



HAL
open science

Optimisation et modélisation de protection intégrées contre les décharges électrostatique, par l'analyse de la physique mise en jeu

David Trémouilles

► **To cite this version:**

David Trémouilles. Optimisation et modélisation de protection intégrées contre les décharges électrostatique, par l'analyse de la physique mise en jeu. Micro et nanotechnologies/Microélectronique. INSA de Toulouse, 2004. Français. NNT: . tel-00010263

HAL Id: tel-00010263

<https://theses.hal.science/tel-00010263v1>

Submitted on 23 Sep 2005

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.

Optimisation et modélisation de protections intégrées contre les décharges électrostatiques, par l'analyse de la physique mise en jeu

THÈSE

présentée et soutenue publiquement le 14 mai 2004

pour l'obtention du

Doctorat de l'Institut National des Sciences
Appliquées de Toulouse
(spécialité électronique)

par

David TRÉMOUILLES

Composition du jury

<i>Président :</i>	J.M. DORKEL
<i>Rapporteurs :</i>	G. GHIBAUDO G. GROESENEKEN
<i>Examineurs :</i>	L. LESCOUZÈRES P. NOUET Ph. PERDU
<i>Invité :</i>	F. BEAUDOIN
<i>Directrice de thèse :</i>	M. BAFLEUR

Mis en page avec la classe thloria.

Remerciements

Le travail présenté dans ce mémoire a été effectué conjointement au sein du groupe "Composant et intégration de puissance" (CIP) du laboratoire d'analyse et d'architecture des systèmes (LAAS) du CNRS et du groupe de technologie de la société ON Semiconductor à Toulouse.

A ce titre, je tiens à remercier Monsieur J.C. LAPRIE et M. GHALLAB, directeurs successifs du LAAS, de bien avoir voulu m'accueillir au sein du Laboratoire.

Je remercie également Monsieur J.L. SANCHEZ, Directeur de recherche au CNRS et responsable du groupe CIP pour m'avoir donné les moyens de mener à bien ce travail. J'ai également une pensée émue pour Monsieur G. CHARITAT, ancien directeur du groupe.

C'est avec une profonde sincérité que j'adresse ma reconnaissance à Madame Marise BAFLEUR, Directrice de Recherche au CNRS, qui m'a dirigé au cours de cette thèse. Je la remercie pour son soutien permanent et pour ces précieux conseils qui m'ont permis de mener à bien ce travail dans les meilleures conditions.

J'adresse mes remerciements à Lionel LESCOUZÈRES, responsable du groupe de technologie de ON Semiconductor pour m'avoir encadré et pour la confiance qu'il m'a accordée tout au long de cette thèse.

Je remercie les membres du jury, et tout d'abord Monsieur Jean-Marie DORKEL Professeur à l'Institut des Sciences Appliquées (INSA) de Toulouse pour m'avoir fait l'honneur de présider ce jury et pour les discussions très intéressantes que nous avons pu échanger.

Je remercie Gérard GHIBAUDO, Directeur de Recherche au CNRS, et Monsieur Guido GROESENKEN, Professeur à l'Université de Louvain (Belgique), qui ont accepté la tâche d'être rapporteurs de mes travaux de thèse.

J'adresse également mes remerciements à Monsieur Pascal NOUET, Maître de conférence à l'Université de Montpellier, Philippe PERDU, Ingénieur Senior au Centre National des Études Spatiales (CNES) et Félix BEAUDOIN Ingénieur au CNES, pour leurs participations au jury de thèse et pour les échanges scientifiques fructueux que nous avons pu avoir.

Je remercie Monsieur Nicolas NOLHIER, maître de conférence à l'Université Paul Sabatier de Toulouse, pour ses compétences et ses grandes qualités humaines qui ont contribué à l'aboutissement de ce travail.

Je remercie particulièrement Monsieur Nicolas MAURAN, responsable de la salle caractérisation du laboratoire, pour sa compétence, sa bonne humeur et sa patience durant les longues heures passées à l'amélioration du banc de caractérisation TLP.

Je remercie également les membres de l'équipe ESD de ON Semiconductor Toulouse : G. BERTRAND, E. STEFANOV, R. ESCOFFIER et P. GEULLE. Je tiens également à remercier les membres de l'équipe de design, pour qui ESD signifie encore trop souvent "Expect Some Delay".

Merci à Christelle FRANCHINI (DELAGE) et Géraldine BERTRAND, qui ont étudié le sujet des ESD au cours des thèses précédant celle-ci. Sans leur compétences et les discussions instructives que nous avons échangées, ces travaux de thèse n'auraient pu atteindre un tel aboutissement.

J'adresse également mes remerciements les plus sincères à Thomas BEAUCHÊNE et Fabien ESSELY, successivement doctorant à l'IXL de Bordeaux, pour leur amitié et le temps qu'il ont bien voulu consacrer à l'étude de nos échantillons.

Je n'oublie pas de remercier Isabelle NOLHIER, secrétaire du groupe CIP pour sa disponibilité, son efficacité et sa sympathie.

J'en profite également pour remercier tous mes amis et collègues de bureau ou de pause. Dans le désordre je pense à : Isabelle B. (Petit ange et reine de l'anacoluthie), Rodolphe DM. (Futur cadre de la métallurgie (au sens liturgie) s'il n'a pas perdu l'ouïe avant la fin de sa thèse...), Patrice B. (hier jeune doctorant dynamique et à cette date jeune cadre dynamique), Jean Philippe L. (Le petit Suisse), Nicolas G. (Disciple), Christophe S. (Dit Bibi), Laurence et Frédéric M. (Fred l'incolable...), Stéphane A. (3D conceptor), Nicolas M. (zob), Christian C. (D.J.), Sandrine A., Eric I., Hervé C. (maçons ...), Isabelle N. (Maman), Nicolas N. (Papa), J. Saint-Martin, Géraldine B., Chritelle D., Sylvie R., Thierry B. (Bah! Te fais pas chier ...), Amaury G. (Le nouveau), Laurent R. (Ca se passe deux têtes plus haut), Steve J., Bruno E., Emmanuel D., Thomas B., Fabien E., Guillaume B. (english man), Laurent B. (Boubou), Adeline F., Abdelhakim B., Eric A., Ghislain T., Karine I., Lionel M., Maxime D., ptifred, Laetitia et Laurent M. (un ami de 15 ans (déjà...)), Julien T., Aurelie C., Julien B., Jérôme P., Stéphane C., Anne Laure et Olivier D. (à qui je souhaite beaucoup de bonheur), les Caretounaires, ... J'adresse à tous mes vœux de réussite les plus sincères.

Santé à tous les membres de l'équipe de rugby du LAAS et du CNRS.

Je salue également toutes les personnes que j'oublie de citer et qui ont pourtant compté au cours de ces années de préparation de thèse.

Je tiens à exprimer ma reconnaissance à l'ensemble du personnel technique du LAAS, de la gestion du personnel, du magasin et du service édition et documentation.

Je tiens à remercier également mon instituteur Monsieur G. Tourtonde qui m'a donné le goût pour les sciences et l'informatique.

Enfin, c'est avec une profonde émotion que je remercie mes proches et ma famille pour le soutien et la confiance qu'ils m'ont toujours accordés.

À mes parents.

*Le savant n'étudie pas la nature parce que cela est utile ;
il l'étudie parce qu'il y prend plaisir
et il y prend plaisir parce qu'elle est belle.
Si la nature n'était pas belle, elle ne vaudrait pas la peine d'être connue,
la vie ne vaudrait pas la peine d'être vécue.
Henri Poincaré, extrait de Science et Méthode*

Table des matières

Introduction générale	11
------------------------------	-----------

Chapitre 1

Décharges électrostatiques et outils de caractérisation

1.1	Les Décharges électrostatiques	13
1.1.1	Généralités	13
1.1.2	ESD et microélectronique	14
1.2	Modèles de décharge et testeurs industriels	15
1.2.1	HBM	15
1.2.2	MM	16
1.2.3	CDM	17
1.2.4	Comparaison de différents types de stress	18
1.3	Outils de caractérisation	19
1.3.1	Le banc de caractérisation TLP	20
1.3.2	La microscopie à émission lumineuse	21
1.4	Techniques d'analyse et de localisation de défaillance	22
1.4.1	La photoémission	22
1.4.2	Les techniques de stimulation laser	22
1.4.3	Analyse par la face arrière	25
1.5	Conclusion	25

Chapitre 2

Étude des transistors bipolaires autopolarisés (TBA)

2.1	Composants de protection ESD basés sur le TBA	27
2.1.1	Les différents types de composants	27
2.1.2	Principe de fonctionnement	29
2.1.3	Éléments d'optimisation	30
2.2	Effets des fortes densités de courant dans les transistors bipolaires	34
2.2.1	Chute du gain en courant	34
2.2.2	Limitation de l'aire de sécurité d'utilisation	34
2.2.3	Effet Kirk	36
2.2.4	Cas des TBA	36

2.3	Éléments d'une approche régionale unidimensionnelle	36
2.3.1	Comportement de l'émetteur	37
2.3.2	Comportement de la région de base	38
2.3.3	Comportement de la région de collecteur	40
2.4	Étude unidimensionnelle fort courant des TBA	42
2.4.1	Mise en équation du problème	43
2.4.2	Caractéristique tension-courant	45
2.4.3	Effets spécifiques des fortes densités de courant	45
2.5	Phénomène de focalisation du courant	49
2.5.1	Étude théorique	50
2.5.2	Vérification expérimentale	51
2.6	Destruction du composant : aspect thermique	52
2.6.1	Chemin vers la destruction du composant	52
2.6.2	Proposition d'un nouveau type de composant	54
2.7	Conclusion et perspectives	55

Chapitre 3

Optimisation de transistors bipolaires autopolarisés

3.1	Utilisation des simulateurs électrothermiques	57
3.1.1	GIGO law	58
3.1.2	Limitation des modèles physiques	58
3.1.3	Approche et solutions retenues	59
3.2	Cas d'un composant bipolaire NPN vertical	60
3.2.1	Description de la structure	60
3.2.2	Modes de fonctionnement lors d'une décharge électrostatique	60
3.2.3	Caractéristiques électriques	61
3.3	Étude de son fonctionnement par la simulation physique bidimensionnelle	61
3.3.1	Surtension avant le repliement	62
3.3.2	Repliement de la tension, élargissement de la région de base	64
3.3.3	Phénomènes de focalisation du courant	66
3.3.4	Comportement selon la longueur de l'émetteur	68
3.3.5	Extrapolations aux très forts courants	73
3.4	Augmentation de la profondeur de la région de base effective	73
3.4.1	Découplage thermique	73
3.4.2	Découplage électrique	74
3.5	Influence du profil de dopage de collecteur sur la tension de maintien	76
3.6	Bilan : Éléments d'optimisation pour la conception	79
3.7	Résultats expérimentaux	80
3.7.1	Description des composants réalisés	80
3.7.2	Coupes technologiques	81

3.7.3	Critère de défaillance	84
3.7.4	Influence de la distance collecteur-émetteur	85
3.7.5	Influence de la géométrie	87
3.7.6	Comparaison des performances avec les composants standards	89
3.7.7	Techniques de déclenchement	91
3.7.8	Analyse de défaillance	95
3.8	Cas d'un TBA PNP	97
3.9	Conclusion	97

Chapitre 4

Modélisation de composants bipolaires autopolarisés

4.1	État de l'art de la modélisation de type SPICE des protections ESD	101
4.1.1	Transistors bipolaires NPN	102
4.1.2	Transistors NMOS	103
4.2	Approche retenue et justification	105
4.3	Modélisation des TBA à collecteur faiblement dopé	106
4.3.1	Présentation générale du modèle	106
4.3.2	Méthode de calcul du facteur de multiplication	107
4.3.3	Variation du coefficient de multiplication avec la densité de courant	109
4.4	Remarques pour la modélisation de transistors NMOS	112
4.5	Modélisation du TBA de la technologie 3	112
4.5.1	Modélisation de la diode D_D	113
4.5.2	Extraction des paramètres statiques	113
4.5.3	Résultats de simulation, limitation du modèle	116
4.6	Étude de la dynamique	118
4.6.1	Dynamique du déclenchement	118
4.6.2	Déclenchement par dv/dt	119
4.6.3	Méthode d'optimisation du couplage de grille	123
4.7	Conclusion	127

Chapitre 5

Étude des stratégies de protections

5.1	Stratégies de protection des circuits	129
5.1.1	Protection d'entrée	129
5.1.2	Protection de sortie	130
5.1.3	Protection du bus d'alimentation	131
5.1.4	Stratégie de protection globale	131
5.2	Étude d'une stratégie de protection	133
5.2.1	Description du circuit de test	133
5.2.2	Confrontation des résultats expérimentaux et simulés	135

5.2.3	Analyse de défaillance	138
5.3	Étude et optimisation des circuits de protection élémentaires	140
5.3.1	Résultats sur les circuits de test	141
5.3.2	Analyse par la simulation physique	143
5.3.3	Règles de dessin et validation	145
5.4	Cas particulier pour le stress MM	147
5.4.1	Résultats expérimentaux	147
5.4.2	Origine du problème	148
5.4.3	Solution et recommandations	149
5.5	Conclusion	150
	Conclusion générale	153
	Bibliographie	157
	Liste des publications	167

Introduction générale

Les décharges électrostatiques ou «electrostatic discharges» (ESD) en anglais sont des phénomènes naturels qui peuvent détruire les circuits intégrés. On peut comparer cette agression à celle de la foudre sur un arbre, laquelle met en jeu une densité d'énergie équivalente. Les décharges électrostatiques sont devenues un problème majeur de la microélectronique à la fin des années 1970, avec l'utilisation croissante des microprocesseurs et des technologie MOS. Aujourd'hui le souci de protection des composants couvre de nombreux domaines, les applications radio-fréquences [1, 2], les composants optoélectroniques [3], les capteurs chimiques ISFET [4], les micro-systèmes [5] ...

Tous les dispositifs ou circuits de la microélectronique moderne, destinés à une utilisation grand public, se doivent d'être protégés contre les ESD. Sans cette précaution, leur fiabilité et en particulier leur durée de vie se trouveraient extrêmement limitées. Pucés des cartes bancaires, connecteurs de recharge des téléphones portables, ports de connexion informatiques (USB, Fire-Wire, ...), sont des exemples d'applications particulièrement critiques. 30 à 50 % des défaillances des dispositifs sont attribuées aux conséquences des ESD, ce qui fait de ce domaine un grand enjeu industriel.

Il est donc impératif de limiter l'impact de ces phénomènes :

- D'une part en contrôlant la génération de charges électriques dans les lieux stratégiques que sont les zones de fabrication et de test des dispositifs par l'utilisation de bracelets antistatiques, le contrôle du degré d'humidité, etc ...
- D'autre part, en intégrant au cœur même du système des composants de protection qui vont détecter et détourner les décharges protégeant ainsi les parties actives.

Aujourd'hui, compte tenu de la vulnérabilité croissante des circuits intégrés et de la réduction des cycles de développement des circuits et des technologies, l'approche utilisée pour concevoir des circuits de protection n'est plus empirique. Les études menées au cours des dernières décennies ont conduit à la création d'outils de caractérisation et de simulation appropriés qui permettent une conception plus méthodique.

