Étude transitoire du déclenchement de protections haute tension contre les décharges électrostatiques

Antoine Delmas

▶ **To cite this version:**

Antoine Delmas. Étude transitoire du déclenchement de protections haute tension contre les décharges électrostatiques. Micro et nanotechnologies/Microélectronique. Université Paul Sabatier - Toulouse III, 2012. Français. NNT : . tel-00720123

HAL Id: tel-00720123

https://theses.hal.science/tel-00720123

Submitted on 23 Jul 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par :
Université Toulouse III Paul Sabatier (UT3 Paul Sabatier)

Discipline ou spécialité :

Électronique

Présentée et soutenue par :
Antoine Delmas

le : 27 Février 2012

Titre :

Étude transitoire du déclenchement de protections haute tension contre les décharges électrostatiques

---

Ecole doctorale :
Génie Electrique, Electronique et Télécommunications (GEET)

Unité de recherche :
LAAS-CNRS

Directeur(s) de Thèse :
Nicolas Nolhier, Professeur à l'Université Paul Sabatier, Toulouse
Marise Bafleur, Directrice de recherche, LAAS-CNRS

Rapporteurs :

M. Dionyz Pogany, Professeur, Université de Technologie de Vienne
Philippe Benech, Professeur, IMEP, Grenoble

Membre(s) du jury :

Alain Reinex, Directeur de recherche CNRS, XLIM, Limoge
Patrice Besse, expert ESD, Freescale Semiconductor
Patrick Austin, Professeur à l'Université Paul Sabatier, Toulouse

Remerciements

Le travail de thèse présenté dans ce mémoire a été réalisé dans le cadre d'une convention CIFRE (Convention Industrielle de Formation par la REcherche) établie entre le LAAS-CNRS et la société Freescale Semiconducteurs SAS. Au LAAS, j'ai rejoins l'équipe de l'Intégration des Systèmes de Gestion de l'Energie (ISGE) et à Freescale la division Technology Solutions Organization (TSO), puis le laboratoire du groupe Analog & Mixed Signal Product (AMPD). Ma thèse a été initiée dans le cadre du laboratoire commun « Laboratoire d'Intégration des systèmes de Puissance Avancée » conclu entre le groupe TSO de Freescale et l'équipe ISGE du LAAS de 2006 à 2009.

A ce titre, je remercie Raja Chatila, Jean-Louis Sanchez ainsi que Jean Arlat, directeurs successifs du LAAS de m'avoir accueilli dans leur laboratoire. J'ai une pensée particulière pour Jean-Louis Sanchez, qui nous a quitté en mai dernier. Je remercie également Marise Bafleur en tant que directrice du groupe ISGE, et Jean-Louis Chaptal, ancien manager de l'équipe R&D de Freescale Toulouse.

Je tiens à remercier Dionyz Pogany et Philippe Benech d'avoir accepté de lire et évaluer mon manuscrit de thèse. Merci également aux examinateurs Alain Reinex et Patrice Besse. Merci à Patrick Austin pour son rôle d'examinateur et pour avoir accepté de présider le jury.

Je remercie chaleureusement mon directeur de thèse Nicolas Nolhier pour son soutien dans les moments de doute et l'autonomie qu'il a su me laisser dans la prise en main de ma thèse. Merci encore à ma co-directrice de thèse Marise Bafleur.

Je remercie Philippe Renaud, maître de stage puis encadrant industriel de ma thèse durant les premiers mois, avant d'être appelé à de plus hautes fonction. Je retiendrai sa verve et son éloquence. Je suis fier de lui avoir fait connaître l'existence du « air guitar ». Un très grand merci à Amaury Gendron, ancien doctorant de Philippe, qui a pris le relai de l'encadrement de ma thèse depuis le milieu de l'Arizona. Ses grandes qualités tant professionnelles qu'humaines m'ont été d'un précieux support. J'ai toutefois dû mettre de côté l'admiration qu'avait suscité en moi la lecture de sa thèse pour pouvoir collaborer parfaitement. Merci à Chai Gill, manager de l'équipe ESD de Tempe pour la confiance qu'elle m'a accordée en m'acceptant dans son équipe dès la fin 2008.

Ma reconnaissance va particulièrement à David Trémouilles, véritable inventeur de la méthode « transient-TLP », dont la curiosité, la disponibilité et la clairvoyance ont été déterminants pour ma thèse. Je retiendrai les discutions passionnantes que nous avons eues et les nombreuses feuilles blanches que nous avons noircies de gribouillis indéchiffrables.

Je remercie Philippe Debosque de m'avoir accueilli dans son équipe à la fin de mon contrat

CIFRE pour installer un banc de mesure « transient-TLP » au laboratoire, et à Oscar De Barros pour voir prolongé mon CDD. Un grand merci à Alain Salles, ingénieur instrumentation au laboratoire pour sa patience et ses conseils. Merci à Matthieu Aribaud pour les moment de détente au café, à Rémy, Mylène, Thierry, Alain C., Philippe T., Barbara, Salim, Alexis, Pierre, Sailith et tous les autres du laboratoire. Merci à Dominique pour ses photos de Chamonix de l'époque où il y avait encore des glaciers.

Merci à mes anciens voisins de bureau de TSO : Evgeniy, René, Jean-Michel, John et Bertrand. Merci à Céline pour son amitié, qui m'a aidé à tenir le coup pendant ces années. Contrairement à moi, tu es capable de parler ET travailler en même temps. Merci à Mikaël pour avoir supporté nos bavardages. Merci à Audrey, Aurélie, Nathalie, Julie et Noé pour les moments de détente à Freescale et à l'extérieur.

Merci à tous mes nouveaux collègues du design à Freescale, et en premier lieu merci à Patrice Besse, dont le rôle dans mon embauche à Freescale a été crucial.

Je remercie Sylvette Cunnac pour sa réactivité et son efficacité dans la gestion des dossier CIFRE.

Je remercie tous ceux que j'ai pu Nicolas Mauran, Alexandre, Nicolas Monnereau, Sandra, Jason, Yuan, Fabrice, Johan, Emmanuel, Benjamen, Emilie, Aude, Christophe, Laurent, Abdelakim, et tous ceux que je ne cite pas ici.

Table des matières

Remerciements **iii**

Introduction générale **xi**

1 Décharges électrostatiques et composants de protection **1**
1.1 Introduction . . . 1
1.2 Fenêtre de conception . . . 1
1.3 Modèles d'évaluation de la robustesse . . . 3
1.3.1 Robustesse des protections ESD et des circuits intégrés . . . 3
1.3.2 Méthodes de test au niveau composant : normes $JEDEC\ ESD$ . . . 4
1.3.2.1 Modèle du corps humain ou human body model (HBM) . . . 4
1.3.2.2 Modèle de machine, ou Machine model (MM) . . . 5
1.3.2.3 Modèle du composant chargé, ou Charged device model (CDM) 6
1.3.3 Normes de test au niveau système et véhicule pour l'automobile . . . 7
1.3.4 Human Metal Model (HMM) . . . 9
1.4 Outils d'investigation . . . 9
1.4.1 Mesures impulsionnelles $TLP/vf-TLP$ . . . 10
1.4.1.1 Principe de la caractérisation quasi-statique . . . 10
1.4.1.2 Fonctionnement d'un banc de mesure TLP/vf-TLP en configuration TDR . . . 11
1.4.2 CC-TLP . . . 12
1.5 Etat de l'art des composants de protection . . . 13
1.5.1 Les diodes . . . 14
1.5.2 Transistors bipolaires autopolarisés . . . 14
1.5.2.1 Modes de déclenchement par « dV/dt » . . . 16
1.5.2.2 Transistors bipolaires latéraux . . . 16
1.5.2.3 Transistors bipolaires verticaux . . . 17
1.5.3 Silicon Controlled Rectifier SCR . . . 18
1.5.4 Composants à base de MOS . . . 18
1.6 Présentation des composants de protection étudiés . . . 20
1.6.1 SCR haute tension . . . 21
1.6.1.1 NPNR . . . 21
1.6.1.2 NPNZ . . . 22
1.6.1.3 NPNB . . . 24
1.6.2 PNP/Diode . . . 25

1.7 Conclusion 26

2 État de l'art sur la caractérisation transitoire des protections ESD haute tension 27
2.1 Introduction 27
2.2 Principe de fonctionnement du vf-TLP Oryx 27
2.2.1 Matériel employé 27
2.2.1.1 Générateur d'impulsion 27
2.2.1.2 Boîte de mesure (measurement pod) 28
2.2.1.3 Câbles et pointes RF 30
2.2.1.4 Oscilloscope 30
2.2.1.5 Merge box 30
2.3 Limites du système pour l'analyse transitoire. 32
2.3.1 Réponse fréquentielle des éléments du banc de mesure 32
2.3.1.1 Filtrage des sondes en courant et tension 34
2.3.1.2 Atténuation apportée par les câbles coaxiaux 35
2.3.1.3 Filtrage de la merge box 36
2.3.1.4 Bilan du filtrage du banc Oryx 36
2.3.2 Réglage du délai de la merge box 39
2.4 Méthode de mesure « indirecte » de la surtension au déclenchement : couplage avec un composant à protéger. 40
2.4.1 Principe de la mesure indirecte de pics de surtension 40
2.4.2 Application à l'étude de la surtension au déclenchement de protections ESD à base de transistors PNP 41
2.4.3 Limites de la méthode de mesure indirecte des pics de surtension 44
2.5 Conclusion 44

3 Nouvelle méthodologie basée sur la correction des données : le $transient-TLP$ (t-TLP) 45
3.1 Introduction 45
3.2 Principe de la méthode 45
3.3 Transformation de Fourier Discrète 49
3.3.1 Utilisation pratique de la Transformation de Fourier discrète sous $Matlab$ 49
3.3.2 Algorithme de FFT 51
3.4 Détermination du spectre fréquentiel nécessaire 52
3.4.1 Limites des analyseurs de réseau 52
3.4.2 Limites de l'oscilloscope 52
3.4.3 Choix des limites fréquentielles 53
3.5 Traitement des données temporelles et fréquentielles 55
3.5.1 Signal non nul à $t < 0$: correction de l'offset 55
3.5.2 Séparation des impulsions incidentes et réfléchies 55
3.5.3 Suppression des réflexions parasites 57
3.5.4 Compromis entre le temps de montée et la précision des résultats 61
3.5.5 Traitement sur les coefficients de paramètres -S mesurés 63
3.5.6 Calcul des formes d'ondes $V_{DUT}(t)$ et $I_{DUT}(t)$ 63
3.6 Résultats de mesure $t-TLP$ 64
3.6.1 Validation sur des cas simples 64
3.6.2 Application à la mesure sur des composants ESD de librairie 67
3.6.3 Limites de la méthode d'étalonnage à l'analyseur de réseau 69
3.7 Outil de mesure avancé avec étalonnage à l'oscilloscope 71

3.7.1 Introduction . . . 71
3.7.2 Mesure des paramètres -S à l'oscilloscope . . . 72
3.7.2.1 Principe . . . 72
3.7.2.2 Mise en œuvre . . . 73
3.7.2.3 Validité de la substitution de S_{inc} et S_{ref} par S'_{inc} et S'_{ref} . . . 75
3.7.3 Problématiques de dé-convolution . . . 77
3.8 Conclusion . . . 78

4 Méthode d'étalonnage pour des mesures $transient - TLP$ sur boîtier 79
4.1 Introduction . . . 79
4.2 Caractéristiques physiques du banc de mesures TLP . . . 81
4.2.1 Banc de mesures TLP . . . 81
4.2.2 Modifications du banc pour des mesures transitoires . . . 81
4.3 Étalonnage du banc de mesure . . . 83
4.3.1 Introduction . . . 83
4.3.2 Parties coaxiales 50 Ω . . . 84
4.3.3 Étalonnage des pinces . . . 84
4.3.3.1 Étape 1 : de $P2$ à $P4$. Mesure du coefficient de transmission de la partie coaxiale . . . 85
4.3.3.2 Étape 2 : de $P4$ à $P5$. Mesure de l'impédance des fils et des pinces 87
4.3.4 Étalonnage des pistes de la carte PCB . . . 88
4.3.4.1 Expression de V_{DUT} et I_{DUT} en fonction des tensions et courants dans P5 . . . 88
4.3.4.2 Évaluation de l'impédance caractéristique des pistes . . . 89
4.3.4.3 Mesure du paramètre γl . . . 90
4.3.5 Expression compacte pour la tension et le courant dans le plan du composant 91
4.3.6 Traitement des réflexions multiples entre les plans 5 et 6 . . . 92
4.3.7 Bilan des étapes d'étalonnage du système de mesure . . . 94
4.4 Résultats . . . 94
4.5 Conclusion. . . . 98

5 Analyse du déclenchement d'un composant de protection ESD haute tension à base de thyristor : le $NPNB$ 101
5.1 Introduction . . . 101
5.2 Conditions nécessaires au retournement . . . 102
5.3 Corrélation entre les mesures $transient - TLP$ et les simulations $TCAD$ . . . 104
5.4 Analyse du déclenchement du composant à protéger . . . 106
5.4.1 Introduction . . . 106
5.4.2 Présentation du composant à protéger . . . 106
5.4.3 Redéfinition de la fenêtre de conception . . . 108
5.5 Présentation du composant $NPNB$ . . . 111
5.5.1 Principe de fonctionnement . . . 111
5.5.2 Caractéristiques quasi-statiques et dynamiques . . . 112
5.6 Dynamique du claquage par avalanche des diodes base-collecteur en inverse : comparaison PIN vs. P^+/N^+ . . . 113
5.6.1 Introduction . . . 113
5.6.2 Simulation de la dynamique du claquage par avalanche des deux diodes . 115
5.6.3 Expression analytique pour l'évolution du courant d'avalanche lors du claquage . . . 118
5.7 Déclenchement transitoire du $NPNB$ . . . 121

5.7.1 Introduction . . . 121
5.7.2 De t_{BV}=2.6 ns à t_{EB}=2.95 ns : courant de génération . . . 122
5.7.3 De t_{EB} = 2.95 ns à t_{NPN} = 3.1 ns : déclenchement du NPN . . . 127
5.7.4 De t_{NPN}=3.1 ns à $t_{max,\ NPN\ lat.}$ = 3.6 ns : étude du repliement du NPN latéral . . . 132
5.7.5 Conduction du transistor PNP et de la diode verticale . . . 137
5.8 Solutions de dessin envisagées . . . 139
5.8.1 Règles de dessin . . . 139
5.8.2 Application des règles de dessin . . . 141
5.9 Conclusion . . . 146

Conclusion générale **i**

A Éléments d'analyse de Fourier **v**
A.1 Décomposition en séries de Fourier (DSF) . . . v
A.2 Transformation de Laplace . . . vi
A.3 Transformation de Fourier complexe . . . vi
A.4 Énergie et densité spectrale d'énergie . . . vii

B Fonctionnement d'un générateur d'impulsions $TLP/vf-TLP$ **ix**

Bibliographie **x**

Liste des publications **xvii**

Résumé **xix**

L'expérience nous apprend que l'homme n'apprend jamais rien de l'expérience.

George Bernard Shaw.

Introduction générale

Le terme « décharge électrostatique » (Electrostatic Discharge, ou ESD) désigne le transfert bref et intense de charges électriques entre deux objets de potentiels différents. Ce phénomène est la conséquence du processus de « charge » des deux objets en question, durant lequel un déséquilibre de charges électriques se crée entre les deux corps. L'électricité statique accumulée ne peut cependant donner lieu à une décharge électrique que si un chemin de conduction se crée entre les deux corps, par exemple par une mise en contact, ou encore par ionisation de la couche d'air isolant les deux corps. Le déplacement des charges d'un corps vers l'autre crée alors un courant électrique dont l'intensité est d'autant plus importante que la décharge est brève.

Ce phénomène intéresse les scientifiques depuis l'antiquité. Dans la vie courante, il se manifeste sous la forme des « châtaignes » que l'on peut ressentir par exemple en effleurant une poignée de porte, ou en sortant d'une voiture. La foudre en est un autre exemple plus spectaculaire.

Dans le domaine de la microélectronique, les défaillances causées par les ESD constituent depuis les années 1970 un sujet de préoccupation de plus en plus important, avec la complexification des circuits intégrés (CI) et la réduction de leurs dimensions. En effet, les hautes tensions mises en jeu et la réduction des dimensions donnent lieu à des champs électriques de plus en plus élevés et des densités de courant électrique de plus en plus fortes, qui peuvent conduire au claquage des diélectriques ou à la destruction thermique des CI.

Historiquement, les efforts des industriels dans la protection des CI contre les ESD se sont d'abord portés sur les risques encourus lors des étapes d'élaboration industrielle (fabrication, test, assemblage). Le premier axe de travail a porté sur la création d'espaces protégés des ESD (Electrostatic Protected Area, EPA). En salle blanche, par exemple, la génération de charges statiques peut être limitée par l'utilisation de matériaux à revêtements antistatiques et d'ioniseurs d'air pour neutraliser les charges. Lors des étapes de test et d'assemblage, la probabilité pour un opérateur d'infliger une ESD à un composant peut être singulièrement réduite si celui-ci se relie électriquement au plan de masse par le biais d'un bracelet conducteur. Ces programmes de prévention ont considérablement réduit les dangers d'accumulation d'électricité statiques dans

les composants.

Les recherches actuelles portent sur la conception de stratégies de protections internes au CI. Celles-ci reposent sur l'action de composants dits de protection dont la fonction est, à la détection d'une ESD, d'absorber tout le courant de décharge tout en maintenant la tension en-deçà d'une valeur critique. Les composants de protection ESD doivent en outre être conçus de telle sorte que le fonctionnement du CI n'est perturbé ni par l'ESD ni par le composant de protection.

De nombreuses stratégies de protection ont été élaborées depuis que cette problématique est prise en considération. Cependant, l'intégration croissante de modules différents sur un même produit, la réduction des dimensions et la sévérité croissante des normes internationales soulèvent aujourd'hui le problème de la rapidité de déclenchement des composants de protection. En effet, les quelques centaines de picosecondes nécessaires au déclenchement des composants les plus rapides suffisent parfois à la décharge électrostatique pour générer une défaillance irréversible dans le circuit à protéger. Cela est dû au fait que la tension de décharge ne peut pas être contrôlée tant que le composant de protection ne s'est pas déclenché. Le pic de surtension qui s'établit alors aux bornes du composant à protéger peut lui être fatal. Le besoin des industriels pour des composants de protection plus rapides est donc bien réel.

Ma thèse s'inscrit dans cette problématique, plus précisément sur la protection d'entrées/sorties (ES) haute tension (40-90 V) pour des technologies *Smart Power* développées à *Freescale Semiconductor*. Deux axes majeurs se distinguent dans cette étude. Le premier, d'ordre expérimental, concerne la possibilité de mesurer des phénomènes électriques extrêmement rapides (plusieurs centaines de Volt sur quelques dizaines de picosecondes). J'ai commencé à travailler sur cette problématique à la fin de l'année 2008 au $LAAS$, et une méthode de mesure a été élaborée, basée sur le post-traitement de données mesurées à l'aide d'un dispositif de mesures ESD classique. Le second axe de recherche concerne l'étude physique, notamment par la simulation électrothermique, de la dynamique de déclenchement de structures de protection existantes, et l'amélioration de leurs performances en termes de rapidité.

Le travail présenté est organisé en cinq chapitres. Le premier présente le cadre dans lequel s'inscrit ma thèse. Le rôle des composants de protection ESD intégrés dans les CI est d'abord défini, à travers la notion de « fenêtre de conception ». Puis les différents modèles permettant d'évaluer la robustesse d'un composant de protection soumis à une agression ESD sont étudiés. Les outils d'analyse permettant de tester la capacité du composant à respecter sa fenêtre de conception, et notamment le « TLP » sont présentés. Un état de l'art des composants de protection les plus couramment utilisés est établi. Enfin, les composants utilisés par Freescale pour

la protection des E/S haute tension en technologie $SmartMos$ 8 MV sont présentés.

Le second chapitre présente deux des méthodes de mesure les plus couramment utilisées pour évaluer l'amplitude du pic de surtension accompagnant le déclenchement des composants de protection haute tension. En premier lieu, une présentation et une analyse du « very-fast TLP » est réalisée, puis une méthode d'analyse « indirecte » est décrite. Les faiblesses de ces deux approches sont analysées.

Dans le troisième chapitre, nous présentons une nouvelle méthode de mesure sous pointes du comportement transitoire des composants de protection ESD, que nous avons nommée « transient-TLP ». Son principe est présenté dans une première section. Les bases mathématiques permettant son application sont ensuite brièvement exposées. Le réglage des appareils de mesure pour une utilisation optimale est déterminé et justifiés. Le traitement mathématique appliqué sur le signal est ensuite détaillé. Des résultats de mesure sont exposés. Enfin, une méthode d'étalonnage alternative, ne nécessitant pas l'emploi d'un analyseur de réseau est présentée.

Le quatrième chapitre étend la méthode présentée au troisième chapitre, et l'applique à la mesure de composants encapsulés en boîtier. Le banc de mesure utilisé est présenté dans une premier temps. Puis, la méthode d'étalonnage employée est détaillée. Enfin, des résultats de mesure sont exposés.

Dans le cinquième chapitre, une analyse du comportement transitoire d'un composant de protection ESD haute tension utilisé à Freescale est réalisée. Dans un premier temps, les conditions générales décrivant le repliement dynamique de la tension aux bornes d'un composant soumis à une décharge TLP sont étudiées. Ensuite, la corrélation que l'on peut établir entre les mesures « transient-TLP » et les simulations $TCAD$ est exposée. Nous étudions alors le comportement du composant à protéger afin de déterminer sa vulnérabilité aux ESD. Le composant de protection faisant l'objet de ce chapitre est ensuite présenté. Une analyse détaillée des étapes de son déclenchement est livrée. Enfin, nous présentons des règles de dessin permettant de réduire l'amplitude du pic de surtension au déclenchement du composant.

Chapitre 1

Décharges électrostatiques et composants de protection

1.1 Introduction

La protection des circuits intégrés (CI) contre les décharges électrostatiques est un enjeu très préoccupant pour l'industrie de la microélectronique. L'impact qu'une ESD peut avoir sur un CI concerne à la fois la destruction d'un composant comme la perte de fonctionnalité temporaire du circuit. Pour cette raison, les stratégies de protection développées par les concepteurs sont le fruit d'une approche globale, partant du circuit entier jusqu'au composant de protection lui-même. Nous allons dans ce chapitre introduire la problématique industrielle dans laquelle s'inscrit cette thèse, en mettant en évidence les principaux défis et enjeux auxquels doivent faire face les fabricants de semi-conducteurs. Nous verrons dans une première section de quelle manière s'appliquent les contraintes de conception des composants de protection, par la définition d'une « fenêtre de conception » pour ces composants. Puis une classification des différente types de décharges pouvant s'appliquer sur un circuit sera présentée, ainsi que les modèles de test qui y sont associés. Des outils de mesure permettant d'examiner le comportement des composants et en particulier le respect de leur fenêtre de conception seront présentés. Nous examinerons ensuite les composants de protections ESD les plus couramment employés. Pour finir, les composants de protection ESD haute tension de la technologie $SmartMos\ 8\ MV$ de Freescale Semiconductor seront présentés.

1.2 Fenêtre de conception

La vulnérabilité des CI vis-à-vis des ESD ne cesse d'augmenter face à la réduction des dimensions technologiques et à l'intégration croissante des composants sur une même puce. On

peut regrouper les processus de destruction d'un composant électronique par une ESD en deux types. Le premier est la destruction thermique, imputable aux fortes densités de courant pouvant circuler dans le composant. Le second est le claquage des diélectriques, qui est causé par les surtensions. L'oxyde de grille d'un transistor MOS, par exemple, est, de par sa faible épaisseur, particulièrement vulnérable aux surtensions.

Pour réaliser correctement son rôle, une protection ESD doit donc absorber toute l'énergie de décharge lors d'une ESD. La fonction d'un composant de protection est celle d'un interrupteur commandé parfait : placée en parallèle avec le circuit à protéger, son impédance doit être infinie en l'absence de décharge de manière à ne pas perturber le circuit, et nulle ou très faible dès qu'une décharge est détectée. La protection doit conduire le courant sur toute la durée de la décharge et maintenir la tension à une valeur inoffensive pour le circuit. Le laps de temps nécessaire à son déclenchement doit être idéalement nul, de sorte que la tension de décharge ne puisse pas atteindre la valeur critique de destruction du CI pendant cette durée.

La figure 1.1 montre un exemple d'intégration de structures ESD protégeant l'entrée d'un CI. Les protections 1 et 2 sont conçues de telle sorte qu'elles se déclenchent quand la *ddp* à leur borne dépasse la tension d'alimentation. Ainsi lors d'un stress ESD sur le plot de signal, si le potentiel de ce plot vient à dépasser la tension d'alimentation, la protection 1 se déclenchera, tandis qu'une tension négative sur le plot de signal déclenchera la protection 2.

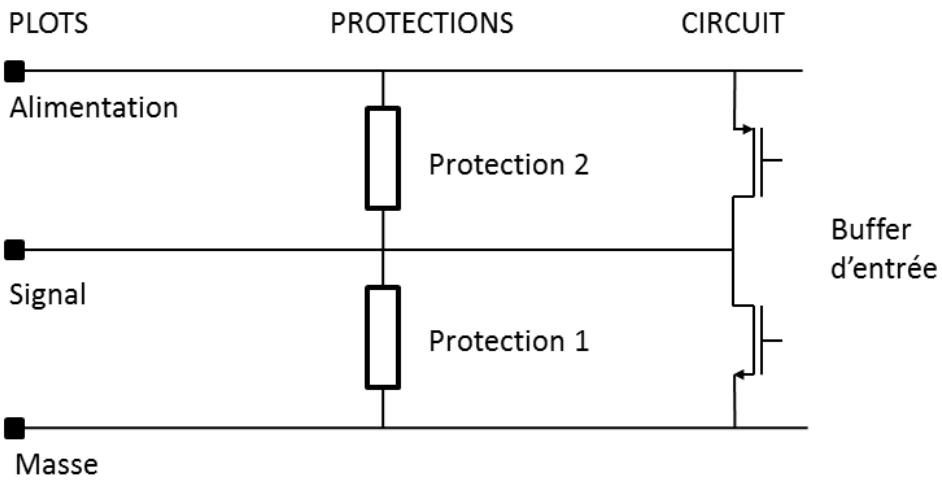

FIGURE 1.1 – Stratégie d'intégration de protections ESD localisées protégeant l'entrée d'un CI.

Les contraintes liées à la protection du circuit imposent à la protection ESD le respect d'une « fenêtre de conception » : il s'agit de la zone « autorisée » du plan $I-V$, à l'intérieur de laquelle doit s'inscrire la courbe cactéristique du composant de protection. Deux limites définissent cette fenêtre de conception. La borne inférieure de cette fenêtre est fixée par la tension d'alimentation

du circuit. En effet, la protection ESD ne doit pas se déclencher pour des tensions inférieures à cette tension d'alimentation, sans quoi elle risquerait de perturber le circuit en utilisation normale. Mais la tension ne doit non plus jamais dépasser la tension V_{max} de destruction du circuit à protéger. C'est cette tension V_{max} qui constitue donc la borne supérieure. A ces deux tensions limites s'ajoutent généralement des marges de sûreté de quelques Volt.

Toutes les protections ESD possèdent les mêmes deux états de fonctionnement : un état de haute impédance, ou état bloqué, dans lequel le composant doit se trouver en l'absence d'agression ESD, et un état de faible impédance, dit état passant, durant lequel il conduit le courant de décharge. La figure 1.2 montre la caractéristique courant-tension typique d'une protection ESD, mesurée avec un outil dédié, le « TLP », dont le fonctionnement sera exposé dans la partie 1.4.1. Cet outil permet d'extraire les paramètres caractéristiques du fonctionnement de la protection ESD, parmi lesquels figurent :

- la tension de déclenchement V_{t1}, aussi appelée tension de repliement. C'est la tension à partir de laquelle le courant commence à croître.
- La tension de maintien V_H. C'est la tension à laquelle le composant se stabilise à faible courant, immédiatement après son déclenchement. Cette tension est dans la plupart des composants plus faible que la tension de déclenchement. Elle doit être supérieure à la tension d'alimentation du circuit (borne inférieure de la fenêtre de conception). Cela permet de s'assurer que la protection ne reste pas verrouillée à l'état passant lors d'un déclenchement intempestif. Nous verrons plus en détail dans les partie 1.5 les processus physiques gouvernant la valeur de V_H.
- la résistance dynamique R_{ON} : c'est l'inverse de la pente de la caractéristique I-V, une fois le composant déclenché.
- La tension et le courant de destruction par claquage thermique : V_{t2}, I_{t2}.

Nous verrons que certains composants de protection, comme les diodes polarisées en inverse, ne présentent pas de mécanisme de repliement. Leur tension de maintien V_H est confondue avec leur tension de déclenchement V_{t1}.

1.3 Modèles d'évaluation de la robustesse

1.3.1 Robustesse des protections ESD et des circuits intégrés

L'une des propriétés essentielles des circuits de protection contre les ESD est leur capacité à conduire des courants très importants sans se dégrader. Cette propriété, appelée robustesse, doit être évaluée précisément à l'aide de testeurs dédiés. Ces testeurs reproduisent la forme d'onde d'une impulsion ESD réelle et permettent un contrôle de l'intégrité du composant entre deux impulsions. La forme d'onde d'une impulsion ESD agressant un composant est complexe car

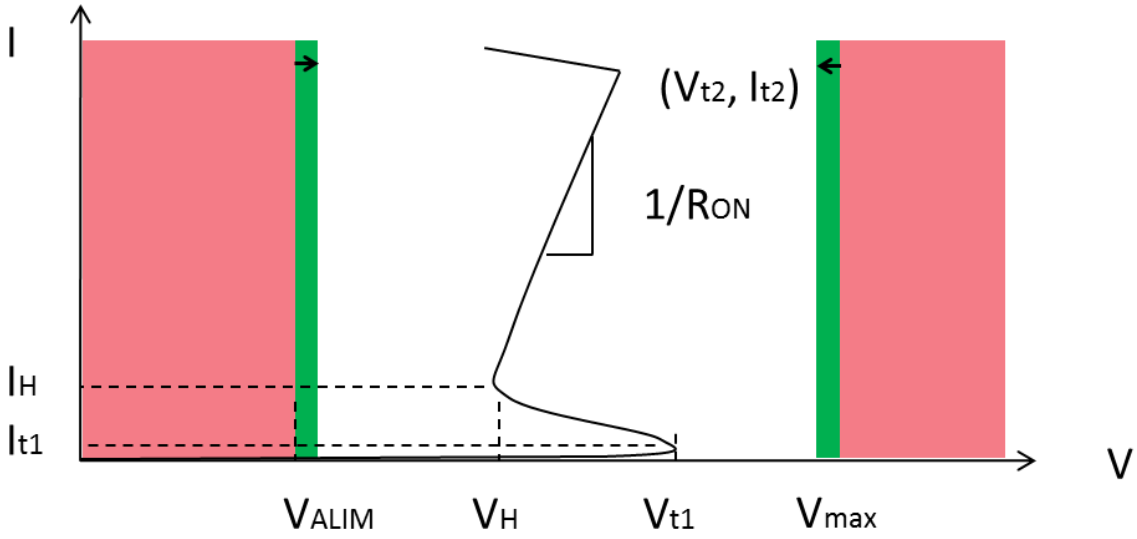

FIGURE 1.2 – Caractéristique I-V TLP typique d'une protection ESD et paramètres électriques associés. Les flèches noires représentent les marges de sûreté.

elle dépend de nombreux paramètres tels que l'impédance du générateur de décharge, les caractéristiques électriques du chemin de décharge, mais aussi l'évolution d'impédance du composant lui-même au cours de son déclenchement. Des normes ont été établies, visant à modéliser et classifier les décharges électrostatiques les plus courantes. À chacun de ces modèles est associé un dispositif de test permettant de reproduire ces conditions en laboratoire. On définit alors la robustesse d'un composant pour un modèle donné comme étant la tension maximale du générateur de décharge qui a pu lui être appliquée sans défaillance du composant.

Les modèles *ESD* les plus anciens se restreignent à des tests au niveau du composant, mais les exigences actuelles sur le marché des composants électroniques poussent désormais les fabricants de semiconducteurs à tester leurs produits au niveau système, voire même à des niveaux supérieurs. Dans l'automobile, les tests s'effectuent jusqu'au niveau véhicule.

1.3.2 Méthodes de test au niveau composant : normes *JEDEC ESD*

1.3.2.1 Modèle du corps humain ou human body model (*HBM*)

Tous les modèles de décharge *ESD* sont constitués d'un condensateur chargé puis déchargé à travers un circuit de décharge. La norme *HBM* est la plus ancienne des normes *ESD* et vise à simuler la décharge que génère un être humain (typiquement un opérateur en laboratoire) manipulant un circuit [1] [2] [3]. L'être humain peut être modélisé par une capacité d'environ 100 pF, se déchargeant à travers une résistance de 1.5 $k\Omega$ et une inductance de 3.5 μH constitués par la peau et le corps. Le modèle correspondant est représenté à la figure 1.3-a), et la forme

d'onde obtenue sur un court-circuit pour une tension de précharge de 2 kV est montré à le figure 1.3-b).

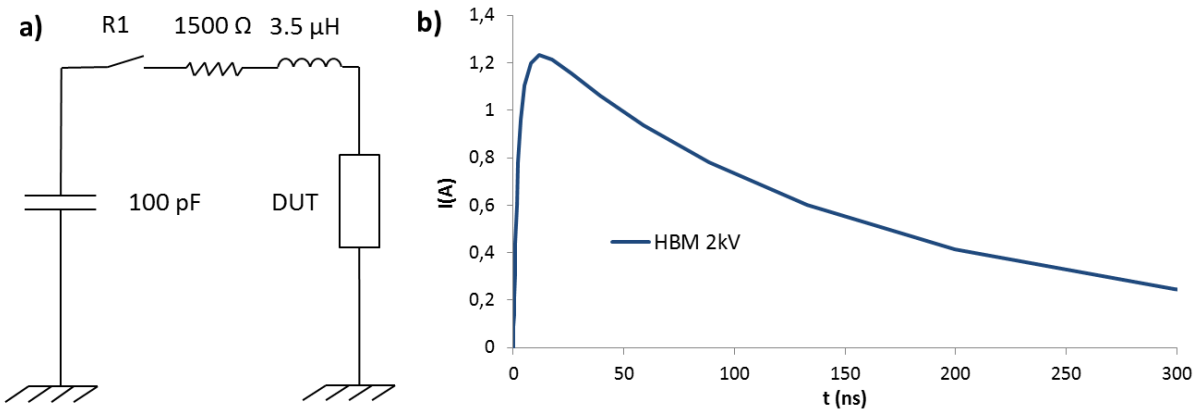

FIGURE 1.3 – Modèle électrique équivalent du corps humain (a) et décharge simulée sur un court-circuit pour une tension de charge de 2 kV (b).

Un condensateur de 100 pF est chargé puis déchargé à travers le composant sous test ou DUT (Device Under Test) à la fermeture du relai $R1$. Le courant maximal imposé est atteint vers 10ns, et suit ensuite une décroissance exponentielle sur plusieurs centaines de nanosecondes. L'énergie du stress est de l'ordre de la dizaine de microjoules. La défaillance des CI soumis à des agressions de type HBM est généralement causée par la fusion locale d'un des éléments du circuit (métallisation, silicium etc)

Pendant de nombreuses années, l'objectif habituel de robustesse à atteindre en HBM est resté constant et égal à 2 kV HBM, niveau de décharge maximal pouvant être infligé à un composant par un opérateur en laboratoire. Mais étant donné les progrès réalisé dans la prévention contre les ESD dans les laboratoires (bracelets antistatiques, matériaux adaptés etc), des voix se sont récemment élevées pour proposer un abaissement de cette limite à 1 kV [4].

1.3.2.2 Modèle de machine, ou Machine model (MM)

Le modèle MM est censé reproduire la décharge d'une machine (outils, fer à souder) de faible résistance à travers un CI [5] [6], [7]. Le schéma est très similaire à celui du modèle HBM, mais la capacité est ici de 200 pF (figure 1.4-a) et la résistance, correspondant à la résistance parasite de l'appareil, y est très faible (ici 10 Ω). La forme d'onde oscillatoire d'une décharge MM de 200 V sur un court-circuit est représentée à la figure 1.4-b. La particularité de ce type de décharge est que le composant y est agressé pour une polarité successivement positive et négative, ce qui permet d'évaluer les robustesses d'un circuit de protection à la fois en direct et en inverse, contrairement au modèle HBM [8].

La valeur de 200 V a longtemps été la référence pour la qualification MM, mais comme pour l'HBM, l'abaissement de cette valeur voire l'élimination pure et simple de l'étape de test MM a été récemment discutée [4].

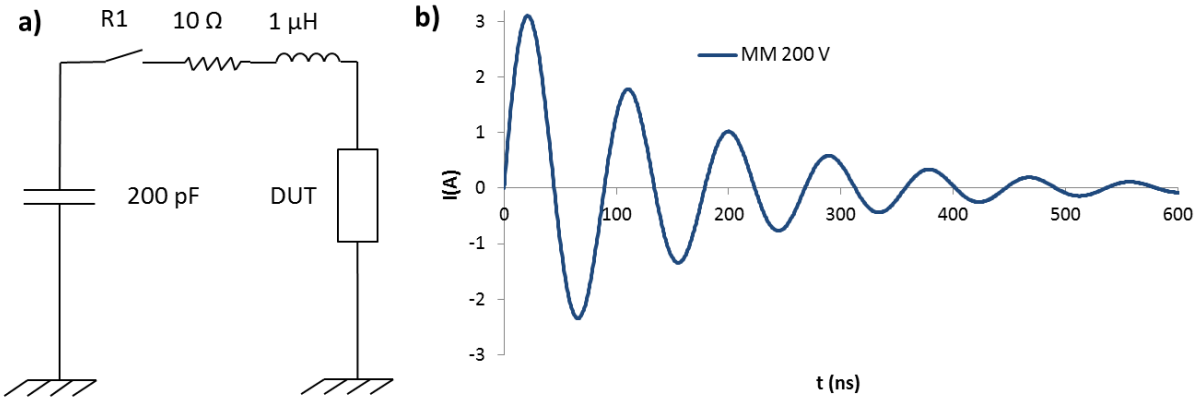

FIGURE 1.4 – Modèle électrique équivalent d'une machine (a) et forme d'onde simulée sur un court-circuit pour une tension de charge de 200 V (b).

1.3.2.3 Modèle du composant chargé, ou Charged device model (CDM)

Ce modèle (Charged Device Model) est très différent des deux premiers. Contrairement aux modèles CDM et MM, il ne vise pas à étudier l'application d'une décharge de forme d'onde prédéfinie à travers un composant, mais à reproduire les conditions de charge puis de décharge du CI lui-même à travers un plan de masse [9], [10], [11]. Les caractéristiques d'une décharge CDM dépendent fortement du composant sous test, ce qui rend très difficile l'évaluation de la valeur des composant parasites [12] [13]. Différentes méthodes ont été proposées pour reproduire ce phénomène, mais la plus courante consiste à charger le composant par induction en connectant une de ses broches à l'extrémité d'un condensateur chargé, puis, à le décharger en reliant la broche testée à la masse (socketed CDM).

La capacité de charge est plus faible que celle d'un opérateur ou d'une machine, de même que l'inductance et la résistance parasite du circuit de décharge (figure 1.5-a). La figure 1.5-b montre la forme d'onde en courant d'une décharge CDM de 500 V. La caractéristique est oscillatoire, mais fortement amortie. L'agression est beaucoup plus courte que celle du HBM et du MM, mais le temps de montée est très rapide (<1 ns). L'énergie dissipée par une telle décharge est inférieure au microjoule, mais son temps de montée très court peut, si le circuit de protection ne se déclenche pas suffisamment rapidement, occasionner une surtension transitoire fatale aux oxydes de grille des MOS. De ce fait, les phénomènes physiques occasionnant la défaillance des CI par des décharges CDM sont souvent très différents de ceux régissant leur endommagement par les décharges HBM et MM.

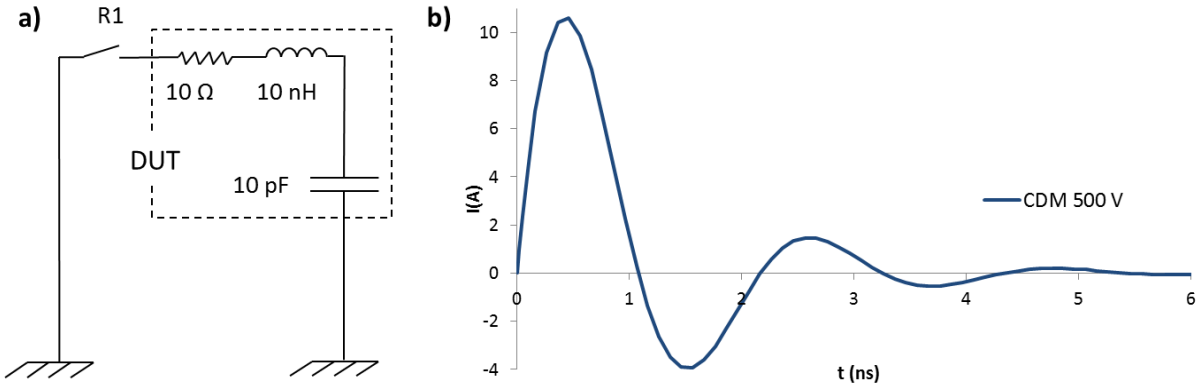

FIGURE 1.5 – Modèle électrique équivalent du composant chargé (a) et forme d'onde obtenue à 50 V de tension de charge.

La figure 1.6 compare les formes d'onde en courant des trois décharges HBM, MM et CDM aux tensions de précharge généralement requises pour la qualification des composants : respectivement 2 kV, 200 V et 500 V. On voit bien que les niveaux de tension ne sont pas corrélés aux courants atteints lors de la décharge : étant donné la forte résistance série du générateur HBM, le courant d'une décharge HBM de 2 kV ne dépasse pas 1.3 A alors qu'il atteint 10.5 A pour une décharge CDM de 500 V. Par soucis de lisibilité, nous avons ajouté un délai artificiel de 60 ns à la courbe représentant la décharge CDM.

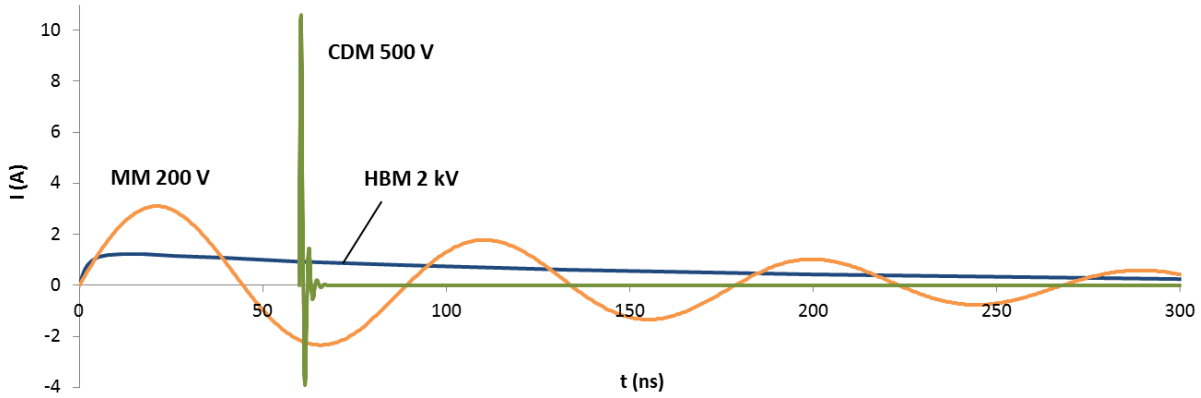

FIGURE 1.6 – Comparatif des formes d'ondes des trois modèles de décharge abordés dans cette section : HBM, MM, CDM.

1.3.3 Normes de test au niveau système et véhicule pour l'automobile

Les modèles présentée ci-dessus sont limités aux décharges que peuvent subir le composant durant les étapes de son élaboration, depuis la fabrication en salle blanche jusqu'à l'assemblage en laboratoire. Une fois assemblés, ils restent cependant très exposés aux ESD provenant de l'environnement extérieur. Des normes ont été établies pour encadrer le test ESD des produits

au niveau système et véhicule, utilisant un générateur d'impulsion appelé « pistolet ESD », ou « Gun ».

La figure 1.7 montre le schéma équivalent souvent utilisé pour décrire le pistolet ESD dans les normes IEC $61000-4-2$ [14] et ISO 10605 [15]. Les deux normes diffèrent par leurs capacités $C1$ et leurs résistances série $R2$. La capacité $C2$ en pointillés, de l'ordre de la dizaine de pF, n'apparaît pas dans les normes, mais est nécessaire pour décrire le couplage entre la pointe de décharge et le plan de masse. Dans la norme IEC $61000-4-2$, la capacité $C1$ vaut 150 pF et la résistance $R2$ 330 Ω. Dans la norme ISO 10605, la résistance $R2$ est de 2 $k\Omega$, et la capacité $C1$ peut valoir 150 pF ou 330 pF suivant l'application simulée. La première des deux normes reproduit le cas d'un humain rentrant à l'intérieur d'un véhicule, et la seconde celui d'un humain assis à l'intérieur du véhicule. Les capacités $C1$ et $C2$ sont chargées à travers une forte résistance $R1$. Des modèles plus complets ont été proposés [16] [17], mais sont tous liés à un appareil de test particulier.

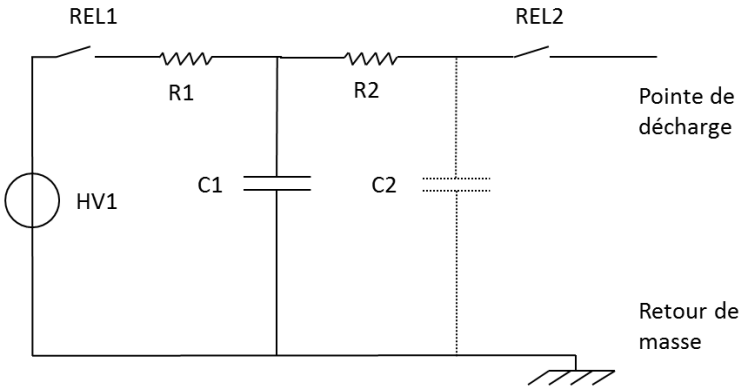

FIGURE 1.7 – Modèle électrique simplifié d'un générateur *Gun*. Seules $C1$ et $R2$ sont spécifiés dans les normes. $R1$ est la résistance de charge, typiquement de 1 $M\Omega$. $C2$ est une capacité parasite de l'ordre de 10 pF, reproduisant le couplage existant entre la pointe de décharge et le plan de masse.

La figure 1.8 montre une forme d'onde simulée sur un court-circuit avec les éléments de la norme IEC $61000-4-2$ à 8 kV. La décharge présente deux pics successifs : le premier, très rapide et intense, est dû à la décharge du condensateur $C2$, et le second à celle du condensateur $C1$ à travers la résistance $R2$ (1.7). Le courant atteint lors du premier pic dépend peu de la valeur de la résistance $R2$, au contraire de celui atteint pendant le deuxième pic. Ainsi, une forte valeur de $R2$ limite le courant du deuxième pic sans modifier la forme du premier pic. De même, la valeur de $C1$ n'a d'effet que sur la forme du second pic, à travers la constante de temps $\tau = R2C1$. Les valeurs de ces deux résistances dans les normes IEC $61000-4-2$ et ISO 10605 permettent donc de modifier l'importance du second pic par rapport au premier. La robustesse de l'élément sous test peut dépendre des valeurs utilisées. Si la défaillance est due au courant de pic, les trois modèles présenteront le même résultat, mais si elle provient d'un claquage thermique, l'effet de $R2$ et $C1$ est déterminant.

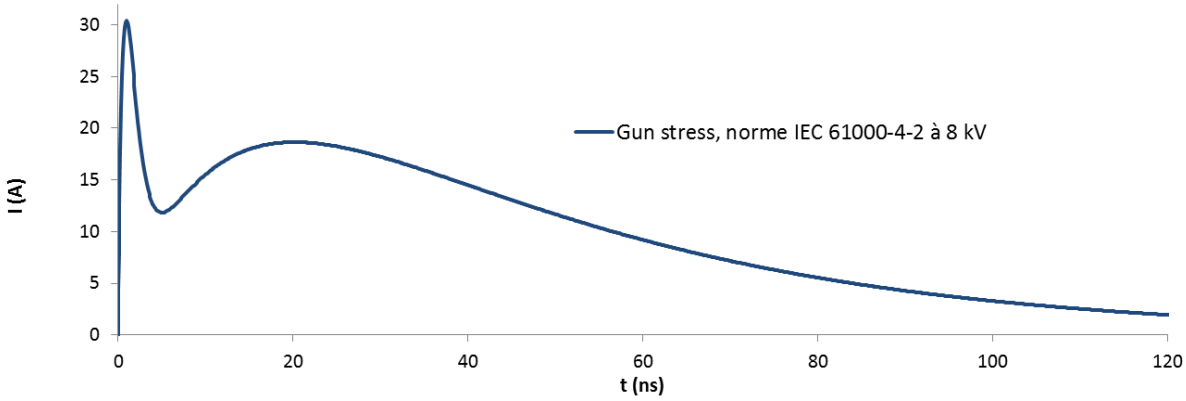

FIGURE 1.8 – Forme d'onde d'une décharge pistolet à 8 kV selon la norme IEC 61000-4-2.

1.3.4 Human Metal Model (HMM)

Une nouvelle norme appelée « Human Metal Model » est en cours de publication. Cette norme vise à établir une procédure de test normalisée au niveau composant, reproduisant la décharge d'une personne tenant dans ses mains un outil en métal. La forme d'onde correspondante s'approche de celle d'un pistolet ESD [18]. Plusieurs procédures sont à l'étude, l'une d'entre elles prévoyant l'utilisation d'un pistolet ESD comme générateur de décharge, et l'autre se basant sur un système d'impédance 50 Ω de type TLP modifié (« Transmission Line Pulsing », voir section 1.4.1).

L'avantage de la première solution est l'impédance élevée du générateur, qui la rapproche du cas réel de décharge, celui d'un humain tenant un outil métallique. Mais de sérieux doutes ont été émis quant à sa reproductibilité pour des mesures sur composant discret [19]. D'importantes disparités ont été constatées sur le courant de pic, le courant à 30 ns et 60 ns, ainsi que le spectre à haute fréquence de l'impulsion générée par différents modèles de pistolets respectant la norme IEC.

L'utilisation d'un outil de type TLP est également envisagé. La très bonne répétabilité du système et l'absence de rayonnement parasites tels que peuvent en générer les pistolets ESD en font un bon candidat [20].

1.4 Outils d'investigation

Les outils de caractérisation HBM, MM et CDM doivent être vus comme de simples outils de mesure de la robustesse des composants. Ils n'apportent pas a priori d'informations sur la capacité du composant à respecter sa fenêtre de conception. Des outils d'analyse plus fine ont donc été développés, reproduisant des formes d'onde semblables à celles des impulsions HBM et CDM, mais permettant en plus de mesurer la tension et le courant du composant durant son

fonctionnement. En effet, dans les testeurs HBM et CDM en particulier, l'ajout des sondes de courant et tension est possible mais peut apporter une importante rupture d'impédance dans le système. Pour ces raisons, des méthodes de caractérisation alternatives ont été proposées. Elles offrent l'avantage d'une meilleure reproductibilité de par la suppression des nombreuses sources de bruit présentes dans les testeurs de robustesse (inductances parasites des fils, couplage entre les lignes etc.). L'énergie dissipée y est comparable à celle d'une décharge électrostatique. Mais l'avantage déterminant de ces outils est la possibilité d'extraire les formes d'onde de courant et tension du composant durant son fonctionnement pour les principaux types d'agression.

1.4.1 Mesures impulsionnelles $TLP/vf-TLP$

1.4.1.1 Principe de la caractérisation quasi-statique

La méthode de Transmission Line Pulsing (TLP) est la plus employée d'entre elles. Elle a été introduite en 1985 par T. Maloney [21], pour pouvoir mesurer la caractéristique I-V des composants à fort courant. Le générateur TLP envoie des impulsions carrées d'une durée limitée (typiquement 100 ns). Cette faible durée permet de limiter l'énergie dissipée dans le composant durant la mesure, et d'élever le courant maximal pouvant être injecté sans détruire le composant par échauffement. Pour la grande majorité des composants, cette durée de 100 ns est également supérieure à la durée d'établissement du régime permanent, ce qui permet aux formes d'ondes V(t) et I(t) d'être stabilisées avant la fin de l'impulsion.

Un moyennage des formes d'onde de courant et de tension sur les dernières nanosecondes de l'impulsion, pour des niveaux de courant d'intensité croissante, permet d'obtenir une caractéristique I-V dite « quasi-statique » du composant, car à mi-chemin entre la mesure dynamique et la mesure statique (fig 1.9). Il importe évidemment de prendre garde au choix de la fenêtre de temps choisie pour le moyennage, de sorte que les phénomènes transitoires liés tant au fonctionnement du composant qu'à celui du dispositif de mesure soient bien terminés. Une mesure statique I-V à faible courant du composant entre chaque mesure impulsionnelle permet de vérifier l'état du composant et d'évaluer sa robustesse. On extrait de ces graphes I-V les grandeurs caractéristiques de la protection, qui suffisent à décrire le fonctionnement quasi-statique de la structure : la tension de déclenchement V_{t1}, de repliement V_H de destruction thermique V_{t2}, ainsi que les niveaux de courant associés (fig 1.2).

Le système TLP génère des impulsions rectangulaires de $-1000\ V$ à $1000\ V$, d'une durée de 100 ns, et d'un temps de montée modulable variant entre 1 ns et 10 ns. Par sa forme d'onde, il se rapproche d'un générateur d'impulsions HBM. Le temps de montée t_r de l'impulsion peut être modifié à l'aide de filtres passe-bas en série avec le générateur. Pour une bande passante

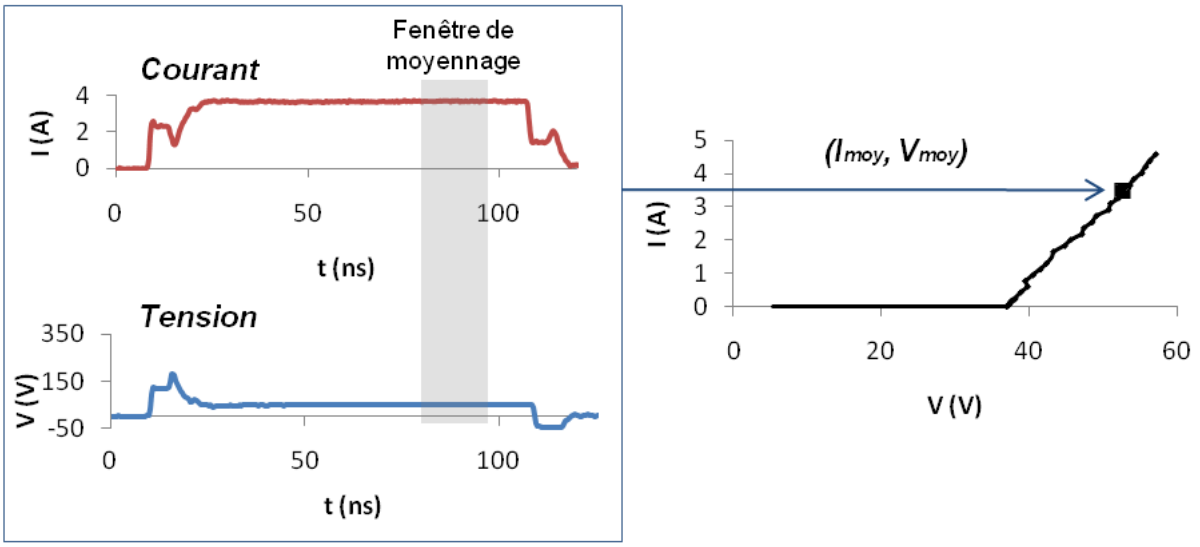

FIGURE 1.9 – Construction point par point d'une caractéristique $I-V$ TLP.

Δf, le temps de montée t_r en sortie peut se retrouver par la relation :

$$t_r = \frac{0.35}{\Delta f} \tag{1.1}$$

Un autre type de système impulsionnel, le very fast-TLP ($vf-TLP$) a été introduit par H. Gieser en 1996 [22] pour reproduire les décharges plus rapides, de types CDM. Les temps de montée générés sont compris entre 100 ns et 1 ns, et les durées d'impulsions entre 1.25 ns et 5 ns. Ce temps de montée très faible permet d'étudier l'effet de fronts de montée abrupts tels ceux rencontrés en CDM. Le $vf-TLP$ permet aussi de caractériser les composants de protections à des niveaux de courants plus élevés qu'en TLP, la plus courte durée d'impulsion garantissant une énergie dissipée plus faible dans le composant. Nombre de composants peuvent ainsi être testés jusqu'à 20 A sans être endommagés. En revanche, le $vf-TLP$ n'est pas un outil de mesure quasi-statique, la mesure ne s'effectuant qu'aux alentours de 3-4 ns (contre 70-80 ns en TLP). Le comportement du composant est à ce stade souvent gouverné par les processus dynamiques durant cette fenêtre temporelle.

1.4.1.2 Fonctionnement d'un banc de mesure TLP/vf-TLP en configuration TDR

La mesure de la tension et du courant du composant sous test (DUT) se fait par la méthode de réflectométrie temporelle (Time Domain Reflectometry). Une impulsion TLP créée par le générateur (l'impulsion dite « incidente »), est envoyée dans le système de mesure via le câble coaxial TL_1 (fig 1.10). Le fonctionnement du générateur TLP est détaillé dans l'annexe B. Les tensions et courants incidents sont mesurés grâce à des sondes dédiées et échantillonnés par un

oscilloscope à haute bande passante. Pour pouvoir capturer les temps de montée spécifiés dans les normes, l'oscilloscope doit avoir une bande passante d'au moins 400 MHz pour le TLP et d'au moins 2 GHz pour le $vf-TLP$.

Après avoir été mesurée par l'oscilloscope, l'impulsion incidente se propage dans TL_2, jusqu'à atteindre le composant sous test (DUT), sur lequel elle est partiellement réfléchie et partiellement absorbée, suivant l'impédance du composant. L'onde réfléchie est une image de l'impédance temporelle du composant, donc de son comportement à son déclenchement.

L'onde réfléchie se propage dans TL_2 et est mesurée à nouveau par l'oscilloscope. Celui-ci fournit donc dans la même acquisition une mesure des ondes incidentes et réfléchies à la suite, séparées par le délai correspondant à un aller-retour dans le câble TL_2. Le courant I_{DUT} et tension V_{DUT} du composant se retrouvent par la relation :

$$V_{DUT} = V_{inc} + V_{ref} \tag{1.2}$$

$$I_{DUT} = I_{inc} + I_{ref} \tag{1.3}$$

avec V_{inc}, I_{inc}, respectivement les tensions et courants incidents, et V_{ref}, I_{ref}, les tensions et courants réfléchis.
Le délai séparant ces deux ondes incidentes et réfléchies est typiquement de quelques nanosecondes. Dans le cas d'une impulsion TLP de largeur 100 ns, elles se superposent donc sur l'écran de l'oscilloscope, ce qui facilite l'extraction de V_{DUT} et I_{DUT} (technique du TDR-Overlap). En vf-TLP, en revanche, un traitement est nécessaire pour reconstituer les formes d'ondes vues par le composant (TDR-Separate).

1.4.2 CC-TLP

Nous avons vu qu'un des objectifs du $vf-TLP$ était de reproduire les caractéristiques des décharges CDM, par leurs impulsions de faible durée et leurs fronts montants extrêmement rapides. Gieser et al. ont donc tenté d'établir une corrélation entre le test CDM et le test $vf-TLP$ en comparant les signatures de défaillance d'un même circuit soumis aux deux tests [23]. La comparaison des seuils de défaillance sur les trois entrées du circuit employé pour les deux types de test n'a pas permis d'établir de corrélation franche. De plus, une analyse détaillée de topographie du circuit a montré que les chemins de courant suivi par les décharges CDM et $vf-TLP$ étaient radicalement différents. Les auteurs concluent qu'aucune corrélation ne peut être établie entre un stress sur une broche, de type CDM, et le stress sur deux broches que constitue le $vf-TLP$. Le $vf-TLP$ reste cependant une bonne méthode pour évaluer la robustesse et les conditions de déclenchement d'une protection ESD isolée.

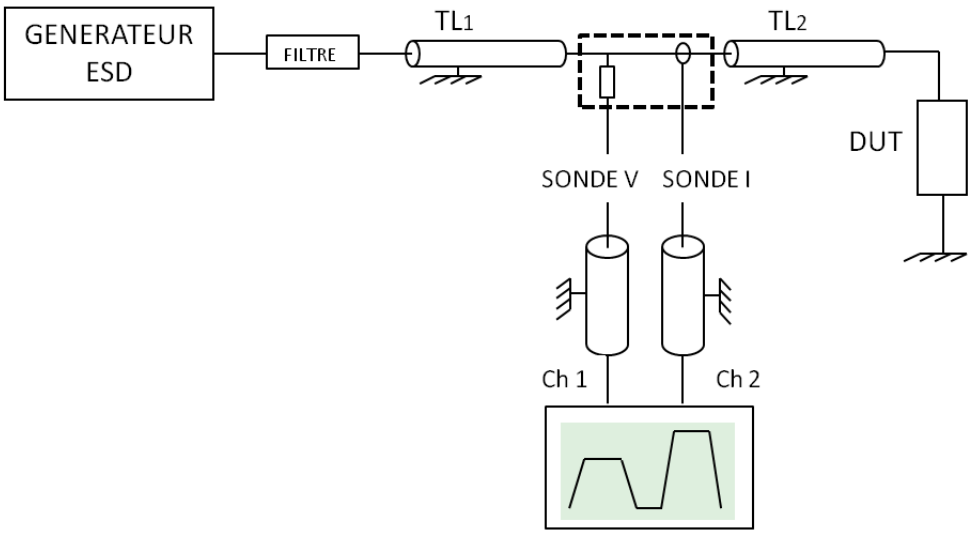

FIGURE 1.10 – Schéma de principe de la caractérisation par réflectométrie. Exemple du vf-TLP, dans lequel les impulsions incidentes et réfléchies apparaissent séparées sur l'écran de l'ocilloscope.

Les mêmes auteurs ont établi une nouvelle méthode de mesure dans laquelle les impulsions sont directement injectées sur la broche à tester, le composant étant par ailleurs relié à la masse via un émulateur de boîtier (« package emulator ») conçu spécifiquement pour chaque puce [24]. Cette méthode, appelée « capacity coupled TLP », ou $CC-TLP$ utilise un banc de mesure $vf-TLP$ pour générer la décharge et l'acheminer vers le composant et bénéficie donc de la bonne reproductibilité du générateur TLP. Une très bonne corrélation a été obtenue avec le CDM sur des circuits montés en boîtier, tant sur le courant de pic qu'au niveau des signatures de défaillance électriques et physiques du composant [25] [26].

1.5 Etat de l'art des composants de protection

Nous allons dans ce chapitre présenter les principaux composants utilisés pour la protection contre les décharges électrostatiques. Une grande partie des composants les plus courant repose sur le principe du transistor bipolaire autopolarisé (TBA). Le fonctionnement du transistor bipolaire latéral, vertical, ainsi que le MOS utilisé en ESD en sont de bons exemples. L'étude des TBA en conditions ESD a fait l'objet d'un traitement abondant dans la littérature [13] [27] [28]. Une analyse poussée de la physique des composants bipolaires en régime de forte injection a également été menée dans [29], [30] et [31]. Nous étudierons également dans le chapitre 5 le fonctionnement des diodes et des transistors bipolaires autopolarisés, en apportant une contribution originale sur leurs dynamiques de déclenchement.

1.5.1 Les diodes

Le composant le plus simple permettant de conduire le courant à partir d'un certain seuil de tension est la diode PN. Elle peut être utilisée tant en direct qu'en inverse, et a pour principal avantage la relative simplicité de conception et l'économie de place réalisé par rapport à des composants plus complexes.

Polarisée en direct, sa tension de seuil est égale à 0.5 V en technologie Silicium. Son déclenchement en direct est relativement lent (de l'ordre de la nanoseconde), car lié aux phénomènes de modulation de conductivité dans les zones quasi-neutres (voir chapitre 5). Cela peut devenir problématique pour des applications CDM ou Gun, dont les fronts de montée sont abrupts.

Polarisée en inverse, son fonctionnement repose sur le phénomène d'avalanche. A une certaine tension appelée tension de claquage, ou « Breakdown Voltage » (BV), les porteurs de charge (électrons et trous) acquièrent suffisamment d'énergie, sous l'effet du fort champ électrique régnant dans la zone de charge d'espace, pour ioniser les atomes du réseau cristallin par impact. Chaque porteur crée par ionisation peut générer à son tour par collision une paire électron-trou et alimenter le phénomène d'« avalanche ». Ainsi, grâce au grand nombre de porteurs présents dans la zone d'avalanche, la diode en inverse peut conduire des courants très élevés. Elle trouve donc naturellement son application en électronique de puissance, même si sa forte résistance à l'état passant R_{ON} restreint son champ d'application.

La tension de claquage d'une diode en inverse ne dépend que des concentrations en dopants de part et d'autre de la jonction, ainsi que des profils de dopage, et peut donc être modifiée selon le besoin.

1.5.2 Transistors bipolaires autopolarisés

Le transistor bipolaire autopolarisé (TBA) est un bipolaire dont les contacts d'émetteur et de base sont reliés soit par une résistance externe, soit par un court-circuit (figure 1.11-a). La polarisation en direct de la jonction émetteur base y est assurée par la chute de potentiel s'établissant à travers la région de base (résistance d'accès) et éventuellement la résistance externe R_B. Nous présenterons dans cette partie le comportement d'un bipolaire NPN, plus couramment utilisé que le PNP pour des applications ESD, mais le principe de fonctionnement de ces deux transistors sont similaires. Il sont généralement constitués de trois régions : un émetteur court et fortement dopé, une base courte et faiblement dopée, et une région de collecteur longue et fortement dopée.

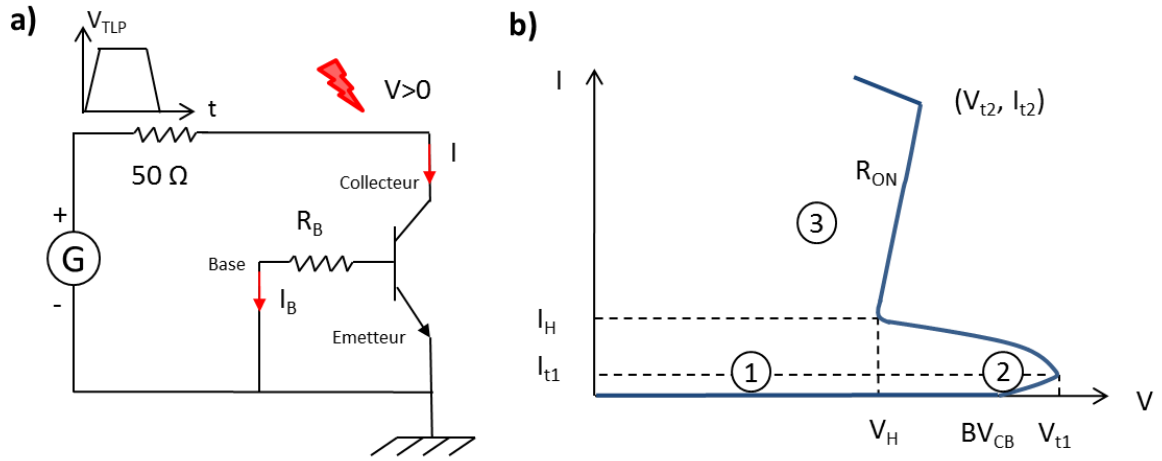

FIGURE 1.11 – Schéma représentatif d'une mesure TLP sur transistor bipolaire autopolarisé NPN (a). Courbe caractéristique $I-V$ correspondante (b). On distingue les trois zones de fonctionnement : état bloqué (1), avalanche de la diode base-collecteur (2), et fonctionnement replié (3).

La caractéristique TLP schématique d'un TBA typique est reportée en figure 1.11-b. Dans la configuration de la figure 1.11-a, une impulsion positive est appliquée sur le contact de collecteur, l'émetteur étant relié à la masse. La jonction collecteur-base est polarisée en inverse. Pour des tensions inférieures à sa tension de claquage par avalanche BV_{CB} (zone 1 de la figure 1.11-b), le courant circulant dans le composant est très faible. Il est en effet égal au courant de saturation de la diode collecteur-base en inverse, négligeable par rapport aux courants mis en jeu lors d'une ESD.

Le claquage par avalanche de la jonction base-collecteur en inverse a lieu quand la tension appliquée au transistor atteint BV_{CB} et un courant d'avalanche circule alors dans le composant. Ce dernier se comporte dans un premier temps comme une diode en inverse, avec une résistance à l'état passant R_{ON} élevée (zone 2). Dans cette zone, le courant circulant dans résistance de base et donc la chute de potentiel à ses bornes sont encore trop faibles pour pouvoir polariser en direct la jonction émetteur-base.

A mesure que le niveau de courant I dans le composant augmente, la chute de potentiel à travers la résistance de base s'accroît, et à un certain niveau de tension $V = V_{t1}$ et de courant $I = I_{t1}$, la tension base-émetteur devient suffisamment importante pour annuler le potentiel de diffusion de la diode émetteur-base en direct. La barrière de diffusion annulée, les nombreux électrons issus de l'émetteur N fortement dopé peuvent diffuser librement à travers la base P du NPN. En atteignant la zone d'avalanche de la jonction base-collecteur, ils subissent une multiplication par impact. Il s'ensuit une brutale augmentation du courant I et une brutale chute de l'impédance du composant. La tension décroît et se stabilise à une faible valeur V_H, dite tension

de maintien. Ce phénomène est appelé « repliement ». La tension dite de repliement V_{t1} est la tension maximale atteinte par le composant avant d'entrer en mode de repliement. Le courant correspondant est noté I_{t1}

Dans la zone 3, le composant est dans un mode de fonctionnement à faible impédance, caractérisé par une résistance dynamique R_{ON}, et limité par le courant I_{t2} de destruction du composant par les phénomènes d'emballement thermique. Plus la tension de maintien est basse, moins l'énergie dissipée est importante. Le phénomène de repliement tend donc à augmenter la limite supérieure I_{t2} en permettant au composant de dissiper plus d'énergie. Ce repliement ne doit pas être trop important, pour éviter de sortir de la fenêtre de conception.

