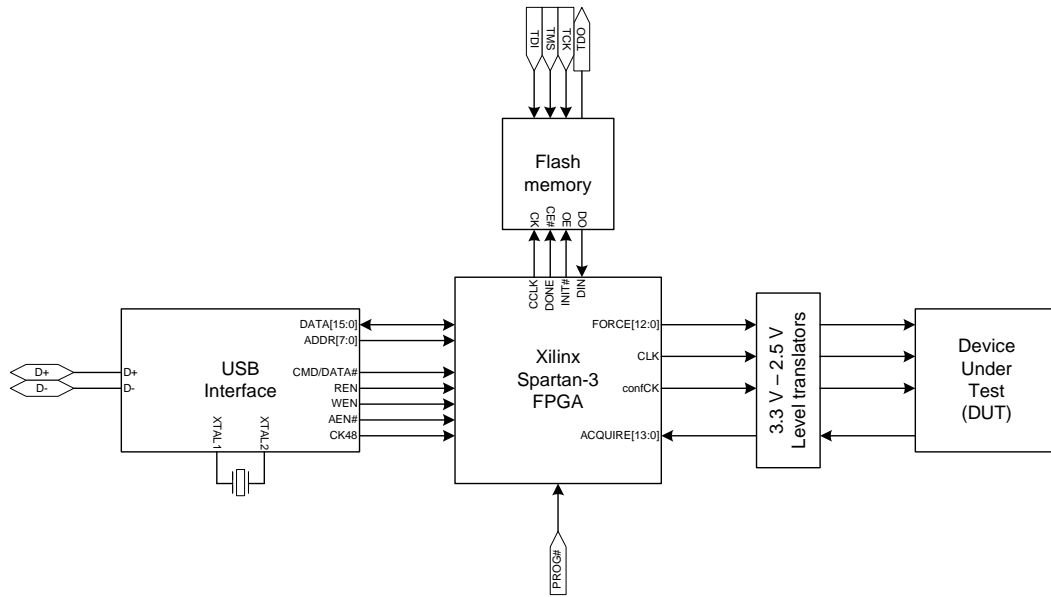


Figure 4.19: Ion beam test board block diagram.

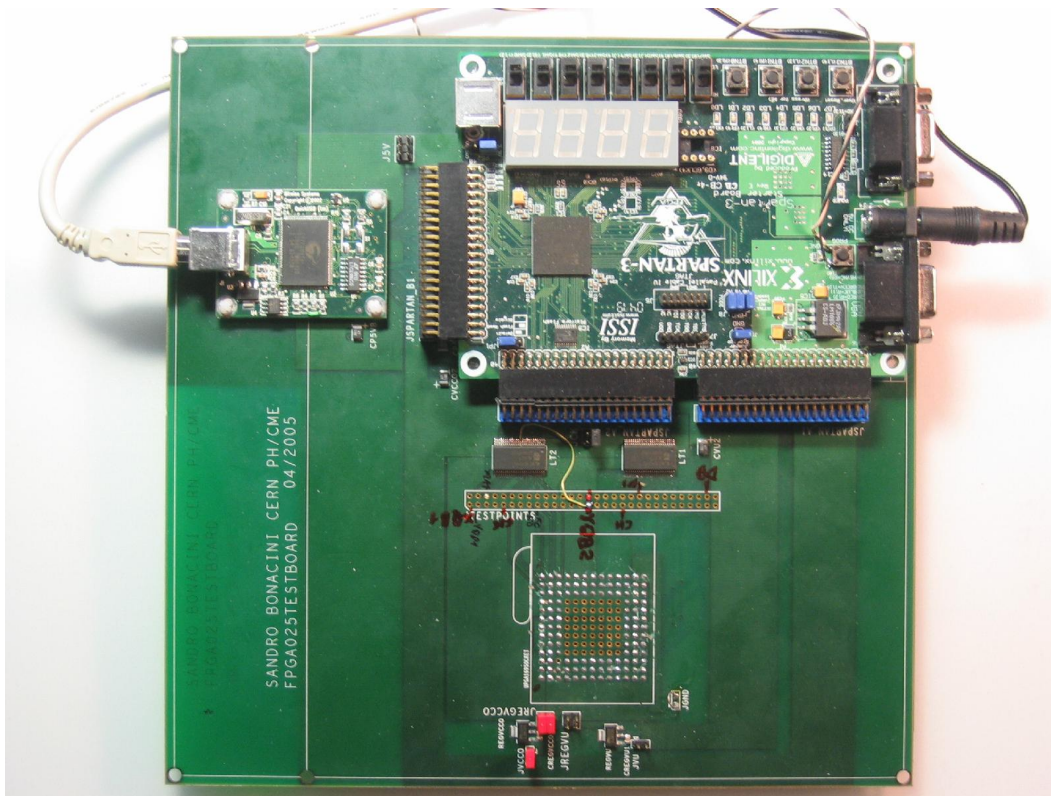


Figure 4.20: Ion beam test board photograph.

Even though the beam can be focused in a tiny spot, the chips other than the DUT are placed away from the latter in order not to be exposed to radiation: they are not radiation qualified components.

The bitstream program of the Spartan-3 is stored in an on-board flash memory which is activated at every bootstrap (power-up) and loads the configuration in the chip. This flash memory is accessed by JTAG interface from the host computer via

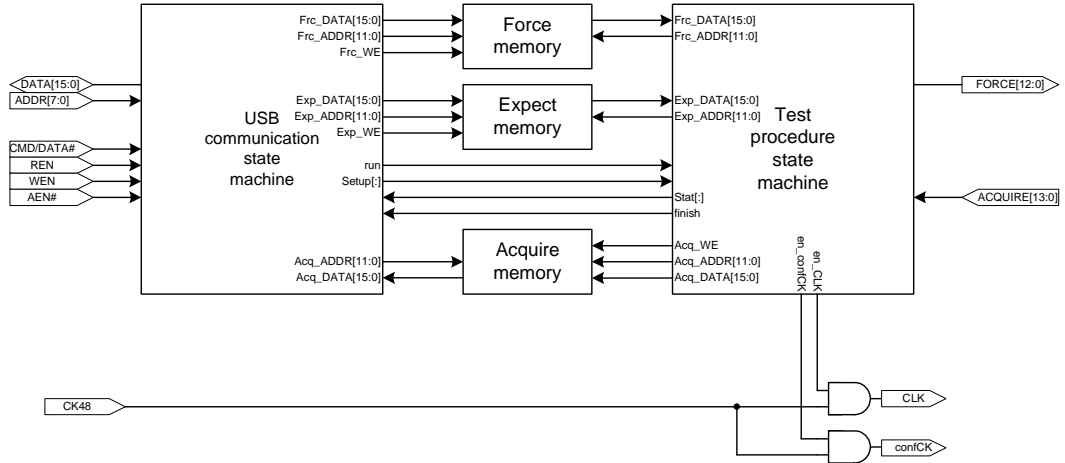


Figure 4.21: Logic block diagram of the programmed configuration in the Spartan-3 chip present on the test board.

parallel cable.

### *Spartan-3 programming*

The Spartan-3 had to be programmed to run the tests, acquire the data and communicate with the USB interface. The program consists basically of two top-level state machines, one for the USB to memory communication and one for the test operation. The two state machines work in two different clock domains and deal through a handshake. A set of setup registers, accessible from USB, was programmed to provide parameters for the test procedure.

The test state machine has two main modes: the parallel vector test mode and the serial vector test mode. Intuitively, the parallel mode is implemented for the dynamic user data test but it is also used for the static test, while the serial mode is implemented for the dynamic configuration test.

**Parallel mode** In parallel mode, the test state machine reads the content of the on-chip memory, one word per clock cycle, and assigns it to the output vector of the Spartan-3, which is the input vector of the DUT and it is also called *force vector*. Meanwhile, the output vector of the DUT, therefore the input vector of the Spartan-3, which is also called *acquire vector*, is stored in the memory or compared to the *expect vector*.

The allocated on-chip memory<sup>3</sup> is in fact split in three blocks of  $4096 \times 16$  bits, one for the force vectors, one for the acquired vectors and one for the expected vectors. All of the three RAM blocks are dual-port: the USB communication state machine can only write to the force and expect memories and only read from the acquire memory, while, conversely, the test state machine can only read from the force and expect memories and only write to the acquire memory. This creates the necessary independence to work in different clock domains.

The configuration clock and/or user clock are run in case two particular bits in the force vectors are set. This is done by gating the test clock.

<sup>3</sup>The Spartan-3 chip has two kinds of on-chip memory: distributed RAM and block RAM. In our case the block RAM was used. Not all the block RAM on the chip was used; besides, the size of the available RAM on the chip is different for different part numbers: the part used is a XS200FT-4.

The memories can be addressed with a 12-bit value. A register keeps track of the number of vectors forced and acquired, and a setup register marks the stopping address.

During the exposition to the beam a recoil particle from the DUT can still upset the rest of the logic on the board. SEU detection mechanisms were therefore implemented in the Spartan-3 program: three checksum registers are provided respectively the three memories. The acquire checksum register is incremented by the test state machine every time a value is stored in the acquire memory; conversely, the force and the expect checksums are incremented when a value is read from the memories. The force, expect and acquire checksums are readable-only via USB and the host program can identify upsets in the memories from any difference in their values and the corresponding memory content.

A USB communication checksum is also provided for reading and writing to the memories on the USB interface state machine side. In this way it is possible to distinguish between transmission errors and logic upsets.

On top of that, any upset in the state machine's logic would most likely bring to a checksum mismatch, therefore these redundancies assure a good enough SEU detection mechanism.

**Serial mode** The serial mode resembles fully the dynamic configuration test. Therefore in this mode the test state machine has 3 sub-modes: alternate 1s and 0s, all-0s and all-1s. The configuration clock is started and the user clock remains steady, while the configuration corresponding to the sub-mode is loaded in the DUT.

The state machine waits 1024 clock cycles, equal to the length of the DUT shift-register chain, and then starts checking the scan output of the DUT, which must be the same as its scan input in all 3 sub-modes (this is true because the shift-register length is even).

In respect to the parallel mode, checking of the acquired bitstream is done only on-board rather than on the host computer. The reason for this choice is that recording the acquired bitstream in the acquired memory, even though possible, would limit the length of the acquired pattern to  $4096 \cdot 16 = 64$  kbit. Since the comparison between the expected value and the acquired value is easy to do, a better solution is to record the errors only instead of the acquired pattern. A longer pattern can thus be run giving a better statistics.

Hence, two error counters are provided, one per DUT chip, as there is one force scan bit but two acquired scan bits. Besides, a record of the position of the error inside the scan chain is stored in the acquire memory: a real-time clock is incremented at every test clock cycle and its least significative bits represent the location in the chain. The start and end timestamps are also saved in a couple of configuration registers for statistics.

For SEU protection in the Spartan-3 the error counters and the real-time clock, which are critical registers, are protected by TMR, therefore triplicated and voted.

An application program for Microsoft Windows XP was made which interfaces via USB with the test board. The software is entirely written in Visual Basic.

The distinction between parallel mode and serial mode are kept in the software: the user interface window has a bigger part dedicated to the parallel mode and a small part for the serial mode. A common part also displays some information like errors and warnings. The main tasks for the parallel mode are to load the test

*Host computer  
programming*

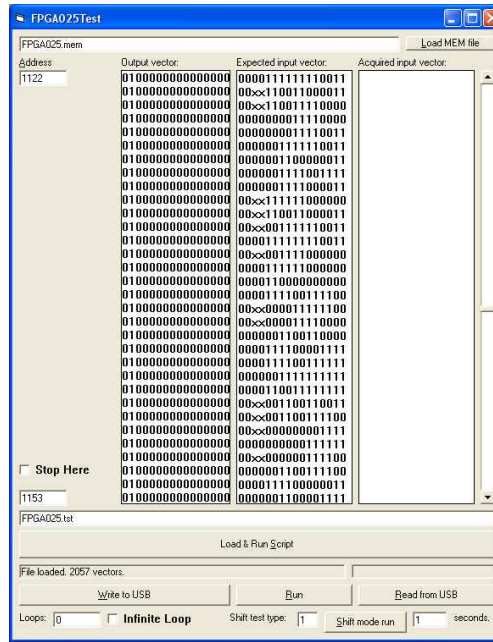


Figure 4.22: View of the beam test software user interface.

patterns from a file on the host computer, transfer them to the test board, get the results and compare them with the expected values.

The file format for the test vectors has been chosen to be identical to the IMS format for the digital tester: this allows to use exactly the same patterns of the functional tests creating some compatibility. A pattern file stores the force vector as well as the expected vector, which is the sequence of values supposed to be acquired. Any bit differing between the acquired vector and the expected vector accounts for an error.

A loaded file is displayed on the main window which is depicted in Figure 4.22: there are 3 vector boxes on the window; the first is the force vector, the second is the expected vector and the third is the acquired vector; the loaded file is therefore shown in the first two boxes, while the third is left empty.

The vectors can then be sent to the test board and run: the software takes care of storing the correct values in all the setup registers of the test board. As soon as the acquire vector is retrieved, it is displayed on the third box, and an error count appears. Each couple of bits differing is highlighted in red in the vector boxes.

The software can also schedule more than one file to be run in sequence: this allows to have, for instance, a file which loads a configuration in the DUT, another which runs some random user data on the DUT and a third file which scans out the configuration for verification. This resembles the dynamic user data test. Substituting the user data file with a simple wait state makes the static test instead.

A log file, which contains timestamps and error count is produced. In case there are errors, also the acquired pattern themselves are stored in the log.

For the serial mode, the user interface gives only the possibility to choose which one of the three sub-modes to exploit and for how much time the test should be run.

### Test setup

Irradiation was performed at the Heavy-Ion Facility (HIF) at CYCLONE in Louvain-La-Neuve, Belgium [Berger 96]. This cyclotron provides several ions covering an LET range from 1.7 to 55.9 cm<sup>2</sup>MeV/mg and with an average flux as high as

Test type	Ion	Energy	Tilt	LET <sub>eff</sub>	Average flux	Fluence	No. errors	$\sigma$
		[MeV]	[deg]	[MeVcm <sup>2</sup> /mg]	[cm <sup>-2</sup> s <sup>-1</sup> ]	[cm <sup>2</sup> ]		[cm <sup>2</sup> /bit]
Static shift-register test	Ar <sup>8+</sup>	150	0	15.1	2.0E+04	1.0E+06	0	$\leq 1.5E-09$
	Kr <sup>17+</sup>	316	0	35.6	1.0E+04	1.0E+06	0	$\leq 1.5E-09$
	Kr <sup>17+</sup>	316	45	51.2	1.5E+04	1.0E+06	0	$\leq 2.1E-09$
	Xe <sup>26+</sup>	459	0	56.3	1.5E+04	1.0E+06	0	$\leq 1.5E-09$
	Kr <sup>17+</sup>	316	60	74.0	1.5E+04	1.0E+06	0	$\leq 2.9E-09$
	Xe <sup>26+</sup>	459	45	79.6	2.0E+04	1.0E+06	0	$\leq 2.1E-09$
	Xe <sup>26+</sup>	459	60	112.0	2.0E+04	1.0E+07	0	$\leq 2.9E-10$
Dynamic shift-register test	N <sup>3+</sup>	62	0	3.5	1.0E+04	1.1E+06	0	$\leq 1.3E-09$
	Ar <sup>8+</sup>	150	0	15.1	2.0E+04	6.0E+06	0	$\leq 2.4E-10$
	Kr <sup>17+</sup>	316	0	35.6	1.0E+04	1.1E+06	0	$\leq 1.3E-09$
	Kr <sup>17+</sup>	316	45	51.2	1.3E+04	2.0E+06	0	$\leq 1.0E-09$
	Xe <sup>26+</sup>	459	0	56.3	1.5E+04	1.0E+06	0	$\leq 1.5E-09$
	Kr <sup>17+</sup>	316	60	74.0	1.5E+04	1.0E+06	0	$\leq 2.9E-09$
	Xe <sup>26+</sup>	459	45	79.6	2.0E+04	6.0E+06	0	$\leq 3.5E-10$
	Xe <sup>26+</sup>	459	60	112.0	2.0E+04	1.1E+07	7	$\leq 1.2E-09$
Dynamic user data test	Kr <sup>17+</sup>	316	0	35.6	1.0E+04	1.0E+06	0	$\leq 4.7E-08$
	Xe <sup>26+</sup>	459	0	56.3	2.0E+04	6.0E+06	0	$\leq 7.8E-09$
	Kr <sup>17+</sup>	316	60	74.0	1.5E+04	1.0E+06	0	$\leq 9.4E-08$
	Xe <sup>26+</sup>	459	45	79.6	2.0E+04	5.0E+06	0	$\leq 1.3E-08$
	Xe <sup>26+</sup>	459	60	112.0	2.0E+04	5.0E+06	0	$\leq 1.9E-08$

Table 4.1: Beam test results summary.