L'objectif de nos travaux est de contribuer à l'amélioration de ces méthodes de conception et des performances des composants de protection.

Pour cela, l'accent est principalement mis sur l'étude de la physique mise en jeu dans un composant soumis à une décharge électrostatique. En effet, les densités de courant atteignent des valeurs très importantes dans un composant soumis à un ESD. Ces régimes de fonctionnement extrêmes vont bien au delà des régimes limites de fonctionnement de composants classiques. Nous nous sommes donc attachés à décrire les phénomènes physiques spécifiques afin d'optimiser et modéliser plus efficacement les composants de protection. Le cas d'une protection largement répandue, le transistor bipolaire autopolarisé, est plus particulièrement étudié. De plus, l'étude et l'optimisation de circuits de protection ESD est également menée.

Le premier chapitre donne une vue générale du problème des décharges électrostatiques et des outils expérimentaux qui permettent d'étudier les composants de protection. La très courte durée d'une décharge électrostatique et la très forte intensité du courant demande l'utilisation d'outils de caractérisation spécifiques. Les modèles de décharges électrostatiques couramment utilisés dans l'industrie et les testeurs associés sont décrits. Le banc de mesure TLP (Transmission Line Pulse) et la technique de microscopie par émission de lumière qui permettent une caractérisation

plus approfondie sont également présentés. Enfin, nous introduisons les techniques de localisation de défaillance qui donnent également de précieux indices sur les mécanismes de dégradation. Les techniques de stimulation laser sont plus particulièrement décrites. Les études présentées dans les chapitres suivants s'appuient largement sur les résultats obtenus grâce à ces outils.

Le second chapitre est consacré à l'étude du fonctionnement des composants de protection qui reposent sur des transistors bipolaires autopolarisés. Un rapide état de l'art de ce type de composant est d'abord présenté. L'étude des phénomènes liés aux très fortes densités de courant s'appuie sur les connaissances issues de l'électronique de puissance et des composants haute fréquence. L'extension des approches adoptées en électronique de puissance permet de décrire le comportement des composants. Nous présentons la construction d'un modèle unidimensionnel adapté au fort courant à partir d'une méthode de modélisation régionale. Il en découle une explication originale de la focalisation du courant dans ce type de composant et l'identification de paramètres critiques pour leur optimisation.

Dans le troisième chapitre, l'optimisation de transistor bipolaire de protection de type NPN est présentée. Les composants étudiés possèdent un collecteur faiblement dopé qui permet d'atteindre une grande robustesse aux ESD. L'étude de leur fonctionnement s'appuie sur l'utilisation de la simulation physique bidimensionnelle. Après avoir mis en évidence ses limitations, l'utilisation astucieuse de cet outil permet de proposer une description approfondie de leur fonctionnement ainsi qu'un ensemble de règles de conception pour obtenir des composants optimisés et particulièrement robustes. Ces règles sont vérifiées expérimentalement sur trois technologies distinctes. De plus, la description du fonctionnement proposée est validée par l'analyse des caractéristiques électriques TLP et les résultats de localisation de défaillance.

La prise en compte des effets de fortes densités de courant dans les modèles compacts de type SPICE est abordée dans le quatrième chapitre. Après avoir présenté un rapide état de l'art des méthodes de modélisation des composants de protection ESD, nous proposons une solution pour enrichir ces modèles et prendre en compte les effets des fortes densités de courant étudiés dans les chapitres précédents. Nous abordons également les aspects de la dynamique du déclenchement des structures. Le cas des composants de protection réalisés à partir de transistors MOS est traité plus en détail. Cette analyse est conclue par la présentation d'une méthode d'optimisation de la technique de déclenchement par le couplage de grille.

Enfin, nous consacrons le dernier chapitre à l'étude de la stratégie globale de protection d'un circuit intégré. Après un rapide aperçu des stratégies de protection couramment utilisées, l'étude complète de la protection d'un circuit de test est présentée. L'efficacité de l'utilisation de la simulation de type SPICE est démontrée. Cependant, nous mettons en évidence que le développement de composants de protection performants et l'utilisation de la simulation pour vérifier la validité d'une stratégie de protection ne sont pas suffisants pour garantir la robustesse d'un circuit. Des phénomènes parasites liés à l'interaction des composants au sein du circuit de protection sont mis en évidence. Les solutions proposées sont validées et la méthodologie de conception enrichie pour éviter ce type de problème.

Chapitre 1

Décharges électrostatiques et outils de caractérisation

La protection des circuits et dispositifs de la micro et nano-électronique est aujourd'hui un enjeu primordial pour garantir des durées de vie raisonnables aux applications. Considérant sa forte intensité en courant et sa durée extrêmement courte, une décharge électrostatique est un phénomène qui est particulièrement violent et peut être destructeur pour un circuit intégré. Aussi, des outils de caractérisation spécifiques ont été développés, tels les testeurs HBM, MM et CDM qui permettent de reproduire en laboratoire, de manière contrôlée, ces phénomènes naturels. Mais ces derniers ne donnent qu'une information limitée sur la capacité d'un dispositif à supporter une décharge d'intensité donnée sans être détruit. Des outils de caractérisation adaptés ont donc été développés. Le banc TLP est un outil qui donne accès à la caractéristique quasi-statique à très fort courant. Les techniques d'analyse classiques, comme la photoémission, permettent de donner des informations sur le fonctionnement des dispositifs. Enfin, de nouvelles techniques d'analyse de défaillance permettent de localiser et caractériser les défauts provoqués par une décharge. La maîtrise de ces outils expérimentaux plus ou moins spécifiques est indispensable pour la conception, l'optimisation et la modélisation des dispositifs de protection.

1.1 Les Décharges électrostatiques

1.1.1 Généralités

Les décharges électrostatiques sont des phénomènes naturels qui se produisent dans une large gamme d'échelles. L'éclair qui déchire le ciel ou la petite étincelle qui surgit au bout de votre doigt touchant la poignée d'une porte en sont les manifestations les plus communes et spectaculaires, parfois destructrices pour la première, souvent douloureuses pour la seconde. Lors de ces deux phénomènes, l'émission de lumière visible témoigne de l'ionisation des molécules contenues dans l'air, le plasma ainsi créé permettant le passage d'un courant de forte intensité. L'origine de ce courant est associée à un déséquilibre de charges électriques entre deux corps électriquement isolés : le corps humain et un objet métallique, ou les nuages et la terre pour les éclairs. La création ou l'apparition d'un chemin de conduction permet de rétablir l'équilibre, par le transfert d'une quantité importante de charges d'un corps à l'autre. Comme ce chemin est très peu résistif, le courant est très intense et la durée du transfert des charges extrêmement courte.

Nous n'entrerons pas dans les détails des mécanismes de la génération du déséquilibre de charges qui, d'une part, est largement traité pour les phénomènes de la vie courante [6, 7] et d'autre part, occupe encore beaucoup de scientifiques concernant la foudre [8]. Nous décrirons simplement les plus significatives dans l'environnement de la microélectronique, la triboélectricité, l'induction et la conduction.

Le premier de ceux-ci est le plus courant. L'exemple typique est une personne qui marche sur une moquette. En effet, lors du contact entre deux matériaux de nature différente (les chaussures et le sol), un transfert d'électrons libres peut avoir lieu entre les objets. Si l'un d'eux au moins est isolant, une charge résiduelle persiste (dans l'isolant) lors de la séparation. Cette charge en excès va induire une différence de potentiel qui est fonction de la capacité entre l'objet ou la personne et la terre. La quantité de charges transférée et leur polarité dépendent des caractéristiques des matériaux (Travaux de sortie). La triboélectrification est d'autant plus importante que les surfaces sont lisses et d'aire importante, et que la pression et la vitesse de frottement entre les matériaux sont grandes.

La charge par induction est un phénomène typique des environnements informatiques. Il se déroule en deux temps. Un objet conducteur placé dans un champ électrique, par exemple d'un autre objet chargé ou d'un écran d'ordinateur, voit une partie de ses charges se séparer comme l'illustre la figure 1.1(a).

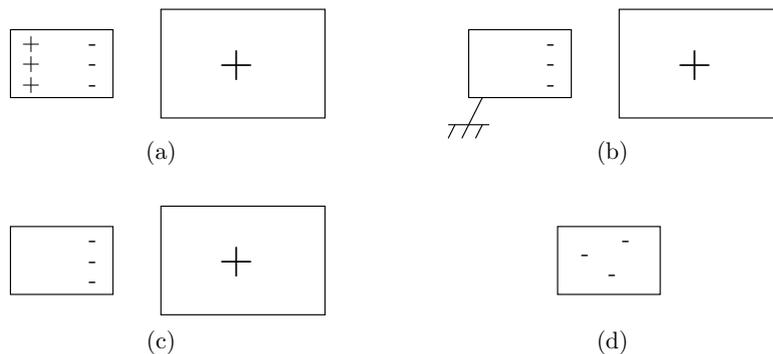


FIG. 1.1 – La charge par induction.

Si l'objet est momentanément mis à la masse, une partie des charges va être évacuée, figure 1.1(b). L'objet est alors chargé, figure 1.1(c) et (d). Hors de la zone d'influence, s'il est de nouveau mis à la masse, ses charges vont de nouveau s'équilibrer. Deux pics de courant peuvent donc avoir lieu dans ce cas.

Le dernier phénomène, la charge par conduction, se déroule principalement sur les chaînes de test. Les composants étant souvent en contact, un composant chargé, par exemple par triboélectrification, peut transférer une partie de ces charges aux autres composants.

L'humidité relative de l'air est un paramètre important qui permet de limiter la génération et le maintien des charges électriques. Un air humide sera favorable pour diminuer les méfaits des décharges électrostatiques, tandis qu'un air sec entraînera des tensions vingt à trente fois plus importantes. Dans un air sec, marcher sur un tapis peut engendrer une tension de 35kV entre une personne et la terre, elle ne sera plus que de 1,5kV avec un air humide [6].

1.1.2 ESD et microélectronique

Les décharges électrostatiques rencontrées dans la vie courante peuvent avoir des conséquences spectaculaires si elles se produisent dans les environnements dangereux où sont manipulés des substances explosives comme les stocks de carburant ou d'armes [9].

Moins spectaculaire dans le monde de la microélectronique, elles n'en restent pas moins la cause de nombreux problèmes.

Problème de fiabilité

Les dispositifs et les circuits de l'électronique moderne deviennent très vulnérables aux décharges électrostatiques compte tenu de la réduction des dimensions qui sont couramment aujourd'hui de l'ordre de la centaine de nanomètres. Ainsi, l'énergie d'une décharge provoquée

par un être humain se trouve dissipée dans des volumes de plus en plus réduits, menant à une destruction prématurée du dispositif, même pour de faibles niveaux de décharges. On constatait en 1995 que le pourcentage des défaillances des dispositifs liées aux ESD était de 30 à 50 % [6]. Il est donc impératif de limiter l'impact de ces phénomènes. D'une part, en contrôlant la génération de charges électriques dans les lieux stratégiques que sont les zones de fabrication et de test des dispositifs, par l'utilisation de bracelets, de vêtements et de matériaux antistatiques, le contrôle du degré d'humidité, ... et d'autre part, en intégrant des composants de protection au cœur même du système, qui vont détecter et détourner les décharges protégeant ainsi les parties actives [10, 11, 12].

Grâce aux protections intégrées, la robustesse des circuits a augmenté malgré la forte réduction des dimensions lithographiques et la diminution des tensions d'alimentation liée à l'aminçissement des oxydes de grille. En conséquence, la protection des circuits face aux décharges électrostatiques est devenue l'une des causes majeures de re-dessin de masques d'un circuit et une étape importante de la conception d'une application. Ainsi, les enjeux du développement de solutions de protection tiennent dans l'optimisation de composants performants, en terme de robustesse par unité de surface de silicium occupée, et la mise au point de nouveaux outils et méthodes permettant de traiter ce problème dès le début de la conception d'un circuit.

Problème de compatibilité électromagnétique

Lors d'une décharge électrostatique, le champ électromagnétique émis peut perturber le fonctionnement des circuits situés à proximité du lieu de la décharge. Il ne s'agit pas ici de l'effet direct du courant de décharge, mais de l'onde électromagnétique émise dans l'environnement proche. Les décharges électrostatiques posent donc des problèmes de compatibilité électromagnétique (CEM). Les mesures réalisées dans le cas de la décharge d'un corps humain pour des niveaux de précharge de 3 kV, montrent que l'intensité du champ rayonné est importante [13]. La variation du champ électrique est de l'ordre de 10^2 à 10^3 V/m crête-à-crête et de 10 à 10^2 A/m pour le champ magnétique, à une distance de 10 cm du lieu de la décharge. La largeur du spectre des fréquences émises est supérieure à 4 GHz.

1.2 Modèles de décharge et testeurs industriels

Beaucoup de paramètres peuvent influencer sur la forme du courant, et la durée totale d'une décharge électrostatique. Plusieurs modèles de décharges ont donc été développés en fonction des différentes situations rencontrées en microélectronique. Le plus courant est le modèle du corps humain, en anglais Human Body Model (HBM) qui, chargé électriquement, se décharge au travers d'un composant. De la même manière, le modèle de la machine (MM) décrit la décharge engendrée par un équipement. Le modèle du composant chargé, en anglais Charged Device Model (CDM), considère le cas où le composant lui-même est chargé, et se décharge en entrant en contact avec un conducteur.

Les testeurs utilisés dans l'industrie reproduisent les modèles de décharge précédents en respectant certaines normes. La caractérisation de la robustesse d'un circuit vis-à-vis des ESD consiste à soumettre le composant à une série de décharges d'intensité croissante, jusqu'à sa destruction.

1.2.1 HBM

Modèle HBM

Considéré comme le principal modèle de décharges électrostatiques, il est également le plus ancien. Il décrit la décharge d'un être humain debout, par l'extrémité d'un de ses doigts. En première approximation, on peut le représenter par un simple réseau RC, composé d'une capacité de 100 pf et d'une résistance de 1500 Ω (Fig. 1.2(a)).

Avant la décharge, la capacité est typiquement chargée à des tensions de l'ordre de quelques kV. Cette tension de précharge est utilisée pour caractériser l'intensité de la décharge HBM.

L'impédance offerte par un circuit intégré au courant de décharge ESD pouvant en général être considérée comme très faible, la décharge HBM est assimilée à une impulsion de courant indépendante du composant testé. La durée totale de l'impulsion est d'environ 300 ns avec un temps de montée qui peut varier entre 2 et 10 ns. Le pic d'intensité a une valeur comprise entre 1 et 10 A.

Sans protection, la robustesse d'un circuit est de l'ordre d'une centaine de Volts et seulement de quelques dizaines de Volts pour les têtes de lecture magnétiques. Ces chiffres sont à comparer au seuil de perception humain d'une décharge qui est d'environ 3 kV.

Pour les circuits, une robustesse minimum de 2 kV est généralement requise pour permettre leur manipulation dans des conditions classiques de stockage et d'assemblage.