1.5.2.1 Modes de déclenchement par « dV/dt »

Nous avons vu que l'effet bipolaire s'enclenchait dès que le courant circulant dans la résistance de base permettait à la tension base-émetteur d'atteindre sa tension de seuil V_S. Dans le cas étudié ci-dessus, ce courant est un pur courant d'avalanche, résultant du claquage de la jonction base-collecteur. Mais dans certains cas, en particulier lors de fronts de montée rapides, ce courant peut également être fourni par la charge de la capacité de jonction base-collecteur [32] [33]. Ce courant capacitif additionnel, s'ajoutant au courant d'avalanche, est d'autant plus important que la pente « dV/dt » de l'impulsion est élevée. Ainsi, pour une tension de claquage base-collecteur BV donnée, le courant permettant à la tension émetteur-base d'atteindre la tension V_S peut être atteint plus tôt si la pente « dV/dt » est plus élevée. Autrement dit, une augmentation du « dV/dt » incident provoque un abaissement de la tension de déclenchement V_{t1}. Ce phénomène est communément désigné sous le nom de déclenchement par « dV/dt » [27].

Le déclenchement par « dV/dt » peut être mis à profit pour accroître la rapidité de réponse d'un composant soumis à des impulsions à front montant abrupt. Cependant il doit être employé avec précautions, et ce pour deux raisons. La première est que la tension de déclenchement d'un composant sujet au déclenchement par « dV/dt » dépend de la forme de l'impulsion incidente et est donc difficile à contrôler. Ceci pose un évident problème au niveau du respect de la fenêtre de conception. La seconde raison est qu'un composant dont la capacité parasite est élevée est en général plus susceptible aux perturbations CEM.

1.5.2.2 Transistors bipolaires latéraux

Une structure bipolaire est dite latérale ou verticale selon la direction prise par le courant traversant la base. Une structure latérale NPN sur substrat P^+ est représentée en figure 1.12. La base est constituée d'un caisson P implanté dans une couche épitaxiée P^-. La tension de claquage d'une telle structure peut être contrôlée par la largeur de la zone P^- de la base située

entre les implantations d'émetteur et de collecteur (S_p sur le schéma) [31] [34]. En effet, avant le déclenchement du composant, la tension appliquée au composant s'étend principalement sur les zones les plus résistives. Donc plus la zone peu dopée P^- est courte, plus la tension se concentre sur la jonction base-collecteur, réduisant alors la tension qu'il est nécessaire d'appliquer au composant entier pour obtenir le claquage de cette jonction.

La faiblesse des bipolaires latéraux tient à leur robustesse, moins importante que celle des bipolaires verticaux. Cela est dû au faible volume sur lequel se concentre le courant une fois le composant déclenché.

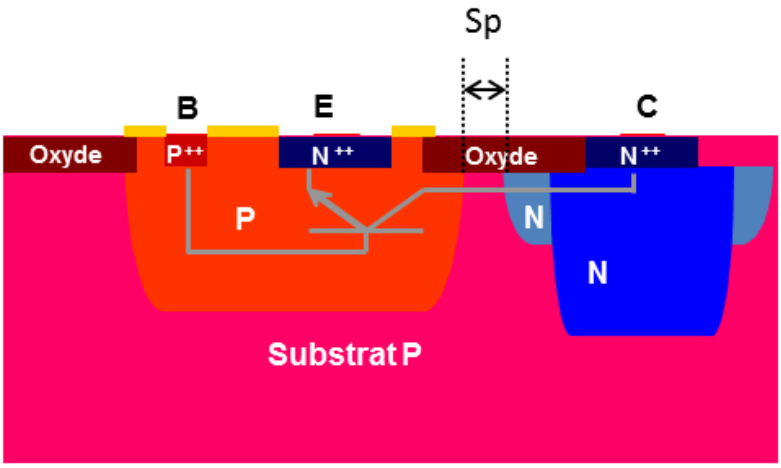

FIGURE 1.12 – Coupe technologique $2D$ simplifiée d'un transistor bipolaire latéral.

1.5.2.3 Transistors bipolaires verticaux

Un exemple de structure verticale est représenté à la figure 1.13. Le courant traverse la base verticalement, le collecteur étant ici constitué par une couche enterrée $N+$. La robustesse du composant peut être augmentée en allongeant la diffusion N^+ d'émetteur pour augmenter le volume dans lequel le courant se répartit. En revanche, les tensions de déclenchement possibles, déterminées par le dopage et la profondeur du caisson P de base et de la couche enterrée, sont limitées par les possibilités offertes par la technologie.

Il peut exister beaucoup plus de variantes possibles selon la technologie utilisée pour fixer la tension de déclenchement des transistors latéraux. Une des solutions courantes consiste à réaliser une combinaison entre un bipolaire latéral et un vertical. Tant que la tension de déclenchement latérale reste inférieure à la tension de déclenchement verticale, on peut profiter de la possibilité de réglage du BV latéral pour ajuster la tension de déclenchement du composant. Nous verrons dans le chapitre 5 qu'une fois le bipolaire latéral déclenché, les effets de forte injection dans la

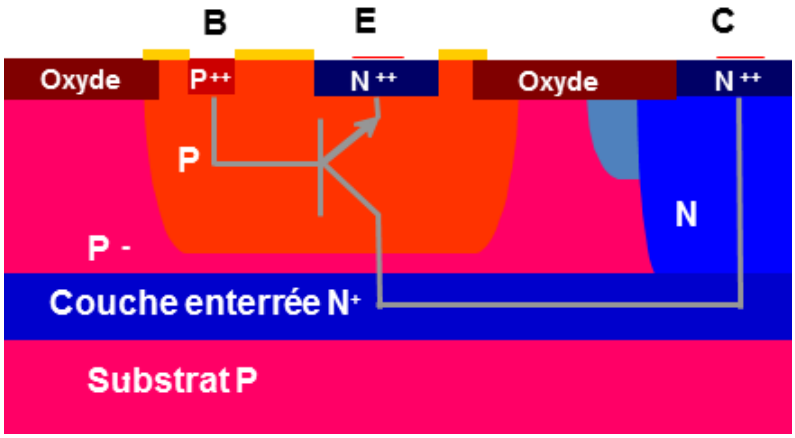

FIGURE 1.13 – Coupe technologique $2D$ simplifiée d'un transistor bipolaire vertical.

base tendent à moduler sa conductivité, et à abaisser la tension de claquage verticale, favorisant alors un déclenchement vertical. Le courant se répartit ensuite sur les deux chemins de courant représentée par les deux bipolaires déclenchés.

1.5.3 Silicon Controlled Rectifier SCR

Le thyristor, ou Silicon Cotrolled Rectifier (SCR) est formé de l'association d'un transistor PNP et d'un NPN imbriqués, de manière à ce que la base de l'un constitue le collecteur de l'autre et réciproquement (figure 1.14). Les deux transistors partagent la même jonction base-collecteur. Le claquage de cette jonction initie le déclenchement d'un des deux transistors, qui provoque à son tour le déclenchement de l'autre transistor en injectant du courant dans sa base.

Le SCR est très utilisé en technologie $CMOS$ comme protection ESD, en particulier pour sa taille relativement réduite et sa faible capacité [35] [36]. En revanche, son temps de déclenchement est relativement important, car il nécessite le déclenchement successif des deux bipolaires. Cette lenteur a longtemps empêché les concepteurs ESD de les utiliser pour protéger les circuits contre les décharges CDM. Des solutions ont été trouvées pour pallier cette lenteur, basées à la fois sur l'optimisation d'un circuit de déclenchement additionnel et sur des solutions au niveau du composant [37] [38].

1.5.4 Composants à base de MOS

Les transistors MOS sont parfois utilisés comme dispositifs de protection contre les ESD, en particulier en technologie $CMOS$. Souvent, ce n'est pas l'effet MOS qui est mis à profit, mais le transistor bipolaire parasite qui lui est associé. La figure 1.15 présente le cas d'un transistor

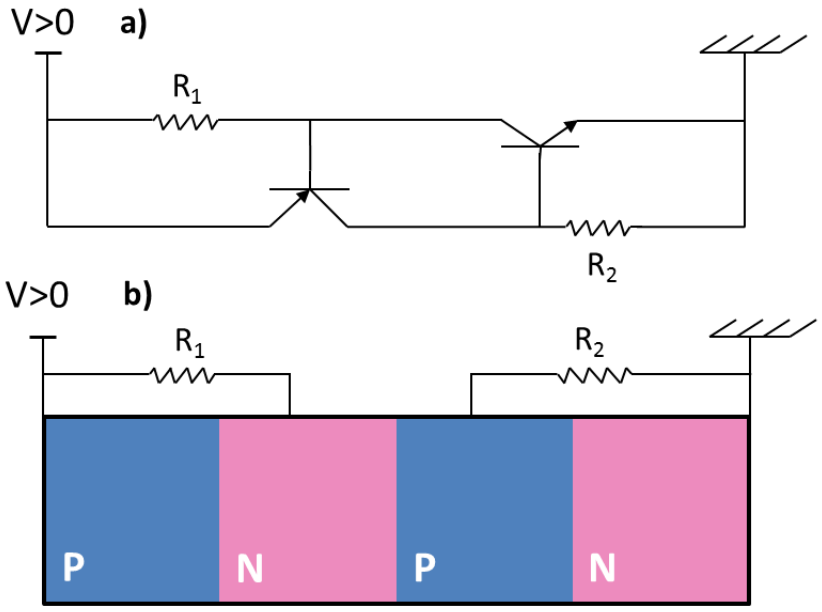

FIGURE 1.14 – Schéma électrique représentatif d'un SCR (a), composé d'un NPN et d'un PNP imbriqués. La structure $PNPN$ correspondante est représentée à la figure (b). Les résistances R_1 et R_2 sont les résistances intrinsèques des bases des deux transistors.

$NMOS$, dont le NPN parasite est utilisé pour conduire la décharge. Pour éviter la formation d'un canal d'électrons sous l'oxyde de grille, la résistance de grille R_G doit être prise nulle. Ainsi, lors d'un stress positif sur le drain, le dispositif se comporte comme un NPN latéral, le drain jouant le rôle de collecteur, la source d'émetteur, et le substrat de base. Comme dans un TBA, c'est la jonction base-collecteur (substrat-drain) qui après claquage par avalanche fournit le courant de polarisation de la jonction base-émetteur (substrat-source) et déclenche l'effet bipolaire.

Un tel composant est assez sensible aux phénomènes de déclenchement par dV/dt évoqués en section 1.5.2.1. Ici l'abaissement de la tension de déclenchement est lié au courant de charge non seulement de la capacité de jonction base-collecteur (substrat-drain), mais également de la capacité de grille, qui fournit une part importante du courant.

Il peut être intéressant de favoriser la formation transitoire d'un canal d'électrons sous l'oxyde de grille pour dériver le courant de décharge pendant le laps de temps nécessaire au déclenchement du NPN parasite. On fixe ici une valeur non nulle à la résistance de grille pour autoriser la formation d'un canal. Lors d'un stress positif abrupt sur le drain, le potentiel du substrat sous l'oxyde de grille augmente brutalement, et l'ensemble « résistance de grille -oxyde de grille » se comporte alors comme un circuit RC soumis à une rampe de tension. Un courant transitoire circule alors à travers R_G, permettant au potentiel de grille d'atteindre la tension de seuil d'ouverture du canal. Un canal d'électrons se forme sous l'oxyde de grille et le composant conduit

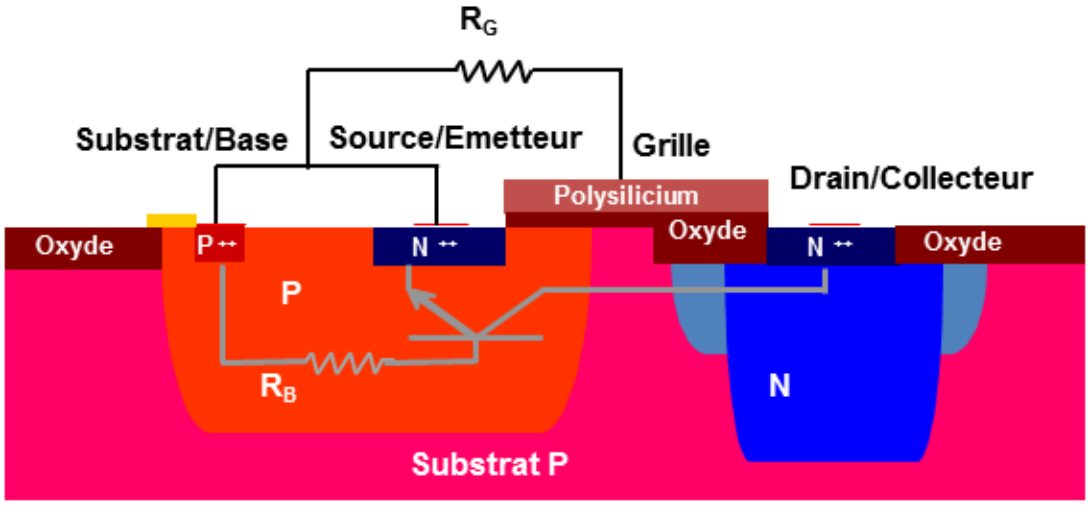

FIGURE 1.15 – Coupe technologique simplifiée d'un transistor $NMOS$ utilisé comme protection ESD, montrant le NPN parasite.

le courant par effet MOS. Si la résistance de grille est correctement choisie, le canal tend à disparaître au cours de l'impulsion, à mesure que la pente « dV/dt » imposée au composant devient moins abrupte.

1.6 Présentation des composants de protection étudiés

Les composants présentés dans cette section ont été développés pour la protection d'E/S haute tension en technologie $SmartMos$ 8 MV 0.25 μm. Cette technologie présente une couche épitaxiée peu dopée P de 6 μm d'épaisseur sur un substrat P fortement dopé. Une isolation latérale est obtenue au moyen de tranchées profondes, et l'isolation verticale est assurée par une couche enterrée N (NBL).

Les critères auxquels doivent répondre les protections ESD pour satisfaire aux exigences de l'industrie automobile sont multiples [39]. Premièrement, la tension de déclenchement des composants de protection doit être facilement contrôlable pour pouvoir mieux s'adapter à l'étendue des domaines de tension intégrées sur la même puce. La variation statistique des grandeurs caractéristiques (V_{t1}, V_H, V_{t2}, I_{t2} etc.) sous différentes températures et orientations cristallines [40] doit être la plus faible possible. Le courant de fuite doit rester bas pour que les composants de protection n'interfèrent pas avec le circuit et ne consomment pas d'énergie à l'état bloqué. Une autre considération concerne leur immunité aux perturbations CEM, particulièrement sévères dans l'environnement automobile [41]. Enfin, le comportement transitoire des protections doit être examiné pour s'assurer que le pic de surtension présent à leur déclenchement ne soit pas un danger pour le CI.

Trois des composants que nous allons étudier dans cette section sont des SCR déclenchés par leur transistor NPN intrinsèque, et le dernier composant décrit est un PNP conçu par Amaury Gendron durant sa thèse [31].

1.6.1 SCR haute tension

Les SCR sont des composants très intéressants pour la protection des E/S haute tension, de par leur faible résistance à l'état passant et leur robustesse élevée [42] [43] [44]. Les trois composants que nous allons étudier présentent des similitudes (figures 1.16, 1.17, 1.18). Ils présentent deux caissons P implantés dans la couche épitaxiée P^-, et trois implantations N de part et d'autres de ces caissons P. Les caissons N, flottants, sont électriquement reliés les uns aux autres par la couche enterrée. Des diffusions peu profondes P^{++} et N^{++} sont effectuées dans les caissons P pour réaliser les contacts.

1.6.1.1 NPNR

Le composant $NPNR$ [45] est composé d'un NPN dont la jonction base-émetteur est déclenchée par une résistance externe R_{BE}, et d'un PNP imbriqué avec le NPN (figure 1.16). Chaque caisson P possède deux diffusions P^+ et N^+. Dans la configuration de la figure 1.16, le NPN a pour collecteur le caisson N central, pour base le caisson P situé du côté négatif (gauche), et pour émetteur la diffusion N^+ de ce caisson. Le PNP, lui, a pour collecteur la base du NPN, et pour base le collecteur du NPN. Son émetteur est constitué du caisson P relié au signal.

Le premier des deux bipolaires à se déclencher est le NPN, qui entraîne ensuite le déclenchement du PNP. La tension de claquage base-collecteur du NPN est bien supérieure à son V_{t1}. Par conséquent, la jonction émetteur-base ne peut pas être polarisée par un courant d'avalanche, c'est la chute de potentiel s'établissant aux bornes de la forte résistance R_{BE} qui permet à la jonction d'atteindre sa tension de seuil. La valeur de cette résistance détermine directement la tension V_{t1} de déclenchement du dispositif. R_{BE} peut varier de 0.5 $k\Omega$ à 60 $k\Omega$, permettant d'obtenir des valeurs de V_{t1} allant de 18 V à 30 V. Des valeurs plus élevées de la tension de déclenchement peuvent être obtenues par la mise en série de deux ou trois de ces composants.

Le fait que sa tension de claquage soit si élevée lui procure une marge de sûreté importante en régime statique. En revanche, du fait de l'importante résistance émetteur-base, le $NPNR$ est très sujet aux phénomènes de déclenchement par dV/dt (section 1.5.2.1) : sa tension de déclenchement peut dépendre de la pente de l'impulsion ESD. De plus, son déclenchement très lent cause l'apparition d'un large pic de surtension dans les premières nanosecondes des impulsions.

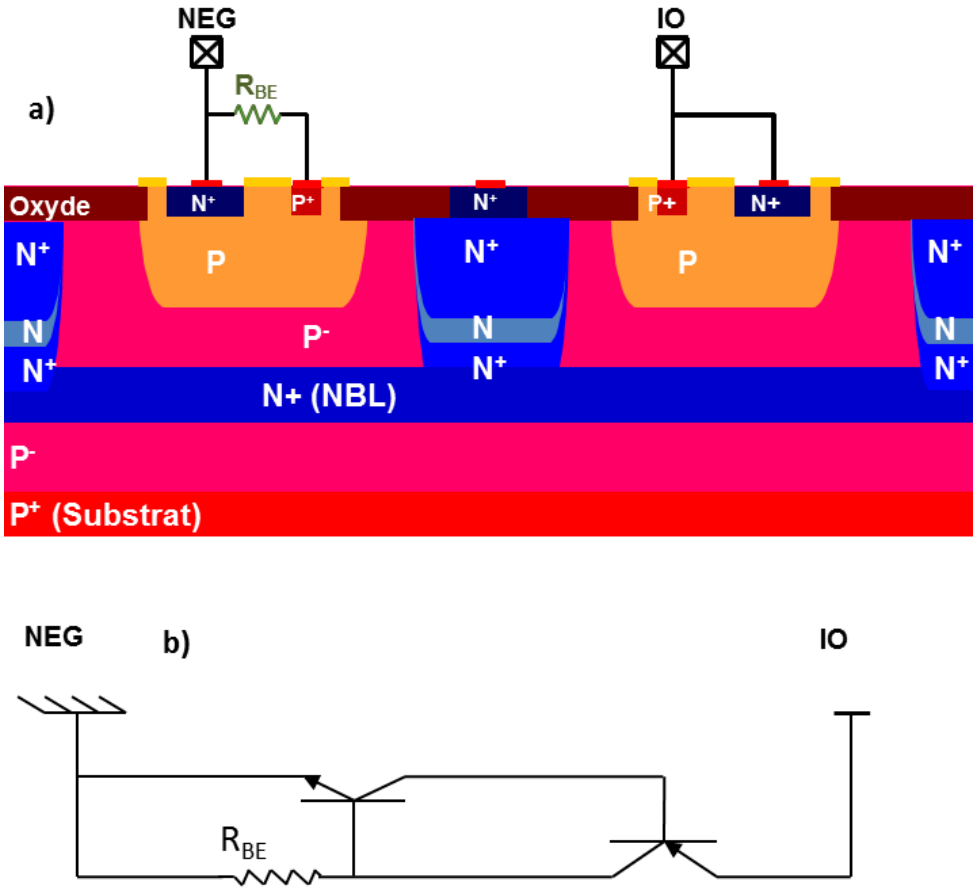

FIGURE 1.16 – Coupe technologique représentative d'un composant de protection $NPNR$ (a) et schéma équivalent (b).

Enfin, il a été démontré que le $NPNR$ était particulièrement susceptibles aux perturbations CEM. Pour toutes ces raisons, ce composant a été mis à l'écart de la librairie.

1.6.1.2 NPNZ

Le $NPNZ$ [46] a été conçu pour répondre aux insuffisances du $NPNR$. Comme on peut le voir sur la figure 1.17, sa structure est très similaire à celle du $NPNR$. Les caissons N latéraux ont été rapprochés des caissons P, et la résistance émetteur-base du PNP a été remplacée par un court-circuit (figure 1.17). Des diffusions P^+ ont été réalisées aux bordures extérieures des caissons P

Le déclenchement est initié par le claquage par avalanche de la diode « Zener 1 » (figure 1.17). Les trous générés par avalanche traversent le caisson P et sont évacués par le contact P^+ de base du NPN, relié à la broche NEG. Ce courant de trous circulant dans la base permet d'augmenter localement son potentiel et de polariser en direct la jonction émetteur-base du

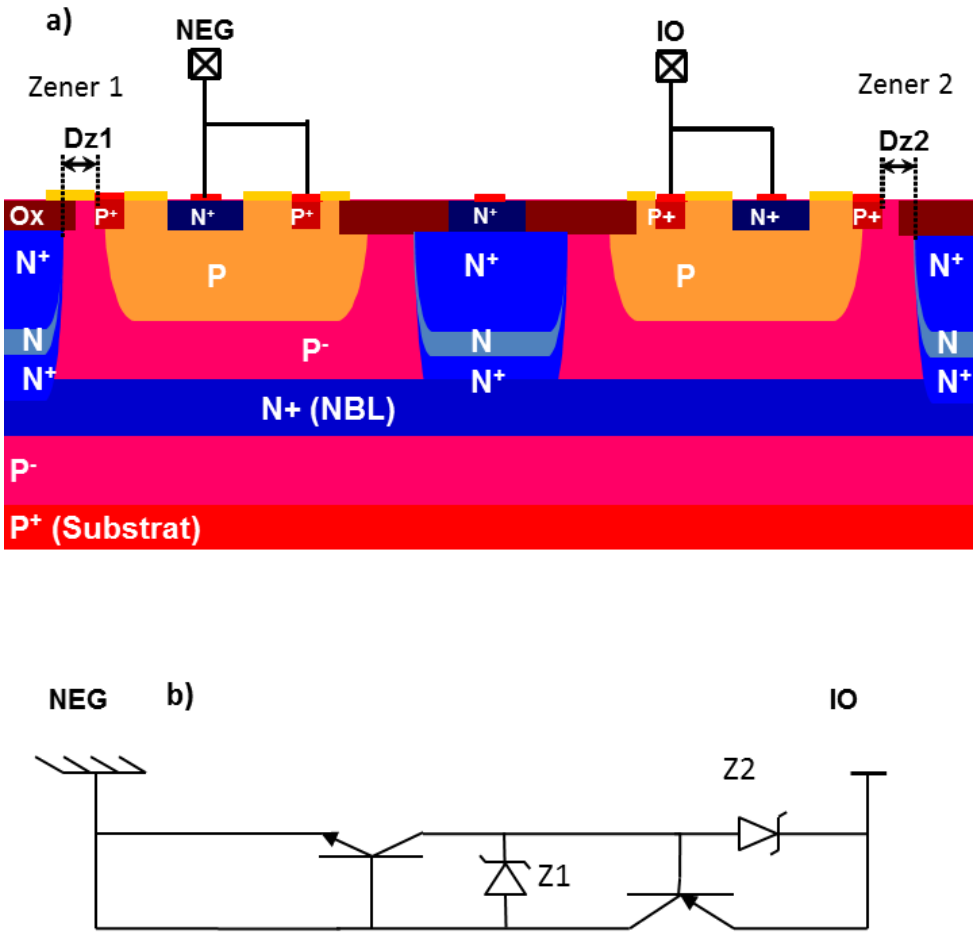

FIGURE 1.17 – Coupe technologique représentative d'un composant de protection $NPNZ$ (a) et schéma équivalent (b).

NPN. Le NPN se déclenche, puis un effet SCR se met en route avec le déclenchement consécutif du PNP. Une deuxième diode « Zener 2 » est utilisée également lors du déclenchement du NPN, pour dériver une partie du courant de décharge.

Contrairement au $NPNR$, le déclenchement par avalanche permet ici d'épargner au composant les principaux inconvénients liés au déclenchement par la résistance émetteur-base. La tension de repliement V_{t1} est très peu dépendante de la pente et de la largeur de l'impulsion. Aussi, la jonction émetteur-base du $NPNZ$ est moins sensible aux agressions CEM. En revanche, la tension de déclenchement V_{t1} varie très vite avec les espacements $Dz1$ et $Dz2$, et est donc assez sensible aux erreurs d'alignement de masques. De plus, les implantations choisies pour les caissons N et P dépendent de l'orientation du composant, et la tension de déclenchement se trouve donc légèrement modifiée par l'orientation du composant.

1.6.1.3 NPNB

Le $NPNB$ [47] [39] [48] est devenu le composant de référence dans la librairie E/S haute tension pour la protection contre les ESD. Il est composé, lui aussi, d'un NPN imbriqué avec un PNP. L'élément déclencheur est le claquage de la jonction base-collecteur du NPN, qui fournit le courant de polarisation de sa jonction émetteur-base et déclenche le NPN. Le courant fourni par le collecteur du NPN déclenche alors le PNP.

Ici, le rôle du NPN est assumé par le caisson P relié à la borne NEG, la diffusion N^+ implantée dans ce caisson, et le caisson N central, qui jouent respectivement le rôle de base, émetteur et collecteur (figure 1.18). Une fois ce NPN latéral déclenché, un déclenchement en NPN vertical vient s'ajouter au NPN latéral. Ce NPN vertical partage sa base et son émetteur avec le latéral, mais a pour collecteur la couche enterrée (NBL). Le PNP, lui, partage sa jonction base-collecteur avec le NPN latéral, et a pour émetteur le caisson P relié au signal. Le fonctionnement et en particulier le déclenchement du $NPNB$ sera étudié en détails dans le chapitre 5.

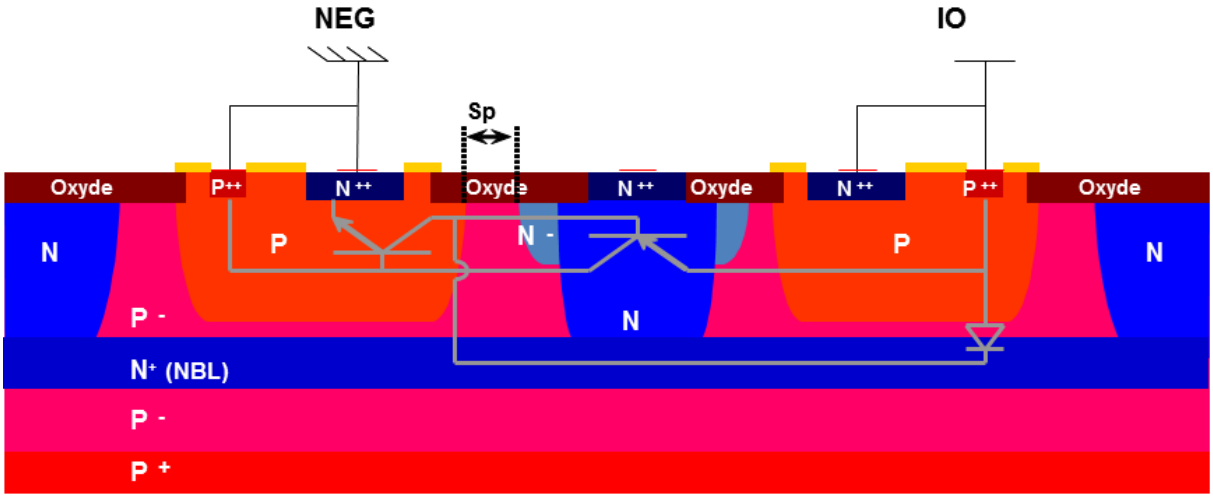

FIGURE 1.18 – Coupe technologique représentative d'un composant de protection $NPNB$.

La tension de déclenchement du composant est contrôlée par l'espacement « Sp » entre le caisson P côté NEG et le caisson N central. Une des avancées par rapport aux composants décrits précédemment est le retour relativement faible du PNP sur le NPN, qui permet d'assurer une tension de maintien particulièrement élevée, et d'accroître la marge de sûreté vis-à-vis du latch-up. Par comparaison, la surtension au déclenchement du $NPNB$ est également plus faible que celle du $NPNR$ et du $NPNZ$.

1.6.2 PNP/Diode

Le $PNP/Diode$ a été conçu par Amaury Gendron durant sa thèse [49] [31]. Il s'agit d'un PNP latéral placé en parallèle avec une diode verticale en inverse (figure 1.19). Le collecteur du PNP est constitué d'un large caisson P implanté dans la couche épitaxiée P^-. Un caisson N implanté dans la même couche épitaxiée joue le rôle de la base flottante, et un caisson P relié au signal, étroit et fortement dopé, joue le rôle de l'émetteur. La tension de déclenchement du PNP latéral est contrôlé par l'espacement Sp entre le caisson P de collecteur et N de base. Mais son R_{ON} est trop important pour qu'il soit utilisé seul, sans l'apport de la diode verticale. Cette diode se déclenche à 35 V, 55 V ou 60 V selon les implantations de collecteur employées. La meilleure configuration consiste à ajuster le V_{t1} latéral de manière à ce qu'il soit égal à la tension de déclenchement de la diode verticale. Le V_{t1} du $PNP/Diode$ ne peut donc prendre que les trois valeurs fixes des tensions de claquage de la diode verticale.

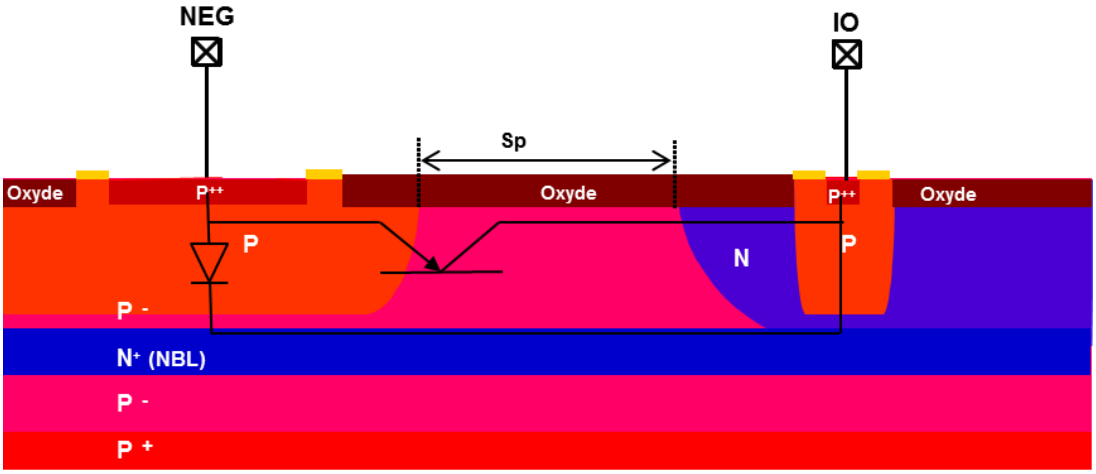

FIGURE 1.19 – Coupe technologique représentative d'un composant de protection $PNP/Diode$.

Une des caractéristiques essentielles de ce composant est son absence de repliement quasi-statique : V_{t1} et V_H sont confondus, et après déclenchement, la tension croît continument avec le niveau de courant. Cette tension de maintien élevée assure une très bonne marge face aux risques de latch-up. De plus, le $PNP/Diode$ présente une faible surtension au déclenchement sous forts « dV/dt ». En effet, avant le claquage, la diode verticale joue le rôle d'une capacité en parallèle qui dérive le courant de décharge et réduit l'amplitude du pic de surtension [50]. Cette capacité de jonction importante du $PNP/Diode$ a pour inconvénient un courant de fuite plus important et une plus grande consommation de puissance à l'état bloqué lors d'évènements de type EMC.

1.7 Conclusion

Dans ce chapitre, nous avons présenté le phénomène des décharges électrostatiques et son impact sur l'industrie de la microélectronique. Le développement de stratégies de protections internes contre les décharges électrostatiques est une étape cruciale de la conception des circuits intégrés.

Les composants de protection doivent pouvoir supporter des courants extrêmement intenses, pouvant aller jusqu'à plusieurs centaines de Watt sur quelques centaines de nanosecondes. Une fois déclenchés, les composants de protection doivent maintenir à leurs bornes une tension inférieure à la tension de claquage des oxydes. Leur tension de repliement ne doit toutefois pas chuter en-deçà de la tension d'alimentation du circuit, afin d'éviter les risques de latch-up. Enfin, la tension de décharge doit être repliée au plus vite par les composants de protection, pour éviter que la tension ne continue d'augmenter et ne représente un danger pour le circuit. Pour cette raison, la vitesse de déclenchement des composants de protection doit être optimisée.

Nous avons examiné les différentes normes industrielles associées aux types de décharges les plus fréquents. Ces normes fixent le cadre de la conception des stratégies de protection ESD.

Les outils de mesure du comportement des composants de protections soumis à des ESD ont été présentés. L'outil de mesure TLP, particulier, a été introduit comme élément d'étude quasi-statique. Nous reviendrons plus longuement sur ses limites dans le chapitre suivant, avant d'introduire une méthode de mesure alternative dans le chapitre 3.

Les composants de protection les plus couramment utilisés par les concepteurs ont été présentés. Un intérêt particulier a été apporté au transistor bipolaire autopolarisé, composant central des protections ESD haute tension.

Enfin, les composants utilisés à Freescale en technologie $SmartMos\ 8\ MV$ pour la protection des E/S haute tension ont été décrits. Le composant le plus performant et polyvalent d'entre eux est le $NPNB$, dont le déclenchement sera étudié au chapitre 5.

Chapitre 2

État de l'art sur la caractérisation transitoire des protections ESD haute tension

2.1 Introduction

La conception de protections ESD intégrées « haute tension » dans l'environnement sévère de l'industrie automobile rend nécessaire l'élaboration d'outils de caractérisation du comportement transitoire des composants pour les décharges rapides de type CDM et Gun stress. Jusqu'à présent, aucun dispositif de mesure n'existait pour accéder de façon précise au comportement au déclenchement des protections « haute tension ». Des dispositifs de mesures transitoires ont certes été rapportés ces dernières années [51], [52]. Mais l'utilisation nécessaire de générateurs d'impulsion « solid state » les rend inapplicables pour les hautes tensions, du fait de la limitation à 50 V de ces générateurs pour les plus évolués d'entre eux.

Nous allons analyser dans ce chapitre les approches classiques utilisées pour accéder au comportement transitoire des protections ESD, que sont la visualisation directe en $vf-TLP$ et l'approche appelée ici « indirecte ». Nous verrons les limites de ces deux approches.

2.2 Principe de fonctionnement du vf-TLP Oryx

2.2.1 Matériel employé

2.2.1.1 Générateur d'impulsion

Le banc de mesure TLP/vf-TLP commercial Oryx/Celestron utilisé au LAAS fonctionne selon le principe énoncé plus haut (1.4.1). Le générateur d'impulsion couvre une gamme de tensions allant de -1000 V à + 1000 V, avec des durées d'impulsions de 1.25 ns, 2.5 ns, 5 ns pour le vf-TLP, et 100 ns pour le TLP. Les impulsions ont un temps de montée très faible (inférieur à

la centaine de picosecondes), lié à la durée de la décharge ayant lieu dans le relai du générateur (annexe B). La résolution temporelle finie des oscilloscopes ne permet pas de caractériser cette durée de décharge, et donc le temps de montée de l'impulsion. Il est seulement permis d'affirmer qu'elle est inférieure ou égale à 100 ns, temps de montée mesuré par l'oscilloscope Tektronix TDS6604C utilisé dans ces travaux. Des filtres passe-bas à insérer à la sortie du générateur permettent de ralentir le temps de montée de l'impulsion, à des valeurs connues : 175 ps, 300 ps, 1 ns et 10 ns. En générant un temps de montée connu, ces filtres permettent de s'assurer que les fréquences contenues dans le spectre de l'impulsion ne sont pas plus élevées que la bande passante des différents éléments du banc de mesure, limitant ainsi les risques de distorsion et garantissant en principe l'intégrité du signal.

Comme on peut le voir à la figure 2.1, le générateur d'impulsion est formé par la mise en série du générateur statique haute tension, de la résistance de charge (10 $M\Omega$), ainsi que la ligne de charge (annexe B). Cette ligne est un câble coaxial de type $RG-401$ de Belden, conçu pour résister à de hautes tension statiques (3 kV). Les concepteurs du générateur ont inséré un atténuateur de 6 dB (division par deux en tension) en série avec la ligne, pour atténuer les réflexions parasites en provenance du composant sous test et retournant vers le testeur. Ces impulsions sont donc atténuées d'un facteur deux en entrant dans le générateur, réfléchies à nouveau sur la résistance de charge R du générateur, et à nouveau atténuées d'un facteur 2 en sortant du générateur. Cet atténuateur a donc pour effet de protéger le composant et le système de mesure contre les agressions multiples dues à des allers-retours de l'impulsion entre la résistance de charge et le composant testé.

2.2.1.2 Boîte de mesure (measurement pod)

L'impulsion produite par le générateur se propage ensuite à travers un câble coaxial de 90 cm (TL_1), et atteint la « boîte de mesure vf-TLP ». Comme le montre la figure 2.2, cette boîte de mesure contient trois éléments : une sonde de tension et de courant pour les mesures TLP, ainsi qu'un système permettant de mesurer la caractéristique statique I-V du composant entre deux impulsions, afin de vérifier son état.

La sonde de tension est un « té de prélèvement » (« pick-off tee »), constitué d'une résistance soudée sur le cœur de la ligne de transmission pour former un pont diviseur de tension (fig 2.3). La résistance R_{pick}, d'une valeur de 1.5 $k\Omega$, assure une division par $\frac{R_{pick}+50}{50} = 31$ entre le point de TL_1 où la tension est prélevée (point A), et le point B de TL_3, où elle est envoyée vers l'entrée d'impédance 50 Ω de l'oscilloscope. Seule une petite partie du signal est donc déviée du chemin de propagation principal formé par les lignes TL_1 et TL_2.

La sonde en courant est un transformateur de courant « CT6 », de la marque Tektronix, dont la bande passante s'étend de 250 kHz à 2 GHz. Le principe des transformateurs de courant repose sur l'induction magnétique créée par le passage du courant dans le cœur de la ligne

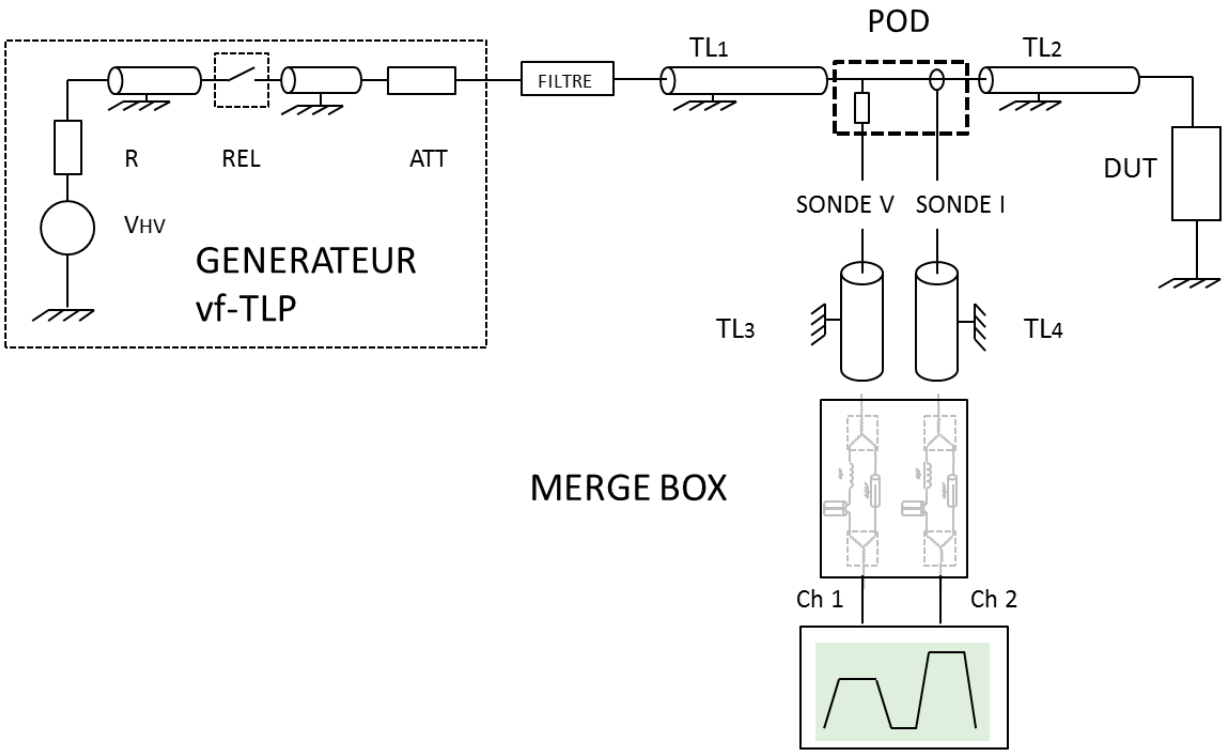

FIGURE 2.1 – Principe de fonctionnement du banc de mesure vf-TLP Oryx. Les lignes de transmission (« Transmission Line ») TL_1, TL_2, TL_3 et TL_4 sont représentées, ainsi que le filtre de temps de montée. Le rôle de la « merge box » sera détaillé plus loin.

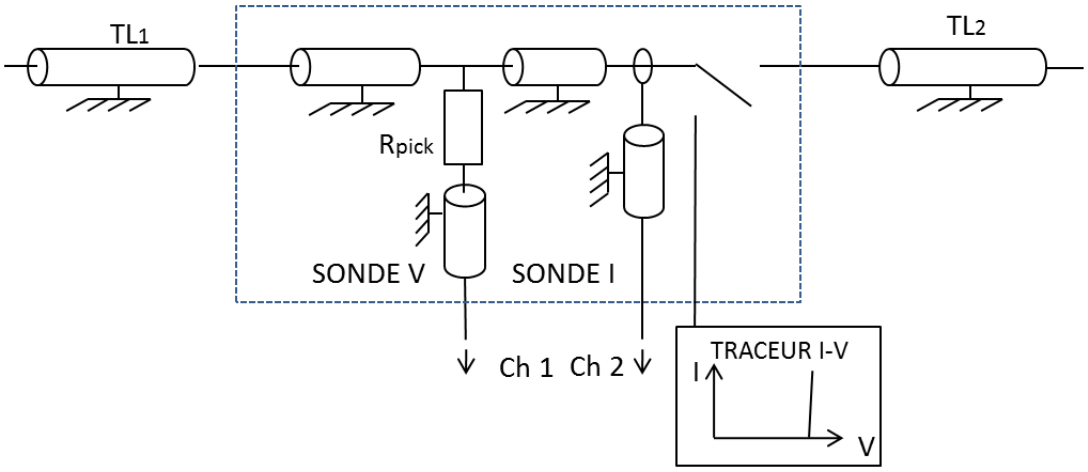

FIGURE 2.2 – La boîte de mesure contient un relai pour commuter entre la mesure TLP et le traceur I-V. Les sondes de courant et de tension sont placées en série.

de transmission. Ce champ magnétique induit lui-même un courant qui est recueilli dans un « circuit secondaire ». C'est ce courant qui transite à travers TL_4, vers l'oscilloscope (figure 2.1. Le coefficient de couplage entre les circuits primaires et secondaires, c'est-à-dire entre l'entrée et la sortie de la sonde est de 0.1, ce qui revient à une sensibilité de 5 Volt par Ampère, dans

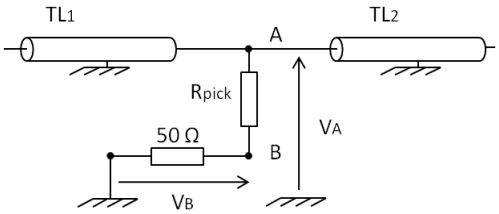

FIGURE 2.3 – Schéma équivalent de la sonde en tension.

notre cas, où le courant est collecté par une ligne d'impédance 50 Ω.

2.2.1.3 Câbles et pointes RF

La ligne TL_2 reliant la boite de mesure au composant sous test a une longueur de 90 cm. Cette longueur détermine le délai entre l'impulsion incidente et l'impulsion réfléchie (voir section 2.1.2.3.1), ce délai étant égal à la durée nécessaire pour réaliser un aller-retour dans la ligne TL_2, soit 9.1 ns.

Les mesures sous pointes s'effectuent à l'aide de pointes RF « picoprobes », produites par GGB, d'une bande passante de 18 GHz.

2.2.1.4 Oscilloscope

L'oscilloscope employé est un oscilloscope numérique TDS6604C de Tektronix, qui a une haute bande passante analogique (6 GHz), et un taux d'échantillonnage de 20 GS/s, permettant d'effectuer des acquisitions impulsionnelles en temps réel. Il est particulièrement adapté à l'observation de phénomènes transitoires rapides.

2.2.1.5 Merge box

Comme on l'a vu dans la partie 1.4.1, le système donne une image des impulsions incidentes et réfléchies sur le composant, séparées par un délai. La superposition entre l'onde incidente et l'onde réfléchie, qui a lieu sur le composant, doit donc être re-créée de façon à accéder aux tensions et courants réellement vus par le composant. La tension aux bornes du composant et le courant le traversant peuvent être calculés par les relations suivantes, après ré-alignement des impulsions incidentes et réfléchies dans le temps :

$$V_{DUT} = V_{inc} + V_{ref} \tag{2.1}$$

$$I_{DUT} = I_{inc} + I_{ref} \tag{2.2}$$

avec V_{inc}, I_{inc}, respectivement les tensions et courants incidents, et V_{ref}, I_{ref}, les tensions et courants réfléchis sur le composant.

La superposition entre l'onde incidente et l'onde réfléchie, qui a lieu sur le composant, doit donc être re-créée de façon à accéder aux tensions et courants réellement vus par le composant. Avec le logiciel permettant de visualiser les impulsions acquises à l'oscilloscope et de créer les courbes I-V quasi-statiques correspondantes, il est possible de sélectionner des fenêtres de temps à partir des impulsions incidentes et réfléchies, sur lesquelles un moyennage est fait. La somme ci-dessus est ensuite réalisée (équations 2.1 et 2.2), pour obtenir les tensions et courants quasi-statiques correspondants.

Cette méthode est limitée à l'extraction de caractéristiques quasi-statiques. Pour pouvoir extraire des formes d'ondes complètes, un dispositif a été inventé par Evan Grund [53] en 2006 pour effectuer ce ré-alignement. La merge box a pour effet d'introduire un délai dans la propagation du signal incident, de sorte que signaux incidents et réfléchis se retrouvent superposés sur l'oscilloscope.

Ce dispositif est construit selon le schéma suivant (figure 2.4) : il possède deux entrées sur lesquelles sont branchées les sondes de courant et de tension et deux sorties correspondantes. A l'entrée de chaque canal se trouve un diviseur de puissance, qui a pour effet de séparer le signal en deux composantes de même amplitude. Une des deux composantes transite alors à travers un système constitué d'une ligne à retard ayant un retard fixe de 9 *ns* en série avec une seconde ligne à retard, dont le retard est réglable. La première ligne est un câble coaxial de 180 *cm* de longueur. La seconde ligne est un composant de type « line stretcher », permettant d'introduire un délai variant de 0 à 630 ps avec une précision de 10 ps environ par quart de tour. La deuxième composante issue du diviseur de puissance n'est pas retardée. Les deux composantes sont alors dirigées vers un autre diviseur de puissance, analogue à celui cité plus haut, mais utilisé en combineur.

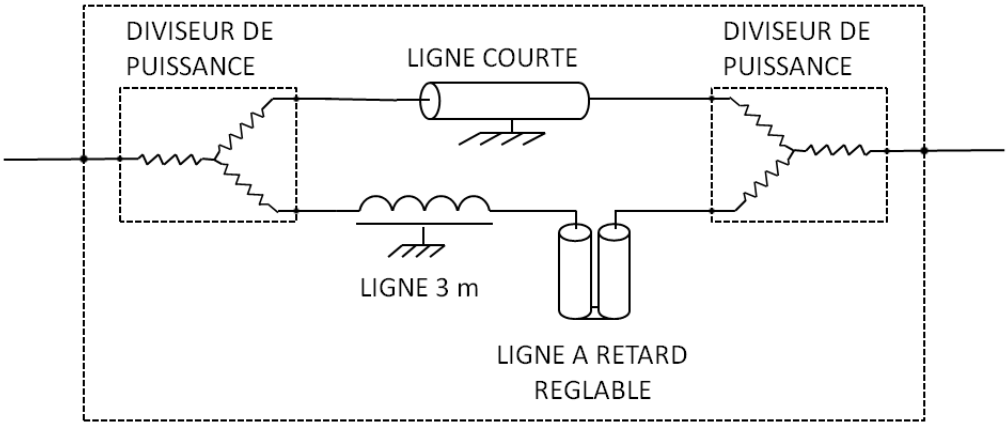

FIGURE 2.4 – Architecture d'un canal de la merge box. Les deux canaux sont identiques.

Ainsi, le signal entrant dans la merge box est divisé en deux signaux égaux séparés par un retard, puis re-combinés, de sorte qu'à la sortie du dispositif, ces deux signaux apparaissent sur le même câble coaxial, l'un à la suite de l'autre. A cause de l'atténuation par 6 dB des diviseurs de puissance, les impulsions mesurées en sortie de merge box présentent une amplitude divisée d'un facteur 4 par rapport aux impulsions arrivant à l'entrée. Pour obtenir les formes d'ondes $V(t)$ et $I(t)$, il est nécessaire de régler le délai de la merge box, de sorte que sur l'oscilloscope, l'impulsion incidente retardée s'affiche en même temps que l'impulsion réfléchie non retardée. La correction par un facteur 4 donne alors accès aux formes d'onde en courant et tension.

Ce délai est réglé manuellement pour chaque canal de la merge box, en mesurant la tension sur un court-circuit supposé parfait et le courant sur un circuit ouvert également parfait. En effet, dans les deux cas, la superposition des signaux incidents et réfléchis doit être nulle. On considère que le délai est bon quand le résultat de cette superposition est le plus faible.

2.3 Limites du système pour l'analyse transitoire.

Ce système est très bon pour l'analyse quasi-statique en vf-TLP. Dans ce cas, les valeurs des tensions et courants sont évaluées en effectuant une moyenne sur les dernières nanosecondes de l'impulsion. La fenêtre de moyennage peut être réglée par l'utilisateur pour éviter tout problème dû au désalignement des impulsions incidente et réfléchie. De plus, il est possible de s'affranchir partiellement de l'erreur due à l'atténuation du signal dans le système de mesure en effectuant un étalonnage complet du banc de mesure. Cet étalonnage peut être effectué entre -1000 V et +1000 V à partir de mesures sur un court-circuit et un circuit ouvert. La mesure des résistances série R_{ser} et parallèle R_{par} à chaque niveau de tension est utilisée comme facteur de correction. Le paramètre R_{ser} est déterminé par une mesure sur court-circuit, de sorte que la tension résultante soit nulle. R_{par} est calculé de manière analogue, en annulant le courant de fuite mesuré sur un circuit ouvert.

Bien que le vf-TLP ait été également proposé [53] pour l'analyse transitoire des protections ESD, nous allons voir dans les sections suivantes les nombreux facteurs d'erreur qui limitent la précision de la mesure transitoire. Parmi ces erreurs, la plus importante concerne le filtrage du signal par le banc de mesure lui-même lors de sa propagation depuis le générateur jusqu'à l'oscilloscope. Une seconde source d'erreur provient de la difficulté à régler visuellement le délai de la merge box.

2.3.1 Réponse fréquentielle des éléments du banc de mesure

La bande passante du banc de mesure est un paramètre à prendre en compte lors de la mesure de transitoires rapides. En effet, les différents éléments du banc de mesure tels que les

sondes de tension et de courant, mais aussi les câbles coaxiaux et la merge box présentent des caractéristiques d'atténuation fréquentielle susceptibles de modifier l'allure des impulsions mesurées de façon importante. De plus, le spectre typique des impulsions ESD peut s'étendre jusqu'à des fréquences bien supérieures à la fréquence de coupure de ces éléments.

La figure 2.5 représente les densités spectrales d'énergie (DSE) de trois impulsions $vf-TLP$ incidentes de même intensité mais ayant des temps de montée différents. La DSE d'un signal peut être vue comme l'analogue dans le domaine fréquentiel de la puissance instantanée dans le domaine temporel : l'intégration sur toute la bande de fréquence de la DSE est égale à l'énergie totale de l'impulsion. Une définition plus complète de cette grandeur est donnée dans annexe A.4. Ces caractéristiques sont assez similaires à la DSE d'une impulsion carrée parfaite. La courbe bleue en trait plein représente la DSE d'une impulsion $vf-TLP$ de 5 ns de largeur, sans aucun filtre de temps de montée à la sortie du générateur. Les courbes rouges pointillées et en tirets verts représentent les DSE de la même impulsion, mais après filtrage par les filtres passe-bas estampillés respectivement « 300 ps » et « 1 ns » (fréquences de coupure respectives de 1.2 GHz et 350 MHz).

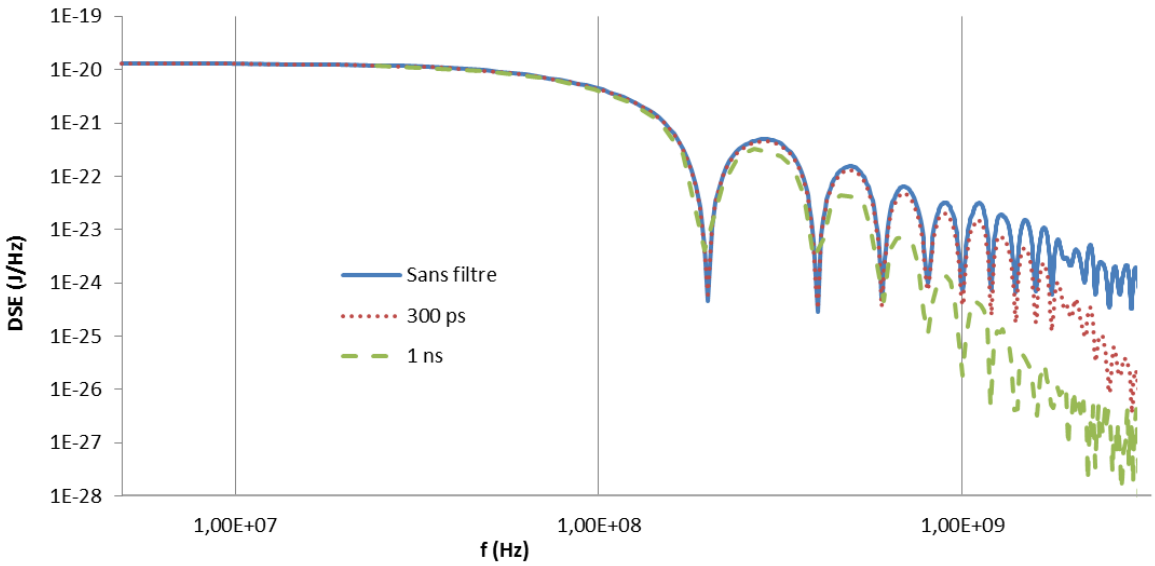

FIGURE 2.5 – Densités spectrales d'énergie d'impulsions ESD de type vf-TLP de 5 ns de largeur d'impulsion. Résultats obtenus à partir de FFT du signal mesuré à l'oscilloscope.

Ces courbes ont été obtenues par le calcul, à partir de mesures à l'oscilloscope. Elles sont le résultat de l'opération A.17 à partir de la transformation de Fourier rapide (FFT) de la tension $v(t)$ acquise à l'oscilloscope. On visualise ainsi la répartition de l'énergie sur toute l'étendue du spectre. Plus le temps de montée est court, plus les composantes fréquentielles élevées sont importantes. En effet, on peut vérifier par intégration des courbes sur la fréquence que 99.8 % de

l'énergie du signal sont concentrés dans les 500 premiers MHz dans le cas du temps de montée de 1 ns, alors que ce pourcentage tombe à 97.6 % dans le cas de l'impulsion non filtrée. Il est important de noter que l'impulsion « non filtrée » subit en réalité le filtrage de l'oscilloscope dont la bande-passante analogique est de 6 GHz.

2.3.1.1 Filtrage des sondes en courant et tension

Les sondes utilisées ont un effet de distorsion important sur le signal, contribuant pour beaucoup à son altération. La figure 2.6 montre les caractéristiques fréquentielles de transmission des sondes en courant et tension utilisées. On retrouve, en basse fréquence, le rapport de division par 31 (-30 dB) du pick-off tee constituant la sonde en tension, et la sensibilité de 5 mV/mA de la sonde en courant, correspondant à une transmission de 0.1, soit -20 dB. La sonde en tension peut être utilisée jusqu'à 3 GHz, ce qui est acceptable pour des impulsions standard (figure 2.6). En revanche, la fréquence de coupure (environ 2 GHz) de la sonde en courant est trop basse, et son atténuation trop importante en hautes fréquences pour qu'elle puisse être utilisée en $vf-TLP$ comme un outil de caractérisation transitoire précis.

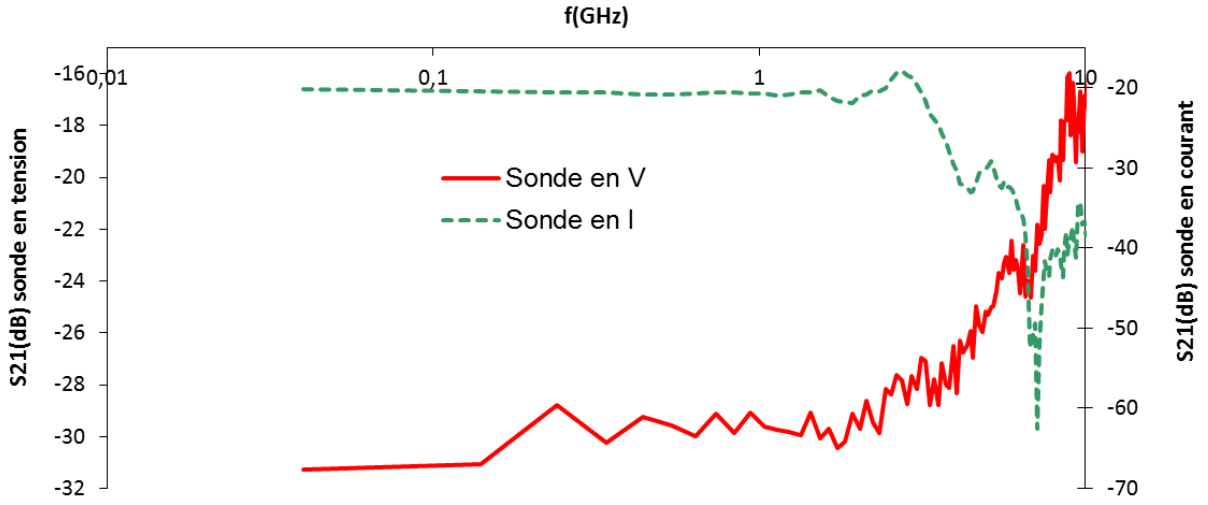

FIGURE 2.6 – Mesure de la transmission fréquentielle des sondes en courant et en tension utilisées. La sonde en courant a une fréquence de coupure de 2 GHz, et la sonde en tension de 3 GHz environ.

On note une augmentation du coefficient de transmission de la sonde en tension dans les hautes fréquences ($> 3\ GHz$), qui est dû à un effet de couplage capacitif entre l'entrée et la sortie du dispositif, celui-ci étant constitué d'une résistance de 1.5 kΩ soudée entre le cœur de la ligne de transmission formant l'« entrée » et de celle formant la « sortie ».

2.3.1.2 Atténuation apportée par les câbles coaxiaux

Le principal désavantage des câbles coaxiaux est un effet d'atténuation fréquentielle, due à la grande longueur de câble utilisée. L'atténuation, proportionnelle à la longueur de câble, augmente avec la fréquence : elle est de 0.6 dB/m à 1 GHz et 1 dB/m à 3 GHz (figure 2.7). Le filtrage résultant sur l'étendue du chemin de propagation est conséquent, particulièrement pour les composantes fréquentielles élevées. En effet, le signal arrivant au composant sous test est atténué par TL_1 puis par TL_2 (figure 2.1), soit $2 \times 91\ cm = 182\ cm$ de câble si on ne tient compte que de ces deux lignes de transmission. Il en résulte une atténuation de 1.1 dB à 1 GHz, et de 2.04 dB à 3 GHz (tableau 2.1). Cela se manifeste par une augmentation du temps de montée des impulsions.

Cette atténuation, du signal du générateur jusqu'au composant sous test, n'est pas un problème en soi sauf si l'on souhaite agresser le composant avec des temps de montée très courts. Le principal désagrément provient de la différence de forme d'onde entre le signal arrivant réellement au composant sous test, et le signal que l'on mesure à l'oscilloscope. En comparaison avec le signal incident au composant sous test, le signal « incident » mesuré à l'oscilloscope est atténué par TL_1, les câbles TL_3 ou TL_4 des sondes en tension ou courant selon le canal sélectionné, puis la ligne à retard de la merge box. Le signal est ensuite conduit vers une boîte d'atténuateurs (non représentée sur le schéma 2.1), et enfin vers l'oscilloscope. Cela représente 4.14 m de câbles, soit une atténuation de 2.6 dB à 1 GHz et 4.6 dB à 3 GHz. Le signal « réfléchi » mesuré ne suit pas le même chemin de propagation, mais parcourt une longueur de câble équivalente, de manière à ce que signaux incidents et réfléchis se superposent à la sortie de la merge box. L'atténuation subie par le signal réfléchi est donc du même ordre. La différence d'atténuation entre les signaux mesurés et les signaux vus par le composant est donc de 1.4 dB à 1 GHz et 2.6 dB à 3 GHz.

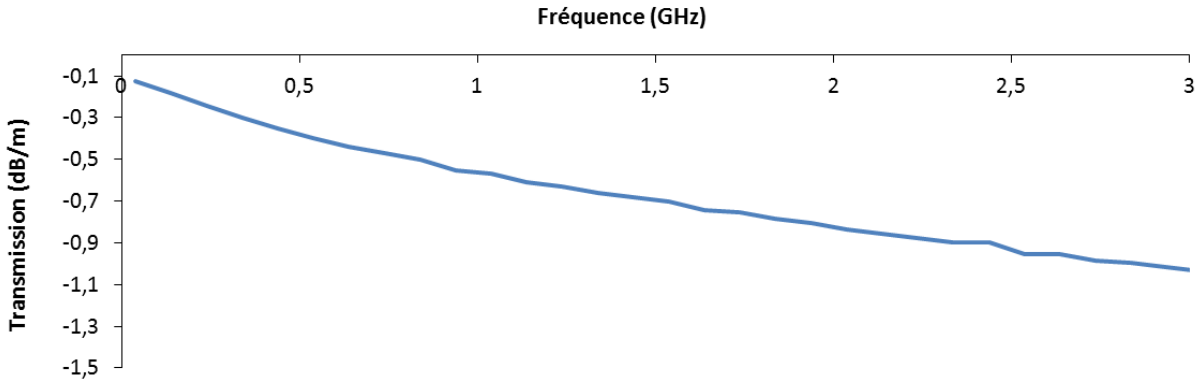

FIGURE 2.7 – Caractéristique de transmission fréquentielle des câbles coaxiaux utilisés dans le vf-TLP Oryx.

	1 GHz	2 GHz	3 GHz
$V_{inc,DUT}, V_{ref,DUT}(dB)$	1.1	1.7	2.0
$V_{inc,osc.}, V_{ref,osc.}(dB)$	2.6	3.8	4.6

TABLEAU 2.1 – Atténuation du signal à travers les câbles coaxiaux, à différentes fréquences. Comparaison entre l'atténuation du signal vu par le DUT et le signal mesuré.

2.3.1.3 Filtrage de la merge box

La merge box est également un élément de distorsion du signal. La documentation technique indique des pertes d'insertion allant jusqu'à 0.5 dB pour les fréquences inférieures à 4 GHz pour les diviseurs de puissance [54]. Les lignes à retard ont également un léger impact sur le signal.

2.3.1.4 Bilan du filtrage du banc Oryx

Domaine fréquentiel

La mesure de transmission fréquentielle du banc entier est représenté fig 2.8. La figure 2.8-a) montre les résultats obtenus avec la sonde en tension et la figure 2.8-b avec la sonde en courant. Le chemin de propagation correspondant à l'impulsion incidente, c'est-à dire issue du générateur et suivant le chemin retardé de la merge box est représenté en bleu foncé, et la transmission le long du chemin « réfléchi », issu du composant et non retardé est représenté en orange clair. Les caractéristiques de tension et courant incidents apparaissent assez proches, si ce n'est une fréquence de coupure légèrement inférieure pour le courant, due à l'influence de la sonde. Le chemin réfléchi semble présenter une coupure plus précoce que le chemin incident. Les pertes atteignent 4 et 5 dB à 2 GHz respectivement sur la mesure de tension incidente et réfléchie, et apparaissent moins élevées sur la mesure de courant.

Pour des raisons pratiques, il n'a pas été envisageable de mesurer directement les paramètres S du système de mesure entier (en incluant la merge box) à l'aide d'un analyseur de réseau vectoriel (VNA). Une des raisons est bien sûr l'encombrement du banc de mesure, mais la principale raison provient de la multiplicité des signaux générés par le système, en particulier lors du passage successif de l'impulsion dans les deux diviseurs de puissance. Ces données ont donc été obtenues sans outils extérieurs au banc de mesure, par une méthode qui sera détaillée et validée dans le partie 4. Cette technique est basée sur le traitement numérique par l'analyse de Fourier des impulsions ESD mesurées à l'aide du générateur Oryx et de l'oscilloscope Tektronix du banc de mesure.

L'ajustement analytique (courbes « fit ») a été réalisé sous « matlab » par des fonctions polynômiales, afin de s'affranchir des artéfacts de mesure (pics) et d'observer une tendance générale. Enfin, la transmission correspondant au cas sans pertes a été tracée, pour comparaison avec la correction uniforme qu'effectue le logiciel Oryx. Les 42 dB d'atténuation de la figure

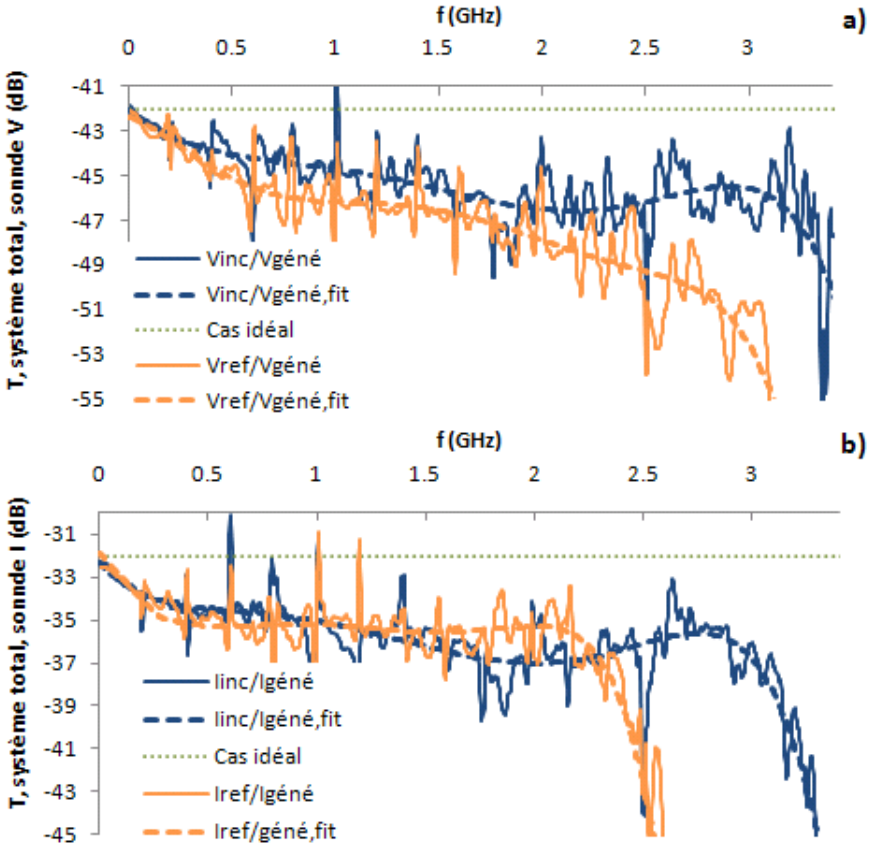

FIGURE 2.8 – Mesure de pertes fréquentielles du système de mesure vf-TLP, depuis le générateur jusqu'à l'oscilloscope. La transmission le long du chemin « incident » est représentée en bleu foncé, et le long du chemin « réfléchi » en orange clair.