$2 \cdot 10^4 \text{ cm}^{-2}\text{s}^{-1}$ . The test board was mounted on a frame in the vacuum chamber and the DUT was delidded.

In order to acquire statistics for more LET values, the test board could be tilted by a 45 degrees and 60 degrees in respect to the beam. Each of the three tests was performed with different ions and at different tilt angles, covering the LET range from 15 to 112 cm<sup>2</sup>MeV/mg.

#### 4.1.13 Ion beam test results

A summary of the beam test results is shown in Table 4.1. Throughout all the explored LET range the collected statistics were very low or null, therefore in most of the cases it is possible to give only an upper bound for the cross-section. This upper bound is given with a 95% confidence level and is calculated as described in [Hagiwara 02].

Experimental data highlights the SEU robustness of the circuit up to an LET of 79.6 cm<sup>2</sup>MeV/mg, since no errors were observed up to this LET.

At the LET of 112 cm<sup>2</sup>MeV/mg the dynamic shift-register test showed a small sensitivity, with a cross-section of  $6.2 \cdot 10^{-10} \text{ cm}^2/\text{bit}$ , which gives an upper bound of  $1.2 \cdot 10^{-9} \text{ cm}^2/\text{bit}$ , while the other tests did not generate errors.

With the available statistics, the limit cross-section is below or equal to  $2.9 \cdot 10^{-9} \text{ cm}^2/\text{bit}$  throughout all the observed LET range for the static test and the dynamic configuration test. In the latter case the cross-section is higher than in the former case because the fluence, and therefore the statistics, is lower. For comparison purposes, a register fabricated in the same technology and with the same TID-hardening techniques but

not protected against SEUs demonstrated a LET threshold of  $14.7 \text{ cm}^2\text{MeV}/\text{mg}$  and a saturation cross-section of  $2.59 \cdot 10^{-7} \text{ cm}^2/\text{bit}$  [Faccio 99].

The user data dynamic test requires some additional considerations, since the number of clocked register is in this case 64. The SET cross-section at 25 MHz has an upper bound which is lower or equal to  $9.4 \cdot 10^{-8} \text{ cm}^2/\text{bit}$  throughout all the observed LET range.

Even though the clock frequency which the test was run at is quite low, the total fluence is enough for the particles to have hit the total sensitive area of the combinatorial logic affecting the user registers in the right time window of vulnerability. With the given configuration in the user data dynamic test, the combinatorial logic affecting each user register is composed of 184 gates, which correspond to a total sensitive area of about  $A_p = 1500 \mu\text{m}^2$  for the p-type transistors and about  $A_n = 750 \mu\text{m}^2$  for the n-type transistors. Given the beam fluence  $\phi$ , the chip tilt angle in respect to the beam  $\theta$ , the typical SET pulse duration  $t_{SET} \approx 200 \text{ ps}$ , the clock period  $T = 40 \text{ ns}$  and the number of registers  $k = 64$ , it is possible to estimate the number of particles crossing the sensitive area in the vulnerable time window with the formula

$$N = k(A_p + A_n)\phi \cos(\theta)t_{SET}/T .$$

This number is above 7 for all the LETs used in the user data dynamic test. The absence of errors in this test confirms the SET robustness of the presented structure.

An explanation for the errors observed in the dynamic shift-register test can be given: the LUT register has a weak region due to the mutual vicinity of the two input multiplexers (see Fig. 4.4), which could cause multiple-node charge collection phenomena. These 2 multiplexers form together a sensitive area of about  $44.1 \mu\text{m}^2$  for the p-type transistors and  $24.3 \mu\text{m}^2$  for the n-type transistors which could very well be responsible for the recorded upsets. It is clear that upsetting together both copies of the same signal results in an error.

It is possible to note that errors were observed only when the board was tilted by 60 degrees: this increases the probability of hitting multiple nodes since the particle moves also along the devices. In future versions of the LB, this problem could be corrected changing the placement of the input multiplexers of the LUT register.

## 4.2 Migration of the LB design to 0.13 micron

Due to the successful results of the first test chip and within the perspective of a long-term production of the radiation-tolerant FPGA, the design effort focussed on the migration of the logic block design to a more advanced technology. The  $0.13 \mu\text{m}$  CMOS technology is currently under evaluation for radiation hardness up to the LHC levels at CERN. The choice therefore fell on this technology, which allows for higher density logic than the previous  $0.25 \mu\text{m}$ .

In addition, there are clear indications that the  $0.13 \mu\text{m}$  technology is intrinsically hard to radiation, not requiring the use of enclosed layout transistor and allowing for regular linear transistor layout (see Section 2.1.2). For the same reasons, guard-rings seem to be unnecessary.

The design began with the research for a SEU-robust storage cell in the new technology, which was then used to port the full LB.

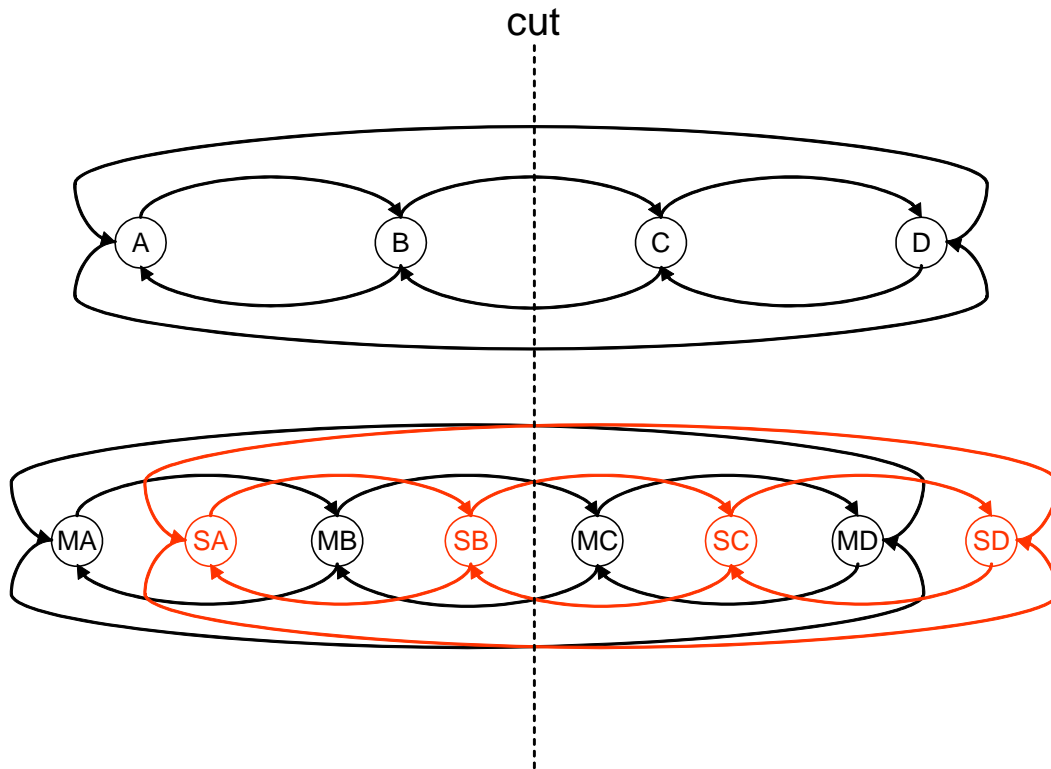


Figure 4.23: Directed graph of SEU-robust latch (top) and flip-flop (bottom) topology.

#### 4.2.1 Single interleaved SEU-robust register

The results obtained with the SEU-robust register in  $0.25\ \mu\text{m}$  cannot easily be scaled down to  $0.13\ \mu\text{m}$ , since the geometry of the cell must change (see Section 2.2.3). Since it is convenient to comply with the standard cells pitch in order to have the possibility to include the SEU-robust register in the library, the cell must be  $3.6\ \mu\text{m}$  high.

This constraint limits the available routing to 9 horizontal tracks on all metal levels between M1 and M6. As mentioned in Section 4.1, interleaving of the master and slave nodes of a DICE flip-flop is a solution for augmented robustness to multiple-node charge deposition SEUs. The trouble sits in the fact that interleaving requires lots of routing resources, quantitatively 2 wires per memory node, therefore 16 wires in total for a simple flip-flop.

At each cut between two nodes of the layout of a latch 4 wires run in parallel (see Figure 4.23), therefore interleaving 2 latches makes each cut to have 8 wires to run in parallel. It is clear that a single metal layer is not sufficient for the implementation of a simple interleaving scheme, since some routing resources are subtracted by the necessary placement of contacts, vias and local routing. On the other hand, the polysilicon layer can be used for local routing.

Two metal layers are nevertheless enough for the routing of a single interleaved flip-flop with a schematic similar to the one in Figure 4.3, but without input multiplexer.

A D flip-flop was designed for testing and the layout is described in Figure 4.24. The cell area is  $14.5 \times 3.6\ \mu\text{m}^2$ , which is about two times the area of the non-SEU-robust standard latch present in the commercial library. Two metal levels are fully

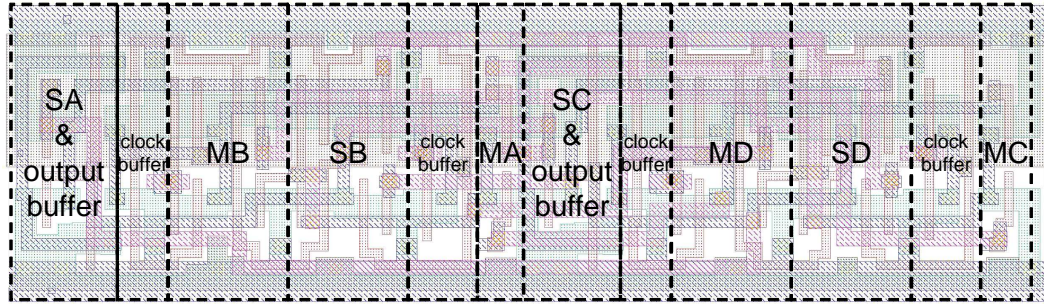


Figure 4.24: Layout of the single interleaved SEU-robust register. MA, MB, MC and MD are the master nodes, while SA, SB, SC and SD are the slave nodes. The cell size is  $14.5 \times 3.6 \mu\text{m}^2$ .

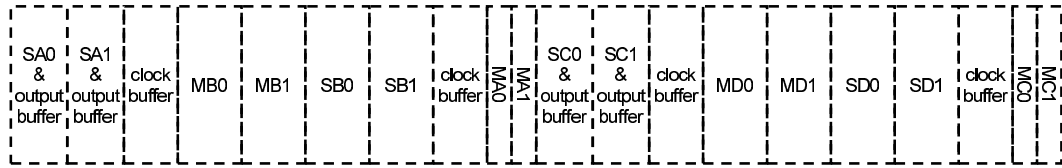


Figure 4.25: Example of a double interleaved D flip-flop. The nodes belonging to the two flip-flops are indicated respectively with the numbers 0 and 1. This layout gives a 3 times distance between sensitive nodes with respect to a single interleaved D flip-flop, but routing is 2 times denser.

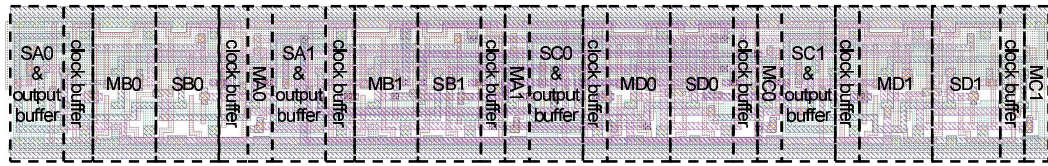


Figure 4.26: Layout of a double interleaved D flip-flop. The nodes belonging to the two flip-flops are indicated respectively with the numbers 0 and 1. This layout enhances the minimum distance between sensitive nodes with respect to Figure 4.24 but uses less routing resources and area than Figure 4.25. The cell size is  $28.7 \times 3.6 \mu\text{m}^2$ , reaching the same density as the single interleaved register.

used in the design, thus the inter-cell routing has to be placed from M3 above.

In order to protect the cell from TID, all the transistors have a width above  $0.3 \mu\text{m}$ , such that the threshold voltage shift should be limited to 100 mV. Wide transistors are however required to guarantee a sufficient driving strength to the components<sup>4</sup>.

The minimum distance between sensitive nodes is  $2.4 \mu\text{m}$  which is between the nodes SB and SC and it is 4 times less than the one obtained in the  $0.25 \mu\text{m}$  technology chip. It is evident that the advancement in technology also weakens a robustness parameter like the distance between sensitive nodes, which is very important to avoid multiple node SEUs. For this reason, a second flip-flop layout was developed.

<sup>4</sup>In fact, a minimum size inverter has enough driving strength only to drive  $\approx 1.4 \text{ fF}$  with a typical inverter delay of 25 ps. The former capacitance is equivalent to a  $5.2 \mu\text{m}$  metal line.



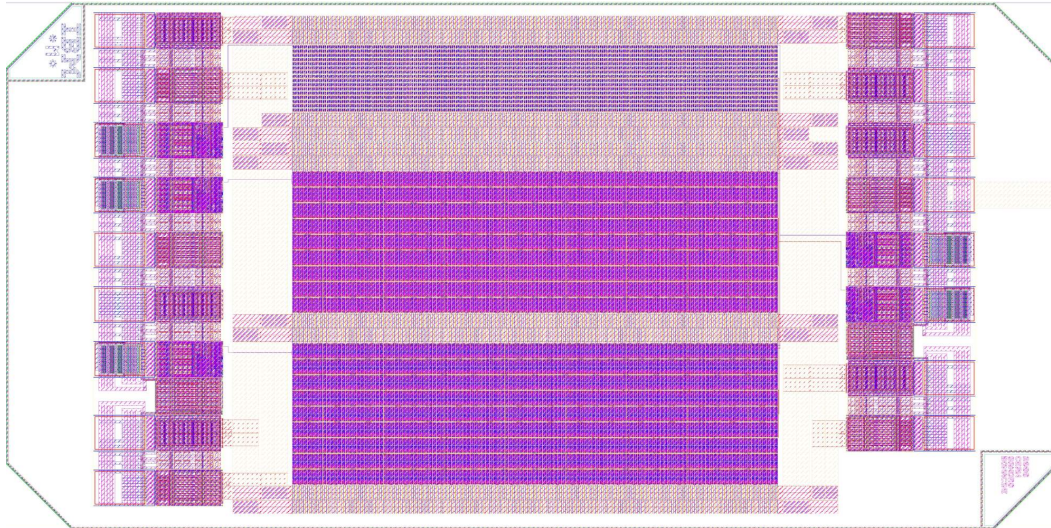


Figure 4.27: Layout of the test chip in 0.13  $\mu\text{m}$  technology for the evaluation of the two SEU-robust register structures. The chip dimensions are  $2 \times 1 \text{ mm}^2$  and it has 5 I/O pads and 12 power and ground pads. The pads are on the sides in the figure, while it is possible to see clearly the 3 shift-registers, respectively, from top to bottom, the non-hardened, the single interleaved and the double interleaved ones.