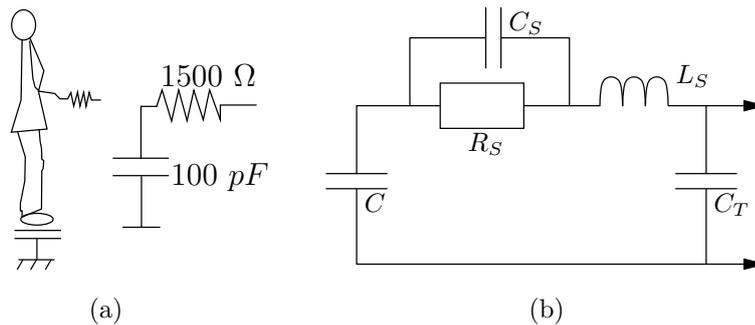


FIG. 1.2 – Modèle du corps humain (HBM) (a) et schéma électrique d'un testeur HBM (b).

Testeur HBM

Très utilisé industriellement, il reproduit le modèle du corps humain, en reprenant ses composants de base ($C=100 \text{ pF}$, $R_S=1500 \Omega$) (Fig. 1.2(a)). Les éléments parasites L_S et C_S sont nécessaires pour produire une forme d'onde réaliste. C_T représente la capacité parasite associée au testeur. L'inductance série L_S dont la valeur est de l'ordre de $7 \mu\text{H}$, détermine le temps de montée de l'impulsion de courant.

Les normes décrivent les valeurs des composants du schéma électrique (Fig. 1.2(b)), ainsi qu'un gabarit de la forme d'onde du courant de décharge dans un court-circuit ou une résistance de 500Ω [14, 15, 16]. La procédure de test ainsi qu'une classification des composants selon leur niveau de robustesse sont également définie.

Toutefois, il semble que les informations fournies par les fabricants de circuit intégrés ne sont pas toujours claires, d'autant plus que des normes HBM spécifiques et différentes sont utilisées pour les systèmes sur carte [17]. Malgré la normalisation, des résultats sensiblement différents peuvent également être observés en fonction du testeur utilisé. Ces écarts peuvent être attribués aux différences de temps de montée de l'impulsion de courant.

Pour les simulations présentées dans ce mémoire, les paramètres du modèle HBM ont été ajustés pour représenter le testeur utilisé au sein du laboratoire ($C = 100 \text{ pF}$, $R_S = 1,5 \text{ k}\Omega$, $L_S = 6,4 \mu\text{H}$, $C_S = 1 \text{ pF}$, $C_T = 61,5 \text{ pF}$).

1.2.2 MM

Modèle MM

Le modèle de décharge Machine Model (MM) ou modèle des machines est une extension du HBM pour le cas où la résistance série R_S est réduite. Cela permet en particulier de rendre compte de la décharge d'une personne tenant un outil métallique (pince, fer à souder), ou par un

système, comme un robot, manipulant des composants (Fig. 1.3(a)). Ce modèle a été développé au Japon comme pire cas du modèle HBM, dont il diffère principalement par sa résistance série quasi nulle.

Avant une décharge, la capacité est typiquement chargée à quelques centaines de Volts. La forme d'onde du courant généré par la décharge est généralement oscillatoire, avec une fréquence comprise entre 5 et 15 Mhz. L'intensité maximale au cours d'une décharge MM est de l'ordre de 1 à 10 A. Les oscillations sont dues à l'existence d'une inductance série non négligeable, en particulier au contact, et à la faible résistance série du modèle.

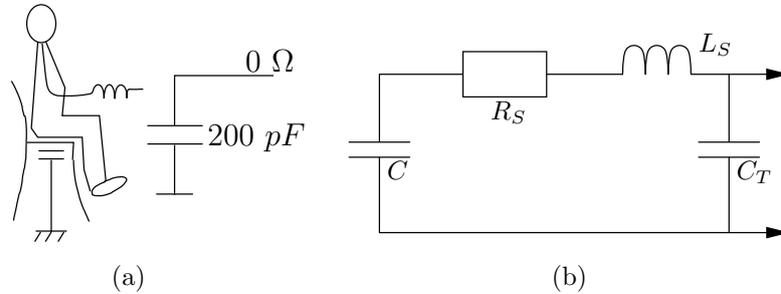


FIG. 1.3 – Modèle des machines (MM) (a) et schéma électrique d'un testeur MM (b).

Testeur MM

D'architecture proche du testeur HBM, il est souvent inclus dans le même appareil de caractérisation. Il repose sur le même schéma électrique (Fig. 1.3(b)), seules les valeurs des composants sont différentes. La capacité est de 200 pF, la résistance série R_S est théoriquement nulle et la valeur de l'inductance L_S de l'ordre de 500 nH.

Les normes décrivent les paramètres du modèle et la forme du courant dans un court-circuit et une résistance de 500 Ω [18, 19, 20].

L'utilisation de ce type de test est moins courante que le HBM car la faible valeur de la résistance série le rend très dépendant du testeur utilisé. De plus, les procédures de test sont peu standardisées.

Les Paramètres du modèle ajustés sur le testeur utilisé au sein du laboratoire sont les suivants : $C = 200 \text{ pF}$, $R_S = 5,65 \text{ } \Omega$, $L_S = 700 \text{ nH}$ et $C_T = 6,5 \text{ pF}$.

1.2.3 CDM

Modèle CDM

Ce modèle plus récent est différent des deux précédents par la nature des phénomènes qu'il décrit. Il représente la décharge d'un composant lui-même chargé, par une seule de ses pattes mise à la masse (Fig. 1.4(a)). Ce type d'événement se rencontre fréquemment sur les chaînes d'assemblage automatisées et pourrait constituer un problème majeur pour les circuits submicroniques.

La position du boîtier par rapport au plan de masse, les conditions atmosphériques (humidité relative, température, ...) influencent fortement le phénomène. Aussi, il est difficile d'évaluer précisément les valeurs des composants parasites, le boîtier ainsi que la puce faisant partie intégrante du modèle.

La forme du courant est elle aussi sensible à ces paramètres. Ce type de décharge se caractérise cependant par de très rapides variations du courant, plusieurs ampères par nano-seconde, et des temps très courts, quelques nano-secondes.

En fonction du composant lui-même, l'impulsion peut osciller ou non. Le pic de courant, d'une dizaine d'ampères, est grand par rapport aux modèles HBM et MM.

Le modèle CDM a été développé afin d'expliquer la rupture de certains oxydes dont l'origine ne peut être expliquée par un stress HBM ou MM. La détérioration est alors due aux chemins internes du courant et aux surtensions créées dans la puce lors de la décharge.

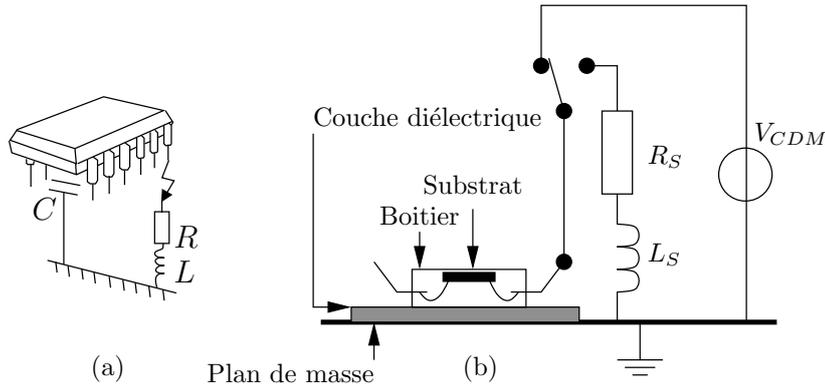


FIG. 1.4 – Modèle du composant chargé (CDM) (a) et schéma d'un testeur CDM (b).

Testeur CDM

Ce type de testeur est le plus complexe que les testeurs HBM et MM (Fig. 1.4(b)). Afin d'obtenir des conditions reproductibles, il doit s'affranchir des facteurs influençant la capacité associée au composant. De plus, la résistance et l'inductance série équivalente du circuit de décharge, doivent être minimisées et contrôlées pour garantir la rapidité de la décharge. Les tests CDM sont définis par les normes JEDEC [21] et de l'ESD Association [22]. Les valeurs typiques de la résistance et de l'inductance du modèles sont de l'ordre de 10Ω et 10 nH , respectivement.

1.2.4 Comparaison de différents types de stress

Le tableau 1.1 montre une synthèse comparative entre les trois modèles de décharges HBM, MM et CDM. Il permet de mettre en évidence la spécificité du modèle CDM par rapport aux modèles HBM et MM. Pour une tension de précharge de 500 V , au cours d'une décharge dans une résistance de 10Ω , le temps de montée et la durée de la décharge CDM sont environ dix fois plus faibles que celles des modèles HBM et MM, pour un pic de courant presque cinquante fois plus fort que celui du test HBM et deux fois plus fort que celui du modèle MM.

Modèle	HBM	MM	CDM
Éléments du modèle			
C(pF)	100	200	10
L_S (μH)	10	0,75	0,0025
R_S (Ω)	1500	10	10
Charge de C à 500 V pour $R_d = 10\Omega$			
$t_{montée}$ (ns)	<10	<10	0,1
$t_{durée}$ (ns)	150	<100	1
I_{pic} (A)	0,3	6,5	14

TAB. 1.1 – Comparaison des modèles de décharges électrostatiques HBM, MM et CDM [23].

L'amplitude plus importante des pics de courant lors de stress MM et CDM peut entraîner des défauts liés aux surtensions, en particulier dans les grilles des composants MOS qui y sont très sensibles.

Ces trois modèles peuvent également être comparés en terme d'énergie. L'énergie totale stockée dans la capacité est donnée par la relation :

$$E_{stock} = \frac{1}{2}CV^2 \quad (1.1)$$

avec V la tension de précharge. Cette énergie va être dissipée dans la résistance série du modèle et du composant testé ainsi que dans les jonctions, en particulier celles polarisées en inverse.

Si l'on considère seulement les résistances, l'énergie stockée doit être égale à l'énergie dissipée :

$$E_{stock} = R \int_0^{t_f} I^2(t)dt \quad (1.2)$$

où $R = R_S + R_d$ est la somme de la résistance du modèle et du composant testé, en intégrant sur la totalité de la durée de la décharge (t_f). L'énergie dissipée dans une résistance R_d est donc une fonction de l'énergie stockée :

$$E_{R_d} = \frac{R_d}{R_S + R_d} E_{stock} \quad (1.3)$$

L'énergie dissipée dans une région de charge d'espace d'une diode ou d'un transistor bipolaire qui soutient une tension V_H est donnée par :

$$E_{V_H} = \int_0^{t_f} V_H I(t)dt \quad (1.4)$$

Si l'on considère que la tension V_H reste constant durant toute la décharge et que $I > 0$, alors

$$E_{V_H} = V_H \int_0^{t_f} I(t)dt \quad (1.5)$$

Or, l'intégrale du courant correspond à la charge stockée dans la capacité du modèle ($q = CV$) on obtient donc :

$$E_{V_H} = V_H CV \quad (1.6)$$

et

$$E_{V_H} = 2 \frac{V_H}{V} E_{stock} \quad (1.7)$$

Cette dernière expression est exacte dans le cas d'un stress HBM pour lequel le courant est toujours positif. Nous pouvons cependant l'utiliser comme estimation dans le cas des stress MM et CDM afin de réaliser des comparaisons.

Le tableau 1.2 permet de comparer pour chacun des modèles, les proportions de l'énergie totale, dissipées dans une résistance de 10Ω et dans une jonction polarisée en inverse sous une tension de 10 V .

Il apparaît que la proportion d'énergie dissipée dans les résistances est très importante dans le cas des stress MM et CDM. C'est pour cette raison que la fusion de pistes métalliques est beaucoup plus fréquente pour ces types de stress.

Dans le cas des modèles HBM et MM, les dégâts occasionnés dans la puce sont généralement dus à l'énergie dissipée dans le composant. La rapidité et la forte valeur du pic de courant lors de stress CDM, durant lesquels une énergie limitée est dissipée vis-à-vis de stress HBM ou MM, va favoriser la création de défauts dans les oxydes de grille causés par les surtensions. Ces défauts, souvent latents car difficilement détectables, peuvent évoluer au cours du temps [24] ou fragiliser par la suite le circuit vis-à-vis de stress HBM [25].

1.3 Outils de caractérisation

Les testeurs HBM, MM et CDM ne donnent qu'une indication sur la robustesse du composant. La mesure en impulsion (TLP) permet, par contre, d'obtenir des caractéristiques plus fines. Cette méthode fut introduite en 1985 par Maloney dans le cadre de l'étude des protections ESD [26].

Modèle	HBM	MM	CDM
Éléments du modèle			
C(pF)	100	200	10
R_S (Ω)	1500	10	10
Charge de C à 500 V pour $R_d = 10 \Omega$ et $V_H = 10 V$			
E_{stock} (μJ)	12,5	25	1,25
E_{V_H} (μJ)	0,5	1	0,05
$\frac{E_{V_H}}{E_{stock}}$ (%)	4	4	4
E_{R_d} (μJ)	0,08	12,5	0,63
$\frac{E_{R_d}}{E_{stock}}$ (%)	0,64	50	50

TAB. 1.2 – Comparaison des modèles de décharges électrostatiques HBM, MM et CDM en termes d'énergie.

1.3.1 Le banc de caractérisation TLP

La méthode utilisée dans un banc de mesure TLP permet d'obtenir la caractéristique fort courant d'un composant en s'affranchissant des problèmes thermiques qui pourraient le détruire. En effet, lors d'une décharge électrostatique, des courants importants sont mis en jeu, mais la durée de l'impulsion étant très faible, l'énergie totale dissipée est finalement limitée. Ainsi un composant de protection ESD est amené à fonctionner sous de forts courants sans pour autant être endommagé.

Le principe du banc de mesure TLP est de générer une impulsion carrée de courant dans le composant testé, pendant une durée suffisamment courte pour ne pas le détruire et assez longue pour obtenir un courant et une tension constants et stables pour permettre de les mesurer. Grâce à plusieurs impulsions successives d'intensité croissante, on obtient point par point la caractéristique I(V) du composant. Le courant et la tension étant constants au cours de chaque mesure, on peut considérer que cette caractéristique est statique ou quasi statique. Entre chaque impulsion, l'évolution des caractéristiques électriques classiques peut être suivie pour mettre en évidence d'éventuelles dégradations du composant et déterminer sa robustesse maximale.

Pour réaliser l'impulsion et contrôler sa durée, on utilise une ligne de transmission en câble coaxial d'où le nom de TLP (Transmission Line Pulse). La figure 1.5 représente le schéma électrique du banc de mesure.