2.8-a correspondent aux 30 dB de division du té de prélèvement (pick-off tee) constituant la sonde en tension et les 6 dB de séparation des deux diviseurs de puissance successifs. La sonde en courant quand à elle, apporte une atténuation de 20 dB, et on retrouve les mêmes 6 dB de séparation des diviseurs de puissance. Ces atténuations idéales sont prises en compte par le logiciel, qui multiplie les signaux mesurés par une valeur fixe. Les corrections quasi-statiques évoquées précédemment affinent ce réglage, mais seules les composantes fréquentielles les plus basses entrent en ligne de compte.

Domaine temporel

La figure 2.9 est une illustration dans le domaine temporel du filtrage évoqué dans les précédents paragraphes. Les figures 2.9-a et 2.9-b présentent respectivement les tensions et courants V_{CO} et I_{CC}, obtenus sur un circuit ouvert et sur un court-circuit pour une impulsion de 10 V TLP. Sur chacun de ces schémas, on a tracé une comparaison entre des courbes obtenues séparément. Les courbes orange clair montrent des mesures classiques avec tout le système, c'est-à-dire les sondes, les câbles et la merge box et les courbes bleu foncé des mesures de la tension et du

courant réel vus par le composant testé.

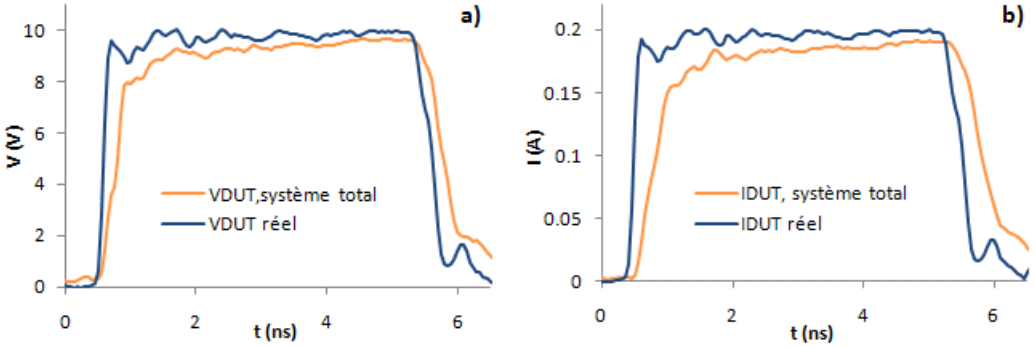

FIGURE 2.9 – Formes d'onde en tension et courant mesurées à l'oscilloscope. Le filtrage du chemin de mesure prend la forme d'une augmentation du temps de montée ainsi qu'une atténuation du pic de surtension.

Ces dernières courbes « réelles » ont été obtenues de la manière suivante : l'extrémité du câble TL_2 de la figure 2.1 a été reliée à l'oscilloscope. Le signal ainsi mesuré correspond alors à la tension incidente $V_{inc,DUT}$ qui arriverait au composant dans une configuration de test classique. Sur un circuit ouvert parfait, cette tension se réfléchit à l'identique, et la superposition de l'incident et du réfléchi donne une tension V_{DUT} égale à deux fois l'impulsion incidente. Sur un court-circuit, ce n'est pas la tension mais le courant incident $I_{inc,DUT} = \frac{V_{inc,DUT}}{50}$ qui se réfléchit à l'identique, et I_{DUT} est alors égal au double du courant incident. Notre mesure de $V_{inc,DUT}$ permet donc de calculer la tension réelle de circuit ouvert et le courant réel de court-circuit en utilisant les relations :

$$\begin{cases} V_{circuit\ ouvert} = 2.V_{inc,DUT} \\ I_{court\ circuit} = 2.\frac{V_{inc,DUT}}{50}. \end{cases} \tag{2.3}$$

Comme dans le logiciel, la tension mesurée avec le système a été multipliée par un facteur $31 \times 4 = 124$ pour compenser l'atténuation de la sonde et de la merge box. Le courant lui, a été multiplié par $\frac{10\times4}{50} = 0.8$ pour la même raison, en divisant par l'impédance de la ligne pour convertir la tension mesurée à l'extrémité de TL_4 en courant.

L'effet de filtrage est bien visible sur les courbes « système total » : l'impulsion réellement vue par le composant sous test présente un pic de surtension (composantes fréquentielles élevées), qui disparait totalement des mesures lors de la propagation de l'impulsion à travers le banc de test. Les courbes obtenues se rapprochent alors de courbes du type « charge de circuit R-C ». On note que la valeur quasi-statique perd également 3.4 % sur l'impulsion en tension et 5.6 % sur l'impulsion en courant.

2.3.2 Réglage du délai de la merge box

Comme on l'a vu, l'ajustement du délai du canal en tension de la merge box peut être réalisé en mesurant la tension de court-circuit. Une vis micrométrique permet un ajustement très précis, de l'ordre de 5 ps pour un quart de tour de la molette. On considère que le réglage est effectué quand les pics positifs et négatifs résultant d'une superposition incomplète des impulsions incidentes et réfléchies s'annulent.

Cependant, une partie du signal est noyée dans le bruit de mesure, ce qui peut rendre difficile la tâche de détection des pics. Plus gênant encore, le générateur TLP est construit de telle sorte qu'une tension résiduelle lentement décroissante est souvent présente parfois longtemps après la fin de l'impulsion (typiquement 6 ns environ pour une impulsion de 5 ns). Cette « traine » peut atteindre jusqu'à 10 % de l'amplitude de l'impulsion. Elle se superpose au bruit de mesure, et les deux phénomènes ensemble contribuent à perturber l'évaluation du délai.

A titre d'exemple, considérons une impulsion de 100 V d'amplitude, altérée par un bruit V_B atteignant $\pm$ 1 V, avec une tension résiduelle introduisant une incertitude V_R de 10 %, soit 10 V, et un temps de montée de 1 ns. L'incertitude sur la lecture de la tension atteint donc 11 V. En terme de précision temporelle, cela revient à un délai de réglage minimum Δt de :

$$\Delta t = \frac{V_R + |V_B|}{dV/dt} = 110\ ps \tag{2.4}$$

Il est donc impossible, dans cet exemple, d'être plus précis que 110 ps dans le réglage du délai.

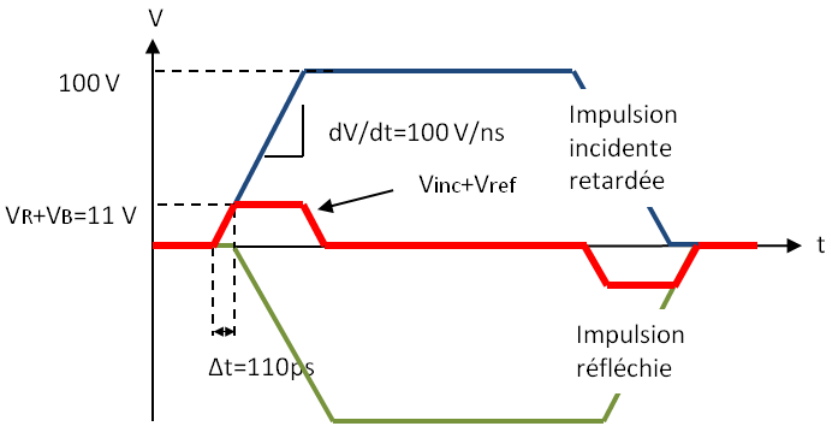

FIGURE 2.10 – Réglage du délai de la merge box. En bleu, l'impulsion incidente, et en vert l'impulsion réfléchie. La superposition des deux impulsions se limite à deux pics que l'utilisateur tente d'annuler.

2.4 Méthode de mesure « indirecte » de la surtension au déclenchement : couplage avec un composant à protéger.

Outre la compréhension des mécanismes dynamiques à l'œuvre lors du déclenchement des structures de protection ESD, le principal objectif de l'analyse transitoire est de déterminer la capacité des protections à ne pas dépasser les fenêtres de tension imposées par les concepteurs de circuits, et particulièrement dans les premières nanosecondes de l'impulsion. En effet, pour de courts temps de montées de l'impulsion incidente, il arrive que le composant n'ait pas le temps d'entrer en mode de basse impédance et de se replier avant que la tension de destruction des composants à protéger ne soit atteinte. Ce retard au déclenchement du composant de protection se traduit par un pic de surtension qu'il est important de surveiller, car dans certaines conditions, il peut à lui seul induire le claquage d'un oxyde de grille.

L'observation directe en $vf - TLP$ n'est pas adaptée à la mesure de pics de surtension au déclenchement d'une protection, car ce pic survient précisément pendant le front de montée de l'impulsion. Cela correspond au moment où les signaux varient le plus vite, et donc où les composantes fréquentielles élevées sont cruciales. Les pics mesurés risquent donc d'être peu fidèles à la réalité. De plus, cet instant est également celui pour lequel se manifeste l'effet du désalignement des impulsions incidentes et réfléchies évoqué dans la section précédente. Dans l'exemple cité plus haut,(équation 2.4), une erreur de 110 ps, (la précision du réglage) dans l'ajustement provoque une incertitude de 11 V sur les 100 V de l'impulsion.

2.4.1 Principe de la mesure indirecte de pics de surtension

Une technique très simple pour évaluer l'amplitude de cette surtension est de coupler la protection à étudier avec un composant à protéger placé en parallèle, si possible dans une configuration qui le rend particulièrement vulnérable à une agression par des pics très courts, c'est-à-dire qu'il se détruit instantanément s'il se déclenche.

Ainsi, selon les niveaux de tension et courant de destruction des deux composants, et la tension de claquage BV du composant à protéger, on peut avoir trois configurations :

- ***Cas 1***

  La protection est endommagée en premier. Cela signifie que pour tous les niveaux de courant inférieurs au courant de destruction de la protection, celle-ci a convenablement rempli sa tâche. Le pic de surtension n'est donc pas une menace.

- ***Cas 2***

  Le composant à protéger est endommagé en premier, à une tension quasi-statique supérieure ou égale à sa tension de destruction V_{t2}. Dans ce cas, sa destruction est prévisible

si l'on connaît la caractéristique TLP des deux composants. Elle n'est pas causée par un pic de surtension, mais par une trop forte résistance dynamique R_{ON} de la protection. Il s'agit d'un phénomène quasi-statique.

– ***Cas 3***

Le composant à protéger est endommagé en premier, mais à une tension quasi-statique inférieure à sa tension de destruction V_{t2}. Selon une logique quasi-statique, il est supposé rester intact. On en déduit l'existence d'un pic de surtension dont l'amplitude est au moins égale à la différence entre la tension quasi-statique V_{t2} de déclenchement du composant et celle pour laquelle il est effectivement détruit.

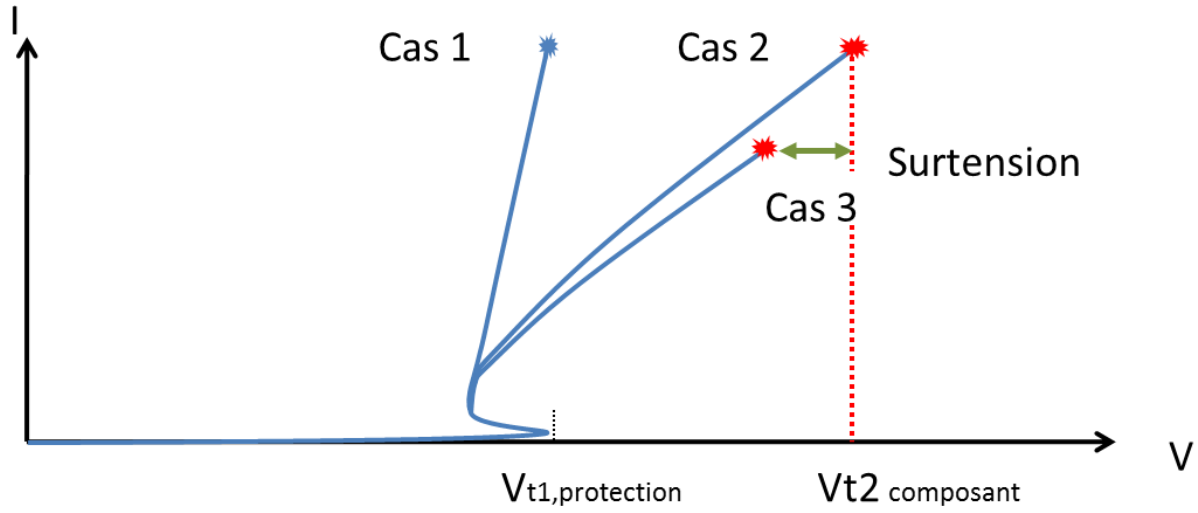

FIGURE 2.11 – Principe de la mesure « indirecte » du pic de surtension.

2.4.2 Application à l'étude de la surtension au déclenchement de protections ESD à base de transistors PNP

Cette technique de caractérisation a été utilisée pour déterminer le comportement au déclenchement de protections développés par A. Gendron lors de sa thèse [31]. Ces composants, appelés « PNP/Diode », sont formés par la mise en parallèle au sein du même composant d'un transistor bipolaire PNP latéral avec un PNP vertical. Le transistor vertical peut cependant être assimilé à sa diode base/collecteur en inverse, du fait du très faible rapport d'injection de sa jonction émetteur/base. Une structure $PNP/Diode$ se déclenche lors du claquage par avalanche de la jonction verticale base/collecteur, qui survient à 35 V (chapitre 1.6.2).

La protection est testée dans cet exemple avec le composant à protéger pour lequel il a été conçu, un composant de puissance $NLDMOS$ dont le claquage par avalanche survient aux alentours de BV = 60 V dans la configuration de la figure 2.12. Le $NLDMOS$ est en effet stressé entre le drain et la source, avec un stress ESD positif sur le drain, cette configuration étant

favorable au déclenchement du transistor NPN parasite si la tension appliquée est suffisante. Le déclenchement du $NLDMOS$ sera étudié plus en détails dans le chapitre 5.4. Il est important cependant de noter que ce composant n'est pas conçu pour fonctionner en avalanche, et le claquage de la jonction base/collecteur de son NPN parasite entraine sa destruction en moins d'une nanoseconde. Une résistance R_G d'une valeur variant de 1 $k\Omega$ jusqu'à 10 $k\Omega$ est insérée entre le contact de grille et la source. Une autre caractéristique de ce composant est sa sensibilité à la valeur de la résistance R_G et au temps de montée de l'impulsion ESD. Une forte résistance de grille et une forte pente $\frac{dV}{dt}$ de l'impulsion entrainent une réduction significative de la tension de déclenchement drain/source V_{t1} du composant.

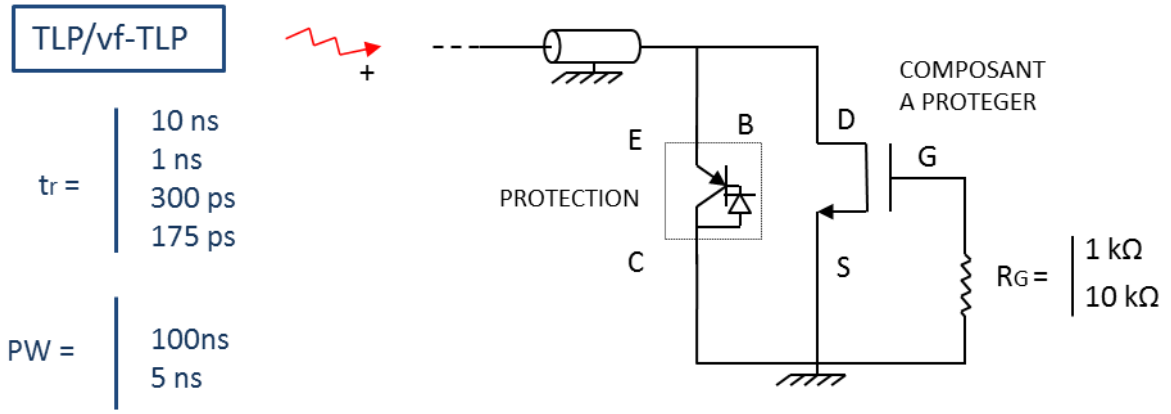

FIGURE 2.12 – Test en conditions de l'efficacité de protection du $PNP/Diode$ pour une entrée 50 V.

Le test a été effectué pour des tensions positives croissantes, jusqu'à ce que l'un des deux composants soit détruit. Le composant détruit est identifié par une observation au microscope, et par des mesures complémentaires I-V statiques. Différentes conditions ont été analysées, comme la durée de l'impulsion (PW), le temps de montée, t_r, et la résistance de grille R_G. La durée d'impulsion varie entre 5 ns en $vf-TLP$ et 100 ns en TLP. Les temps de montée ont pour valeurs 175 ps, 300 ps, 1 ns et 10 ns. Les mêmes mesures ont été répétées sur plusieurs échantillons différents afin d'écarter les artéfacts. Les résultats sont reportés dans le tableau 2.2. Les différents cas évoqués plus haut correspondent à des conditions particulières.

Le cas n^o 1, dans lequel le composant de protection $PNP/Diode$ est détruit en premier, est obtenu pour des temps de montée longs (1 ns, 10 ns), qui ne provoquent pas de surtensions importantes de la part du $PNP/Diode$, et ne déclenchent pas le NLDMOS. Ainsi, ce dernier est protégé efficacement. Il est transparent pour le $PNP/Diode$, qui atteint donc son courant de claquage thermique I_{t2}, à 7 A environ pour des impulsions de 100 ns.

Le cas n^o 2 (déclenchement « quasi-statique » du $NLDMOS$) est obtenu pour des impulsions d'une durée plus courte, pour lesquelles le risque de détruire le $PNP/Diode$ est beaucoup

plus faible ($I_{t2} = 16\ A$). On atteint donc la tension de déclenchement et de claquage thermique du $NLDMOS$.

Le cas n^o 3 est obtenu dans une configuration similaire à la précédente, mais avec des temps de montée courts et des résistances de grilles élevées, qui rendent le NLDMOS beaucoup plus susceptible aux pics de surtension provenant du PNP/Diode. Du point de vue du PNP/Diode, cette configuration n'est pas différente de la précédente, et la surtension générée est donc la même. Cependant, si elle est invisible dans le cas 2, elle est ici capable de déclencher le NLDMOS du fait de la forte résistance de grille, qui abaisse sa tension de déclenchement.

	Cas 1	Cas 2	Cas 3
PW	100 ns	5 ns	5 ns, 100 ns
t_r	1 ns, 10 ns	175 ps à 10 ns	175 ps, 300 ps
R_G	1 $k\Omega$, 10 $k\Omega$	1 $k\Omega$	10 $k\Omega$

TABLEAU 2.2 – Bilan des configurations pour lesquelles les différents cas évoqués ont lieu sur la protection PNP/Diode et le composant NLDMOS. Le cas 3 n'est observé que pour des temps de montée faibles et des fortes résistances de grille.

La figure ci-dessous montre les courbes TLP obtenues dans une configuration correspondant au cas n^o 3.

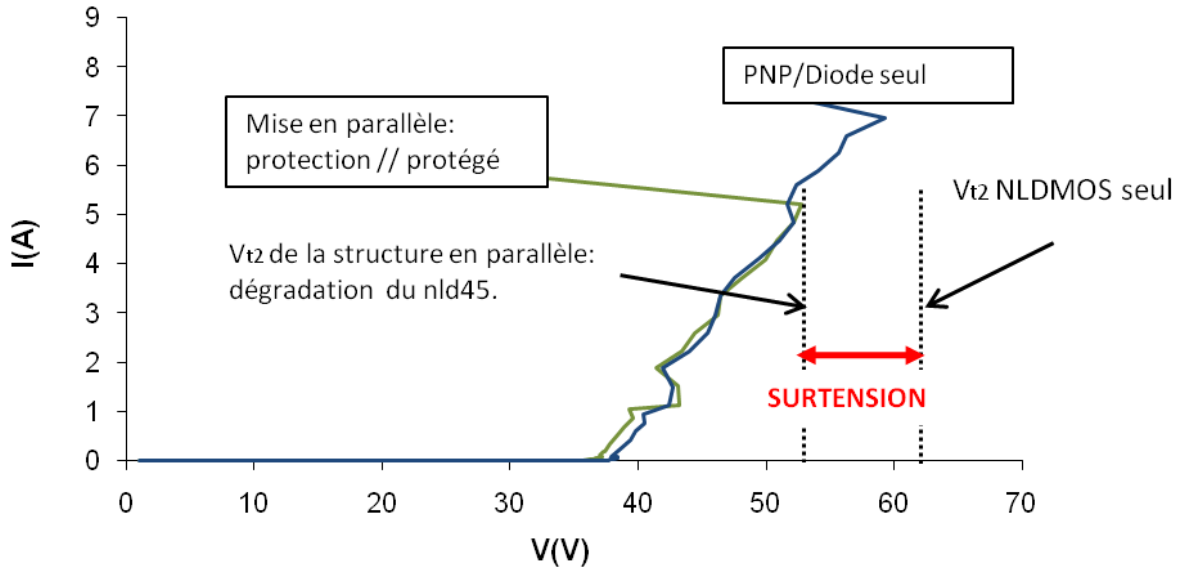

FIGURE 2.13 – Calcul de la surtension générée par le PNP/Diode à l'aide d'une mesure TLP indirecte. La largeur d'impulsion est de 100 ns, le temps de montée de 300 ps, et la résistance de grille est de 10 $k\Omega$. La caractéristique TLP de la structure couplée protection//NLDMOS est représentée en vert, et la comparaison avec une protection seule en bleu. La superposition des deux courbes montre que le NLDMOS ne se déclenche pas avant d'atteindre sa tension de dégradation.

2.4.3 Limites de la méthode de mesure indirecte des pics de surtension

Un avantage de cette approche est qu'elle peut servir à la fois de méthode d'investigation, pour évaluer les éventuelles surtensions, mais aussi de test de validation, car la protection est placée directement dans la configuration d'un CI à protéger. Mais l'usage reste limité à cause notamment de la difficulté de mise en œuvre de la mesure. En effet, il s'agit de comparer des tensions mesurées à très bas courant (tensions de déclenchement de composants seuls) avec d'autres mesurées à fort courant (plusieurs Ampere pour le déclenchement de composants à protéger placé en parallèle avec la protection). Une telle comparaison nécessite un très bon étalonnage du banc de mesure, qui doit être effectué avant chaque série de mesure.

De plus, dans le cas précédent comme dans de nombreux cas, la tension de déclenchement du composant à protéger varie en fonction de la pente du front montant de l'impulsion ESD (phénomène de « $\frac{dV}{dt}$ triggering »). Or, à temps de montée égal, cette pente est proportionnelle à la tension de l'impulsion envoyée. Il est donc délicat de reproduire les conditions de déclenchement du composant à protéger en tenant compte de cette pente une fois le composant placé en parallèle avec la protection.

Enfin, la mise en œuvre de cette méthode est coûteuse du fait de la nécessité d'ajouter, dans les véhicules de test silicium, des structures couplées de chaque protection avec l'ensemble des composants à protéger. De plus, la qualification d'une protection se trouve retardée par le nombre de tests à effectuer, d'autant qu'il est nécessaire de répéter les mesures plusieurs fois afin de déterminer la précision du résultat.

2.5 Conclusion

Nous avons vu dans cette section les méthodes les plus couramment utilisées pour évaluer le pic de surtension atteint par les protections ESD lors de leur déclenchement. La visualisation en $vf-TLP$ est assez simple à utiliser, mais a pour désavantage principal un manque de précision dans l'alignement des impulsions incidentes et réfléchies, ainsi qu'une importante distorsion des signaux mesurés par les sondes de tension et courant. Les mesures indirectes ne sont pas à abandonner totalement. Il est intéressant de pouvoir tester les protections ESD en même temps que les composants à protéger. Cela permet notamment de déterminer l'éventuelle tolérance que peut avoir le composant à protéger lorsque sa tension de seuil est dépassée pendant une durée très courte, ce qui arrive fréquemment lorsque la protection atteint son pic de surtension.

Chapitre 3

Nouvelle méthodologie basée sur la correction des données : le $transient - TLP$ (t-TLP)

3.1 Introduction

La solution que nous avons développée au LAAS reprend le principe du $vf - TLP$ en y intégrant un traitement mathématique des données mesurées à l'oscilloscope. La précision de la mesure s'en trouve grandement améliorée, si bien que la limitation de la mesure correspond alors aux 6 GHz de bande passante analogique de l'oscilloscope, soit à une résolution de 58 ps dans le domaine temporel. Cette méthode s'applique tant à l'extraction précise de formes d'onde V(t) et I(t) qu'à l'obtention de graphes $I - V$ TLP et $vf - TLP$ complets, comme on pourra le voir.

Nous allons voir dans une première partie le principe de la méthodologie, puis nous aborderons le traitement qu'il est nécessaire de faire subir aux signaux pour pouvoir l'appliquer. Des résultats de mesure sur quelques composants seront présentés. Enfin, une méthode d'étalonnage plus simple à utiliser sera présentée.

3.2 Principe de la méthode

Le dispositif de mesure a été réduit aux éléments mentionnés sur la figure 3.1. Le générateur d'impulsions $vf - TLP$ est conservé, ainsi que l'oscilloscope. On écarte la boîte de mesure « measurement pod » pour la remplacer par sa sonde en tension, s'affranchissant ainsi des parasites apportés par la sonde en courant et le relai (figure 2.2). La sonde en courant n'est plus nécessaire, car comme on va le voir, les formes d'onde en courant se calculent à partir des formes d'onde en tension. La merge box non plus n'est plus nécessaire et est retirée, ainsi que la boîte d'atténuateurs, contenant beaucoup d'éléments susceptibles de filtrer le signal. Les atténuateurs sont insérés manuellement à l'entrée de l'oscilloscope. Trois câbles coaxiaux TL_1, TL_2 et TL_3,

acheminent l'impulsion respectivement vers la sonde en tension, le dispositif sous test (DUT) et l'oscilloscope.

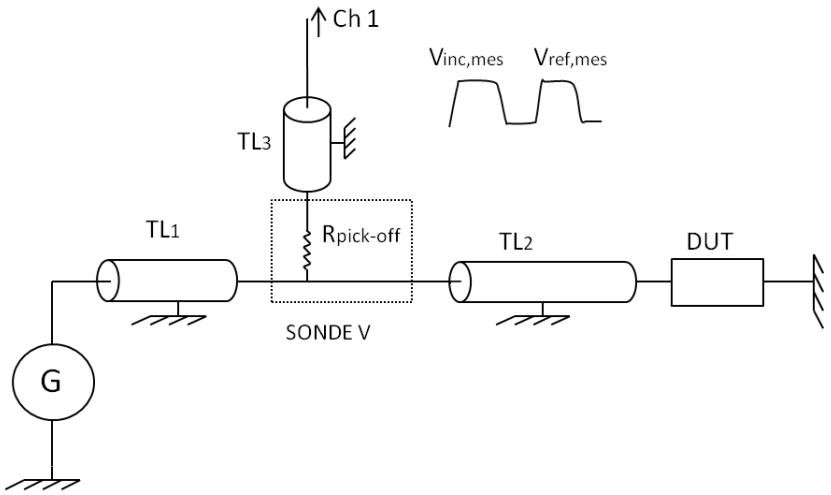

FIGURE 3.1 – Banc de mesure simplifié utilisé pour les mesures $transient - TLP$.

Cette méthode est inspirée de la technique proposée par David Trémouilles sur un banc de mesure TLP [55], à la différence essentielle que la séparation temporelle des impulsions incidentes et réfléchies en $vf - TLP$ modifie profondément le traitement à appliquer aux données. L'emploi de pointes RF à bande passante élevée simplifie également le problème, car on s'affranchit ainsi des parasites apportés par les pointes classiques, qui ont l'inconvénient d'être particulièrement inductives.

On rappelle que le signal mesuré à l'oscilloscope est constitué d'une impulsion incidente et d'une impulsion réfléchie, qui sont la conséquence du passage de l'impulsion issue du générateur pour la première et de celle réfléchie sur le composant testé pour la seconde. La tension vue par ce dernier se trouve être la somme des deux impulsions dans le plan du composant. Dans le plan de l'oscilloscope en revanche, les deux impulsions apparaissent séparées par un délai, correspondant à la propagation aller-retour dans la ligne TL_2 (figure 3.1). La tension vue par le composant est alors obtenue en combinant correctement ces ondes incidente et réfléchie.

Le principe est le suivant : le système de mesure peut être considéré comme un système RF à trois ports (figure 3.2), le port 1 étant situé à la sortie du générateur, ou de la ligne TL_1, le port 2 à la sortie de TL_2 et le port 3 à l'entrée de l'oscilloscope. On peut mesurer les paramètres S du système ainsi constitué en utilisant un analyseur de réseau vectoriel(VNA). Les informations d'atténuation fréquentielle et de déphasage sont entièrement contenues dans les paramètres de transmission S_{31}, S_{32} et S_{21} correspondant aux chemins pris par l'impulsion en $vf - TLP$.

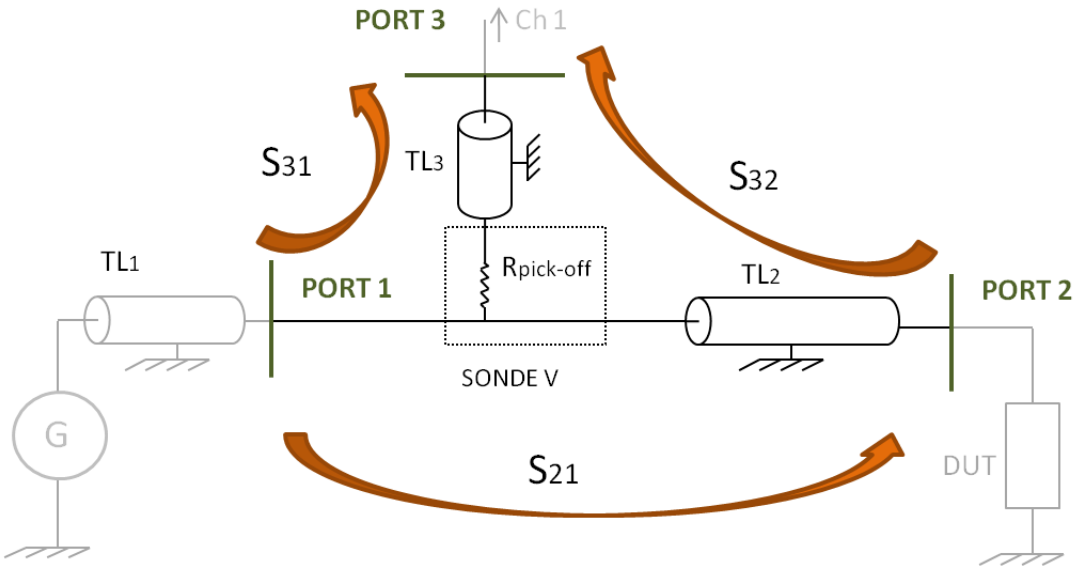

FIGURE 3.2 – Principe de la méthode de mesure $transient - TLP$: mesure des trois paramètres de transmission S_{31}, S_{32} et S_{21}.

Ces coefficients peuvent ensuite être utilisés pour recalculer les formes d'onde réellement vues par le composant sous test. Ce calcul, simplement réalisé dans le domaine fréquentiel, est ensuite transposé dans le domaine temporel grâce à l'analyse de Fourier [56]. Dans le domaine fréquentiel, l'impulsion issue du générateur $V_{gen}(f)$, les tensions incidentes et réfléchies vues par le composant, $V_{inc,DUT}(f)$ et $V_{ref,DUT}(f)$ sont obtenues à partir des impulsions incidentes et réfléchies acquises à l'oscilloscope, $V_{inc,oscillo.}(f)$ et $V_{ref,oscillo.}(f)$, et des coefficients de paramètres S, S_{31}, S_{32} et S_{21} :

$$V_{inc,DUT}(f) = S_{21}.V_{gen}(f). \tag{3.1}$$

Avec

$$V_{gen}(f) = \frac{1}{S_{31}}.V_{inc,oscillo.}(f), \tag{3.2}$$

on en déduit :

$$V_{inc,DUT}(f) = \frac{S_{21}}{S_{31}}.V_{inc,oscillo.}(f). \tag{3.3}$$

De même,

$$V_{ref,DUT}(f) = \frac{1}{S_{32}}.V_{ref,oscillo.}(f). \tag{3.4}$$

Avec (équation 1.2)

$$V_{DUT}(f) = V_{inc,DUT}(f) + V_{ref,DUT}(f), \tag{3.5}$$

on obtient la tension dans le plan du composant, dans le domaine fréquentiel :

$$\boxed{V_{DUT}(f) = \frac{S_{21}}{S_{31}}.V_{inc,oscillo.}(f) + \frac{1}{S_{32}}.V_{ref,oscillo.}(f).} \tag{3.6}$$

Les courants incidents et réfléchis sont liés aux tensions correspondantes par les relations suivantes, avec Z_C l'impédance caractéristique des câbles coaxiaux ($Z_C = 50\Omega$) :

$$I_{inc,DUT}(f) = \frac{V_{inc,DUT}(f)}{Z_C} \tag{3.7}$$

et

$$I_{ref,DUT}(f) = -\frac{V_{ref,DUT}(f)}{Z_C}. \tag{3.8}$$

Avec (équation 1.3)

$$I_{DUT}(f) = I_{inc,DUT}(f) + I_{ref,DUT}(f), \tag{3.9}$$

On obtient l'expression suivante, du courant dans le plan du composant sous test, dans le domaine fréquentiel :

$$\boxed{I_{DUT}(f) = \frac{1}{Z_C}.\left(\frac{S_{21}}{S_{31}}.V_{inc,oscillo.}(f) - \frac{1}{S_{32}}.V_{ref,oscillo.}(f)\right).} \tag{3.10}$$

Les algorithmes de transformation de Fourier rapide (FFT) et transformation de Fourier rapide inverse permettent de passer du domaine temporel au domaine fréquentiel et vice-versa. Les formes d'onde mesurées à l'oscilloscope sont transformées par FFT en signal fréquentiel, ce qui permet d'appliquer les formules 3.6 et 3.10. Après traitement, les formes d'onde incidentes et réfléchies vues par le composant sous test s'obtiennent par transformation inverse.

Remarque : Les mesures de paramètres $-S$ faites à l'analyseur de réseau donnent accès aux coefficients de transmission entre les différents ports de la figure 3.2. Le port 1 pourrait être situé n'importe où sur TL_1 entre le générateur et la sonde en tension. En effet, le coefficient de transmission de la ligne TL_1 est présent à la fois au numérateur et au dénominateur de l'expression 3.3, et n'entre donc pas en ligne de compte dans le calcul final. Le port 2, lui, devrait idéalement être situé dans le plan de la plaquette de silicium. Or ce n'est pas réalisable car le plan de la plaquette coïncide avec l'extrémité des pointes RF, qu'il est impossible de relier à l'entrée de l'analyseur de réseau. Les coefficients de paramètres -S ne sont donc mesurés que jusqu'à l'extrémité du câble TL_2. La caractéristique de transmission fréquentielle des pointes de mesure est évaluée de façon séparée et intégrée dans le calcul. Il est très difficile de les caractériser dans le domaine fréquentiel, mais leur bande passante très élevée ($12\ GHz$) nous autorise à les assimiler à un délai pur $\tau_{pointes}$ de 80 ps, mesuré dans le domaine temporel. Ce délai apparaît simplement

comme un terme de déphasage $e^{-i2\pi f.\tau_{pointes}}$ dans les coefficients de paramètres -S.

Toutes ces opérations sont réalisées par un script que j'ai développé sous « $Matlab$ », qui prend en charge le traitement des données acquises à l'oscilloscope, depuis la charge des fichiers de points bruts jusqu'à l'écriture et l'affichage des formes d'onde corrigées. Une partie importante du traitement consiste en une manipulation des fichiers de points pour qu'ils puissent être utilisés de manière optimale par les algorithmes effectuant la FFT sous $Matlab$. Afin de mieux en saisir les tenants et les aboutissants, nous allons nous intéresser à présent aux opérations réalisées par le logiciel lors des calculs de transformation de Fourier.

3.3 Transformation de Fourier Discrète

3.3.1 Utilisation pratique de la Transformation de Fourier discrète sous $Matlab$

La transformation de Fourier continue d'un signal $v(t)$ temporel en un signal fréquentiel $V(f)$ et son opération inverse sont respectivement définies par les expressions suivantes (cf. annexe A)

$$V(f) = \int_{-\infty}^{+\infty} v(t).e^{-2\pi ift}dt \tag{3.11}$$

et

$$v(t) = \int_{-\infty}^{+\infty} V(f).e^{2\pi ift}df. \tag{3.12}$$

On voit que pour calculer V, il faudrait faire une acquisition de $t = -\infty$ à $t = +\infty$ à l'oscilloscope avec une résolution infinie, ce qui est impossible en pratique. En mesures impulsionnelles de type $vf - TLP$, l'oscilloscope ne peut qu'effectuer un nombre fini N d'acquisitions sur une fenêtre de temps T_W, avec une durée Δt entre chaque acquisition, définie par la fréquence d'échantillonnage de l'oscilloscope. On a :

$$T_W = N.\Delta t \tag{3.13}$$

Une information dans le domaine fréquentiel à partir de cet ensemble fini de N mesures peut tout de même être obtenue, grâce à la transformation mathématique dite « transformation de Fourier discrète » (DFT). La DFT s'approche plus d'une décomposition en séries de Fourier (DSF) que d'une transformation de Fourier (Annexe A). Malgré son nom, il s'agit en fait d'une DSF dont chaque terme est défini de façon discrète, et non par une intégrale de $-\infty$ à $+\infty$. Comme les signaux ne sont pas périodiques, la DFT considère le signal échantillonné comme un signal périodique de période $T_p = T_W$.

Les valeurs discrètes f_k et t_k de fréquence et de temps peuvent s'exprimer comme ci-dessous :

$$\begin{cases} f_n = n.f_0 = \frac{n}{T_W} = \frac{n}{N.\Delta t} \\ t_k = k.\Delta t \end{cases} \tag{3.14}$$

La DFT d'un signal échantillonné v s'écrit alors, par définition :

$$\boxed{V(f_n) = V_n = \sum_{k=0}^{k=N-1} v_k.e^{-2\pi i.\frac{nk}{N}}}. \tag{3.15}$$

La DFT inverse est définie par :

$$\boxed{v(t_k) = v_k = \frac{1}{N} \sum_{n=0}^{n=N-1} V_n.e^{2\pi i.\frac{nk}{N}}}, \tag{3.16}$$

On note que ces équations peuvent se retrouver en partant de la DSF du signal $v(t)$ (Annexe A), en remplaçant A_n par $\frac{V(f_n)}{N}$ et en abrégeant $v(t_k)$ en v_k et $V(f_n)$ en V_n dans l'équation A.3.

L'équation 3.15 définit un tableau **V** de N points dans le domaine fréquentiel. Il est l'image du tableau **v** dans le domaine temporel, chaque point du tableau correspondant à une fréquence $f_n = nf_0 = n/T_W$, avec f_0 la fréquence fondamentale. On remarque aisément que la valeur correspondant à la fréquence nulle ($n = 0$),

$$V(0) = \sum_{k=0}^{k=N-1} v_k = N.\bar{v}, \tag{3.17}$$

s'exprime simplement à partir de la valeur moyenne $\bar{v}$ du tableau **v** exprimé dans le domaine temporel.

Le tableau des temps est constitué de N points tels que $T_W = N\Delta t$, tout comme les tableaux **f** et **V** du domaine fréquentiel. La limitation constituée par la résolution temporelle Δt s'exprime donc en terme de spectre fréquentiel par l'existence d'une valeur maximale finie f_{ny}, dite « fréquence de Nyquist », ne dépendant que de Δt :

$$f_{ny} = \frac{1}{2\Delta t} = \frac{N}{2}.f_0. \tag{3.18}$$

Le rang n pour lequel cette fréquence atteint son maximum f_{ny} correspond à la moitié du nombre de points $n = N/2$, si N est pair. Un tableau du domaine temporel **v** se transforme par DFT en un tableau **V** dont la première moitié contient une composante DC suivie des $N/2$ harmoniques de f_0, de f_0 à f_{ny}. La seconde moitié est constituée des complexes conjuguées des $\frac{N}{2}$ premières valeurs sans le DC, ce qui se traduit par l'apparition d'harmoniques « négatives », depuis $-f_{ny}$ jusqu'à $-f_0$.

En effet, considérons une valeur de **V** de rang $N/2 + m$, donc appartenant à la deuxième moitié, avec m un entier positif inférieur à $N/2$. On a, d'après 3.15 :

$$V_{\frac{N}{2}+m} = \sum_{k=0}^{N-1} v_k . e^{-\frac{2\pi ik}{N}.\left(\frac{N}{2}+m\right)} \tag{3.19}$$

$$= \sum_{k=0}^{N-1} v_k . e^{-\pi ik} . e^{-\frac{2\pi ikm}{N}} \tag{3.20}$$

$$V_{\frac{N}{2}+m} = \sum_{k=0}^{N-1} v_k . e^{\frac{2\pi ik}{N}.\left(\frac{N}{2}-m\right)} \tag{3.21}$$

avec, $e^{-i\pi k} = e^{+i\pi k}$, car k est entier. Le complexe conjugué de cette somme est égal à la somme des complexes conjugués de ses termes. Les termes v_k étant réels, on a alors :

$$V_{\frac{N}{2}+m} = \overline{\sum_{k=0}^{N-1} v_k . e^{-\frac{2\pi ik}{N}.\left(\frac{N}{2}-m\right)}}, \tag{3.22}$$

où la barre horizontale désigne l'opération de conjugaison.

$$\boxed{V_{\frac{N}{2}+m} = {V_{\frac{N}{2}-m}}^*} \tag{3.23}$$

Ainsi pour résumer, les tableaux des temps **t** et des tension **v** se transforment en tableaux fréquence **f** et tension **V** dans le domaine fréquentiel, tels que :

$$\boxed{\begin{cases} (t_1 \ ... \ t_N) & \xrightarrow{DFT} & (f_1 \ ... \ f_N) = \left(0,\ f_0,\ 2f_0,\ ...\ f_{\frac{N}{2}} = f_{ny},\ ...\ f_0\right) \\ (v_1 \ ... \ v_N) & \xrightarrow{DFT} & (V_1 \ ... \ V_N) = \left(V_1 \ ... \ V_{\frac{N}{2}+1},\ {V_{\frac{N}{2}}}^*,\ {V_{\frac{N}{2}-1}}^* \ ... \ V_2^*\right) \end{cases}} \tag{3.24}$$

3.3.2 Algorithme de FFT

Il est souvent nécessaire d'augmenter le nombre de points N des tableaux **v** et **t**, notamment pour pouvoir abaisser la fréquence fondamentale f_0, celle-ci étant définie par :

$$f_0 = \frac{1}{N.\Delta t}. \tag{3.25}$$

Ainsi, selon le besoin de l'utilisateur en termes de résolution en basses fréquences et les capacités de l'oscilloscope, ce nombre N peut facilement atteindre les 8000 points ou plus. Le temps de calcul pour évaluer directement les N termes de l'expression 3.15 peut vite s'avérer très long. En effet, le nombre d'opérations correspondantes s'élève à N^2 multiplications et N^2 additions, soit $2.N^2$ opérations.

On appelle « transformation de Fourier rapide » (Fast Fourier Transformation) toute méthode réduisant ce nombre d'opérations à $Nlog(N)$ au lieu de $2N^2$. Parmi ces méthodes, la plus courante est de loin l'algorithme de Cooley-Tukey, qui réduit le nombre d'opérations à $2.Nlog_2(N)$ [57]. Il se base sur une technique algorithmique surnommée « diviser pour régner », qui consiste à diviser un problème en deux ou plusieurs sous-problèmes plus simples à résoudre. De façon pratique l'algorithme de Cooley-Tukey divise l'équation 3.15 en deux sommes, l'une contenant les termes de rang impairs et l'autre les termes de rang pairs. Chacune de ces sommes peut ensuite être divisée de la même manière en sous-sommes, jusqu'à ce que le nombre de termes de la somme obtenue ne soit plus divisible par deux.

La conclusion immédiate de ce qui précède est que le nombre de points doit être une puissance de 2 pour que l'algorithme de Cooley-Tukey soit le plus puissant :

$$N = 2^m. \tag{3.26}$$

Si le nombre de points acquis ne respecte pas cette condition, il peut être utile, avant d'effectuer le calcul de FFT, de rajouter artificiellement des points au signal temporel échantillonné. Le double intérêt de cette opération est d'une part de réduire d'autant la fréquence fondamentale $f_0 = 1/N.\Delta t$ du spectre correspondant, et d'optimiser la vitesse de calcul de la simulation en ajustant le nombre de points.

3.4 Détermination du spectre fréquentiel nécessaire

Le choix des paramètres de calcul (nombre de points N, durée d'acquisition T_W, fréquence d'échantillonnage f_{ny} etc.) est limité par les performances de l'équipement disponible au laboratoire. L'application de la méthode décrite plus haut nécessite l'utilisation d'un oscilloscope et d'un analyseur de réseau dont les bandes passantes sont les plus larges possible.

3.4.1 Limites des analyseurs de réseau

Aucun analyseur de réseau ne peut mesurer avec précision les coefficients de paramètres -S d'un système à la fois dans les basses et les hautes fréquences. Cependant, en combinant les mesures opérées avec deux analyseurs de réseau différents disponibles au LAAS, on a pu obtenir une mesure sur une plage de fréquences allant de 9 kHz à 4 GHz avec un pas de 9 kHz, puis jusqu'à 40 GHz avec un pas de 50 MHz.

3.4.2 Limites de l'oscilloscope

L'oscilloscope choisi pour le banc de caractérisation $vf - TLP$ est le modèle TDS6604C de Tektronix. Sa fréquence d'échantillonnage est de 20 Gsamples/s, ce qui correspond à un pas en

temps de 50 ps. Ces 50 ps ne constituent pas cependant la vraie limite physique de l'appareil. Celle-ci est fixée par le filtrage des composants internes à l'oscilloscope, qui déterminent sa bande passante analogique, donnée à 6 GHz par le constructeur. Le temps de montée minimal mesurable est défini dans la littérature par la correspondance fréquentielle suivante :

$$t_r = \frac{0.35}{BW}, \tag{3.27}$$

avec BW la bande passante de l'appareil de mesure. Il est à noter que cette formule est une approximation obtenue en assimilant le système filtrant à un circuit à un pôle. Dans le cas de l'oscilloscope choisi, cela correspond à un temps de montée de 58 ps. La valeur donnée par le constructeur est la plage 60 $ps - 100$ ps.

3.4.3 Choix des limites fréquentielles

Limite supérieure

Dans les hautes fréquences, on voit que la limite est déterminée par l'oscilloscope, dont la bande passante est inférieure à celle du VNA. L'utilisateur voulant calculer la DFT d'un signal temporel a tout intérêt à tirer pleinement partie de la fréquence d'échantillonnage de l'oscilloscope en conservant son pas en temps Δt minimal. Le signal fréquentiel n'en sera que mieux défini dans les hautes fréquences, car comme on l'a vu, les spectres du signal échantillonné s'étend alors jusqu'à $f_{max} = f_{ny} = \frac{1}{2\Delta t}$. On peut donc d'ores et déjà s'arrêter aux paramètres suivants :

$$\boxed{\Delta t = 50\ ps,} \tag{3.28}$$

ce qui amène à :

$$\boxed{f_{ny} = 10\ GHz} \tag{3.29}$$

Limite inférieure

L'étape qui suit consiste à déterminer la durée d'acquisition T_W du signal à l'oscilloscope, et le pas en fréquence $f_0 = \Delta f$ de l'acquisition au VNA, ces deux grandeurs étant liées par la relation $f_0 = \frac{1}{T_W}$, comme on l'a vu plus haut. On pourrait choisir, pour profiter de la bonne résolution fréquentielle du VNA d'abaisser la fréquence fondamentale à la résolution de l'analyseur de réseau, c'est-à-dire 9 kHz. Mais la durée de la fenêtre d'acquisition à l'oscilloscope correspondant à un tel pas en fréquence serait alors de 111 μs, et le nombre de points serait supérieur à 2.2 millions en conservant $\Delta t = 50$ ps.

Si par chance, le nombre de points maximal que l'oscilloscope peut stocker en une acquisition est justement égal à 2 millions, il est cependant inutile dans notre cas de dépasser les 400 ns

de durée d'acquisition. En effet, le générateur TLP n'est pas conçu pour créer des impulsions d'une durée supérieure à 100 ns. On verra plus loin 3.5.4) que la plage de temps nécessaire pour mesurer tout le signal ne dépasse pas 200 ou 300 ns.

Pour abaisser la fréquence fondamentale f_0 du spectre, plutôt que de faire de très longues acquisitions, on peut étendre manuellement le nombre de points des signaux temporels en ajoutant des points de valeur nulle à la suite du signal. Cependant, on peut montrer qu'il est illusoire de vouloir augmenter la précision en réalisant une telle opération, car en réalité les points ainsi ajoutés aux tableaux du domaine fréquentiel n'apportent aucune nouvelle information.

En conclusion, pour ce qui est du choix de la durée d'acquisition, on peut se limiter à la fenêtre de mesure contenant le signal utile. Il n'est pas nécessaire de dépasser la valeur $T_W = 400\ ns$. Avec une résolution Δt de 50 ps, le nombre de points correspondant est égal à 8000.

On rappelle cependant que le nombre de points sur lequel un signal doit être échantillonné pour pouvoir effectuer le calcul de la DFT d'un signal doit être une puissance de deux pour que le calcul soit le plus efficace. On peut donc allonger légèrement la durée d'acquisition de manière à ramener N à la puissance de deux la plus proche, soit $2^{13} = 8192$. Cette opération a pour effet d'allonger T_W à 409.6 ns et de réduire f_0 à 2.44 MHz. Si l'oscilloscope ne permet pas de choisir manuellement le nombre de points des acquisitions, on comble les fichiers de points avec autant de zéros qu'il est nécessaire, de manière à atteindre une puissance de 2. On peut maintenant fixer le pas en fréquence idéal de l'acquisition au VNA :

$$\boxed{f_0 = 2.44\ MHz,} \tag{3.30}$$

pour :

$$\boxed{T_W = 409.6\ ns} \tag{3.31}$$

Bilan

Pour résumer, les paramètres optimaux d'acquisition à l'analyseur de réseau sont les suivants :

$$\boxed{\begin{cases} f_0 = 2.44\ MHz \\ f_{max} = f_{ny} = 10\ GHz \\ N = 4096\ points \end{cases}} \tag{3.32}$$

Les paramètres d'acquisition à l'oscilloscope sont quant à eux :

$$\begin{cases} \Delta t = 50\ ps \\ T_W = 409.6\ ns \\ N = 8192\ points \end{cases} \tag{3.33}$$

3.5 Traitement des données temporelles et fréquentielles

Comme on l'a vu, le calcul de DFT se base sur une approximation : un signal temporel unique est considéré par la DFT comme un signal périodique dont la période est égale à la durée d'acquisition T_W. Cette durée d'acquisition doit être optimisée pour améliorer la résolution fréquentielle des signaux obtenus et la vitesse de calcul des algorithmes. Mais aussi, il est indispensable que la valeur prise par le signal échantillonné soit égale aux instants $t = 0$ et à $t = T_W$, de manière à ce que cette approximation d'un signal périodique soit justifiée. Dans le cas contraire, les discontinuités crées dans les signaux « périodiques » seraient source de composantes parasites à hautes fréquences. Typiquement, en mesures impulsionnelles $vf - TLP/transient - TLP$, les signaux sont nuls avant l'impulsion et nuls après, ce qui rend le calcul plus facile que dans le cas d'une fonction échelon par exemple.

3.5.1 Signal non nul à $t < 0$: correction de l'offset

Si cependant le signal n'est pas nul à $t < 0$, cela est généralement révélateur d'un défaut de synchronisation de l'oscilloscope. Ce dernier est en effet synchronisé sur le front montant de l'impulsion, et le temps $t = 0$ est choisi au moment où la tension atteint un certain seuil. Il arrive que les formes d'ondes fournies par l'oscilloscope soient affectées d'un léger décalage, positif ou négatif, à cause de la résolution en tension non idéale de l'oscilloscope. On peut facilement s'en débarrasser en retranchant à chacun des signaux mesurés une valeur moyenne calculée sur les premières nanosecondes de l'acquisition, avant l'apparition de la première impulsion.

3.5.2 Séparation des impulsions incidentes et réfléchies

Les expressions 3.6 et 3.10 traitent les ondes incidentes et réfléchies comme s'il s'agissaient de deux impulsions séparées. Or, dans la réalité, ces deux impulsions apparaissent à la suite dans la même mesure, ou même superposées si le retard dans TL_2 est plus court que la durée des impulsions. En effet, une tension résiduelle est toujours présente à la suite de l'impulsion incidente mesurée. Elle peut prendre la forme d'une lente décroissance ou d'une réplique miniature de l'impulsion ESD (figure 3.3), ce dernier phénomène étant probablement dû à une réflexion parasite à l'intérieur du générateur d'impulsions. Dans tous les cas, elle se superpose à l'impulsion réfléchie sur l'écran de l'oscilloscope.

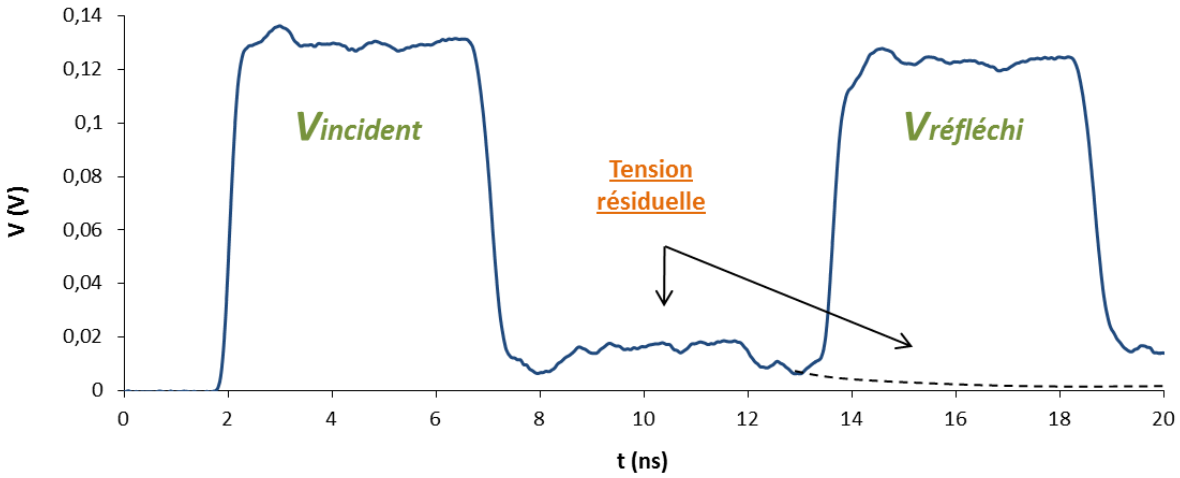

FIGURE 3.3 – Impulsions $vf - TLP$ mesurées à l'oscilloscope avec le dispositif de la figure 3.2. La tension résiduelle présente à la suite de l'impulsion incidente se superpose partiellement avec l'impulsion réfléchie.

Ces tensions résiduelles font partie intégrante de l'impulsion incidente injectée dans le système de mesure par le générateur TLP, et il est donc exclu de s'en débarrasser : elles se répercutent sur les signaux réfléchis, et les supprimer reviendrait à introduire une importante erreur dans le re-calcul des formes d'onde.

Il est donc nécessaire d'utiliser des méthodes de séparation de ces deux impulsions. Deux solutions expérimentales ont été retenues pour résoudre ce problème, solutions que dans la pratique on tente de combiner. La première consiste à allonger la ligne TL_2 (figure 3.1) de manière à augmenter le délai entre les deux impulsions mesurées, et la seconde à procéder en deux étapes pour acquérir le signal à l'oscilloscope.

Solution 1

En allongeant la ligne TL_2, on retarde l'apparition de la seconde impulsion sur l'oscilloscope, ce qui limite l'étendue de la superposition entre les deux impulsions. Mais le rallongement du câble a également pour effet d'introduire une atténuation fréquentielle se traduisant par un ralentissement du front montant. Cette solution doit donc être utilisée en connaissance de cause, notamment en ce qui concerne le temps de montée que l'on désire appliquer au composant sous test.

Solution 2

L'autre solution permet de limiter l'atténuation fréquentielle de $V_{ref,oscillo}$ due à l'allongement du câble. Elle consiste à mesurer au préalable la totalité de l'impulsion incidente $V_{inc,oscillo}$, visible à l'oscilloscope, sans qu'elle ne soit perturbée par l'impulsion réfléchie $V_{ref,oscillo}$. Pour ce faire, on remplace simplement le DUT par une charge adaptée de 50 Ω. Ainsi, le signal incident est intégralement absorbé et il n'y a pas d'impulsion réfléchie. On ne mesure qu'un signal incident $V_{load,oscillo}$, et le signal réfléchi s'obtient ensuite en retranchant au signal mesuré sur le DUT ce signal incident « complet » :

$$V_{incident,oscillo.} = V_{load,oscillo.} \tag{3.34}$$

$$V_{réfléchi,oscillo.} = V_{DUT,oscillo.} - V_{load,oscillo.} \tag{3.35}$$

De cette manière, la séparation des signaux incidents et réfléchis mesurés est très précise, aucun des deux signaux ne venant interférer sur l'autre.

De plus, dans le cas où les impulsions générées sont si longues qu'elles se superposent sur l'oscilloscope (par ex. impulsions de type « TLP »), cette approche devient le seul moyen de séparer les deux signaux.

Dans le cas des impulsions $vf - TLP$, il est judicieux, pour ce qui est de la définition de « $V_{inc,oscillo.}$ » de combiner les impulsions incidentes mesurées sur 50 Ω et sur le composant. En effet, le générateur TLP n'est pas idéalement répétitif, et autant que possible, on évitera d'utiliser dans la formule 3.6 des impulsions mesurées à partir de deux agressions différentes à la même tension. On mesure successivement les impulsions obtenues sur une charge 50 Ω (acquisition 1, figure 3.4-a) et sur le DUT (acquisition 2, figure 3.4-b). « $V_{inc,oscillo.}$ » sera égal à l'acquisition 2 sur toute la durée de l'impulsion incidente, puis à l'acquisition 1 sur le reste de la mesure.

Connaissant l'impulsion incidente, il est alors facile de remonter à l'impulsion réfléchie par soustraction du signal total par le signal incident. (figure 3.5).

3.5.3 Suppression des réflexions parasites

Un autre cas plus préoccupant est celui des réflexions parasites ayant lieu dans le banc de mesure, typiquement sur la forte résistance de charge R_1 du générateur d'impulsions (figure B.1). Ce signal, artéfact en modèle réduit de l'impulsion ESD d'origine, est issu de la réflexion de l'impulsion V_{ref} sur le générateur, elle-même réflexion de l'impulsion incidente sur le DUT (figures 3.4 et 3.5). Un atténuateur de 6 dB prévu à cet effet (section 2.2.1.1) divise par deux

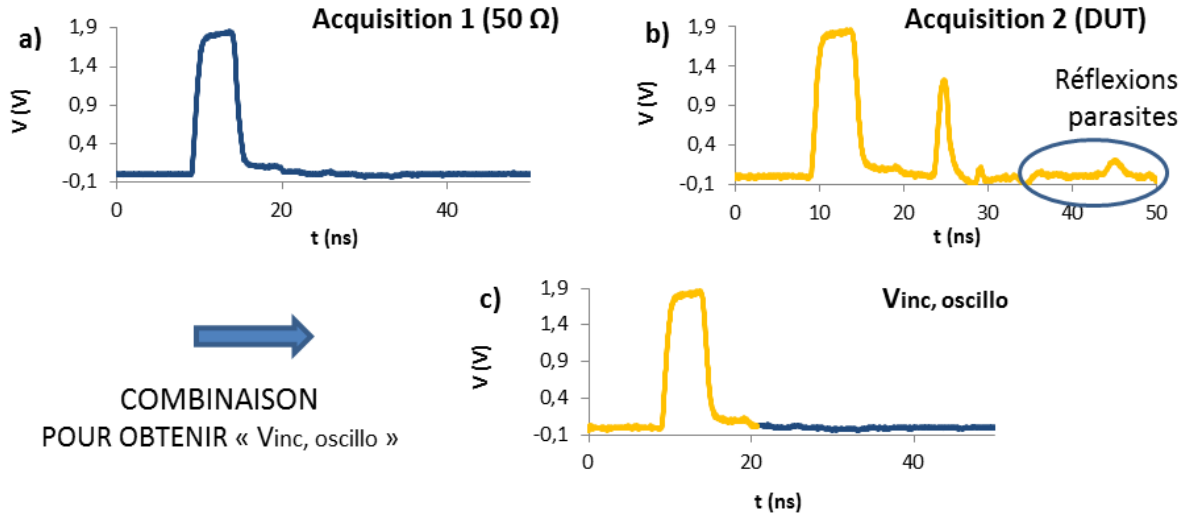

FIGURE 3.4 – Exemple d'acquisitions successives sur une charge 50 Ω (acquisition 1) et sur une protection *ESD* haute tension à 40 *V TLP* (acquisition 2), après correction des offsets . On construit ainsi le fichier $V_{inc,oscillo}$, qui sera utilisé par la suite. L'avantage de cette opération est de faire en sorte que les signaux incidents et réfléchis utilisés dans les calculs soient issus de la même impulsion d'origine. Ils sont alors liés par une transformation mathématique. Les valeurs non nulles dans l'acquisition 2 après l'impulsion réfléchie sont le résultat de réflexions parasites sur le générateur. On note qu'à ce stade, les amplitudes des impulsions n'ont pas encore été corrigées.

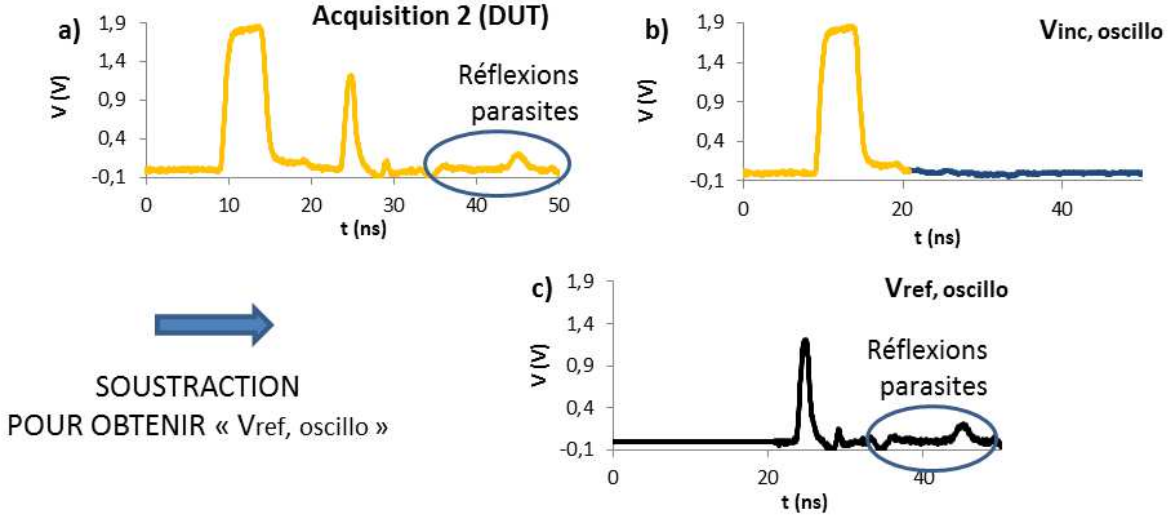

FIGURE 3.5 – Construction du fichier de points $V_{ref,oscillo.}$, défini comme étant la différence entre la mesure effectuée sur le DUT et le fichier de points précédemment construit $V_{inc,oscillo.}$.

ce signal parasite pour éviter qu'il ne se réfléchisse indéfiniment aux deux extrémités du banc de mesure. Malgré cette atténuation, l'apparition et la superposition de ce signal sur le signal « utile » vient perturber le traitement qu'on applique aux impulsions mesurées.

Solution 1 : retarder l'apparition des réflexions parasites

Un moyen efficace de s'en débarrasser est d'augmenter la longueur de la ligne TL_1 (figure 3.1), afin de retarder le plus possible son apparition. Ainsi, on laisse un laps de temps suffisant après l'impulsion réfléchie mesurée $V_{ref,osc.}$ pour que la tension retombe à zéro sans être perturbée par la réflexion parasite. Les signaux parasites sont ensuite supprimés des fichiers de points bruts sans danger de suppression de signaux « utiles ».

Solution 2 : le fenêtrage des signaux

En complément de l'allongement de la ligne TL_1, il est nécessaire de trouver un moyen de se débarrasser des impulsions réfléchies parasites. La solution consistant à insérer un isolateur en série avec le générateur d'impulsion pour s'affranchir des signaux revenant au générateur après réflexion n'est pas envisageable car aucun de ces dispositifs n'est conçu pour fonctionner sur une telle bande de fréquences, en particulier pour les plus basses. Il reste donc la possibilité de supprimer de manière mathématique ces signaux en multipliant les formes d'ondes par des fonctions dites de « fenêtrage ». Ces fonctions ont pour but de sélectionner une fenêtre déterminée d'un signal (fenêtre d'observation), tout en s'assurant que le signal retombe à zéro sans discontinuités sur les bords de la fenêtre. On réalise ainsi une « apodisation » des signaux (littéralement « couper les pieds »).

Les fonctions de fenêtrage sont donc égales à 1 sur la fenêtre à sélectionner, et à des fonctions décroissantes sur les bords de la fenêtre. Elles prennent une valeur nulle en $t = -\infty$ et $t = +\infty$. Une des préoccupations principales de ces fonctions est la gestion de leur impact fréquentiel sur le signal échantillonné. En effet, si $e(t)$ est le signal d'origine et $h(t)$ la fonction de fenêtrage, le signal apodisé $s(t)$ est défini par :

$$s(t) = h(t).e(t) \tag{3.36}$$

Cette opération de multiplication dans le domaine temporel correspond à une convolution dans le domaine fréquentiel :

$$S(f) = H(f) * E(f) \tag{3.37}$$

avec $E(f)$, $H(f)$ et $S(f)$ les transformées de Fourier respectives de $e(t)$ $h(t)$ et $s(t)$.

$$S(f) = \int_{-\infty}^{+\infty} H(f - f').E(f')df', \tag{3.38}$$

On voit alors qu'idéalement, pour que la fonction de fenêtrage ait un impact minimal sur le signal échantillonné, elle doit autant que possible se rapprocher de la distribution de Dirac dans le domaine fréquentiel. La fonction de Dirac est représentée dans le domaine temporel par

une fonction constante infinie, c'est-à-dire ne réalisant aucun fenêtrage. Donc dans la pratique, le choix d'une bonne fonction de fenêtrage est guidé par deux aspects principaux :

- L'efficacité du fenêtrage : dans notre cas précis, on s'assure que les motifs parasites sont bien supprimés.
- Le spectre fréquentiel de la fonction : les composantes fréquentielles doivent au maximum être regroupées dans les basses fréquences.

La transformée de Fourier de ces fonctions se présente sous la forme d'un lobe principal centré sur la fréquence nulle, que l'on souhaite le plus étroit possible, et de lobes secondaires, que l'on souhaite les plus faibles possibles.

La fonction de fenêtrage la plus élémentaire est la fonction de fenêtrage rectangulaire, ou fonction « porte », égale à 1 sur la fenêtre sélectionnée et nulle en-dehors. Cette fonction est représentée à la figure 3.6-a. Si T est la largeur de la fenêtre, on a :

$$h(t) = \begin{cases} 1 & si\ t \in [0, T] \\ 0 & sinon. \end{cases} \tag{3.39}$$

Elle a l'avantage d'être la plus simple à régler, mais ses fortes discontinuités sur les bords de la fenêtre en font un mauvais candidat en termes d'impact fréquentiel. En effet, son spectre, représenté figure 3.6-b, se présente sous la forme d'un sinus cardinal. Il montre une série d'oscillations dont l'amplitude décroit progressivement pour atteindre 1 % de la valeur DC au bout de quelques GHz. Elle peut cependant être intéressante si l'on arrive à faire en sorte que le signal à traiter soit nul aux aux instants correspondant aux discontinuités.

Une autre fonction élémentaire est la fonction triangulaire, définie comme suit :

$$h(t) = \begin{cases} \frac{2t}{T} & si\ t \in [0, \frac{T}{2}] \\ \frac{2.(T-t)}{T} & si\ t \in [\frac{T}{2}, T] \\ 0 & sinon. \end{cases} \tag{3.40}$$

Une combinaison des fonctions rectangulaires et triangulaires est représentée figure 3.6-a. La fonction est égale à 1 sur la fenêtre de mesure et retombe à zéro en 5 ns de façon linéaire. Ici encore, les discontinuités aux raccordements ont un impact sur les signaux convolués.