#### 4.2.2 Double interleaved SEU-robust register

A solution which can increase the distance between sensitive nodes without waste of area is to further enhance interleaving alternating the nodes of two registers instead of one. Even though, this solution requires more routing resources than the one presented in the previous section, since 16 wires need to run in parallel at each cut between sensitive nodes. In other words 3 metal levels are mandatory for this strategy.

A double interleaved register will contain two independent registers which will be only mixed together in layout. The two registers can be laid out like in Figure 4.25, having an about  $9 \mu\text{m}$  minimum distance between sensitive nodes. On the other hand this has an impact on the cell density since nodes that could share active areas in the previous case cannot anymore. Moreover, many polysilicon connections have to be substituted with metal connections due to their increased length. The layout is eased by using the same clock source for both registers, but this constrains the clock distribution.

A compromise solution can be the one depicted in Figure 4.26 which was fully developed for testing. This solution brings to a minimum distance between sensitive nodes of  $3.1 \mu\text{m}$ , giving an improvement over the single register, and uses less routing resources than the layout shown in Figure 4.25. The compromise consists in interleaving blocks of logic bigger than nodes. The interleaved blocks form a half latch each.

The cell size is  $28.7 \times 3.6 \mu\text{m}^2$  and since the cell contains 2 registers, the density is the same as the single interleaved register.

#### 4.2.3 Test chip for evaluation of SEU-robust structures

The two register architectures described in the previous sections were assembled in a test chip together with a non-hardened library register. Each one of the three

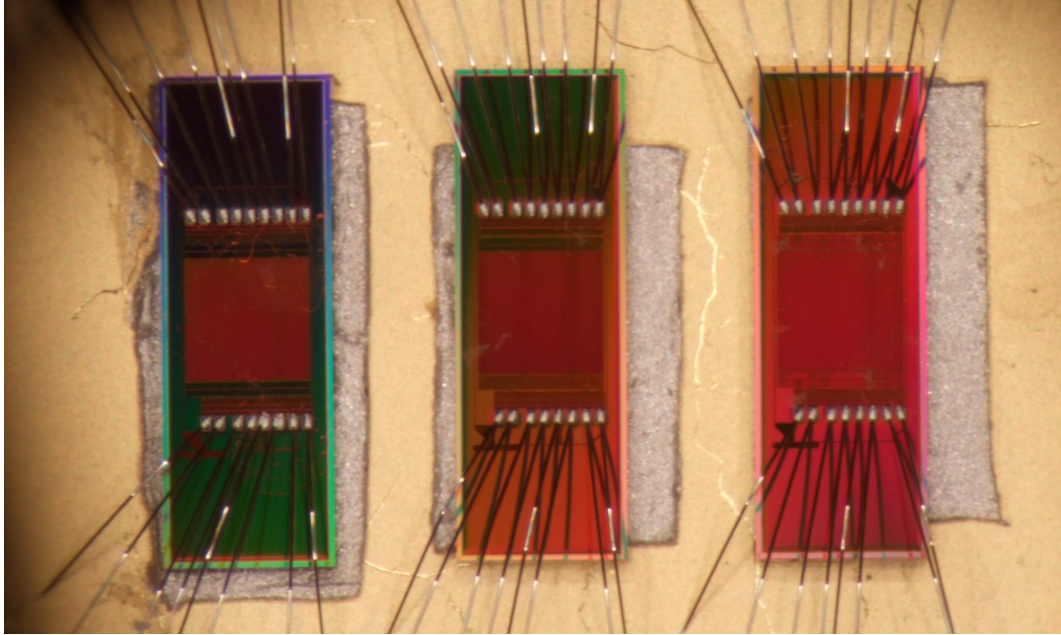


Figure 4.28: Three test chips in  $0.13 \mu\text{m}$  technology packaged in a PGA-100 box.

registers is replicated in order to form a long shift-register chain suitable for testing. The non-hardened shift register is composed of 4096 cells, while both the hardened shift-registers are composed of 4608 cells each.

The layout of the test chip appears in Figure 4.27. The 3 shift-registers get the input from a pair of common input pads, clock and data, while their outputs are fed separately to 3 output pads. The non-hardened shift-register is powered separately from the others in order to observe different changes in the power supply current due to leakage. The core is powered at 1.2 V while the pads at 2.5 V.

The two clock inputs of the SEU-robust registers are connected to two separate clock trees which unite only at the root, which coincides with the input clock pad.

#### 4.2.4 Testing procedures

The testing procedures are similar to those discussed for the  $0.25 \mu\text{m}$  test chip. In order to acquire a good amount of statistics, 3 chips were packaged in a single PGA-100 (see Figure 4.28) and exposed together to the test beam.

Dynamic tests and static tests were performed using three different bitstreams respectively composed of alternate 0s and 1s, all 0s and all 1s. Shift operations were run at 24 MHz frequency. A new test board was designed for this purpose, capable of powering the DUT at 1.2 V.

The ion-beam tests were performed at the HIF in Louvain-La-Neuve.

#### 4.2.5 Ion-beam test results

The test results are summarized in Table 4.2 and depicted in the plot in Figure 4.29. The tests proved the two SEU-robust structures to have great robustness in static mode, while they showed sensitivity in dynamic mode.

In static mode, errors were observed only in the single-interleaved cell and at an LET of  $45.8 \text{ cm}^2\text{MeV}/\text{mg}$ , which is well above the target LET threshold. On

TestType	Ion	Tilt	LETeff	Fluence	n. errors	n. errors	n. errors	Cross section	Cross section	Cross section	
		[deg]	[MeVcm <sup>2</sup> /mg]	[cm <sup>2</sup> ]	standard	SEUR-single	SEUR-double	standard	SEUR-single	SEUR-double	
								[cm <sup>2</sup> /bit]	[cm <sup>2</sup> /bit]	[cm <sup>2</sup> /bit]	
Dynamic test alternate 1s and 0s pattern	Ne	0	3.3	5.0E+06	555	0	0	9.03E-09	≤ 4.34E-11	≤ 4.34E-11	
		45	4.7	2.0E+06	294	0	0	1.20E-08	≤ 1.09E-10	≤ 1.09E-10	
		60	6.6	5.0E+06	1026	47	7	1.67E-08	≤ 6.80E-10	≤ 1.01E-10	
	Ar	0	10.1	3.5E+06	1027	0	0	2.39E-08	≤ 6.20E-11	≤ 6.20E-11	
		45	14.3	6.0E+06	2456	84	88	3.33E-08	≤ 1.01E-09	≤ 1.06E-09	
		60	20.2	4.0E+06	1866	132	133	3.80E-08	≤ 2.39E-09	≤ 2.41E-09	
	Kr	0	32.4	4.0E+06	2699	39	20	5.55E-08	≤ 7.12E-10	≤ 3.65E-10	
		30	37.4	1.4E+06	1157	25	22	6.73E-08	≤ 1.29E-09	≤ 1.14E-09	
		45	45.8	2.0E+06	2243	95	50	9.13E-08	≤ 3.44E-09	≤ 1.81E-09	
		60	64.8	1.8E+06	3558	245	170	1.61E-07	≤ 9.85E-09	≤ 6.83E-09	
	Dynamic test all 0s pattern	Ar	0	10.1	2.0E+06	333	0	0	1.35E-08	≤ 1.09E-10	≤ 1.09E-10
			60	20.2	1.0E+06	352	0	0	2.86E-08	≤ 2.17E-10	≤ 2.17E-10
Kr		0	32.4	8.0E+05	295	0	0	3.00E-08	≤ 2.71E-10	≤ 2.71E-10	
		45	45.8	8.0E+05	434	0	0	4.41E-08	≤ 2.71E-10	≤ 2.71E-10	
		60	64.8	2.2E+06	1588	0	0	5.87E-08	≤ 9.86E-11	≤ 9.86E-11	
Dynamic test all 1s pattern		Ar	0	10.1	2.0E+06	351	0	0	1.43E-08	≤ 1.09E-10	≤ 1.09E-10
	60		20.2	1.0E+06	345	0	0	2.81E-08	≤ 2.17E-10	≤ 2.17E-10	
	Kr	0	32.4	8.0E+05	344	0	0	3.50E-08	≤ 2.71E-10	≤ 2.71E-10	
		45	45.8	8.0E+05	480	0	0	4.88E-08	≤ 2.71E-10	≤ 2.71E-10	
		60	64.8	2.2E+06	1732	18	1	6.41E-08	≤ 5.92E-10	≤ 3.29E-11	
	Static test alternate 1s and 0s pattern	Ne	0	3.3	4.0E+06	453	0	0	9.22E-09	≤ 5.43E-11	≤ 5.43E-11
60			6.6	2.0E+06	427	0	0	1.74E-08	≤ 1.09E-10	≤ 1.09E-10	
Ar		0	10.1	3.0E+06	890	0	0	2.41E-08	≤ 7.23E-11	≤ 7.23E-11	
		45	14.3	1.0E+06	422	0	0	3.43E-08	≤ 2.17E-10	≤ 2.17E-10	
Kr		0	20.2	2.0E+06	1066	0	0	4.34E-08	≤ 1.09E-10	≤ 1.09E-10	
		30	32.4	1.5E+06	957	0	0	5.19E-08	≤ 1.45E-10	≤ 1.45E-10	
Kr		30	37.4	1.0E+06	867	0	0	7.06E-08	≤ 2.17E-10	≤ 2.17E-10	
		45	45.8	1.3E+06	1452	5	0	9.09E-08	≤ 2.78E-10	≤ 1.67E-10	
		60	64.8	3.1E+06	4941	0	0	1.30E-07	≤ 7.00E-11	≤ 7.00E-11	

Table 4.2: Results of the beam test on the 0.13  $\mu\text{m}$  chip. Cross-section upper bounds with 95% confidence level are shown in the cases where no errors were recorded and are indicated with ‘ $\geq$ ’.

the other hand, in dynamic mode both SEU-robust cells had errors with a cross-section strongly dependent on the tilt angle of the DUT. This is a clear indication that multiple-node charge collection plays a role in the upset mechanism, since more errors are produced when ions have a high angle of incidence.

Analyzing the layout topology of the circuits it is possible to find a SEU mode which is present only in dynamic mode and in both registers. An explanation for the errors observed could be found in the region of the register which has a clock inverter next to a slave inverter (node SA in Figure 4.24). Due to the test procedure, which kept the clock high when static, in static mode the slave latch of the register was open (transparent) and therefore did not store any value, which was instead held in the master latch. In dynamic mode instead both latches were used 50% of the time (assuming 50% duty cycle). Due to the chosen topology, the slave latch has an inverter adjacent to a correlated clock buffer, while the master latch nodes are adjacent only to uncorrelated clock buffers, therefore the slave latch is less robust to multiple-node charge collection than the master.

Still, a small difference between the two SEU-robust register is observed at LETs above 30  $\text{cm}^2\text{MeV}/\text{mg}$ , suggesting an additive SEU mechanism present in the single-interleaved register but not in the double-interleaved register. This mechanism could involve the clock buffer correlated with the master inverter MA, whose distance from each to the other is smaller in the single-interleaved register than in the double-interleaved register.

Nevertheless, both SEU-robust cells can be used for configuration storage in the target application since their static mode robustness is more than sufficient. A harder circuit should instead be used for the user-register.

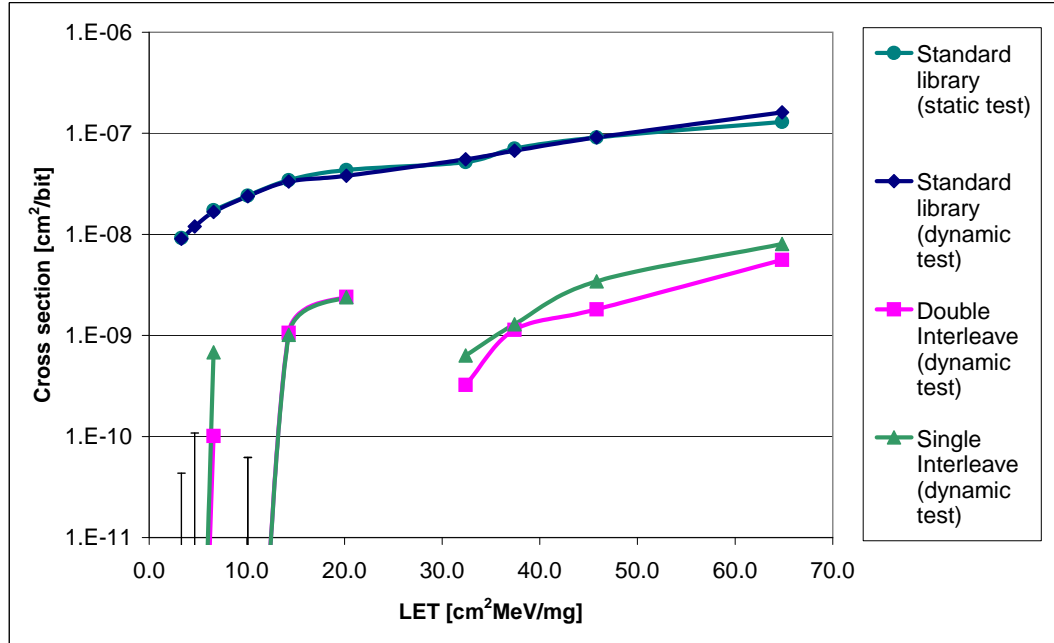


Figure 4.29: Cross-section vs LET plot of the three irradiated registers. Error bars for the upper bounds with 95% confidence level are shown where no errors were observed. The cross-section of the two SEU-robust registers show a strong dependence on the angle of incidence of the ion.

### 4.3 Development of the FPGA interconnectivity

After the LB design has been ported to  $0.13 \mu\text{m}$ , the design of the switch matrix remains to be done. As mentioned in Section 4.1.5, the chosen granularity for the FPGA corresponds to a LB pair, which exposes to the switch matrix a number of I/Os equal to 17. The rest of the I/Os of each LB pair is directly connected to the other pairs, therefore it does not participate in the configurable routing and it will not be considered in this section.

#### 4.3.1 Switch matrix architecture

The FPGA interconnectivity must be a balanced combination of local connections, which bring signals between neighbouring or close cells, and long connections, which bring signals between distant places on the chip. A LB pair together with its adjacent routing forms a *tile*, which is the basic structure repeated in two dimensions to form an array.

#### Wires

In order to let the user implement a non-congested routing the number of horizontal and vertical wires should be about the same of the total LB pair I/Os which have to be connected. In this design the number of wires is 18 per direction, with a ratio of 1 : 2 between long-distance lines and short-distance lines. There are therefore 6 long lines and 12 short-distance lines. Each LB shall thus preferably be connected with neighbouring or nearby LBs, in order to reduce congestion and delay.

The short-distance lines are divided into local lines and double lines. The former are interrupted by a switch at every tile, while the latter are interrupted only at every second tile. The ratio between double and local lines is again 1 : 2, having 4 double lines and 8 local lines. Figure 4.30 depicts the wiring architecture designed.