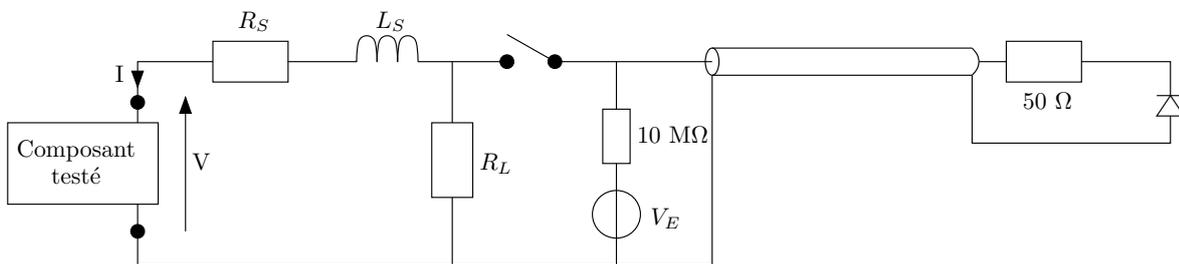


FIG. 1.5 – Schéma électrique du banc de mesure TLP [27].

Dix mètres de câble d'impédance caractéristique de 50Ω permettent d'obtenir une impulsion d'une durée de 100 ns. La ligne est chargée à une tension V_E au travers d'une résistance de très forte valeur. La résistance R_L de 50Ω évite les réflexions dans la ligne. La décharge de tension est convertie en courant par une résistance R_S de 500 à $1 \text{ k}\Omega$. Cette résistance permet également de limiter l'interaction entre le composant à tester et la ligne et d'obtenir un courant stable.

Enfin, une inductance L_S peut être ajoutée en série avec la résistance pour contrôler le temps de montée de l'impulsion [27] d'une façon identique au testeur HBM.

Les caractéristiques de l'impulsion de courant sont proches de celles du modèle HBM. Des signatures de défaillance identiques entre ces stress ont été rapportées pour une durée d'impulsion TLP de 100 ns [28]. Cependant, il n'existe pas de constante de corrélation précise entre ces deux types de stress. La tension HBM maximale (en kV) est généralement comprise entre 1,5 et 2 fois le courant maximum (en A) obtenu par la mesure TLP [29, 30, 31, 32, 33]. D'une manière générale, la corrélation entre les tests ESD dépend des caractéristiques des impulsions, temps de montée et durée, ainsi que des technologies dans lesquelles sont réalisés les composants. Trouver une corrélation entre les différents modèles n'est évidemment possible que dans les cas où ils engendrent le même type de défaillance [34, 35].

Récemment, afin de réaliser un outil de caractérisation plus adapté aux décharges de type CDM, un banc de caractérisation appelé VF-TLP (Very Fast TLP) a été développé [36]. Il permet de générer des impulsions carrées de courant avec un temps de montée inférieur à 500 ps et une durée de 3,5 à 10 ns qui sont comparables aux modèles CDM. Il faut noter cependant qu'il n'est pas envisageable d'obtenir une corrélation entre VF-TLP et CDM puisque ce type de décharge est résolument différent. Le banc VF-TLP apportera seulement des informations sur la dynamique des structures de protection et leur capacité à protéger contre un stress CDM.

1.3.2 La microscopie à émission lumineuse

La microscopie à émission lumineuse, plus communément appelée photoémission, a d'abord été utilisée comme technique performante de localisation et d'analyse de défaillance pour les circuits intégrés [37, 38]. Elle permet également d'observer le fonctionnement des composants.

Cette technique est basée sur la détection de photons associés aux phénomènes physiques dans diverses parties du composant. Deux grandes familles de photons peuvent être distinguées :

- Les photons générés par des collisions de porteurs accélérés par un champ électrique
 - Fuite dans une diode en inverse
 - Diode en inverse (à l'avalanche)
 - Transistor bipolaire : régime non saturé
 - Fuite dans les oxydes
 - Transistor MOS en saturation
- Les photons générés par la recombinaison radiative de paires électron-trou
 - Diode en direct
 - Transistor bipolaire : mode saturé
 - Thyristor à l'état passant et latch-up

De plus, la longueur d'onde de la lumière émise peut donner une indication sur le mécanisme physique sous-jacent comme en témoigne la figure 1.6.

Cette technique, lorsqu'elle est couplée à un générateur d'impulsions TLP, permet d'observer le comportement d'un composant de protection [39, 40], donnant ainsi une information complémentaire sur son fonctionnement à divers niveaux de courant. Pour obtenir un contraste suffisant, l'impulsion TLP doit être répétée à une fréquence qui est généralement de l'ordre de 10 Hz, pendant une durée de quelques minutes (1 à 10 min), afin que la caméra de photoémission accumule un nombre suffisant de photons.

De la même manière, le chemin suivi par le courant de décharge dans un circuit peut être obtenu [41]. Cependant, dans ce cas, seuls les composants de protection fonctionnant dans des régimes d'avalanche seront visibles. En effet, la génération de photons est beaucoup plus importante lors de phénomènes associés à la collision de porteurs accélérés par un champ électrique qu'il ne l'est pour les phénomènes de recombinaison, comme le montre la figure 1.6. La très faible durée d'une impulsion de courant TLP permet difficilement d'observer le fonctionnement d'une diode polarisée en direct alors que les transistors bipolaires dont la jonction collecteur-base est polarisée en inverse seront plus nettement visibles.

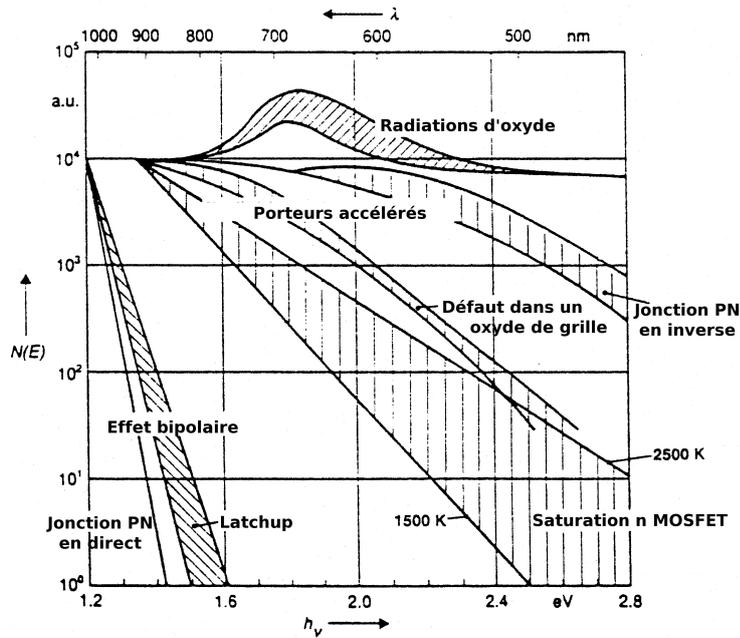


FIG. 1.6 – Spectre des longueurs d'onde d'émission en fonction des mécanismes physiques [37].

1.4 Techniques d'analyse et de localisation de défaillance

L'ensemble des méthodes d'analyse de défaillance courantes sont utiles pour étudier la nature des défauts engendrés par les décharges électrostatiques [42]. La localisation de points chauds par des cristaux liquides, la microscopie optique et électronique, la découpe par les techniques FIB (focus ion beam), la microscopie à force atomique (AFM), en constituent une liste non exhaustive.

Parmi ces techniques, la photoémission est particulièrement adaptée à la localisation de défauts. La technique des cristaux liquides, qui permet de mettre en évidence des points chauds, est rarement efficace pour localiser des défauts créés par les décharges électrostatiques. En effet, les courants de fuite de faible intensité associés au défaut n'engendrent pas d'élévation de température suffisante. Les performances de nouvelles techniques de localisation basées sur l'utilisation des lasers, ont été récemment mises en évidence pour la détection et l'analyse de défauts, même de taille réduite, créés par les décharges électrostatiques [43, 44]. Ces méthodes de localisation présentent l'avantage de ne pas être destructrices.

1.4.1 La photoémission

Cette technique de localisation est bien adaptée aux défauts créés par les décharges électrostatiques [45, 46]. Elle permet de localiser rapidement le lieu de défaillance et de déterminer l'endroit où une analyse de défaillance plus poussée doit être réalisée.

Pour localiser le(s) défaut(s), le composant est polarisé dans une configuration défaillante et placé sous un microscope équipé d'une caméra spécifique pour la photoémission. La technique est particulièrement efficace lorsque le champ électrique est fort dans la région défectueuse. Le contraste peut ainsi être amélioré en augmentant la tension de polarisation, au risque cependant de faire évoluer la nature ou la taille du défaut.

1.4.2 Les techniques de stimulation laser

De nombreuses méthodes de localisation de défauts internes (situés au cœur de la puce de silicium) ont été développées ces dernières années, notamment les techniques basées sur la

microscopie optique à balayage utilisant des lasers infrarouges. Parmi celles-ci, les méthodes utilisant un laser comme source de perturbation du fonctionnement d'un circuit ont connu un succès remarquable en tant qu'outil d'analyse de défaillance [47, 48].

Le faisceau laser est utilisé comme une source d'énergie qui modifie le fonctionnement du circuit sous test. L'énergie déposée est suffisamment faible pour ne pas dégrader le circuit. Le faisceau laser interagit localement avec les différents matériaux qui constituent le circuit intégré qui est lui même détecteur de la perturbation. Deux effets induits par la photoexcitation sont principalement exploités :

- l'échauffement (effet photothermique)
- la génération de paires électron-trou par absorption de photons (effet photoélectrique)

La longueur d'onde du laser utilisé permet de déterminer la nature de l'excitation obtenue. Le minimum d'absorption lumineuse du silicium se situe à une longueur d'onde de $1,1 \mu\text{m}$. Les longueurs d'onde supérieures à $1,1 \mu\text{m}$, qui correspondent à des photons d'énergie inférieure au gap du silicium, sont principalement absorbées par les porteurs libres. Ces porteurs se thermalisent et l'énergie optique est donc convertie principalement en énergie thermique. Pour des longueurs d'onde inférieures à $1,1 \mu\text{m}$, l'absorption est liée au phénomène de génération interbande. L'énergie optique est donc principalement convertie en porteurs libres. Le choix de la longueur d'onde pour la stimulation thermique et photoélectrique repose sur le compromis entre la profondeur d'absorption recherchée et la quantité d'énergie déposée. Les longueurs d'onde proches de $1,1 \mu\text{m}$ sont peu absorbées mais pénètrent profondément dans le substrat. Le choix est également limité par les possibilités des lasers disponibles. Des longueurs d'onde de l'ordre de $1,3 \mu\text{m}$ sont généralement utilisées pour la stimulation thermique alors que la génération photoélectrique est réalisée avec des longueurs d'onde de l'ordre de 900 nm .

La stimulation thermique laser (STL)

Le principe de la stimulation thermique laser repose sur l'échauffement localisé de la surface de silicium. Les effets thermoélectriques vont entraîner une modification des caractéristiques électriques qui va être détectée aux bornes du circuit. Le balayage de la surface de la puce permet de réaliser une cartographie des variations mesurées.

Différentes techniques peuvent être différenciées en fonction de la polarisation appliquée au circuit testé.

- La technique OBIRCH (Optical Beam Induced Resistance Change) consiste à observer les variations spatiales de résistance sous l'effet de l'échauffement du laser [49, 50, 51]. Une tension constante est imposée au circuit. La variation de courant aux bornes du circuit est alors liée à la variation locale de résistance induite par le faisceau laser.
- La technique TIVA (Thermally Induced Voltage Alteration) est l'approche duale à la technique OBIRCH. La variation de résistance est étudiée pour un courant constant imposé au circuit [52, 53]. La variation de tension aux bornes de l'alimentation est alors analysée.
- La technique SEI (Seebeck Effect Imaging). Le composant n'est pas polarisé et seule la tension aux bornes des broches du circuit est mesurée. Les gradients locaux de température induits par le faisceau laser dans les différentes jonctions entre matériaux (métal–métal, métal–semi-conducteur, semi-conducteur–semi-conducteur) induisent des gradients de potentiels par effet Peltier et Seebeck. L'absence de polarisation est un avantage dans cette technique car le(s) défaut(s) ne seront pas aggravés ou modifiés au cours du test.

Les défauts créés par les décharges électrostatiques conduisent à l'apparition de courants de fuite. Ils sont généralement très localisés et d'une taille réduite. Les courants qu'ils entraînent ont des valeurs relativement faibles, de l'ordre ou inférieure au μA . Par la technique OBIRCH, la présence d'un défaut se traduit par une variation du courant de fuite au passage du laser. La cartographie des variations de courant en fonction de la position du laser permet de localiser le(s) défaut(s). Le retraitement informatique de l'information permet de superposer l'image optique du circuit avec l'information de variation du courant. En quelques minutes, la position des

défauts dans le circuit est ainsi obtenue. Cette technique permet de localiser très efficacement des défauts induisant des courants de l'ordre du micro-Ampère, indétectables par la technique de photoémission. Les défauts induisant des courants inférieurs au μA restent cependant indétectables.

La stimulation photoélectrique laser (SPL)

Le principe de la stimulation photoélectrique laser repose sur la création localisée de paires électron-trou dans le silicium. L'interaction du laser avec la puce est plus complexe que dans le cas de la stimulation thermique. Comme pour la stimulation thermique, la technique repose sur la détection de la variation des caractéristiques électriques aux bornes du composant, sous l'effet de l'excitation laser. Une cartographie des variations mesurées est réalisée en balayant la surface du composant à caractériser.

Différentes techniques peuvent être différenciées en fonction de la polarisation appliquée au circuit testé.

- La technique OBIC (Optical Beam Induced Current) où la grandeur électrique analysée est la variation du courant d'alimentation induite par le laser, la tension d'alimentation étant maintenue constante [54, 55].
- La technique LIVA (Light Induced Voltage Alteration), duale de la technique OBIC, consiste à mesurer les variations de tension d'alimentation induites par effet photoélectrique, le courant d'alimentation étant maintenu constant [52, 56].
- La technique NB-OBIC (Non Biased OBIC). Le composant n'est pas polarisé et seul le courant engendré par la photogénération est mesuré.

L'amplitude du signal électrique obtenu dépend de l'intensité du champ électrique dans la région de génération de porteurs provoquée par l'excitation laser. En effet, la présence du champ électrique va engendrer le mouvement des porteurs libres générés et induire l'apparition d'un courant (photocourant). Plus le champ électrique est important, plus le courant sera important. L'amplitude du photocourant dépend également du taux de recombinaison des porteurs dans la région de génération. Plus les recombinaisons sont importantes, plus faible sera l'amplitude du courant observé sur les broches. Enfin, si le champ électrique dans la région de génération est proche du champ de claquage par avalanche, les porteurs pourront être multipliés par avalanche. L'amplitude du courant observé sera alors plus importante.

La présence d'un défaut peut entraîner la modification de la distribution du champ électrique et induire une augmentation du taux de recombinaison des porteurs qui vont influencer l'amplitude du signal électrique.

Les photocourants induits par un faisceau laser (qui possède une faible section) ont une amplitude faible. L'utilisation d'un laser impulsif couplé à l'utilisation d'une détection synchrone permet d'augmenter significativement la sensibilité du dispositif [48]. L'utilisation de cette technique diminue cependant la vitesse de balayage qui est de l'ordre d'une dizaine de minutes. Le schéma de principe du banc de test OBIC à détection synchrone est représenté dans la figure 1.7.