La fonction de Tukey est constituée d'une combinaison de la fonction porte et de fonctions cosinus décroissantes. La durée pendant laquelle elle est égale à 1 est réglable par le paramètre α, de même que la durée de retour à zéro, par le paramètre T. On a ici fixé la durée de retour à zéro à 5 ns comme dans le cas précédent. Mais la dérivée nulle du signal aux bords de la fenêtre ainsi qu'aux retours à zéro assurent un bien meilleur raccordement que dans les cas précédents. On note que ses harmoniques sont également bien plus faibles (division par respectivement 10^4

et 10^3 à 3 GHz par rapport aux deux cas précédents).

$$h(t) = \begin{cases} \frac{1+cos\left[\frac{2\pi}{\alpha T}.(t-\frac{\alpha T}{2})\right]}{2} & si\ t \in [0, \frac{\alpha T}{2}[\\ 1 & si\ t \in [\frac{\alpha T}{2}, T - \frac{\alpha T}{2}[\\ \frac{1+cos\left[\frac{2\pi}{\alpha T}.(t-1+\frac{\alpha T}{2})\right]}{2} & si\ t \in [T - \frac{\alpha T}{2}, T]. \end{cases} \tag{3.41}$$

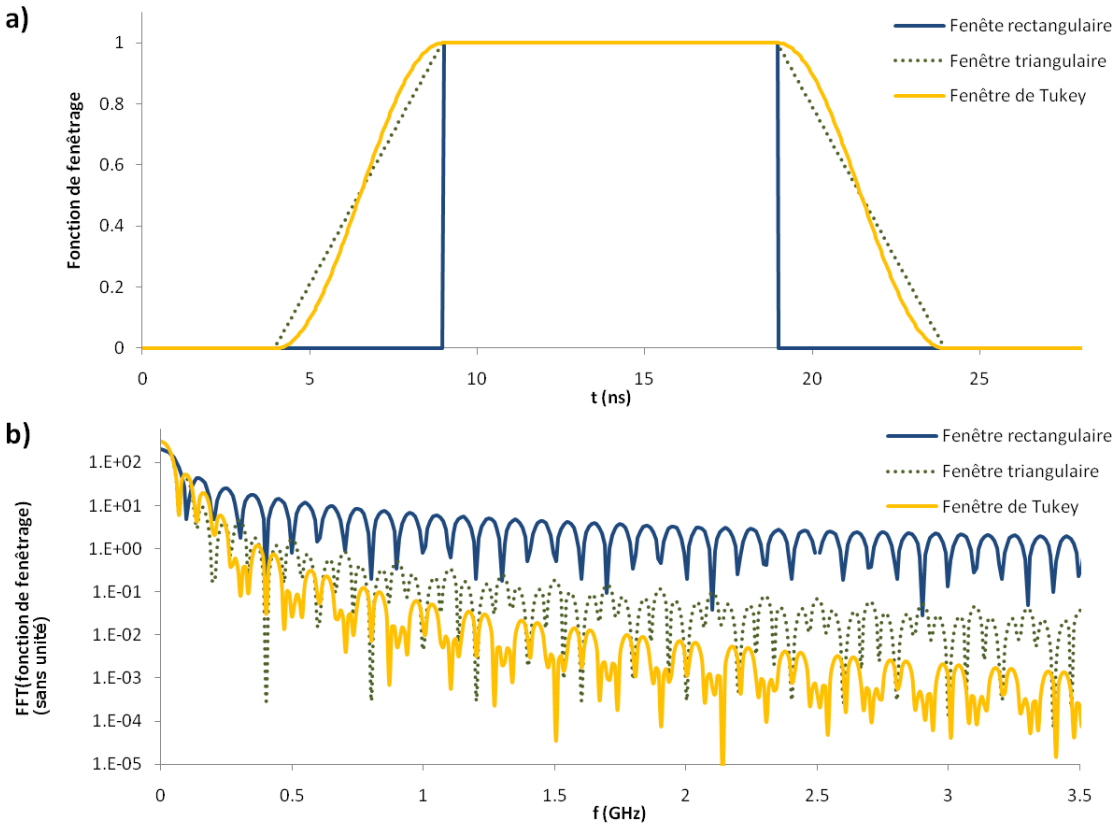

FIGURE 3.6 – Exemple de fonctions de fenêtrage (a) et leurs spectres fréquentiels respectifs (b).

Les fonctions de fenêtrage utilisées dans le script *Matlab* que j'ai développé sont les « fonctions de Tukey ». Le résultat du fenêtrage est représenté dans la figure 3.7. On choisit de couper l'impulsion peu avant l'apparition des réflexions parasites. Les impulsions $V_{inc,oscillo}$ ne sont pas concernées par les problèmes de réflexions parasites, et cette étape ne leur est donc pas nécessaire.

3.5.4 Compromis entre le temps de montée et la précision des résultats

L'allongement des lignes TL_1 et TL_2, qu'il vise respectivement à écarter les réflexions parasites (section 3.5.3) ou à séparer les impulsions incidentes et réfléchies, (section 3.5.2) doit résulter d'un compromis entre le temps de montée désiré en bout de ligne, aux bornes du composant testé, et la précision des résultats. En effet, comme on l'a vu, l'incidence sur le signal des

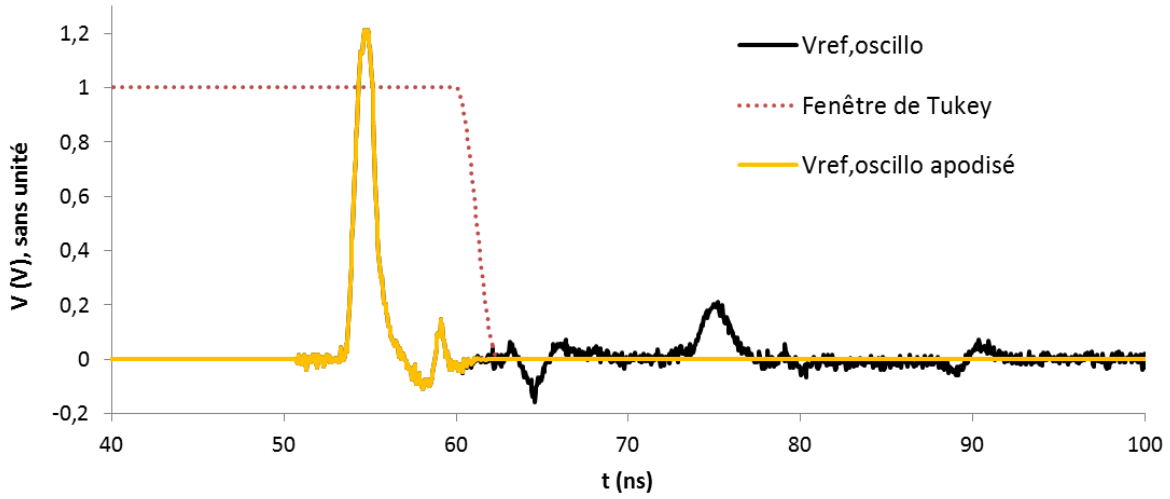

FIGURE 3.7 – Opération de fenêtrage sur le fichier de points $V_{ref,oscillo}$ précédemment construit. En noir, l'impulsion d'origine avant le fenêtrage, en rouge pointillé, la fenêtre de Tukey employée et en orange le résultat de la multiplication des deux.

fonctions mathématiques d'apodisation est loin d'être anodine et on tente d'en réduire l'utilisation. De même, l'utilisation de deux mesures séparées (sur 50 Ω et sur le DUT) pour obtenir la séparation de l'incident et du réfléchi est source d'erreur s'il s'avère que le générateur n'est pas suffisamment répétitif. Ainsi, plus les lignes TL_1 et TL_2 sont longues et moins l'impact de ces deux opérations sur le signal est important, car elles s'appliquent alors sur des signaux presque nuls.

D'un autre côté, ces allongements sont également responsables de l'atténuation fréquentielle des impulsions en bout de ligne, et en particulier du temps de montée. Supposons par exemple que l'on veuille faire apparaître un délai de $\tau_1 = 5\ ns$ à la suite de l'impulsion réfléchie pour retarder les réflexions parasites, et un second délai τ_2 de la même valeur entre les deux impulsions afin de limiter leur superposition. La longueur de câble à rajouter correspond à un aller-retour dans TL_1 et un aller retour dans TL_2 :

$$L = \frac{\nu.(\tau_1 + \tau_2)}{2} = \frac{0.66c.(\tau_1 + \tau_2)}{2} \tag{3.42}$$

$$L = 1\ m, \tag{3.43}$$

avec ν la vitesse de phase de l'impulsion dans le câble coaxial, égale à $0.66c$ dans le cas de câbles coaxiaux à isolant en polyéthylène.

Évaluons l'impact de cet allongement en termes d'atténuation fréquentielle. A 3 GHz, par exemple, l'atténuation des câbles est de 1 dB/m. On obtient donc une atténuation totale de 1 dB.

Le meilleur compromis à réaliser consiste à fixer la longueur des lignes de telle sorte que le temps de montée obtenu corresponde naturellement à celui désiré sans qu'il soit nécessaire d'insérer un filtre passe-bas en série. Les méthodes de fenêtrage et de soustraction d'impulsions incidentes ne sont alors utilisées qu'en complément.

3.5.5 Traitement sur les coefficients de paramètres -S mesurés

Les fichiers de points comportant les coefficients de paramètres -S sont également optimisés, de manière à ce que le nombre de points, le pas en fréquence et la fréquence maximale respectent les conditions des équations (3.32) si ce n'est pas déjà le cas après l'acquisition au VNA. Une interpolation est réalisée sur les courbes pour réajuster le pas en fréquence et le nombre de points.

On construit ensuite de nouveaux fichiers de points S_{31}, S_{32}, S_{21}, et f à partir de ces coefficients de paramètres -S et du tableau des fréquences. En effet, comme on l'a vu plus haut (équations 3.24), le résultat d'une DFT d'un tableau de N points dans le domaine temporel est un tableau de N points dont les $\frac{N}{2}+1$ premiers points contiennent la composante continue et les composantes fréquentielles « positives », et les $\frac{N}{2}-1$ points restant les composantes fréquentielles « négatives ». La première moitié correspond donc au fichier de points obtenu par la mesure au VNA, tandis que la seconde moitié est constituée de ses complexes conjugués. Pour ce qui est du tableau des fréquences, on construit sa seconde moitié à partir des valeurs opposées aux fréquences de la première moitié. En procédant ainsi, on fait en sorte que les coefficients de paramètres $-S$ et les transformées de Fourier des tableaux du domaine temporel soient de la même forme.

3.5.6 Calcul des formes d'ondes $V_{DUT}(t)$ et $I_{DUT}(t)$

On est à présent en mesure d'appliquer les algorithmes de FFT aux tableaux $V_{inc,oscillo}(t)$ et $V_{ref,oscillo}(t)$. Les résultats $V_{inc,oscillo}(f)$ et $V_{ref,oscillo}(f)$ de ces opérations sont reportés dans les équations 3.6 et 3.10 pour le calcul de $V_{DUT}(f)$ et $I_{DUT}(f)$. Les formes d'onde correspondantes dans le domaine temporel sont calculées par algorithmes de FFT inverse. Le résultat de transformation inverse correspond à un évènement physique et doit donc obligatoirement être réel. La partie imaginaire qui subsiste est due à l'erreur de mesure lors des acquisitions à l'analyseur de réseau et à l'oscilloscope. Elle doit être supprimée dans les formes d'ondes finales.

La figure 3.8 montre le résultat final du calcul des formes d'ondes dans le plan du composant, pour un composant de protection haute tension $NPNZ$ (section 1.6.1.2) destiné à la protection des E/S 20 V. Les mesures ont été effectuées pour une impulsion $vf-TLP$ de 100 V, une des premières tensions de pré-charge pour laquelle la protection se replie. Les phénomènes physiques sont encore relativement lents à ce niveau de tension, et on voit bien à travers les

formes d'onde l'évolution de l'impédance instantanée du composant, passant de l'état bloqué à l'état passant. À l'état bloqué, son impédance est supérieure à celle du générateur, qui se comporte alors comme un générateur de tension. Le courant, essentiellement capacitif, est constant et proportionnel à la pente de la tension imposée. Après déclenchement, l'impédance du composant devient inférieure à celle du générateur, qui se transforme alors en générateur de courant.

La tension de repliement quasi-statique V_{t1} est de 25.4 V, et la tension maximale atteinte ici de 29.8 V environ. La surtension est donc de tout de même de plus de 4 V malgré la faible tension TLP injectée, si l'on définit la surtension comme étant la différence entre la tension maximale du pic et la tension de repliement quasi-statique.

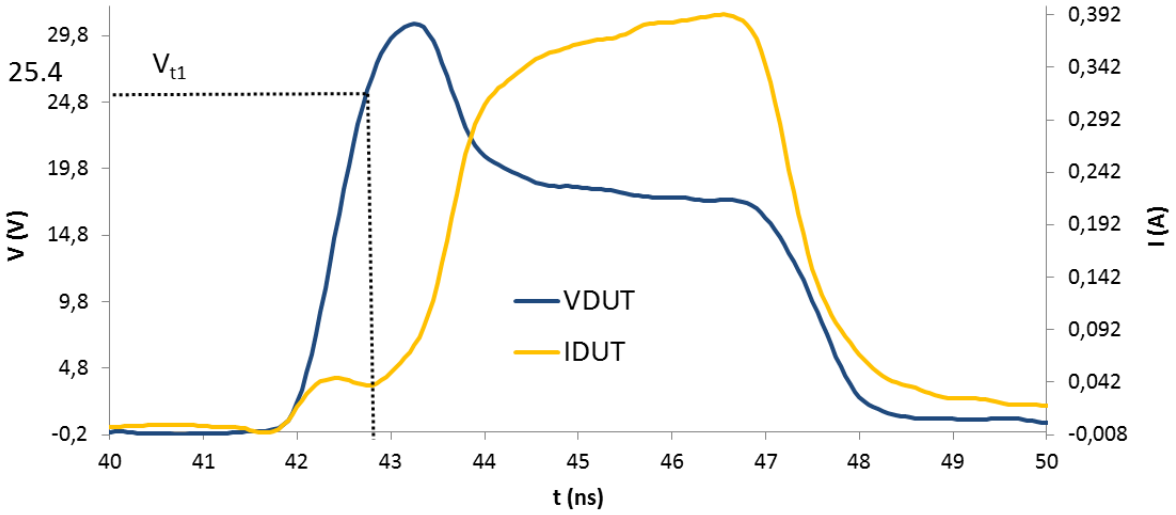

FIGURE 3.8 – Résultat du calcul de $V_{DUT}(t)$ (en bleu foncé) et $I_{DUT}(t)$ (en orange clair) sur une protection ESD haute tension pour 100 V $vf - TLP$, avec 1 ns de temps de montée. Les deux impulsions débutent au même moment, ce qui indique que les délais de propagation dans le système de mesure sont correctement corrigés.

3.6 Résultats de mesure $t - TLP$

3.6.1 Validation sur des cas simples

Formes d'ondes

La façon la plus simple de tester la précision de la méthode de mesure est de la valider sur des cas dont le résultat est connu, typiquement un court-circuit et un circuit ouvert. On vérifie que les impulsions incidentes et réfléchies sont bien égales et superposées sur le circuit ouvert (coefficient de réflexion égal à 1), et opposées sur le court-circuit (coefficient de réflexion égal à -1). Les résultats sont montrés dans les figures 3.9, pour le circuit ouvert et 3.11, pour le court-circuit. Sur le circuit ouvert, la bonne superposition de ces deux courbes donne un aperçu

de la précision de la mesure. En effet, il est à noter que les impulsions incidentes et réfléchies sont calculées en utilisant des coefficients de paramètres -S différents, correspondant à des chemins de propagation indépendants (figure 3.10), et il n'y a donc aucune redondance dans l'obtention des résultats. L'erreur maximale obtenue sur ce circuit ouvert est de l'ordre de 1 %, calculée d'après l'expression suivante, pour une tension de précharge V_{TLP} de 100 V :

$$Erreur_{CO}(\%) = 100.\frac{V_{incident} - V_{réfléchi}}{V_{TLP}} \tag{3.44}$$

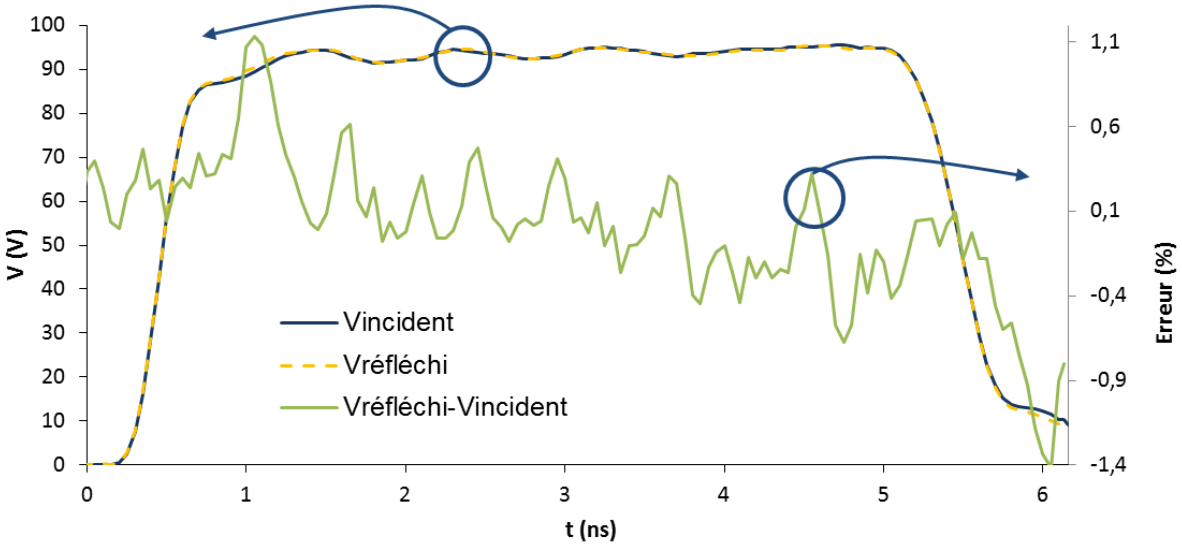

FIGURE 3.9 – Mesure des impulsions incidentes (bleu) et réfléchies (orange pointillés) sur un circuit ouvert (pointes RF laissées en l'air) à 200 V TLP, 5 ns de durée d'impulsion et 300 ps de temps de montée. La différence entre les deux est tracée en vert pâle sur l'axe de droite.

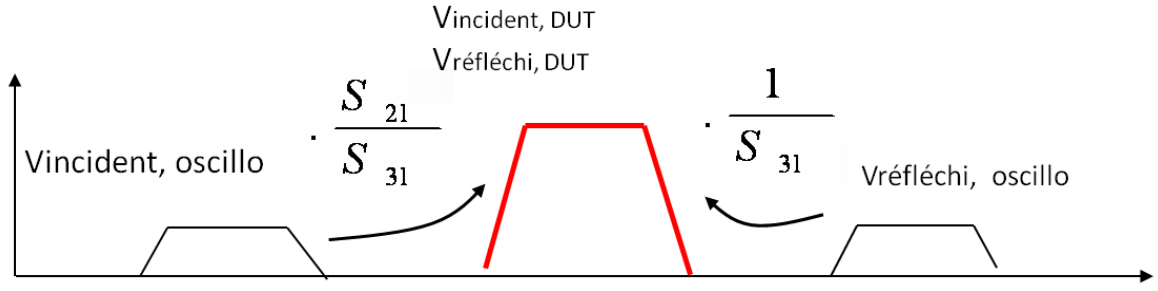

FIGURE 3.10 – Calcul des impulsions incidentes et réfléchies sur un circuit ouvert parfait. Dans ce cas, les deux formes d'onde doivent être identiques et parfaitement alignées dans le temps.

Concernant le court circuit, l'erreur est calculée d'après :

$$Erreur_{CC}(\%) = 100.\frac{V_{incident} + V_{réfléchi}}{V_{TLP}} \tag{3.45}$$

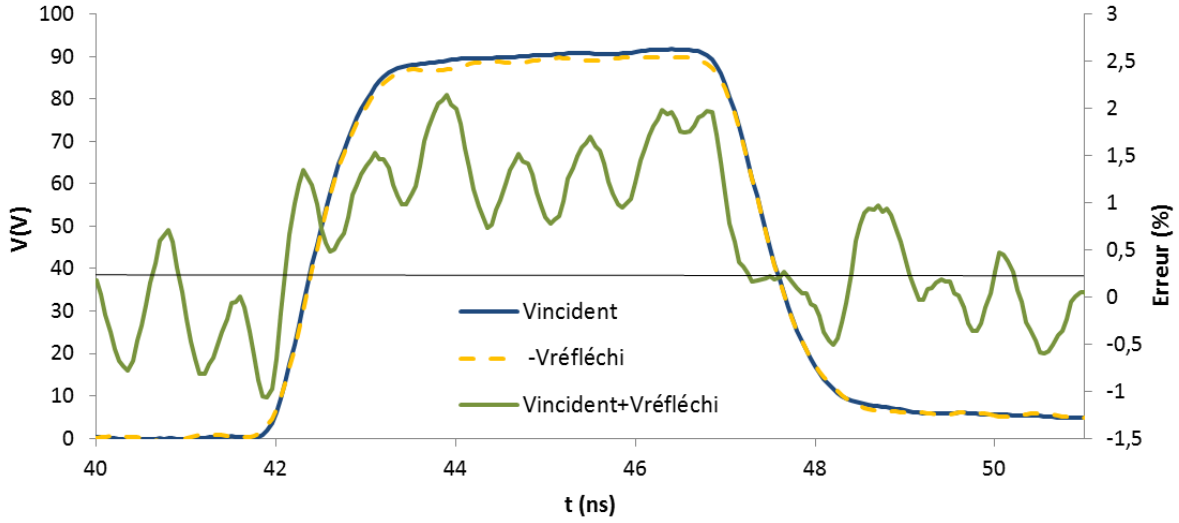

FIGURE 3.11 – Impulsions mesurées sur un court circuit à 200 V TLP, $t_r = 1$ ns, 5 ns de durée d'impulsion. L'impulsion réfléchie (orange pointillés) a été multipliées par -1 pour comparaison avec l'incident (bleu). La somme de l'incident et du réfléchi est tracée en vert pâle, et sa moyenne est reportée en trait noir continu.

L'erreur maximale est de l'ordre de 2 %, soit deux fois plus que dans le cas du circuit ouvert. Mais si le circuit ouvert peut être considéré comme parfait (il est obtenu avec les pointes en l'air), le court-circuit mesuré est une ligne de métal tracée sur wafer entre deux plots de contact. La chute de potentiel visible aux bornes de cette ligne de métal peut lors s'expliquer par plusieurs facteurs, comme la non-idéalité des contacts (résistances de contact), la résistance intrinsèque de la ligne de métal, et les capacités parasites ramenées par l'ensemble « plots+lignes ».

Mode quasi-statique $TLP/vf - TLP$

On peut également tester la pertinence des résultats sur les mêmes court-circuits et circuits ouverts par des mesures quasi-statiques entre 1 V et 1000 V. Une moyenne effectuée dans les dernières nanosecondes de l'impulsion permet, comme en mesures TLP classiques de tracer les graphes $I-V$ correspondant. La figure 3.12 montre la valeur absolue du courant et de la tension obtenus sur court-circuit et circuit ouvert dans ces conditions. La résistance du circuit ouvert apparaît supérieure à 7 $k\Omega$ et celle du court-circuit inférieure à 300 $m\Omega$ entre 0 et 1000 V TLP. Le moyennage sur plusieurs nanosecondes permet ici de réduire fortement les erreurs liées au bruit de mesure.

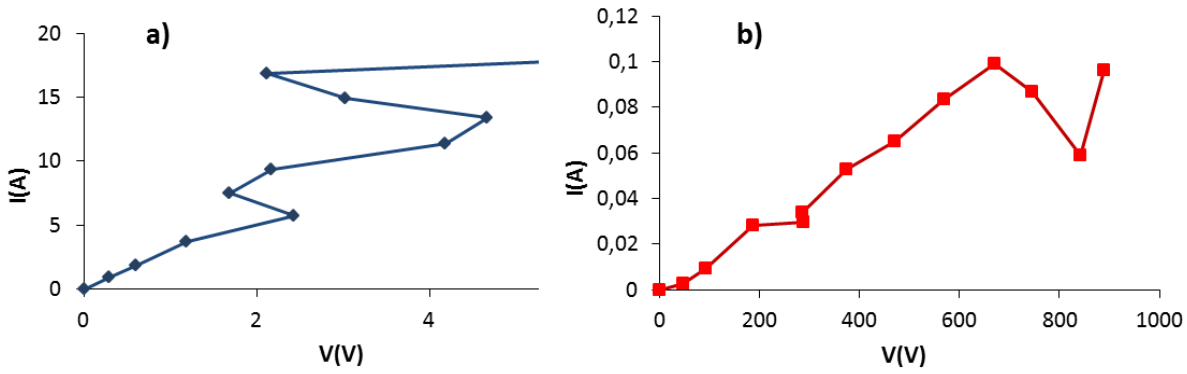

FIGURE 3.12 – Valeur absolue des caractéristiques quasi-statiques obtenues avec l'outil de mesure $transient-TLP$ sur un court-circuit (a) et un circuit ouvert (b).

3.6.2 Application à la mesure sur des composants ESD de librairie

Dans la figure 3.13, on a tracé les caractéristiques de tension et courant en fonction du temps pour une structure de protection $NPNZ$ (section 1.6.1.2), dédiée aux E/S 20 V. La tension de précharge TLP est de 250 V dans les trois cas, seul change le temps de montée ($<$ 100 ps, 300 ps, 1 ns) ajusté au moyen de filtres passe-bas insérés en sortie du générateur. Le temps de montée inférieur à 100 ps correspond au cas sans filtre. Le filtrage est alors celui de l'oscilloscope, qui ne permet pas de mesurer des temps de montée inférieurs à 100 ps. Pour des temps de montée décroissants, on observe une augmentation de l'amplitude du pic de surtension au déclenchement de la protection, variant de 60 V à 95 V environ par rapport à la tension de retournement quasi-statique V_{t1}, qui est de 25.4 V. Ce résultat, intuitif, sera expliqué en détails à l'aide de résultats de simulation (chapitre 5). On note que le temps de montée n'a pas d'impact sur la tension et le courant de stabilisation, qui apparaissent superposés dans les trois cas.

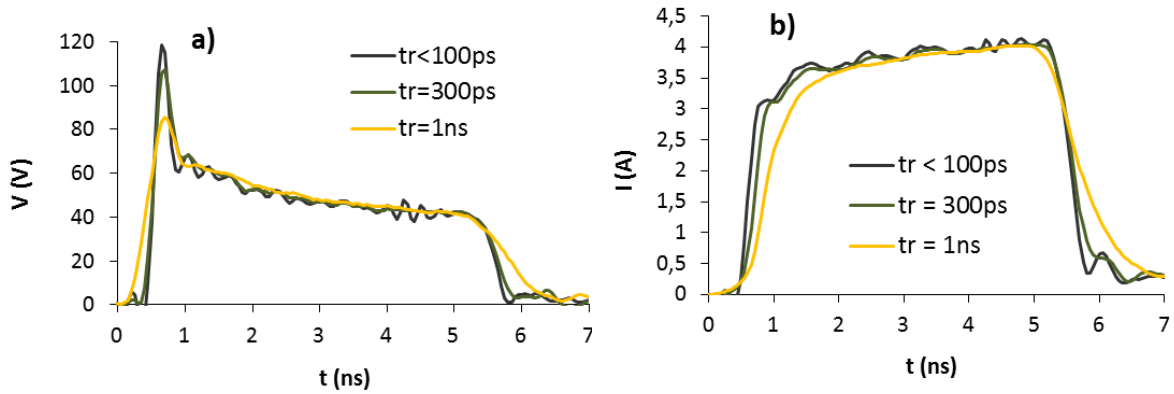

FIGURE 3.13 – Mesure $transient-TLP$ sur un $NPNZ$ dédié aux E/S 20 V pour des impulsions de 5 ns et 250 V de tension de charge avec trois temps de montée différents. La tension est reportée figure a) et le courant figure b).

Courbes « $I - V_{max}$ »

On peut répéter ces mêmes mesures à plusieurs niveaux de tension, et en extrayant pour chaque impulsion la tension de pic V_{max}, on construit des courbes « $I - V_{max}$ ». V_{max} est reporté en abscisse et le courant de stabilisation quasi-statique, mesuré en TLP de façon classique est reporté sur l'axe des ordonnées. La courbe $I - V$ quasi-statique est généralement reportée sur le même graphe pour comparaison [58]. Ainsi, pour chaque niveau de courant, le comportement du composant est décrit simplement et de façon complète. Le dessin de ces structures est facilité par la possibilité de superposer ces courbes avec la fenêtre de conception.

La figure 3.14 montre les courbes $I - V_{max}$ obtenues sur des composants $NPNB$ dédiés aux E/S 20 V (section 1.6.1.3) jusqu'à 6 A de courant quasi-statique, pour deux temps de montées de 1 ns et 10 ns. On remarque que pour toutes les tensions inférieures à V_{t1}, les tensions de pic V_{max} sont égales aux tensions quasi-statiques, ce qui signifie que la surtension est nulle. Les surtensions ne commencent à croître qu'aux niveaux de courant supérieurs, et augmentent d'autant plus vite que le temps de montée est court.

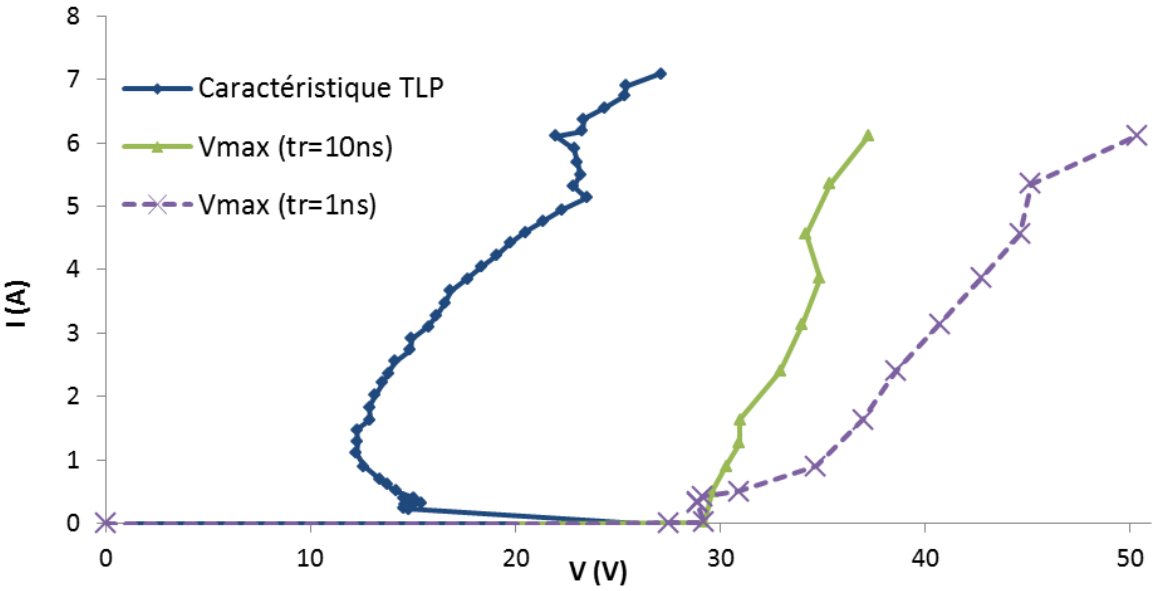

FIGURE 3.14 – Courbes $I - V_{max}$ obtenues par caractérisation $transient - TLP$ sur des $NPNB$ dédiés aux E/S 20 V avec des temps de montée de 1 ns et 10 ns.

On note que les valeurs de V_{max} pour le temps de montée de 10 ns ont été obtenues avec une largeur d'impulsion de 100 ns, et non 5 ns comme pour les autres temps de montée. Le dispositif de mesure n'est alors pas fondamentalement différent du dispositif décrit dans les paragraphes précédents, si ce n'est que l'on a rallongé la ligne de charge du générateur d'impulsion pour augmenter la durée d'impulsion, et les lignes TL_1 et TL_2 de la figure 3.1 pour respectivement éloigner les réflexions parasites et séparer au mieux les impulsions incidentes et réfléchies. Le

choix de rallonger ces deux dernières lignes est justifié par le fait que le temps de montée désiré est suffisamment lent pour n'être que peu affecté par l'atténuation fréquentielle liée à la longueur des câbles coaxiaux.

Comparaison avec des mesures transitoires $vf-TLP$

La figure 3.15 montre une comparaison entre des mesures ($V_{mesuré\ 1}$ et $V_{mesuré\ 2}$) effectuées avec le dispositif $vf-TLP$ d'Oryx incluant la merge box (section 2.2.1.5), et les données du $transient-TLP$ ($V_{corrigé}$). Les deux acquisitions brutes correspondent à deux réglages différents du délai de la merge box, séparés de 150 ps, ce qui correspond à la précision approximative de réglage (section 2.4). $V_{mesuré\ 1}$ et $V_{mesuré\ 2}$ donnent un bon aperçu des effets de filtrage fréquentiel du signal : la pente dV/dt de 240 V/ns des données brutes est augmentée de 25 % dans les données corrigées, et la tension de stabilisation est réduite de 15 %. L'erreur de tension de stabilisation peut certes être corrigée par l'étalonnage pour des mesures quasi-statiques, mais pas le temps de montée. Par ailleurs, la différence d'ajustement du délai de la merge box conduit à d'importantes erreurs sur la mesure de la surtension (erreur de 15 %).

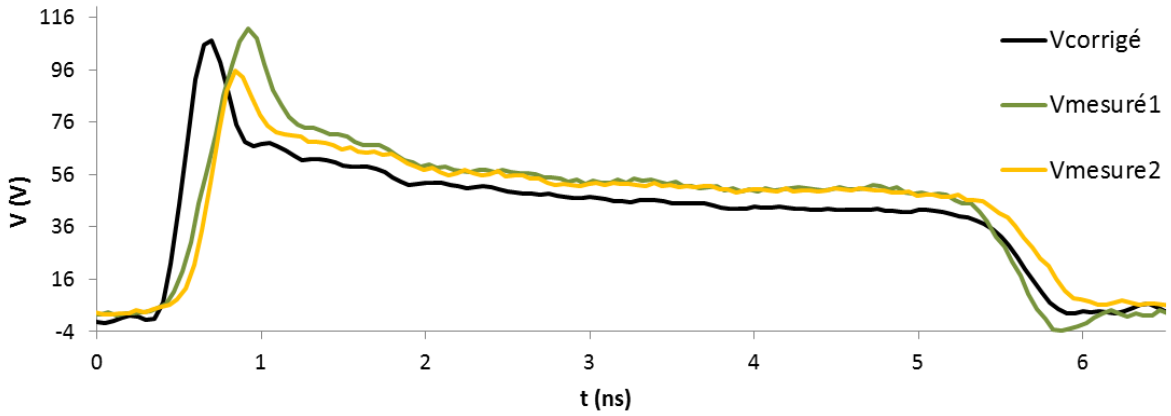

FIGURE 3.15 – Mesure de formes d'onde en tension obtenues avec la méthode décrite ($V_{corrigé}$), et comparaison avec l'outil standard $vf-TLP$ Oryx ($V_{mesuré1}$ et $V_{mesuré2}$). Un délai de 150 ps a été introduit entre les deux mesures $vf-TLP$ dans le réglage de la merge box pour visualiser son effet sur le pic de tension.

3.6.3 Limites de la méthode d'étalonnage à l'analyseur de réseau

Nous allons examiner ici deux des principales limites de la méthode présentée.

Problèmes de linéarité en puissance

La pertinence des résultats pourrait être limitée par d'éventuels problèmes de linéarité du banc de mesure pour des puissances croissantes. En effet, la mesure des coefficients de paramètres

-S est effectuée à l'analyseur de réseau, qui injecte des signaux sinusoïdaux dont la puissance n'excède pas 1 mW, alors que les signaux ESD concernés peuvent, eux, atteindre des puissances instantanées de 20 kW.

Pour valider la méthode $transient - TLP$, il est nécessaire de tester l'hypothèse selon laquelle les coefficients de paramètres -S mesurés à basse puissance sont valables à forte puissance. Celle-ci ne peut bien sûr pas être validée par des mesures au VNA, mais une autre approche peut être employée pour tester la linéarité en puissance du système à des niveaux correspondant à des impulsions TLP. Le montage simple proposé en figure 3.16 permet de comparer les tensions envoyées aux différentes extrémités du système. La comparaison des spectres de V_{ch1} et V_{ch2} aux différents niveaux de tension permet de visualiser la répartition de puissance entre les deux chemins de propagation.

Les éventuels problèmes de puissance ne semblent pas provenir des câbles coaxiaux, mais de la sonde en tension (té de pick-off) utilisée. Pour les mesures présentées sur la figure 3.17, le té utilisé est constitué de câbles coaxiaux reliés par une résistance CMS de 1.3 $k\Omega$. La répartition de la puissance est testée de 1 V à 180 V, soit de 20 mW à 648 W. Un atténuateur est placé à l'entrée de Ch 2 pour protéger l'oscilloscope. Les spectres des tensions mesurées sont calculés par FFT, et le rapport de ces deux spectres est tracé pour les différents niveaux de tension. Les résultats pour les différents niveaux de tension sont identiques, ce qui semble indiquer qu'il est pertinent d'utiliser des mesures à basse puissance (VNA) pour traiter les formes d'onde $vf - TLP$ et TLP.

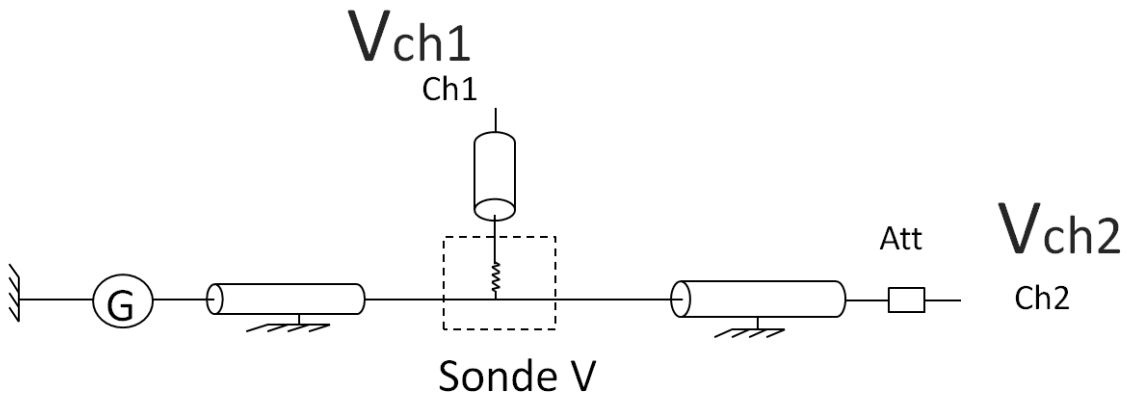

FIGURE 3.16 – Dispositif de test de la répartition en puissance des deux sorties des tés de pick-off. L'atténuateur de 30 dB en série avec l'entrée 2 de l'oscilloscope est nécessaire pour rester dans la plage de tension spécifiée.

Dans d'autres cas non détaillés ici, des effets de non-linéarité en puissance ont cependant été constatés, notamment en utilisant un té de pick-off imprimé sur PCB. Ces non-linéarités

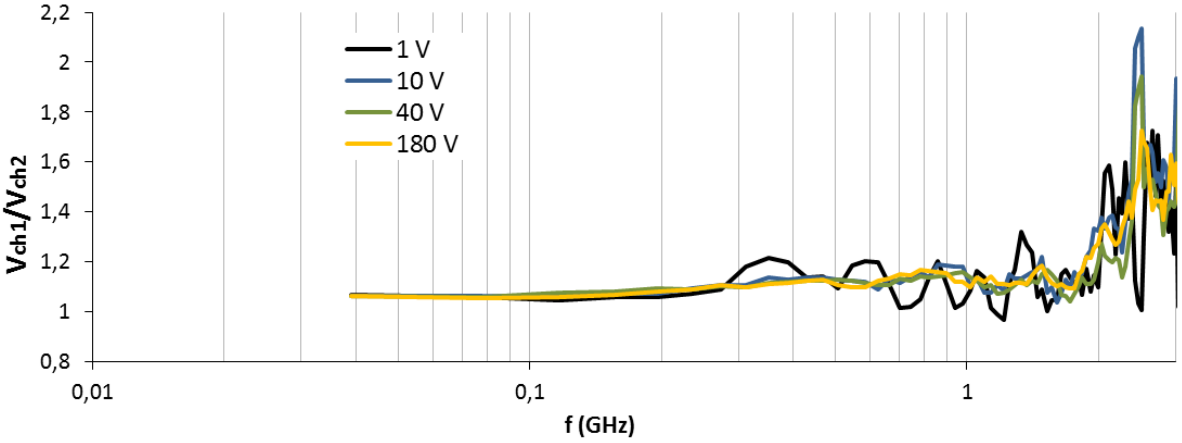

FIGURE 3.17 – Mesure de la linéarité en puissance entre 1 V et 180 V, de 340 MHz à 3 GHz sur le système utilisé le plus fréquemment. Les mesures sont effectuées dans le domaine temporel (impulsions $vf-TLP$), puis transférées dans le domaine fréquentiel par FFT.

peuvent sans doute être attribuables à une désadaptation des pistes 50 Ω à forte puissance.

Modulabilité du set-up

En plus d'éventuels problèmes de linéarité en puissance, viennent s'ajouter des préoccupations d'ordre pratique. Il est en effet difficile dans cette configuration, de modifier les éléments du banc, car chaque changement modifie les coefficients de paramètres -S. L'utilisateur peut pourtant y être amené pour diverses raisons. La principale est la nécessité d'adapter la longueur des câbles coaxiaux pour réaliser le meilleur compromis longueur de ligne/temps de montée (section 3.5.4). Une autre raison provient du fait que le matériel utilisé au laboratoire n'est pas spécifiquement dédié aux mesures $transient-TLP$, et doit donc être réutilisable pour d'autres systèmes de mesure. Enfin, l'instabilité du matériel employé dans le temps (câbles, sondes) oblige l'utilisateur à de fréquentes caractérisations des coefficients de paramètres -S, devant idéalement être effectuées avant chaque séance de mesures $transient-TLP$, comme pour toute procédure d'étalonnage.

Pour toutes ces raisons, une nouvelle approche dans l'acquisition des coefficients de paramètres -S est proposée dans la section suivante.

3.7 Outil de mesure avancé avec étalonnage à l'oscilloscope

3.7.1 Introduction

Cette partie traite d'une méthode d'étalonnage du banc de mesure ne faisant pas intervenir de mesures à l'analyseur de réseau. En effet, les étapes d'étalonnage faites à l'analyseur de réseau peuvent être également réalisées avec un oscilloscope pourvu que sa bande passante soit

suffisamment élevée. Tous types de générateurs peuvent être utilisés, mais les signaux crées doivent avoir une étendue spectrale au moins égale à celle des impulsions TLP. Le plus simple est de faire l'étalonnage avec le générateur TLP lui-même : l'étendue spectrale est alors la même pour l'étalonnage et la mesure, et on évite également d'avoir à changer de générateur entre l'étalonnage et la mesure, ce qui simplifie grandement la procédure.

3.7.2 Mesure des paramètres -S à l'oscilloscope

3.7.2.1 Principe

On a vu au chapitre 3 que la tension V_{DUT} aux bornes du composant et le courant I_{DUT} le traversant s'exprimaient entièrement dans le domaine fréquentiel à partir des tensions incidentes $V_{inc,DUT}$ et réfléchies $V_{ref,DUT}$, par les relations :

$$\begin{cases} V_{DUT}(f) = V_{inc,DUT}(f) + V_{ref,DUT}(f) \\ I_{DUT}(f) = \frac{1}{Z_C}(V_{inc,DUT}(f) - V_{ref,DUT}(f)), \end{cases} \tag{3.46}$$

avec $Z_C = 50\ \Omega$ l'impédance caractéristique de la ligne de transmission TL_2 (figure 3.2).

Ces tensions incidentes et réfléchies dans le plan du composant sont elles-même le résultat des opérations suivantes :

$$\begin{cases} V_{inc,DUT}(f) = \frac{S_{21}}{S_{31}}.V_{inc,oscillo}(f) = S_{inc}.V_{inc,oscillo}(f) \\ V_{ref,DUT}(f) = \frac{1}{S_{32}}.V_{ref,oscillo}(f) = S_{ref}.V_{ref,oscillo}(f), \end{cases} \tag{3.47}$$

avec $S_{inc} = \frac{S_{21}}{S_{31}}$ et $S_{ref} = \frac{1}{S_{32}}$ les deux grandeurs à évaluer pour caractériser complètement le système en vue de l'application de la méthode $t - TLP$. Inversons les expressions du système 3.47 :

$$\boxed{\begin{cases} S_{inc} = \frac{S_{21}}{S_{31}} = \frac{V_{inc,P2}(f)}{V_{inc,P3}(f)} \\ S_{ref} = \frac{1}{S_{32}} = \frac{V_{ref,P2}(f)}{V_{ref,P3}(f)} \end{cases}} \tag{3.48}$$

Cette équation montre que l'on peut calculer S_{inc} à partir de n'importe quelles mesures simultanées des tensions incidentes $V_{inc,P3}$ et $V_{inc,P2}$ respectivement dans les plans 3 de l'oscilloscope et 2 du composant (figure 3.2), et S_{ref} à partir de n'importe quelles mesures simultanées des tensions réfléchies $V_{ref,P3}$ et $V_{ref,P2}$.

Ceci fait, pour une impulsion incidente quelconque mesurée dans le plan 3, on peut ensuite calculer la tension incidente dans le plan du composant à partir de S_{inc} en utilisant la première ligne du système 3.47. De même, on calcule l'impulsion réfléchie dans le plan du composant à partir de S_{ref} et l'impulsion $V_{ref,3}$ mesurée en $P3$.

3.7.2.2 Mise en œuvre

Dans la pratique, des mesures simultanées dans les plans 2 et 3 sont impossibles. On ne peut pas effectuer de mesures d'impulsions incidentes ou réfléchies dans le plan 2, car celui-ci coïncide avec l'extrémité des pointes, que l'on ne peut pas connecter telles quelles à l'oscilloscope. On effectue alors les mesures à l'entrée des pointes, dans le plan appelé 4 dans la figure 3.18. On montrera dans la section suivante que les mesures obtenues de cette manière sont équivalentes à des mesures en $P2$.

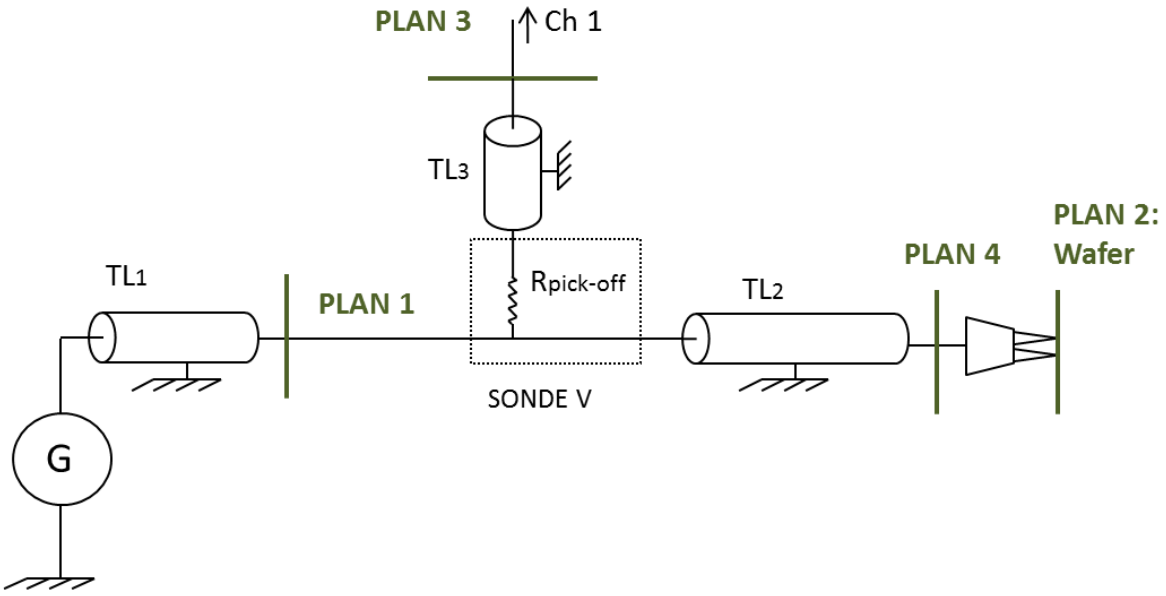

FIGURE 3.18 – Plans de propagation 1, 2 et 3 mis en jeu pour l'application de la méthode d'étalonnage « transient-TLP ». Le plan 4 coïncide avec l'extrémité du câble coaxial TL_2. Le plan 2 ne peut pas être connecté à l'entrée de l'oscilloscope, mais une méthode alternative est proposée en en utilisant le plan 4 à la place du plan 2.

On note S'_{inc} et S'_{ref} les coefficients mesurés de cette manière. S'_{inc} et S'_{ref} peuvent être utilisés à la place de S_{inc} et S_{ref} dans les équations 3.47 pour calculer les tensions incidentes et réfléchies dans le plan du composant. On verra que le seul effet de cette substitution est d'avancer $V_{inc,DUT}$ et $V_{ref,DUT}$ d'un même délai τ correspondant à la propagation dans les pointes RF.

Mesure 1 : mesure de S'_{inc}
Le paramètre S'_{inc} est défini par :

$$\boxed{S'_{inc} = \frac{V_{inc,P4}}{V_{inc,P3}} = \frac{S_{41}}{S_{31}}} \tag{3.49}$$

Il se mesure tel que montré dans la figure 3.19 : le câble coaxial TL_3 est relié à la voie 1 de l'oscilloscope, et le câble TL_2 à la voie 2. L'impulsion envoyée par le générateur est mesurée

simultanément dans le plan 3, donnant accès à $V_{inc,P3}$ et dans le plan 4, donnant accès à $V_{inc,P4}$. Le paramètre S'_{inc} correspond simplement au rapport de leurs transformées de Fourier (équation 3.49). Les formes d'onde acquises doivent au préalable subir les traitements de la section 3.5 (correction d'offset, extrapolation, fenêtrage, DFT etc ...).

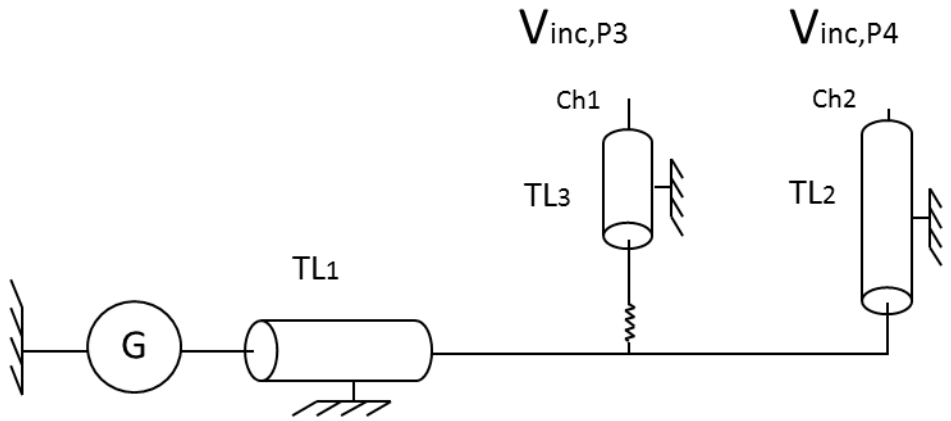

FIGURE 3.19 – Procédure de mesure du paramètre S'_{inc}.

Mesure 2 : mesure de S'_{ref}

La définition de S'_{ref} est la suivante : il s'agit du rapport entre l'impulsion incidente en $P4$ par l'impulsion réfléchie sur $P2$, mesurée dans le plan 3, avec un circuit ouvert en $P2$ (configuration de la figure 3.20).

$$\boxed{S'_{ref} = \frac{V_{inc,P4}}{(V_{ref,P3})_{P2=OPEN}}} \tag{3.50}$$

$V_{inc,P4}$ a déjà été mesuré lors de l'étape précédente, il suffit donc de récupérer la forme d'onde correspondante et de l'intégrer dans l'équation 3.50. Quant à la mesure de $(V_{ref,P3})_{P2=OPEN}$, elle s'effectue sur un circuit ouvert, selon le schéma de la figure 3.20.

Remarques :

- S'_{ref} se calcule à partir des impulsions acquises lors de deux mesures séparées, et non simultanées comme c'est le cas pour S'_{inc}. On fait l'hypothèse que le générateur d'impulsion est suffisamment répétitif pour que le paramètre calculé ne soit pas trop faussé. Cette hypothèse est vérifiée avec une erreur de l'ordre de 2 %.

- On note que l'on pourrait mesurer simultanément les impulsions incidentes et réfléchies dans le plan 3 en insérant un buffer haute impédance en série avec TL_2. Mais la bande passante de ces appareils se limite en général à 500 MHz dans le meilleur des cas, ce qui

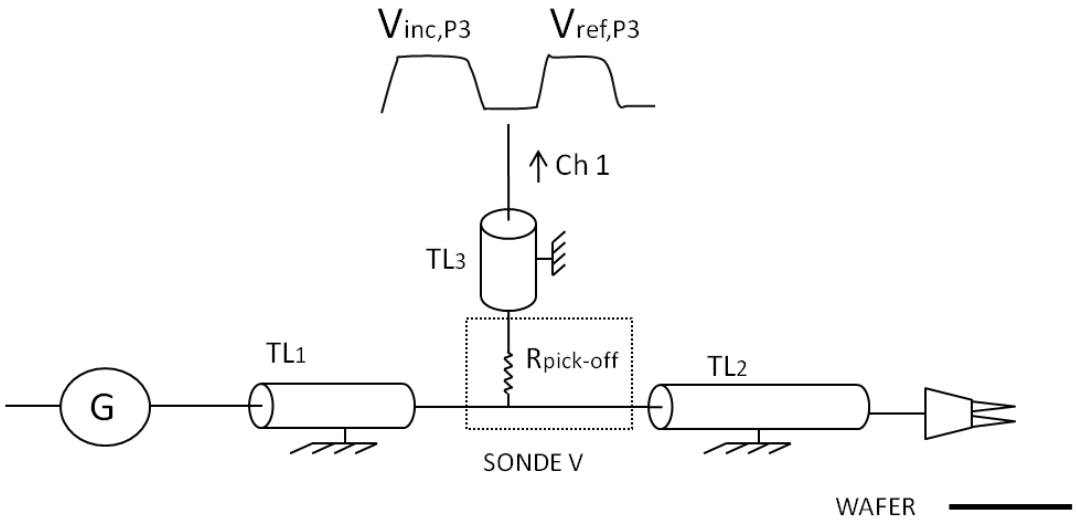

FIGURE 3.20 – Procédure de mesure du paramètre S'_{ref}. Le set-up s'apparente à celui d'une mesure $transient - TLP$ sur un circuit ouvert parfait. Le câble coaxial TL_3 est relié comme précédemment à la voie 1 de l'oscilloscope, et le câble TL_2 aux pointes RF, ces dernières étant laissées en l'air.

n'est pas suffisant dans notre application.

3.7.2.3 Validité de la substitution de S_{inc} et S_{ref} par S'_{inc} et S'_{ref}

Nous allons montrer ici pourquoi la connaissance des paramètres S'_{inc} et S'_{ref} suffit à caractériser le système. Nous verrons que l'on peut utiliser les paramètres S'_{inc} et S'_{ref} à la place de S_{inc} et S_{ref} dans les équations 3.47 sans introduire d'erreur.

S'_{inc} s'écrit, par définition (équation 3.49) :

$$S'_{inc} = \frac{S_{41}}{S_{31}} \tag{3.51}$$

Avec S_{24} la fonction de transfert des pointes RF pour un signal se propageant du plan 4 au plan 2,

$$S'_{inc} = \frac{S_{21}}{S_{24}} . \frac{1}{S_{31}} \tag{3.52}$$

Soit

$$\boxed{S'_{inc} = \frac{S_{inc}}{S_{24}}} \tag{3.53}$$

De même, le paramètre S'_{ref} s'exprime comme ci-dessous (équation 3.50) :

$$S'_{ref} = \frac{V_{inc,P4}}{(V_{ref,P3})_{P2=OPEN}} \tag{3.54}$$

Or l'impulsion incidente traverse les pointes du plan 4 vers le plan 2, avec le coefficient de transmission S_{24} correspondant :

$$V_{inc,P2} = S_{24}.V_{inc,P4} \tag{3.55}$$

Donc

$$S'_{ref} = \frac{1}{S_{24}}.\frac{V_{inc,P2}}{(V_{ref,P3})_{P2=OPEN}} \tag{3.56}$$

Soit, avec $V_{inc,P2} = (V_{inc,P2})_{P2=OPEN} = (V_{ref,P2})_{P2=OPEN}$, pour une réflexion sur le plan $P2$,

$$S'_{ref} = \frac{1}{S_{24}}.\frac{(V_{ref,P2})_{P2=OPEN}}{(V_{ref,P3})_{P2=OPEN}} \tag{3.57}$$

$$\boxed{S'_{ref} = \frac{S_{ref}}{S_{24}}} \tag{3.58}$$

Pour résumer, les paramètres S'_{inc} et S'_{ref} s'expriment à partir de S_{inc} et S_{ref} selon les relations :

$$\begin{cases} S'_{inc} = \frac{S_{inc}}{S_{24}} \\ S'_{ref} = \frac{S_{ref}}{S_{24}} \end{cases} \tag{3.59}$$

Réintégrées dans les équations 3.47, ces expressions donnent :

$$\boxed{\begin{cases} V_{inc,DUT} = S_{24}.S'_{inc}.V_{inc,oscillo} \\ V_{ref,DUT} = S_{24}.S'_{ref}.V_{ref,oscillo} \end{cases}} \tag{3.60}$$

On a vu au chapitre précédent que ces dernières pouvaient être approximées par un délai pur, c'est-à-dire que leur atténuation pouvait être négligée dans la bande de fréquence concernée (DC-3 GHz). On pourrait caractériser ce délai $\tau_{pointes}$ à l'oscilloscope, avec une précision de l'ordre de 50-60 ps. La fonction de transfert des pointes serait alors représentée dans les calculs par un terme de déphasage $e^{-i2\pi f.\tau_{pointes}}$. On voit qu'en fait il n'est pas utile de connaître ce délai, car les impulsions $V_{inc,DUT}$ et $V_{ref,DUT}$ de l'équation ci-dessus sont déphasées du même délai, et se retrouvent donc toujours superposées. Or la phase de chacune des impulsions n'est pas intéressante en soi, seul compte le fait qu'elles soient correctement superposées. On peut donc effectuer les mesures 1 et 2 décrites plus haut, et remplacer sans problèmes S_{inc} et S_{ref} par S'_{inc} et S'_{ref} dans les calculs de 3.6 et 3.10.

3.7.3 Problématiques de dé-convolution

Le problème du calcul des paramètres S'_{inc} et S'_{ref} à partir des équations 3.49 et 3.50 semble extrêmement simple et direct. Pourtant, la solution est loin d'être simple. Le problème est lié à l'opération de division dans le domaine fréquentiel. Quand le dénominateur devient très faible et se rapproche du bruit de mesure, les divisions $\frac{V_{inc,P4}}{V_{inc,P3}}$ et $\frac{V_{inc,P4}}{(V_{ref,P3})_{P2=OPEN}}$ tendent vers l'infini.

Pour nous, ce problème survient dans deux types de configuration : le numérateur et le dénominateur de ces divisions sont tous deux des transformées de Fourier d'impulsions $TLP/vf-TLP$. Dans le domaine fréquentiel, le module de telles impulsions se rapproche de fonction sinus cardinal, dont l'enveloppe décroît en $1/x$ (figure 2.5). Donc le bruit du dénominateur tend à devenir prépondérant quand la fréquence augmente, et les paramètres S_{inc} et S_{ref} prennent des valeurs artificiellement élevées. De plus, ces transformées de Fourier ont la particularité de s'annuler de façon périodique, avec une pseudo-période égale à l'inverse de la largeur de l'impulsion TLP dans le domaine temporel, comme on peut le voir dans la figure 2.5. Le problème du dénominateur nul ou très faible se présente donc aussi de façon périodique pour des valeurs discrètes de la fréquence.

Ces valeurs divergentes de S'_{inc} et S'_{ref} ont un impact important sur le signal calculé. Le résultat du calcul 3.60 donne, après FFT inverse, des formes d'onde extrêmement bruitées. Le processus permettant d'obtenir S'_{inc} à partir de $V_{inc,P4}$ et $V_{inc,P3}$ (formule 3.49), et S'_{ref} à partir de $V_{inc,P4}$ et $(V_{ref,P3})_{P2=OPEN}$ (formule 3.50) se ramène au problème de « déconvolution » [59]. La déconvolution est en effet la problématique consistant à déterminer la fonction de transfert d'un système à partir des signaux mesurés en entrée et en sortie. L'opération mathématique réalisée ici est bien une déconvolution, même si le signal au dénominateur n'est pas un signal d'entrée.

La méthode de déconvolution que nous avons retenue consiste à appliquer aux rapports $\frac{V_{inc,P4}}{V_{inc,P3}}$ et $\frac{V_{inc,P4}}{(V_{ref,P3})_{P2=OPEN}}$ à calculer des filtres mathématiques de type «Nahman-Guillaume» [60] [61]. Ces filtres sont optimisés pour les signaux impulsionnels, car ils maintiennent la causalité des signaux lors de l'opération de déconvolution. Les signaux impulsionnels ont en effet la particularité d'être « causaux », c'est-à-dire nuls à $t < 0$, et ils requièrent l'utilisation de méthodes de déconvolution causales, sans quoi d'importantes erreurs sont introduites. En pratique, l'effet des filtres de Nahman-Guillaume est d'atténuer les valeurs aberrantes que sont les pics observés dans les spectres de S'_{inc} et S'_{ref} à chaque zéro du dénominateur.

3.8 Conclusion

La nouvelle méthode de mesure présentée permet de caractériser des impulsions TLP et $vf - TLP$ allant jusqu'à 1000 V TLP (20 A sur un court-circuit), avec des durées d'impulsion de 1.25 à 100 ns, et des temps de montée de 175 ps à 10 ns. Un des avantages de cette méthode est son applicabilité aux systèmes standard $vf - TLP$, sans ajout de matériel.

La possibilité de réaliser l'étalonnage et la mesure aux mêmes niveaux de tension permet d'assurer la même fiabilité sur toute la plage de tension. En outre, l'étalonnage peut être répété avant chaque séance de mesure si besoin. Le travail de filtrage sur les spectres des signaux manipulés permet entre autres de diminuer fortement le bruit de mesure, rendant inutile le moyennage des signaux acquis à l'oscilloscope. Les mesures « mono-coup » ont pour avantage, outre un gain de temps de pouvoir visualiser des phénomènes instables, comme la dégradation d'un composant soumis à un stress TLP.

La principale limite provient de la bande passante analogique à 6 GHz de l'oscilloscope, qui détermine la résolution temporelle de 60 ps. Par ailleurs, les signaux mesurés sont codés sur 8 bit, soit 256 valeurs pour l'ensemble de la plage de mesure. En $transient - TLP$, cette plage doit s'étendre sur une largeur égale à la tension de précharge TLP du générateur pour pouvoir visualiser l'incident et le réfléchi sur une même fenêtre, ce qui amène la résolution en tension à $V_{TLP}/256$.

Les étapes d'étalonnage et de mesure ont été automatisées par Nicolas Mauran, ingénieur instrumentation au $LAAS$. Une interface *Labview* a été développée, réalisant les fonctions d'acquisition et de stockage des données, guidant l'utilisateur pas à pas à travers les étapes de mesure, et lançant le script $Matlab$.

Chapitre 4

Méthode d'étalonnage pour des mesures $transient - TLP$ sur boîtier

4.1 Introduction

A la fin de l'année 2010, le laboratoire de validation de Freescale Semiconducteurs Toulouse a souhaité se doter d'un banc de mesure $transient - TLP$ afin de compléter les étapes de test des produits par des mesures de surtension de protections ESD. J'ai rejoins cette équipe pour assumer cette tâche en y concevant un banc de mesure similaire à celui créé au $LAAS$. Une des demandes qui m'ont été faites fut de n'apporter au banc TLP existant qu'un minimum de modifications dans la conception du banc $transient - TLP$. La méthode développée a donc dû se concentrer sur un important travail de post-traitement pour pouvoir compenser les faiblesses du banc de mesure.

Contrairement au dispositif décrit dans les précédents chapitre, le banc de mesures TLP de Freescale est dédié à des caractérisations sur boîtier uniquement, pour la validation des composants finis. Les composants à valider sont des puces montées en boîtiers plats de type R-PQFP (Rectangular Plastic Quad FlatPack). Comme on peut le voir sur la photo de la figure 4.1, les boîtiers sont montés sur des cartes PCB via un « socket », permettant ainsi le test de toutes les broches des produits. Chaque broche est reliée par une ligne micro-ruban à une prise de contact pouvant être utilisée pour des mesures TLP. La connexion entre le câble coaxial par lequel arrive l'impulsion TLP et les prises de contact de la carte PCB se fait à l'aide d'un système de « grip-fils » montés directement sur un connecteur coaxial.

Le nombre de broches des puces à tester en ESD est extrêmement variable, allant de 2 à 128 selon les produits. Ainsi, une carte de test spécifique est crée pour chaque produit, cette carte des test servant aussi bien pour les mesures ESD que pour les autres types de test. Étant

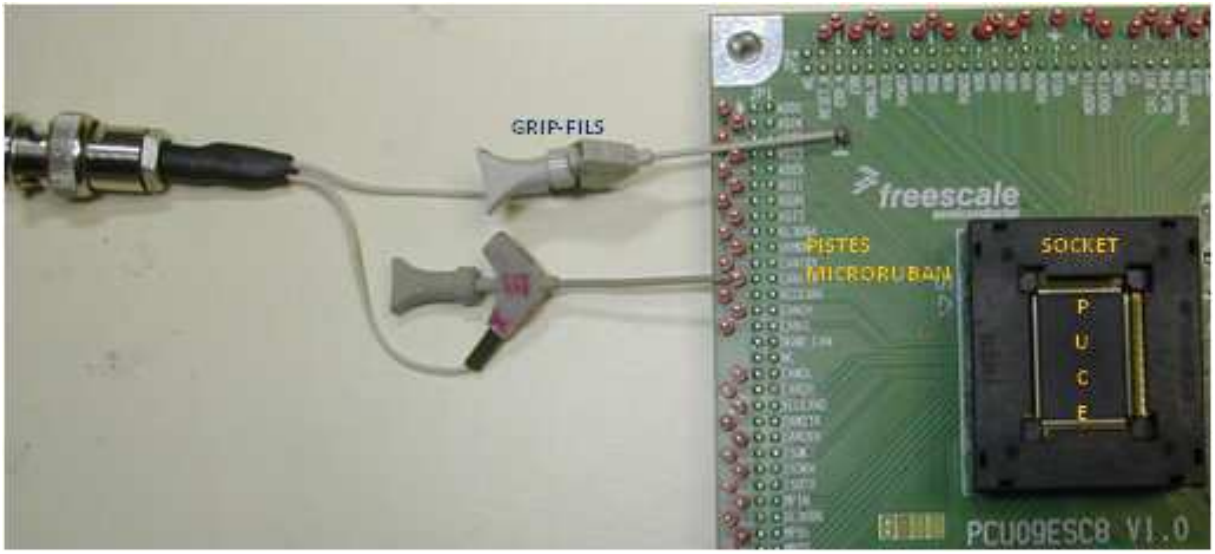

FIGURE 4.1 – Système utilisé en caractérisations TLP pour relier le câble coaxial d'acheminement de l'impulsion TLP au composant sous test. Les pinces sont constituées de fils reliés l'un à l'âme et l'autre à la gaine du connecteur coaxial, et terminés par des pinces « grip-fil ».

donné la spécificité des mesures $transient - TLP$, la solution la plus simple aurait été de créer de nouvelles cartes PCB dédiées à ces mesures. Cette solution a été écartée pour des raisons pratiques (rapidité et facilité d'utilisation) et de coûts, la caractérisation transitoire ne devant venir qu'en complément des mesures ESD standard. Dans la mesure du possible, il a donc été demandé d'utiliser les cartes existantes.

Cette demande est très contraignante, les cartes en question n'étant pas conçues pour des mesures transitoires ni fréquentielles. En effet, l'impédance caractéristique des pistes micro-ruban d'accès au composant varie de 68 Ω à 120 Ω, et sont donc fortement désadaptées par rapport aux câbles coaxiaux 50 Ω par lesquels sont acheminées les impulsions TLP. De plus, contrairement aux cartes dédiées aux mesures CEM ou à certaines cartes de mesure TLP au niveau boîtier, ces cartes ne comportent pas de port RF (SMA ou coaxial) pour y connecter les câbles coaxiaux. La connexion avec les câbles coaxiaux se fait par de simples fils électriques (grip-fils) très inductifs, qui sont source d'importantes erreurs en fréquence. L'inductance de chacun de ces fils est de l'ordre de 90 nH, et le module de leur impédance atteint 56 Ω à 100 MHz. Chacune de ces ruptures d'impédance (fils, piste micro-ruban) est une source de réflexion partielle des impulsions TLP. De plus, des réflexions multiples peuvent avoir lieu entre deux plans de réflexion, comme par exemple entre le plan du composant sous test et celui de l'entrée de la piste micro-ruban.

On voit donc que les conditions de mesure sont très différentes de celles prévalant au $LAAS$, et que le traitement appliqué au banc $vf - TLP$ du $LAAS$ ne peut être simplement reproduit

ici. Une méthode de caractérisation originale doit donc être inventée.

Nous allons nous intéresser dans une première partie aux caractéristiques du banc de mesure TLP de $Freescale$, puis la méthode d'étalonnage sera présentée en détails, et enfin quelque résultats seront présentés.

4.2 Caractéristiques physiques du banc de mesures TLP

4.2.1 Banc de mesures TLP

Le banc de mesure TLP du laboratoire de validation de Freescale Toulouse a été conçu par Nicolas Mauran sur la base de celui qu'il a développé au LAAS pendant son projet de fin d'étude [62]. Il n'est pas fondamentalement différent du banc décrit dans les précédents chapitres.

Comme on peut le voir dans la figure 4.2, le générateur d'impulsions est constitué d'une ligne de transmission chargée à travers une résistance de 100 $M\Omega$ par une alimentation haute tension. Cette ligne de charge est terminée par une série de diodes en inverse et une résistance de 50 Ω, permettant en principe d'absorber les réflexions négatives lors du retour de l'impulsion vers le générateur après réflexion sur le composant. Les impulsions générées ont un temps de montée de l'ordre de 4 ns, et une durée fixée à 100 ns. L'utilisateur peut insérer s'il le souhaite une résistance en série avec le générateur d'impulsion pour réduire l'intensité du courant injecté, et donc augmenter la sensibilité à bas courant (mesure de V_{t1}). La sonde en tension est une sonde active d'une impédance d'entrée de 10 $M\Omega$ et la sonde en courant un transformateur de courant $CT1$ de $Tektronix$. Un système de relais permet d'activer/désactiver les différentes fonctions du système : mesure de tension, ajout ou non d'une résistance série, et de mesurer la réponse $I-V$ statique entre deux impulsions.