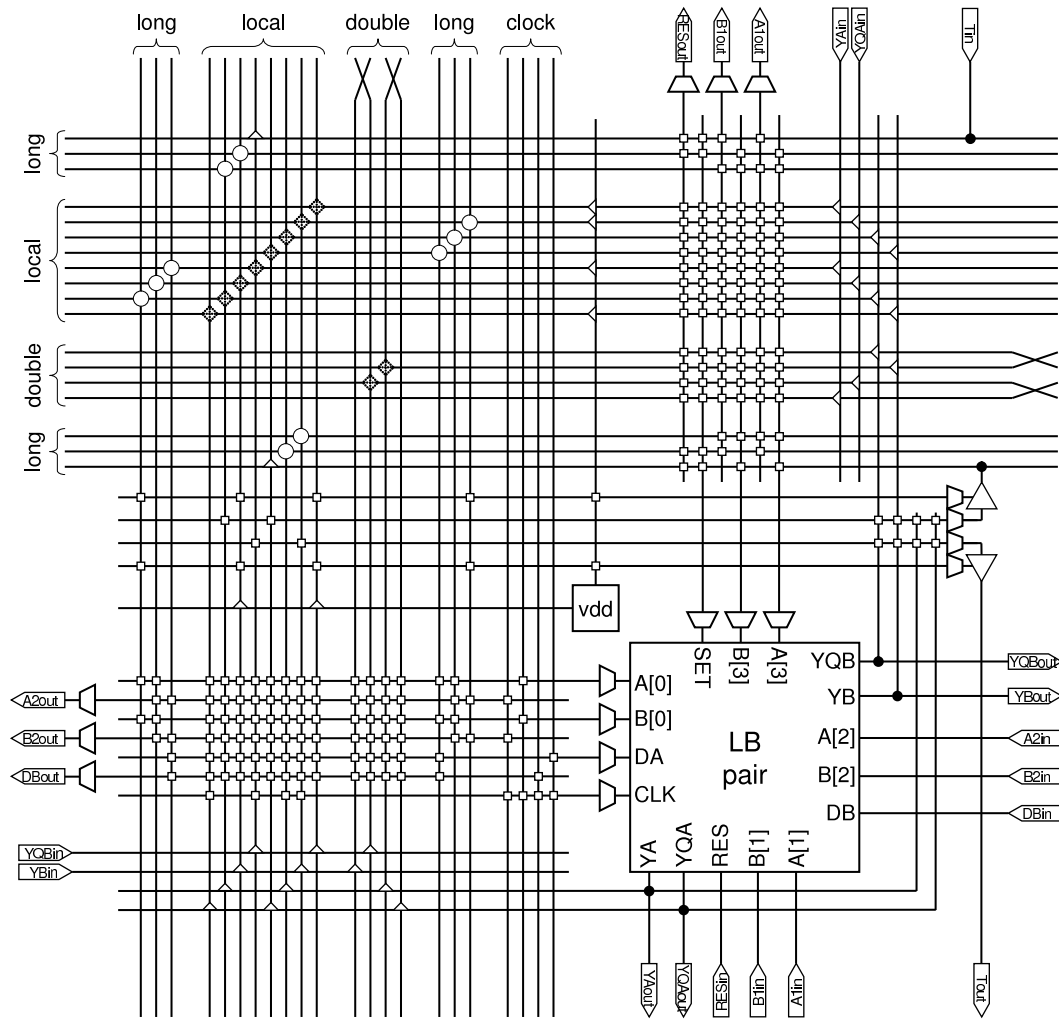


Figure 4.30: Periodical structure for the interconnectivity. The block is replicated in 2 dimensions in order to create a large array. Carry and wide-fanin connections between LBs are not shown. The symbols ‘ $\triangle$ ’, ‘ $\diamond$ ’, ‘ $\circ$ ’ and ‘ $\square$ ’ stand for special connection circuits explained in the following figures, while the symbol ‘ $\bullet$ ’ indicates a regular circuit node.

Four clock tree lines are available for the **CLK** input and other inputs of the LBs. Each clock tree line shall be connected as a global network coming from a dedicated pad.

In addition, adjacent tiles share a number of direct connections. In fact, the **YB** and **YQB** outputs of each LB reach directly the neighbouring tile on the right, while the **YA** and **YQA** outputs reach the neighbouring tile on the bottom.

The inputs of each LB pair are physically divided among the four sides of the block in order to distribute their load. There are 3 inputs per side, except for the left side which has the clock pin as fourth input. The placement of the inputs favours their direct connection from a specific neighbouring cell. For instance, output **YQB** of a cell can be connected with minimal routing resources utilization to the inputs of the rightwards neighbouring cell.

The several wires present in a tile are connected in different ways depending on the length of the lines and their purpose. The schematic in Figure 4.30 contains

*Switches*

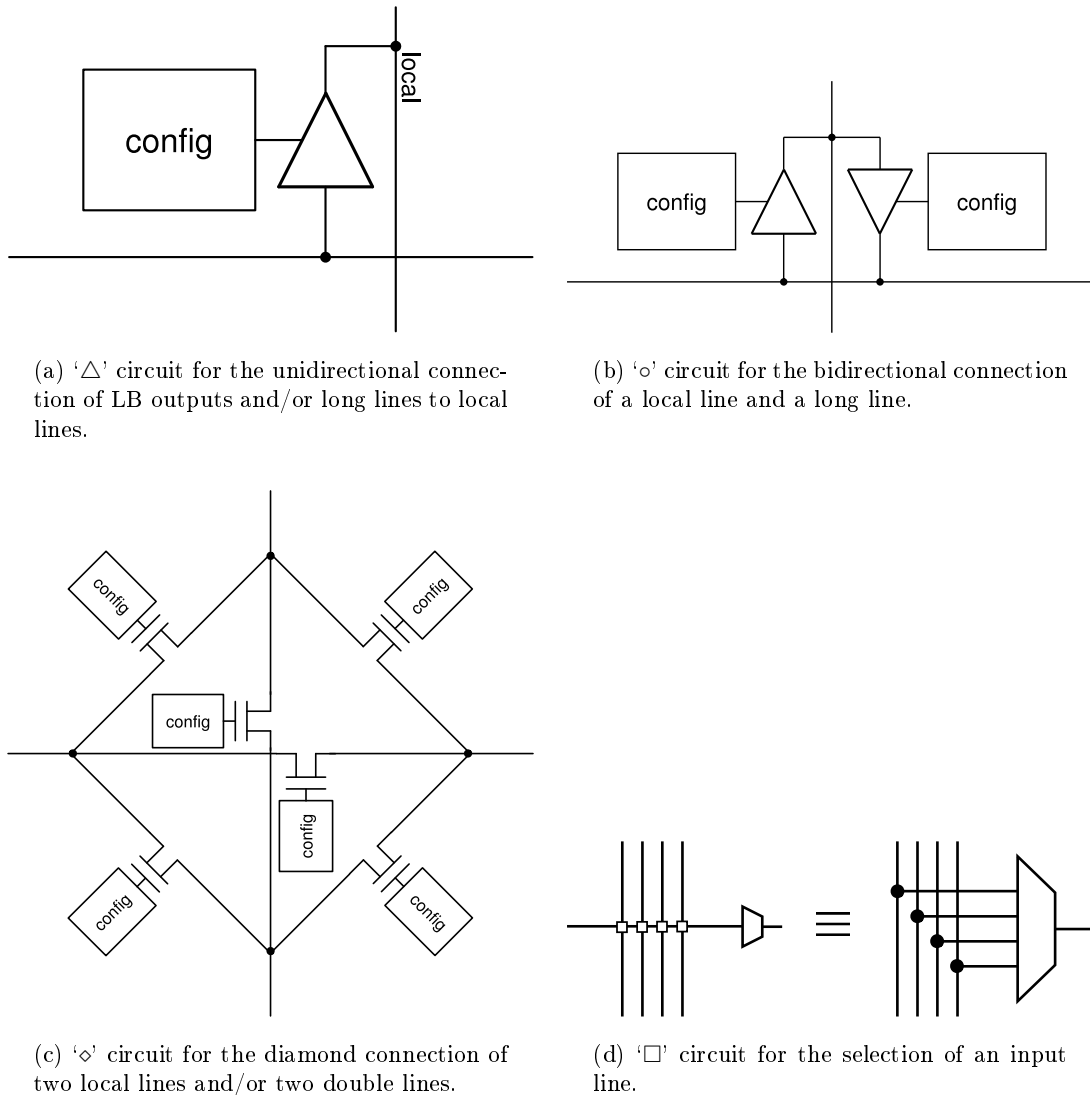


Figure 4.31: Schematics for special connections in the switch matrix.

various symbols which are explained in Figure 4.31.

Connections of long lines and LB outputs toward local or double lines are implemented with a tristate buffer whose enable terminal is driven by a configuration setting (see Figure 4.31(a)). The possibility of tying a few local lines to the logic high value is provided and in this case the connection is simply made by a p-channel transistor driven by a configuration bit.

Bidirectional connections between long lines and local lines are made by two opposed tristate buffer which can be turned on alternatively by a couple of configuration registers (see Figure 4.31(b)). Two horizontal long lines are nevertheless driven in a special fashion by a tristate buffer whose enable pin can be controlled by user data and not only by configuration. These tristate buffers are visible in Figure 4.30. Through these tristate buffers it is also possible to feed the output of the LBs directly to the two long lines.

Local and double lines are interrupted, respectively at every tile and at every second tile, by a *diamond* switch. The diamond switches are situated at the in-

tersections of corresponding local and double lines, thus there are 8 switches for the local lines and 2 switches for the double lines. A diamond switch is composed by 6 pass-transistors, controlled by 6 respective configuration bits, which allow the connection of the 4 lines segments reaching the switch from the four directions to each other (see Figure 4.31(c)). Nonetheless, diamond switches do not provide any buffering of the signals, therefore the number of diamond switches a signal can cross is limited.

The inputs of the LBs are selected among the available wires by multiplexers. Each LB input has its own multiplexer connected to a set of the lines present on its side of the tile.





## Chapter 5

# A radiation-tolerant PLD

In some HEP experiments or accelerators designs a few glue logic or simple function between ASICs is often necessary, for example for adaptation or bug fixing. In these cases a PAL/PLD is often of help since it can be tuned to the user needs. Obviously, within radiation environment, a radiation-tolerant PLD is mandatory; hence work was done for the development of such device. This chapter focuses on this work.

### 5.1 Structure

The main constraint for the design is the cost, which has to be as low as possible, and translates directly to a requirement in terms of area utilization. The aim was to build a  $2 \times 2$  mm<sup>2</sup> chip.

All PLDs are based on non-volatile cells since they cannot have a bootstrap sequence and they have to be functional already at power-up. The choice on the kind of memory cell for configuration falls therefore on fuse-based cells. In the available technology at CERN electrically-programmable anti-fuses are not present. However, laser-programmable fuses are part of the process and they are what will be used in this design.

*Storage cell*

As will be explained later, the fuse storage cell could not be smaller than about  $32 \times 14$   $\mu\text{m}$  and this imposed a constraint in the design.

The PLD is composed of a traditional AND/OR architecture, which in other words consists in a programmable AND matrix followed by a fixed OR wiring (see Section 3.1.2). The inputs of the PLD enter vertically in the matrix together with their negated counterpart, as shown in Figure 5.1 which depicts a section of the PLD corresponding to a single output.

*AND/OR  
architecture*

Each row of the matrix constitutes an AND of the programmed vertical lines, generating a minterm. Thus, depending on the program, any minterm can contain any of the input signals in its positive or negated form. The OR block sums the minterms, generating the boolean function.

The output of the OR goes to a simple configurable logic block which has a bidirectional port connected to an I/O pad. Each logic block feeds back a signal to the AND array, which can be either the output of the logic block or the input from the pad. The feedback comes to the AND matrix in its positive and negated form, occupying therefore 2 vertical lines.

It follows that the AND matrix has a number of vertical lines equal to  $m = 2(j + k)$ , where  $j$  is the number of inputs and  $k$  is the number of outputs. This

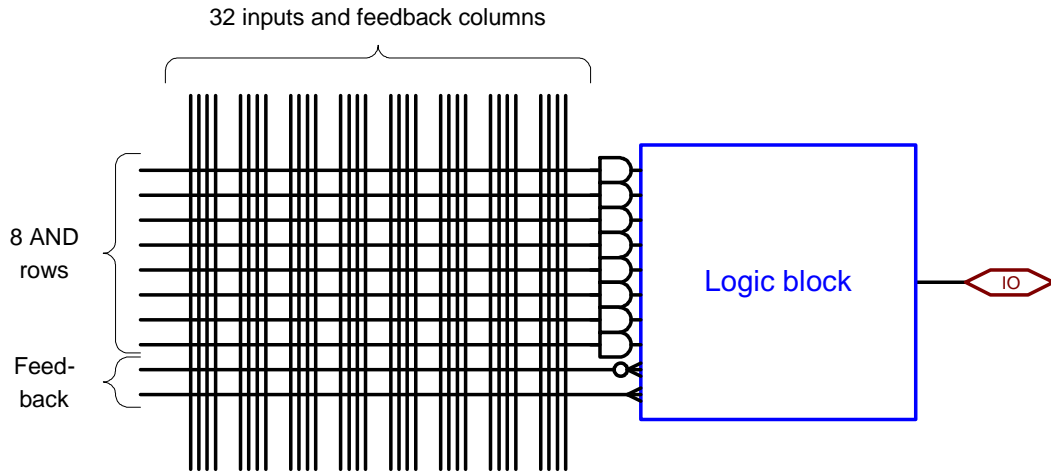


Figure 5.1: Section corresponding to one output of the PLD. Each intersection can be programmed to form a connection.

architecture gives the possibility of using the feedback of some logic blocks to create more complex functions. Commonly, the number of inputs is the same as the number of outputs, which means  $j = k$ , therefore  $m = 4k$ .

The number of horizontal lines in the AND array is instead  $n = pk$  which depends on  $p$ , the number of minterms per output.

With the given storage cell size it is possible to estimate the size of the AND array which can fit in the available area. Considering an area of about  $1 \text{ mm}^2$  for the core, it is feasible to fit about 2048 fuse cells which correspond to a  $64 \times 32$  matrix suitable for a PLD with  $j = 8$  inputs,  $k = 8$  outputs and  $p = 8$  minterms per output. Figure 5.2 represents the structure of the PLD.

The PLD has two additional inputs which interact directly with the logic blocks. These are the clock and output enable inputs, which are connected to each logic block and run vertically in the figure. These inputs can however be used as normal inputs since they are connected to an alternate input of the first and the last logic block; the choice depends on the configuration.

### 5.1.1 The logic block

As mentioned before each logic block accepts 8 minterms coming from the AND matrix. Each logic block is connected to an I/O pad and it has an alternate input connected to a second pad. The behaviour of each logic block depends on 4 configuration bits which select among 3 different modes of operation: registered, simple and complex. The mode of operation decides whether the logic block uses the primary pad or the alternate pad and whether it uses it as input, output or I/O.

The schematic of a logic block is shown in Figure 5.3. The multiplexers present in the schematic are controlled by the configuration bits. A configurable XOR provides negation of the output of the OR in case the output of the block needs to be inverted. Many times, in fact, depending on the number of necessary minterms, it is convenient to synthesize the negated value of a logic function rather than its positive value and then invert it.