L'étude menée en collaboration avec l'équipe du laboratoire IXL à Bordeaux a permis de montrer, à partir de la simulation numérique corrélée aux résultats expérimentaux, l'impact de la présence d'un défaut de silicium fondu à proximité ou au travers d'une jonction. Ces travaux sont présentés dans la thèse de Thomas Beauchêne [48]. La sensibilité du banc de mesure obtenu grâce à la détection synchrone a permis de caractériser des défauts de faible dimension entraînant des courants de fuite inférieurs au micro-Ampères, de l'ordre du nano-Ampères. Ces défauts étaient indétectables par les techniques de photoémission et de stimulation thermique. Les résultats obtenus montrent principalement que la technique NBOBIC permet de détecter un défaut quelle que soit sa position. Qu'il court-circuite ou non la jonction, le défaut va induire une diminution du signal électrique car il est le siège d'importantes recombinaisons des porteurs. En revanche, lorsque la jonction est polarisée, l'amplitude plus importante du photocourant

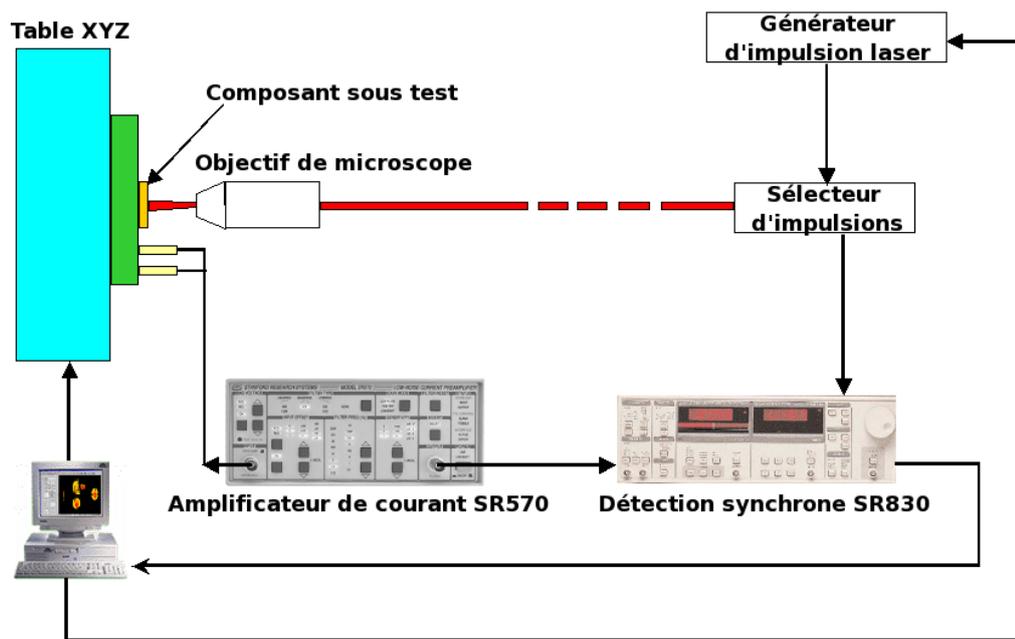


FIG. 1.7 – Schéma de principe du banc de test OBIC [48].

masque souvent l'effet des recombinaisons et la présence d'un défaut qui ne court-circuite pas une jonction n'est pas toujours facilement détectable. Par contre, si le défaut court-circuite la jonction, la déformation du champ électrique qu'il induit va favoriser la génération par avalanche de porteurs et la forte augmentation du courant de photogénération met en évidence sa présence.

1.4.3 Analyse par la face arrière

La principale limitation des techniques de stimulation laser et de photoémission provient de la difficulté à réaliser des analyses par la face avant des circuits. En effet, la complexité croissante des circuits, liée à l'augmentation de la densité et du nombre de niveaux des pistes métalliques en surface, ainsi que l'apparition de nouvelles méthodes d'encapsulation (flip-chip, lead on chip, ...), rendent de moins en moins accessibles les zones actives dans le substrat du circuit. Pour contourner cette limitation, il est possible de réaliser une analyse par la face arrière du composant préalablement amincie [52]. Les techniques de stimulation laser et de photoémission sont compatibles avec ce type d'analyse [57, 48].

1.5 Conclusion

Ce chapitre nous a permis de présenter le phénomène de décharge électrostatique et ses conséquences dans le monde de la microélectronique. La protection des circuits contre les décharges électrostatique constitue aujourd'hui un aspect incontournable de la conception. L'optimisation des performances des composants de protection et le développement de nouveaux outils et méthodes spécifiques d'aide à la conception, sont des enjeux importants pour assurer la compétitivité des fabricants.

Les principaux modèles de décharge ont été décrits, ainsi que les techniques de caractérisation utilisées dans l'industrie pour mesurer la robustesse des circuits intégrés.

Le principe de fonctionnement du banc de mesure TLP qui constitue un outil indispensable de la conception de composants et circuits de protection a été abordé.

Enfin, les techniques de photoémission et de stimulation laser permettent de localiser rapidement le lieu de défaillance d'un circuit ou d'un composant. Les techniques récentes de stimulation

laser appliquées à la localisation de défauts créés par des décharges électrostatique, ont été plus particulièrement décrites. Nous avons également montré comment les résultats obtenus par les techniques OBIC et NBOBIC permettent de caractériser finement la position et la nature d'un défaut de silicium fondu à proximité d'une jonction.

Cet ensemble d'outils plus ou moins spécifiques, constitue la base de l'étude expérimentale des structures de protection ESD. L'objet de cette thèse qui est principalement l'étude des phénomènes physiques relativement complexes et spécifiques aux composants de protection soumis aux décharges électrostatiques, repose sur les informations précieuses qu'ils fournissent.

Chapitre 2

Étude des transistors bipolaires autopolarisés (TBA)

Beaucoup de composants de protection ESD intégrés sur les puces électroniques reposent sur le fonctionnement du transistor bipolaire autopolarisé (TBA). Le transistor MOS à grille couplée ou non, le transistor bipolaire vertical ou latéral en constituent les principaux exemples. Une bonne compréhension de leur fonctionnement est donc requise pour en tirer le meilleur parti, en particulier en termes de robustesse vis-à-vis de la surface de semi-conducteur utilisée. Bien que leur comportement soit depuis quelques années largement étudié [6, 58, 59, 60, 61], il n'est pas apparu jusqu'alors d'explication générale de leur fonctionnement s'appuyant sur une approche physique approfondie et adaptée. Le propos de ce chapitre est donc d'offrir un cadre de réflexion solide pour l'analyse de ces structures, en essayant de mettre en évidence des points clefs pour la compréhension.

À cette fin, nous débuterons par la présentation des composants de protection ESD, basés sur le TBA, les plus répandus. L'état de l'art de leur principe de fonctionnement sera décrit.

Il apparaît, à la lecture des publications dans le domaine, que le défaut majeur des approches adoptées est la non-prise en compte ou de manière superficielle des effets des fortes densités de courant rencontrées lors du fonctionnement. Aussi, nous nous attacherons à baser notre réflexion sur les phénomènes déjà connus et largement analysés, dont la plupart proviennent de l'électronique de puissance et des composants haute fréquence. Cependant, nous verrons que la physique des TBA est très particulière et se situe finalement dans un régime de fonctionnement en second claquage électrique, régime limite et jamais atteint dans les composants classiques, sauf à l'instant de leur destruction. L'extension des approches adoptées par ailleurs en électronique de puissance, nous permettra de décrire simplement ces structures en régime de forte densité de courant.

Nous verrons que le paramètre critique de l'optimisation est l'uniformisation de la répartition de température, et par extension du courant, dans le composant. Les différents phénomènes de focalisation seront étudiés. La clarification des solutions adoptées jusqu'à présent aux problèmes de focalisation ainsi qu'une nouvelle approche de conception, seront finalement présentées.

2.1 Composants de protection ESD basés sur le TBA

2.1.1 Les différents types de composants

Jusqu'à présent les composants de protection intégrés dans le silicium sont réalisés au moyen des différentes couches présentes dans une technologie. Aucun procédé ou étape technologique spécifique à la réalisation de composants de protection ESD n'est disponible, sauf exception, comme dans le cas des drains et sources faiblement dopés (LDD), où une implantation supplémentaire spécifique est *parfois* ajoutée pour améliorer les performances des protections [62]. Les

composants ESD sont donc basés sur les composants de la technologie, auxquels des règles de dessin très spécifiques sont appliquées [6]. Nous allons décrire les plus couramment utilisés.

Les transistors MOS

Dans les technologies CMOS, certains composants de protection sont basés sur les transistors NMOS et PMOS. Le fonctionnement du dispositif ne repose pas sur le transistor MOS mais sur le transistor bipolaire parasite associé. Ce dernier est représenté sur la figure 2.1 dans le cas d'un transistor NMOS. Nous verrons que les transistors PMOS, ou plus exactement les transistors PNP, présentent de moins bonnes performances que les transistors NPN.

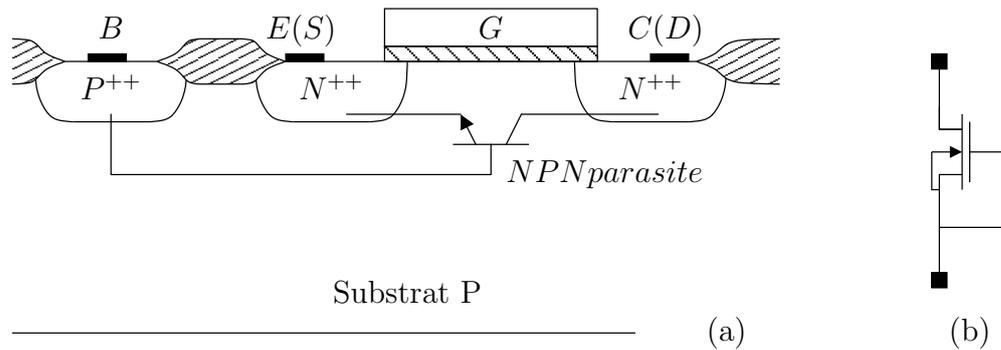


FIG. 2.1 – Vue en coupe du transistor NMOS (a) et sa représentation schématique (b).

Les transistors bipolaires latéraux

Une coupe technologique d'un transistor bipolaire NPN latéral est représentée sur la figure 2.2. Son aspect très similaire au MOS, a conduit à le nommer FOD, pour Field Oxide Device. En effet, la grille du MOS se voit remplacée par un oxyde de champ. Il en découle que pour une technologie donnée, la profondeur de sa base sera très supérieure à celle des structures basées sur les transistors MOS. Cependant, l'absence d'oxyde de grille fragile, le rend souvent plus robuste que ces derniers.

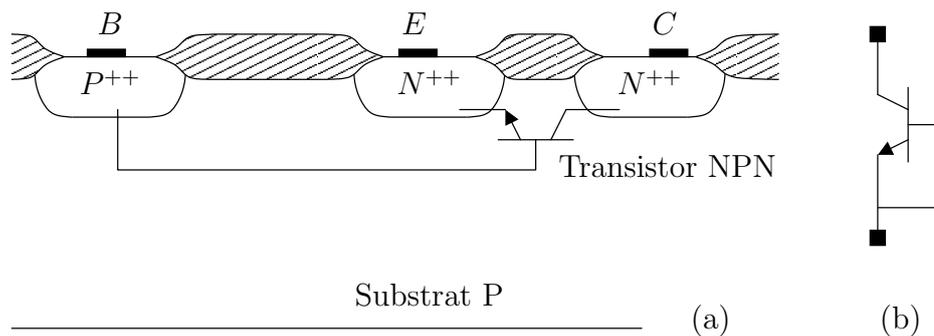


FIG. 2.2 – Vue en coupe du transistor bipolaire FOD (a) et sa représentation schématique (b).

Les transistors bipolaires verticaux

Dans les technologies BiCMOS ou bipolaires, on peut utiliser des transistors bipolaires verticaux. Le plus courant est le transistor NPN réalisé avec une couche enterrée qui permet de prendre le contact de collecteur (Fig.2.3).

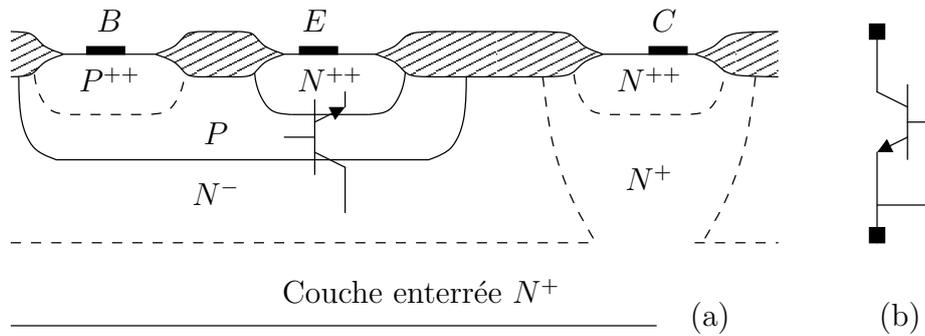


FIG. 2.3 – Vue en coupe du transistor bipolaire NPN vertical (a) et sa représentation schématique (b).

Il existe beaucoup de variantes pour l'ensemble de ces composants en fonction des différentes couches disponibles dans la technologie. Leur fonctionnement repose cependant toujours sur le même principe.

2.1.2 Principe de fonctionnement

Lorsqu'ils sont utilisés en tant que protections ESD, les transistors bipolaires présentés auparavant voient leurs émetteur, base et le cas échéant grille court-circuités. Le composant de protection qu'ils constituent est alors un dipôle qui va pouvoir conduire les courants engendrés par les décharges électrostatiques.

Une caractéristique TLP $I_C(V_C)$ typique est schématiquement présentée sur la figure 2.4 (le potentiel commun d'émetteur, base et grille sert de référence). Elle n'est pas symétrique et l'on distingue deux cas, courant direct et courant inverse (respectivement à gauche et à droite de l'axe des ordonnées) :

- Cas d'un courant direct : Pour un courant direct, vis-à-vis de la jonction collecteur-base, le composant se comporte en diode et la tension à ses bornes est relativement basse. La puissance qu'il dissipe est faible, son échauffement limité. Il pourra conduire des forts courants avant d'être détruit. Ce cas est favorable pour le composant.
- Cas d'un courant inverse : Un courant inverse traverse la jonction collecteur-base. Le composant conduit le courant sous une tension beaucoup plus grande que dans le cas précédent car la jonction collecteur-base est polarisée en inverse. La puissance dissipée est plus importante, ce cas est donc critique.