4.2.2 Modifications du banc pour des mesures transitoires

Des modifications mineures ont été apportées au système pour pouvoir appliquer une méthode de traitement. En particulier, la longueur des impulsions a été réduite à 20 ns et une ligne de délai de 100 ns a été introduite entre la boîte de mesure et le composant sous test. Ces deux modifications visent à séparer les impulsions incidentes et réfléchies, et ainsi simplifier le traitement (voir section 3.5.2). Une deuxième ligne de délai de 100 ns a également été insérée entre le générateur d'impulsions et la boîte de mesure, de manière à retarder l'apparition des réflexions parasites sur l'écran de l'oscilloscope (voir section 3.5.3). La pente dV/dt des impulsions est assez peu affectée par la propagation à travers les câbles, étant relativement faible à l'origine. En revanche, l'ensemble de l'impulsion est atténué d'environ 25 % lors de sa propagation dans la ligne de délai. Mais cette atténuation n'est pas problématique dans la mesure où elle est connue

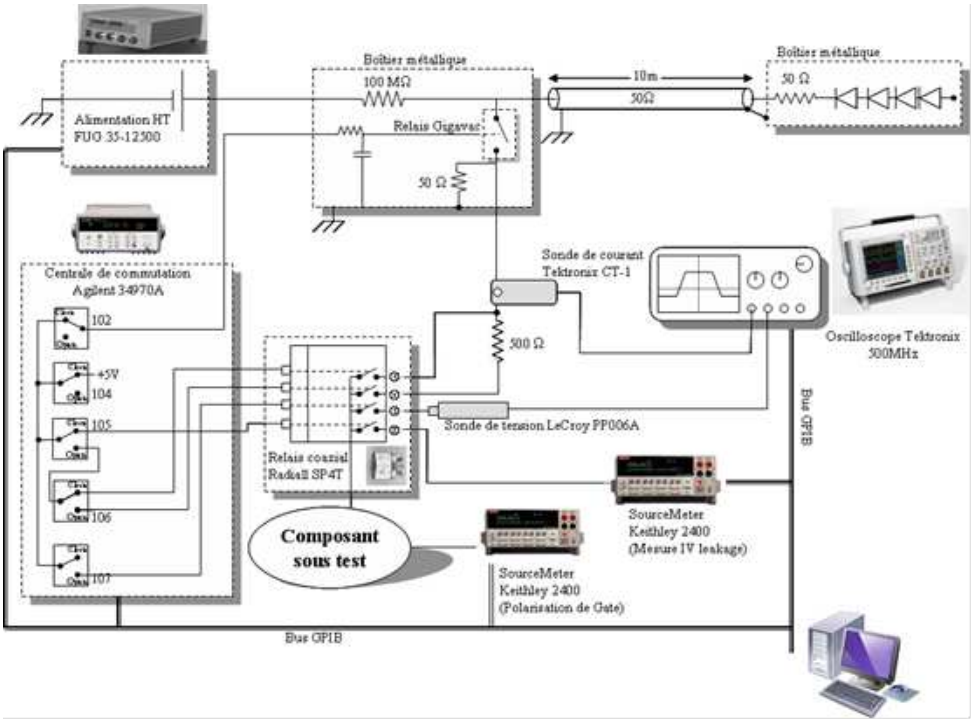

FIGURE 4.2 – Schéma global du système de mesure *TLP* utilisé à Freescale [62].

et prise en compte.

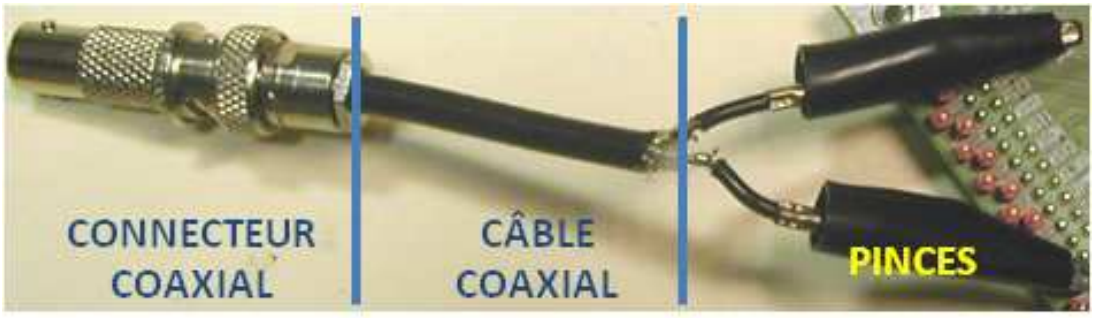

FIGURE 4.3 – Pinces fabriquées pour remplacer les pinces d'origine, trop inductives et instables.

La sonde active de 10 $M\Omega$ en tension est remplacée par un simple câble coaxial connecté sur l'entrée 50 Ω l'oscilloscope (ligne TL_2 de la figure 4.4). Les câbles TL_1 et TL_2 acheminant l'impulsion respectivement vers le *DUT* et vers l'oscilloscope sont reliés au même point en sortie du générateur.

Le câble coaxial issu du générateur, la ligne TL_1 et la ligne TL_2 (figure 4.4) forment un «té de pick-off » dont les entrées/sorties sont désadaptées : l'impédance vue depuis chacune de ces entrées/sorties est en effet réduite à 25 Ω. Par conséquent, l'impulsion incidente issue du générateur se sépare en trois composantes d'égale amplitude en arrivant au té de pick-off. L'une

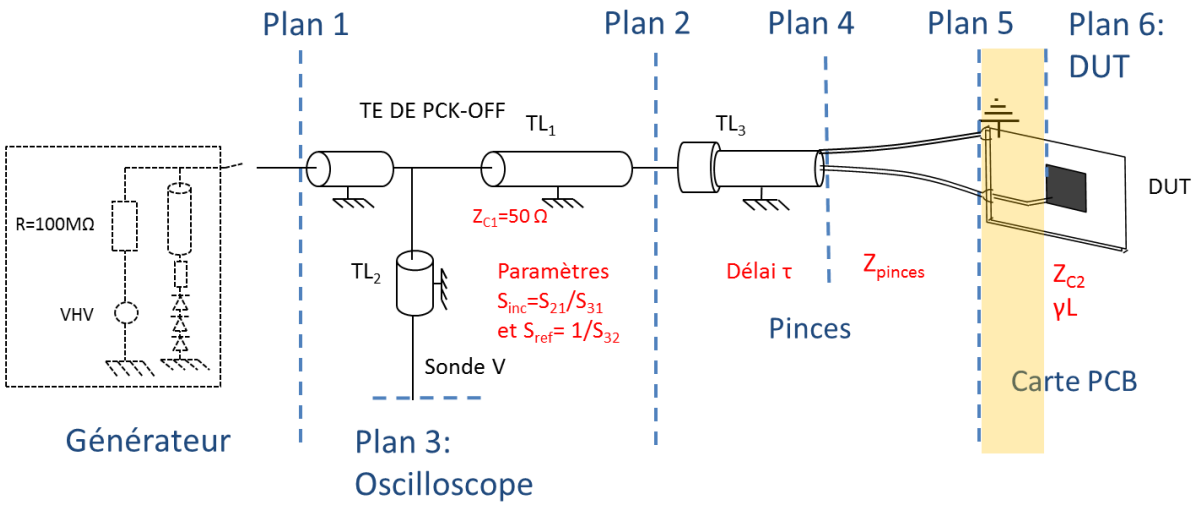

FIGURE 4.4 – Schéma équivalent représentant les plans de propagation retenus pour décomposer le système différents segments, chacun des segment pouvant être caractérisé indépendamment des autres.

d'entre elles est réfléchie vers le générateur avec un coefficient de réflexion égal à $-1/3$, et les deux autres sont conduites de manière égale sur les lignes TL_1 et TL_2. Les réflexions parasites supplémentaires induites par cette désadaptation ne posent pas problème. Elles sont en effet traitées de la même manière que celles ayant lieu sur le générateur : les lignes de 100 ns évoquées plus haut permettent de les retarder de manière à ne pas perturber le signal « utile » sur l'oscilloscope.

Les pinces reliant le câble coaxial à la carte PCB, visibles sur la figure 4.1, ont été remplacées par des pinces moins inductives (figure 4.3), de par leur plus grande section et leur longueur moins importante. Leur plus grande rigidité leur confère par ailleurs une meilleure stabilité dans le temps de leur caractéristiques électriques. Les « grip-fils » ont été remplacés par des pinces crocodile, également moins inductives.

4.3 Étalonnage du banc de mesure

4.3.1 Introduction

Le schéma de la figure 4.2 est représenté de manière simplifiée à la figure 4.4). Les éléments situés « en amont » du té de pick-off n'ont pas à être pris en compte dans l'étalonnage et peuvent être considérés comme faisant partie du générateur (voir la remarque de la section 3.2). En particulier, on va voir dans ce qui suit que la sonde en courant est devenue inutile. Elle n'est donc pas représentée ici.

L'étalonnage du banc de mesure s'effectue en trois étapes (voir schéma 4.4) :

1. Mesure des paramètres S_{inc} et S_{ref} tels que définis dans les équations 3.48. Les impulsions incidentes et réfléchies mesurées dans le plan 3 (oscilloscope) peuvent alors être calculées dans le plan 2 à partir de ces deux paramètres. Cette étape est similaire à celle décrite dans la section 3.7.2.

2. Mesure des paramètres caractéristiques des pinces. Cette tâche comporte deux étapes :
   - Mesure du coefficient de transmission de la partie en câble coaxial des pinces. Du plan 2, on se ramène alors au plan 4. Étant donnée la faible atténuation de cette portion de câble coaxial, elle peut être modélisée par une ligne de délai sans pertes, de délai τ.
   - Mesure de l'impédance Z_{pinces} des fils et des pinces crocodile. Cette donnée permet d'accéder aux ondes de tension et courant dans le plan 5. La connaissance de l'impédance caractéristique Z_{C2} de la ligne micro-ruban de la carte PCB permet d'en déduire les impulsions incidentes et réfléchies dans le même plan.

3. Mesure du paramètre « γl » de la ligne de transmission micro-ruban de la carte PCB, γ étant le paramètre de propagation dans la piste, et l la longueur de ces pistes. L'impédance caractéristique Z_{C2} de ces lignes peut, lui se calculer à partir des paramètres géométriques et physiques donnés par le layout de la carte. La connaissance de γl donne alors accès aux impulsions incidentes et réfléchies dans le plan 6, c'est-à-dire dans le plan du DUT. Connaissant Z_{C2}, on en déduit les tensions et courants V_{DUT} et I_{DUT} du composant.

4.3.2 Parties coaxiales 50 Ω

L'acquisition des paramètres S_{inc} et S_{ref} a déjà été traité en détails dans la partie 3.7.2. Elle s'effectue en deux étapes : S_{inc} s'obtient en mesurant simultanément les tensions incidentes dans les plans 2 et 3. On en déduit le paramètre S_{inc} basculant dans le domaine fréquentiel :

$$S_{inc} = \frac{S_{21}}{S_{31}} = \frac{V_{inc,P2}(f)}{V_{inc,P3}(f)} \tag{4.1}$$

Pour calculer S_{ref}, on termine la ligne TL_1 par un circuit ouvert. Les impulsions incidentes et réfléchies étant égales sur un tel circuit, on en déduit le paramètre qui nous intéresse :

$$S_{ref} = \frac{1}{S_{32}} = \frac{V_{ref,P2}(f)}{V_{ref,P3}(f)} = \frac{V_{inc,P2}(f)}{V_{ref,P3}(f)} \tag{4.2}$$

4.3.3 Étalonnage des pinces

Les pinces sont constituées d'une portion de câble coaxial et de deux fils reliés l'un à l'âme du câble, et l'autre à la gaine (figure 4.5), et terminés par des pinces « crocodile ». D'un point de vue électrique, il est pertinent de décomposer le dispositif en deux sections : l'une, représentée par le connecteur et la portion de câble, et l'autre par l'ensemble des deux fils et des pinces. En effet, si la première section peut être entièrement caractérisée par son coefficient de transmission

S_{42} et son impédance caractéristique ($Z_{C1} = 50\ \Omega$), la seconde ne peut pas être traitée de cette manière, car il n'y a pas dans les fils de phénomène de propagation comme il peut y en avoir dans une ligne de transmission. Cette dernière section doit donc être traitée comme un élément localisé, représentée par une impédance Z_{pinces}.

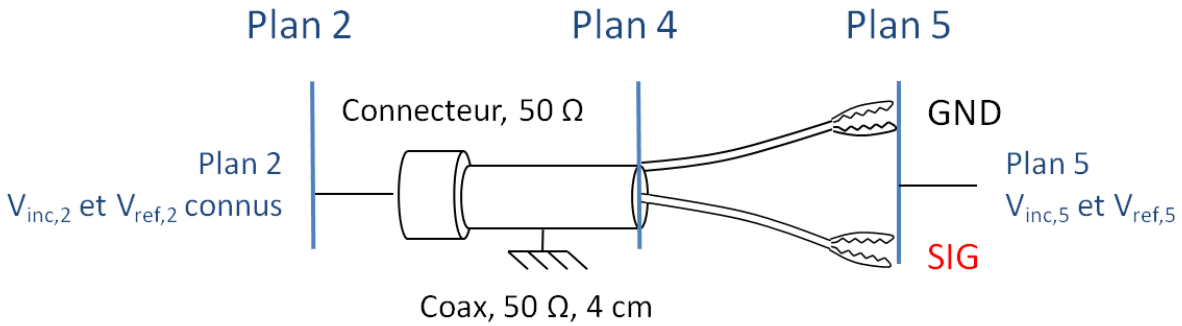

FIGURE 4.5 – Représentation schématique des pinces. Les deux étapes d'étalonnage permettent de calculer les tensions et courants en $P5$ à partir des tensions incidentes en $P2$. Dans le cas où les pinces sont connectées à une ligne de transmission, comme dans la figure 4.4, on peut calculer les tensions incidentes et réfléchies dans la ligne en $P5$.

4.3.3.1 Étape 1 : de $P2$ à $P4$. Mesure du coefficient de transmission de la partie coaxiale

La mesure s'effectue selon le schéma de la figure 4.6-a : les pinces sont laissées à l'air libre (circuit ouvert). De cette manière le courant circulant à travers les fils est nul, ainsi que la différence de potentiel le long des fils. La réflexion a donc lieu sur le plan 4, avec une coefficient de réflexion égal à 1.

On peut mesurer le coefficient de réflexion à l'entrée du dispositif (plan 2), soit avec un analyseur de réseau, soit à l'oscilloscope. Dans le cas de mesures temporelles à l'oscilloscope, le coefficient de réflexion $\Gamma(P2) = \frac{V_{ref,P2}(f)}{V_{inc,P2}(f)}$ s'obtient par des transformations de Fourier. Le résultat permet d'extraire la fonction de transfert de cette partie et donc le délai de propagation à travers la partie coaxiale. On suppose ce coefficient de transmission égal dans les deux directions de propagation :

$$S_{42} = S_{24}. \tag{4.3}$$

On peut exprimer le coefficient de réflexion au plan 2 dans la configuration de la figure 4.6-a par :

$$\Gamma(P2)_{P4=OPEN} = S_{42}.\Gamma(P4)_{P4=OPEN}.S_{24} \tag{4.4}$$

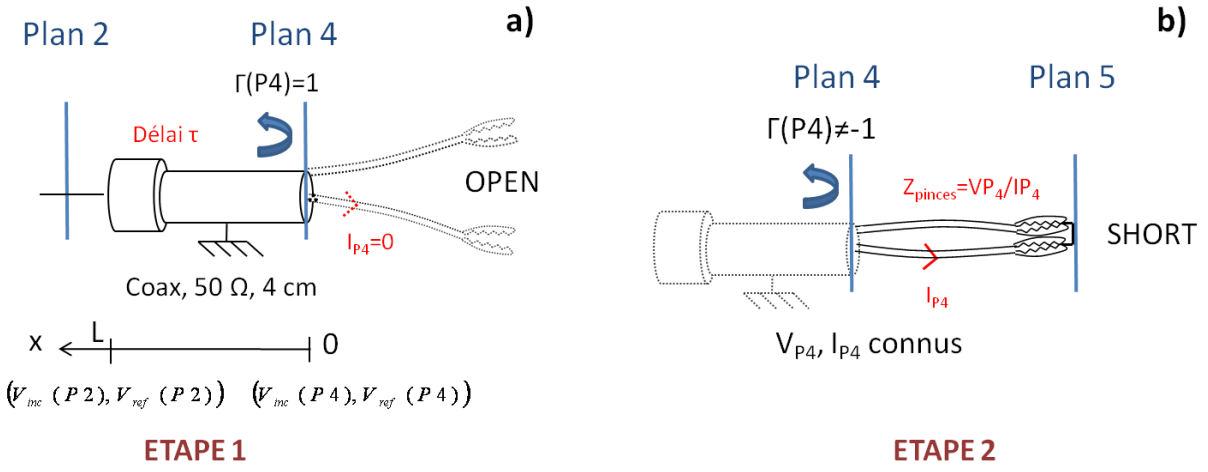

FIGURE 4.6 – Etapes d'étalonnage des pinces. Mesure du coefficient de transmission du câble coaxial (a) et de l'inductance Z_{pinces} des fils (b)

$$\Gamma(P2)_{P4=OPEN} = S_{42}^2 \tag{4.5}$$

avec $\Gamma(P4)_{P4=OPEN} = 1$. On obtient ainsi le coefficient de transmission de la partie coaxiale :

$$\boxed{S_{42} = [\Gamma(P2)_{P4=OPEN}]^{\frac{1}{2}}} \tag{4.6}$$

Connaissant S_{42}, $V_{inc,P2}$ et $V_{ref,P2}$, on en déduit simplement les tensions incidentes et réfléchies dans le plan 4 :

$$\boxed{V_{inc,P4} = S_{42}.V_{inc,P2}} \tag{4.7}$$

$$\boxed{V_{ref,P4} = \frac{1}{S_{24}}.(V_{ref,P2})_{P4=OPEN} = \frac{1}{S_{42}}.(V_{inc,P2})_{P4=OPEN}} \tag{4.8}$$

Remarque :
Comme évoqué plus haut, la faible longueur de la partie coaxiale autorise à l'approximer sans trop d'erreur par un délai pur τ, qui se calcule par :

$$\tau(\omega) = -\frac{arg(\Gamma(P2)_{P4=OPEN})}{2\omega} \tag{4.9}$$

et réécrire le coefficient de transmission :

$$S_{42} = S_{24} = e^{-j\omega\tau}. \tag{4.10}$$

On note qu'il est également possible d'éliminer l'étape de mesure du paramètre S_{42} ou du paramètre τ en remplaçant la mesure de S_{ref} (section 4.3.2) par celle du paramètre S'_{ref} tel que

décrit à la section 3.7.2.2. Le paramètre S_{inc} peut être conservé car conforme à la procédure décrite dans cette section. Ainsi, les impulsions incidentes et réfléchies mesurées dans le plan 3 sont calculées dans le plan 4 à l'aide de ces deux paramètres. Elles sont cependant déphasées d'un facteur $e^{+j\omega\tau}$ par rapport aux impulsions réelles (voir section 3.7.2.3). Ce déphasage n'est pas gênant, car la phase en elle-même de chacun des signaux incidents et réfléchis n'est d'aucune importance : seule compte la différence de phase entre ces deux signaux.

4.3.3.2 Étape 2 : de $P4$ à $P5$. Mesure de l'impédance des fils et des pinces

On l'a vu, on considère les fils et les pinces comme des éléments localisés, dont la somme des impédances est notée Z_{pinces}. Selon cette approche, les fils, les pinces, et le composant connecté à l'extrémité des pinces (plan 5), sont tous situés dans le plan 4, dans lequel a lieu la réflexion. En termes de plans de propagation, le plan 5 n'est pas distinct du plan 4. La différentiation entre les plans 4 et 5 doit être considérée comme une notation d'intérêt pratique.

Quel que soit le dispositif que l'on connecte à l'extrémité des pinces (plan 5), il se retrouve en série avec les fils, et on peut écrire :

$$\boxed{I_{P5} = I_{P4}}. \tag{4.11}$$

La tension dans le plan P5 s'écrit :

$$\boxed{V_{P5} = V_{P4} - Z_{pinces}.I_{P4}} \tag{4.12}$$

Or, les tensions et courants totaux dans le plan 4 sont calculées à partir des tensions incidentes et réfléchies $V_{inc,P4}$ et $V_{ref,P4}$ dans ce même plan :

$$\begin{cases} V_{P4} = V_{inc,P4} + V_{ref,P4} \\ I_{P4} = \frac{V_{inc,P4} - V_{ref,P4}}{Z_{C1}} \end{cases} \tag{4.13}$$

avec $Z_{C1} = 50\ \Omega$ l'impédance caractéristique du câble coaxial.

L'impédance Z_{pinces}, quant à elle, peut se mesurer en réalisant le schéma de la figure 4.6-b : on relie l'extrémité des pinces à un court-circuit, et on mesure dans le plan 3 les impulsions injectées dans le système. L'impédance recherchée s'obtient de façon immédiate à partir des tensions et courants recalculés dans le plan 4 :

$$\boxed{Z_{pinces} = \frac{(V_{P4})_{P5=SHORT}}{(I_{P4})_{P5=SHORT}}} \tag{4.14}$$

On a alors tous les éléments nécessaires pour calculer les tensions et courants dans le plan 5 avec les formules 4.11 et 4.12.

4.3.4 Étalonnage des pistes de la carte PCB

4.3.4.1 Expression de V_{DUT} et I_{DUT} en fonction des tensions et courants dans P5

Les lignes de transmission que constituent les pistes de la carte PCB peuvent être entièrement décrites par deux paramètres : leur impédance caractéristique Z_{C2} et le produit γl de leur longueur l par le paramètre de propagation complexe γ.

γ est le coefficient qui décrit la propagation des ondes progressives et régressives à travers une ligne de transmission. Le coefficient de transmission S_{65} des pistes microruban s'exprime à partir de ce dernier de la manière suivante :

$$S_{65} = e^{-\gamma l} \tag{4.15}$$

$$S_{65} = e^{-\alpha l}.e^{-j\beta l} \tag{4.16}$$

Avec α et β des coefficients réels. Le facteur $e^{-\alpha l}$, réel, est une fonction exponentielle décroissante de l. Il décrit les pertes d'atténuation à travers la ligne de transmission. $e^{-j\beta l}$ est le facteur de phase. Il contient l'information du délai de propagation des ondes à travers la piste ($\beta l = \omega\tau$).

En considérant les coefficients de transmission égaux dans les deux directions de propagation ($S_{65} = S_{56}$), on peut calculer les tensions incidentes et réfléchies dans le plan 6 à partir de celles du plan 5 et du paramètre γl de la manière suivante :

$$\boxed{\begin{cases} V_{inc,P6} = e^{-\gamma l}.V_{inc,P5} \\ V_{ref,P6} = e^{+\gamma l}.V_{ref,P5} \end{cases}}. \tag{4.17}$$

Calculons alors les tensions et courants totaux dans le plan du composant en fonction des tensions et courant du plan 5. Le composant étant situé sur une ligne d'impédance caractéristique Z_{C2}, V_{DUT} et I_{DUT}, sont données par :

$$\begin{cases} V_{DUT} = V_{P6} = V_{inc,P6} + V_{ref,P6} \\ I_{DUT} = I_{P6} = \frac{V_{inc,P6} - V_{ref,P6}}{Z_{C2}} \end{cases}, \tag{4.18}$$

soit

$$\begin{cases} V_{DUT} = e^{-\gamma l}.V_{inc,P5} + e^{+\gamma l}.V_{ref,P5} \\ I_{DUT} = \frac{e^{-\gamma l}.V_{inc,P5} - e^{+\gamma l}.V_{ref,P5}}{Z_{C2}} \end{cases}. \tag{4.19}$$

On a à présent une expression de V_{DUT} et I_{DUT} en fonction des tensions incidentes et réfléchies dans le plan 5. Pour évaluer $V_{inc,P5}$ et $V_{ref,P5}$, on part de V_{P5} et I_{P5} :

$$\begin{cases} V_{P5} = V_{inc,P5} + V_{ref,P5} \\ I_{P5} = \frac{V_{inc,P5} - V_{ref,P5}}{Z_{C2}} \end{cases} \tag{4.20}$$

En inversant le système, on obtient :

$$\begin{cases} V_{inc,P5} = \frac{V_{P5} + Z_{C2} I_{P5}}{2} \\ V_{ref,P5} = \frac{V_{P5} - Z_{C2} I_{P5}}{2} \end{cases} \tag{4.21}$$

On injecte alors 4.21 dans 4.19 :

$$\begin{cases} V_{DUT} = e^{-\gamma l}.\frac{V_{P5} + Z_{C2} I_{P5}}{2} + e^{+\gamma l}.\frac{V_{P5} - Z_{C2} I_{P5}}{2} \\ I_{DUT} = \frac{1}{Z_{C2}}.\left[e^{-\gamma l}.\frac{V_{P5} + Z_{C2} I_{P5}}{2} - e^{+\gamma l}.\frac{V_{P5} - Z_{C2} I_{P5}}{2}\right] \end{cases} . \tag{4.22}$$

Soit, après développement :

$$\boxed{\begin{cases} V_{DUT} = V_{P5}.cosh(\gamma l) - Z_{C2} I_{P5}.sinh(\gamma l) \\ I_{DUT} = I_{P5}.cosh(\gamma l) - \frac{V_{P5}}{Z_{C2}}.sinh(\gamma l) \end{cases}} . \tag{4.23}$$

4.3.4.2 Évaluation de l'impédance caractéristique des pistes

L'impédance caractéristique Z_{C2} des pistes micro-ruban peut être calculée à partir des paramètres géométriques de la ligne et du substrat, et de la constante diélectrique du substrat. Un article a été publié en 1967, qui constitue une très bonne introduction aux lignes de transmission microruban [63]. L'auteur débute avec les équations de propagation à travers un fil au-dessus d'un plan de masse. En y incorporant un travail théorique et expérimental, il calcule les équations décrivant un circuit rectangulaire séparé d'une ligne de cuivre par un diélectrique.
L'équation suivante donne l'impédance caractéristique d'une telle ligne en fonction de l'épaisseur H du substrat, l'épaisseur T de la ligne métallique, sa largeur W, et la permittivité relative ϵ_r du substrat (figure 4.7)

$$Z_c = \frac{87}{\epsilon_r + 1.41}.ln\left(\frac{5.98H}{0.8W + T}\right) \tag{4.24}$$

On peut facilement extraire ces paramètres du layout de la carte, et calculer l'impédance caractéristique correspondante.

On note qu'il nous serait également possible de retrouver cette valeur par le calcul, mais seulement au prix d'une étape d'étalonnage supplémentaire. En effet l'ajout d'une inconnue rend nécessaire celui d'une équation supplémentaire. Une mesure sur un court-circuit dans le plan 6 pourrait apporter cette équation. Mais dans notre cas, l'impédance caractéristique est en général la même pour toutes les pistes d'une même carte PCB. Il suffit d'effectuer ce calcul une fois pour toutes les pistes de la carte.

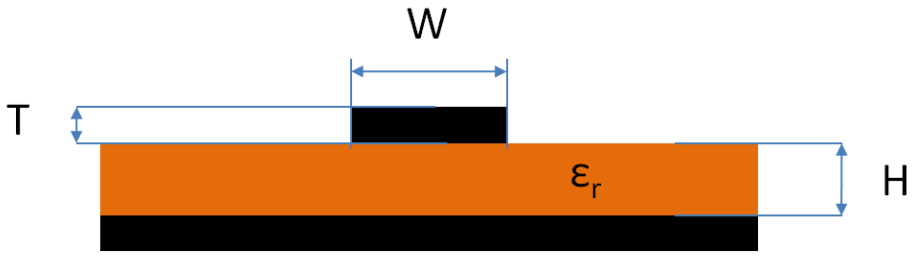

FIGURE 4.7 – Paramètres géométriques et physiques des lignes de transmission microruban permettant le calcul de leur impédance caractéristique.

4.3.4.3 Mesure du paramètre γl

Pour évaluer le paramètre γl, on choisit d'effectuer une mesure sur circuit ouvert, c'est-à-dire avec tout les éléments de la figure 4.4, sans composant. En utilisant les résultats des sections 4.3.2 et 4.3.3, on calcule les tensions et courants dans P5, P6 étant ouvert. γl sera exprimé en fonction de ces grandeurs.

On connait également les tensions incidentes et réfléchies dans le plan 5 grâce à l'équation 4.21. L'équation 4.17 nous donne alors l'expression des impulsions incidentes et réfléchies dans le plan 6 du DUT.

Sur un circuit ouvert $(P6)$, les impulsions incidentes et réfléchies sont égales, ce qui conduit, d'après 4.17 à :

$$e^{-\gamma l}.(V_{inc,P5})_{P6=OPEN} = e^{\gamma l}.(V_{ref,P5})_{P6=OPEN} \tag{4.25}$$

Soit, avec $\gamma = \alpha + j\beta$, et α, β réels,

$$e^{2\alpha l}.e^{2j\beta l} = \left|\frac{(V_{inc,P5})_{P6=OPEN}}{(V_{ref,P5})_{P6=OPEN}}\right| .e^{j.arg\left(\frac{(V_{inc,P5})_{P6=OPEN}}{(V_{ref,P5})_{P6=OPEN}}\right)} \tag{4.26}$$

$$\Leftrightarrow \begin{cases} \alpha l = \frac{1}{2}.ln\left(\left|\frac{(V_{inc,P5})_{P6=OPEN}}{(V_{ref,P5})_{P6=OPEN}}\right|\right) \\ \beta l = \frac{1}{2}.arg\left(\frac{(V_{inc,P5})_{P6=OPEN}}{(V_{ref,P5})_{P6=OPEN}}\right) \end{cases} \tag{4.27}$$

Soit :

$$\gamma l = \frac{1}{2}.ln\left(\left|\frac{(V_{inc,P5})_{P6=OPEN}}{(V_{ref,P5})_{P6=OPEN}}\right|\right) + \frac{1}{2}.j.arg\left(\frac{(V_{inc,P5})_{P6=OPEN}}{(V_{ref,P5})_{P6=OPEN}}\right) \tag{4.28}$$

En utilisant 4.21 dans 4.28, on obtient :

$$\boxed{\begin{array}{rcl} \gamma l & = & \frac{1}{2}.ln\left(\left|\frac{(V_{P5})_{P6=OPEN}+Z_{C2}(I_{P5})_{P6=OPEN}}{(V_{P5})_{P6=OPEN}-Z_{C2}(I_{P5})_{P6=OPEN}}\right|\right) \\ & + & \frac{1}{2}.j.arg\left(\frac{(V_{P5})_{P6=OPEN}+Z_{C2}(I_{P5})_{P6=OPEN}}{(V_{P5})_{P6=OPEN}-Z_{C2}(I_{P5})_{P6=OPEN}}\right) \end{array}} \tag{4.29}$$

4.3.5 Expression compacte pour la tension et le courant dans le plan du composant

On a à présent décomposé les étapes permettant de caractériser les différentes parties du banc de mesure. On se propose de synthétiser ces résultats en une expression plus globale. L'équation 4.23 exprime les tensions et courants de composant à partir des formes d'onde ramenées dans le plan 5, évaluées grâce aux étapes précédentes. Ces mêmes formes d'onde s'expriment également en fonctions de celles des plans précédents. Nous allons trouver une formulation donnant directement V_{DUT} et I_{DUT} à partir des impulsions mesurées à l'oscilloscope et des paramètres du système. On peut réécrire l'équation 4.23 avec $V_{P5} = V_{P4} - Z_{pinces}.I_{P4}$ (équation 4.12) et $I_{P5} = I_{P4}$ (équation4.11) :

$$\begin{cases} V_{DUT} = (V_{P4} - Z_{pinces}.I_{P4}).cosh(\gamma l) - Z_{C2}I_{P4}.sinh(\gamma l) \\ I_{DUT} = I_{P4}.cosh(\gamma l) - \frac{(V_{P4}-Z_{pinces}.I_{P4})}{Z_{C2}}.sinh(\gamma l) \end{cases} \tag{4.30}$$

Évaluons $V_{P4} - Z_{pinces}.I_{P4}$ à l'aide des décompositions du système d'équations 4.13 en ondes incidentes et réfléchies en $P4$. Après développement, on a :

$$V_{P4} - Z_{pinces}.I_{P4} = \left(1 - \frac{Z_{pinces}}{Z_{C1}}\right).V_{inc,P4} + \left(1 + \frac{Z_{pinces}}{Z_{C1}}\right).V_{ref,P4}, \tag{4.31}$$

Reporté dans 4.30, cela donne, après calculs :

$$V_{DUT} = \left[\left(1 - \frac{Z_{pinces}}{Z_{C1}}\right)cosh(\gamma l) - \frac{Z_{C2}}{Z_{C1}}sinh(\gamma l)\right]V_{inc,P4} \tag{4.32}$$

$$+ \left[\left(1 + \frac{Z_{pinces}}{Z_{C1}}\right)cosh(\gamma l) + \frac{Z_{C2}}{Z_{C1}}sinh(\gamma l)\right]V_{ref,P4} \tag{4.33}$$

et :

$$I_{DUT} = \left[\frac{cosh(\gamma l)}{Z_{C1}} + \left(\frac{Z_{pinces}}{Z_{C1}.Z_{C2}} - \frac{1}{Z_{C2}}\right)sinh(\gamma l)\right]V_{inc,P4} \tag{4.34}$$

$$- \left[\frac{cosh(\gamma l)}{Z_{C1}} + \left(\frac{Z_{pinces}}{Z_{C1}.Z_{C2}} + \frac{1}{Z_{C2}}\right)sinh(\gamma l)\right]V_{inc,P4} \tag{4.35}$$

On rappelle que par définition des paramètres S_{inc} et S_{ref} :

$$\begin{cases} V_{inc,P2} = S_{inc}.V_{inc,P3} \\ V_{ref,P2} = S_{ref}.V_{ref,P3} \end{cases} \tag{4.36}$$

Avec les équations ci-dessus et les formules 4.7 et 4.8 exprimant les tensions incidentes et réfléchies dans le plan 4 en fonction de celles du plan 2, on a :

$$\begin{cases} V_{inc,P4} = S_{42} S_{inc} . V_{inc,P3} \\ V_{ref,P4} = \frac{S_{ref}}{S_{42}} . V_{ref,P3} \end{cases} \tag{4.37}$$

On en déduit alors l'expression des tensions et courants de composant en fonction des impulsions mesurées à l'oscilloscope et des paramètres du système de mesure : Z_{C1}, S_{inc}, S_{ref}, S_{42}, Z_{pinces}, Z_{C2} et γl :

$$\begin{cases} V_{DUT} = S_{inc} S_{42} . \left[\left(1 - \frac{Z_{pinces}}{Z_{C1}} \right) cosh(\gamma l) - \frac{Z_{C2}}{Z_{C1}} sinh(\gamma l) \right] V_{inc,P3} \\ \qquad + \frac{S_{ref}}{S_{42}} . \left[\left(1 + \frac{Z_{pinces}}{Z_{C1}} \right) cosh(\gamma l) + \frac{Z_{C2}}{Z_{C1}} sinh(\gamma l) \right] V_{ref,P3} \\ I_{DUT} = S_{inc} S_{42} . \left[\frac{cosh(\gamma l)}{Z_{C1}} + \left(\frac{Z_{pinces}}{Z_{C1} . Z_{C2}} - \frac{1}{Z_{C2}} \right) sinh(\gamma l) \right] V_{inc,P3} \\ \qquad - \frac{S_{ref}}{S_{42}} . \left[\frac{cosh(\gamma l)}{Z_{C1}} + \left(\frac{Z_{pinces}}{Z_{C1} . Z_{C2}} + \frac{1}{Z_{C2}} \right) sinh(\gamma l) \right] V_{ref,P3} \end{cases} \tag{4.38}$$

4.3.6 Traitement des réflexions multiples entre les plans 5 et 6

Les possibles réflexions de $V_{ref,DUT}$ sur le plan 5 puis à nouveau le plan 6 n'ont pas été abordées dans les sections précédentes. Elles existent pourtant et ne doivent pas être négligées. Elles sont dues à la rupture d'impédance représentée par les pinces et le câble coaxial d'impédance caractéristique $Z_{C1} = 50\ \Omega$ par rapport à l'impédance caractéristique des lignes microruban. Vu depuis la ligne microruban, le coefficient de réflexion sur le plan des pinces est égal à :

$$S_{55} = \frac{Z_{C1} + Z_{pinces} - Z_{C2}}{Z_{C1} + Z_{pinces} + Z_{C2}} \tag{4.39}$$

Plus l'inductance des pinces est élevée, mais surtout plus la désadaptation $|Z_{C1} - Z_{C2}|$ est grande, et plus la réflexion sur le plan 5 est importante. Par exemple, même en négligeant l'impédance Z_{pinces} des pinces et des fils, une impédance caractéristique de 61 Ω de la ligne microruban suffirait à introduire un coefficient de réflexion de 10 % à l'interface pinces-piste, vue depuis la piste.

On montre cependant qu'en fait, les réflexions multiples sont prises en compte implicitement dans toutes les étapes de calcul précédentes, et en particulier l'étape 4.3.4. Ainsi, dans l'équation 4.21, l'onde « incidente » $V_{inc,P5}$ est en réalité la somme exprimée dans $P5$ de toutes les ondes progressives issues des réflexions multiples entre les plans 5 et 6, et l'onde « réfléchie » la somme des ondes régressives dans le même plan.

Il convient alors d'effectuer une précision : contrairement aux ondes incidentes calculées dans les plans 2, 3 et 4, $V_{inc,P5}$ n'est pas une image directe de l'onde générée dans le plan 1. En effet,

elle est aussi fonction des réflexions successives sur le composant, qui elles-mêmes dépendent de son impédance instantanée Z_{DUT}. Il en résulte que l'impulsion « incidente » totale arrivant sur le composant à un instant donné dépend de sa réponse antérieure. Il reste cependant pertinent de parler ici aussi d'onde incidente pour qualifier $V_{inc,P5}$, car en définitive, c'est bien cette superposition d'ondes progressives qui vient réellement agresser le composant. De même, on parle dans la suite d'onde réfléchie $V_{ref,P5}$ pour la superposition d'ondes régressives dans $P5$.

Dans le paragraphe suivant, on établit une formulation mathématiquement équivalente de $V_{inc,P6}$ et $V_{ref,P6}$ mettant en évidence la contribution des réflexions multiples dans les calculs de la partie 4.3.4.

La matrice de paramètres S décrivant le passage du plan 4 au plan 5 est définie par :

$$\begin{pmatrix} V_{ref,P4} \\ V_{inc,P5} \end{pmatrix} = [S_{4\to5}] \begin{pmatrix} V_{inc,P4} \\ V_{ref,P5} \end{pmatrix}, \tag{4.40}$$

avec

$$[S_{4\to5}] = \begin{bmatrix} S_{44} & S_{45} \\ S_{54} & S_{55} \end{bmatrix}. \tag{4.41}$$

On peut montrer en utilisant les relations 4.11, 4.12, 4.13 et 4.20 que cette matrice s'écrit :

$$[S_{4\to5}] = \begin{bmatrix} \frac{Z_{C2}+Z_{pinces}-Z_{C1}}{Z_{C2}+Z_{pinces}+Z_{C1}} & \frac{2.Z_{C1}}{Z_{C2}+Z_{pinces}+Z_{C1}} \\ \\ \frac{2.Z_{C2}}{Z_{C2}+Z_{pinces}+Z_{C1}} & \frac{Z_{C1}+Z_{pinces}-Z_{C2}}{Z_{C2}+Z_{pinces}+Z_{C1}} \end{bmatrix} \tag{4.42}$$

La tension incidente dans le plan 5 se décompose de la manière suivante :

$$V_{inc,P5} = S_{54}.V_{inc,P4} + S_{55}.V_{ref,P5} \tag{4.43}$$

On peut également écrire $V_{inc,P5}$ comme étant la somme d'une onde incidente « primaire » ${V_{inc,P5}}^{(0)}$ et d'une infinité de réflexions secondaires ${V_{inc,P5}}^{(k)}$ sur le plan 5 :

$${V_{inc,P5}} = {V_{inc,P5}}^{(0)} + \sum_{k=1}^{\infty} {V_{inc,P5}}^{(k)} \tag{4.44}$$

L'onde « primaire » non encore perturbée par les réflexions sur le DUT et sur les pinces est une image de l'onde incidente mesurée à l'oscilloscope. Par identification, on obtient :

$${V_{inc,P5}}^{(0)} = \frac{2.Z_{C2}}{Z_{C1}+Z_{C2}+Z_{pinces}}.V_{inc,P4} \tag{4.45}$$

$${V_{inc,P5}}^{(0)} = \frac{2.Z_{C2}}{Z_{C1}+Z_{C2}+Z_{pinces}}.S_{42}.S_{inc}.V_{inc,P3} \tag{4.46}$$

et le second terme de 4.44 est égal à l'onde réfléchie en $P5$, $V_{ref,P5}$, réduite d'un facteur S_{55} :

$$\sum_{k=1}^{\infty} {V_{inc,P5}}^{(k)} = S_{55}.V_{ref,P5} \tag{4.47}$$

avec S_{55} le coefficient de réflexion sur $P5$ vu depuis la ligne microruban.

4.3.7 Bilan des étapes d'étalonnage du système de mesure

Les cinq paramètres à mesurer pour pouvoir effectuer des mesures dans le plan 6 et les étapes d'étalonnage associées à ces paramètres sont résumées dans la figure 4.8.

Toutes les opérations sur les formes d'ondes menant au calcul des tension et courant aux bornes du composant à partir des tensions acquises à l'oscilloscope sont traitées par des fonctions *Matlab* que j'ai implémentées. Un important travail a été fourni par Alain Salles, ingénieur instrumentation *ESD* à Freescale, pour automatiser ces étapes d'étalonnage et de mesure. Un programme *Labview* a été développé par ses soins pour guider l'utilisateur pas à pas à travers les étapes de mesures, et appeler les fonctions *Matlab*.

4.4 Résultats

Comme dans la partie 3.6, on peut évaluer l'erreur associée à la méthode d'étalonnage en effectuant des mesures sur des structures connues : typiquement des circuits ouverts et court-circuits.

Les mesures $transient - TLP$ doivent pouvoir s'adapter à deux types de boîtiers : le cas de boîtier le plus simple est celui des boîtiers « DIL », dont les broches sont suffisamment grandes pour que l'on puisse venir y connecter directement les pinces. Dans ce cas, l'étalonnage s'arrête au plan 5, qui coïncide avec le plan du composant sous test. On peut tester l'efficacité des étapes d'étalonnage correspondantes en mesurant, comme dans la figure 4.9 le courant sur un circuit ouvert et la tension sur un court-circuit. Contrairement au cas de la partie 3.6, le plan de mesure n'est ici pas situé dans une ligne de transmission, mais sur un élément localisé (pinces). Il n'y a pas dans ce cas de propagation d'ondes incidentes et réfléchies, et on ne peut donc pas définir l'erreur à partir de la superposition ou non de ces deux impulsions. La définition retenue pour l'erreur instantanée sur un circuit ouvert est la suivante :

$$Erreur_{CO}(\%) = 100. \left| \frac{Z_{C1}.I_{CO}}{max(V_{CO})} \right| \tag{4.48}$$

Dans le cas du court-circuit, le pourcentage d'erreur est donné par :

$$Erreur_{CC}(\%) = 100. \left| \frac{V_{CC}}{Z_{C1}.max(I_{CC})} \right| \tag{4.49}$$

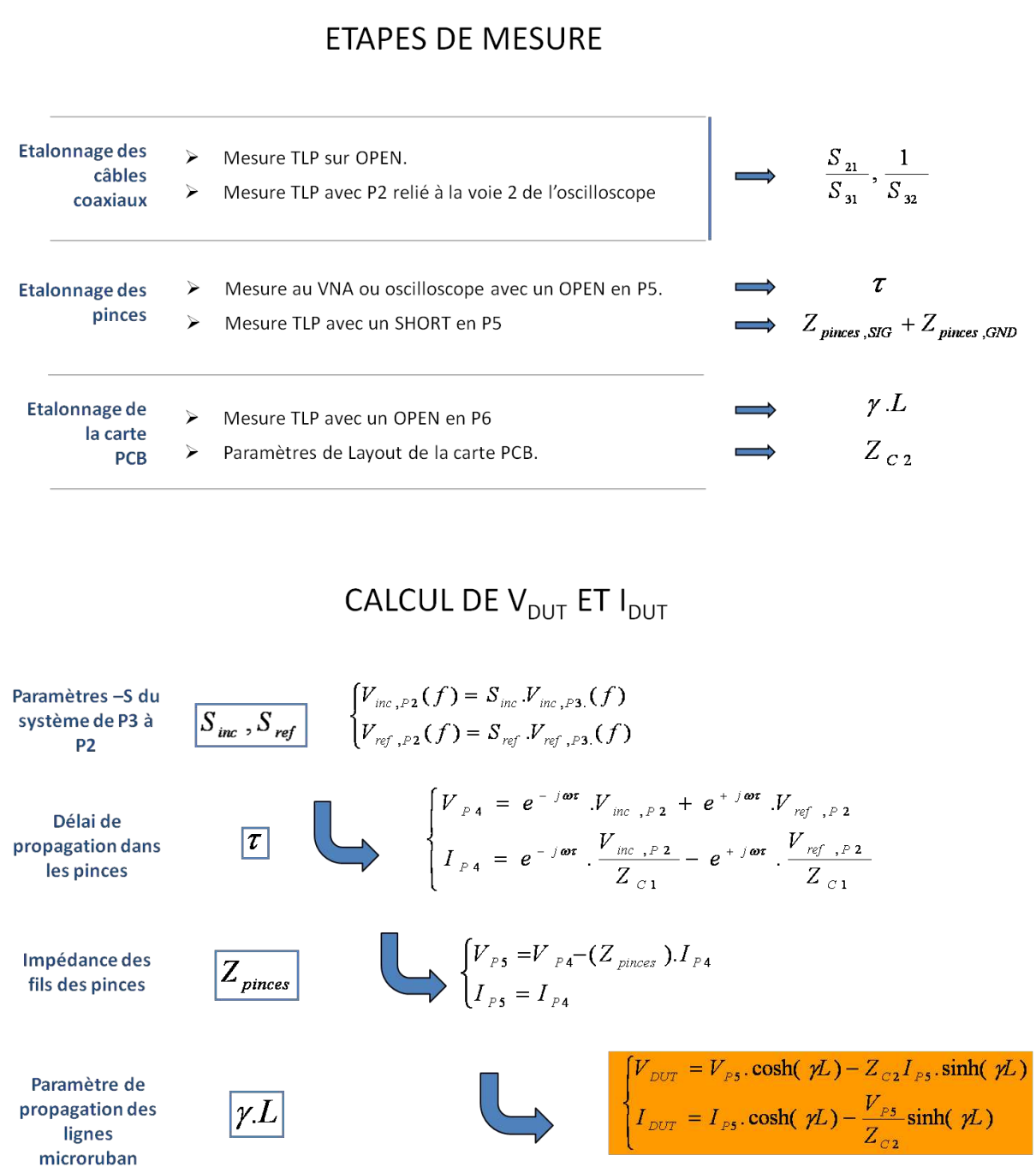

FIGURE 4.8 – Bilan des étapes de mesure permettant l'étalonnage du banc TLP. Les cinq étapes de mesure présentées permettent d'apporter les équations nécessaires à la détermination des cinq inconnues que sont les paramètres S_{inc}, S_{ref}, τ, Z_{pinces} et γl. Le cas de l'impédance caractéristique Z_{C2} des pistes microruban est traité à part, celle-ci étant extraite des paramètres de layout de la carte. On recalcule ensuite les tensions et courants dans les plans successifs en utilisant chacun de ces paramètres.

L'erreur obtenue est inférieure à 3 % dans un cas comme dans l'autre.

La deuxième configuration à tester concerne la mesure sur PCB pour des boîtiers de type

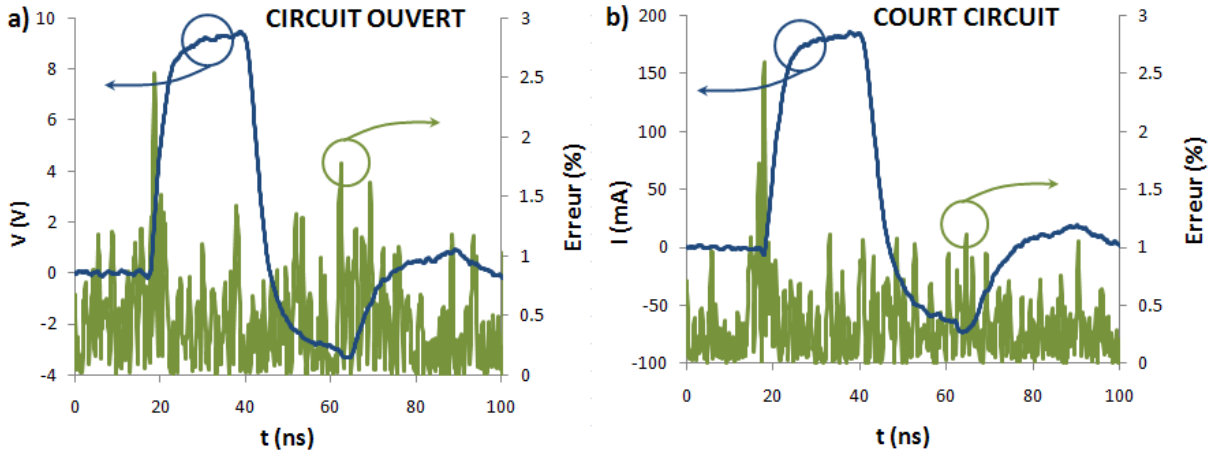

FIGURE 4.9 – Mesure de la tension et du courant sur un circuit ouvert (a) et sur un court-circuit (b) en bout des pinces (plan 5) pour une impulsion de 10 V TLP. Le courant de circuit ouvert et la tension de court-circuit servent à calculer le pourcentage d'erreur dans chacun des cas. Celui-ci est de l'ordre de 2.5 % et atteint sa valeur maximale pendant le front montant de l'impulsion.

$R-PQFP$, qui met en jeu toutes les étapes d'étalonnage. Dans ce cas-là, le plan de mesure est situé sur un ligne de transmission, à travers laquelle se propagent une impulsion incidente et une impulsion réfléchie. Ces dernières sont comme on le sait égales en $P6$ si le composant sous test est un circuit ouvert, et opposées s'il s'agit d'un court-circuit. On se sert de ceci pour définir le pourcentage d'erreur. Sur un circuit ouvert, on le calcule par :

$$Erreur_{CO}(\%) = 100. \left| \frac{V_{inc,P6} - V_{ref,P6}}{max(V_{inc,P6} + V_{ref,P6})} \right|, \tag{4.50}$$

et sur un court-circuit par :

$$Erreur_{CC}(\%) = 100. \left| \frac{V_{inc,P6} + V_{ref,P6}}{max(V_{inc,P6} - V_{ref,P6})} \right| \tag{4.51}$$

La figure 4.10 décrit le résultat d'une mesure sur un circuit ouvert et sur un court-circuit dans le plan 6. Le court-circuit est obtenu au moyen d'une bande adhésive conductrice (cuivre) collée sur la face arrière de la carte, là où ressortent les contacts métalliques de chaque broche. Le contact entre la broche testée et la masse est réalisé par l'intermédiaire de cette bande adhésive, et n'est donc pas parfait.

Sur le circuit ouvert en $P6$, on obtient une erreur de 1 % environ. Ce résultat, inférieur à celui obtenu avec un circuit ouvert en $P5$ peut paraître surprenant. Il est en réalité compréhensible, puisque le paramètre extrait lors de l'étalonnage de la carte PCB est lui aussi calculé à partir d'une mesure sur circuit ouvert en $P6$. Ainsi, lors d'une mesure sur un autre circuit ouvert, il est normal, par réciprocité de la formule 4.29 que les tensions mesurées apparaissent correctes. L'erreur obtenue ne peut être imputable qu'à une répétitivité non parfaite du générateur et du

système.

L'erreur obtenue sur le court-circuit, en revanche est plus élevée : elle est de l'ordre de 5 %. Trois facteurs peuvent expliquer ce résultat :

- Premièrement, l'étalonnage des pistes de la carte PCB n'est pas réalisé sur un court-circuit mais sur un circuit ouvert, et donc on ne bénéficie pas ici de l'effet décrit au-dessus.
- Aussi, le résultat sur un circuit fermé est beaucoup plus affecté par les éventuelles erreurs dans la mesure de l'impédance Z_{pinces} des pinces que ne l'est le circuit ouvert, la différence de potentiel aux bornes des pinces étant dans ce cas bien plus importante.
- Enfin, comme on l'a vu, le court-circuit réalisé n'est pas parfait.

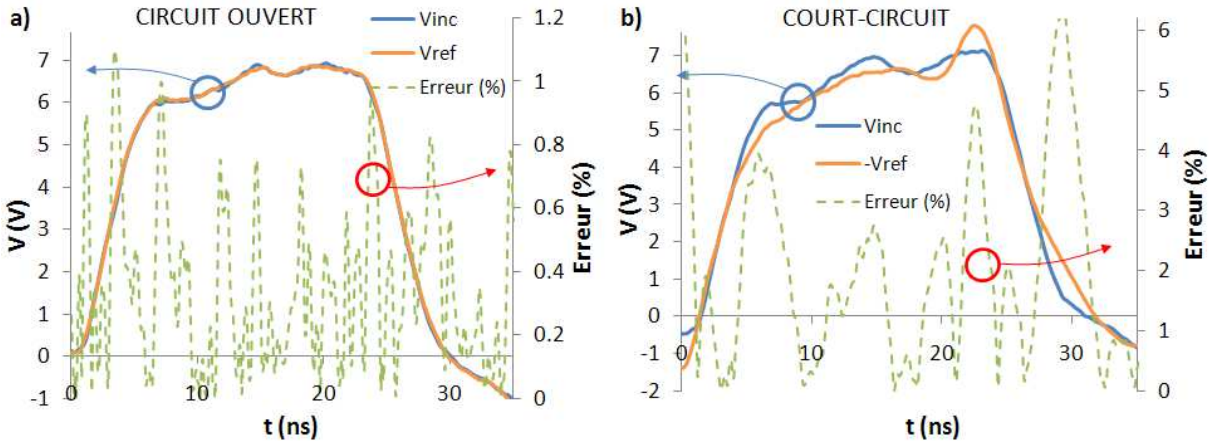

FIGURE 4.10 – Tension incidentes et réfléchies calculées dans le plan du composant lors de mesures sur un circuit ouvert (a) et fermé (b) dans le plan 6. La tension de précharge est ici de 10 V TLP. Dans le cas du court-circuit, la tension réfléchie a été inversée afin de mieux visualiser sa superposition avec la tension incidente. Les erreurs correspondantes sont calculées au moyen des formules 4.50 et 4.51

La figure 4.11 montre les formes d'ondes en tension et en courant obtenues sur des composants $NPNB$ $Pad40$ V (section 1.6.1.3) de deux longueurs différentes : 75 μm et 250 μm. Ces composants sont des structures de tests isolées, sans que soit mis en parallèle un circuit à protéger. Cela permet de s'assurer que l'on mesure bien la protection et non la réponse du circuit. Le boîtier sur lequel sont montées les protections est de type DIL, et la mesure s'effectue donc dans le plan 5. L'agrandissement du composant réduit la surtension de 5 % environ. Cette différence peut être attribuée à une injection de jonction émetteur/base supérieure dans le composant le plus grand.

En réitérant cette mesure à plusieurs niveaux de tension, on trace les courbes « I-Vmax » évoquées dans les précédentes chapitres. Ces courbes sont comparées dans la figure 4.12 aux courbes quasi-statiques TLP mesurées sur les mêmes composants avec des durées d'impulsion de 100 ns. Cette différence de 5 % entre les surtensions du composant de 250 μm et celui de 75

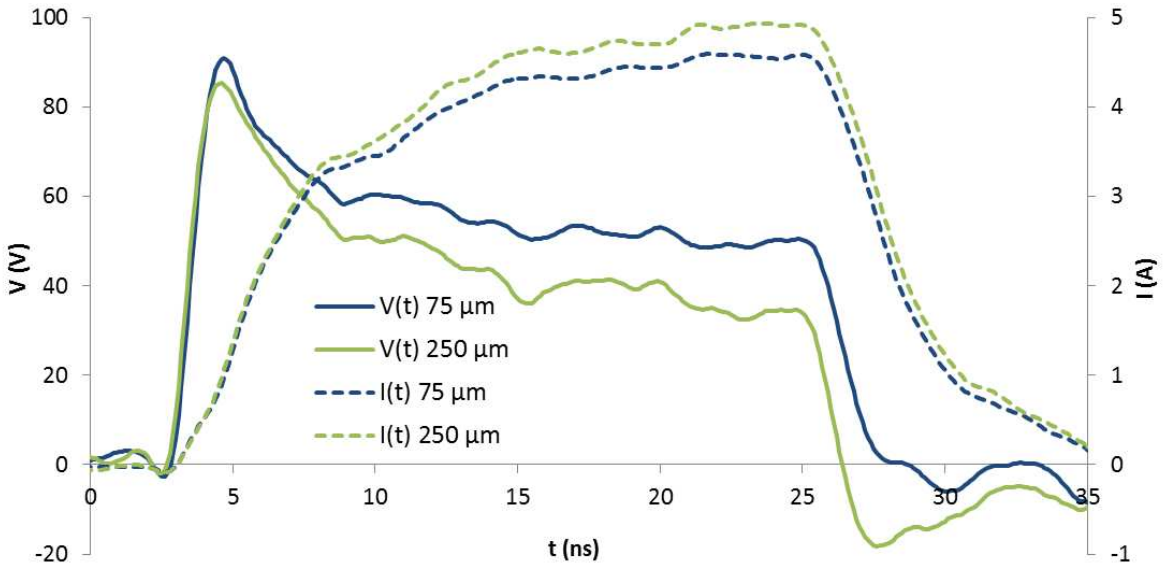

FIGURE 4.11 – Mesure de la tension (trait plein) et du courant (pointillés) sur des composants de protection $NPNB\ Pad20\ V$, de deux longueurs d'émetteur différentes, à 300 V TLP (6 A).

μm est visible sur l'ensemble de la courbe.

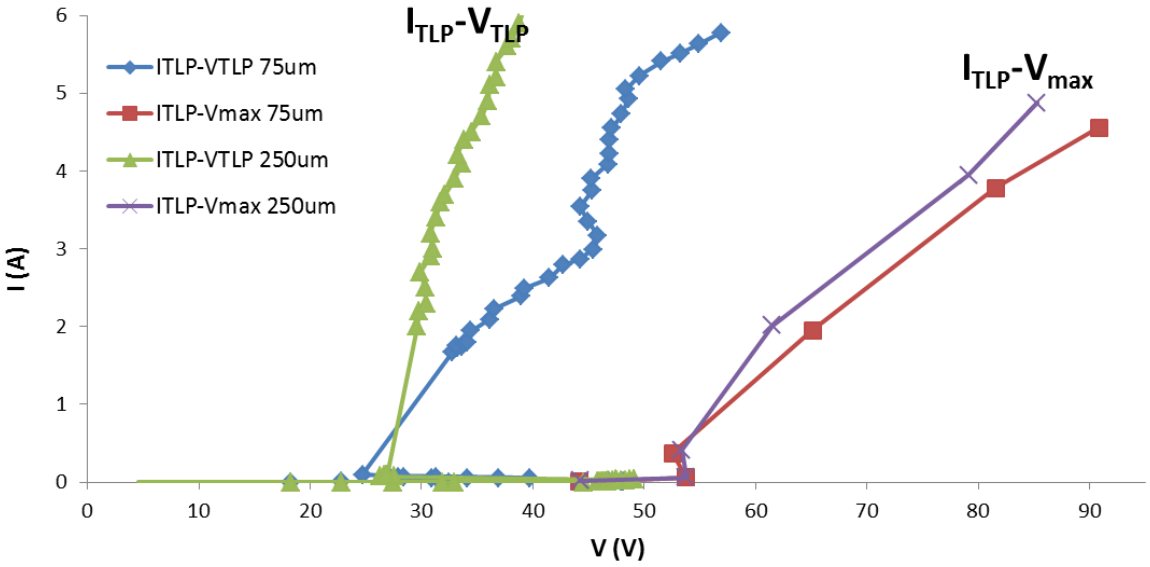

FIGURE 4.12 – Comparaison entre les courbes $I-V$ obtenues par des mesures TLP standard (courbes $I_{TLP}-V_{TLP}$) et les courbes $I-Vmax$ des $NPNB\ Pad40$ des deux dimensions différentes.

4.5 Conclusion.

Dans cette partie, nous avons présenté une méthode d'étalonnage du banc de mesure TLP du laboratoire de validation de Freescale. Le banc de mesure en lui-même n'a été que peu modi-

fié, seule une ligne de transmission (TL_1 de la figure 4.4) a été allongée pour obtenir un bonne séparation des impulsions incidentes et réfléchies à l'oscilloscope, et la sonde (haute impédance) en tension a été remplacée par un câble coaxial.

Le système génère des impulsions d'une largeur de 20 ns, avec un temps de montée de 3.3 ns environ de l'impulsion incidente sur le composant. Les tensions TLP pouvant être traitées vont de 1 V à 1500 V (6 A). L'étalonnage et la mesure peuvent s'effectuer à des niveaux de tension différents, les paramètres du système étant linéaires en puissance, mais une attention particulière a due être portée à l'effet des diodes en inverse présentes à l'extrémité de la ligne de charge pour absorber les réflexions négatives. En effet, ces diodes constituent le seul élément non-linéaire du système, et il convient de s'assurer que les éventuelles ondes réfléchies sur le relai atteignant ces diodes ne viennent pas perturber le signal utile.

L'erreur mesurée ne dépasse pas 5 %. On peut attribuer cette erreur à de multiples facteurs :

- La résolution temporelle de l'oscilloscope est de 100 ps, déterminée par sa fréquence d'échantillonnage de 10 GHz. Cependant sa bande passante analogique de 3 GHz ne permet pas de mesurer des temps de montée inférieurs à 120 ps.
- La résolution en tension de l'oscilloscope est déterminée par le codage sur 8 *bit* de sa voie verticale, ce qui donne une précision égale à $V_{TLP}/256$, soit 0.4 % pour une fenêtre en tension s'étalant sur une largeur égale à V_{TLP}. Mais l'incertitude sur le temps se répercute sur la tension de par le manque de précision dans l'alignement des impulsions incidentes et réfléchies. Dans notre cas, avec un temps de montée de 3.3 ns environ, un décalage de 120 ps entre les deux impulsions conduit à une erreur de 1.7 % de la mesure de la tension de circuit ouvert. On voit donc que pour chaque acquisition, les seules résolutions horizontales et verticales de l'oscilloscope apportent une erreur sur la tension de 2.1 % de V_{TLP}. En revanche, on note que le *jitter* de l'oscilloscope est très faible, de l'ordre de la picoseconde seulement, et ne vient donc pas perturber les mesures.
- Cette erreur est répétée sur toutes les acquisitions nécessaires à l'étalonnage du banc de mesure et aux étapes de mesure en elles-mêmes. L'accumulation des erreurs se répercute sur le résultat final.
- Les paramètres d'étalonnage sont le fruit d'opérations sur des formes d'ondes qui ont la plupart du temps été mesurées séparément et dont les impulsions incidentes ne sont pas exactement identiques. Cette non-répétitivité du générateur et du système (câbles, connecteurs ..) est également source d'erreur dans l'étalonnage.

Tous ces paramètres contribuent à limiter la précision des mesures.

Chapitre 5

Analyse du déclenchement d'un composant de protection ESD haute tension à base de thyristor : le $NPNB$

5.1 Introduction

L'étude menée dans les chapitres précédents nous a conduit à proposer et valider une nouvelle méthode de mesure de la réponse transitoire de protections ESD soumises à une décharge TLP. Son application sur des composants de protection a permis entre autres de confirmer la réalité physique du pic de tension observé en simulation $TCAD$ pendant le déclenchement de la protection.

C'est à ce pic de tension que nous allons nous intéresser dans ce chapitre. Nous verrons que sa forme dépend autant des caractéristiques du générateur que des paramètres physiques propres au composant. Après avoir exposé la relation entre les paramètres du générateur et le courant dans le composant quand la tension entame son repliement, nous étudierons la corrélation entre les mesures $transient - TLP$ et les simulations $TCAD$. Nous nous intéresserons ensuite au comportement transitoire du composant à protéger, dans le but d'affiner la fenêtre de conception ESD. Le déclenchement du $NPNB$ sera ensuite étudié en détails, et un ensemble de règles de dessin sera proposé pour réduire l'amplitude du pic en tension.

Note : pour des raisons de confidentialité, dans ce chapitre, les unités de distances seront données en « unités de mesure » (udm).

5.2 Conditions nécessaires au retournement

Dans cette section, nous allons établir une relation qui caractérise l'interaction de deux des paramètres du générateur d'impulsion, à savoir sa résistance interne et la pente de l'impulsion générée, avec la pente du courant I_{DUT} circulant dans le composant, qui caractérise sa réponse à l'impulsion. Nous verrons quelle relation doit être vérifiée par ces grandeurs pour que le composant puisse se replier.

Le banc de mesure TLP fonctionne comme rappelé en figure 5.1-a) : un générateur d'impulsions TLP délivre une onde incidente $V_{inc,géné}$. Le fonctionnement interne du générateur d'impulsions est évoqué plus en détail dans l'annexe B. Cette onde incidente se propage à travers la ligne de transmission d'impédance caractéristique $Z_C = 50\ \Omega$, et arrive dans le plan du composant après avoir subi l'atténuation et la distorsion propre au système de mesure. On rappelle que la tension et le courant du composant s'expriment par :

$$\begin{cases} V_{DUT} = V_{inc,DUT} + V_{ref,DUT} \\ I_{DUT} = \frac{V_{inc,DUT} - V_{ref,DUT}}{Z_C} \end{cases}, \tag{5.1}$$

avec $V_{inc,DUT}$ et $V_{ref,DUT}$ les ondes en tension incidentes et réfléchies dans le plan du composant. De ces équations, on tire :

$$V_{DUT} = 2V_{inc,DUT} - (V_{inc,DUT} - V_{ref,DUT}), \tag{5.2}$$

$$\boxed{V_{DUT} = 2V_{inc,DUT} - Z_C.I_{DUT}} \tag{5.3}$$

On peut alors représenter le système de mesure par un circuit électrique équivalent (figure 5.1-b), constitué de l'impédance caractéristique réelle Z_C de la ligne de transmission et d'un générateur de tension générant une impulsion TLP d'amplitude $2.V_{inc,DUT}(t)$. Ce circuit équivalent est celui utilisé pour les simulations électrothermiques $TCAD$. On voit qu'il n'est pas utile de simuler la ligne de transmission pour pouvoir reproduire avec exactitude les conditions de mesures TLP. Il est cependant nécessaire de connaître la forme d'onde de l'impulsion incidente $V_{inc,DUT}$ dans le plan du composant. Cette impulsion peut être mesurée en TLP en reliant l'extrémité de la ligne de transmission à l'oscilloscope.

Examinons à présent, en se basant sur le modèle de la figure 5.1-b) les conditions nécessaires pour que la tension aux bornes du composant puisse commencer à décroître. Le repliement a lieu quand la tension atteint son maximum. Cette condition se traduit mathématiquement par l'équation :

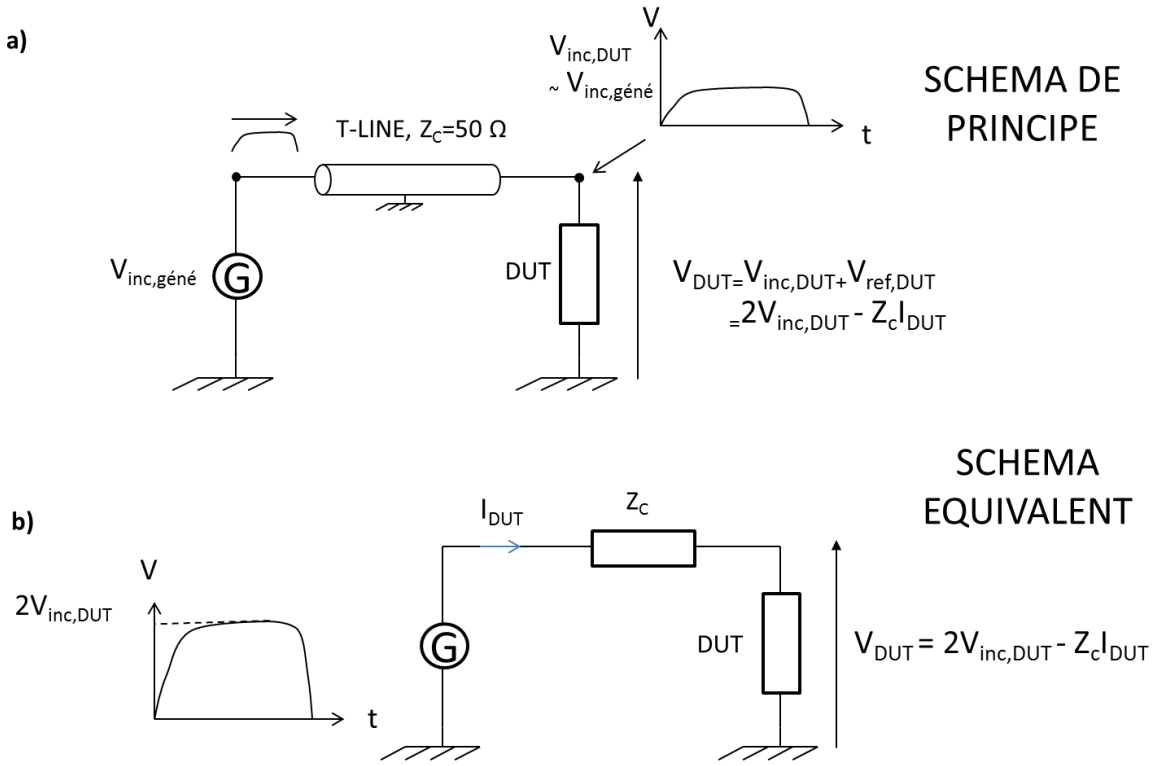

FIGURE 5.1 – Équivalence entre le modèle du banc de caractérisation TLP (a) et le modèle utilisé en simulation (b).

$$\frac{dV_{DUT}}{dt} = 0. \tag{5.4}$$

D'après l'équation 5.3, cela équivaut à :

$$\frac{d(2.V_{inc,DUT})}{dt} - \frac{d(Z_C.I_{DUT})}{dt} = 0, \tag{5.5}$$

soit :

$$\boxed{\frac{d(Z_C.I_{DUT})}{dt} = \frac{d(2.V_{inc,DUT})}{dt}}. \tag{5.6}$$

Cette condition signifie que le repliement a lieu lorsque l'augmentation de tension imposée par le générateur est compensée par l'augmentation de la chute de potentiel à travers la résistance Z_C (figure 5.1-b). Deux constatations peuvent être tirées de ce résultat :

– La première est que la forme d'onde de l'impulsion incidente et en particulier sa pente instantanée $\frac{dV_{inc,DUT}}{dt}$ a un impact direct sur la surtension de la protection au moment de son déclenchement.
– La deuxième constatation concerne l'effet de la résistance série Z_C. On voit qu'à pente

incidente égale, plus cette résistance est importante et plus le retournement sera atteint tôt, et donc moins la surtension sera élevée.