#### *Registered mode*

In registered mode, the logic block exploits its user flip-flop, which is clocked by the dedicated clock pin **CK**. The input of the flip-flop is the OR among the 8

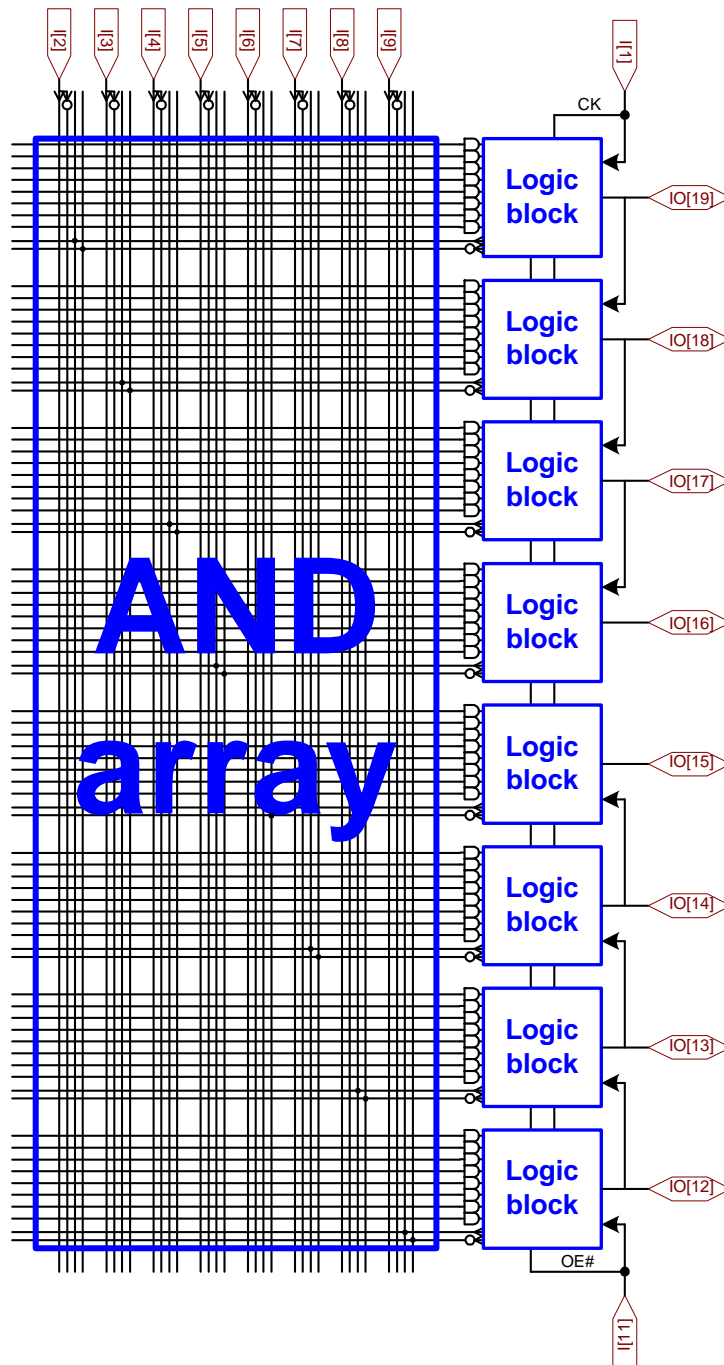


Figure 5.2: Internal PLD structure diagram.

minterms coming from the AND matrix while the output goes to a tristate buffer controlled by the dedicated output enable pin OE.

The feedback to the AND matrix is the output of the flip-flop and not the value present on the I/O pad. The alternate input pad of the block is ignored. Figure 5.4(a) clarifies this configuration mode.

In complex mode, the logic block behaves asynchronously and is configured for bidirectional operation. The tristate buffer is controlled by the first minterm entering in the logic block, which does not participate in the OR. Only 7 of the minterms are

*Complex mode*

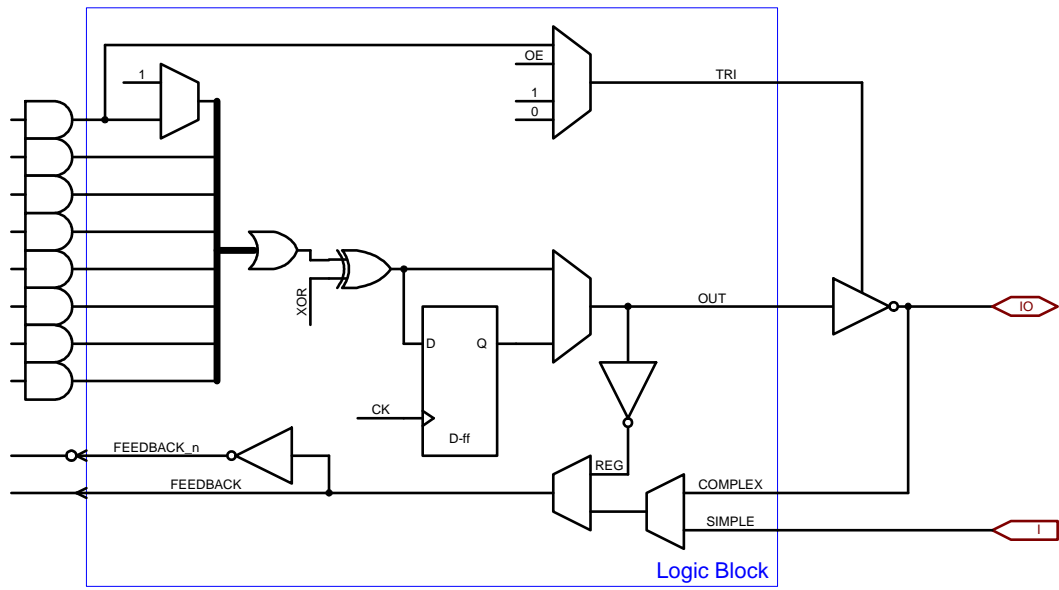
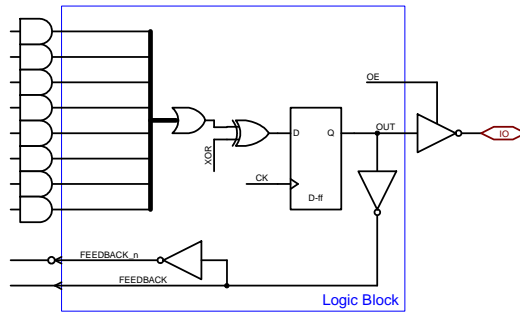
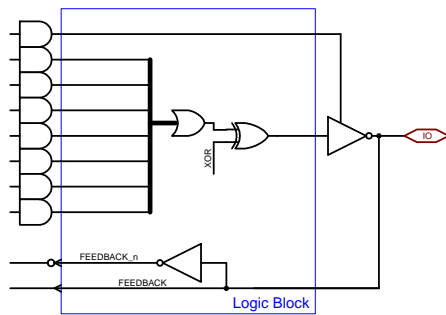


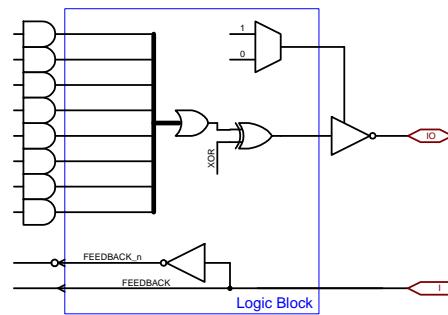
Figure 5.3: Logic block schematic.



(a) Registered mode



(b) Complex mode



(c) Simple mode

Figure 5.4: Behavioural block diagrams for the three possible configurations of the logic block.

ORed and are output through the tristate buffer.

The feedback comes directly from the I/O pad, therefore it corresponds to the output value of the logic block when the tristate control is in output mode. When the tristate is in input mode instead the feedback is the input from the pad. The first and the last logic block are nevertheless an exception since their feedback path is in this case generated respectively by the clock and output enable pins, which behave in this case as normal inputs. Figure 5.4(b) depicts this configuration mode.

In simple mode, the logic block is again asynchronous but it is configured for either input or output operation. The tristate buffer is in fact controlled by a configuration bit, which does not change after programming. All the minterms participate in the ORing which is then forwarded to the tristate buffer through the XOR.

*Simple mode*

The feedback path is generated by the alternate input which comes either from another logic block configured as output or from an input pad. Once again in this case the clock and output enable inputs behave as normal inputs and drive respectively two feedbacks to the matrix. Figure 5.4(c) shows this configuration mode.

The configuration bits of each logic block are stored in 4 fuses cells, therefore there are totally 32 fuses in addition to the AND matrix content.

*Configuration bits*

### 5.1.2 The fuse storage cell

In the CMOS technology used fuses are available with a specific set of layout rules. These rules regard the passivation opening size and the placement of the metal layers within the opening area. A minimum 14  $\mu\text{m}$  size is specified for the fuse opening in the direction along the fuse, while a minimum 39  $\mu\text{m}$  size is specified for the orthogonal direction. More than one fuse is allowed to sit in the same passivation opening in a parallel fashion provided they respect a fixed distance of  $\approx 6 \mu\text{m}$  among each other and a minimum distance of 9  $\mu\text{m}$  from the passivation opening perimeter. All this means that the smallest passivation opening allowed can fit about 4 fuses.

Fuse layout rules also impose the presence of a substrate contact guard ring around the passivation opening and no metals apart from the fuses can run over its enclosed area. This rule prevents the possibility of placing the horizontal and vertical lines of the AND matrix over the fuse area, therefore some space has to be reserved for their allocation. The guard ring has to be 6  $\mu\text{m}$  far from the opening, thus the storage cell pitch is increased by this amount in all directions.

Since the number of horizontal lines is two times the number of vertical lines, it is necessary to choose a placement for the fuses which compacts the vertical size occupancy and evens off the aspect ratio. The target is to have a square chip core. For this reason 4 horizontal lines are routed along each fuse group of 4 fuses. Each horizontal line connects to a fuse and each fuse is in turn connected to the ground potential through a transistor. Nevertheless, this routing choice still leads to an unbalanced aspect ratio with a bigger horizontal size. In order to get an even more balanced ratio the fuses are placed horizontally (long side horizontal), forming a group of 4 fuses which is bigger in the vertical dimension. Figure 5.5 shows the placement of the groups of fuses.

The area utilization of a group of 4 fuses is about  $32 \times 56 \mu\text{m}^2$ . Some area has to be added for the routing of ground and power supply lines.

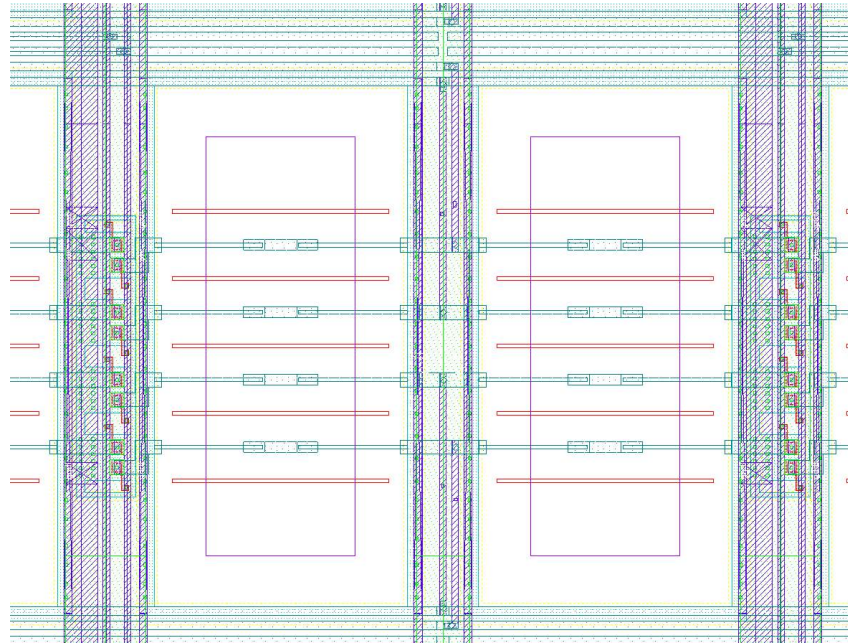


Figure 5.5: Layout of the fuses in the AND matrix. The violet line represents the passivation opening and the cyan rectangles inside are the fuses (metal 2). Each fuse is attached on one side to a short segment which connects an horizontal line running on the top or on the bottom. A transistor is placed on the other side of each fuse, acting as a pull-down. The transistors are connected in vertical lines (blue, metal 1) which correspond to the vertical lines of the AND matrix.

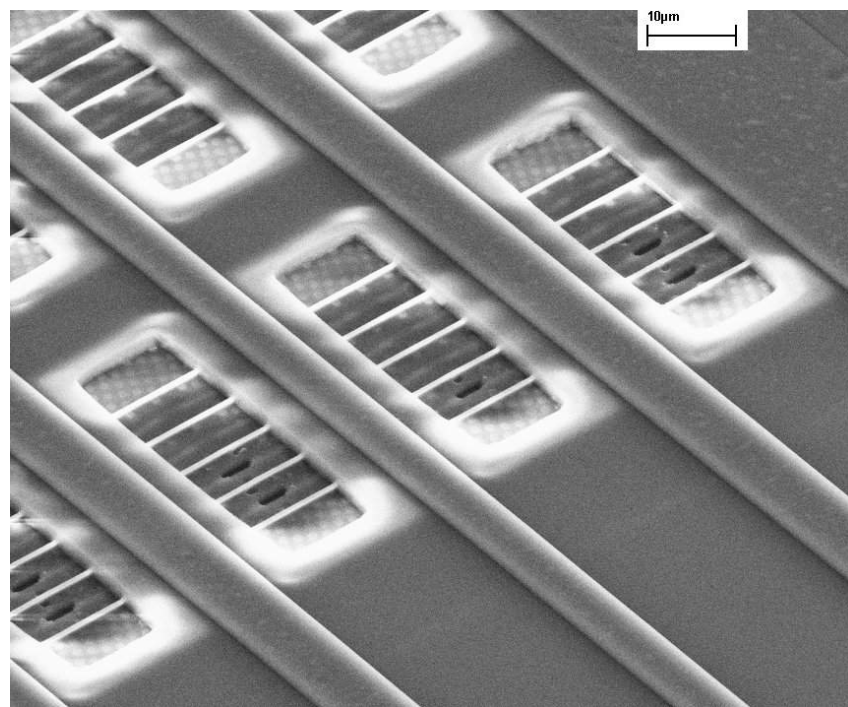


Figure 5.6: Microscope picture of a few fuse groups. The difference between burnt and not burnt fuses is clearly visible.

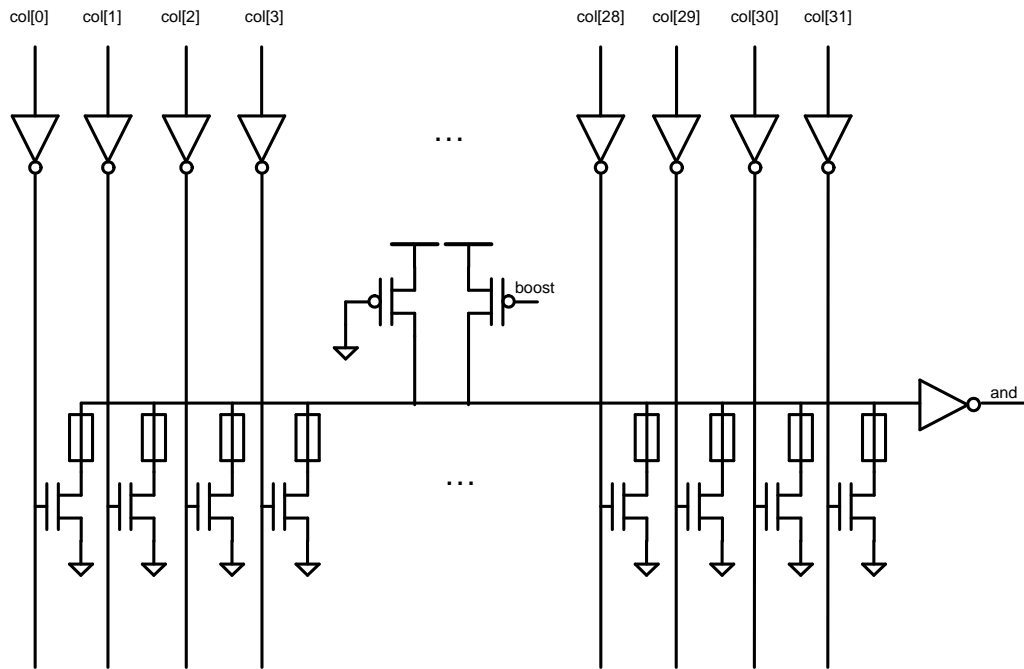
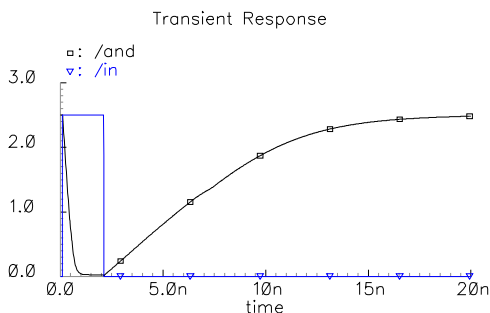


Figure 5.7: Wired-AND structure of the AND matrix. Several pull-down transistors are connected to an horizontal line. Two pull-up transistors are provided: the first operates as constant pull-up, while the second is activated only when the logic inputs change.