Décrivons le comportement du composant pour un courant de décharge inverse, par exemple dans le cas d'un stress HBM. Au début de la décharge, la capacité de la jonction collecteur-base se charge, et le potentiel à ses bornes augmente jusqu'à atteindre sa tension de claquage (BV_{CB}). Dans le cas d'un transistor NPN (Fig.2.5), le courant d'avalanche qui apparaît est un courant de trous qui traverse la résistance de substrat R_b^{int} . Lorsque la chute de potentiel dans cette dernière (V_b^{int}) dépasse la tension de seuil directe de la jonction émetteur-base, l'émetteur injecte un courant d'électrons dans la base. On assiste alors au phénomène de repliement de la tension de collecteur : les électrons injectés par l'émetteur sont multipliés par avalanche lorsqu'ils franchissent la jonction collecteur-base, permettant ainsi la diminution de la tension de collecteur nécessaire pour obtenir un courant de trous (issus de l'avalanche) suffisant dans R_b^{int} pour maintenir la jonction émetteur-base en direct. Le courant engendré par la multiplication du courant de collecteur permet donc de polariser la base du transistor bipolaire, d'où l'appellation de transistor bipolaire autopolarisé. Grâce au phénomène de repliement de la tension, le composant dissipe moins d'énergie, et peut ainsi supporter de forts courants. Plusieurs niveaux remarquables de courant et de tension sont présentés dans la figure 2.4 :

- La tension de claquage de la jonction collecteur-base BV_{CB}

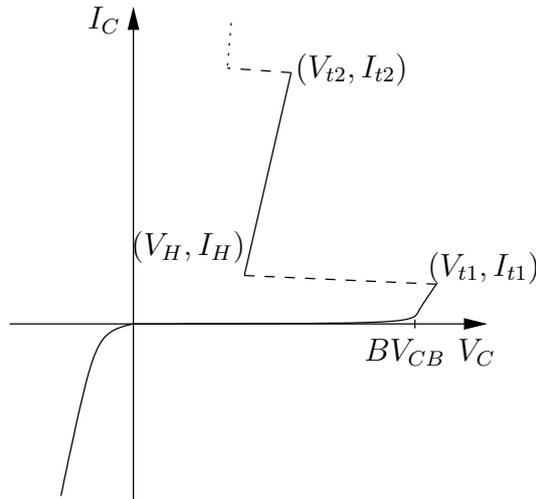
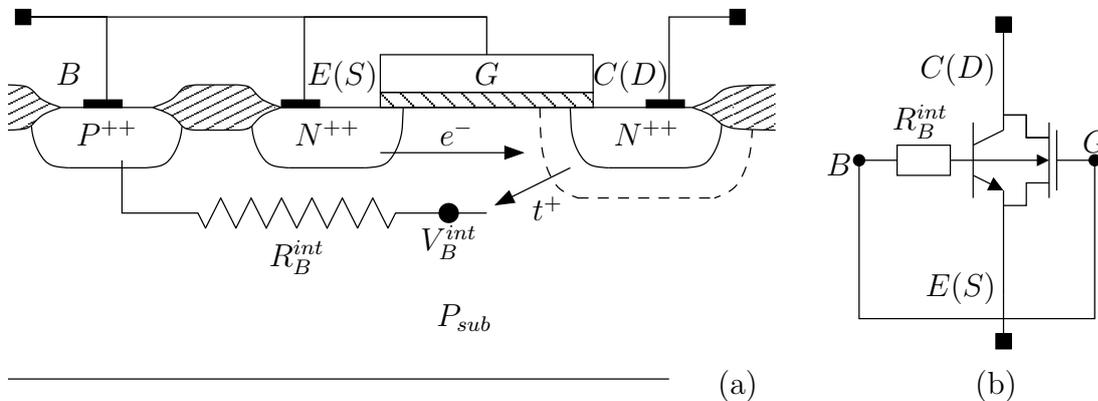

 FIG. 2.4 – Caractéristique $I_c(V_c)$ typique d'un transistor bipolaire NPN autopolarisé.


FIG. 2.5 – Principe de fonctionnement du transistor NPN autopolarisé (a) et schéma électrique équivalent (b).

- Le seuil de repliement du composant (V_{t1}, I_{t1}) (tension et courant de repliement)
- La tension et le courant de maintien lorsqu'il est replié (V_H, I_H) (tension et courant de maintien)
- Le point de second claquage (V_{t2}, I_{t2}) qui indique le maximum de courant que peut supporter le composant, au-delà duquel il est détruit par les phénomènes thermiques

Le fonctionnement d'un transistor PNP est tout à fait similaire, il suffit en particulier d'échanger le rôle des électrons et des trous.

2.1.3 Éléments d'optimisation

Ajustement de la tension de déclenchement

La caractéristique TLP $I_c(V_c)$ d'un composant de protection ESD doit s'insérer dans un gabarit de conception [59], qui assurera une protection efficace des circuits sans interaction parasite avec leurs fonctionnalités. Aucun déclenchement parasite durant le fonctionnement du circuit ne doit avoir lieu. Il convient d'exclure toutes possibilités de déclenchement sur les transitions de tension ou de courant et sur les fréquences des signaux, appliqués sur les broches du circuit et qui correspondent à son fonctionnement normal. Le seuil de déclenchement statique doit être supérieur aux tensions d'utilisations, comme par exemple la tension d'alimentation. Il

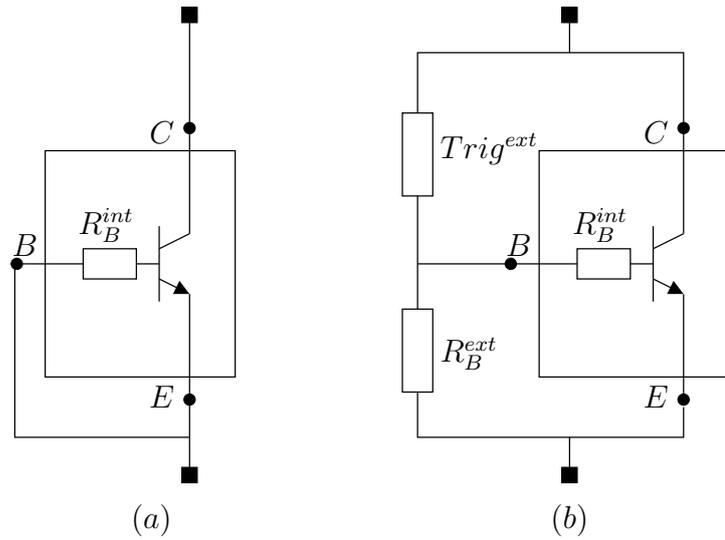


FIG. 2.6 – Schéma équivalent d'un TBA (a) avec circuit de déclenchement associé (b).

faut également éliminer les surtensions, en particulier sur les oxydes de grille des étages d'entrée. Pour respecter l'ensemble de ces contraintes, il est souvent nécessaire d'ajouter un circuit de déclenchement au composant de protection, comme décrit dans la figure 2.6. Cet élément de déclenchement extérieur ($Trig^{ext}$) permet de diminuer la tension maximale sur le collecteur tout en assurant la mise en marche du TBA à travers la résistance R_b^{ext} . Sa taille est faible vis-à-vis du TBA car il n'est actif qu'au tout début de la décharge, l'essentiel du courant circulera dans le TBA. Citons comme exemple de déclencheur une diode en inverse dont la tension de claquage est inférieure à BV_{CB} . Cette diode peut parfois être physiquement intégrée dans le transistor ce qui réduit l'encombrement de l'ensemble [60].

Pour diminuer la tension de déclenchement des TBA basés sur le transistor MOS, il est possible d'utiliser les propriétés propres au transistor MOS. Dans ce cas, on parlera de couplage de grille (ou gate coupling) qui est réalisé par l'ajout d'une résistance entre les contacts de grille et de source du MOS (R_G^{ext} , Fig.2.7). Cette résistance permet l'élévation du potentiel de la

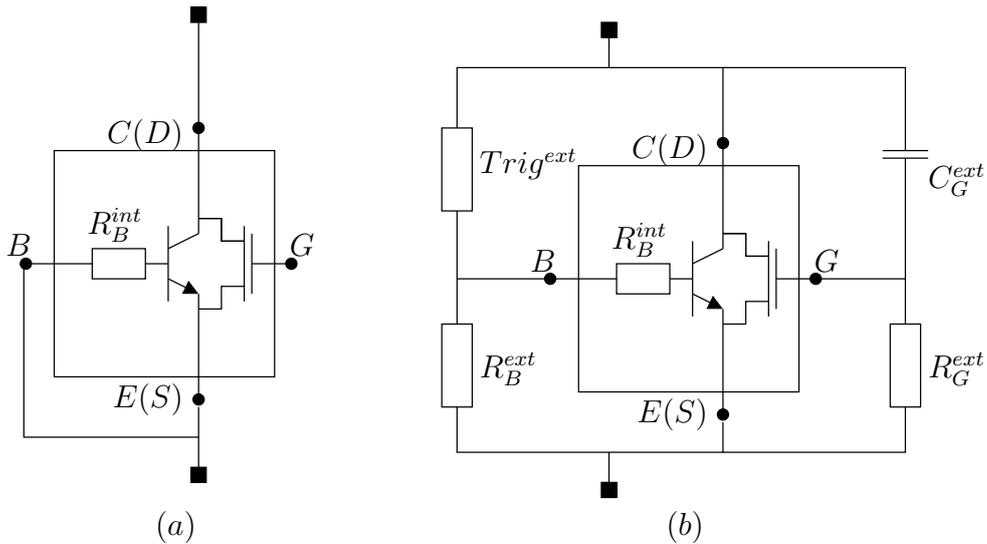


FIG. 2.7 – Schéma équivalent d'un TBA associé au MOS (a) avec un circuit de déclenchement par la base et par couplage de grille (b).

grille au début du transitoire de courant, grâce au couplage capacitif engendré par la capacité intrinsèque grille-drain ou une capacité connectée en externe C_G^{ext} . Lorsque la tension sur la grille dépasse la tension de seuil, le transistor MOS devient passant. Le courant du transistor MOS est alors multiplié à la jonction drain-substrat (ou collecteur-base), ce qui va déclencher le transistor bipolaire, comme décrit précédemment, pour des tensions inférieures à BV_{CB} . Le composant MOS ne devra fonctionner qu'en début de décharge et laisser place au TBA sous peine de voir les performances ESD dégradées [63]. Pour cela, la constante de temps du couple RC, constitué par la capacité de couplage et la résistance associée à la grille, doit être correctement ajustée.

La tension de maintien

L'ajustement du seuil de déclenchement est une part importante de la conception des structures de protection. La robustesse intrinsèque du composant repose cependant essentiellement sur la tension de maintien, qui détermine la puissance dissipée. Par ailleurs, cette tension doit être supérieure à la tension d'alimentation du circuit à protéger, pour éviter tout risque de déclenchement de la structure de protection lors du fonctionnement normal. On peut estimer la valeur de la tension de maintien en se basant sur une étude simple et unidimensionnelle [6, 59]. La figure 2.8 est une représentation mêlant la structure physique d'un transistor bipolaire NPN

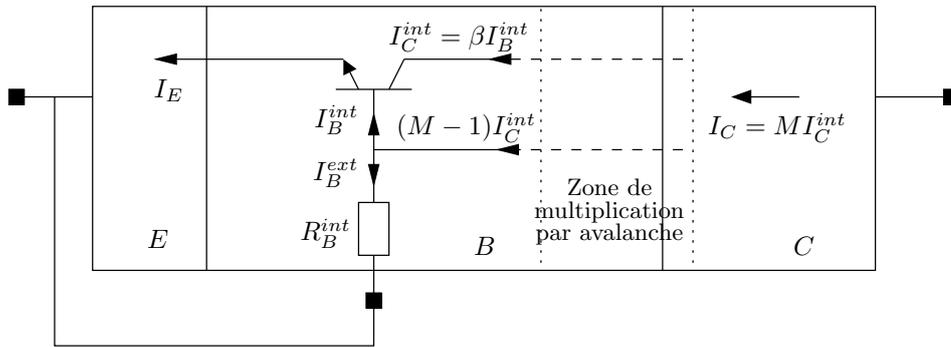


FIG. 2.8 – Modèle unidimensionnel du transistor NPN autopolarisé.

et son schéma électrique. Deux différences majeures apparaissent par rapport au fonctionnement «classique» du composant :

- la zone de multiplication par avalanche à la jonction collecteur-base
- la résistance interne de base

qui permettent en particulier l'autopolarisation du composant. Le courant de collecteur associé à l'effet bipolaire est donné par la relation classique :

$$I_C^{int} = \beta I_B^{int} \quad (2.1)$$

avec β le gain en courant du transistor bipolaire. Ce courant d'électrons est multiplié dans la zone d'avalanche pour donner le courant qui traverse effectivement le contact de collecteur $I_C = M I_C^{int}$, avec M le facteur de multiplication du courant, qui dépend de la tension appliquée à cette zone et de la tension de claquage de la jonction collecteur-base. La multiplication du courant de collecteur donne naissance à un courant de trous égal à $(M - 1) I_C^{int}$, qui se répartit entre le courant de base du transistor I_B^{int} et le courant I_B^{ext} , sortant du contact de base, qui polarise la base «interne» au travers de la résistance R_B^{int} . On a donc :

$$I_B^{int} + I_B^{ext} = (M - 1) I_C^{int} \quad (2.2)$$

Le courant I_B^{ext} devient vite négligeable par rapport à I_B^{int} lorsque le courant augmente. En effet, ces deux courants sont liés par la relation :

$$I_B^{int} = I_S \left(\exp \left(\frac{R_B^{int} I_B^{ext}}{u_T} \right) - 1 \right) \quad (2.3)$$

qui est issue de l'expression classique du courant due aux trous dans une jonction N^+P en direct [64], avec I_S le courant de saturation des trous à la jonction émetteur-base. Plus la résistance intrinsèque de base R_B^{int} est grande, plus I_B^{ext} est négligeable devant I_B^{int} . En pratique, R_B^{int} a toujours une valeur relativement importante. Dans le cas contraire, on n'obtiendrait le déclenchement du transistor bipolaire que pour de fortes valeurs du courant I_{t1} , ce qui réduirait les performances de la structure car l'échauffement de la jonction collecteur-base, polarisée en inverse, serait trop important.

De plus, la relation 2.3 montre que le courant de base interne croît très vite lorsque le courant externe augmente, ce qui implique la rapide prépondérance du courant interne sur le courant externe pour des densités de courant croissantes dans la structure. Le comportement du composant tend donc vers un fonctionnement équivalent à un transistor en base ouverte. Le calcul de la tension de maintien est donc identique à celui de la tension de claquage émetteur-collecteur avec la base ouverte (BV_{CEO}), pour un transistor bipolaire classique [65, 66]. Ainsi, en négligeant I_B^{ext} dans la relation 2.2 et en considérant l'équation 2.1, on déduit l'expression régissant la tension de maintien :

$$(M - 1)\beta = 1 \quad (2.4)$$

En utilisant la formule empirique de Miller [67] pour l'expression du facteur de multiplication :

$$M = \frac{1}{1 - \left(\frac{V}{BV} \right)^m} \quad (2.5)$$

avec ($2 < m < 6$) en fonction du type de jonction, BV la tension de claquage de la jonction et V la tension appliquée. Si on néglige les chutes de tension à la jonction émetteur-base et dans les résistances de collecteur et d'émetteur, on obtient la valeur de la tension de maintien :

$$V_H = \frac{BV_{CB}}{(1 + \beta)^{1/m}} \quad (2.6)$$

avec BV_{CB} la tension de claquage de la jonction collecteur-base. La tension de maintien apparaît comme une fonction du gain du transistor : plus il sera élevé, plus elle sera basse. Elle est, de plus, proportionnelle à la tension de claquage de la jonction collecteur-base. Le concepteur de structures de protection devra donc choisir un transistor bipolaire dont la tension BV_{CB} est proche de la tension de maintien recherchée (tout en restant compatible avec la tension de déclenchement voulue). La valeur de V_H pourra être ajustée par le contrôle du gain qui dépend en particulier de la profondeur de la base, facilement ajustable pour les composants MOS et FOD.