Une conséquence directe de ces deux constatations est qu'il est particulièrement difficile de comparer des pics de tension obtenus en $transient - TLP$ et en HBM. En effet, même si leurs temps de montée sont similaires, la différence importante entre les impédances des générateurs TLP et HBM (50 Ω en TLP contre 1.5 $k\Omega$ en HBM) modifie les conditions pour lesquelles l'équation 5.6 est atteinte dans les deux cas. Il en résulte différentes valeurs de surtensions.

L'utilité du $transient - TLP$ se justifie cependant en tant que moyen d'analyse du comportement transitoire des protections ESD. En effet, d'une part, les phénomènes qui contrôlent le retournement d'un composant de protection ESD sont les mêmes pour une décharge HBM et en TLP. D'autre part, plus l'impédance de générateur est élevée, et plus celui-ci a tendance à « forcer » le courant dans le composant. A contrario, la faible impédance de générateur du $transient - TLP$ permet de laisser les variations de courant au déclenchement s'opérer plus « librement », ce qui facilite l'analyse physique.

5.3 Corrélation entre les mesures $transient - TLP$ et les simulations $TCAD$

Pour pouvoir analyser le comportement physique des composants lors de la surtension, on a recours aux outils de simulation $TCAD$ de Synopsis (*sprocess* et *sdevice*). Le profil $2D$ du composant est d'abord générée par simulation du procédé technologique. Puis, une attention particulière a été portée lors de la sélection de certains modèles physiques. En particulier pour les simulations électrothermiques, les durées de vie des électrons et des trous ont été calibrées pour simuler avec précision la tension de maintien à bas courant, comme recommandé dans [64].

Examinons la corrélation que l'on peut établir entre les mesures $transient - TLP$ et les simulations $TCAD$ $2D$. Les mesures et simulations sont faites sur le composant $NPNB$ Pad 20 V ($V_{t1} = 28.9$ V), dont le déclenchement sera étudié en détails dans les sections suivantes. On effectue d'abord les mesures $transient - TLP$ pour une charge de 100 V, avec des temps de montée de 300 ps, 1 ns, et 10 ns, et des largeurs d'impulsions de 5 ns et 100 ns. De ces mesures, on extrait les formes d'onde incidentes $V_{inc,DUT}$ dans le plan du composant. Ces formes d'onde sont ensuite utilisées pour reproduire en simulation les conditions de mesure $transient - TLP$, conformément au schéma de la figure 5.1-b.

Les corrélations sont très bonnes durant les phases de déclenchement, et ce jusqu'au pic de

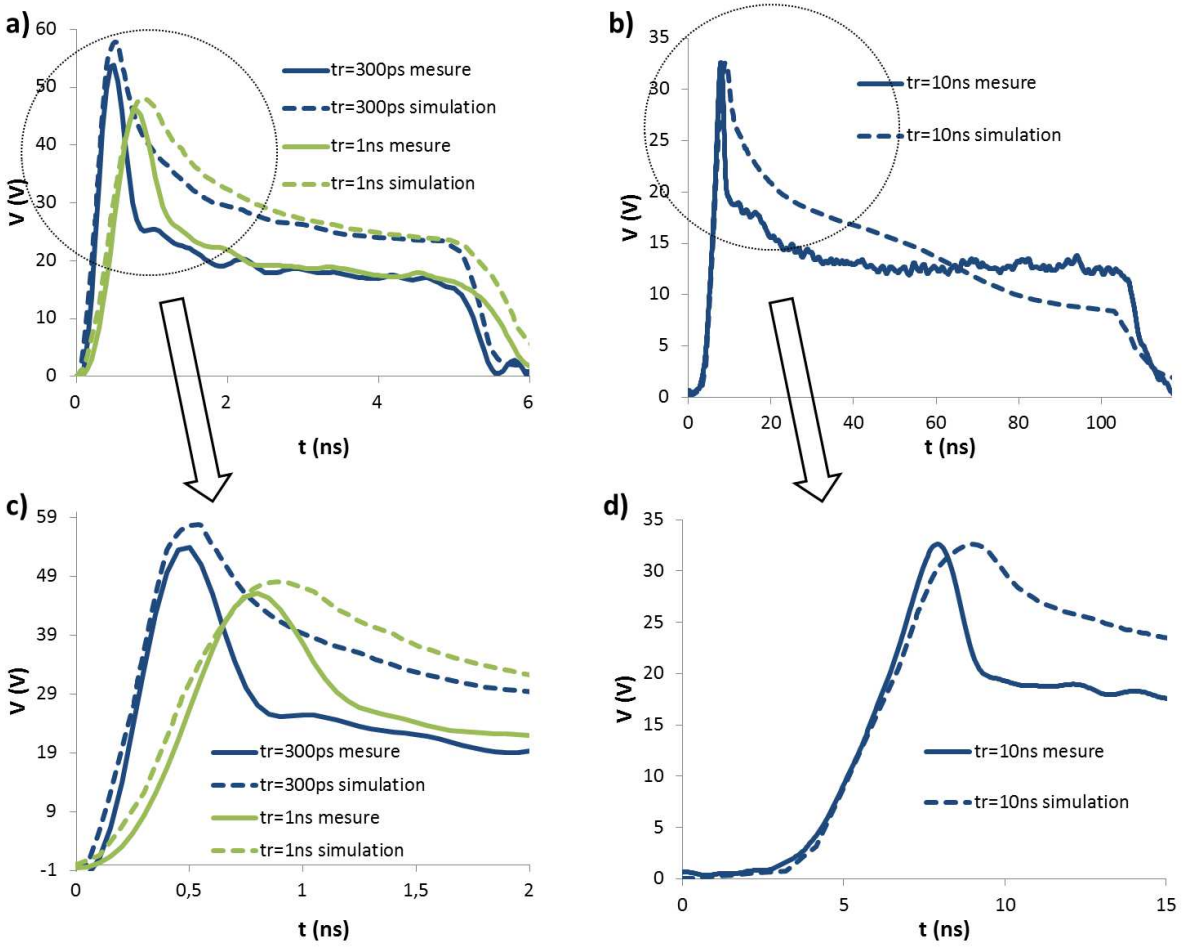

FIGURE 5.2 – Comparaison entre les mesures $transient - TLP$ (trait plein) et les simulations $TCAD$ (tirets). Figure a) : mesure et simulations sur le composant pour une impulsion de 5 ns de longueur, avec 300 ps de temps de montée (bleu foncé) et 1 ns (trait vert clair). Figure b) : mesures et simulations sur une impulsion de 100 ns de long avec 10 ns de temps de montée. Figures c) et d) : agrandissements des figures a) et b) autour du pic de tension.

tension. La différence entre les tensions maximales V_{max} simulées et mesurées est égale à 7 %, 2,3 % et 1 % respectivement pour les temps de montée de 300 ps, 1 ns, et 10 ns.

La pente dV_{DUT}/dt est elle aussi assez bien simulée. Mais à la différence de V_{max}, les écarts entre les valeurs mesurées et simulées s'accroît quand le temps de montée s'allonge. Elles sont respectivement de 5 %, 9 % et 20 % pour les temps de montée de 300 ps, 1 ns et 10 ns.

La corrélation se dégrade quand le composant entame son repliement, la tension ne chutant pas aussi vite en simulation qu'en mesure (jusqu'à 30 % d'erreur). Cette divergence peut être attribuable à des effets tridimensionnels non simulés, qui prennent de plus en plus d'importance à mesure que le composant entre en régime de forte injection. Cependant, tant que l'on se concentre sur les phénomènes physiques contrôlant le pic de tension, on peut raisonnablement

faire confiance aux simulations.

5.4 Analyse du déclenchement du composant à protéger

5.4.1 Introduction

La structure de protection dont nous étudierons le déclenchement dans la partie 5.7 est conçue pour protéger les E/S 20 V dans la technologie $SmartMos$ 8 MV (section 1.6). Le composant qu'il est amené à protéger de façon privilégiée est un $LDMOS$ de puissance (figure 5.3-a)), avec source et body P court-circuités (NBL isolant le substrat).

C'est sur ce $LDMOS$ que nous allons porter notre attention dans cette section, et plus particulièrement sur son comportement dynamique lors d'un stress positif sur le drain. Le but est de connaître sa tolérance à des agressions de type « pic de tension » afin de pouvoir éventuellement re-définir la fenêtre de conception ESD, en y incluant les données dynamiques. Une étude similaire a été menée en 2009 par Liou et. al., introduisant la notion de « zone de fonctionnement sûr transitoire » (Transient Safe Operating Area), qui étend la notion classique de zone de fonctionnement sûr (Safe Operating Area) aux phénomènes dynamiques [65].

5.4.2 Présentation du composant à protéger

Le $LDMOS$ à protéger n'est pas conçu pour fonctionner en avalanche, et la protection a pour fonction d'empêcher la tension de drain de dépasser la tension de claquage drain-body. La configuration dans laquelle sont placées la protection « NPNB » et le composant à protéger est schématisé dans la figure 5.3-b). L'impédance du circuit de pilotage de la grille est modélisée par une résistance pouvant varier de 0 Ω à 10 $k\Omega$.

La caractéristique TLP du $LDMOS$ avec 10 ns de temps de montée est reportée sur la figure 5.4. Dans ce cas précis, le contact de grille a été court-circuité à la source. On note que le $LDMOS$ est endommagé dès qu'il se replie, ce qui confirme la nécessité pour la protection de limiter la tension en-dessous du V_{t1} du $NLDMOS$.

Tension de maintien et robustesse

La cause de cette faible robustesse du $LDMOS$ lorsqu'il fonctionne en avalanche drain-body est à chercher du côté des caractéristiques de la zone N formant le drain. En effet, le repliement est initié par le déclenchement du NPN parasite constitué par la source, le body P et le drain,

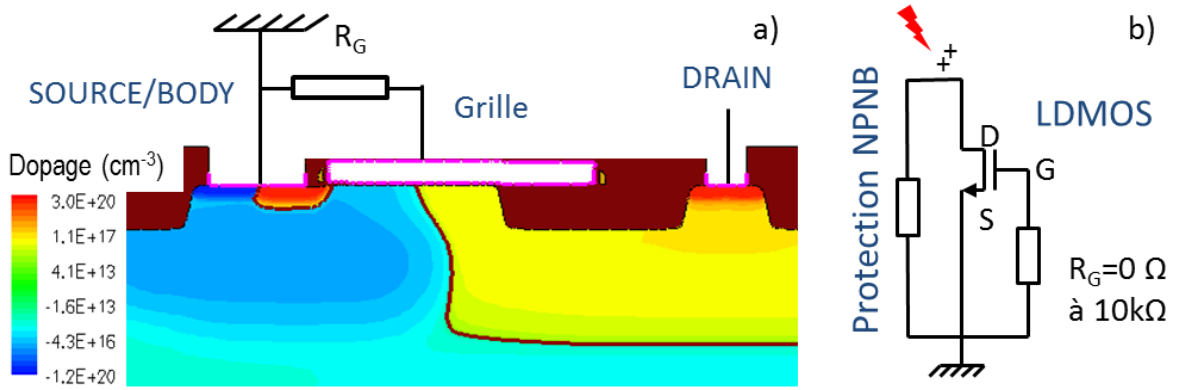

FIGURE 5.3 – Profil $2D$ du $LDMOS$ à protéger (a), et configuration dans laquelle sont placées la protection $NPNB$ et le $LDMOS$ (b).

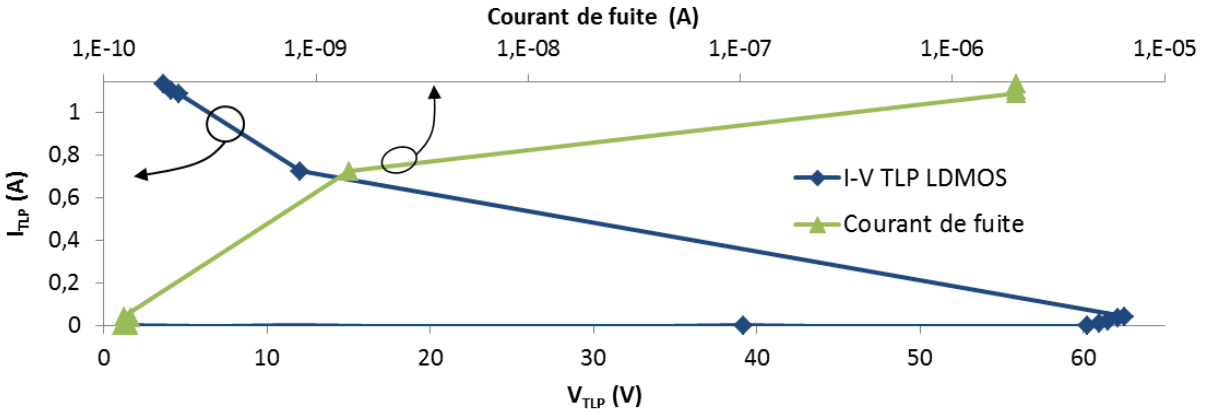

FIGURE 5.4 – Caractéristique $I-V$ TLP du $LDMOS$ à protéger, avec une résistance grille-source nulle et un temps de montée de 10 ns. En bleu foncé sont représentés les points $I-V$, et en vert clair le courant de fuite entre chaque mesure.

la source jouant alors le rôle de l'émetteur, le body de la base, et le drain de collecteur. Le faible dopage N du caisson de drain le rend particulièrement sensible aux effets de forte injection : lorsque le NPN parasite se déclenche, les fortes densités de courant circulant dans le drain tendent à repousser la jonction collecteur-base depuis la jonction métallurgique vers la zone N^+/N^- sous le contact de drain [29] (figure 5.5-b). Cette jonction N^+/N^- possède une tension de claquage très basse, ce qui explique la faible tension de maintien du composant, comme constaté par Mergens (« extraordinary strong snapback » [66]). La densité de courant sous la diffusion N^+ de drain atteint alors sa valeur critique $J_{crit} = qN_D v_{sat} \approx 1.10^5\ A.cm^{-2}$ (figure 5.5-a)) au-delà de laquelle le niveau de dopage n'est plus suffisant pour supporter le courant, ce qui mène à un second claquage par avalanche. Ce claquage par avalanche entraîne un repliement à fort courant conduisant à la destruction du $LDMOS$.

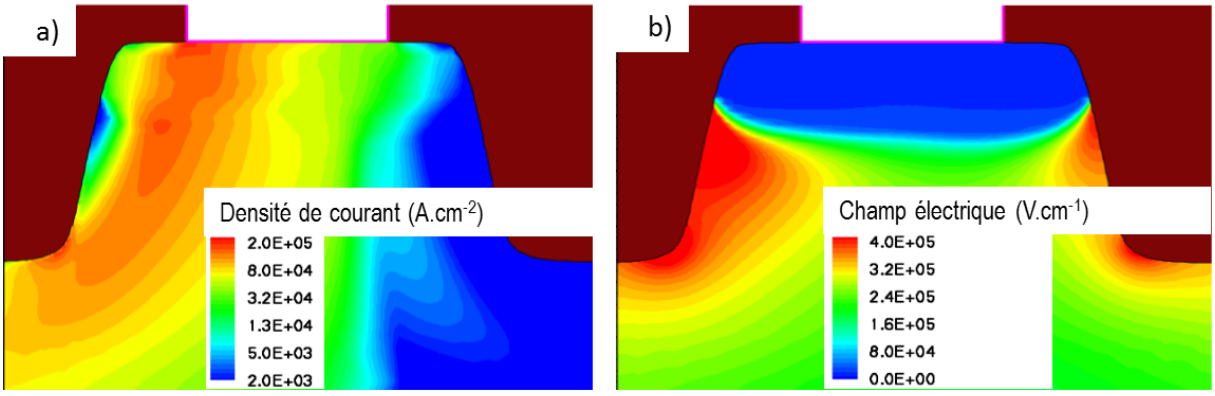

FIGURE 5.5 – Densité de courant (a) et champ électrique (b) simulés dans la région de drain après repliement de la tension. On note l'apparition d'une zone de fort champ électrique à la jonction N^+/N^- sous le contact de drain.

Tension de repliement : déclenchement par « dV/dt »

Ce composant est par ailleurs assez sensible aux phénomènes de « déclenchement par dV/dt » abordés dans la section 1.5.2.1, qui se manifestent par un abaissement de la tension de repliement V_{t1} sous l'effet des courants transitoires. Ces courants transitoires sont d'autant plus importants que la pente dV/dt est forte et que la résistance de grille R_G est élevée. On peut constater, dans le tableau 5.1, l'effet de la pente imposée et de la résistance de grille sur la tension de repliement quasi-statique du composant.

		Pente dV/dt	
		6 V/ns	200 V/ns
R_G	0 Ω	62.5	53,7
	1 $k\Omega$	62.5	52.7
	10 $k\Omega$	62.3	51.1

TABLEAU 5.1 – Tensions de repliement quai-statique du *LDMOS* pour deux pentes $dV_{Générateur}/dt$: 6 V/ns et 200 V/ns, correspondant respectivement à des temps de montée de 10 ns et 300 ps environ, et pour trois valeur de la résistance de grille : 0 Ω, 1 $k\Omega$ et 10 $k\Omega$.

5.4.3 Redéfinition de la fenêtre de conception

Nous avons vu au-dessus que le *LDMOS* était détruit dès son repliement, à la tension $V_{t1,LDMOS}$. Par conséquent il faut veiller, lors de la conception de la protection, à ce qu'elle se replie avant d'atteindre la tension $V_{t1,LDMOS}$. Or, on voit d'après le tableau 5.1 que cette tension de repliement du *LDMOS* dépend de la pente dV/dt appliquée et de la résistance grille-source R_G. La fenêtre de conception des protections doit donc être adaptée de manière à intégrer ces particularités. A première vue, on peut définir la borne supérieure de la fenêtre de conception de la manière suivante : il faut s'assurer que quelque soit la pente de l'agression *ESD*, la forme

d'onde générée par la protection n'atteigne jamais la plus basse des tensions de repliement du $LDMOS$ (tableau 5.1).

Cette définition est très contraignante, puisqu'elle concerne l'intégralité de la forme d'onde $V_{NPNB}(t)$ aux bornes de la protection, y compris le pic de tension. Or on ne sait pas a priori si le $LDMOS$ est sensible à des surtensions brèves telles que celles constatées aux bornes du $NPNB$. Examinons cette question, et tentons de voir s'il est possible d'assouplir la règle édictée au précédent paragraphe.

Quelle est la durée associée à la destruction du $LDMOS$?

Le composant va-t-il se dégrader si sa tension de repliement est dépassée pendant quelques nanosecondes ? Autrement dit, la surtension est-elle suffisamment longue pour que le $LDMOS$ ait le temps de se déclencher et de se dégrader ?

Pour le savoir, nous avons fait subir une série d'agressions $vf-TLP$ à des composants dont la résistance de grille a été choisie à $R_G = 10\ k\Omega$, afin qu'ils soient particulièrement vulnérables aux effets dynamiques. Le temps de montée est de 300 ps, et la tension est celle correspondant au premier point $I-V\ TLP$ après repliement pour ce temps de montée particulier et cette résistance de grille. Cela correspond à une tension d'impulsion de 55 $V\ TLP$, pour laquelle la pente est d'environ 200 V/ns.

Nous avons utilisé les possibilités du banc de mesure Oryx permettant de générer des impulsions de 1.25 ns, 2.5 ns et 5 ns, à l'aide des différentes lignes de charge. La largeur d'impulsion la plus courte, à 1.25 ns, est de l'ordre de grandeur de la largeur du pic de surtension du $NPNB$.

Les formes d'ondes incidentes sont relevées en figure 5.6. On peut voir que les impulsions de 1.25 et 2.5 ns ne s'annulent pas immédiatement après la fin de l'impulsion, mais retombent à zéro par paliers successifs. En effet, contrairement à la ligne de 5 ns, elles ne sont pas conçues comme la ligne TL_1 de la figure B.1 (annexe B). Dans le cas des impulsions de 1.25 ns et 2.5 ns, la ligne TL_1 est scindée en deux parties reliées entre elles par une résistance de 50 Ω. Ainsi, les ondes se réfléchissent sur deux résistances successives, avec des coefficients de réflexion de 1 et $\frac{1}{2}$. Cette réflexion multiple explique la présence de cette décroissance par paliers, dont on peut apercevoir les deux premiers dans la figure 5.6.

Les trois mesures ont bien évidemment été faites sur des composants distincts, et la caractéristique $I-V$ statique mesurée après l'agression dans chacun des trois cas a été comparée (figure 5.7). On constate que quelle que soit la durée de l'impulsion, la caractéristique statique du composant n'est plus la même après l'agression ESD, ce qui dénote une dégradation. De

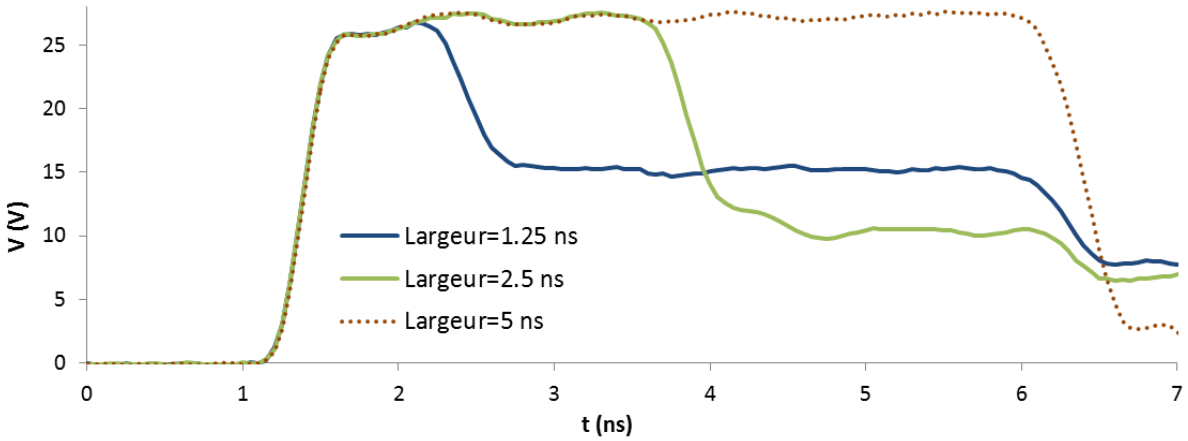

FIGURE 5.6 – Formes d'ondes utilisées pour tester la robustesse du *LDMOS* sous différentes largeurs d'impulsion. Pour les impulsions à 1.25 et 2.5 *ns*, on peut voir les paliers correspondant aux réflexions multiples entre la résistance R_1 (figure B.1) et la résistance de 50 Ω. En bleu foncé est tracée l'impulsion incidente à 1.25 *ns*, en vert clair celle à 2.5 *ns* et en rouge pointillés celle à 5 *ns*.

plus, le défaut créé semble d'autant plus important que l'impulsion est longue. Ce défaut, sans doute situé dans la région de drain, apparaît en moins de 1.25 *ns*, puis croît à mesure que le composant s'échauffe.

On voit donc que la durée nécessaire au *LDMOS* pour se replier et être endommagé est de l'ordre de grandeur de la largeur du pic de tension présent lors du déclenchement d'un *NPNB*. On ne peut donc pas compter sur la brièveté du pic pour ignorer son effet sur le *LDMOS*.

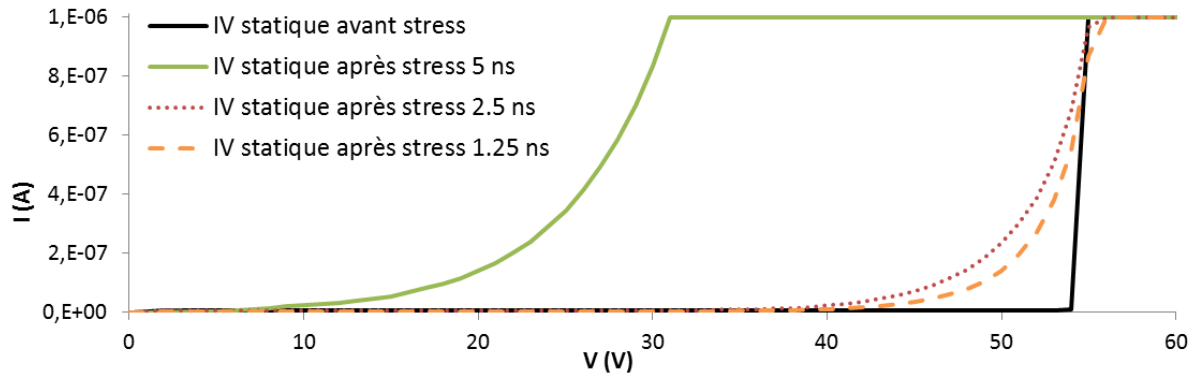

FIGURE 5.7 – Courbes $I-V$ statiques mesurées avant l'agression *ESD* (en noir), après l'impulsion de 5 *ns* (en trait plein vert clair), après celle de 2.5 *ns* (rouge pointillés) et après l'impulsion de 1.25 *ns* (tirets oranges). Une limitation de 1 μA a été appliquée lors de la mesure.

De plus, on verra dans la section 5.7 que le courant circulant dans le *NPNB* avant son déclenchement est très faible et que par conséquent, la tension du générateur se reporte quasi-intégralement sur le composant. Ainsi, on peut considérer comme égales les pentes dV_{NPNB}/dt

et $dV_{Générateur}/dt$ avant repliement du $NPNB$. On ne peut donc pas compter non plus sur une réduction de la pente appliquée sur le $LDMOS$ par l'action du $NPNB$, ni sur une augmentation de sa tension de repliement du fait de l'abaissement de cette pente dV/dt (voir tableau 5.1).

En conclusion, la règle énoncée plus haut ne peut pas être assouplie. Il faut que le pic de tension du $NPNB$ soit toujours inférieur à la plus faible des tensions de repliement du $LDMOS$ sous différentes pentes et résistances de grille.

5.5 Présentation du composant $NPNB$

5.5.1 Principe de fonctionnement

Le composant $NPNB$ (voir section 1.6.1.3) est développé à partir de transistors bipolaires (figure 5.8). Il est destiné à la protection contre les décharges tant positive que négatives . Néanmoins, dans cette étude, nous nous limiterons au cas des décharges positives.

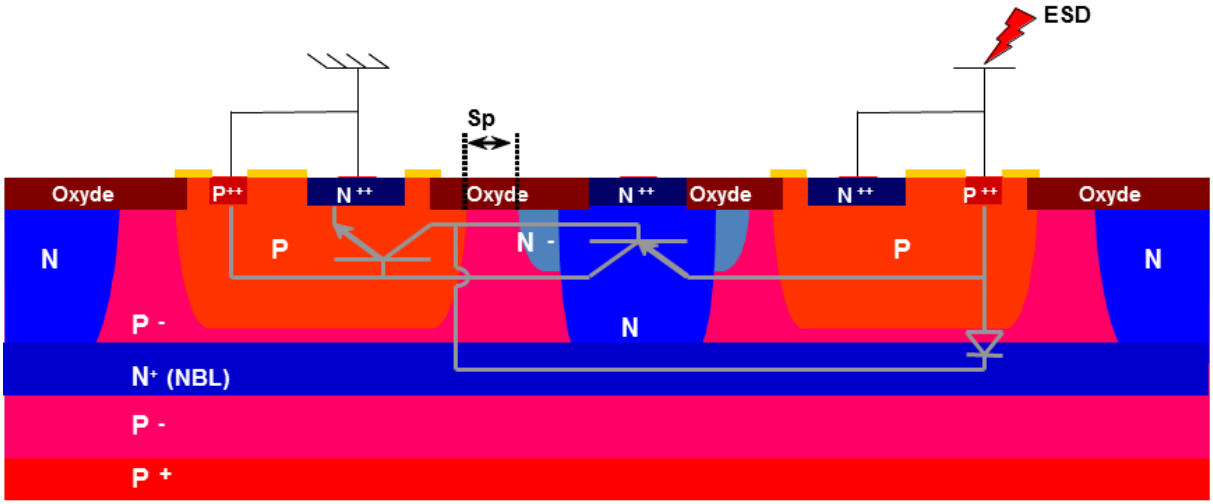

FIGURE 5.8 – Profil du composant $NPNB$ et schéma équivalent.

Il est composé de deux caissons P isolés l'un de l'autre par un caisson N profond relié à une couche enterrée N^+ (NBL). Des diffusions peu profondes et très dopées de N^+ et P^+ sont réalisées dans chacun des caissons P. Tout le composant est entouré par un caisson N d'isolation.

Trois transistors bipolaires doivent être considérés dans le cas d'une décharge positive. Le premier est un NPN dont le caisson P relié à la masse constitue la base, la diffusion N^+ de ce même caisson l'émetteur, et le caisson N du milieu le collecteur. Ce NPN partage sa base et son émetteur avec un NPN vertical, dont le collecteur est constitué de la couche enterrée N^+ (NBL). Le troisième transistor est un PNP ayant le caisson P relié au signal comme émetteur, le caisson N du milieu comme base et le caisson P relié à la masse comme collecteur. Le NPN

latéral et le PNP sont connectés en configuration SCR, dans laquelle le courant de collecteur de l'un des deux transistors constitue le courant de base de l'autre. À ces trois transistors vient s'ajouter une diode verticale constituée par l'émetteur du PNP et la couche enterrée.

La protection est déclenchée par l'effet d'autopolarisation du NPN latéral. Quand la tension atteint la tension de claquage de la jonction base/collecteur en inverse, un courant de trous est injecté dans la base, qui tend à polariser en direct la jonction émetteur-base. L'émetteur injecte alors des électrons dans la base, et l'effet bipolaire se déclenche quand ceux-ci atteignent la jonction base-collecteur, où ils sont multipliés par effet d'avalanche. La jonction émetteur-base du NPN vertical est commune avec celle du NPN latéral, et le NPN vertical se déclenche à son tour quand les électrons injectés par l'émetteur atteignent sa jonction collecteur-base en inverse. Le courant injecté par ces deux transistors NPN augmente alors rapidement, et la tension aux bornes de la protection peut se retourner quand la condition de repliement de l'équation 5.6 est atteinte.

La tension de déclenchement quasi-statique V_{t1}, fortement liée à la tension de claquage, peut être ajustée à travers le paramètre Sp, distance entre les masques du caisson P et N.

Le déclenchement des NPN latéraux et verticaux tend à polariser le PNP latéral et induire un effet SCR. Cependant, le gain du PNP est trop faible à bas courant, sa base étant plus dopée que son émetteur, pour que l'effet SCR puisse être complet. Ce gain maintenu volontairement faible à bas courant permet d'éviter les risques de latch-up propres aux SCR, en conservant une tension de maintien (V_H) élevée. Il se déclenche cependant sous fortes densités de courant, permettant d'abaisser la résistance à l'état passant R_{ON} et d'améliorer la robustesse du $NPNB$.

5.5.2 Caractéristiques quasi-statiques et dynamiques

Des caractérisations TLP et $transient-TLP$ ont été menées sur un composant $NPNB$ de 75 μm de long en technologie $SmartMos$ $8MV$ (figure 5.9). La courbe TLP correspond à la caractéristique quasi-statique sous des impulsions de 100 ns de durée, avec un temps de montée de 10 ns. Les mesures de tensions maximales ont été effectuées avec 1 ns et 10 ns de temps de montée.

Les résultats montrent un V_{t1} de 28.9 V et un V_H de 12.5 V. Le composant est détruit pour un courant (I_{t2}) de 6.9 A, comme on peut le constater par l'augmentation du courant de fuite.

La tension de pic croît continument depuis V_{t1} jusqu'à 37 V à 6 A pour le temps de montée de 10 ns et 50 V pour 1 ns de temps de montée. Comparées au V_{t1}, ces tensions maximales correspondent à des surtensions de 8 V et 21 V, qui doivent être prises en compte dans la fenêtre

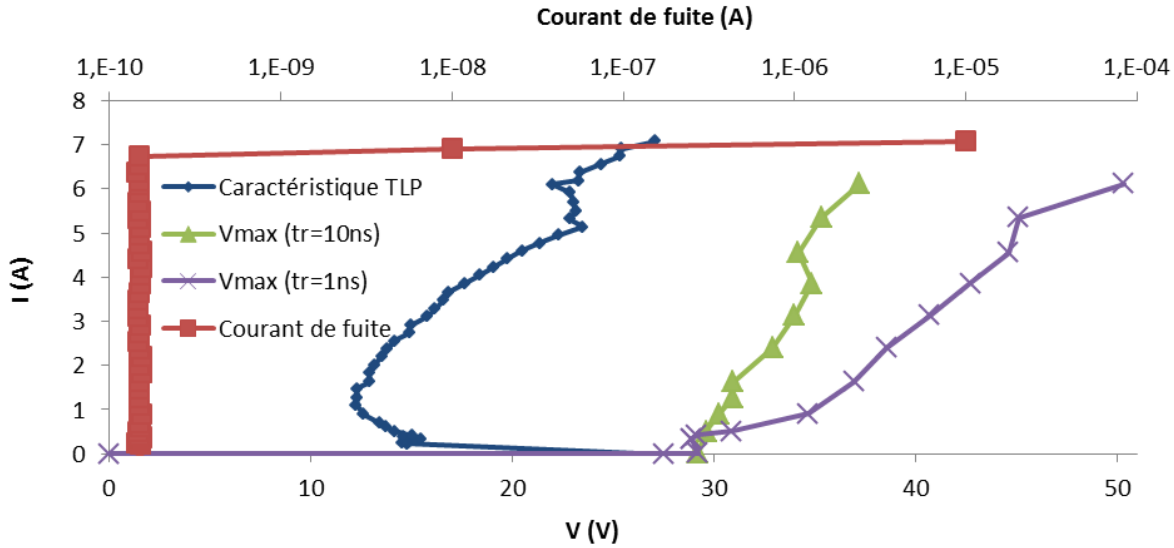

FIGURE 5.9 – Caractérisation TLP du composant (courbe bleu foncé) et $transient - TLP$ avec 10 ns et 1 ns de temps de montée (courbes vert pâle et violette). Le courant de fuite est représenté en rouge, permettant d'identifier le courant I_{t2} à 6.9 A.

de conception.

5.6 Dynamique du claquage par avalanche des diodes base-collecteur en inverse : comparaison PIN vs. P^+/N^+

5.6.1 Introduction

Quand une jonction $P - N$ est polarisée en inverse, un phénomène d'ionisation par impact peut avoir lieu. Il survient lorsque les porteurs libres dans la zone de charge d'espace acquièrent une énergie cinétique suffisante, sous l'action du champ électrique, pour extraire des électrons depuis la bande de valence jusqu'à la bande de conduction. L'impact entre un porteur libre et un électron de valence donne lieu à la création d'une paire électron-trou. Chaque porteur ainsi créé est accéléré par le champ électrique, et peut générer à son tour d'autres paires électron-trou si la longueur de la ZCE est suffisante. Ce phénomène est appelé "avalanche".
Nous nous proposons dans cette partie 5.6 d'évaluer le temps associé à l'établissement du courant d'avalanche lorsqu'une tension constante égale à la tension de claquage BV est appliquée sur le composant.

Singularité des deux diodes étudiées

Nous avons décidé d'isoler les deux diodes qui constituent la jonction base-collecteur, afin d'étudier séparément leur dynamique de déclenchement.

En effet, les deux jonctions sont très différentes : la jonction latérale est une jonction abrupte P^+/N^+, constituée de deux zones de dopage équivalent, tandis que la jonction verticale peut être assimilée à une diode PIN. Elle est composée de deux zones P^+ et N^+ de mêmes dopages que dans la jonction latérale, mais espacées par une large zone de faible dopage P, se comportant comme un zone intrinsèque par rapport aux deux autres (figure 5.10). Ces deux diodes se distinguent également par le profil de champ électrique en inverse. La jonction latérale étant symétrique, le champ électrique se répartit de façon égale dans les zones N et P et se présente comme un pic localisé sur la jonction métallurgique. Pour la jonction verticale, du fait de la forte dissymétrie du dopage, la zone de champ électrique s'étend principalement du côté P^-, moins dopé, où il peut être considéré comme constant sur l'étendue de la zone.

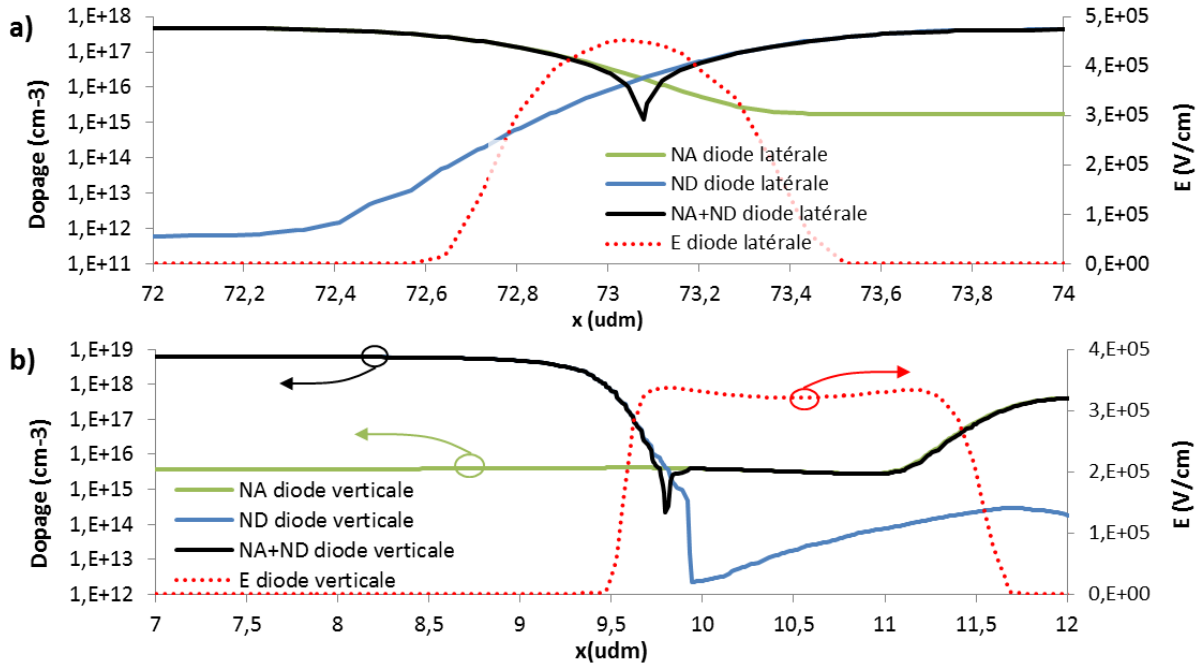

FIGURE 5.10 – Concentrations en atomes donneurs (bleu foncé), accepteurs (vert pâle), dopage total (noir) et profil de champ électrique à bas courant (rouge pointillés) pour la diode latérale (a) et la diode verticale (b) formant les deux jonctions base-collecteur du $NPNB$.

Phénomène d'avalanche : notion d'intégrale d'ionisation

Le phénomène d'ionisation par impact s'enclenche quand le champ électrique dans le silicium atteint localement sa valeur critique ($E_{crit} = 3.10^5\ V.cm^{-1}$), pour laquelle l'énergie des porteurs leur permet d'ioniser les atomes du réseau cristallin. Cependant, il ne suffit pas que le champ critique soit atteint pour déclencher le phénomène d'avalanche, il est également nécessaire que les porteurs accélérés par le champ électrique le soient sur une distance suffisante pour

avoir une chance de rentrer en collision avec un autre porteur, et ainsi générer une autre paire électron-trou. Cette notion est représentée par les paramètres α_e et α_h, les coefficients d'ionisation associés respectivement aux électrons et aux trous. Ils représentent le nombre moyen de paires électron-trou crées par unité de distance par un électron (respectivement un trou) traversant la zone de charge d'espace. Ils peuvent également être vus comme l'inverse de la distance moyenne parcourue par le porteur avant de générer une paire électron-trou. La condition d'avalanche peut être exprimée par la relation suivante, avec L la largeur de la zone de charge d'espace :

$$\Im_e = 1 \; ou \; \Im_h = 1, \tag{5.7}$$

où

$$\Im_e = \int_0^L \alpha_e . exp\left(\int_L^x (\alpha_e - \alpha_h).dx'\right) dx \tag{5.8}$$

et

$$\Im_h = \int_0^L \alpha_h . exp\left(\int_L^x (\alpha_e - \alpha_h).dx'\right) dx \tag{5.9}$$

sont les intégrales d'ionisation des électrons et des trous.

En régime permanent, on définit un facteur de multiplication par avalanche de la façon suivante :

$$M_e = \frac{1}{1 - \Im_e} \tag{5.10}$$

$$M_h = \frac{1}{1 - \Im_h} \tag{5.11}$$

M_e est le facteur par lequel est multiplié le courant d'électrons lors de sa traversée de la zone d'avalanche, et M_h le facteur par lequel est multiplié le courant de trous.

5.6.2 Simulation de la dynamique du claquage par avalanche des deux diodes

Nous nous sommes limités dans un premier temps à une situation simplifiée dans laquelle la tension appliquée est constante dans le temps, et le courant suffisamment faible pour que l'on puisse s'affranchir des effets de forte injection, qui sinon moduleraient l'extension de la ZCE. Ainsi, on peut considérer le profil de champ électrique et la largeur de la ZCE comme étant constants dans le temps. Pour simplifier les conditions aux limites, nous avons également réduit à une quantité négligeable les courants de déplacement issus de l'application d'une rampe de tension dV/dt aux bornes de la capacité équivalente de jonction base-collecteur avant claquage. Pour obtenir une telle situation en simulation, nous avons appliqué aux deux diodes des rampes de tension très lentes de 0 V jusqu'à la tension de claquage BV, (avec 100 ns de temps de

montée), suivies d'un plateau à $V = BV$ (figure 5.11), comme réalisé dans [67].

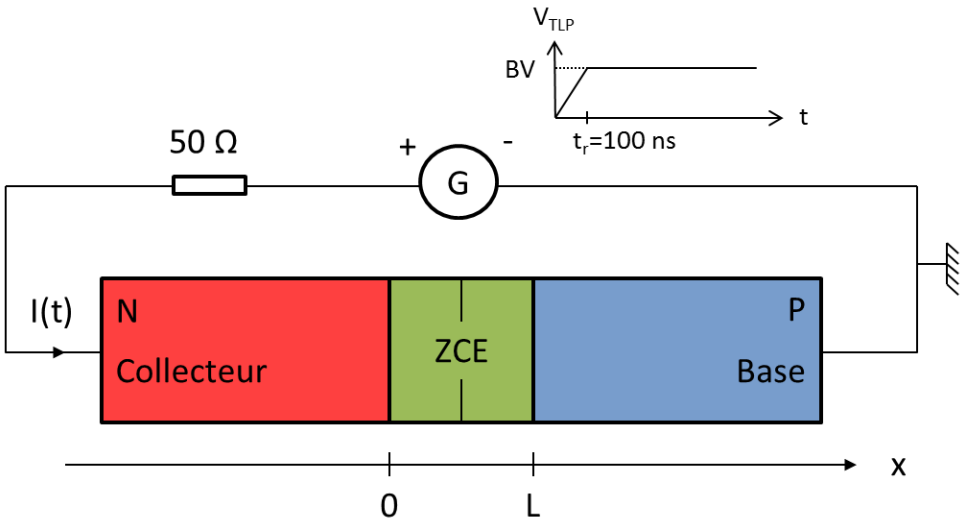

FIGURE 5.11 – Conditions de simulation employées pour étudier la dynamique du processus de claquage par avalanche. Une rampe de tension est appliquée à travers une charge de 50 Ω avec un temps de montée de 100 ns jusqu'à la tension de claquage, puis la tension est maintenue constante.

Les diodes latérales et verticales ont des tensions de claquage respectives BV_{lat} et BV_{vert} de 24.4 V et 62.8 V. Cette différence de tension est due à la différence des largeurs de la zone intrinsèque. Avec $E = -\frac{dV}{dx}$, plus la zone intrinsèque est large, plus la différence de potentiel doit être élevée pour atteindre le champ critique $E_{crit} = 3.10^5\ V/cm$.

Les profils de champ électrique à la tension de claquage sont reportés en figure 5.10. On peut constater que si la diode verticale se déclenche à $E = E_{crit}$, ce n'est pas le cas de la diode latérale, pour laquelle le champ électrique atteint 4.5 $10^5\ V/cm$. Cela est dû à la faible largeur de la zone de charge d'espace de la jonction latérale : la probabilité pour un porteur libre d'ioniser un atome du réseau pendant son passage à travers la ZCE est trop faible à $E = E_{crit}$, rendant nécessaire une augmentation du champ électrique pour que la condition d'ionisation 5.7 soit respectée. La figure 5.12 montre les coefficients d'ionisation des porteurs le long de la ZCE pour les deux jonctions étudiées. Les coefficients d'ionisation ont un profil similaire à celui du champ électrique : on verra dans la suite que leur évolution en fonction du champ électrique est en E^7. Les intégrales d'ionisations correspondantes sont respectivement égales à 0.7 et 1.0 pour les trous et les électrons dans la jonction latérale, contre 0.3 et 0.9 pour les trous et les électrons dans la jonction verticale.

La figure 5.13 montre une comparaison de l'évolution du courant d'avalanche dans les deux diodes, chacune à sa tension de claquage. Aux temps inférieurs au temps de montée (100 ns),

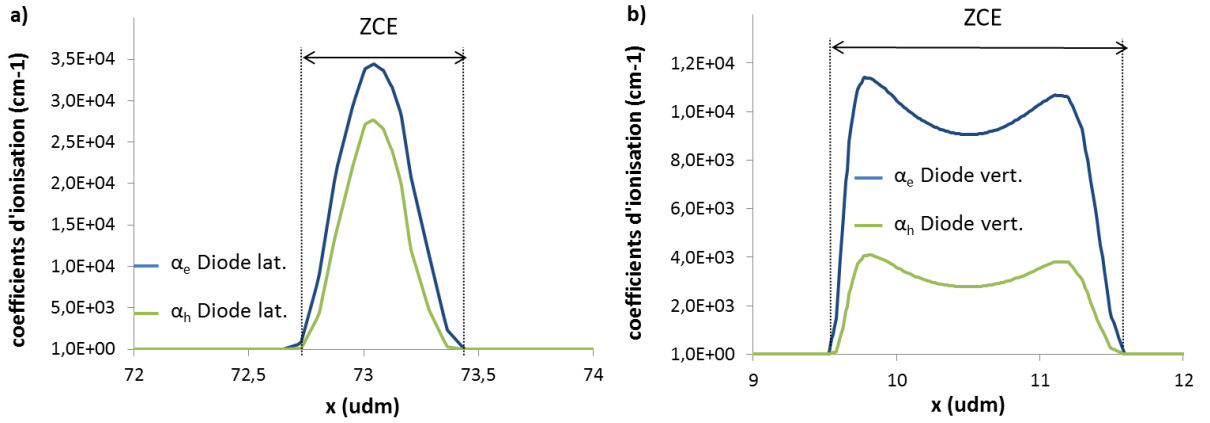

FIGURE 5.12 – Coefficients d'ionisation des électrons (bleu foncé) et des trous (vert clair) à travers la diode latérale (a) et verticale (b).

on note la présence de courants constants non nuls. Ces courants sont la somme des courants de déplacement $C\frac{dV}{dt}$ dûs à la charge de la capacité de transition des jonctions et des courants de saturation en inverse de la diode. Les courants de déplacement s'annulent quand la tension atteint le plateau, à $V = BV$. Une fois la tension de claquage atteinte, le courant généré suit clairement une loi exponentielle en $exp(\frac{t}{\tau})$, dont le paramètre τ varie de 80 ps à 140 ps entre la jonction latérale et la jonction verticale. Le temps d'établissement du régime permanent est, lui, de 700 ps et 2 ns respectivement les jonctions latérales et verticales.

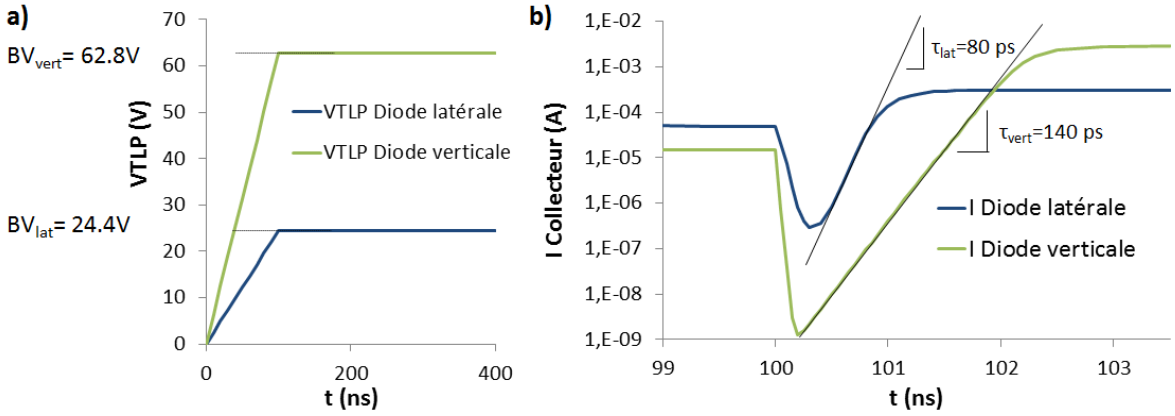

FIGURE 5.13 – Résultats de l'expérience décrite en figure 5.11 pour la diode latérale (bleu foncé) et verticale (vert clair). En a), les tensions appliquées, en b) les courants circulant dans chacune des diodes en échelle logarithmique. On constate une pente plus importante pour la jonction latérale que verticale, suivie d'un plateau correspondant à un régime permanent.

5.6.3 Expression analytique pour l'évolution du courant d'avalanche lors du claquage

Évaluons l'évolution de la densité de courant dans la zone de charge d'espace au moment du claquage par avalanche. L'évolution de la population de trous et d'électrons est régie par les équations de continuité, en une dimension :

$$\begin{cases} \frac{\partial n(x,t)}{\partial t} = \frac{1}{q}\frac{\partial J_n(x,t)}{\partial x} + G(x,t) \\ \frac{\partial p(x,t)}{\partial t} = -\frac{1}{q}\frac{\partial J_p(x,t)}{\partial x} + G(x,t) \end{cases} \tag{5.12}$$

avec $G(x,t)$ le taux de génération par avalanche.
Les électrons et les trous voyageant à leur vitesse de saturation respective v^e_{sat} et v^h_{sat}. G s'écrit :

$$G(x,t) = \alpha_e v^e_{sat} n(x,t) + \alpha_h v^h_{sat} p(x,t). \tag{5.13}$$

On note que l'on a négligé le taux de recombinaisons par rapport au taux de génération par avalanche. On néglige également les courants de diffusion dans la ZCE, et alors les densités de courant d'électrons et de trous s'écrivent :

$$J_n(x,t) = q v^e_{sat} n(x,t) \tag{5.14}$$

et

$$J_p(x,t) = q v^h_{sat} n(x,t). \tag{5.15}$$

On peut alors réécrire les équations 5.12 :

$$\begin{cases} \frac{1}{q v^e_{sat}}\frac{\partial J_n(x,t)}{\partial t} = \frac{1}{q}\frac{\partial J_n(x,t)}{\partial x} + \frac{\alpha_e}{q} J_n(x,t) + \frac{\alpha_h}{q} J_p(x,t) \\ \frac{1}{q v^h_{sat}}\frac{\partial J_p(x,t)}{\partial t} = -\frac{1}{q}\frac{\partial J_p(x,t)}{\partial x} + \frac{\alpha_e}{q} J_n(x,t) + \frac{\alpha_h}{q} J_p(x,t) \end{cases} \tag{5.16}$$

La somme des deux lignes de ce système donne, multipliées par q :

$$\frac{1}{v^e_{sat}}\frac{\partial J_n(x,t)}{\partial t} + \frac{1}{v^h_{sat}}\frac{\partial J_p(x,t)}{\partial t} = \frac{\partial (J_n(x,t) - J_p(x,t))}{\partial x} + 2\left[\alpha_e J_n(x,t) + \alpha_h J_p(x,t)\right] \tag{5.17}$$

Dans la suite, on fait l'approximation suivante : on considère les vitesses de saturation des électrons et des trous égales de même que leurs coefficients d'ionisation : $v^e_{sat} = v^h_{sat} = v_{sat}$ et $\alpha_e = \alpha_h = \alpha$.
On réécrit alors l'équation précédente, en multipliant par v_{sat} :

$$\frac{\partial J(x,t)}{\partial t} = 2 v_{sat} \alpha J(x,t) + v_{sat}\frac{\partial (J_n(x,t) - J_p(x,t))}{\partial x} \tag{5.18}$$

La densité totale de courant totale est supposée constante le long de la zone de charge

d'espace. Par intégration sur x entre 0 et L, on obtient :

$$\int_0^L \frac{\mathrm{d}J(t)}{\mathrm{d}t}dx = 2v_{sat}J(t)\int_0^L \alpha(x)dx + v_{sat}\left[J_n(x,t) - J_p(x,t)\right]_0^L \tag{5.19}$$

$$\frac{\mathrm{d}J(t)}{\mathrm{d}t}L = 2v_{sat}J(t)\int_0^L \alpha(x)dx + v_{sat}\left[J_n(L,t) - J_p(L,t) - J_n(0,t) + J_p(0,t)\right] \tag{5.20}$$

Évaluons ce dernier terme $J_n(L,t) - J_p(L,t) - J_n(0,t) + J_p(0,t)$. La densité de courant à $V = BV$ avant claquage par avalanche est égale à :

$$J(0) = -J_S(e^{\frac{-qBV}{kT}} - 1) + J_{GR}, \tag{5.21}$$

soit

$$J(0) = J_S, \tag{5.22}$$

car on peut négliger le terme exponentiel aux tensions considérées et le courant de génération-recombinaisons J_{GR}.

Ce courant J_S se décompose en un courant de trous et un courant d'électrons :

$$J_S = J_{nS} + J_{pS}. \tag{5.23}$$

Les électrons arrivant en bordure de la ZCE depuis la zone P n'ont pas encore subi la multiplication par impact, et on peut donc considérer le courant d'électrons en $x = L$ comme égal au courant de saturation des électrons :

$$J_n(L,t) = J_{nS}. \tag{5.24}$$

De même,

$$J_p(0,t) = J_{pS}. \tag{5.25}$$

Le courant total, s'exprime donc, en $x = 0$ et $x = L$ de la manière suivante :

$$\begin{cases} J(t) = J_n(0,t) + J_{pS} \\ J(t) = J_p(L,t) + J_{nS} \end{cases} \tag{5.26}$$

La somme des deux lignes nous donne :

$$2J(t) = J_n(0,t) + J_p(L,t) + J_{pS} + J_{nS} = J_n(0,t) + J_p(L,t) + J_S. \tag{5.27}$$

On déduit alors de 5.24, 5.25 et 5.27 :

$$J_n(L,t) - J_p(L,t) - J_n(0,t) + J_p(0,t) = 2\left[J_S - J(t)\right]. \tag{5.28}$$

L'équation 5.20 se réécrit donc, en divisant par L :

$$\frac{\mathrm{d}J(t)}{\mathrm{d}t} - \left(\frac{2v_{sat}}{L}\int_0^L \alpha(x)dx\right) J(t) + \frac{2v_{sat}}{L}J(t) - \frac{2v_{sat}}{L}J_S = 0 \tag{5.29}$$

L'équation différentielle régissant l'évolution de la densité de courant d'avalanche est donc :

$$\boxed{\frac{\mathrm{d}J(t)}{\mathrm{d}t} - \frac{2v_{sat}}{L}\left(\int_0^L \alpha(x)dx - 1\right) J(t) - \frac{2v_{sat}}{L}J_S = 0.} \tag{5.30}$$

La solution est :

$$\boxed{J(t) = \frac{J_S}{1 - \int_0^L \alpha(x)dx}.\left[1 - \int_0^L \alpha(x)dx.e^{-\frac{2v_{sat}}{L}\left(1-\int_0^L \alpha(x)dx\right)t}\right]} \tag{5.31}$$

D'après l'équation 5.31, la densité de courant suit une courbe exponentielle croissante, avec une constante de temps τ_{ava} égale à :

$$\tau_{ava} = \frac{L}{2v_{sat}\left(1 - \int_0^L \alpha(x)dx\right)}. \tag{5.32}$$

On reconnaît l'intégrale d'ionisation $\Im = \int_0^L \alpha(x)dx$ dans le cas particulier où $\alpha_e = \alpha_h = \alpha$ dans les équations 5.8 ou 5.9.

On voit bien que l'établissement du courant d'avalanche est d'autant plus rapide que la ZCE est courte (L faible) et que l'intégrale d'ionisation est élevée, c'est-à-dire que les coefficients d'ionisation des porteurs sont suffisamment élevés pour compenser une faible longueur de ZCE.

Régime permanent

Cette évolution exponentielle caractérise le régime transitoire. En régime permanent, la densité de courant circulant à travers la diode peut se trouver en écrivant que la dérivée temporelle de la densité de courant s'annule, dans l'équation différentielle 5.30. On obtient alors la valeur suivante :

$$J(+\infty) = \frac{J_S}{1 - \int_0^L \alpha(x)dx}. \tag{5.33}$$

Le facteur

$$M = \frac{1}{1 - \int_0^L \alpha(x)dx} \tag{5.34}$$

est le facteur de multiplication par avalanche (équations 5.10 et 5.10) dans le cas où $\alpha_e = \alpha_h = \alpha$ et $\Im_e = \Im_h = \Im$.

Le paramètre α varie très fortement avec le champ électrique. Sa dépendance avec le champ électrique a fait l'objet d'études empiriques et théoriques. Pour les simulations, on utilise le

modèle de Van Overstraeten et de Man [68]. Dans le cas du silicium, il donne, à température ambiante :

$$\begin{cases} \alpha_e(E) = 3,8.10^6.e^{-(\frac{1,8.10^6}{E})} \\ \alpha_h(E) = 2,2.10^7.e^{-(\frac{3,3.10^6}{E})} \end{cases} \tag{5.35}$$

Une décomposition en séries entières donne une approximation des coefficients d'ionisation des porteurs [69] :

$$\alpha_{eff} = a_\alpha E^7, \tag{5.36}$$

Avec $\alpha_{eff} = 1,8.10^{-35}\ cm^6V^{-7}$ pour le silicium.

On peut alors réécrire la densité de courant en régime permanent (équation 5.33) :

$$J(+\infty) = \frac{J_S}{1 - \int_0^L a_\alpha E^7 dx}. \tag{5.37}$$

Pour simplifier, on se place dans le cas où le champ électrique est constant le long de la ZCE, comme dans la diode PIN. Avec $E = \frac{V_{Diode}}{L}$, V_{Diode} étant la tension aux bornes de la jonction L la longueur de la ZCE, l'équation 5.37 se réécrit :

$$\boxed{J(+\infty) = \frac{J_S}{1 - a_\alpha L(\frac{V_{diode}}{L})^7}} \tag{5.38}$$

et

$$M = \frac{1}{1 - a_\alpha L(\frac{V_{diode}}{L})^7} \tag{5.39}$$

5.7 Déclenchement transitoire du $NPNB$

5.7.1 Introduction

Dans cette section, nous allons porter notre attention sur les phases de déclenchement du $NPNB$. Nous nous placerons dans le cas particulier d'un temps de montée de 10 ns à 100 V TLP. La durée associée à chacune de ces étapes de déclenchement est analysée, depuis le claquage par avalanche jusqu'au repliement de la tension. En revanche les phénomènes associés au comportement du composant après le repliement ne sont étudiés que de manière qualitative, car comme on l'a vu les simulations deviennent moins précises une fois passé le pic de tension (section 5.3).

La tension et le courant aux bornes du composant sont représentés à la figure 5.14. Les

instants t_{BV}, t_{EB}, t_{NPN} et t_{max} correspondent respectivement aux instants du claquage par avalanche de la jonction base-collecteur, du déclenchement de la jonction émetteur-base en direct, du déclenchement du NPN latéral, et du repliement de la tension. L'instant t_r correspond au temps de montée de l'impulsion. A $t < t_{BV} = 2.6\ ns$, aucun courant ne circule dans le composant, et la tension V_{DUT} suit exactement la rampe de tension imposée V_{TLP}. Le claquage par avalanche de la jonction latérale a lieu à t_{BV}. L'effet bipolaire s'enclenche à t_{NPN}=3.1 ns, mais ce n'est qu'à 4.3 ns que le courant crée augmente assez rapidement pour que la condition de repliement (équation 5.6) soit respectée et que la tension puisse commencer à décroître. Il s'écoule donc 1.7 ns entre le claquage par avalanche et le repliement, durant lesquelles la tension s'élève de plus de 8 V.

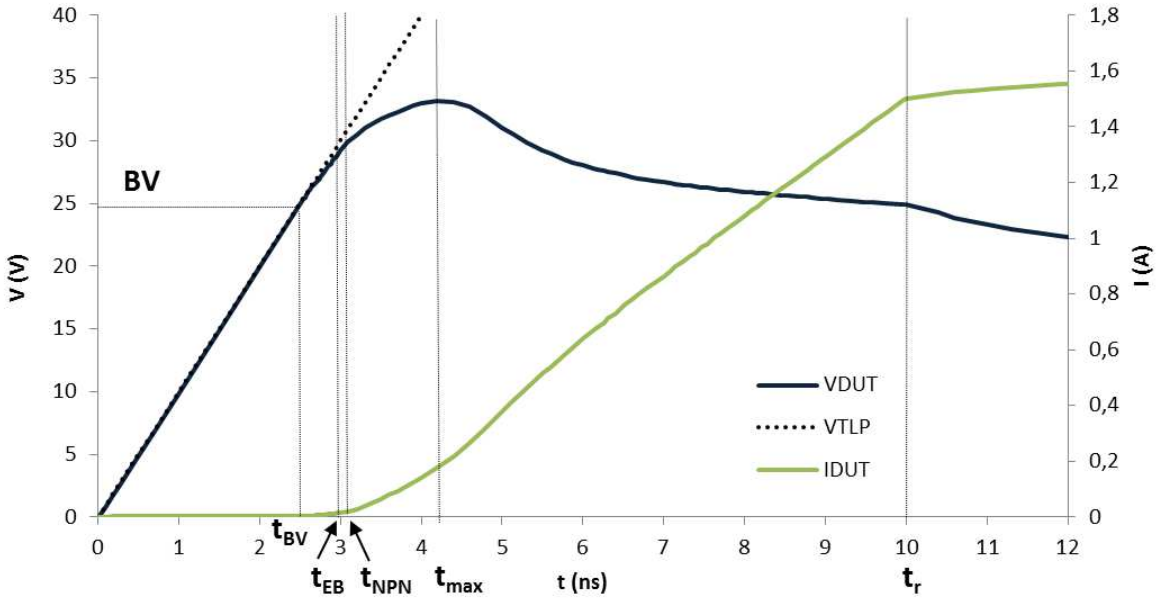

FIGURE 5.14 – Formes d'onde simulées en tension (bleu foncé) et courant (vert clair) du composant sous une rampe TLP de 10 ns de temps de montée.

5.7.2 De t_{BV}=2.6 ns à t_{EB}=2.95 ns : courant de génération

Deux régimes de fonctionnement

Quand la tension atteint la valeur BV, les phénomènes décrits dans la section 5.6 entrent en jeu. La figure 5.15 décrit l'évolution de la valeur maximale du taux de génération par impact G dans l'ensemble de la ZCE. On observe un régime transitoire dont la durée est de l'ordre de 100 ps (voir section 5.6), puis le taux de génération de paires électrons-trous atteint un plateau. Le courant généré suit les deux mêmes régimes (figure 5.16). En régime transitoire, le courant est décrit par la formule 5.31, tandis que sur le plateau du taux de génération G, on peut considérer qu'il est en chaque point solution de l'équation 5.38. Dans cette équation, la tension V_{Diode}

dépend elle-même du courant, par la relation :

$$V_{Diode} = V_{TLP} - (R_{off} + R_{ser}).I_{DUT}, \tag{5.40}$$

avec R_{off} la somme des résistances de collecteur, de base et d'émetteur, et R_{ser} la résistance interne de 50 Ω du générateur TLP (figure 5.1-b).

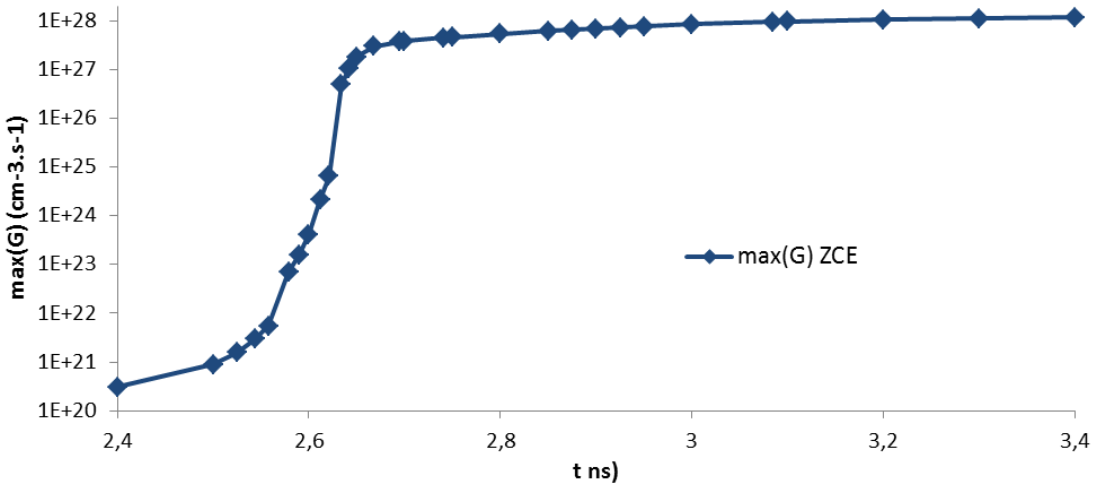

FIGURE 5.15 – Valeur maximale du taux de génération par avalanche dans la zone d'avalanche de la jonction base-collecteur latérale en fonction du temps. On distingue deux régimes, l'un correspondant à un régime transitoire, d'une durée de 100 ps environ, suivi d'une stabilisation à une valeur constante (régime permanent).

Pour le courant, on peut qualifier ce régime de « quasi-permanent » car sa valeur instantanée ne dépend que de la tension appliquée aux bornes du composant.

On note que dans la figure 5.16, le courant de trous au point $B1$ a été obtenu par intégration de la densité de courant de trous sur toute l'épaisseur du faisceau de courant latéral. Les courants relevés dans les figures suivantes ont tous été obtenus de la même manière.

Continuité du courant

Le courant généré par avalanche est un courant de trous du côté P de la jonction base-collecteur, et d'électrons du côté N. On considère que dès le claquage, le courant augmente brutalement dans tout le composant de la même manière que dans la ZCE. En effet, il ne peut y avoir que continuité du courant le long du composant. L'augmentation du courant aux électrodes n'est pas lié à un déplacement de porteurs, mais à un déplacement de « l'information ». Le phénomène est analogue à la propagation d'une onde, et la durée associée à cette propagation

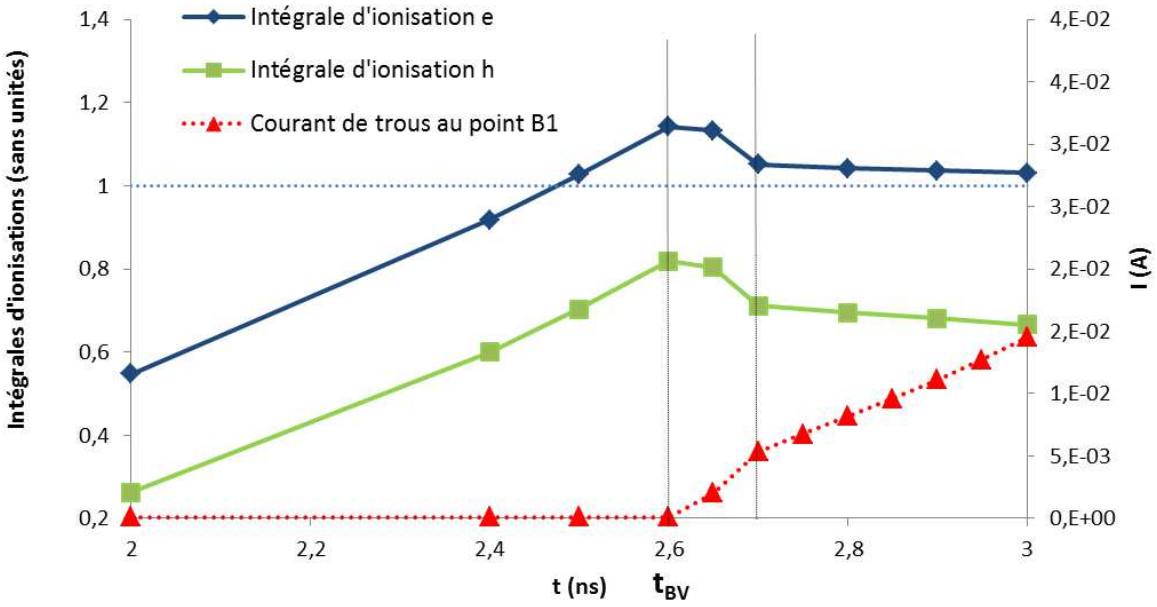

FIGURE 5.16 – Intégrales d'ionisation des électrons (bleu foncé) et des trous (vert clair) calculés le long de la jonction base-collecteur latérale en fonction du temps. Le courant d'avalanche est tracé en rouge pointillés.

est de l'ordre de grandeur du temps de relaxation diélectrique $\tau_{diel} = \frac{\epsilon}{\sigma}$. Pour le Silicium, cette constante vaut environ 1 *ps* et est donc négligeable par rapport aux autres phénomènes. La figure 5.17 montre les courants de trous en deux points de la base : un point $B1$, en bordure de ZCE et un point $B2$ au bord de l'émetteur (figure 5.18). On voit qu'effectivement les courants d'avalanche augmentent de façon simultanée en ces deux points.