Every transistor connected to a fuse acts as a pull-down. In the AND matrix, all transistors in the same fuse column have their gate connected to the same vertical line, which is one of the inputs to the AND matrix.

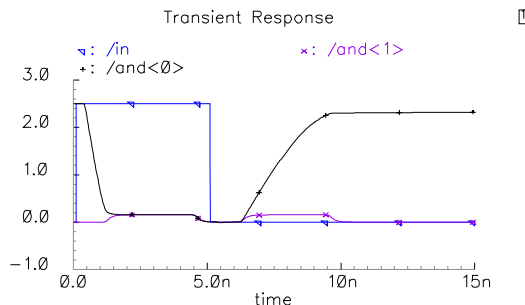
Fuses are fabricated in the second metal layer and will be burnt by a laser.

testPAL\_RT test\_RC schematic : Dec 20 10:02:31 2006



(a) with no secondary pull-up and a primary constant pull-up of 100  $\mu$ A.

testPAL\_RT testATD schematic : Dec 20 09:44:41 2006



(b) with secondary pull-up of 500  $\mu$ A and a primary constant pull-up of 10  $\mu$ A.

Figure 5.8: Wired AND transient responses simulations.

### 5.1.3 The AND matrix

As mentioned previously, the AND matrix is basically a set of wired-AND gates, laid in horizontal lines. Each horizontal line has therefore several pull-down transistors, one per vertical input line, and a constant pull-up provided by an always-on p-channel transistor having the gate tied to ground. Figure 5.7 represents an horizontal line and its connections.

Each horizontal line realizes a NOR gate, which is then inverted to become an AND. The inputs have to be obviously inverted as well, but this comes with no cost, since each input enters in the matrix with its positive and negated value and it's therefore only necessary to swap the two.

Horizontal lines form groups of 8 elements which connect to a LB, forming a section. Power and ground lines are routed along horizontally for each section, which has in addition a couple of feedback lines coming from the LB. Figure 5.1 depicts this structure.

Each horizontal line has a wiring capacitance of about 400 fF. The delay of the wired AND is directly related to the value of this capacitance and to the strength of the pull-up and pull-down drivers. For this reason the drivers should be strong, but, on the other hand, the pull-up cannot be strong since it is always on and it leads to static power consumption when the logic value on the line is low. On top of that, the size of the pull-down has to be proportional to the size of the pull-up, in order to keep the logic low output level within an acceptable limit.

The size of the pull-up is therefore a trade-off between speed and power consumption. As an example, with a constant pull-up of  $100\mu\text{A}$  and minimum size pull-down ELT transistors, the switching transient response looks like the one in Figure 5.8(a), with a propagation delay of about 10 ns. This result is obtained by simulation, which accounts also for the drain capacitance of the transistors connected to the line whose fuse is not burnt. With these conditions, the chip overall power consumption would be about 16 mW for a very slow device.

#### *Secondary pull-up*

For this reason, each horizontal line has a second pull-up transistor, stronger than the primary, which is activated only when the inputs of the logic change. In this way the primary pull-up can be made weaker, decreasing the static power consumption. At the same time the performance can be increased, since it is related to the secondary pull-up only. The secondary pull-up is then designed to deliver a current of  $500\mu\text{A}$  when activated, while the primary pull-up gives a constant  $10\mu\text{A}$ . Figure 5.8(b) shows the transient response of the wired AND in this latter case.

With this configuration, the propagation delay becomes about 3.2 ns and the static power consumption is 1.6 mW. Nevertheless, the dynamic power consumption is increased, since at every time the inputs change the secondary pull-up is activated, potentially wasting energy. In fact, the secondary pull-up is activated regardless of the kind of transition on the inputs, therefore also when not needed.

#### *SEU/SET considerations*

The horizontal lines have a high capacitance which should be enough to resist SETs coming from particles with an LET below  $25\text{ cm}^2\text{MeV/mg}$ , which is more than sufficient in the foreseen application (neutron and proton environment). Each horizontal line feeds two inverters which generate two redundant copies of the same wired AND value. The signals run duplicated from these inverters to the outputs.



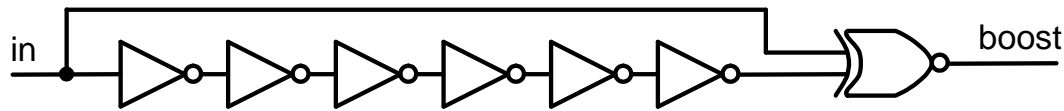
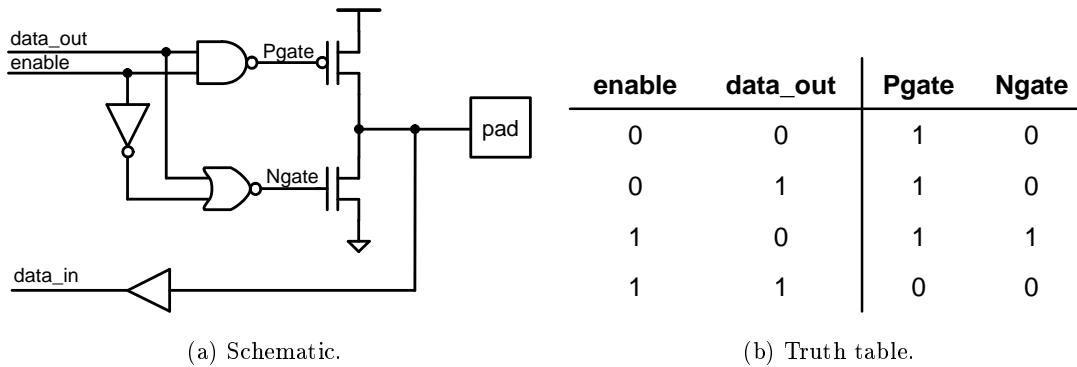


Figure 5.9: Transition detector schematic. The inverters are weak fat for increased delay.



(a) Schematic.

(b) Truth table.

Figure 5.10: Simple tri-state pad with no slew-rate control and ESD protection.

#### 5.1.4 The transition detector

In order to activate the secondary pull-up when the inputs change, a transition detector block is necessary. A transition detector is quite simple to implement: it is composed by a delay line of inverters and an XNOR gate. The XNOR gate receives the input and a delayed copy of the input. The XNOR compares the two signals and gives an high output when they are equal, while low when they differ. The result is a signal which is normally low and becomes high only when the input changes. The signal remains high for an amount of time equal to the delay of the delay line, and then returns to zero.

In order to have a delay line which takes only a small area and has a low power consumption, the inverters employed as delay elements are weak and fat, which means that their transistors have a long and narrow gate. This increases the gate capacitance and decreases the driving strength, which in turn give a longer delay. The total delay of the line is about 3.3 ns.

Each input line and each feedback line coming from the LBs needs a transition detector. All the outputs of these detectors have to be ORed to form a single detect signal which is then inverted and sent vertically throughout the whole AND matrix to all the secondary pull-ups. In other words, whenever an input changes, all the secondary pull-ups are activated.

#### 5.1.5 Tri-state I/O pad design

Since the standard I/O cell library did not have a tri-state I/O pad, the design focused also on its development. The specifications expect an Electro-Static Discharge (ESD) protected, slew-rate controlled, 20 mA output current, input/output tri-state pad.

A tri-state pad can be implemented in several ways; the choice in this work was to control separately the pull-up and the pull-down of the output inverter. Figure 5.10(a) represents a simple tri-state buffer with no slew-rate control, implemented

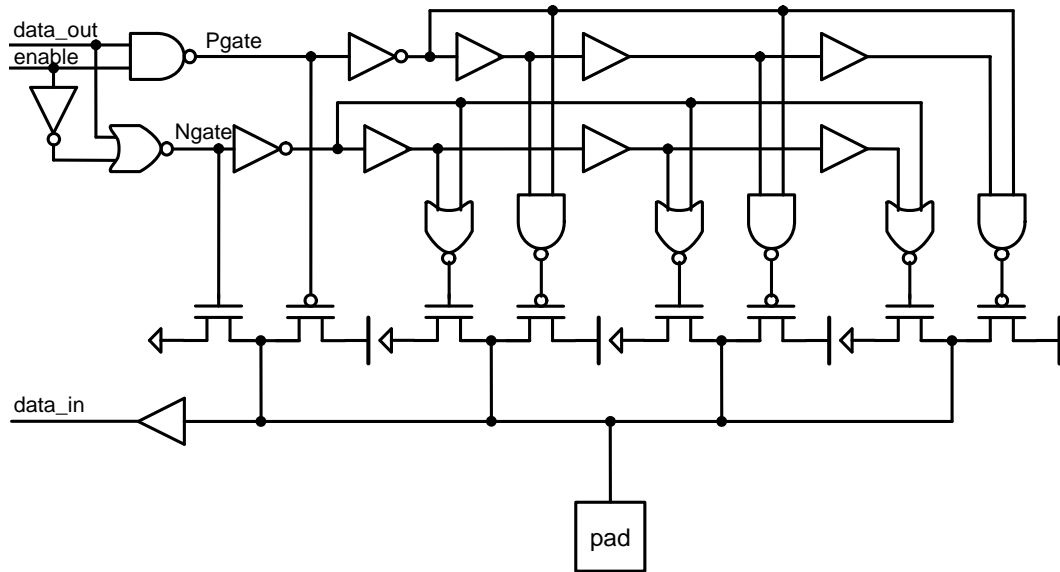


Figure 5.11: 4-stages tri-state pad buffer with slew-rate control.

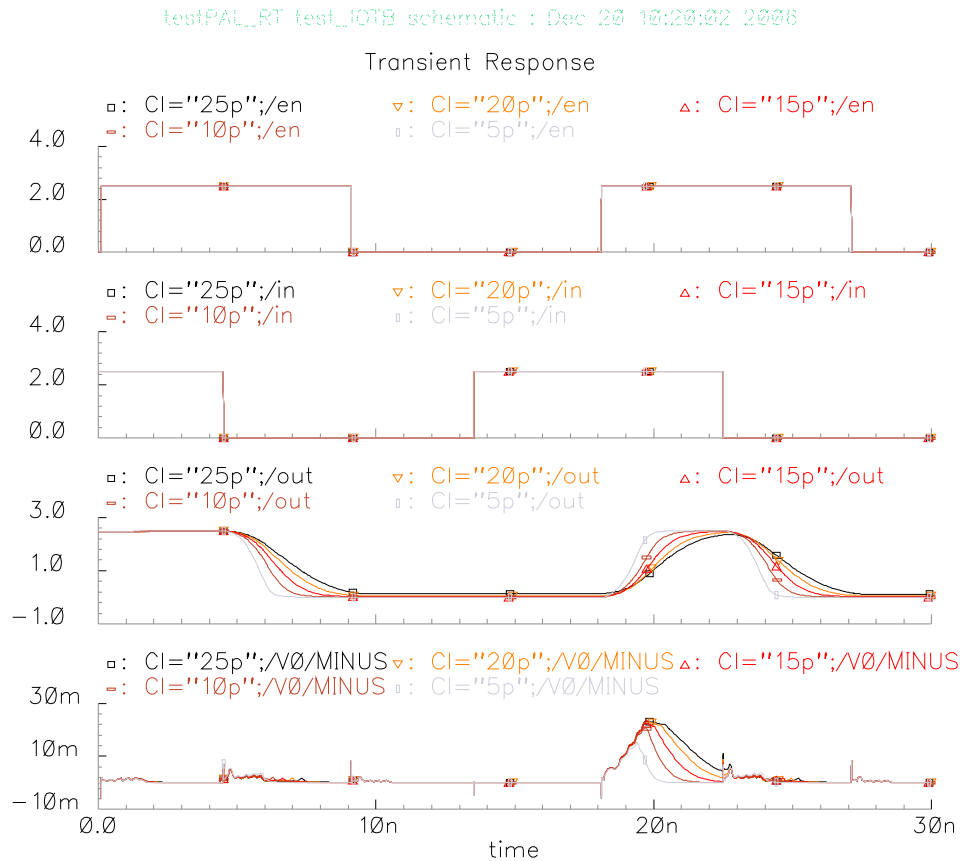


Figure 5.12: Simulation of tri-state slew rate controlled output buffer transient response. The output capacitance is swept between 5 pF and 25 pF. The plots are, respectively from top to bottom, the enable signal, the data output signal, the pad voltage and the pad current. The output current raises with a slew rate of about 10 mA/ns, reaching a maximum of 20 mA.

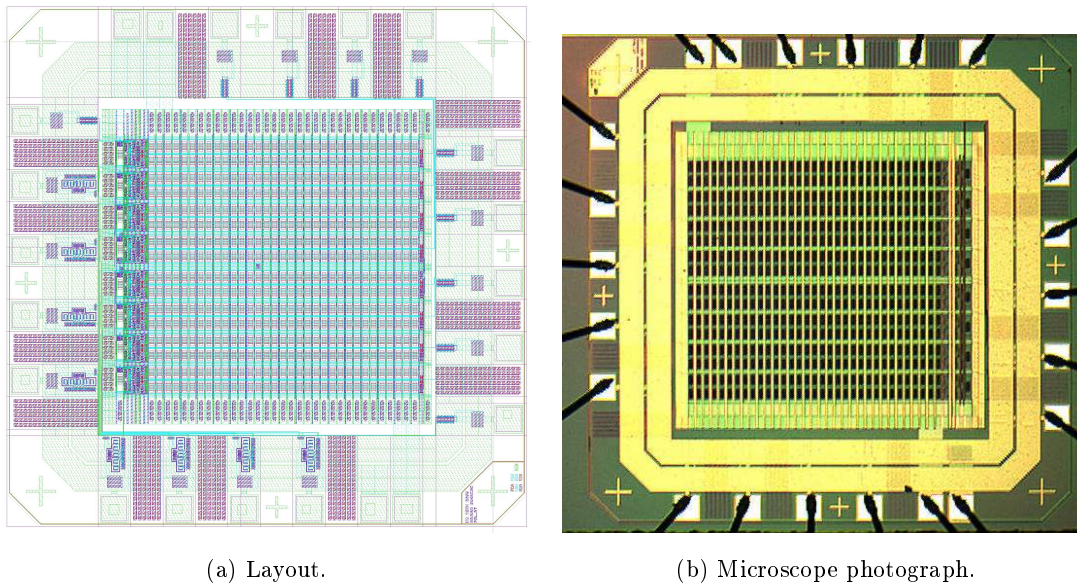


Figure 5.13: PLD chip.

with the chosen technique. The two transistor in the final output inverter can be in three possible states: both off, pulling-up or pulling-down. The NAND and the NOR, visible in the schematic, take care of generating the correct signals for the final inverter transistors, according to the truth table in Figure 5.10(b). The NAND and the NOR have to be sized in order to be able to drive the final inverter and their outputs might need to be buffered.

The slew-rate control is introduced by splitting the final inverter into several parallel inverters. When the input data changes all the inverters have to be turned off together and then turned on in sequence, each one after some delay in respect to the other. In this way the current circulating on the pad and on the power supply changes slowly favouring a smaller  $Ldi/dt$  voltage drop due to parasitic inductance, avoiding switching noise.

*Slew-rate control*

The structure in Figure 5.10(a) is therefore used again to generate the control of the pull-up and the pull-down, but then these two control signals enter two respective delay chains composed by weak fat buffers (about 300 ps delay each). Figure 5.11 shows the schematic of a 4 stages slew-rate-controlled tri-state I/O buffer.