Pour un transistor PNP autopolarisé, l'expression de la tension de maintien s'obtient de la même façon et est identique à celle du transistor NPN. Cependant, à l'inverse du transistor bipolaire NPN, c'est un courant de trous qui est multiplié dans la zone d'avalanche et qui donne naissance au courant de base qui est alors un courant d'électrons. On a donc respectivement dans le cas d'un transistor bipolaire PNP et NPN :

$$I_n = (M_p - 1)I_p \quad \text{et} \quad I_p = (M_n - 1)I_n \quad (2.7)$$

¹Remarquons qu'une résistance de base externe, ajoutée pour relier base et émetteur, joue le même rôle que la résistance interne. Sa valeur vient simplement s'ajouter à la valeur de cette dernière, sans modification supplémentaire du comportement du composant.

pour les courants injectés dans la base, en fonction du courant entrant dans la zone d'avalanche, avec I_n un courant d'électrons, I_p un courant de trous et M_p et M_n les facteurs de multiplication d'un courant de trous et d'un courant d'électrons, respectivement.

Pour des transistors bipolaires NPN et PNP équivalents, le facteur de multiplication du courant de trous sera toujours inférieur à celui du courant d'électrons à tensions égales [68, 69]. Ceci se traduit dans la formule de Miller (Equ.2.5), par une valeur du paramètre m plus grande dans le cas d'un transistor PNP. La tension de maintien est donc toujours plus grande dans ce type de transistor par rapport à un transistor NPN ayant des paramètres physiques et géométriques équivalents, d'autant plus que le gain est généralement plus faible pour un transistor PNP.

L'échauffement provoqué par leur tension de maintien plus importante rend les transistors bipolaires PNP moins robustes. Aussi, ils ont été très peu utilisés jusqu'à présent. Il semble cependant qu'ils présentent de bonnes caractéristiques dans des technologies plus avancées [70].

L'approche qui vient d'être présentée montre l'essentiel des outils théoriques utilisés à l'heure actuelle pour la conception des TBA. Son défaut majeur est de laisser de côté les effets des fortes densités de courant présentes dans ces composants lors des décharges électrostatiques. Certains aspects des fortes densités de courant plus particulièrement centrés sur les diodes et les résistances diffusées, sont traités par G. Bosselli [71]. Nous proposons ici une étude plus spécifique, axée sur les TBA.

2.2 Effets des fortes densités de courant dans les transistors bipolaires

Les décharges électrostatiques sont des transitoires de courant de forte intensité, de l'ordre de quelques ampères pour les modèles HBM et MM et jusqu'à quelques dizaines d'ampère pour le CDM. La taille des protections ESD dans lesquelles transitent ces courants est extrêmement réduite comparée à celle de composants utilisés en électronique de puissance. Évidemment, l'énergie dissipée dans un composant de puissance, pendant quelques heures d'utilisation, est de plusieurs ordres de grandeur supérieure à celle dissipée lors d'une décharge électrostatique dont la durée est extrêmement courte, de l'ordre de la centaine de nanoseconde. Il n'en reste pas moins que les densités de courants rencontrées dans un transistor de protection, jusqu'à quelques $10^6 A/cm^2$ sont gigantesques vis-à-vis de celle des composants de puissance qui sont de l'ordre de $200 A/cm^2$. Les composants pour l'électronique haute fréquence fonctionnent à de très fortes densités de courants, mais les technologies employées mettent en jeu des dopages de valeur élevée qui retardent les effets des fortes densités de courant à des niveaux de quelques $10^5 A/cm^2$ [72]. La connaissance des phénomènes spécifiques aux fortes densités de courant dans ces deux domaines, l'électronique de puissance et haute fréquence, facilite l'accès à la compréhension du comportement des dispositifs de protection contre les ESD.

2.2.1 Chute du gain en courant

La chute du gain des transistors bipolaires à forte densité de courant est un phénomène très connu en microélectronique et n'est pas spécifique à un domaine d'application particulier. C'est un des principaux paramètres qui conditionne le dimensionnement des composants dans les étages d'amplification de puissance basse fréquence. La chute du gain correspond à l'entrée dans un régime de forte injection, où la densité des porteurs minoritaires dans la base n'est plus négligeable par rapport à son dopage. Le gain décroît alors comme l'inverse du courant. Ce phénomène est couramment traité dans les ouvrages généralistes portant sur la physique de composants actifs de la microélectronique [64, 73, 74].

2.2.2 Limitation de l'aire de sécurité d'utilisation, dans les transistors bipolaires pour l'électronique de puissance

Longtemps utilisé comme composant semi-conducteur permettant la commutation de puissance, le transistor bipolaire se voit aujourd'hui remplacé par les transistors MOS, les IGBT et les thyristors GTO. Les études approfondies le concernant ont été principalement réalisées avant les années 1980. De ces avancées aujourd'hui un peu oubliées, on apprend beaucoup sur le comportement du transistor bipolaire en particulier dans ses régimes limites de fonctionnement.

Phénomène de second claquage thermique

Hormis les limitations classiques, telle l'hyperbole de dissipation qui garantit des températures assurant le bon fonctionnement du composant et sa fiabilité à long terme, deux mécanismes d'instabilité limitant le domaine d'utilisation du composant peuvent être distingués : l'un purement thermique, l'autre de nature purement électrique. Ces deux mécanismes conduisent à la destruction du composant par la focalisation (ou filamentation) du courant [75, 76]. En effet, lorsque la température s'élève localement dans le composant au point que la concentration intrinsèque de porteurs dépasse celle des porteurs injectés, et se substitue à celle de ces derniers pour assurer le passage du courant, on assiste à une chute brutale de la tension aux bornes du dispositif et à la concentration du courant dans une zone extrêmement localisée. Ce processus est appelé second claquage thermique, bien que le qualificatif second n'ait pas aujourd'hui de réels fondements, autres qu'historiques [77]. L'emballement thermique associé conduit à la fusion locale du silicium et à la destruction du composant.

Instabilité d'origine thermique

Si l'issue des deux types d'instabilité est le claquage thermique, leur origine est fondamentalement différente. L'instabilité de type thermique provient de la non-uniformité de la distribution de température dans le composant. Le gain d'un transistor étant une fonction croissante de la température, le courant tend à se localiser dans les zones de plus haute température [78]. L'augmentation locale de la densité de courant va encore échauffer cette zone et confiner d'autant plus le courant, ce qui à terme conduit au processus de second claquage thermique. La solution de ce problème consiste principalement à disposer un dissipateur thermique sur le composant, en prenant particulièrement soin à l'uniformité de l'évacuation de la chaleur permettant d'éviter les déséquilibres de température. La réalisation de résistance de ballast d'émetteur est également utilisé pour contrôler ce phénomène.

Instabilité d'origine électrique

Le second type d'instabilité est associé aux fortes densités de courant dans le composant. La densité des porteurs mobiles dans la zone de charge d'espace de la jonction collecteur-base modifie la répartition du champ électrique et la tenue en tension. En effet, au-delà de densités de courant critiques, la densité des porteurs libres qui assurent le passage du courant dans la zone dépeuplée de la jonction collecteur-base, n'est plus négligeable face à celle des impuretés ionisées. La densité de charges électriques dépend non seulement de celle portée par les dopants, mais aussi des charges associées aux porteurs libres. La répartition du champ électrique est donc modifiée, ainsi que la tension de claquage de la jonction [79]. Sous de très fortes densités de courant, le maximum du champ électrique peut être déplacé contre la couche fortement dopée de collecteur (Fig.2.9), menant à la définition d'une jonction délocalisée dont les caractéristiques sont très différentes de la jonction métallurgique collecteur-base. La tenue en tension du transistor est alors faible car la tension de claquage de la jonction délocalisée diminue rapidement lorsque la densité de courant augmente, ce qui limite le domaine d'utilisation du composant. Ce phénomène a été décrit pour la première fois en 1966 dans [80] où il a été dénommé second claquage électrique, car il est

associé, comme dans le cas du second claquage thermique, à une brutale diminution de la tension associée à l'apparition de la multiplication par avalanche. Une étude plus approfondie a été menée dans [81], mettant en avant le lien entre les caractéristiques électriques et les paramètres physiques et structurels du transistor.

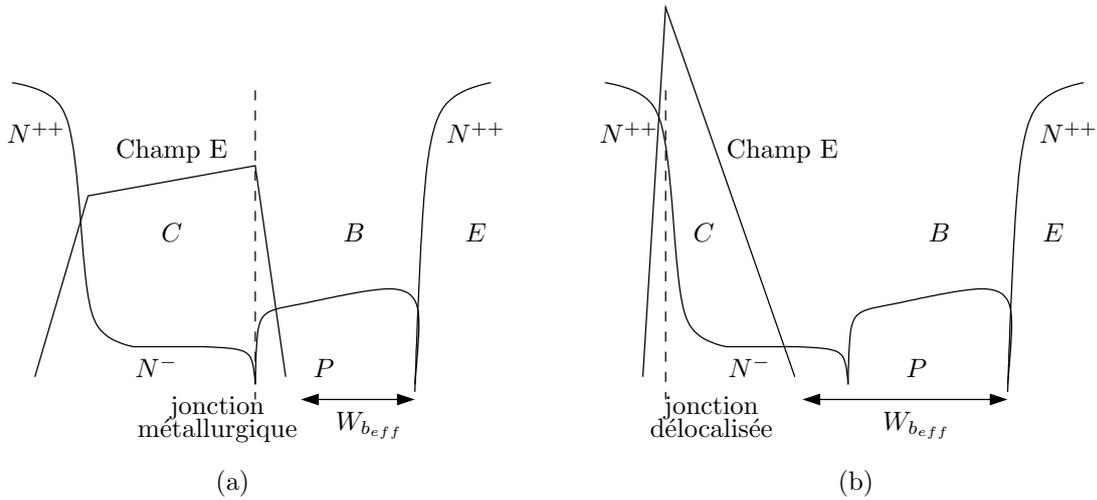


FIG. 2.9 – Profil de dopage typique dans un transistor bipolaire NPN, superposé au profil de champ électrique, avant apparition de l'effet Kirk ou de second claquage électrique (a) et après leur apparition (b).

2.2.3 Effet Kirk, dans les transistors bipolaires pour l'électronique haute fréquence

Découvert en 1962, bien avant le second claquage électrique, l'effet Kirk repose lui aussi sur la prise en compte de la charge électrique des porteurs qui assurent le passage du courant dans la charge d'espace de la jonction collecteur-base d'un transistor bipolaire (qui est polarisée en inverse dans la plupart des applications). La première étude sur le second claquage électrique se base d'ailleurs sur les résultats présentés par Kirk [82]. Comme dans le cas du second claquage électrique présenté au paragraphe précédent, l'effet de la charge électrique des porteurs, non négligeable à forte densité de courant, induit le déplacement du maximum de champ électrique de la jonction métallurgique collecteur-base à une jonction délocalisée située à la transition du collecteur faiblement et fortement dopé. Dans le même temps, la profondeur effective de la base du transistor (Fig.2.9) se voit augmentée, entraînant l'augmentation du temps de transit dans celle-ci et la chute de la fréquence de transition du transistor.

2.2.4 Cas des TBA

Les deux phénomènes, effet Kirk et second claquage électrique, reposent sur le même effet dû aux porteurs libres dans la zone de charge d'espace de la jonction collecteur-base. Leur différence ne repose que sur les conséquences de cet effet : l'élargissement de la base qui entraîne la chute de la fréquence de transition dans le cas de l'effet Kirk, la diminution de la tension maximale supportable à la jonction collecteur base pour le second claquage électrique. Toutes deux définissent des limites d'utilisation du transistor bipolaire dans ces domaines.

Nous allons voir que le régime de fonctionnement des TBA utilisés en ESD se situe au-delà de ces limites. En effet, l'élargissement de la base et la modulation de la tension de claquage de la jonction collecteur-base par les fortes densités de courant se trouvent associés au phénomène de multiplication par avalanche indispensable à l'autopolarisation de ces composants.

2.3 Éléments d'une approche régionale unidimensionnelle

Lors de stress ESD, les composants sont soumis à de fortes densités de courant, même au sens de l'électronique de puissance. Les équations utilisées pour les régimes classiques de fonctionnement dans les domaines comme le traitement du signal ou l'amplification, ne sont pas adaptées pour ces régimes. La physique des dispositifs de l'électronique de puissance est la plus utile à l'étude des TBA aux fortes densités de courant. Aussi, nous baserons principalement notre étude sur les approches et outils qui ont été développés dans ce domaine [83, 84, 85]. L'étude suivante repose sur d'importantes approximations, la complexité des phénomènes rencontrés ne permettant pas de traitement analytique simple. Elle met cependant en évidence les phénomènes clés et l'influence des paramètres structuraux et géométriques sur le comportement des composants, ce qui présente un grand intérêt pédagogique et pratique.

Nous allons considérer un composant élémentaire unidimensionnel en régime stationnaire, dans lequel les aspects thermiques ne seront pas pris en compte. L'étude repose sur un calcul semi-analytique de la densité des porteurs libres dans le composant, s'appuyant sur une approche régionale. Comme cela a pu être justifié au paragraphe 2.1.3, le courant sortant du contact de base sera négligé ce qui facilite l'approche unidimensionnelle. Notons que l'aspect stationnaire est riche d'information à l'instar de la caractéristique quasi-statique TLP.

L'approche régionale de l'étude des semi-conducteurs s'appuie sur la décomposition des structures en régions délimitées par des frontières abruptes dans lesquelles on peut consentir des approximations pour les équations fondamentales de transport de charge. Ces régions sont qualifiées d'électriques par opposition aux régions physiques (dopage) auxquelles elles ne correspondent pas forcément. En première approximation, les transistors bipolaires peuvent être découpés en trois régions électriques distinctes : la région d'émetteur, de base et de collecteur (Fig.2.10). La principale originalité des TBA réside dans la présence de multiplication par avalanche dans la zone de collecteur, indispensable pour fournir le courant de base.