Polarisation de la jonction émetteur-base

Le courant d'avalanche n'est pas suffisant à lui seul pour initier un repliement de la tension, et cette dernière continue de croître. A ce stade, la jonction base-émetteur n'est pas encore polarisée en direct. Cependant, à mesure que le courant d'avalanche augmente, une chute de potentiel apparaît à travers la zone P formant la base. Lorsque cette différence de potentiel est suffisante pour polariser en direct la jonction émetteur-base, cette dernière commence à émettre un courant significatif. Étant donnée la disposition de la diode base-émetteur (5.18), la diode se déclenche d'abord en un point $B2$, à l'extrémité droite de la diffusion d'émetteur, puis la zone d'émission s'étend à toute la jonction à mesure que le courant augmente.

Pour mieux saisir l'effet de la résistance de base sur le temps de déclenchement du NPN latéral, considérons le circuit équivalent représenté sur la figure 5.18. Ce dernier, basé sur un modèle d'Ebers-Moll [70] [71] [37], décrit le comportement du composant entre l'instant t_{BV}

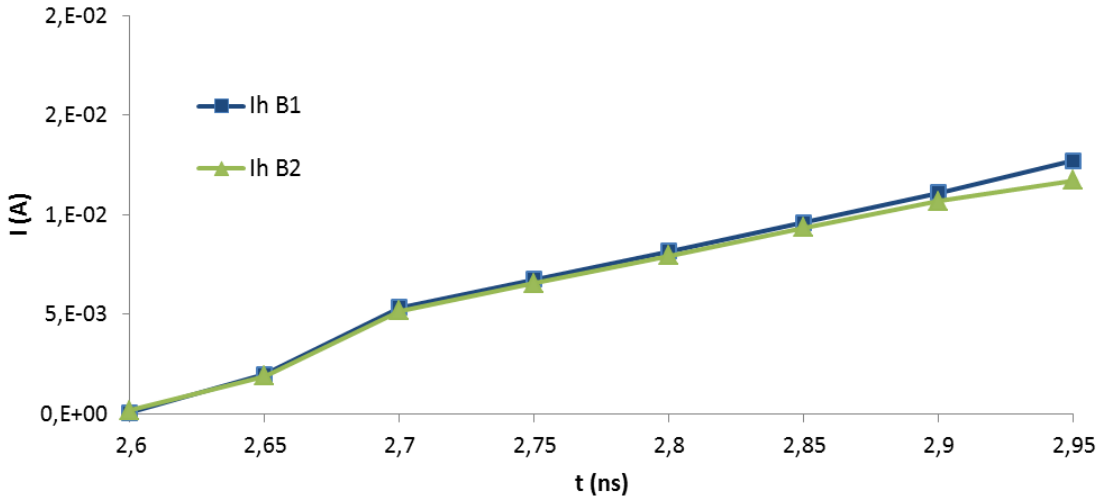

FIGURE 5.17 – Courants d'avalanche (courant de trous) relevés dans la base en deux points : l'un, $B1$ est situé près de la ZCE, et l'autre, $B2$, jouxte la zone de transition de la jonction base-émetteur. Les deux courants sont égaux depuis le claquage à 2.6 ns jusqu'au déclenchement en direct de la jonction émetteur-base à 2.95 ns.

pour lequel le claquage a lieu et l'instant t_{EB} pour lequel la jonction émetteur-base commence à émettre du courant. Les contacts de base et d'émetteurs sont court-circuités, et sont symbolisés par le seul contact d'émetteur E. La jonction émetteur-base est symbolisée par sa capacité de jonction C_J. La résistance de base est scindée en deux parties : R_{B2} représente la résistance de la portion de base reliant le point $B2$ au contact de base, R_{B1} représente la résistance du la portion de la base située entre $B2$ et $B1$.

Le courant d'avalanche I_{ava}, le courant I_{B2} dans la résistance $B2$ et la tension aux bornes de la capacité de jonction C_J vérifient la relation :

$$I_{B2} = I_{ava} - C_J \frac{\mathrm{d}V_{B_2E}}{\mathrm{d}t}. \tag{5.41}$$

Compte-tenu de la faible valeur de la capacité de jonction lorsqu'elle est polarisée en direct, on peut négliger le terme capacitif devant le courant d'avalanche. Le circuit de la figure 5.18 se réduit à celui présenté en (figure 5.19) :

Conséquemment, la tension aux bornes de la résistance R_{B2} devient directement proportionnelle au courant d'avalanche I_{ava}. On déduit que le courant d'avalanche pour lequel la jonction émetteur-base est polarisée en direct est atteint à l'instant $t = t_{EB}$, tel que :

$$I_{ava}(t_{EB}) = \frac{V_{bi}}{R_{B2}} \tag{5.42}$$

La validité du modèle simplifié proposé en figure 5.19 est vérifié par l'évolution du courant

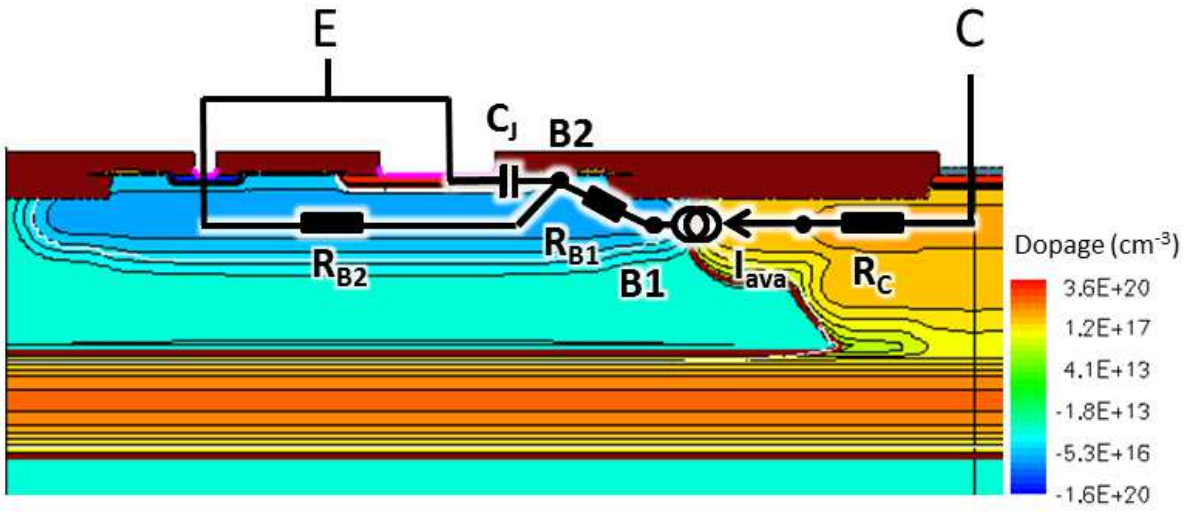

FIGURE 5.18 – Schéma équivalent du NPN latéral entre t_{BV} et t_{EB}. L'établissement de la chute de potentiel aux borne de R_{B2} contrôle le déclenchement en direct de la jonction émetteur-base. Le composant est polarisé suivant les conditions de la figure 5.8 : émetteur et base court-circuités, et collecteur flottant. L'impulsion est appliquée côté collecteur.

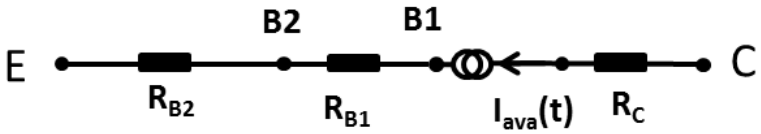

FIGURE 5.19 – Modèle simplifié du NPN avant le déclenchement bipolaire. La capacité de jonction C_J de la figure 5.18 a été supprimée.

d'avalanche à travers la résistance R_{B1} et du potentiel électrostatique au point $B2$ (figure 5.20). On constate un rapport de proportionnalité entre le courant et la tension, correspondant à la résistance R_{B2}, que l'on évalue à environ 25 Ω. Ceci n'est valable qu'à $t < t_{EB}$.

Le déclenchement de la jonction émetteur-base est donc un phénomène qui ne dépend que du courant instantané, qui, comme on l'a vu, ne dépend lui-même, à $t = t_{EB}$, que de la tension appliquée sur la jonction en inverse (équation 5.38), qui elle-même, si on néglige les courants de déplacement, ne dépend que de la tension appliquée sur le composant (équation 5.40). Toutes ces étapes, depuis le claquage par avalanche jusqu'à la polarisation en direct de la jonction émetteur-base, peuvent être considérées comme quasi-statiques. L'instant t_{EB} est le temps pour lequel la tension de composant atteint la tension de repliement quasi-statique $V_{t1} = 28.9\ V$ et le courant I_{ava} atteint I_{t1}.

Cette approximation quasi-statique n'est valable que si le temps de montée de l'impulsion est suffisamment lent pour que les effets transitoires liés au claquage par avalanche soient stabilisés

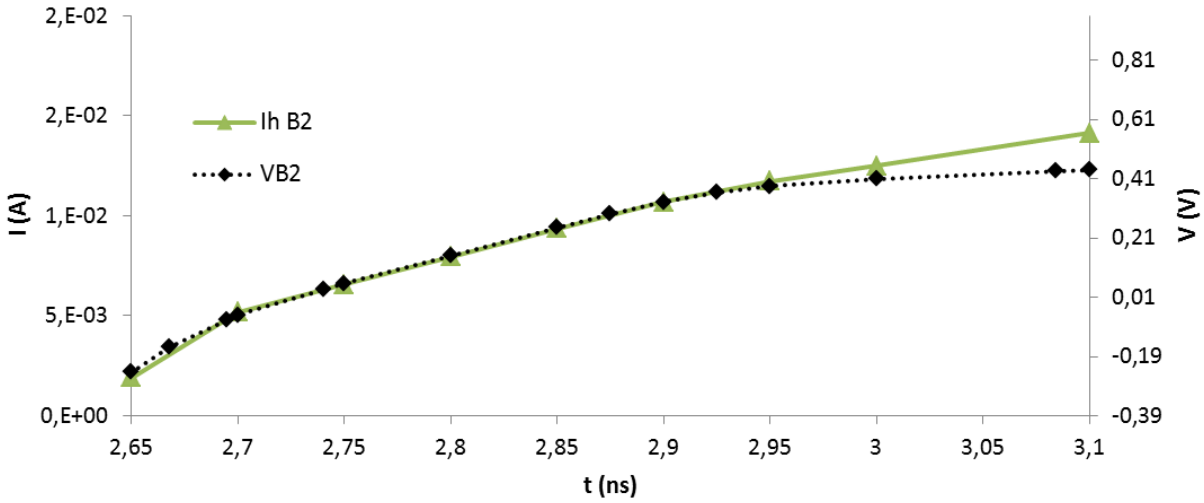

FIGURE 5.20 – Courant dans la résistance de base R_{B1} (noir pointillés) et potentiel dans la base au point $B2$ (trait continu vert clair). Le rapport de proportionnalité entre les deux grandeurs valide l'hypothèse selon laquelle la capacité de jonction émetteur-base est négligeable.

quand le courant atteint sa valeur I_{t1}. Cette condition sur la pente de l'impulsion TLP s'écrit :

$$\frac{V_{t1} - BV}{d(V_{TLP})/dt} > \tau_{ava}. \tag{5.43}$$

Dans le cas présent d'une impulsion de 10 ns de temps de montée à 100 V TLP, le rapport $\frac{V_{t1}-BV}{d(V_{TLP})/dt}$ est de l'ordre de 400 ps, ce qui est supérieur aux 100 ps nécessaires pour installer le régime permanent de la génération par avalanche. De fait, on peut constater sur la figure 5.14, qu'à l'instant t_{EB}=2.95 ns, la tension aux bornes du composant est bien égale à sa tension de déclenchement quasi-statique $V_{t1} = 28.9\ V$.

La pente incidente à partir de laquelle il est nécessaire de prendre en compte la durée d'instauration du régime permanent pour le courant d'avalanche est de 40 V/ns, correspondant à la pente pour laquelle la tension de générateur passe de BV à V_{t1} en 100 ps.

5.7.3 De $t_{EB} = 2.95\ ns$ à $t_{NPN} = 3.1\ ns$: déclenchement du NPN

A l'instant t_{EB}, même si la tension a atteint le seuil V_{t1}, le transistor bipolaire NPN n'est pas encore déclenché. On a vu que ce transistor bipolaire était scindé en deux transistors, l'un latéral et l'autre vertical, partageant la même jonction base-émetteur. Le transistor NPN ne se déclenche que quand le courant d'électrons issus de l'émetteur aura atteint l'une ou l'autre des jonctions base-collecteur.

Déclenchement du NPN latéral

Dans un transistor bipolaire NPN, une fois la jonction émetteur-base à l'état passant, le paramètre limitant la rapidité de déclenchement est le temps de diffusion $\tau_{B,e}$ des électrons à travers la base. Il est donné par la relation suivante [72], [73], tenant compte des effets de forte injection :

$$\tau_{B,e} = \frac{W_B^2}{4D_n}\left(1 + \frac{N_A}{N_A + \Delta p_0}\right), \tag{5.44}$$

où N_A est la concentration en atomes accepteurs, W_B la largeur de la base, D_n le coefficient de diffusion des électrons dans la base, et Δp_0 la concentration de trous en excès dans la base.

Ainsi, en faible injection, quand l'excès de trous n'atteint pas la concentration en accepteurs, on retrouve l'expression connue :

$$\tau_{B,e} = \frac{W_B^2}{2D_n} \tag{5.45}$$

alors qu'en forte injection, celle-ci se rapproche de :

$$\tau_{B,e} = \frac{W_B^2}{4D_n}. \tag{5.46}$$

A l'instant du déclenchement du NPN, la base est en faible injection sur tout son volume. On considère que l'instant t_{NPN} correspondant est :

$$t_{NPN} = t_{EB} + \tau_{B,e} = 3.1\ ns. \tag{5.47}$$

$\tau_{B,e}$ étant défini par la relation 5.45. Petit à petit cependant, la concentration en électrons augmente dans la base, et lors du repliement de la tension, la base est en mode de forte injection sur la moitié de son volume environ.

Déclenchement du NPN vertical : effets de forte injection dans la zone intrinsèque

A partir de t_{EB}, les électrons diffusent dans la base, préférentiellement suivant les forts gradients de concentration. Un certain nombre d'entre eux sont donc amenés à diffuser vers la zone intrinsèque de la diode PIN formant la jonction base-collecteur verticale. On a vu que le champ électrique y était constant en faible injection (figure 5.10-b). Or dès l'instant t_{EB}, la zone intrinsèque entre en régime de forte injection. La charge induite par le courant d'électrons vient alors moduler le profil de champ électrique. Suivant l'équation de Gauss :

$$\frac{\mathrm{d}E}{\mathrm{d}x} = \frac{\rho}{\epsilon}, \tag{5.48}$$

avec ρ la densité de charges, induite dans notre cas par le courant d'électrons. Cette densité de charges négatives a pour effet de diminuer la pente $\frac{\mathrm{d}E}{\mathrm{d}x}$ du côté P et à l'augmenter du côté N. Ce phénomène, communément désigné sous le nom d'« effet Kirk » a tendance à concentrer le champ électrique sur une plus courte distance, du côté du collecteur. On peut constater ce phénomène sur la figure 5.21, qui montre l'évolution du profil de champ électrique de la jonction verticale entre $t = 3\ ns$ et $t = 10\ ns$.

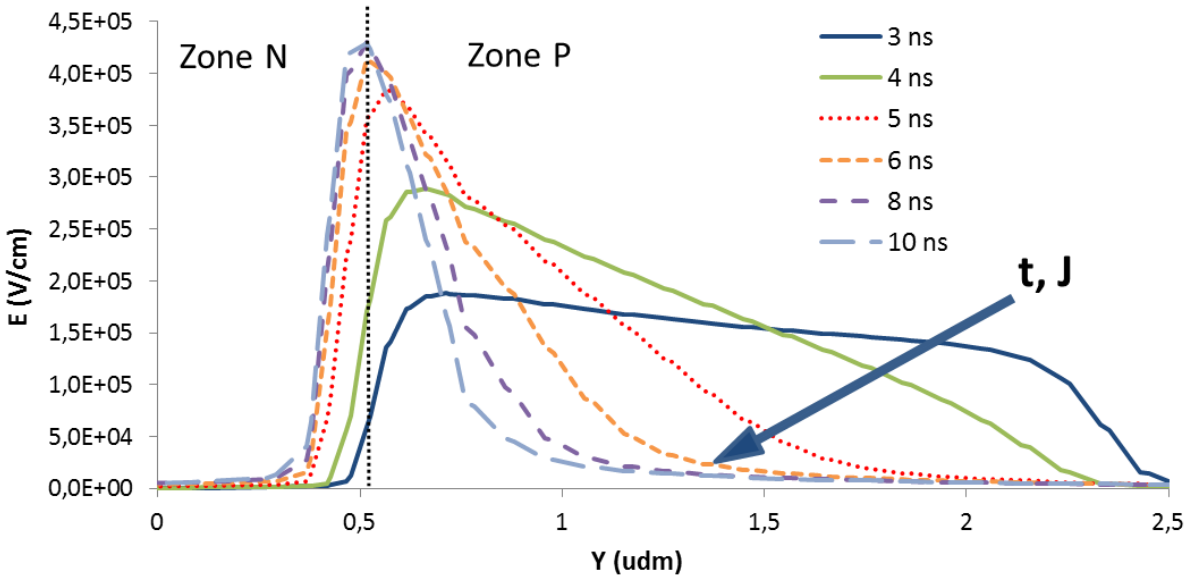

FIGURE 5.21 – Modulation du profil de champ électrique dans la base du NPN vertical. Le zone de fort champ électrique adopte peu à peu le profil d'une jonction abrupte.

La zone de fort champ électrique, vient progressivement se concentrer sur une largeur inférieure à 0.5 udm, pour adopter le profil « triangulaire », centré sur la jonction métallurgique, propre aux jonctions abruptes. Cette réduction dans l'espace va de pair avec une augmentation du champ électrique maximal, et donc des coefficients d'ionisation des porteurs. Les intégrales d'ionisation peuvent ainsi atteindre des valeurs allant jusqu'à 0.39 et 0.18 respectivement pour les électrons et les trous, correspondant à des facteurs de multiplication de 1.64 et 1.21 (figure 5.22). Les électrons injectés par l'émetteur sont multipliés, et le NPN vertical peut être considéré comme déclenché.

Contrairement au NPN latéral, ce n'est pas prioritairement le claquage par d'avalanche qui entraîne le déclenchement du transistor bipolaire, mais l'injection d'émetteur qui abaisse la tension de claquage de la jonction base-colleteur verticale, permettant un déclenchement bipolaire. Le temps de déclenchement est fixé ici aussi par le temps de transit des électrons injecté dans la base verticale, qui est de l'ordre de 150 ps. Cependant, il faut attendre que les effets de forte injection aient modifié le profil de champ électrique pour que le déclenchement soit complet.

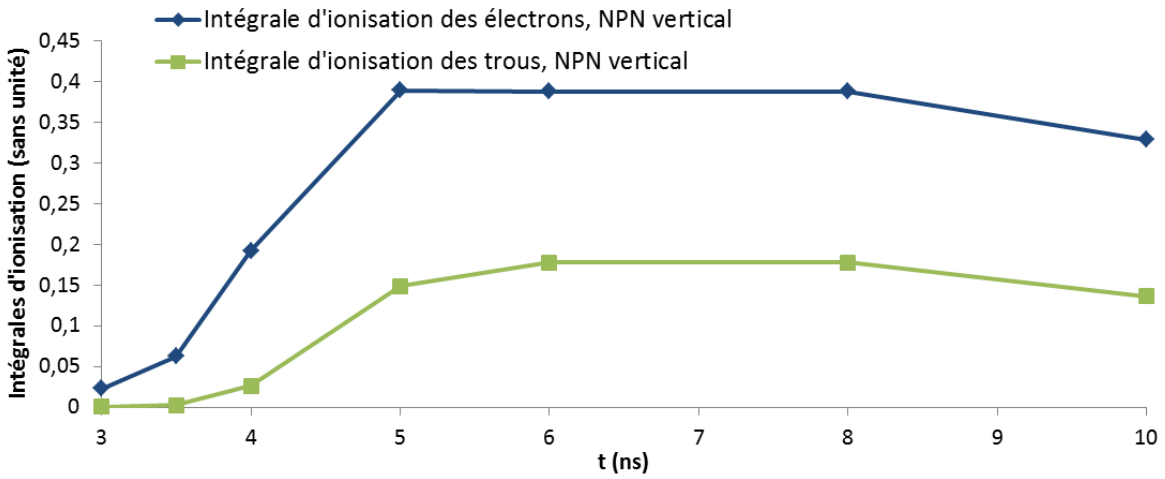

FIGURE 5.22 – Évolution des intégrales d'ionisation des électrons (bleu foncé) et des trous (vert clair) sous l'action de la forte injection. Le maximum est atteint entre 5 et 6 *ns* environ.

Il est à noter que la jonction base-collecteur latérale étant abrupte, elle est moins influencée par les modulations du profil de champ électrique en forte injection.

NPN latéral *vs.* *NPN* vertical

Les courants des *NPN* latéraux et verticaux sont représentés à la figure 5.23. Les transistors bipolaires se déclenchent à peu près au même moment, étant donné les temps de transit τ_{BE} voisins dans leurs bases respectives. Cependant, le coefficient de multiplication à la jonction base-collecteur verticale étant faible à 3.2 *ns*, le courant latéral augmente plus rapidement.

Le courant d'avalanche continue d'augmenter, et la jonction émetteur-base, qui s'était déclenchée au point $B2$ (figure 5.18), émet alors des électrons sur une largeur de plus en plus importante. Ces électrons émis loin du point $B2$ ont une probabilité plus faible de rejoindre la jonction base-collecteur latérale que verticale et ont tendance à diffuser verticalement. A mesure que la zone d'injection de la jonction base-émetteur s'étend et que le facteur de multiplication à la jonction verticale augmente sous l'effet des phénomènes de forte injection, le courant vertical augmente plus rapidement (figure 5.23).

La figure 5.24 montre l'évolution des temps de transit $\tau_{B,e,lat}$ et $\tau_{B,e,vert}$ dans les bases latérales et verticales entre 3 et 4 *ns*. La base latérale étant plus longue que la base verticale, $\tau_{B,e,lat}$ est supérieur à $\tau_{B,e,vert}$. On constate également une légère augmentation de $\tau_{B,e,vert}$ dans le temps, due à l'élargissement de la base effective.

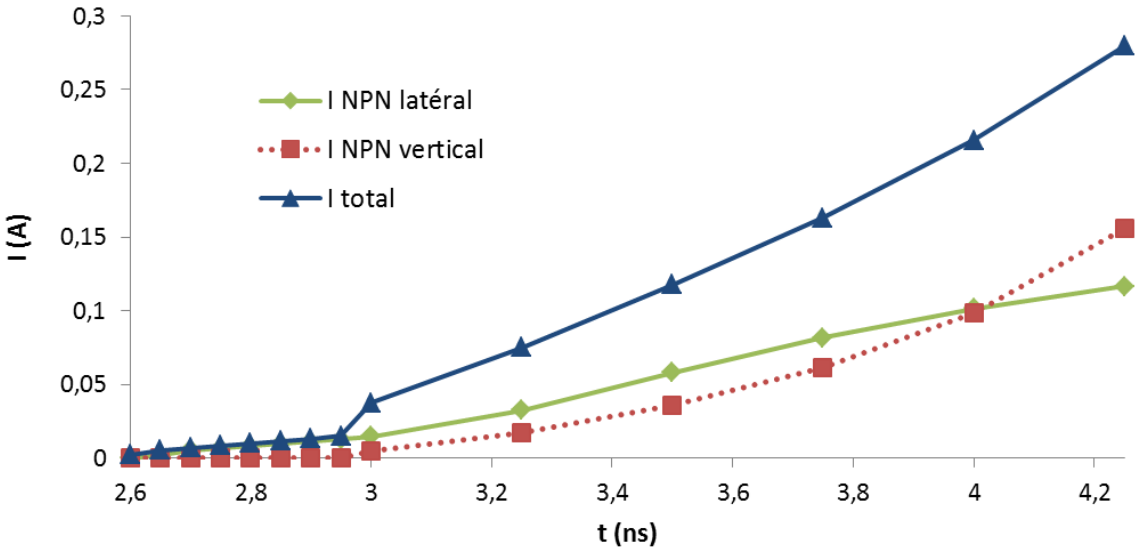

FIGURE 5.23 – Courant du NPN latéral (vert), vertical (rouge pointillé) et courant total (trait plein bleu).

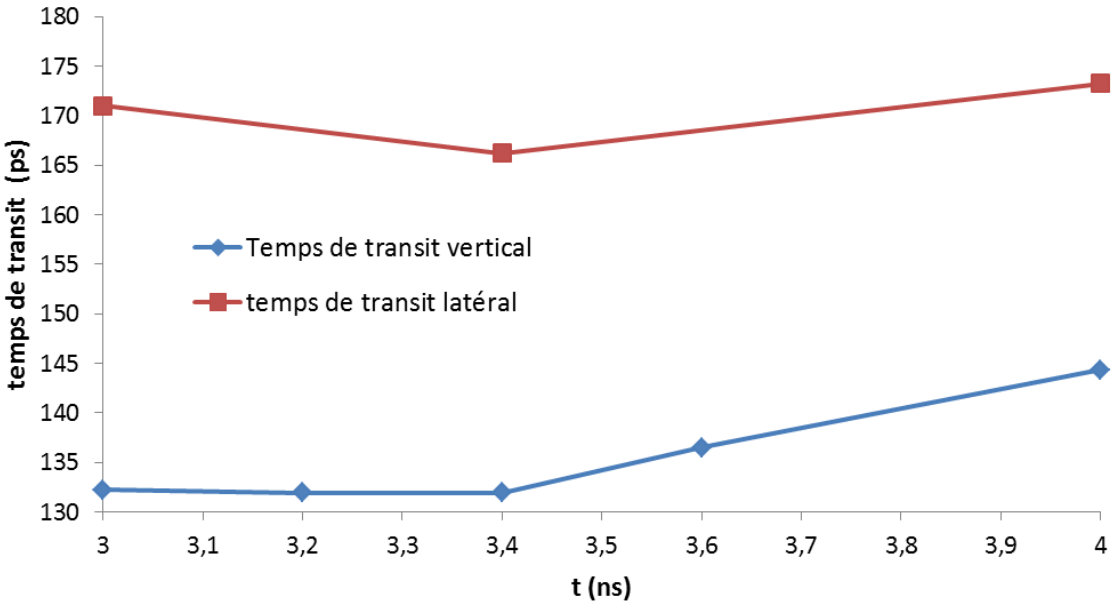

FIGURE 5.24 – Évolution des temps de transit des électrons dans la base latérale (bleu foncé) et verticale (vert clair). Pour la base latérale, le calcul a été fait le long de lignes de courant rejoignant la zone d'avalanche depuis l'extrémité droite de l'émetteur (trajet $[B2 - B1]$). Les valeurs relevées ne prennent donc pas en compte l'augmentation moyenne du temps de transit due à l'élargissement de la zone d'injection d'émetteur le long de la jonction émetteur-base.

5.7.4 De t_{NPN}=3.1 ns à $t_{max,\ NPN\ lat.}$ = 3.6 ns : étude du repliement du NPN latéral

Les phénomènes responsables du repliement du NPN latéral sont intimement liés à la dynamique de diffusion des électrons de l'émetteur vers la jonction base-collecteur. En effet, nous allons voir que l'augmentation du courant d'électrons au point $B1$ conditionne à la fois la modulation de conductivité de la base et celle du profil de champ électrique à la jonction base-collecteur. En d'autre termes, la rapidité d'augmentation du courant en $B1$ détermine directement la surtension observée à la fois aux bornes de la base et à la jonction base-collecteur.

La dynamique de ce phénomène est régie par l'équation de contrôle de charge, que nous allons examiner dans un premier temps. Puis nous allons analyser séparément la surtension apparaissant dans les deux régions sur lesquelles elle se concentre principalement : la base et la jonction base-collecteur.

Équation de contrôle de charge

L'équation de contrôle de charge décrit la dynamique de la diffusion des porteurs minoritaires dans la zone quasi-neutre.

$$J_n(B2) - J_n(B1) = \frac{\mathrm{d}Q_S}{\mathrm{d}t} + \frac{Q_S}{\tau_n}, \tag{5.49}$$

avec $J_n(B2)$ et $J_n(B2)$ les densités d'électrons aux points $B2$ et $B1$ respectivement, τ_n la durée de vie des électrons dans la base, et Q_S la charge stockée par la population d'électrons par unité de surface :

$$Q_s = q \int_{B1}^{B2} n.dx, \tag{5.50}$$

L'équation 5.49 écrit que la différence du courant d'électrons entre la jonction émetteur-base et la jonction base-collecteur ne peut s'expliquer que par un stockage de charges ou par la recombinaison des électrons.

Dans notre cas, la durée de vie des électrons dans la base étant de l'ordre de 100 ns, on néglige le courant de recombinaisons :

$$\boxed{J_n(B2) - J_n(B1) = \frac{\mathrm{d}Q_S}{\mathrm{d}t}}. \tag{5.51}$$

Avant que le NPN vertical ne se déclenche et ne commence à générer un courant significatif, on peut considérer que la majorité du courant passe par le collecteur latéral. Et avant que le

PNP ne se déclenche, ce courant est principalement un courant d'électrons. Autrement dit, avec J_{tot} la densité de courant totale dans le composant et M_e le coefficient de multiplication des électrons dans la jonction émetteur-base, on peut écrire :

$$J_n(B1) = \frac{J_n(Collecteur)}{M_e} = \frac{J_{tot}}{M_e}. \tag{5.52}$$

De plus, au point $B2$, on peut relier le courant d'électrons au rapport d'injection α de la jonction émetteur-base. Le rapport d'injection étant par définition le rapport de la densité d'électrons traversant la jonction par la densité de courant totale, on peut en effet écrire :

$$J_n(B2) = \alpha J_{tot} \tag{5.53}$$

L'expression 5.51 s'écrit alors :

$$\boxed{\left(\alpha - \frac{1}{M_e}\right) J_{tot} = \frac{\mathrm{d}Q_S}{\mathrm{d}t}.} \tag{5.54}$$

L'équation 5.54 est une équation différentielle linéaire du premier ordre. Malgré son apparente simplicité, elle n'a pas de solution analytique simple, ses coefficients étant tous dépendant du temps, et la charge stockée étant définie par une intégrale sur la largeur de la base (équation 5.50). On note cependant que la charge stockée dans la base, et donc sa conductivité, augmente avec le courant. De plus, la conductivité d'équilibre est atteinte d'autant plus rapidement que le rapport d'injection de la jonction émetteur-base et le coefficient de multiplication des électrons dans la zone d'avalanche sont élevés.

A l'équilibre, l'équation 5.54 exprime l'égalité entre le rapport d'injection et l'inverse du coefficient de multiplication des électrons. On retrouve l'équation définissant la tension de maintien V_H d'un NPN dans [31] :

$$\alpha = \frac{1}{M_e}. \tag{5.55}$$

La figure 5.25 compare les courants d'électrons aux points $B1$ et $B2$ dans le temps. On note un délai d'environ 200 ps entre l'élévation de courant aux deux extrémités de la base, correspondant au temps de transit $\tau_{B,e}$. La différence entre les deux courbes, ensuite provient du stockage de charges dans la base (équation 5.51).

Repliement de la tension aux bornes de la base

A t_{NPN}=3.1 ns, le NPN latéral est déclenché, mais sa base est encore en régime de faible injection. La conductivité $\sigma = q(n\mu_n + p\mu_p)$ de la base augmente progressivement sous l'afflux

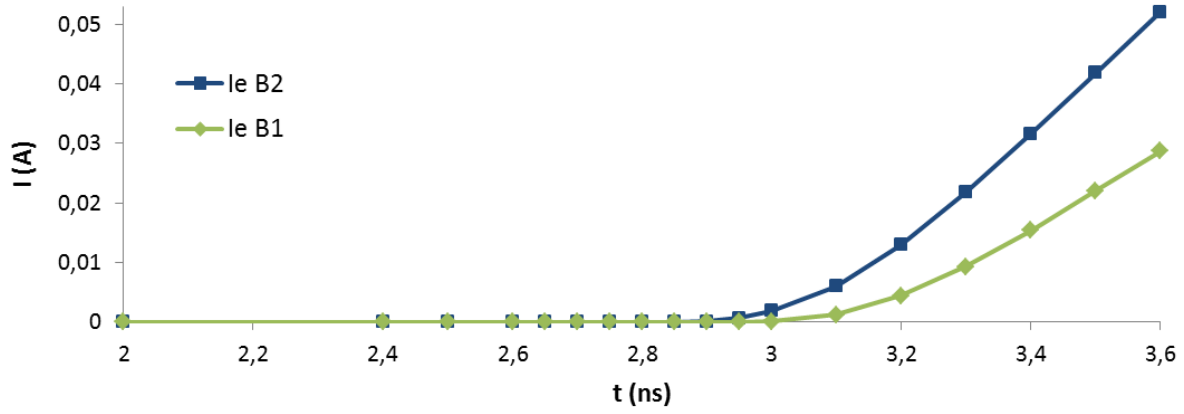

FIGURE 5.25 – Courants d'électrons relevés aux points $B1$ et $B2$ de la base. La charge stockée totale correspond à l'aire entre les courbes $I_{e,B2}$ et $I_{e,B1}$.

des électrons et des trous. La chute de résistance de la base provoque un repliement de la tension à ses bornes.

Nous nous proposons d'estimer la durée nécessaire au repliement de la tension V_B aux bornes de la base. Pour un courant total $I(t)$ et une résistance de base instantanée $R_B(t)$, cette tension s'écrit :

$$V_B(t) = R_B(t).I(t). \tag{5.56}$$

Ici, I est une fonction croissante du temps, et R_B est une fonction décroissante du courant, et donc du temps. Fixons l'origine des temps en $t = t_{NPN}$. Pour simplifier, considérons le courant comme nul avant cet instant. Le courant imposé est une rampe, et l'on écrit :

$$I = \frac{\mathrm{d}I}{\mathrm{d}t}.t, \tag{5.57}$$

et

$$V_B(t) = R_B(t).\frac{\mathrm{d}I}{\mathrm{d}t}.t, \tag{5.58}$$

avec $\frac{\mathrm{d}I}{\mathrm{d}t}$ constant. La dérivée de la tension par rapport au temps est :

$$\frac{\mathrm{d}V_B}{\mathrm{d}t} = \frac{\mathrm{d}R_B}{\mathrm{d}t}\frac{\mathrm{d}I}{\mathrm{d}t}t + R_B\frac{\mathrm{d}I}{\mathrm{d}t}. \tag{5.59}$$

$$\frac{\mathrm{d}V_B}{\mathrm{d}t} = \frac{\mathrm{d}I}{\mathrm{d}t}\left[\frac{\mathrm{d}R_B}{\mathrm{d}t}t + R_B(t)\right]. \tag{5.60}$$

La tension V_B ne pourra commencer à décroître que quand le terme négatif $\frac{\mathrm{d}R_B}{\mathrm{d}t}t$ sera de-

venu suffisamment important et $R(t)$ suffisamment faible pour que les deux termes s'annulent mutuellement. Le repliement de la tension V_B aux bornes de la base a lieu à un instant τ_R tel que :

$$\tau_R = -\frac{R_B(\tau_R)}{\frac{\mathrm{d}R_B}{\mathrm{d}t}(\tau_R)} \tag{5.61}$$

Émettons alors l'hypothèse selon laquelle le rapport $\frac{R_B(t)}{\frac{\mathrm{d}R_B}{\mathrm{d}t}(t)}$ varie peu pendant la durée τ_R. Cela revient alors à résoudre l'équation différentielle suivante au voisinage de τ_R :

$$\boxed{\frac{\mathrm{d}R_B}{\mathrm{d}t}(t) = -\frac{R_B(t)}{\tau_R}} \tag{5.62}$$

La solution est de la forme :

$$R_B(t) = R_{B,off}.e^{-\frac{t}{\tau_R}}, \tag{5.63}$$

avec $R_{B,off}$ la valeur de la résistance de base non modulée. Si la concentration en accepteurs est constante dans la base, on a :

$$R_{B,off} = \frac{1}{l_B L}\int_{B2}^{B1}\frac{\mathrm{dx}}{\sigma(x)}, \tag{5.64}$$

soit :

$$R_{B,off} = \frac{W_B}{q.l_B L.N_A\mu_p}. \tag{5.65}$$

avec l_B la profondeur du caisson P de base, L la longueur du doigt, et W_B la largeur de base. L'équation 5.63 s'écrit alors :

$$\boxed{R_B(t) = \frac{W_B}{q.l_B L.N_A\mu_p}.e^{-\frac{t}{\tau_R}}.} \tag{5.66}$$

Si l'hypothèse 5.62 est valide, la résistance de base est une fonction exponentielle décroissante du temps. La tension aux bornes de la base est définie par l'expression 5.58, et se replie en $t = \tau_R$.

La figure 5.26 montre l'évolution de la tension et de la résistance de la base à partir de t_{BV}. Tant que le transistor n'est pas déclenché ($t < t_{NPN}$), la résistance de la base reste égale à $R_{off} \approx 27\ \ \Omega$, puis la base stocke petit à petit des charges, et R_B chute de façon exponentielle. La constante de temps associée à cette décroissance correspond environ à la durée s'écoulant entre le déclenchement du NPN latéral à t_{NPN} et le repliement de la base à $t_{max,base}$, ce qui valide le résultat 5.66. La surtension accumulée sur la base pendant ce laps de temps est de 0.4 V.

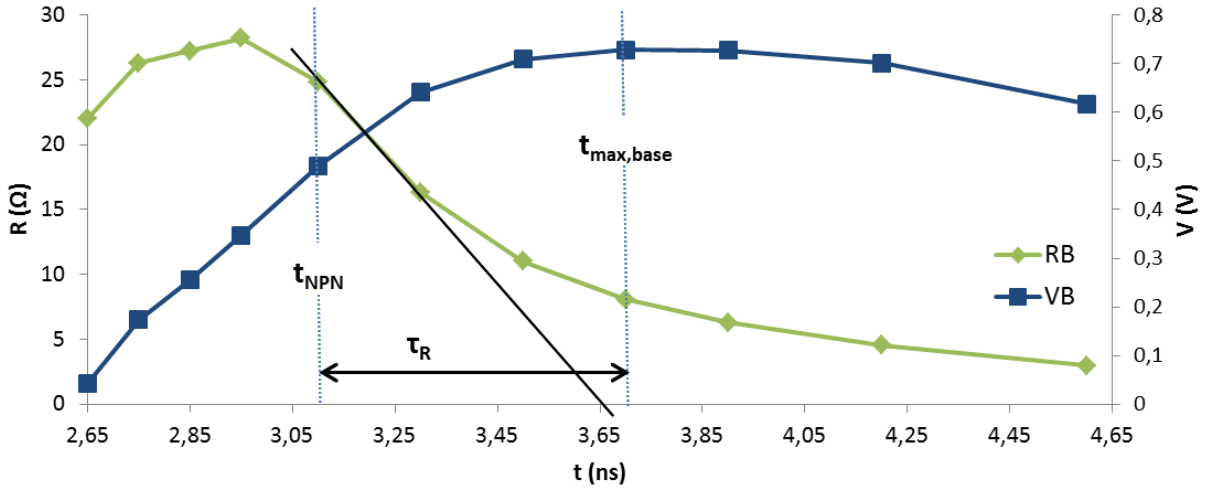

FIGURE 5.26 – Évolution de la tension aux bornes de la base (tension V_{B1B2}, en bleu foncé) et de la résistance de cette portion (en vert clair) en fonction du temps. Le temps s'écoulant entre le déclenchement de l'effet bipolaire à t_{NPN} et le repliementde la tension à t_{max} correspond approximativement à la constante de temps caractéristique τ_R de la décroissance exponentielle de la résistance.

Repliement de la chute de potentiel aux bornes de la ZCE

Examinons à présent la surtension qui se crée aux bornes de la ZCE. On va voir qu'elle résulte directement du temps nécessaire pour charger la base. En effet, le processus provoquant le repliement de la jonction base-collecteur découle de l'effet Kirk, autrement dit de la modulation du champ électrique par le courant d'électrons. On a vu précédemment comment dans le NPN vertical, la charge d'espace négative induite par le courant d'électrons dans la base influait sur la pente dE/dx du champ électrique (équation 5.48). Le phénomène est le même ici, bien que de moins grande ampleur. La charge d'espace induite par le courant d'électrons a tendance à concentrer le champ électrique sur une plus faible largeur, et la tension aux bornes de la jonction base-collecteur, définie par :

$$V_{BC} = \int_{x_p}^{x_n} E.dx, \tag{5.67}$$

s'en trouve diminuée, avec x_p et x_n les bornes de la zone de charge d'espace.

A mesure que l'émetteur injecte des électrons, l'augmentation de la densité d'électrons dans la base réduit la charge d'espace au voisinage du point $B1$, ce qui tend à réduire l'extension de la zone de fort champ électrique et la tension V_{BC}. Parallèlement, l'excès d'électrons injectés dans la base peut participer à l'avalanche, ce qui permet également de réduire le facteur de multiplication, donc la tension V_{BC}. Quand ces effets deviennent suffisamment importants, un repliement de la tension V_{BC} aux bornes de la ZCE peut avoir lieu.

La figure 5.27 montre une comparaison entre la tension aux bornes de la base V_B et celle aux bornes de la jonction base-collecteur V_{BC} en fonction du temps. En l'absence d'effets transitoires, ces deux tensions n'auraient pas dû dépasser leur valeur quasi-statique, atteinte à $t = t_{EB}$. On définit la surtension dans la base et sur la jonction base-collecteur par rapport à ces valeurs quasi-statiques. La surtension totale de 1.3 V aux bornes du NPN latéral par rapport à son V_{t1} se répartit en 0.9 V sur la tension base-collecteur et 0.4 V sur la base.

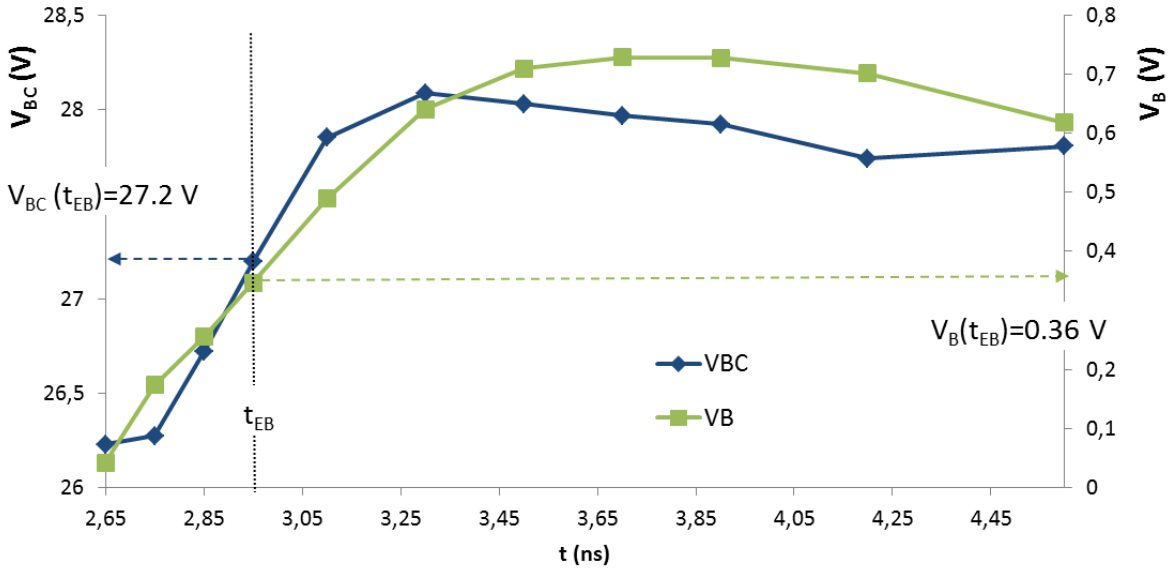

FIGURE 5.27 – Tension aux bornes de la base (vert clair) et aux bornes de la jonction base-collecteur (bleu foncé) en fonction du temps. La valeur atteinte pour chacune des parties à l'instant t_{EB} sert de point de référence pour le calcul de la surtension. On note que la plus grande partie de la surtension est prise en charge par la jonction base-collecteur, et que celle-ci se replie en premier.

5.7.5 Conduction du transistor PNP et de la diode verticale

La partie du composant connectée à la borne positive est constituée de deux transistors bipolaires PNP (figure 5.8). L'un, latéral, a sa jonction base-collecteur commune avec celle du NPN latéral, le caisson P relié au signal constituant l'émetteur. L'autre, vertical est constitué du même caisson P (émetteur), de la couche enterrée N^+ (base) et partage la jonction base-collecteur du NPN vertical. Du fait du fort dopage de la couche enterrée, les trous s'y recombinent rapidement, ce qui empêche le déclenchement du transistor. Le gain bipolaire du PNP vertical est très faible, et il se comporte comme une diode en direct PN^+. Le PNP latéral, en revanche, peut se déclencher et forme un thyristor avec le NPN latéral.

Du fait de la forte dissymétrie de dopage de la diode verticale, le rapport d'injection de la NBL dans le caisson P est élevé, et c'est elle qui se déclenche en premier, prenant en charge le

courant, ensuite relayée par le PNP latéral.

Diode verticale

Au claquage de la jonction base-collecteur du NPN latéral, un important courant d'électrons circule à travers le collecteur latéral et la couche enterrée NBL. Ce courant transite ensuite à travers la diode PN^+ en direct, puis rejoint le contact sur lequel est appliquée la décharge électrostatique. Un délai de 400 ps est nécessaire à partir de $t = t_{BV}$ pour déclencher la diode en direct, après quoi elle conduit le courant, et la tension à ses brones est maintenue constante et égale à 0.7 V. La figure 5.28 montre le courant circulant dans la zone P de la diode, et le rapport d'injection en fonction du temps. On peut identifier le courant injecté par le NPN, augmentant brusquement à partir de l'instant t_{BV}.

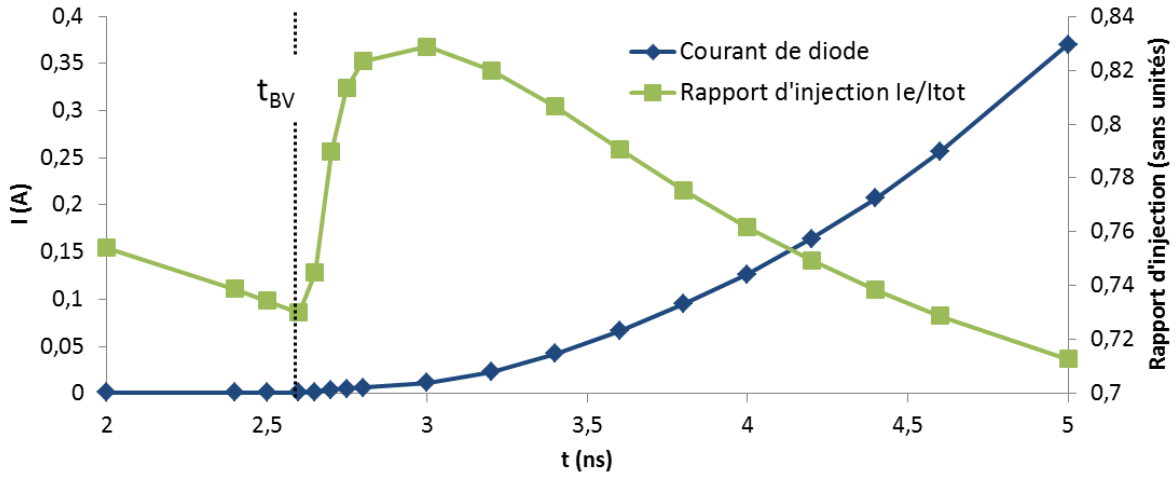

FIGURE 5.28 – Courant circulant dans la diode P^+N et rapport d'injection $\alpha = \frac{I_e}{I}$ en fonction du temps. L'augmentation du rapport d'injection à partir de l'instant t_{BV} montre le déclenchement de la diode en direct.

PNP latéral

La polarisation en direct de sa jonction émetteur-base induit le déclenchement du transistor PNP à 3.4 ns par effet thyristor (figure 5.29). Le courant de trous circulant alors à travers le collecteur flottant du NPN contribue également à l'effet d'avalanche dans la jonction base-collecteur. L'effet thyristor est fortement limité par la longueur et le dopage élevé de la base du PNP.

La figure 5.29 présente le courant de trous dans la base du PNP, au bord de la jonction émetteur-base : on voit ce courant augmenter à partir de 3.4 ns, témoignant du déclenchement de la jonction émetteur-base.

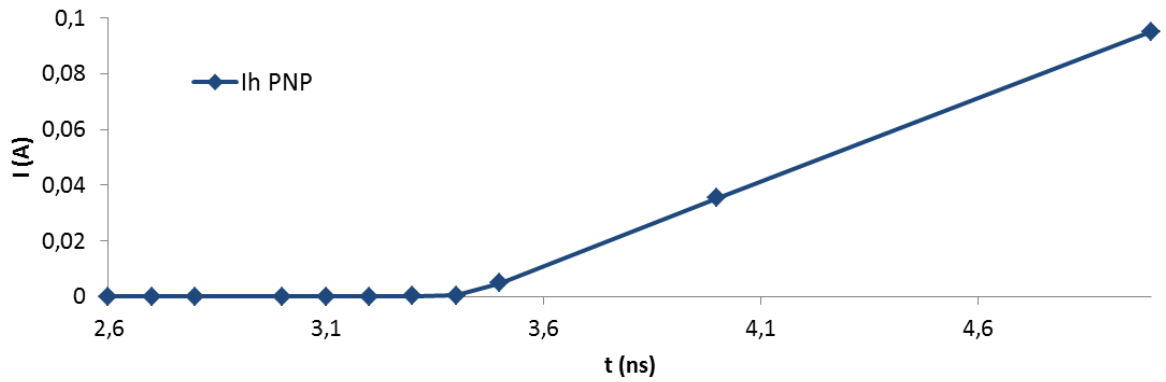

FIGURE 5.29 – Courant de trous dans la base du PNP latéral. La jonction émetteur-base commence à injecter des trous à partir de 3.4 ns.

5.8 Solutions de dessin envisagées

Dans cette partie, nous allons exposer quelques règles de dessin applicables au layout du $NPNB$ pour favoriser son déclenchement et réduire la surtension. Nous avons vu que durant son déclenchement, la tension maximale avant repliement ne pouvait pas être inférieure à sa tension de repliement quasi-statique V_{t1}. Il est donc naturel de choisir V_{t1} comme valeur de référence par rapport à laquelle on définit la surtension. La surtension ST s'écrit donc :

$$ST = V_{max} - V_{t1}, \tag{5.68}$$

avec V_{max} la tension maximale atteinte lors du pic de surtension.

En conséquence, les règles de dessin permettant d'abaisser la tension de claquage BV ou de repliement quasi-statique V_{t1} ne seront pas abordées, celles-ci n'abaissant pas la surtension.

5.8.1 Règles de dessin

Optimiser la jonction collecteur-base latérale

Cette règle concerne soit l'augmentation du courant d'avalanche I_{ava}, soit la diminution de la durée du régime transitoire du claquage par avalanche latéral.

Augmenter le facteur de multiplication par avalanche de la jonction latérale permettrait d'augmenter le courant conduit par le NPN latéral et le PNP, dès leur déclenchement. On peut obtenir une augmentation du facteur de multiplication en modifiant la géométrie de la jonction base-collecteur latérale de manière à étendre la zone de fort champ électrique. Pour la jonction verticale, en revanche, le facteur de multiplication est plus difficile à contrôler étant donné qu'il dépend des caractéristiques de la technologie utilisée, telles que l'épaisseur de l'épi-

taxie et le dopage de la couche enterrée.

On a vu dans la partie « Polarisation de la jonction émetteur-base » de la section 5.7.2 que la durée du régime transitoire n'influait sur la surtension que pour des pentes incidentes dépassant 40 V/ns. Réduire cette durée pourrait se faire cependant en rapprochant les zones P^+ et N^+ de base et collecteur pour concentrer la zone de fort champ électrique sur une plus petite distance (formule 5.32).

Optimiser la jonction émetteur/base du NPN

Augmenter le rapport d'injection de la jonction émetteur-base du NPN augmente le courant d'électrons injectés dans la base, et donc diminue le temps nécessaire pour moduler sa conductivité et initier le repliement (équation 5.54). Le dopage pour cela doit être très dissymétrique entre l'émetteur et la base (base peu dopée, émetteur très dopé).

Dopage et géométrie de la base du NPN

La base du NPN latéral est essentielle pour son déclenchement. Elle doit être faiblement dopée, courte et peu profonde. Un dopage faible permet d'augmenter le coefficient de diffusion des électrons dans la base, et donc de diminuer leur temps de transit (équation 5.44). Or, un faible temps de transit autorise un déclenchement rapide de l'effet bipolaire (équation 5.47), ainsi qu'un faible temps de stockage des charges dans la base (section 5.7.4).

Réduire le temps de transit des électrons dans la base peut se faire en réduisant sa largeur. De manière générale, le volume de la base doit être faible. Ainsi, les charges y sont stockées rapidement, ce qui permet d'abaisser sa résistance et la chute de potentiel à ses bornes. De plus, cela permet d'abaisser le temps de transit dans la base (équations 5.45 et 5.46).

Diminuer la résistance des parties ohmiques du composant

Les zones purement résistives du $NPNB$ sont responsables pour une part significative de la surtension. Principalement, la couche enterrée NBL, bien que fortement dopée présente une résistance d'environ 70 Ω du fait de sa longueur (20 udm). Or, avant que le PNP latéral ne se déclenche, un partie du courant traverse la couche enterrée pour rejoindre la diode verticale. La solution la plus simple pour diminuer cette résistance consiste à réduire l'espace entre les connecteurs. Une autre solution consiste à augmenter la taille totale du composant en allongeant les doigts, ou par un dessin « multidoigts ». Ces solutions ont pour effet de diminuer le R_{ON} du composant, ce qui est globalement bénéfique en termes de réduction de la surtension.

Augmenter la rapidité de déclenchement du PNP et de la diode verticale

Une solution pour accélérer le repliement du $NPNB$ consiste à renforcer l'effet SCR en augmentant la rapidité de déclenchement du PNP. Les règles de dessin associées portent principalement sur la longueur de base et d'émetteur. La concentration du dopage de la base varie de 10^{17} à 10^{18} cm^{-3}, et la durée de vie des trous y est de 10 ns, soit dix fois plus faible que celle des électrons dans la base du NPN. Réduire la longueur de la base du PNP permet donc de diminuer le taux de recombinaisons. L'émetteur est long (environ 10 udm) et faiblement dopé, ce qui le rend très résistif (75 Ω) avant que les effets de forte injection ne viennent moduler sa conductivité. De plus, le processus de modulation de conductivité est d'autant plus lent que l'émetteur est long. Donc réduire la longueur de l'émetteur du PNP peut également favoriser le repliement du composant.

5.8.2 Application des règles de dessin

Cette section présente des résultats de mesures $transient-TLP$ sur des composants optimisés suivant les règles de dessin proposées. Celles-ci sont souvent les mêmes que celles permettant de diminuer la résistance à l'état passant et la tension de maintien. Si la réduction du R_{ON} est souhaitable, il n'en est pas de même pour la tension de maintien. Une faible tension V_H augmente le risque de latch-up pour le composant.

Une première série de mesures a été effectuée sur boîtier, avec l'outil de mesure $transient-TLP$ développé à $Freescale$ (chapitre 4), générant des fronts de montée de 3.3 ns. Les séries de mesures suivantes ont été réalisées avec l'outil développé au $LAAS$ (chapitre 3), pour un temps de montée de 10 ns.

Taille du composant

La figure 5.30 montre les résultats pour différentes longueurs de doigt, 75 μm et 250 μm d'un $NPNB$ $Pad40$ en technologie $SmartMos$ 8 MV. On remarque que la tension de repliement est identique pour les deux composants, ce qui simplifie la comparaison de leurs surtensions (formule 5.68). L'augmentation d'un facteur 3.3 de l'aire du composant se traduit par une réduction d'un facteur 2 du R_{ON}, et d'un facteur entre 1.2 et 1.5 pour la surtension. La réduction du R_{ON} n'est pas proportionnelle à l'augmentation de l'aire du fait d'un phénomène de focalisation du courant.

Le R_{ON} plus faible du composant long autorise la tension à se replier plus bas que celle du composant court. Pour ce qui est de la surtension, elle est abaissée par la diminution de la résistance à l'état passant. Mais cette réduction est limitée, du fait de la diminution de la densité de courant et donc des effets de forte injection. A cela vient s'ajouter le fait qu'une partie de la surtension provient de délais indépendants de la longueur des doigts, tels que le temps de transit

des électrons dans la base du NPN.

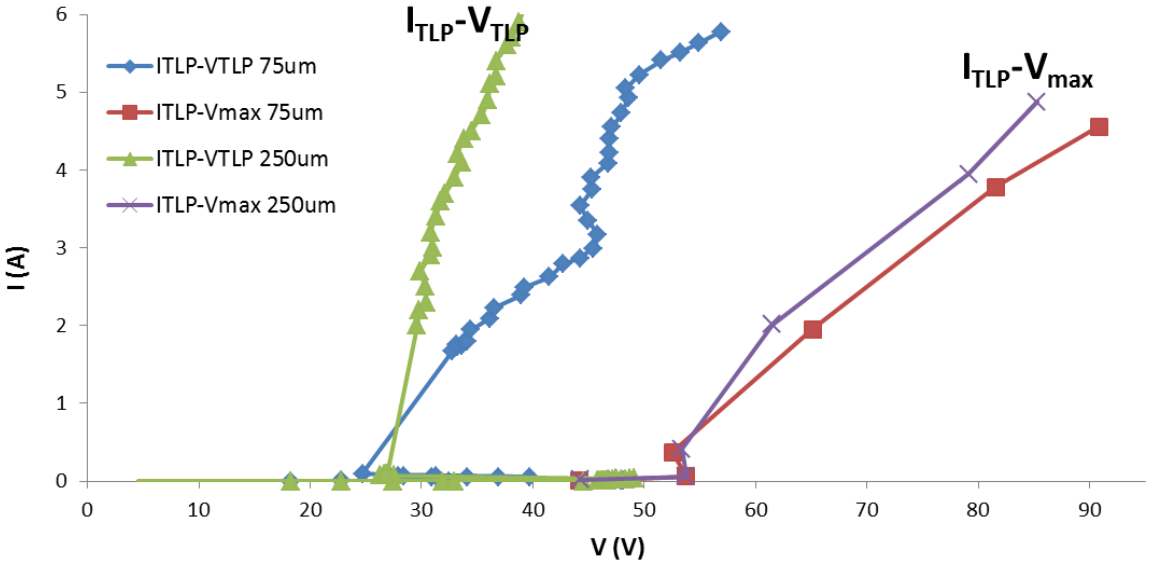

FIGURE 5.30 – Caractéristiques $I-V$ quasi-statiques de composants $NPNB$ $Pad40$ pour deux longueurs de doigt : 75 μm et 250 μm (courbes bleue foncé et vert clair respectivement). Les tensions maximales sont reportées en abscisses, avec le courant quasi-statique en ordonnée (carrés rouges et croix violettes). Les deux composants ayant la même tension V_{t1}, leurs surtensions se comparent directement à partir des valeurs de V_{max}.

Suppression de la tranchée d'isolation peu profonde (STI)

La réduction du temps associé au stockage de charges dans la base est limitée par les caractéristiques de la technologie. Le dopage de la base, par exemple ne peut pas être modifié. Il n'est pas possible non plus d'augmenter la profondeur de l'émetteur afin de réduire le volume de la base et diminuer le temps de stockage des charges.

En revanche, il est possible de réduire la longueur de la base et sa concentration en dopants en supprimant la tranchée d'isolation située au-dessus de la jonction base-collecteur latérale (figure 5.8). Cette tranchée rallonge le trajet des électrons dans la base de 0.2 udm environ. De plus, elle se substitue à une zone de la base plus faiblement dopée.

La réduction du temps de transit lié à la suppression de cette tranchée a pour effet de favoriser la diffusion des électrons vers la jonction base-collecteur latérale, dont le facteur de multiplication est plus important que celui de la jonction verticale. Par ailleurs, cette réduction de la largeur de base a pour effet de réduire le V_{t1} d'environ 4 V.

La figure 5.31 montre les caractéristiques TLP d'un composant de référence avec tranchée d'isolation et d'un composant sans tranchée d'isolation, en technologie $SmartMos$ 8 MVS. La technologie $SmartMos$ 8 MVS est une version low cost de $SmartMos$ 8 MV, offrant un panel réduit de solutions technologiques. On remarque que le V_{t1} du $NPNB$ de référence est plus faible en 8 MVS qu'en 8 MV.

La différence entre les tensions de repliement quasi-statiques des composants avec et sans la tranchée d'oxyde est de 4 V et celle entre leur tension de maintien de 1.3 V. La faible tension de maintien du composant « sans tranchée » est préjudiciable vis-à-vis du latch-up. La figure 5.32 montre une comparaison des surtensions ST pour des tensions de générateur croissantes. La réduction de la surtension ne devient apparente qu'à partir de 80 V TLP environ, et atteint 40 % environ à 160 V.

Notons que la robustesse n'est que peu affectée par la suppression de la tranchée. Le courant de défaillance est de 7.2 A pour le composant de référence contre 7 A pour le composant sans tranchée.

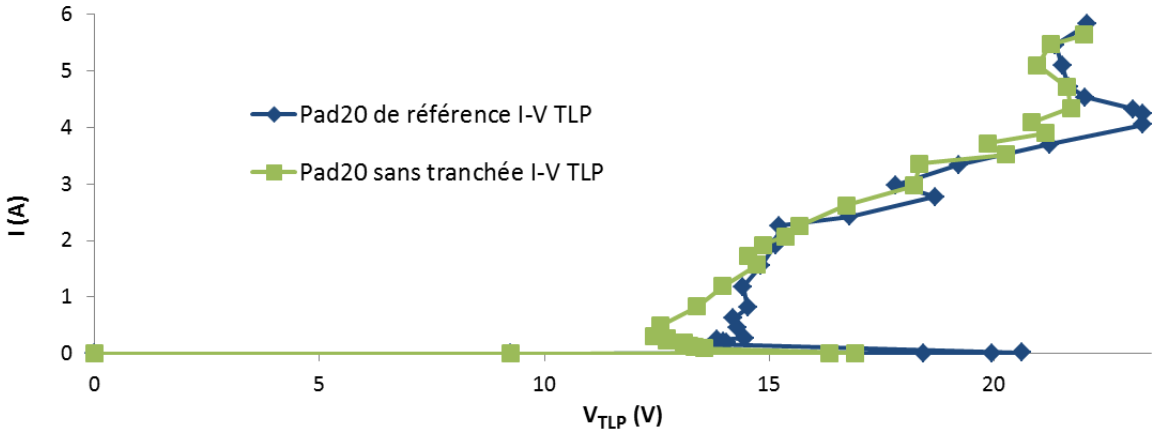

FIGURE 5.31 – Caractéristiques TLP des $NPNB$ standard (courbe bleu foncé) et sans tranchée (vert clair) en technologie $SmartMos$ 8 MVS. Un différence de 4 V est visible entre leurs tensions de repliement.

Longueur de la base du PNP

La figure 5.33 montre une comparaison entre les caractéristiques TLP de deux composants en technologie $SmartMos$ 8 MVS, ne différant que par la longueur de la base de leur PNP. La base du PNP, dans le $NPNB$ $Pad20$ de référence a une longueur de 4.7 udm, contre 3.3 udm dans l'autre layout testé. L'effet sur le V_{t1} comme sur le V_H est très faible (moins de 0.5 V de différence). Cela est probablement dû à la faiblesse de l'effet SCR à bas courant, le PNP

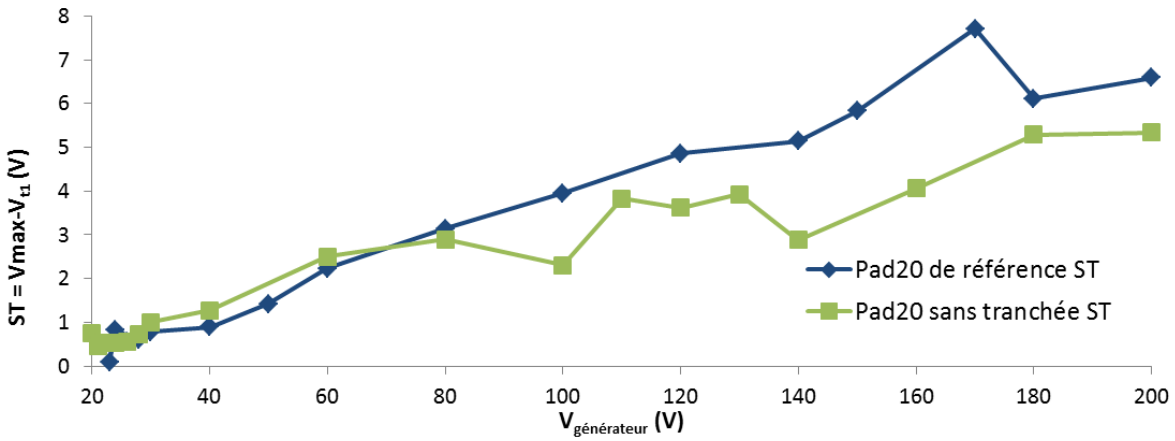

FIGURE 5.32 – Mesure des surtensions des deux composants précédents de 20 *V LP* à 200 *V TLP*. La différence entre les deux surtensions au-delà de 200 *V TLP* est de l'ordre de grandeur de la précision de l'outil de mesure, et elles ne sont donc pas représentées ici.

ne se déclenchant que sous de fortes densités de courant. La figure 5.34 montre l'évolution de la surtension *ST* pour les deux mêmes composants, pour des tensions de générateur croissantes. La diminution de la surtension atteint presque 50 %.

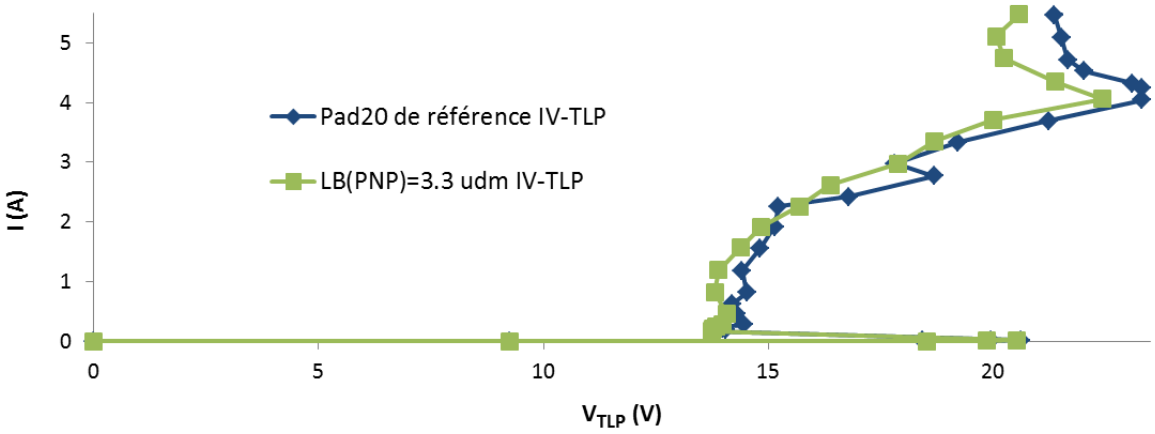

FIGURE 5.33 – Caractéristiques *TLP* de deux *NPNB* en technologie *SmartMos* 8 *MVS* pour deux longueurs de base du *PNP* différentes.

Longueur de l'émetteur du *PNP*

On a vu à la section 5.8.1 que la longueur de l'émetteur du *PNP*, de par sa résistivité élevée pouvait avoir un effet important sur la surtension. Deux « Pads » 15 *V NPNB* de longueurs d'émetteur différentes sont comparés pour la technologie *SmartMos* 8 *MV*. Ceux-ci sont très proches des *Pad* 20, mais ont une tension de repliement quasi-statique plus faible, de par la distance réduite entre les implantations de base et de collecteur du *NPN*. Les autres paramètres, et en particulier ceux influant sur le comportement du *SCR* sont inchangés.