The NANDs and NORs connected directly to the final stage are driven with a delayed and a non-delayed copy of the control signal for their final stage. This makes sure that '1's prevail for the pull-ups and '0's prevail for the pull-downs, in such a way that the transistors are turned off immediately after a change in the input. This is necessary in order to avoid any conflict among the different stages.

The pad buffer designed in the present work has 5 stages, each one capable of delivering 4 mA. Figure 5.12 represents a simulation of the tri-state output buffer and shows that the slew rate is about 10 mA/ns.

Clamping diodes are provided for ESD protection. In addition, the active areas connected directly to the pad are surrounded by a double guard ring. The inner guard ring is a separate n-well together with its n-well contact connected to the power supply, while the outer guard ring is a substrate contact connected to ground.

*ESD protection*

*Input buffer*

The input buffer consists simply in a buffer connected to the pad. The input buffer is always active and has no control signals.

**5.1.6 Chip layout**

The chip size is  $2 \times 2 \text{ mm}^2$ , while the core size is approximately  $950 \times 1150 \mu\text{m}^2$  and the rest of the area is taken by I/O pads and power routing. Figure 5.13(a) depicts the layout of the chip while Figure 5.13(b) shows its microscope picture.

Several p-channel transistors used as capacitors are employed for decoupling of the power supply, for a total of about 90 pF. These capacitors are placed under the power rails where there is no other active area.

The chip has 10 input pads, 8 input/output pads and 4 power supply and ground pads, for a total of 22 pads. The pads are distributed evenly across the perimeter of the chip, in order to allow an easy wire bonding. The two power supply and ground couples are laid at two opposite sides of the chip. The I/O unused area is also exploited for the layout of cross markers used as a reference for the spatial calibration of the fuse laser burning tool.

## Chapter 6

# Conclusions

This work demonstrates the feasibility of the design of SEU-tolerant radiation-hard PLD and an FPGA devices. The complete PLD device was fabricated and will soon undergo functional and radiation testing. The design of the LB for the FPGA device in the CMOS 0.13  $\mu\text{m}$  process is now finalized and work is ongoing to complete the interconnection infrastructure.

In order to reach the desired specifications, several SEU- and TID-hardening techniques were evaluated, and a final approach was chosen and implemented in several test chips for assessment.

An SEU-robust register structure was designed and tested in a CMOS 0.25  $\mu\text{m}$  technology as well as in a CMOS 0.13  $\mu\text{m}$  technology. The SEU-robust register is tailored in order to be used as a memory element in the design of programmable logic circuits.

The irradiation test results obtained in the CMOS 0.25  $\mu\text{m}$  technology demonstrate good robustness of the circuit up to an LET of 79.6  $\text{cm}^2\text{MeV}/\text{mg}$ , which make it suitable for the target environment.

The CMOS 0.13  $\mu\text{m}$  circuit instead showed robustness up to an LET of 37.4  $\text{cm}^2\text{MeV}/\text{mg}$  in the static test mode but had increased sensitivity in the dynamic test mode. The SEU-tolerance of the 0.13  $\mu\text{m}$  register is sufficient for the implementation of a configuration register but not for a user register, therefore additional strengthening work is necessary for the latter purpose.

A TID assessment of both programmable logic structures is also foreseen in the short period.

Future plans of the project include the development of software capable of generating a programming bit-stream for the FPGA and PLD components.



# Appendix A

## Memory cell layout for SEU-robustness

An efficient way to place the nodes of a latch in layout is to lay them out along a line and not in the usual stacked position which has the n-well running all along the cell.

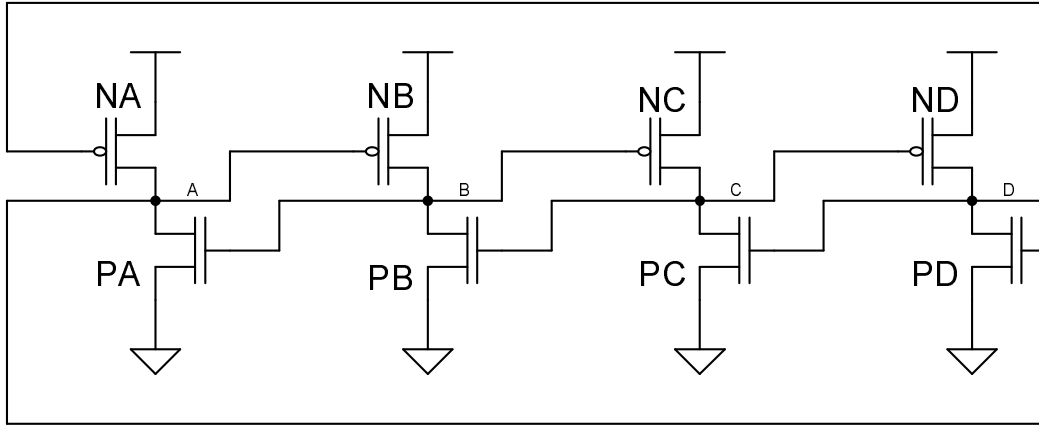


Figure A.1: DICE cell.

If we consider figure A.1, which represents a DICE cell, studying the possible particle hits on the transistors, it is possible to see that some multiple particle hits are allowed and do not cause an upset. For instance, if the drain of PB is hit while the cell is in state  $(A, B, C, D) = (1, 0, 1, 0)$ , node B will collect charge and go to 1. In this situation, a second hit on transistors PA, NA, NB, PC, ND won't cause any upset since it won't affect the nodes which are not already affected. Table A.1 resumes all the allowed multiple hits on transistors of the DICE.

From the table it is possible to infer that an optimal placement can be obtained by maximizing the distance of the couples  $(PA, NB)$ ,  $(PB, NC)$ ,  $(PC, ND)$ ,  $(PD, NA)$

when $(A, B, C, D) = (0, 1, 0, 1)$	when $(A, B, C, D) = (1, 0, 1, 0)$
PA, PD, ND, NA, PB, NC	PB, PA, NA, NB, PC, ND
PC, PB, NB, NC, PD, NA	PD, PC, NC, ND, PA, NB

Table A.1: Allowed multiple hits on the DICE without generation of upset.

— meaning that the first member of each couple should be far away from the second member.

On the other hand, the couples (PA,ND), (PB,NA), (PC,NB), (PD,NC) and (PA,NA), (PB,NB), (PC,NC), (PD,ND) can have a small distance, since they are always present in the table.



Figure A.2: DICE optimal layout.

It follows that an optimal placement would be the one in figure A.2, which takes into consideration the mentioned constraints and puts close together p-type transistors couples in order to form a common n-well.



# Bibliography

- [Actel 97] Actel Corporation. *Design Techniques for Radiation-Hardened FPGAs*, September 1997. Application note. 75
- [Actel 04] Actel Corporation. *RTAX-S RadTolerant FPGA*, May 2004. Data sheet v0.5. 74, 75
- [Admrel 02] Architectures and Methodologies for Dynamic Reconfigurable Logic (ADMREL) - Information Societies Technology (IST) Program. *Survey of existing fine-grain reconfigurable hardware platforms*, November 2002. v2.0. 70
- [Anelli 97] G. Anelli *et al.* *Total dose behavior of submicron and deep submicron CMOS technologies*. In 3th Workshop on electronics for LHC experiments, London, September 1997. 40
- [Anelli 00] G. Anelli. *Design and characterization of radiation tolerant integrated circuits in deep submicron CMOS technologies for the LHC experiments*. PhD thesis, Insitut National Polytechnique de Grenoble, France, December 2000. 5, 37, 39, 41, 42
- [Baumann 04] R.C. Baumann. *Soft Errors in Commercial Integrated Circuits*. International Journal of High Speed Electronics and Systems, vol. 14, no. 2, pages 299–309, 2004. 45
- [Baze 97] M.P. Baze. *Attenuation of Single Event Induced Pulses in CMOS Combinational Logic*. IEEE Transactions on Nuclear Science, vol. 44, no. 6, pages 2217–2223, December 1997. 56
- [Berger 96] G. Berger, G. Ryckewaert, R. Harboe-Sorensen & L. Adams. *The heavy ion irradiation facility at CYCLONE - a dedicated SEE beam line*. In Radiation Effects Data Workshop, pages 78–83. IEEE, July 1996. 100
- [Blum 05] D.R. Blum, M.J. Myjak & J. Delgado-Frias. *Enhanced Fault-Tolerant Data Latches for Deep Submicron CMOS*. In Proceedings of the 2005 International Conference on Computer Design (CDES), pages 28–34, Las Vegas, Nevada, USA, June 2005. CSREA Press. 54
- [Boesch 85] H.E.Jr. Boesch & F.B. McLean. *Hole transport and trapping in field oxides*. IEEE Transactions on Nuclear Science, vol. 32, no. 6, December 1985. 36

- [Boesch 86] H.E.Jr. Boesch *et al.* *Saturation of threshold voltage shift in MOSFETs at high total dose.* IEEE Transactions on Nuclear Science, vol. 33, no. 6, pages 1191–1197, December 1986. 37
- [Bonacini 03] S. Bonacini. Design of two digital radiation tolerant integrated circuits for high energy physics experiments data readout. Master's thesis, Università di Modena e Reggio Emilia, March 2003. 58, 63
- [Bonacini 06] S. Bonacini, F. Faccio, K. Kloukinas & A. Marchioro. *An SEU-robust Configurable Logic Block for the Implementation of a Radiation-Tolerant FPGA.* IEEE Transactions on Nuclear Science, no. 6, December 2006. 137
- [Bonacini 07] S. Bonacini, K. Kloukinas & A. Marchioro. *Development of SEU-robust, radiation-tolerant and industry-compatible programmable logic components.* Journal of Instrumentation (JINST), vol. 2, September 2007. 137
- [Braunig 93] D. Braunig. *Ionization and Displacement.* In Notes of the Short Course of the 2nd European Conference on Radiation and its Effects on Components and Systems, number 2 in , Saint-Malo, France, September 1993. 35
- [Buchner 97] S. Buchner, M. Baze, D. Brown, D. McMorrow & J. Melinger. *Comparison of Error Rates in Combinational and Sequential Logic.* IEEE Transactions on Nuclear Science, vol. 44, no. 6, pages 2209–2216, December 1997. 7, 48
- [Calin 96] T. Calin, M. Nicolaidis & R. Velazco. *Upset Hardened Memory Design for Submicron CMOS Technology.* IEEE Transactions on Nuclear Science, vol. 43, no. 6, pages 2874–2878, December 1996. 7, 50
- [Cellere 04] G. Cellere & A. Paccagnella. *A review of ionizing radiation effects in Floating-Gate memories.* IEEE Transactions on Device and Materials Reliability, vol. 4, no. 3, pages 359–370, September 2004. 75
- [Clark 81] G.C. Clark & J.B. Cain. Error-correction coding for digital communications. Plenum, New York, 1981. ISBN:0-306-40615-2. 9, 61
- [CMS 94] CMS. *The Compact Muon Solenoid.* Technical proposal, CERN, December 1994. CERN/LHCC/94-38. 30
- [CMS 97] CMS. *The Electromagnetic Calorimeter Project.* Technical Design Report 4, CERN/CMS, December 1997. CERN/LHCC/97-33. 31
- [Degalahal 04] V. Degalahal, R. Rajaram, N. Vijaykrishan, Y. Xie & M.J. Irwin. *The effect of threshold voltages on soft error rate.* In 5th International Symposium on Quality Electronic Design, Pennsylvania, March 2004. EMC<sup>2</sup>. 49

- [Dingwall 77] A.G.F. Dingwall & R.E. Stricker. *C<sup>2</sup>L: A new high-speed high-density bulk CMOS technology*. IEEE Journal of Solid-State Circuits, vol. 12, August 1977. 41
- [Dooley 94] J.G. Dooley. *SEU-immune latch for gate array, standard cell, and other ASIC applications*. U.S. Patent No. 5311070, May 1994. 54
- [Faccio 98] F. Faccio, G. Anelli, M. Campbell, M. Delmastro, P. Jarron, K. Kloukinas, A. Marchioro, T. Calin, J. Cosculluela, R. Velazco, M. Nicolaidis & A. Giraldo. *Total dose and Single Event Effects (SEE) in a 0.25  $\mu$ m CMOS technology*. In 4th Workshop on electronics for LHC experiments, Roma, September 1998. Università di Roma “La Sapienza”. 5, 38, 39
- [Faccio 99] F. Faccio, K. Kloukinas, A. Marchioro, T. Calin, J. Cosculluela, M. Nicolaidis & R. Velazco. *Single Event Effects in Static and Dynamic Registers in a 0.25 $\mu$ m CMOS Technology*. IEEE Transactions on Nuclear Science, vol. 46, no. 6, pages 1434–1439, December 1999. 102
- [Faccio 04] F. Faccio. *Radiation Issues in the New Generation of High Energy Physics Experiments*. International Journal of High Speed Electronics and Systems, vol. 14, no. 2, pages 379–399, 2004. 47
- [Faccio 05] F. Faccio & G. Cervelli. *Radiation-induced edge effects in deep submicron CMOS transistors*. IEEE Transactions on Nuclear Science, vol. 52, no. 6, pages 2413–2420, December 2005. 5, 38, 40
- [Fuja 88] T. Fuja, C. Heegard & R. Goodman. *Linear Sum Codes for Random Access Memories*. IEEE Transactions on Computers, vol. 37, no. 9, pages 1030–1042, September 1988. 63
- [Gagliardi 03] G. Gagliardi. *Measurement of SEU on the Module Controller Chip of the ATLAS Pixel Detector*. In 5th International Meeting on Front-End Electronics, Snowmass, Colorado, USA, June 2003. 82
- [Gambles 03] J. Gambles, L. Miles, J. Hass, W. Smith & S. Whitaker. *An Ultra-Low-Power, Radiation-Tolerant Reed Solomon Encoder for Space Applications*. In Custom Integrated Circuits Conference, pages 631–634, San Jose, California, September 2003. IEEE. 54
- [Giraldo 98] A. Giraldo. *Evaluation of Deep Submicron Technologies with Radiation Tolerant Layout for Electronics in the LHC Environments*. PhD thesis, University of Padova, Italy, December 1998. URL: <http://wwwcdf.pd.infn.it/cdf/sirad/giraldo/tesigiraldo.html>. 32, 41
- [Hagiwara 02] K. Hagiwara *et al.* Physical Review D, volume 66 of III, chapter 31-Statistics, pages 010001–229. The American Physical Society, July 2002. 101