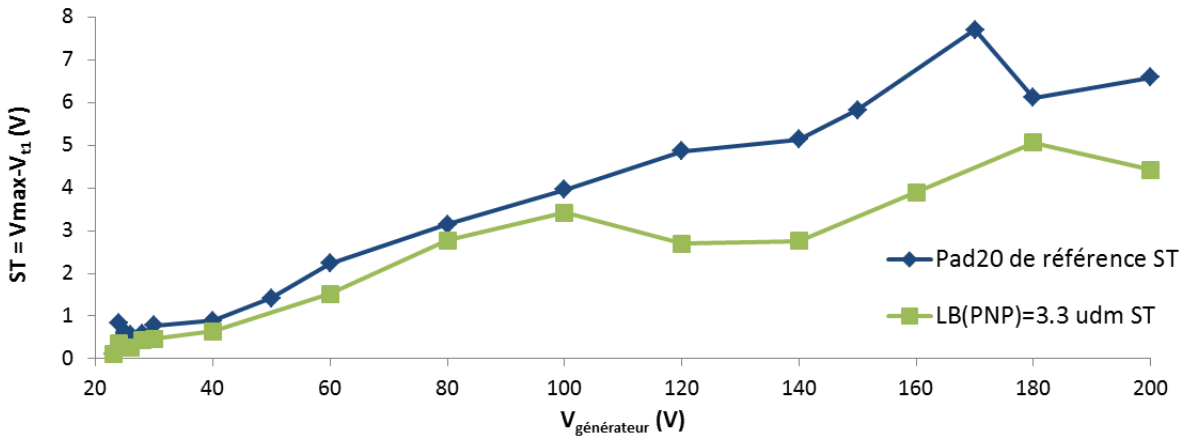

FIGURE 5.34 – Surtensions comparées des deux $NPNB$ précédents entre 20 V et 200 V.

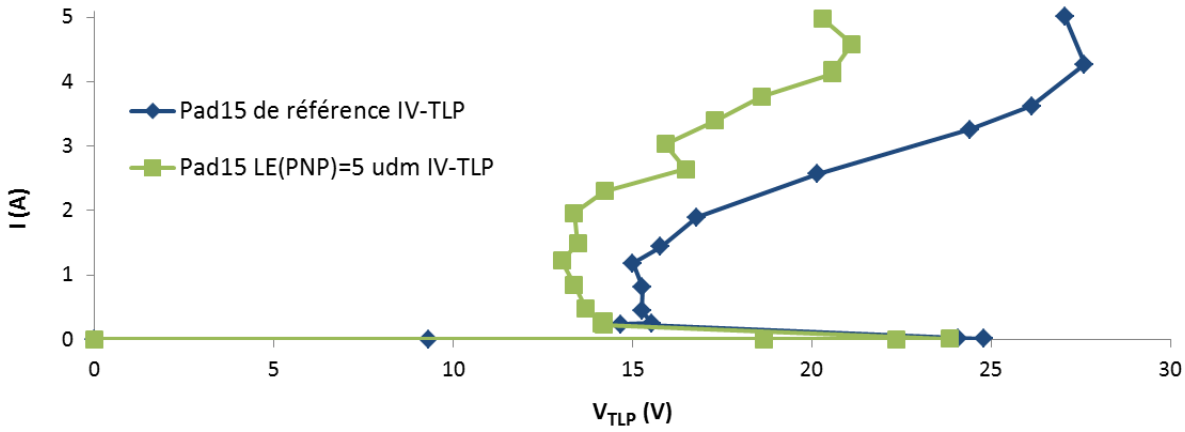

FIGURE 5.35 – Courbes quasi-statiques de deux $Pad15$ $NPNB$ en technologie $SmartMos$ 8 MV, pour des longueurs d'émetteur du PNP de 8 udm (bleu foncé) et de 5 udm (vert clair).

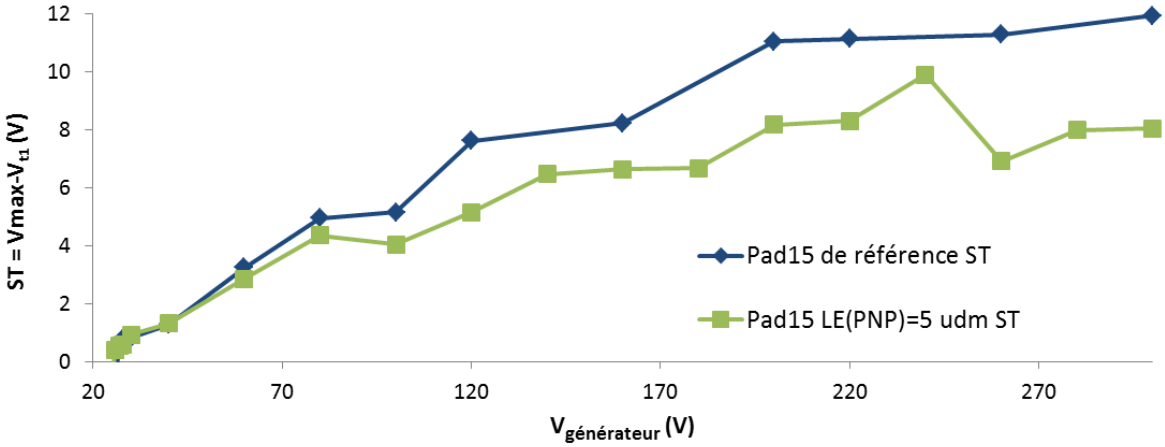

FIGURE 5.36 – Surtensions mesurées sur les deux composants de la figure 5.35.

La seule différence entre les deux $Pad15$ étudiés est la longueur de l'émetteur du PNP. Le $Pad15$ de référence possède un émetteur de 8 udm et l'autre $Pad15$ un émetteur de 5 udm. Le changement a été obtenu en permutant les diffusions N^+ et P^+ reliés au signal. La figure 5.35 montre les caractéristiques TLP de ces deux $NPNB$ $Pad15$. Une réduction d'un peu moins de 1 V est observée sur la tension V_{t1}, et de 0.5 V sur la tension de maintien V_H. La résistance à l'état passant, en revanche, est nettement plus faible pour le composant avec une longueur d'émetteur de 5 udm.

La figure 5.36 montre l'effet de la longueur d'émetteur du PNP sur la surtension : la réduction de surtension atteint 33 % à 300 V.

5.9 Conclusion

Dans un premier temps, nous avons déterminé l'influence de paramètres tels que la résistance interne du générateur et la pente de l'impulsion incidente sur la durée de la surtension. Ensuite, nous avons montré une bonne corrélation entre les simulations $TCAD$ et les mesures $transient-TLP$ pendant le déclenchement d'une protection ESD. Une brève étude sur le composant à protéger a montré que le temps nécessaire pour l'endommager était du même ordre que la durée du pic de surtension généré par le $NPNB$. D'où l'importance de la réduction de l'amplitude de ce pic. Une analyse détaillée des étapes de déclenchement du $NPNB$ a été menée, débutant par le déclenchement transitoire des deux jonctions en inverse, suivi des NPN latéraux et verticaux et enfin du PNP latéral et de la diode verticale. Autant que possible, une évaluation quantitative des durées associées à chaque processus a été proposée.

Le repliement des NPN est directement contrôlé par les phénomènes de stockage de charges dans la base. Le repliement du NPN latéral initie ensuite le déclenchement des autres parties du composant : NPN vertical, PNP et diode verticale. Nous avons pu constater que la contribution principale à la surtension provenait de la forte résistance à l'état passant des parties du composant connectées au signal, et principalement la couche enterrée et le PNP latéral. Basé sur ces constatations, un certain nombre de règles de dessin ont pu être proposées. Celles-ci portent sur la rapidité de repliement des NPN que la résistance des parties passives, la rapidité de déclenchement et la résistance à l'état passant du PNP. La validation de ces règles sur des composants optimisés a montré que la réduction la plus importante de la surtension était obtenue par l'optimisation de la base et de l'émetteur du PNP.

Conclusion générale

Nous avons présenté dans le chapitre 1 la problématique associée à la protection des circuits intégrés contre les ESD. Les modèles décrivant les principaux types de décharges ont été introduits, et les instruments de mesure permettant d'étudier le comportement d'un composant soumis à une ESD ont été exposés. Une attention particulière a été apportée aux outils TLP et $vf-TLP$, permettant de générer des impulsions carrées de temps de montée et durée d'impulsion réglables. Les types de composants de protection ESD les plus couramment employés dans l'industrie ont été présentés, ainsi qu'une partie de ceux utilisés à Freescale en technologie $SmartMos\ 8\ MV$. Le fonctionnement du $NPNB$ est ici abordé, avant d'être étudié plus profondément dans le chapitre 5.

Dans le chapitre 2, nous nous sommes intéressé aux méthodes de mesure les plus couramment utilisées pour analyser le comportement transitoire des protections ESD haute tension, ou du moins pour évaluer l'amplitude du pic de surtension à leur déclenchement. Les limites du système de mesure $vf-TLP$ pour des mesures transitoires ont été mises à jour. En particulier, la forte atténuation fréquentielle des composants du banc de mesure est responsable d'une distorsion importante du signal mesuré. Les difficultés liées à l'alignement des impulsions incidentes et réfléchies à l'oscilloscope sont une source d'incertitude supplémentaire.
Une seconde méthode est étudiée. Il s'agit de la méthode « indirecte », consistant à placer le composant de protection à tester en parallèle avec un composant dont la tension de destruction est connue, et à étudier la robustesse de l'ensemble. Cette opération permet de mettre en évidence l'existence d'un pic de surtension et d'en évaluer l'amplitude. Cette méthode souffre de nombreuses limitations d'ordre pratique (coût des structures à ajouter sur les véhicules de test, nécessité de répéter les mesures plusieurs fois pour confirmer les résultats). Elle nécessite en outre un excellent étalonnage du banc de mesure, puisque des mesures TLP réalisées à bas courant sont comparées à des mesures à haut courant. Elle est en outre relativement mal adaptée à la caractérisation de composants sensibles aux effets de déclenchement par « dV/dt ».

Dans le chapitre 3, nous avons présenté la méthode de mesure originale, nommée $transient-TLP$, permettant de capturer les formes d'onde en courant et tension d'un composant soumis à une agression de type $TLP/vf-TLP$. Le principe de cette méthode est relativement simple, mais

sa mise en œuvre nécessite un travail complexe de post-traitement. Une méthodologie est tout d'abord proposée pour la détermination des paramètres optimaux d'acquisition de l'oscilloscope (résolution temporelle, durée d'acquisition) et de l'analyseur de réseau (résolution fréquentielle, fenêtre de fréquences), afin de tirer le meilleur parti des possibilités de ces appareils tout en limitant la taille des fichiers à traiter. On voit par exemple qu'il est inutile d'abaisser la résolution fréquentielle du VNA en-deçà de 2.4 MHz, les points correspondant se trouvant en dehors du spectre fréquentiel des impulsions TLP et $vf-TLP$. A l'inverse, la bonne résolution temporelle de l'oscilloscope doit être exploitée : le pas en temps est fixé à 50 ps.
Le traitement appliqué au signaux mesurés est ensuite exposé en détails. Le signal acquis à l'oscilloscope doit être nul avant et après l'impulsion de manière à respecter l'approximation du signal périodique, fondamental pour l'application des algorithmes de DFT.
Une étape cruciale du traitement concerne la séparation des impulsions incidentes et réfléchies mesurées à l'oscilloscope. Deux méthodes sont proposées pour isoler ces deux signaux superposés. L'une consiste à allonger la ligne de transmission responsable du délai entre les deux impulsions. L'autre consiste à mesurer séparément l'impulsion incidente, et à la retrancher au signal total pour obtenir l'impulsion réfléchie. Dans la pratique, nous avons tiré parti de ces deux approches. La première méthode a pour inconvénient d'accentuer le filtrage imposé à l'impulsion incidente de par la plus grande longueur de câble coaxial, et en particulier à atténuer son front montant. L'allongement de la ligne de transmission doit donc prendre en compte le temps de montée maximal que l'on souhaite appliquer au composant sous test. La seconde méthode requiert, pour être appliquée, un générateur d'impulsion aussi répétitif que possible. Une alternative est proposée cependant pour pallier l'effet d'un générateur non parfait.
Les signaux parasites, issus de la réflexion de l'impulsion réfléchie sur le générateur sont également supprimés. Pour ce faire deux approches sont ici aussi combinées. L'une consiste à allonger la ligne de transmission située en amont de la sonde en tension, de manière à retarder l'apparition de ces parasites. Cette méthode souffre de l'inconvénient évoqué plus haut, lié à l'atténuation du front montant de l'impulsion à travers une plus grande longueur de câble coaxial, et est donc employée avec précautions. Le signal parasite est ainsi isolé, puis supprimé à l'aide de fonctions de fenêtrage, choisies de manières à ce que leur incidence sur le spectre soit minimale.
Les résultats de mesure présentés sur des structures de calibration (court-circuit, circuit ouvert) montrent une erreur inférieure à 2 %. Le $transient-TLP$ est également utilisable pour des mesures quasi-statiques, pour lesquelles il présente une précision bien supérieure aux mesures TLP et $vf-TLP$. Des exemples de mesure sont ensuite présentés sur des composants de protection ESD de $Freescale$.
Enfin, une méthode d'étalonnage avancée est présentée, dans laquelle les étapes de mesure à l'analyseur de réseau sont remplacées par une série de mesures à l'oscilloscope. Outre les aspects pratiques liés à la réduction de l'équipement nécessaire, cette méthode permet de pallier les éventuels problèmes de linéarité des coefficients de paramètres -S pour des niveaux courant

croissants. De plus, l'étalonnage plus facile du banc de mesure peut être réalisé avant chaque étape de mesure, et les éléments du banc peuvent être plus aisément remplacés.

Le chapitre 4 montre une application de la méthode $transient - TLP$ pour des composants montés en boîtier. La problématique est plus complexe que pour les mesures sur plaque de silicium, car le système de mesure n'est pas dans ce cas adapté 50 Ω jusque dans le plan du composant. Chaque rupture d'impédance est source de réflexions multiples qui doivent être prises en compte dans le calcul. Le problème peut être présenté comme un système à cinq inconnues, nécessitant au moins cinq équations pour pouvoir être résolu. Ces cinq équations sont fournies par une série de cinq mesures effectuées à l'aide de l'oscilloscope, le banc de mesure étant placé dans différentes configurations. Ces étapes d'étalonnage réalisées, la mesure sur un composant en conditions TLP « classique » permet, par le calcul, de retrouver les tensions et courants vus par le composant.
Les mesures sur des structures d'étalonnage montrent une erreur comprise entre 1 % et 5 %.

Le chapitre 5 est une étude du déclenchement du $NPNB$ et des phénomènes donnant lieu à l'établissement d'un pic de surtension à son déclenchement. Tout d'abord nous démontrons que les conditions pour lesquelles la tension aux bornes d'un composant entame son repliement ne concernent pas seulement le comportement du composant, mais également la pente de l'impulsion incidente et l'impédance du générateur. Il en résulte qu'une valeur de surtension mesurée ou simulée sur un composant doit toujours être associée non seulement à un niveau de tension appliqué, mais aussi au type de décharge choisi (TLP, HBM) et au temps de montée de l'impulsion.
Une corrélation entre les mesures $transient - TLP$ et les simulations $TCAD$ est ensuite effectuée sur un composant de protection ESD en conditions TLP pour différents temps de montée. La corrélation s'avère bonne jusqu'au sommet du pic, puis se dégrade singulièrement. Entre autres, ces résultats permettent de confirmer la réalité physique du pic de surtension déjà observé en simulations : ce pic n'est pas un artéfact de simulation, et il est donc pertinent d'étudier les étapes conduisant à son apparition.
Une analyse de la robustesse vis-à-vis de pics de tension du composant que le $NPNB$ doit protéger est livrée. Nous voyons que le laps de temps nécessaire pour détruire ce composant est de l'ordre de grandeur de la durée du pic de surtension aux bornes du $NPNB$. On ne peut donc pas considérer le pic de surtension comme inoffensif pour le composant à protéger du fait de sa brièveté.
La section suivante prépare l'étude du déclenchement du $NPNB$, en proposant une analyse de la dynamique du processus d'avalanche dans ses deux jonctions $P - N$ en inverse. Une étude analytique est proposée, et l'expression en fonction du temps de la densité de courant dans les jonctions en avalanche est livrée, permettant d'identifier les facteurs gouvernant cette dyna-

mique. Nous voyons qu'à la fois la longueur de la zone d'avalanche comme la valeur de l'intégrale d'ionisation sont des paramètres déterminants. En régime permanent, une relation entre la densité de courant et la tension appliquée est donnée.
Nous effectuons ensuite une analyse des étapes du déclenchement du $NPNB$ soumis à un stress TLP depuis le claquage par avalanche de sa jonction base-collecteur latérale jusqu'au sommet du pic de surtension. Nous démontrons entre autres que pour peu que la pente incidente reste inférieure à 40 V/ns, les phénomènes ayant lieu avant l'instant de polarisation de la jonction émetteur-base du NPN latéral n'ont aucun impact sur la rapidité de déclenchement du composant. Nous montrons que les facteurs régissant le déclenchement transitoire du $NPNB$ sont d'une part la rapidité de charge de la base du NPN latéral, et d'autre part l'efficacité de l'effet SCR du NPN avec le PNP. Le premier des deux facteurs peut être contrôlé par le dopage et la géométrie de la zone de base du PNP (base peu dopée et courte). Le second est amélioré significativement par la réduction de la longueur de la base et de l'émetteur du PNP.

Ces travaux ouvrent différentes perspectives. La méthode de mesure développée pour des impulsions $TLP/vf-TLP$ doit pouvoir être appliquée pour des décharges de type HBM, permettant ainsi de tester le comportement du composant dans des conditions réelles. Des mesures HMM transitoires sont également envisageables pour peu que le générateur utilisé soit un générateur 50 Ω de type TLP modifié. L'application au test pistolet reste pour l'instant plus hypothétique, à cause, notamment de la mauvaise répétabilité du dispositif.
Il serait également intéressant d'intégrer les simulations $TCAD$ transitoires du composant de protection dans un modèle plus global du circuit intégré, prenant en compte la propagation de l'impulsion depuis la broche testée jusqu'au composant lui-même.

Annexe A

Éléments d'analyse de Fourier

Les mathématiciens Fourier et Laplace ont été les premiers, au XVIII^{e} siècle, à théoriser les différentes relations de transformations réciproques existant entre le domaine fréquentiel et le domaine temporel.

A.1 Décomposition en séries de Fourier (DSF)

La décomposition en séries de Fourier peut s'appliquer aux signaux infiniment périodiques, c'est-à-dire à des signaux temporels définis de $t = -\infty$ à $t = +\infty$, et qui répètent le même signal un nombre infini de fois avec une période T_p. Fourier a montré qu'un tel signal pouvait être décomposé en une somme de séries de fonctions sinus et cosinus, qu'on nomme « séries de Fourier ». Chaque terme de rang n de cette somme représente une fréquence f_n unique. Le spectre fréquentiel s'étend de la fréquence f_0, la plus basse après la composant continue, égale à l'inverse de la période :

$$f_0 = \frac{1}{T_p}, \tag{A.1}$$

jusqu'à $f = +\infty$. Les fréquences intermédiaires sont composées des harmoniques de la fréquence fondamentale, c'est-à-dire de tous les multiples de cette fréquence :

$$f_n = n.f_0 = \frac{n}{T_p} \tag{A.2}$$

Les termes des séries de Fourier forment un ensemble de composantes définies par leur amplitude et leur phase, pour des valeurs discrètes de la fréquence (équation A.2). Sur un analyseur de spectre, les amplitudes de ces termes forment un ensemble de pics. Une fonction v de t vérifiant les conditions ci-dessus est décomposée de la manière suivante, selon la formulation complexe :

$$v(t) = \sum_{n=-\infty}^{n=+\infty} A_n.e^{2\pi i f t} = \sum_{n=-\infty}^{n=+\infty} A_n.e^{i.\frac{2\pi n}{T_p}.t} \tag{A.3}$$

en exprimant la fréquence comme définie dans l'expression A.2. Le coefficient A_n est définit comme suit :

$$A_n = \frac{1}{T_p} \int_{-\frac{T_p}{2}}^{+\frac{T_p}{2}} v(t).e^{i.\frac{2\pi n}{T_p}.t} dt. \tag{A.4}$$

A.2 Transformation de Laplace

La décomposition en séries de Fourier est un outil puissant pour les signaux périodiques, mais ne peut pas s'appliquer aux signaux temporels uniques sans périodicité. La transformation de Laplace s'applique à des tels signaux, à la condition cependant qu'ils soient causals, c'est-à-dire nuls pour $t < 0$, comme on le voit dans l'équation A.5. Cependant, comme précédemment, ces signaux doivent être définis jusqu'à $t = +\infty$.

$$\mathcal{L}\left[v(t)\right](s) = \int_0^{\infty} v(t).e^{-st} dt \tag{A.5}$$

La transformation de Laplace convertit les données du domaine temporel à un domaine fréquentiel complexe. Le résultat de la transformation est une fonction continue, contrairement au résultat d'une décomposition en série de Fourier, qui fournit un ensemble discret de « pics ».

A.3 Transformation de Fourier complexe

La transformation de Fourier est une généralisation de la décomposition en séries de Fourier complexe pour une largeur de période tendant vers l'infini : $T_p \to \infty$, résolvant le problème posé par la non-périodicité des signaux en mesures impulsionnelles. En remplaçant A_n par une formulation continue $V(f)df$, la formule A.3 devient :

$$v(t) = \int_{-\infty}^{+\infty} V(f).e^{2\pi i f t} df \tag{A.6}$$

et

$$V(f) = \int_{-\infty}^{+\infty} v(t).e^{-2\pi i f t} dt. \tag{A.7}$$

L'opération définie par l'équation A.6 est appelée « transformation de Fourier inverse », et celle définie par l'équation A.7 « transformation de Fourier directe » ou « transformation de Fourier ». Du point de vue de l'électronicien, la transformation de Fourier est un intermédiaire entre les séries de Fourier et la transformation de Laplace. Si le signal dans le domaine temporel n'est pas un signal périodique s'étendant de $t = -T_p/2$ à $t = +T_p/2$, alors on peut effectuer une transformation de Fourier, en étendant ces limites à $-\infty$ et $+\infty$. Ici encore, le résultat dans le domaine fréquentiel est une fonction continue, et non un ensemble discret de « pics », comme on peut l'obtenir avec un signal périodique.

A.4 Énergie et densité spectrale d'énergie

L'énergie portée par un signal temporel se calcule en intégrant sa puissance instantanée $p(t)$ de $-\infty$ à $+\infty$:

$$E = \int_{-\infty}^{+\infty} p(t)dt \tag{A.8}$$

$$E = \frac{1}{R}.\int_{-\infty}^{+\infty} v^2(t)dt \tag{A.9}$$

avec

$$p(t) = \frac{v^2(t)}{R}, \tag{A.10}$$

si le signal est mesuré sur une résistance de charge R. D'après le théorème de Plancherel, le signal $v(t)$ et sa transformée de Fourier $[\mathcal{F}(v)](f)$ vérifient la relation :

$$\int_{-\infty}^{+\infty} |[\mathcal{F}(v)](f)|^2 df = \int_{-\infty}^{+\infty} |v(t)|^2 dt \tag{A.11}$$

ou encore :

$$\frac{1}{R}\int_{-\infty}^{+\infty} |[\mathcal{F}(v)](f)|^2 df = \frac{1}{R}\int_{-\infty}^{+\infty} v^2(t)dt \tag{A.12}$$

car v(t) est réel. On en déduit une formulation de l'énergie du signal dans le domaine fréquentiel :

$$\boxed{E = \frac{1}{R}\int_{-\infty}^{+\infty} |[\mathcal{F}(v)](f)|^2 df} \tag{A.13}$$

Du point de vue du physicien, le théorème de Plancherel n'exprime donc rien d'autre que la conservation de l'énergie lors de la transformation du domaine temporel au domaine fréquentiel

La grandeur :

$$\Phi(f) = \frac{|[\mathcal{F}(v)](f)|^2}{R}, \tag{A.14}$$

est appelée « densité spectrale d'énergie », et s'exprime en J/Hz.

Formulation discrète pour E et Φ :

La transformée de Fourier continue $\mathcal{F}(v)(f) = \int_{-\infty}^{+\infty} v(t).e^{-2\pi i f t}dt$ du signal $v(t)$ s'exprime en $V.s$, ou V/Hz. Or la DFT d'un signal échantillonné (équation 3.15) s'exprime elle en Volt, et ne constitue donc pas à proprement parler la formulation discrète de la transformée de Fourier continue.

Pour calculer la densité spectrale d'énergie d'un signal échantillonné à partir de sa DFT, une équivalence discrète de l'équation A.7 doit être définie :

$$V_n^{'} = \sum_{k=0}^{k=N-1} v_k.e^{-2\pi i.\frac{nk}{N}}.\Delta t = V_n.\Delta t \tag{A.15}$$

avec $V_n^{'}$ le terme de rang n, V_n la DFT définie par l'équation 3.15, et Δt le pas en temps du signal échantillonné. Cette grandeur s'exprime en V/Hz tout comme son équivalente continue. La densité spectrale d'énergie, définie de manière continue par l'équation A.14, peut alors s'exprimer selon une formulation discrète :

$$\Phi_n = \frac{|V'_n|^2}{R} \tag{A.16}$$

$$\Phi_n = \frac{|V_n.\Delta t|^2}{R} \tag{A.17}$$

Selon la même logique, l'énergie contenue dans le signal s'exprime alors selon une formulation discrète :

$$E = \sum_{n=0}^{n=N-1} \Phi_n.\Delta f \tag{A.18}$$

Soit :

$$\boxed{E = \sum_{n=0}^{n=N-1} \frac{|V_n|^2}{R}.\Delta t^2.\Delta f} \tag{A.19}$$

Annexe B

Fonctionnement d'un générateur d'impulsions $TLP/vf - TLP$

Les générateurs d'impulsion TLP comme $vf - TLP$ sont basés sur le même principe. Une ligne de transmission 50 Ω de longueur déterminée est chargée par une source de tension statique haute tension à travers une forte résistance. La fermeture d'un interrupteur connecte brutalement cette ligne au reste du système, et une décharge est alors envoyée à travers des câbles coaxiaux dans le système. Dans l'exemple de la figure B.1, avant la fermeture du relai REL_1, tous les points du cœur du câble coaxial TL_1 sont chargé au potentiel V et tous les points du cœur de TL_2 sont au potentiel zéro. La fermeture du relai REL_1 entraine une mise en contact de ces deux lignes. TL_1 commence alors à se décharger dans TL_2, et le potentiel se stabilise à $V/2$ de part et d'autre du relai. Du côté droit du relai (TL_2), cette brusque augmentation du potentiel gagne petit à petit la ligne de transmission, et une onde de tension positive, d'amplitude $V/2$ se propage vers la droite avec une vitesse de phase ν.

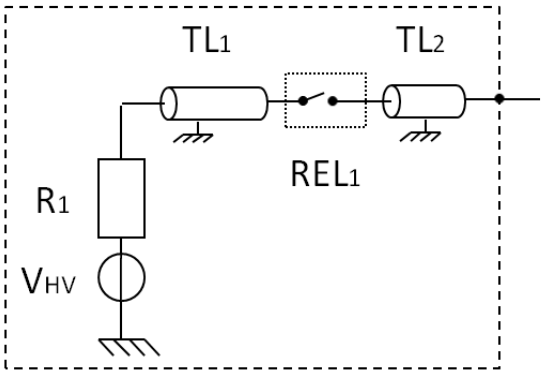

FIGURE B.1 – Schéma représentatif d'un générateur d'impulsion TLP/vf-TLP

Dans le même temps, un processus identique s'opère du côté gauche du relai (TL_1). Il s'agit

ici d'une diminution du potentiel, de V à $V/2$. Cette diminution de $V/2$ gagne progressivement tous les points du cœur de TL_1, de la droite vers la gauche, telle une onde de tension négative d'amplitude $-V/2$. Cette onde de tension se réfléchit sur la résistance R_1, d'une valeur typique de 10 $M\Omega$. Le coefficient de réflexion sur une telle résistance est très proche de 1, et l'onde se réfléchit telle quelle, superposant alors au potentiel $V/2$ précédemment établi, un potentiel $-V/2$. Le potentiel de TL_1 est alors petit à petit ramené à zéro. Cette onde d'amplitude $-V/2$ se propage de gauche à droite à travers TL_1, puis TL_2 et le reste du système, ramenant à zéro le potentiel sur tous les points qui avaient été chargés à $V/2$ par l'onde de tension. La durée de l'impulsion correspond donc à un aller-retour $REL_1 - R_1$.

L'onde de tension TLP générée est donc une impulsion rectangulaire, d'amplitude $V/2$, et d'une durée égale à

$$\tau = \frac{2L}{\nu} \tag{B.1}$$

avec L la longueur de la ligne, et ν la vitesse de propagation dans cette ligne. ν est égal à $0.66.c$ pour des câbles coaxiaux ayant une gaine en polyéthylène, comme ceux utilisés au laboratoire.

Le temps de montée de l'impulsion est inférieur à 100 ps. Il est fonction de la résistance de l'arc électrique généré lors de la fermeture du relai REL_1, et des inductances parasites ramenées par ce relai. Ce temps de montée peut être augmenté par l'ajout de filtre passe-bas à la sortie du générateur, comme expliqué plus haut.

Les largeurs d'impulsions sont fixées directement par la longueur du câble de charge, et on ne peut donc générer qu'un nombre limité de largeurs d'impulsions. Les valeurs standard pour la largeur d'impulsion sont respectivement 100 ns pour le TLP, 5 ns et 2.5 ns pour le vf-TLP.

Bibliographie

[1] JEDEC. Electrostatic discharge (ESD) sensitivity testing human body model (HBM), JESD22-A114D. Rapport technique, 2006.

[2] ESDA. Standard test method for electrostatic discharge sensitivity testing human body model (HBM) component level, ANSI/ESD STM5.1-2001. Rapport technique, 2001.

[3] Automotive Electronic Council. Human body model electrostatic discharge test AEC-Q100-002-REV-D. Rapport technique.

[4] Industry Council on ESD Target Levels. White paper 1 : A case for lowering component level HBM/MM ESD specifications and requirements. Rapport technique, 2007.

[5] JEDEC. Electrostatic discharge (ESD) sensitivity testing machine model (MM), JESD22-A115C. Rapport technique, 2010.

[6] ESDA. Standard test method for the electrostatic discharge sensitivity testing machine model - component level, ESD STM5.2-1999. Rapport technique, 2001.

[7] Automotive Electronic Council. Machine model (MM) electrostatic discharge test, AEC-Q100-003 -Rev-E. Rapport technique, 1998.

[8] Hiroyasu Ishizuka, Kousuke Okuyama, Katsuhiko Kubota, Masamichi Komuro, and Yuji Hara. A study of ESD protection devices for input pins discharge characteristics of diode, lateral bipolar transistor, and thyristor under MM and HBM tests. *IEEE Transactions on Components, Packaging and Manufacturing Technology, Part C*, VOL. 21, no 4, octobre 1998.

[9] JEDEC. Field-induced charged-device model test method for electrostatic-discharge-withstand thresholds of microelectronic components, JESD22-C101. Rapport technique, 1995.

[10] ESDA. Standard test method for electrostatic discharge sensitivity testing -charged device model (CDM) component level, ANSI ESD STM5.3.1 :1999. Rapport technique, 1999.

[11] Automotive Electronic Council. Charged device model (CDM) electrostatic discharge test, AEC-Q100-011 -Rev-B. Rapport technique, 1999.

[12] Leo G. Henry, Jon Barth, Hugh Hyatt, Tom Diep, and Michael Stevens. Charged device model metrology : limitations and problems. In *Microelectronics Reliability 42*, 2002.

[13] Ajith Amerasekera and Charvaka Duvvury. *ESD in Silicon Integrated Circuits*. John Wiley & Sons, LTD, 2002.

[14] IEC. Electromagnetic compatibility (EMC) - part 4-2 : Testing and measurement techniques - electrostatic discharge immunity test. Rapport technique, 2001.

[15] ISO. Road vehicles - test methods for electrical disturbances from electrostatic discharges. Rapport technique, 2001.

[16] H. Tanaka, O. Fujiwara, and Y. Yamanaka. A circuit approach to simulate discharge current injected in contact with an ESD-gun. In *3rd Symposium on Electromagnetic Compatibility*, 2002.

[17] S. Caniggia and F. Maradei. Circuit and numerical modeling of electrostatic discharge generators. *IEEE Transactions on Industry Application*, 2006.

[18] ANSI/ESD. Sp5.6 - electrostatic discharge sensitivity testing - human metal model (HMM) - component level. Rapport technique, 2009.

[19] Kathleen Muhonen, Nate Peachey, and Al Testin. Human metal model (HMM) testing, challenges to using ESD guns. In *EOS/ESD*, 2009.

[20] Yiqun Cao, David Johnsson, Bastian Arndt, and Matthias Stecher. A TLP-based human metal model ESD-generator for device qualification according to IEC 61000-4-2. In *Asia-Pacific International Symposium on Electromagnetic Compatibility*, 2010.

[21] Tim Maloney and N. Khurana. Transmission line pulsing techniques for circuit modeling of ESD phenomena. In *EOS/ESD Symposium Proceedings*, pages 140–146, 1985.

[22] H. Gieser and M. Haunschild. Very-fast transmission line pulsing of integrated structures and the charge device model. In *EOS/ESD Symposium Proceedings*, pages 85–94, 1996.

[23] Horst Gieser and Markus Haunschild. Very-fast transmission line pulsing of integrated structures and the charged device model. In *EOS/ESD*, 1996.

[24] Heinrich Wolf, Horst Gieser, Wolfgang Stadler, and Wolfgang Wilkening. Capacitively coupled transmission line pulsing CC-TLP - a traceable and reproducible stress method in the CDM-domain. *EOS/ESD*, 2003.

[25] Heinrich Wolf, Horst Gieser, and Dirk Walter. Investigating the CDM susceptibility of IC's at package and wafer level by capacitive coupled TLP. *EOS/ESD*, 2007.

[26] Heinrich Wolf, Horst Gieser, Karlheinz Bock, Agha Jahanzeb, Charvaka Duvvury, and Yen-Yi Lin. Capacitive coupled TLP (CC-TLP) and the correlation with the CDM. *EOS/ESD*, 2009.

[27] C. Russ. *ESD protection devices for CMOS technologies : Processing, impact, modeling and testing issues.* Thèse de doctorat, Technical University, Munich, 1999.

[28] Géraldine Bertrand. *Conception et modélisation électrique de structures de protection contre les décharges électrostatiques en technologies BiCMOS et CMOS analogique.* Thèse de doctorat, Institut National des Sciences Appliquées de Toulouse, 2001.

[29] David Trémouilles. *Optimisation et modélisation de protections intégrées contre les décharges électrostatiques, par l'analyse de la physique mise en jeu.* Thèse de doctorat, Institut National des Sciences Appliquées de Toulouse, 2004.

[30] Patrice Besse. *Tenue en énergie de structures LDMOS avancées de puissance intégrées dans les domaines temporels de la nanoseconde à la milliseconde.* Thèse de doctorat, Université Paul Sabatier de Toulouse, 2004.

[31] Amaury Gendron. *Structures de protection innovantes contre les décharges électrostatiques, dédiées aux Entrées/Sorties hautes tensions de technologies SmartPower.* Thèse de doctorat, Université Paul Sabatier de Toulouse, 2007.

[32] Gianluca Boselli. *On High Injection Mechanisms in Semiconductor Devices under ESD Conditions.* Thèse de doctorat, Universiteit Twente, 2001.

[33] Agnès Guilhaume. *Evaluation de la robustesse de circuits intégrés vis-à-vis des décharges électrostatiques.* Thèse de doctorat, Institut National des Sciences Appliquées de Lyon, 2002.

[34] V. De Heyn, G. Groeseneken, B. Keppens, M. Natarajan, L. Vacaresse, and G. Gallopyn. Design and analysis of new protection structures for smart power technology with controlled trigger and holding voltage. In *IRPS*, 2001.

[35] Chung-Yu Wu, Ming-Dou Ker, Chung-Yuan Lee, and Joe KO. A new on-chip ESD protection circuit with dual parasitic SCR structures for CMOS VLSL. *IEEE Journal of Solid-State Circuits*, pp. 274-280, 1992.

[36] Ker Ming-Dou and K-C Hsu. Overview of on-chip electrostatic discharge protection design with SCR-based devices in CMOS integrated circuits. *IEEE transactions on Device Reliability*, 2005.

[37] James Di Sarro, Kiran Chatty, Robert Gauthier, and Elyse Rosenbaum. Sudy of design factors affecting turn-on time of silicon controlled rectifiers (SCRs) in 90 nm bulk CMOS. In *44th Annual International Reliability Physics Symposium, San Jose*, page 163, 2006.

[38] Johan Bourgeat, Christophe Entringer, Philippe Galy, Pascal Fonteneau, and Marise Bafleur. Local ESD protection structure based on silicon controlled rectifier achieving very low overshoot voltage. In *EOS/ESD Symposium*, 2009.

[39] Amaury Gendron, Chai Gill, Carol Zhan, Mike Kaneshiro, Bill Cowden, Changsoo Hong, Richard Ida, and Dung Nguyen. New high voltage ESD protection devices based on bipolar transistors for automotive applications. In *EOS/ESD Symposium*, 2011.

[40] D. Trémouilles, Y. Gao, and M. Bafleur. Pushing away the silicon limits of ESD protection structures : exploration of crystallographic orientation 2008. In *IEEE Bipolar / BiCMOS Circuits and Technology Meeting (BCTM 2008), Monterey, CA (USA), October 13-16,*, 2008.

[41] R. Ida, M. Majerus, and C. Gill. A method for characterizing the DPI immunity of an ESD clamp. In *IEW*, pages pages 261–272, 2011.

[42] N. Jensen, G. Groos, M. Denison, J. Kuzmik, D. Poganyand E. Gornik, and M. Stecher. Coupled bipolar transistors as very robust ESD protection devices for automotive applications. In *EOS/ESD*, 2003.

[43] V. Vashchenko, V. Kuznetsov, and P. Hopper. Implementation of dual-direction SCR devices in analog CMOS process. In *EOS/ESD*, 2007.

[44] Z. Liu, J. Liou, and J. Vinson. Novel silicon-controlled rectifier (SCR) for high-voltage electrostatic discharge (ESD) applications. *Electron Device Letters*, volume 29 :pages 753–755, 2008.

[45] Rouying Zhan, Chai Ean Gill, James D Whitfield, and Hongzhong Xu. Resistor triggered electrostatic discharge protection. Freescale Semiconducor, 2009.

[46] Hongzhong Xu, Chai Ean Gill, James D Whitfield, and Jinman Yang. Complementary zener triggered bipolar ESD protection. Freescale Semiconducor, April 2010.

[47] Vadim A. Kushner, Amaury Gendron, and Chai Ean E. Gill. ESD protection structure. Freescale Semiconducor, May 2009.

[48] Antoine Delmas, Amaury Gendron, Marise Bafleur, Nicolas Nolhier, and Chai Gill. Transient voltage overshoots of high voltage ESD protections based on bipolar transistors in smart power technology. In *IEEE Bipolar/BiCMOS Circuits and Technology Meeting (BCTM)*, 2010.

[49] Philippe Renaud, Patrice Besse, and Amaury Gendron. Semiconductor device structure and integrated circuit therefor. Freescale Semiconducor, March 2006.

[50] Antoine Delmas. Etude des phénomènes physiques transitoires en régime de très forte injection dans les semi-conducteurs appliquée aux protections contre les décharges électrostatiques (ESD). Projet de fin d'études, INSA Rennes, 2007.

[51] Jean-Robert Manouvrier. *Comportement des diodes de protection lors d'évènements transitoires rapides dûs aux ESD : outils de caractérisation, physique du semi-conducteur et modélisation.* Thèse de doctorat, Université Montpellier II, 2008.

[52] Heinrich Wolf, Horst Gieser, Wolfgang Stadler, Wolfgang Wilkening, Petra Rose, and Ning Qu. Transient analysis of ESD protection elements by Time Domain Transmission using repetitive pulses. In *EOS/ESD Symposium Proceedings*, pages 304–310, 2006.

[53] Evan Grund. Deriving the DUT current and voltage waveforms by merging vf-TLP incident and reflected signals. In *EOS/ESD/EMI Worshop Proceedings, Toulouse France*, mai 2006.

[54] Picosecond Pulse Labs. *Model 5331 6 dB Power Divider Product Specification.*

[55] D. Trémouilles, S. Thijs, Ph. Roussel, M.I. Natarajana, V. Vassilev, and G. Groeseneken. Transient voltage overshoot in TLP testing : Real or artifact ? In *Microelectronics Reliability*, volume 47, Issue 7, Special Issue : EOS/ESD Symposium 2005, pages 1016–1024, Juillet 2007.

[56] Antoine Delmas, David Trémouilles, Nicolas Nolhier, Marise Bafleur, Nicolas Mauran, and Amaury Gendron. Accurate transient behavior measurement of high-voltage ESD protections based on a very fast transmission-line pulse system. In *EOS/ESD Symposium Proceedings*, 2009.

[57] J. W. Cooley and J. W. Tukey. An algorithm for the machine calculation of complex Fourier series. *Mathematics of Computation*, vol 19 :pp 297–301, 1965.

[58] Geert Wybo, Stefaan Verleye, Benjamin Van Camp, and Olivier Marichal. Characterizing the transient device behavior of SCRs by means of VFTLP waveform analysis. In *EOS/ESD Symposium Proceedings*, pages 366–375, 2007.

[59] Sedki M. Riad. The deconvolution problem : an overview. *Proceedings of the IEEE*, 74 :82–85, Janvier 1986.

[60] Abdelhak Bennia and Norris S. Nahman. Deconvolution of causal pulse and transient data. *IEEE Transactions on Instrumentation and Measurement*, VOL 79. NO 6 :933–939, Décembre 1990.

[61] Bidyut Parruck. Study and performance evaluation of two iterative frequency-domain deconvolution techniques. *IEEE Transactions on Instrumentation and Measurement*, IM-33, NO. 4 :281–287, 1984.

[62] Nicolas Mauran. Conception et réalisation d'un banc de caractérisation sous pointes pour mesures impulsionnelles haute énergie. Mémoire cnam, Conservatoire National des Arts et Métiers, 2003.

[63] H.R. Kaupp. Characteristics of microstrip transmission lines. *IEEE Transactions*, EC-16, no. 6 :pp. 241–276, Juin 1967.

[64] K. Esmark. *Device Simulation of ESD Protection Elements.* Thèse de doctorat, Swiss federal Institute of Technology, 2OO2.

[65] Juin J. Liou, Slavica Malobabic, David F. Ellis, Javier A. Salcedo, Jean-Jacques Hajjar, and Yuanzhong Zhou. Transient Safe Operating Area (TSOA) definition for ESD applications. In *EOS/ESD Symposium*, 2009.

[66] Markus Mergens, Wolfgang Wilkening, Stephan Mettler, Heinrich Wolf, Andreas Stricker, and Wolfgang Fichtner. Analysis and compact modeling of lateral DMOS power devices under ESD stress conditions. In *EOS/ESD Symposium*, 1999.

[67] J. Willemen, A. Andreini, V. De Heyn, K. Esmark, M. Etherton, H. Gieser, G. Groeseneken, S. Mettler, E. Morena, W. Soppa N. Qu, W. Stadler, R. Stella, W. Wilkening, H. Wolf, and L. Zullino. Characterization and modeling of transient device behavior under CDM ESD stress. *Journal of Electrostatics*, 62 :pp 133–153, 2004.

[68] R. VanOverstraeten and H. DeMan. Measurement of the ionization rates in diffused silicon P-N junctions. *Solid-State Electron.*, vol. 13 :pp 583–60, 1970.

[69] W. Fulop. Calculation of avalanche breakdown of silicon P-N junctions. *Solid-State Electronics*, vol. 10 :pp 39–43, 1967.

[70] G. Massobrio and P. Antognetti. *Semiconductor Device Modeling with SPICE*. Seconde Edition, 1993.

[71] C. Russ, Koen Verhaege, Karlheinz Bock, Philippe J. Roussel, Guido Groeseneken, and Herman E. Maes. A compact model for the grounded gate nMOS behaviour under CDM ESD. In *Proceedings of the EOS/ESD Symposium*, 1996.

[72] Heinrich Wolf, Horst Gieser, and Wolfgang Wilkening. Analyzing the switching behavior of ESD protection transistors by very fast transmission line pulsing. In *EOS/ESD Symposium*, 1999.

[73] B. El-Kareh and R. J. Bombard. *Introduction to VLSI Silicon Devices*. Kluwer Academic Publishers, 1996.

Liste des publications

Conférences internationales

Antoine Delmas, David Trémouilles, Nicolas Nolhier, Marise Bafleur, Nicolas Mauran, Amaury Gendron : Accurate transient behavior measurement of High-Voltage ESD Protections Based on a Very Fast Transmission-Line Pulse system, *EOS/ESD Symposium Proceedings*, 2009.

Antoine Delmas, Amaury Gendron, Maryse Bafleur, Nicolas Nolhier, Chai Gill : Transient voltage overshoots of high voltage ESD protections based on bipolar transistors in smart power technology, *IEEE Bipolar/BiCMOS Circuits and Technology Meeting (BCTM)*, 2010

Conférences nationales

Antoine Delmas, David Trémouilles, Nicolas Nolhier : Méthode précise de mesure du comportement transitoire de protections ESD haute tension, basée sur un système vf-TLP, *Poster, journée GEET*, mars 2010.

Résumé

Résumé

Les travaux présentés dans ce mémoire visent à analyser et optimiser le comportement des composants de protection haute tension contre les décharges électrostatiques (ESD) à leur déclenchement. Pour cela, deux approches ont été suivies :

Un outil de mesure dédié, le «transient-TLP », a été développé. Cet outil est basé sur la correction mathématique des données mesurées à l'oscilloscope avec un système de mesure vf-TLP standard. L'erreur de mesure est inférieure à 2 %. La méthode, d'abord conçue pour des mesures sur wafer, a ensuite été appliquée pour mesurer des composants sur boîtier.

A l'aide de cet outil, le comportement transitoire des protections ESD utilisées à *Freescale* a pu être analysé. En particulier, la simulation physique a permis de mieux comprendre l'origine physique de l'apparition d'un pic de surtension au déclenchement de ces composants et des solutions de dessin ont été proposées pour en réduire l'amplitude.

Mots clés : ESD, TLP, vf − TLP, surtension, transitoire, instrumentation, TCAD.

Abstract

The research work presented in this thesis is aimed to analyze and optimize the triggering behavior of high voltage devices against electrostatic discharges (ESD). Two approaches were used :

A dedicated tool was created. The « transient-TLP » tool is based on mathematical post-processing of oscilloscope raw data, measured using a standard vf-TLP system. Measurement error is less than 2 %. This method, first designed for on-wafer measurement were then adapted to packaged device measurements.

The transient behavior of ESD devices used at *Freescale* was analyzed using this tool. TCAD simulations were then performed to analyze the turning-on of high voltage ESD devices. The physical origins of the strong voltage overshoot peak observed at the triggering of this component was explained, and design solutions were proposed to significantly reduce the magnitude of this peak.

Keywords : ESD, TLP, vf − TLP, overshoot, transient, instrumentation, TCAD.

Table des figures

1.1 Stratégie d'intégration de protections ESD localisées protégeant l'entrée d'un CI. 2

1.2 Caractéristique I-V TLP typique d'une protection ESD et paramètres électriques associés. Les flèches noires représentent les marges de sûreté. 4

1.3 Modèle électrique équivalent du corps humain (a) et décharge simulée sur un court-circuit pour une tension de charge de 2 kV (b). 5

1.4 Modèle électrique équivalent d'une machine (a) et forme d'onde simulée sur un court-circuit pour une tension de charge de 200 V (b). 6

1.5 Modèle électrique équivalent du composant chargé (a) et forme d'onde obtenue à 50 V de tension de charge. 7

1.6 Comparatif des formes d'ondes des trois modèles de décharge abordés dans cette section : HBM, MM, CDM. 7

1.7 Modèle électrique simplifié d'un générateur Gun. Seules $C1$ et $R2$ sont spécifiés dans les normes. $R1$ est la résistance de charge, typiquement de 1 $M\Omega$. $C2$ est une capacité parasite de l'ordre de 10 pF, reproduisant le couplage existant entre la pointe de décharge et le plan de masse. 8

1.8 Forme d'onde d'une décharge pistolet à 8 kV selon la norme IEC 61000-4-2. . . 9

1.9 Construction point par point d'une caractéristique $I-V$ TLP. 11

1.10 Schéma de principe de la caractérisation par réflectométrie. Exemple du vf-TLP, dans lequel les impulsions incidentes et réfléchies apparaissent séparées sur l'écran de l'ocilloscope. 13

1.11 Schéma représentatif d'une mesure TLP sur transistor bipolaire autopolarisé NPN (a). Courbe caractéristique $I-V$ correspondante (b). On distingue les trois zones de fonctionnement : état bloqué (1), avalanche de la diode base-collecteur (2), et fonctionnement replié (3). 15

1.12 Coupe technologique $2D$ simplifiée d'un transistor bipolaire latéral. 17

1.13 Coupe technologique $2D$ simplifiée d'un transistor bipolaire vertical. 18

1.14 Schéma électrique représentatif d'un SCR (a), composé d'un NPN et d'un PNP imbriqués. La structure $PNPN$ correspondante est représentée à la figure (b). Les résistances R_1 et R_2 sont les résistances intrinsèques des bases des deux transistors. 19
1.15 Coupe technologique simplifiée d'un transistor $NMOS$ utilisé comme protection ESD, montrant le NPN parasite. 20
1.16 Coupe technologique représentative d'un composant de protection $NPNR$ (a) et schéma équivalent (b). 22
1.17 Coupe technologique représentative d'un composant de protection $NPNZ$ (a) et schéma équivalent (b). 23
1.18 Coupe technologique représentative d'un composant de protection $NPNB$. . . . 24
1.19 Coupe technologique représentative d'un composant de protection $PNP/Diode$. 25

2.1 Principe de fonctionnement du banc de mesure vf-TLP Oryx. Les lignes de transmission (« Transmission Line ») TL_1, TL_2, TL_3 et TL_4 sont représentées, ainsi que le filtre de temps de montée. Le rôle de la « merge box » sera détaillé plus loin. 29
2.2 La boîte de mesure contient un relai pour commuter entre la mesure TLP et le traceur I-V. Les sondes de courant et de tension sont placées en série. 29
2.3 Schéma équivalent de la sonde en tension. 30
2.4 Architecture d'un canal de la merge box. Les deux canaux sont identiques. 31
2.5 Densités spectrales d'énergie d'impulsions ESD de type vf-TLP de 5 ns de largeur d'impulsion. Résultats obtenus à partir de FFT du signal mesuré à l'oscilloscope. 33
2.6 Mesure de la transmission fréquentielle des sondes en courant et en tension utilisées. La sonde en courant a une fréquence de coupure de 2 GHz, et la sonde en tension de 3 GHz environ. 34
2.7 Caractéristique de transmission fréquentielle des câbles coaxiaux utilisés dans le vf-TLP Oryx. 35
2.8 Mesure de pertes fréquentielles du système de mesure vf-TLP, depuis le générateur jusqu'à l'oscilloscope. La transmission le long du chemin « incident » est représentée en bleu foncé, et le long du chemin « réfléchi » en orange clair. 37
2.9 Formes d'onde en tension et courant mesurées à l'oscilloscope. Le filtrage du chemin de mesure prend la forme d'une augmentation du temps de montée ainsi qu'une atténuation du pic de surtension. 38
2.10 Réglage du délai de la merge box. En bleu, l'impulsion incidente, et en vert l'impulsion réfléchie. La superposition des deux impulsions se limite à deux pics que l'utilisateur tente d'annuler. 39
2.11 Principe de la mesure « indirecte » du pic de surtension. 41
2.12 Test en conditions de l'efficacité de protection du $PNP/Diode$ pour une entrée 50 V. 42

2.13 Calcul de la surtension générée par le PNP/Diode à l'aide d'une mesure TLP indirecte. La largeur d'impulsion est de 100 ns, le temps de montée de 300 ps, et la résistance de grille est de 10 $k\Omega$. La caractéristique TLP de la structure couplée protection//NLDMOS est représentée en vert, et la comparaison avec une protection seule en bleu. La superposition des deux courbes montre que le NLDMOS ne se déclenche pas avant d'atteindre sa tension de dégradation. 43

3.1 Banc de mesure simplifié utilisé pour les mesures $transient - TLP$. 46

3.2 Principe de la méthode de mesure $transient - TLP$: mesure des trois paramètres de transmission S_{31}, S_{32} et S_{21}. 47

3.3 Impulsions $vf - TLP$ mesurées à l'oscilloscope avec le dispositif de la figure 3.2. La tension résiduelle présente à la suite de l'impulsion incidente se superpose partiellement avec l'impulsion réfléchie. 56

3.4 Exemple d'acquisitions successives sur une charge 50 Ω (acquisition 1) et sur une protection ESD haute tension à 40 V TLP (acquisition 2), après correction des offsets . On construit ainsi le fichier $V_{inc,oscillo}$, qui sera utilisé par la suite. L'avantage de cette opération est de faire en sorte que les signaux incidents et réfléchis utilisés dans les calculs soient issus de la même impulsion d'origine. Ils sont alors liés par une transformation mathématique. Les valeurs non nulles dans l'acquisition 2 après l'impulsion réfléchie sont le résultat de réflexions parasites sur le générateur. On note qu'à ce stade, les amplitudes des impulsions n'ont pas encore été corrigées. 58

3.5 Construction du fichier de points $V_{ref,oscillo.}$, défini comme étant la différence entre la mesure effectuée sur le DUT et le fichier de points précédemment construit $V_{inc,oscillo.}$. 58

3.6 Exemple de fonctions de fenêtrage (a) et leurs spectres fréquentiels respectifs (b). 61

3.7 Opération de fenêtrage sur le fichier de points $V_{ref,oscillo}$ précédemment construit. En noir, l'impulsion d'origine avant le fenêtrage, en rouge pointillé, la fenêtre de Tukey employée et en orange le résultat de la multiplication des deux. 62

3.8 Résultat du calcul de $V_{DUT}(t)$ (en bleu foncé) et $I_{DUT}(t)$ (en orange clair) sur une protection ESD haute tension pour 100 V $vf - TLP$, avec 1 ns de temps de montée. Les deux impulsions débutent au même moment, ce qui indique que les délais de propagation dans le système de mesure sont correctement corrigés. . . . 64

3.9 Mesure des impulsions incidentes (bleu) et réfléchies (orange pointillés) sur un circuit ouvert (pointes RF laissées en l'air) à 200 V TLP, 5 ns de durée d'impulsion et 300 ps de temps de montée. La différence entre les deux est tracée en vert pâle sur l'axe de droite. 65

3.10 Calcul des impulsions incidentes et réfléchies sur un circuit ouvert parfait. Dans ce cas, les deux formes d'onde doivent être identiques et parfaitement alignées dans le temps. 65

3.11 Impulsions mesurées sur un court circuit à 200 V TLP, $t_r = 1$ ns, 5 ns de durée d'impulsion. L'impulsion réfléchie (orange pointillés) a été multipliées par -1 pour comparaison avec l'incident (bleu). La somme de l'incident et du réfléchi est tracée en vert pâle, et sa moyenne est reportée en trait noir continu. 66

3.12 Valeur absolue des caractéristiques quasi-statiques obtenues avec l'outil de mesure $transient - TLP$ sur un court-circuit (a) et un circuit ouvert (b). 67

3.13 Mesure $transient - TLP$ sur un $NPNZ$ dédié aux E/S 20 V pour des impulsions de 5 ns et 250 V de tension de charge avec trois temps de montée différents. La tension est reportée figure a) et le courant figure b). 67

3.14 Courbes $I - V_{max}$ obtenues par caractérisation $transient - TLP$ sur des $NPNB$ dédiés aux E/S 20 V avec des temps de montée de 1 ns et 10 ns. 68

3.15 Mesure de formes d'onde en tension obtenues avec la méthode décrite ($V_{corrigé}$), et comparaison avec l'outil standard $vf - TLP$ Oryx ($V_{mesuré1}$ et $V_{mesuré2}$). Un délai de 150 ps a été introduit entre les deux mesures $vf - TLP$ dans le réglage de la merge box pour visualiser son effet sur le pic de tension. 69

3.16 Dispositif de test de la répartition en puissance des deux sorties des tés de pick-off. L'atténuateur de 30 dB en série avec l'entrée 2 de l'oscilloscope est nécessaire pour rester dans la plage de tension spécifiée. 70

3.17 Mesure de la linéarité en puissance entre 1 V et 180 V, de 340 MHz à 3 GHz sur le système utilisé le plus fréquemment. Les mesures sont effectuées dans le domaine temporel (impulsions $vf - TLP$), puis transférées dans le domaine fréquentiel par FFT. 71

3.18 Plans de propagation 1, 2 et 3 mis en jeu pour l'application de la méthode d'étalonnage « transient-TLP ». Le plan 4 coïncide avec l'extrémité du câble coaxial TL_2. Le plan 2 ne peut pas être connecté à l'entrée de l'oscilloscope, mais une méthode alternative est proposée en en utilisant le plan 4 à la place du plan 2. . 73

3.19 Procédure de mesure du paramètre S'_{inc}. 74

3.20 Procédure de mesure du paramètre S'_{ref}. Le set-up s'apparente à celui d'une mesure $transient - TLP$ sur un circuit ouvert parfait. Le câble coaxial TL_3 est relié comme précédemment à la voie 1 de l'oscilloscope, et le câble TL_2 aux pointes RF, ces dernières étant laissées en l'air. 75

4.1 Système utilisé en caractérisations TLP pour relier le câble coaxial d'acheminement de l'impulsion TLP au composant sous test. Les pinces sont constituées de fils reliés l'un à l'âme et l'autre à la gaine du connecteur coaxial, et terminés par des pinces « grip-fil ». 80
4.2 Schéma global du système de mesure TLP utilisé à Freescale [62]. 82
4.3 Pinces fabriquées pour remplacer les pinces d'origine, trop inductives et instables. 82
4.4 Schéma équivalent représentant les plans de propagation retenus pour décomposer le système différents segments, chacun des segment pouvant être caractérisé indépendamment des autres. 83
4.5 Représentation schématique des pinces. Les deux étapes d'étalonnage permettent de calculer les tensions et courants en $P5$ à partir des tensions incidentes en $P2$. Dans le cas où les pinces sont connectées à une ligne de transmission, comme dans la figure 4.4, on peut calculer les tensions incidentes et réfléchies dans la ligne en $P5$. 85
4.6 Etapes d'étalonnage des pinces. Mesure du coefficient de transmission du câble coaxial (a) et de l'inductance Z_{pinces} des fils (b) 86
4.7 Paramètres géométriques et physiques des lignes de transmission microruban permettant le calcul de leur impédance caractéristique. 90
4.8 Bilan des étapes de mesure permettant l'étalonnage du banc TLP. Les cinq étapes de mesure présentées permettent d'apporter les équations nécessaires à la détermination des cinq inconnues que sont les paramètres S_{inc}, S_{ref}, τ, Z_{pinces} et γl. Le cas de l'impédance caractéristique Z_{C2} des pistes microruban est traité à part, celle-ci étant extraite des paramètres de layout de la carte. On recalcule ensuite les tensions et courants dans les plans successifs en utilisant chacun de ces paramètres. 95
4.9 Mesure de la tension et du courant sur un circuit ouvert (a) et sur un court-circuit (b) en bout des pinces (plan 5) pour une impulsion de 10 V TLP. Le courant de circuit ouvert et la tension de court-circuit servent à calculer le pourcentage d'erreur dans chacun des cas. Celui-ci est de l'ordre de 2.5 % et atteint sa valeur maximale pendant le front montant de l'impulsion. 96
4.10 Tension incidentes et réfléchies calculées dans le plan du composant lors de mesures sur un circuit ouvert (a) et fermé (b) dans le plan 6. La tension de précharge est ici de 10 V TLP. Dans le cas du court-circuit, la tension réfléchie a été inversée afin de mieux visualiser sa superposition avec la tension incidente. Les erreurs correspondantes sont calculées au moyen des formules 4.50 et 4.51 97
4.11 Mesure de la tension (trait plein) et du courant (pointillés) sur des composants de protection $NPNB$ $Pad20$ V, de deux longueurs d'émetteur différentes, à 300 V TLP (6 A). 98

4.12 Comparaison entre les courbes $I-V$ obtenues par des mesures TLP standard (courbes $I_{TLP}-V_{TLP}$) et les courbes $I-Vmax$ des $NPNB$ $Pad40$ des deux dimensions différentes. 98

5.1 Équivalence entre le modèle du banc de caractérisation TLP (a) et le modèle utilisé en simulation (b). 103

5.2 Comparaison entre les mesures $transient-TLP$ (trait plein) et les simulations $TCAD$ (tirets). Figure a) : mesure et simulations sur le composant pour une impulsion de 5 ns de longueur, avec 300 ps de temps de montée (bleu foncé) et 1 ns (trait vert clair). Figure b) : mesures et simulations sur une impulsion de 100 ns de long avec 10 ns de temps de montée. Figures c) et d) : agrandissements des figures a) et b) autour du pic de tension. 105

5.3 Profil $2D$ du $LDMOS$ à protéger (a), et configuration dans laquelle sont placées la protection $NPNB$ et le $LDMOS$ (b). 107

5.4 Caractéristique $I-V$ TLP du $LDMOS$ à protéger, avec une résistance grille-source nulle et un temps de montée de 10 ns. En bleu foncé sont représentés les points $I-V$, et en vert clair le courant de fuite entre chaque mesure. 107

5.5 Densité de courant (a) et champ électrique (b) simulés dans la région de drain après repliement de la tension. On note l'apparition d'une zone de fort champ électrique à la jonction N^+/N^- sous le contact de drain. 108

5.6 Formes d'ondes utilisées pour tester la robustesse du $LDMOS$ sous différentes largeurs d'impulsion. Pour les impulsions à 1.25 et 2.5 ns, on peut voir les paliers correspondant aux réflexions multiples entre la résistance R_1 (figure B.1) et la résistance de 50 Ω. En bleu foncé est tracée l'impulsion incidente à 1.25 ns, en vert clair celle à 2.5 ns et en rouge pointillés celle à 5 ns. 110

5.7 Courbes $I-V$ statiques mesurées avant l'agression ESD (en noir), après l'impulsion de 5 ns (en trait plein vert clair), après celle de 2.5 ns (rouge pointillés) et après l'impulsion de 1.25 ns (tirets oranges). Une limitation de 1 μA a été appliquée lors de la mesure. 110

5.8 Profil du composant $NPNB$ et schéma équivalent. 111

5.9 Caractérisation TLP du composant (courbe bleu foncé) et $transient-TLP$ avec 10 ns et 1 ns de temps de montée (courbes vert pâle et violette). Le courant de fuite est représenté en rouge, permettant d'identifier le courant I_{t2} à 6.9 A. . . . 113

5.10 Concentrations en atomes donneurs (bleu foncé), accepteurs (vert pâle), dopage total (noir) et profil de champ électrique à bas courant (rouge pointillés) pour la diode latérale (a) et la diode verticale (b) formant les deux jonctions base-collecteur du $NPNB$. 114

5.11 Conditions de simulation employées pour étudier la dynamique du processus de claquage par avalanche. Une rampe de tension est appliquée à travers une charge de 50 Ω avec un temps de montée de 100 ns jusqu'à la tension de claquage, puis la tension est maintenue constante. 116

5.12 Coefficients d'ionisation des électrons (bleu foncé) et des trous (vert clair) à travers la diode latérale (a) et verticale (b). 117

5.13 Résultats de l'expérience décrite en figure 5.11 pour la diode latérale (bleu foncé) et verticale (vert clair). En a), les tensions appliquées, en b) les courants circulant dans chacune des diodes en échelle logarithmique. On constate une pente plus importante pour la jonction latérale que verticale, suivie d'un plateau correspondant à un régime permanent. 117

5.14 Formes d'onde simulées en tension (bleu foncé) et courant (vert clair) du composant sous une rampe TLP de 10 ns de temps de montée. 122

5.15 Valeur maximale du taux de génération par avalanche dans la zone d'avalanche de la jonction base-collecteur latérale en fonction du temps. On distingue deux régimes, l'un correspondant à un régime transitoire, d'une durée de 100 ps environ, suivi d'une stabilisation à une valeur constante (régime permanent). 123

5.16 Intégrales d'ionisation des électrons (bleu foncé) et des trous (vert clair) calculés le long de la jonction base-collecteur latérale en fonction du temps. Le courant d'avalanche est tracé en rouge pointillés. 124

5.17 Courants d'avalanche (courant de trous) relevés dans la base en deux points : l'un, $B1$ est situé près de la ZCE, et l'autre, $B2$, jouxte la zone de transition de la jonction base-émetteur. Les deux courants sont égaux depuis le claquage à 2.6 ns jusqu'au déclenchement en direct de la jonction émetteur-base à 2.95 ns. . . . 125

5.18 Schéma équivalent du NPN latéral entre t_{BV} et t_{EB}. L'établissement de la chute de potentiel aux borne de R_{B2} contrôle le déclenchement en direct de la jonction émetteur-base. Le composant est polarisé suivant les conditions de la figure 5.8 : émetteur et base court-circuités, et collecteur flottant. L'impulsion est appliquée côté collecteur. 126

5.19 Modèle simplifié du NPN avant le déclenchement bipolaire. La capacité de jonction C_J de la figure 5.18 a été supprimée. 126

5.20 Courant dans la résistance de base R_{B1} (noir pointillés) et potentiel dans la base au point $B2$ (trait continu vert clair). Le rapport de proportionnalité entre les deux grandeurs valide l'hypothèse selon laquelle la capacité de jonction émetteur-base est négligeable. 127

5.21 Modulation du profil de champ électrique dans la base du NPN vertical. Le zone de fort champ électrique adopte peu à peu le profil d'une jonction abrupte. . . . 129

5.22 Évolution des intégrales d'ionisation des électrons (bleu foncé) et des trous (vert clair) sous l'action de la forte injection. Le maximum est atteint entre 5 et 6 ns environ. 130

5.23 Courant du NPN latéral (vert), vertical (rouge pointillé) et courant total (trait plein bleu). 131

5.24 Évolution des temps de transit des électrons dans la base latérale (bleu foncé) et verticale (vert clair). Pour la base latérale, le calcul a été fait le long de lignes de courant rejoignant la zone d'avalanche depuis l'extrémité droite de l'émetteur (trajet $[B2-B1]$). Les valeurs relevées ne prennent donc pas en compte l'augmentation moyenne du temps de transit due à l'élargissement de la zone d'injection d'émetteur le long de la jonction émetteur-base. 131

5.25 Courants d'électrons relevés aux points $B1$ et $B2$ de la base. La charge stockée totale correspond à l'aire entre les courbes $I_{e,B2}$ et $I_{e,B1}$. 134

5.26 Évolution de la tension aux bornes de la base (tension V_{B1B2}, en bleu foncé) et de la résistance de cette portion (en vert clair) en fonction du temps. Le temps s'écoulant entre le déclenchement de l'effet bipolaire à t_{NPN} et le repliementde la tension à t_{max} correspond approximativement à la constante de temps caractéristique τ_R de la décroissance exponentielle de la résistance. 136

5.27 Tension aux bornes de la base (vert clair) et aux bornes de la jonction base-collecteur (bleu foncé) en fonction du temps. La valeur atteinte pour chacune des parties à l'instant t_{EB} sert de point de référence pour le calcul de la surtension. On note que la plus grande partie de la surtension est prise en charge par la jonction base-collecteur, et que celle-ci se replie en premier. 137

5.28 Courant circulant dans la diode P^+N et rapport d'injection $\alpha = \frac{I_e}{I}$ en fonction du temps. L'augmentation du rapport d'injection à partir de l'instant t_{BV} montre le déclenchement de la diode en direct. 138

5.29 Courant de trous dans la base du PNP latéral. La jonction émetteur-base commence à injecter des trous à partir de 3.4 ns. 139

5.30 Caractéristiques $I-V$ quasi-statiques de composants $NPNB$ $Pad40$ pour deux longueurs de doigt : 75 μm et 250 μm (courbes bleue foncé et vert clair respectivement). Les tensions maximales sont reportées en abscisses, avec le courant quasi-statique en ordonnée (carrés rouges et croix violettes). Les deux composants ayant la même tension V_{t1}, leurs surtensions se comparent directement à partir des valeurs de V_{max}. 142

5.31 Caractéristiques TLP des $NPNB$ standard (courbe bleu foncé) et sans tranchée (vert clair) en technologie $SmartMos$ 8 MVS. Un différence de 4 V est visible entre leurs tensions de repliement. 143

5.32 Mesure des surtensions des deux composants précédents de 20 V LP à 200 V TLP. La différence entre les deux surtensions au-delà de 200 V TLP est de l'ordre de grandeur de la précision de l'outil de mesure, et elles ne sont donc pas représentées ici. 144
5.33 Caractéristiques TLP de deux $NPNB$ en technologie $SmartMos$ 8 MVS pour deux longueurs de base du PNP différentes. 144
5.34 Surtensions comparées des deux $NPNB$ précédents entre 20 V et 200 V. 145
5.35 Courbes quasi-statiques de deux $Pad15$ $NPNB$ en technologie $SmartMos$ 8 MV, pour des longueurs d'émetteur du PNP de 8 udm (bleu foncé) et de 5 udm (vert clair). 145
5.36 Surtensions mesurées sur les deux composants de la figure 5.35. 145

B.1 Schéma représentatif d'un générateur d'impulsion TLP/vf-TLP ix

Liste des tableaux

2.1 Atténuation du signal à travers les câbles coaxiaux, à différentes fréquences. Comparaison entre l'atténuation du signal vu par le DUT et le signal mesuré. 36

2.2 Bilan des configurations pour lesquelles les différents cas évoqués ont lieu sur la protection PNP/Diode et le composant NLDMOS. Le cas 3 n'est observé que pour des temps de montée faibles et des fortes résistances de grille. 43

5.1 Tensions de repliement quai-statique du $LDMOS$ pour deux pentes $dV_{Générateur}/dt$: 6 V/ns et 200 V/ns, correspondant respectivement à des temps de montée de 10 ns et 300 ps environ, et pour trois valeur de la résistance de grille : 0 Ω, 1 $k\Omega$ et 10 $k\Omega$. 108