- [Hamming 50] R.W. Hamming. *Error detecting and error correcting codes*. Bell System Technical Journal, vol. 29, pages 147–160, 1950. 9, 61
- [Hass 98] K.J. Hass, J.W. Gambles, B. Walker & M. Zampaglione. *Mitigating Single Event Upsets From Combinational Logic*. In 7th NASA VLSI Design Symposium, Albuquerque, New Mexico, October 1998. 56
- [Hass 03] K.J. Hass & J.R. Piepmeier. *An Ultra-Low Power, Radiation Tolerant, High Speed Correlator*. In 11th NASA VLSI Design Symposium, Coeur D’Alene, Idaho, May 2003. University of Idaho. 56, 61
- [Hazucha 04] P. Hazucha, T Karnik, S. Walstra, B.A. Bloechel, J.W. Tschanz, J. Maiz, K. Soumyanath, G.E. Dermer, S. Narendra, V. De & S. Borkar. *Measurement and Analysis of SER-Tolerant Latch in a 90-nm Dual-V<sub>T</sub> CMOS Process*. IEEE Journal of Solid-State Circuits, vol. 39, no. 9, pages 1536–1543, September 2004. 51
- [Hopkins 71] A.L. Hopkins. *A Fault-Tolerant Information Processing Concept for Space Veichles*. IEEE Transactions on Computers, vol. C-20, no. 11, pages 1394–1403, November 1971. 56
- [Hsiao 70] M.Y. Hsiao. *A Class of Optimal Minimum Odd-weight-column SEC-DED Codes*. IBM Journal of Research and Development, vol. 14, pages 395–401, July 1970. 63
- [Huhtinen 97] M. Huhtinen. *Method for estimating dose rates from induced radioactivity in complicated hadron accelerator geometry*. Divisional report, CERN/TIS, 1997. 47
- [Huhtinen 00] M. Huhtinen & F. Faccio. *Computational method to estimate Single Event Upset rates in an accelerator environment*. Nuclear Instruments and Methods in Physics Research A, no. 450, pages 155–172, June 2000. 47
- [Jarron 99a] P. Jarron, G. Anelli, T. Calinet *al.* *Deep submicron CMOS technologies for the LHC experiments*. Nuclear Physics B (Proceedings Supplement), vol. 78, pages 625–634, August 1999. issues 1–3. 33
- [Jarron 99b] P. Jarron, G. Anelliet *al.* *Study of the radiation tolerance of IC’s for LHC*. RD49 status report 2, CERN/MIC, March 1999. CERN/LHCC/99-8. 33
- [Jarron 00] P. Jarron, G. Anelliet *al.* *Study of the radiation tolerance of IC’s for LHC*. RD49 status report 3, CERN/MIC, January 2000. CERN/LHCC/2000-03. 33
- [Kerns 89] S.E. Kerns & B.D. Shafer. *Ionizing Radiation Effects in MOS Devices & Circuits*, chapter Transient-Ionization and Single-Event Phenomena. J.Wiley & Sons, New York, 1989. 5, 43

- [Kloukinas 98] K. Kloukinas, F. Faccio, A. Marchioro & P. Moreira. *Development of a radiation tolerant 2.0 V standard cell library using a commercial deep submicron CMOS technology for the LHC experiments*. In 4th Workshop on electronics for LHC experiments, Roma, September 1998. Università di Roma “La Sapienza”. 5, 42
- [Kloukinas 03] K. Kloukinas, P. Aspell, D. Barney, S. Bonacini & S. Reynaud. *Kchip: A Radiation Tolerant Digital Data Concentrator chip for the CMS Preshower Detector*. In 9th Workshop on Electronics for LHC Experiments, Amsterdam, The Netherlands, October 2003. 58, 63
- [Kloukinas 05] K. Kloukinas, S. Bonacini & A. Marchioro. *Characterization and production testing of a quad 12 bit 40 Ms/sec A/D converter with automatic digital range selection for calorimetry*. In 11th Workshop on Electronics for LHC and Future Experiments, Heidelberg, September 2005. 137
- [Koga 98] R. Koga. *Single Event Functional Interrupt (SEFI) Sensitivity in EEPROMs*. In Military and Aerospace Programmable Logic Device (MAPLD) International Conference, Greenbelt, Maryland, September 1998. NASA Goddard Space Flight Center. 7, 48
- [Kumar 04] N. Kumar & D. Zacher. *Automated FSM Error Correction for Single Event Upsets*. In Military and Aerospace Programmable Logic Device (MAPLD) International Conference, Washington, D.C., September 2004. Ronald Reagan Building and International Trade Center. 63
- [Kuo 99] J.B. Kuo & J.H. Luo. *Low Voltage CMOS VLSI Circuits*. Wiley Interscience, J. Wiley and Sons, 1999. 39
- [Langley 04] T.E. Langley & P. Murray. *SEE and TID Test Results of 1Gb Flash Memories*. In IEEE Radiation Effects Data Workshop, pages 58–61, July 2004. 75
- [Larsen 72] R.W. Larsen & I.S. Reed. *Redundancy by Coding Versus Redundancy by Replication for Failure-Tolerant Sequential Circuits*. IEEE Transactions on Computers, vol. C-21, no. 2, pages 130–137, February 1972. 63
- [Liu 92] M.N. Liu & S. Whitaker. *LOW POWER SEU IMMUNE CMOS MEMORY CIRCUITS*. IEEE Transactions on Nuclear Science, vol. 39, no. 6, pages 1679–1684, December 1992. 53
- [Maki 01] G.K. Maki, J.K. Hass, Q. Shi & J. Murguia. *Conflict Free Radiation Tolerant Storage Cell*. U.S. Patent No. 6573773, February 2001. 54
- [Marchioro 98] A. Marchioro. *Deep submicron technologies for HEP*. In 4th Workshop on electronics for LHC experiments, Roma, September 1998. Università di Roma “La Sapienza”. 5, 42

- [Marple 92] D. Marple & L. Cooke. *An MPGA compatible FPGA architecture*. In ACM First International Workshop on Field-Programmable Gate Arrays, pages 39–44, Berkley, California, February 1992. 70
- [Mavis 00] D.G. Mavis & P.H. Eaton. *SEU and SET Mitigation Techniques for FPGA Circuit and Configuration Bit Storage Design*. In Military and Aerospace Programmable Logic Device (MAPLD) International Conference, Laurel, Maryland, September 2000. The Johns Hopkins Univerisity - Applied Physics Laboratory. 56, 58
- [McLean 89] F.B. McLean, H.E.Jr. Boesch & T.R. Oldham. *Ionizing Radiation Effects in MOS Devices & Circuits*, chapter Electron-Hole generation, transport and trapping in SiO<sub>2</sub>. J.Wiley & Sons, New York, 1989. 36
- [McWhorter 90] P.J. McWhorter, S.L. Miller & W.M. Miller. *Modeling the anneal of radiation-induced trapped holes in a varying thermal environment*. IEEE Transactions on Nuclear Science, vol. 37, no. 6, pages 1682–1688, December 1990. 37
- [Messenger 97] G.C. Messenger & M.S. Ash. *Single event phenomena*. Chapman & Hall, New York, 1997. ISBN: 0-412-09731-1. 46, 47
- [Meyer 71] J.F. Meyer. *Fault-Tolerant Sequential Machines*. IEEE Transactions on Computers, vol. C-20, no. 10, pages 1167–1177, October 1971. 63
- [Nguyen 99] D.N. Nguyen, S.M. Guertin, G.M. Swift & A.H. Johnston. *Radiation Effects on Advanced Flash Memories*. IEEE Transactions on Nuclear Science, vol. 46, no. 6, pages 1744–1750, December 1999. 75
- [Niranjan 96] S. Niranjan & J.F. Frenzel. *A Comparison of Fault-Tolerant State Machines Architectures for Space-Borne Electronics*. IEEE Transactions on Reliability, vol. 45, no. 1, pages 109–113, March 1996. 63
- [Reed 96] R.A. Reed, M.A. Carls, P.W. Marshall, C.J. Marshall, S. Buchner, M. LaMacchia, B. Mathes & D. McMorrow. *Single Event Upset Cross Sections at Various Data Rates*. IEEE Transactions on Nuclear Science, vol. 43, no. 6, pages 2862–2867, December 1996. 7, 48
- [Roche 05] P. Roche, F. Jacquet, G. Gasiot, C. Caillat, B. Borot & J.P. Schoellkopf. *High Density SRAM robust to radiation-induced Soft Errors in 90nm CMOS technology*. In 1st International Conference on Memory Technology and Design, Giens, May 2005. Provence Materials and Microelectronics Laboratory (L2MP). 49
- [Rockett 88] L. Rockett. *An SEU Hardened CMOS Data Latch Design*. IEEE Transactions on Nuclear Science, vol. 35, no. 6, pages 1682–1687, December 1988. 54

- [Rose 93] J. Rose, A. El Gamal & A. Sangiovanni-Vincetelli. *Architecture of Field-Programmable Gate Arrays*. Proceedings of the IEEE, vol. 81, no. 7, pages 1013–1028, July 1993. 70, 71
- [Shuler 05] R.L. Shuler, C. Kouba & P.M. O'Neill. *SEU Performance of TAG Based Flip-Flops*. IEEE Transactions on Nuclear Science, vol. 52, no. 6, pages 2550–2553, December 2005. 64
- [Snoeys 00] W. Snoeys, G. Anelli, M. Campbellet *al.* *Integrated circuits for particle physics experiments*. IEEE Journal of Solid-State Circuits, vol. 35, no. 12, December 2000. 33
- [Speers 99] T. Speers, J.J. Wang, B. Cronquist, J. McCollum, H. Tseng, R. Katz & I. Kleyner. *0.25  $\mu\text{m}$  FLASH Memory Based FPGA for Space Applications*. In Military and Aerospace Programmable Logic Device (MAPLD) International Conference, Laurel, Maryland, September 1999. The Johns Hopkins University - Applied Physics Laboratory. 75
- [Swift 95] G. Swift & R. Katz. *An Experimental Survey of Heavy Ion Induced Dielectric Rupture in Actel Field Programmable Gate Arrays (FPGAs)*. In Radiation and its Effect on Components and Systems (RADECS), pages 425–430, September 1995. 75
- [Velazco 96] R. Velazco, T. Calin, M. Nicolaidis, S.C. Moss, S.D. La Lumondiere, V.T. Tran & R. Koga. *SEU-Hardened Storage Cell Validation Using a Pulsed Laser*. IEEE Transactions on Nuclear Science, vol. 43, no. 6, December 1996. 51
- [Von Neumann 56] J. Von Neumann. *Probabilistic logics and the synthesis of reliable organisms from unreliable components*. In Automata Studies, number 34 in Annals of Mathematical Studies, pages 43–98. Princeton University Press, Princeton, New Jersey, 1956. 8, 56
- [Wang 03a] J.J. Wang. *Radiation Effects in FPGAs*. In 9th Workshop on Electronics for LHC Experiments, Amsterdam, The Netherlands, October 2003. 74, 75
- [Wang 03b] J.J. Wang, W. Wong, S. Wolday, B. Cronquist, J. McCollum, R. Katz & I. Kleyner. *Single Event Upset and Hardening in 0.15  $\mu\text{m}$  Antifuse-Based Field Programmable Gate Array*. IEEE Transactions on Nuclear Science, vol. 50, no. 6, pages 2158–2166, December 2003. 13, 75
- [Wang 04] W. Wang & H. Gong. *Edge Triggered Pulse Latch Design With Delayed Latching Edge for Radiation Hardened Application*. IEEE Transactions on Nuclear Science, vol. 51, no. 6, pages 3626–3630, December 2004. 7, 48, 56
- [Whitaker 91] S. Whitaker, J. Canaris & K. Liu. *SEU Hardened Memory Cells for a CCSDS Reed Solomon Encoder*. IEEE Transactions on Nuclear Science, vol. 38, no. 6, pages 1471–1477, December 1991. 52

- [Winokur 89] P.S. Winokure *et al.* Ionizing Radiation Effects in MOS Devices & Circuits, chapter Radiation-Induced Interface Traps. J.Wiley & Sons, New York, 1989. 37
- [Xilinx 00] Xilinx Inc. *Correcting Single-Event Upsets Through Virtex Partial Configuration*, June 2000. Application note v1.0. 13, 76
- [Xilinx 01] Xilinx Inc. *Triple Module Redundancy Design Techniques for Virtex FPGAs*, November 2001. Application note v1.0. 75
- [Xilinx 04] Xilinx Inc. *QPro Virtex-II 1.5V Radiation Hardened QML Platform FPGAs*, January 2004. Data sheet v1.5. 75
- [Yui 03] C.C. Yui, G.M. Swift, C. Charmichael, R. Koga & J.S. George. *SEU Mitigation Testing of Xilinx Virtex II FPGAs*. In Radiation Effects Data Workshop, pages 92–97. IEEE, July 2003. 13, 76



# List of publications

The work carried out for the present thesis brought to the following publications [Kloukinas 05, Bonacini 06, Bonacini 07]:

S. Bonacini, F. Faccio, K. Kloukinas & A. Marchioro. *An SEU-robust Configurable Logic Block for the Implementation of a Radiation-Tolerant FPGA*. IEEE Transactions on Nuclear Science, no. 6, December 2006.

S. Bonacini, K. Kloukinas & A. Marchioro. *Development of SEU-robust, radiation-tolerant and industry-compatible programmable logic components*. Journal of Instrumentation (JINST), vol. 2, September 2007.

K. Kloukinas, S. Bonacini & A. Marchioro. *Characterization and production testing of a quad 12 bit 40 Ms/sec A/D converter with automatic digital range selection for calorimetry*. In 11th Workshop on Electronics for LHC and Future Experiments, Heidelberg, September 2005.

In addition, the content of the first two articles was presented, respectively, at the IEEE Nuclear and Space Radiation Effects Conference (NSREC) 2006, in Ponte Vedra Beach, Florida, USA, and at the Topical Workshop on Electronics for Particle Physics (TWEPP) 2007, in Prague, Czech Republic.

---

**TITRE en français**

Développement de circuits logiques programmables résistants aux aléas logiques en technologie CMOS submicrométrique

---

**RÉSUMÉ en français**

L'électronique associée aux détecteurs de particules du grand collisionneur d'hadrons (LHC), en construction au CERN, fonctionnera dans un environnement très radioactif. La plupart des composants microélectroniques développés pour la première génération des expériences du LHC ont été conçus avec des buts spécifiques et très précis, non adaptables pour d'autres applications. Les composants commerciaux ne peuvent pas être employés en proximité du point de collision des particules, car ils ne sont pas tolérants aux radiations. Cette thèse contribue à couvrir le besoin en composants programmables résistants aux rayonnements et aux aléas logiques pour les expériences de physique des hautes énergies. Dans ce sens, deux composants sont en cours de développement : un dispositif logique programmable (PLD) et un réseau de portes programmables in-situ (FPGA). Le PLD est configuré par fusibles et il possède 10 entrées et 8 I/O. Le PLD est fabriqué dans une technologie CMOS 0.25  $\mu\text{m}$ . Le FPGA est composé d'une rangée de  $32 \times 32$  blocs logiques, ce qui équivaut approximativement à 25k portes, et il a été conçu dans une technologie CMOS 0.13  $\mu\text{m}$ . Ce travail s'est concentré également sur la recherche d'un registre résistant aux aléas logiques dans les deux technologies mentionnées. Le registre est utilisé comme bascule pour les données d'utilisateur dans le FPGA et le PLD, mais aussi comme cellule de configuration dans le FPGA.

---

**TITRE en anglais**

Development of Single-Event Upset hardened programmable logic devices in deep submicron CMOS

---

**RÉSUMÉ en anglais**

The electronics associated to the particle detectors of the Large Hadron Collider (LHC), under construction at CERN, will operate in a very harsh radiation environment. Most of the microelectronics components developed for the first generation of LHC experiments have been designed with very precise experiment-specific goals and are hardly adaptable to other applications. Commercial Off-The-Shelf (COTS) components cannot be used in the vicinity of particle collision due to their poor radiation tolerance. This thesis is a contribution to the effort to cover the need for radiation-tolerant SEU-robust programmable components for application in High Energy Physics (HEP) experiments. Two components are under development: a Programmable Logic Device (PLD) and a Field-Programmable Gate Array (FPGA). The PLD is a fuse-based, 10-input, 8-I/O general architecture device in 0.25  $\mu\text{m}$  CMOS technology. The FPGA under development is instead a  $32 \times 32$  logic block array, equivalent to  $\approx 25\text{k}$  gates, in 0.13  $\mu\text{m}$  CMOS. This work focussed also on the research for an SEU-robust register in both the mentioned technologies. The SEU-robust register is employed as a user data flip-flop in the FPGA and PLD designs and as a configuration cell as well in the FPGA design.

---

**SPÉCIALITÉ** : Micro et Nano électronique

---

**MOTS-CLÉS** : Circuits intégrés, effets des radiations, aléa logique, SEU, circuits programmables, FPGA, PLD, Large Hadron Collider.

---

**INTITULE ET ADRESSE DU LABORATOIRE DE RATTACHEMENT** :

CERN, Laboratoire Européen pour la Recherche Nucléaire, CH-1211 Genève 23, Suisse.

---



ISBN : 978-2-84813-111-5 9 782848 131115 >