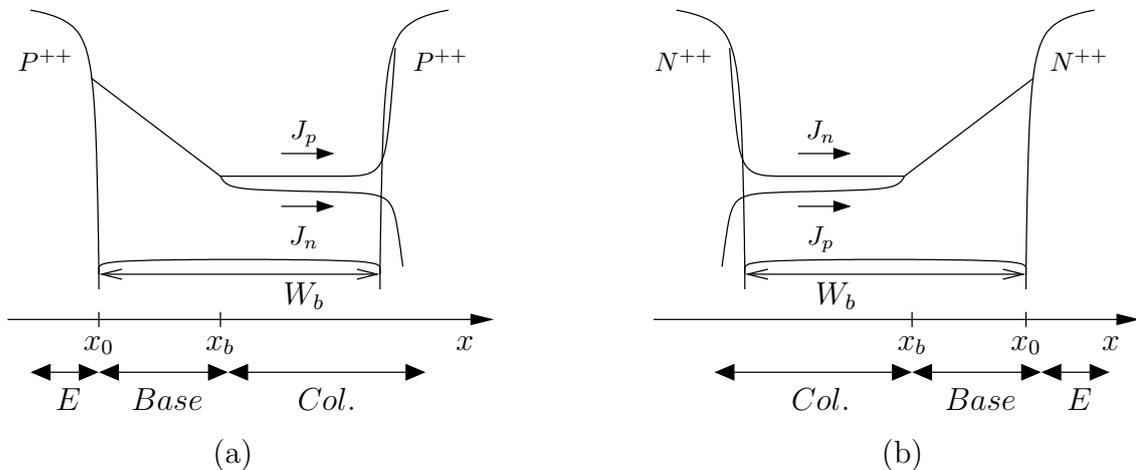


FIG. 2.10 – Profils des porteurs libres dans la base physique de largeur W_b : (a) dans le cas d'un transistor PNP et (b) dans le cas d'un transistor NPN. Les régions électriques sont également représentées.

2.3.1 Comportement de l'émetteur

La jonction émetteur-base qui se trouve polarisée en direct se comporte idéalement comme un injecteur de porteurs minoritaires dans la base. Il existe cependant un courant de diffusion de porteurs minoritaires dans l'émetteur qui est la principale composante du courant de base lors du fonctionnement classique du transistor bipolaire. Ce courant s'exprime commodément

au moyen du coefficient de recombinaison d'émetteur h_E par l'équation [86, 87] :

$$J_{n,p_{x_0}} = qh_E[(pn)_{x_0} - n_i^2] \quad (2.8)$$

Cette équation donne la densité de courant de trous (J_p) et d'électrons (J_n) à la jonction émetteur-base, respectivement dans le cas des transistors bipolaires NPN et PNP. Elle se simplifie à forte densité de courant, lorsque la base quasi-neutre est en régime de forte injection ($n \approx p \gg n_i$) pour donner l'équation ² :

$$J_{n,p_{x_0}} = qh_E p_{x_0}^2 \quad (2.9)$$

Il est remarquable que le coefficient de recombinaison d'émetteur h_E est peu dépendant du niveau de courant et des caractéristiques de l'émetteur lorsqu'il est fortement dopé ($N > 10^{18} \text{ cm}^{-3}$) [88]. Ceci provient de la conjonction d'effets des forts dopages (recombinaison Auger, diminution apparente de la bande interdite, réduction de la mobilité) qui tend à réduire l'influence des paramètres de structure : profondeur, dopage moyen, profil... [89]. Dans la littérature, les valeurs de h_E se situent toujours dans une gamme étroite de valeurs, comprises entre 1 et $3 \cdot 10^{-14} \text{ cm}^4/\text{s}$.

2.3.2 Comportement de la région de base

Caractéristiques d'une région quasi-neutre en forte injection

L'élimination du champ électrique entre les expressions habituelles des densités de courant de trous et d'électrons dans le cas unidimensionnel :

$$J_n = qn\mu_n E + qD_n \frac{\partial n}{\partial x} \quad \text{et} \quad J_p = qp\mu_p E - qD_p \frac{\partial p}{\partial x} \quad (2.10)$$

ainsi que les hypothèses de quasi-neutralité et de forte injection, conduisent à la relation [83] :

$$\frac{J_n(x)}{D_n} - \frac{J_p(x)}{D_p} \approx 2q \frac{\partial p}{\partial x} \quad (2.11)$$

J_n et J_p désignant respectivement les densités de courant d'électrons et de trous, D_n et D_p les coefficients de diffusion des électrons et des trous, q la charge électrique élémentaire.

En régime stationnaire, le courant de déplacement est nul et l'expression du courant total J_T s'écrit dans la base :

$$J_T = J_n(x) + J_p(x) \quad (2.12)$$

On peut alors exprimer les densités de courant J_n et J_p en fonction de la densité de courant totale et du gradient des porteurs à partir de l'équation (2.11). On obtient ainsi :

$$J_n(x) = \frac{D_n}{D_n + D_p} J_T + qD \frac{\partial p}{\partial x} \quad \text{et} \quad J_p(x) = \frac{D_p}{D_n + D_p} J_T - qD \frac{\partial p}{\partial x} \quad (2.13)$$

avec :

$$D = 2 \frac{D_n D_p}{D_n + D_p} \quad (2.14)$$

le coefficient de diffusion ambipolaire.

Sachant que pour le silicium $D_n \approx 3D_p$ on a :

$$J_n(x) = \frac{3}{4} J_T + qD \frac{\partial p}{\partial x} \quad \text{et} \quad J_p(x) = \frac{1}{4} J_T - qD \frac{\partial p}{\partial x} \quad (2.15)$$

²La forte injection et la quasi neutralité entraînant l'égalité des concentrations de porteurs dans la région de base $n \approx p$, il est d'usage de désigner cette concentration par p , quel que soit le type de composant considéré.

Les densités de courant J_n et J_p ne pouvant évidemment pas être supérieures à la densité de courant totale J_T , on déduit des équations précédentes un encadrement pour les valeurs possibles du gradient de concentration des porteurs :

$$\frac{-3}{4qD}J_T < \frac{\partial p}{\partial x} < \frac{1}{4qD}J_T \quad (2.16)$$

On remarquera de plus que $J_n = J_p$ pour :

$$\frac{\partial p}{\partial x} = \frac{-1}{4qD}J_T \quad (2.17)$$

Dans le cadre des conventions fixées à la figure 2.10, la figure 2.11 représente la variation des densités de courant d'électrons et de trous en fonction du gradient de concentration des porteurs, données par les équations (2.15). Le cas des transistors NPN correspond aux valeurs positives du gradient (à droite de l'axe des ordonnées), celui des PNP aux gradients négatifs (à gauche de l'axe).

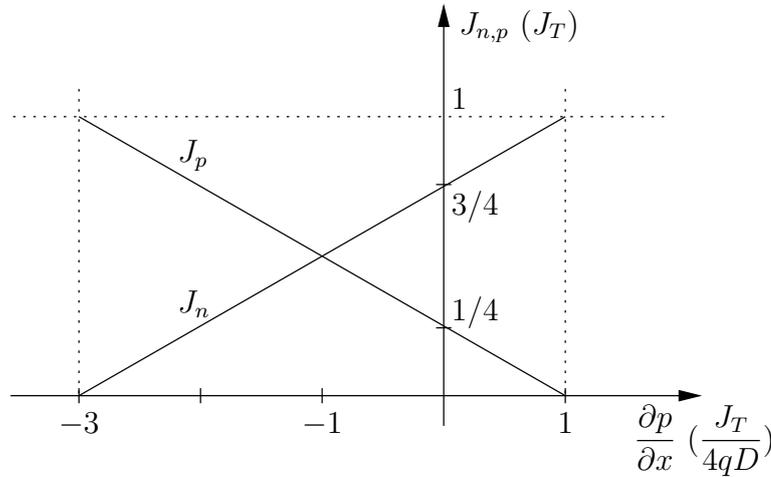


FIG. 2.11 – Évolution des densités de courant de trous et d'électrons en fonction du gradient de porteur.

Connaissant la valeur du courant total, les équations (2.15) permettent donc de déterminer en tout point de la base, la part du courant d'électrons et du courant de trous au courant total, en fonction de la valeur de la dérivée de la concentration de porteurs en ce point. La figure 2.11 met en évidence une dissymétrie du comportement des transistors bipolaires NPN et PNP, qui va induire une différence marquée du comportement des deux types de transistor.

Équation régissant le profil des porteurs

Les hypothèses de quasi-neutralité et de forte injection conduisent à l'expression de l'équation de diffusion ambipolaire [83] :

$$D \frac{\partial^2 p}{\partial x^2} = \frac{p}{\tau} + \frac{\partial p}{\partial t} \quad (2.18)$$

avec τ la durée de vie des porteurs. Cette équation gouverne d'une manière générale la répartition des porteurs dans la base. En régime stationnaire, on obtient :

$$\frac{\partial^2 p}{\partial x^2} = \frac{p}{D\tau} \quad (2.19)$$

ou :

$$\frac{\partial^2 p}{\partial x^2} = \frac{p}{L^2} \quad (2.20)$$

avec $L = \sqrt{D\tau}$ la longueur de diffusion ambipolaire.

Comportement de la région de base d'un TBA

Pour des profondeurs de base (W_b) faibles vis-à-vis de L , on pourra négliger, en première approximation, les recombinaisons dans la base. La dérivée seconde du profil de porteurs peut alors être considérée comme nulle et donc sa dérivée constante, ce qui implique un profil linéaire des porteurs dans la région de base. Cette approximation est justifiée pour les composants de protection ESD qui ont des profondeurs de base faibles. Les équations (2.15) indiquent dans ce cas que les courants d'électrons et de trous sont constants sur toute la profondeur de la base et ont donc les mêmes valeurs à chacune des deux extrémités (en x_0 et x_b Fig.2.10). Indépendamment du type du transistor, on a $J_{n_{x_0}} = J_{n_{x_b}}$ et $J_{p_{x_0}} = J_{p_{x_b}}$. De plus, la valeur de ces courants est donnée par les équations (2.15) lorsque l'on connaît la densité de courant totale et la valeur de la dérivée du profil de porteurs.

Étant donné le profil linéaire des porteurs, la valeur de sa dérivée peut aussi être calculée connaissant la concentration de porteurs aux extrémités de la base, du côté émetteur (p_{x_0}) et collecteur (p_{x_b}). On a alors :

$$\frac{\partial p}{\partial x} = \frac{p_{x_0} - p_{x_b}}{x_0 - x_b} \quad (2.21)$$

2.3.3 Comportement de la région de collecteur

Dans cette région, les courants sont essentiellement des courants de conduction. Le fort champ électrique qui y règne implique que les porteurs transitent à leur vitesse limite ³ $v_{ln} = 1,0710^7 cm/s$ pour les électrons et $v_{lp} = 8,3710^6 cm/s$ pour les trous [90]. Il est courant de considérer que les porteurs se déplacent à leur vitesse limite dans la totalité de la zone, ce qui permet d'établir une relation directe entre les courants et les densités de porteurs :

$$n(x) = \frac{J_n(x)}{qv_{ln}} \quad \text{et} \quad p(x) = \frac{J_p(x)}{qv_{lp}} \quad (2.22)$$

La valeur de la charge d'espace qui donne naissance au champ électrique est alors donnée localement par la relation :

$$\rho(x) = q(N(x) - \frac{J_n(x)}{qv_{ln}} + \frac{J_p(x)}{qv_{lp}}) \quad (2.23)$$

avec N la valeur du dopage, positive pour un type N et négative pour un type P.

Au-delà d'une valeur critique de ces courants, la charge électrique due aux porteurs libres n'est plus négligeable vis-à-vis de celle des impuretés ionisées :

$$J_n^{crit} = |N|qv_{ln} \quad \text{et} \quad J_p^{crit} = |N|qv_{lp} \quad (2.24)$$

Ces seuils sont d'autant plus bas que la zone est faiblement dopée. Au-delà de ces valeurs, la densité de charge électrique (ρ), et donc le comportement de la jonction collecteur-base, dépendront principalement de la valeur des densités de courant.

En particulier, la valeur du facteur de multiplication par avalanche, qui dépend essentiellement des profils et des valeurs de dopage à faible courant, sera fortement influencée. Dans le cas d'un transistor bipolaire NPN, le facteur de multiplication M_n est associé à un courant d'électrons, contrairement à un courant de trous pour celui M_p d'un PNP :

$$M_p = \frac{1}{1 - \mathfrak{S}_p} \quad \text{et} \quad M_n = \frac{1}{1 - \mathfrak{S}_n} \quad (2.25)$$

avec \mathfrak{S}_n et \mathfrak{S}_p les intégrales d'ionisation :

$$\mathfrak{S}_n = \int_0^W \alpha_n \exp\left(\int_W^x (\alpha_n - \alpha_p) dx'\right) dx \quad \text{et} \quad \mathfrak{S}_p = \int_0^W \alpha_p \exp\left(\int_0^x (\alpha_n - \alpha_p) dx'\right) dx \quad (2.26)$$

³La saturation de la vitesse de dérive des porteurs est due à la réduction de leur mobilité pour des champs supérieurs à $10^4 V/cm$.

intégrales dans la zone de charge d'espace d'extension W , avec α_n et α_p les coefficients d'ionisation des porteurs qui ne dépendent que du champ électrique E :

$$\alpha_n = \alpha_n^\infty \exp\left(-\frac{E_n^{crit}}{E}\right) \quad \text{et} \quad \alpha_p = \alpha_p^\infty \exp\left(-\frac{E_p^{crit}}{E}\right) \quad (2.27)$$

avec :

- $\alpha_n^\infty = 7.10^5 \text{ cm}^{-1}$, $E_n^{crit} = 1,2.10^6 \text{ V/cm}$ pour les électrons
- $\alpha_p^\infty = 6,7.10^5 \text{ cm}^{-1}$, $E_p^{crit} = 1,7.10^6 \text{ V/cm}$ pour les trous.

L'approximation des coefficients d'ionisation par des expressions en E^7 [91] :

$$\alpha_n = A_n E^7 \quad \text{et} \quad \alpha_p = A_p E^7 \quad (2.28)$$

avec :

- $A_n = 3,6.10^{-35}$
- $A_p = 0,3.10^{-35}$

permet de calculer analytiquement les intégrales d'ionisation dans le cas d'une jonction plane abrupte [92] :

$$\mathfrak{S}_n = \frac{A_p}{A_p - A_n} \left[\exp\left(2(A_p - A_n) \left(\frac{qC_{eff}}{\varepsilon_{Si}}\right)^3 V_{av}^4\right) - 1 \right] \quad (2.29)$$

$$\mathfrak{S}_p = \frac{A_p}{A_p - A_n} \left[1 - \exp\left(2(A_n - A_p) \left(\frac{qC_{eff}}{\varepsilon_{Si}}\right)^3 V_{av}^4\right) \right] \quad (2.30)$$

avec $C_{eff} = \frac{N_a N_d}{N_a + N_d}$ la concentration effective de la jonction (N_a et N_d les valeurs (positives) des dopages côté P et côté N respectivement), V_{av} la tension à ses bornes et ε_{Si} la permittivité diélectrique du silicium. Il apparaît tout de suite qu'à forte densité de courant, l'expression de C_{eff} devra être corrigée au moyen de l'équation (2.23), pour tenir compte de la densité des porteurs libres.

La prise en compte de la densité des porteurs libres dans le calcul de la concentration effective correspond aux conditions rencontrées dans l'effet Kirk et le second claquage électrique. Les composants étudiés ici (Fig. 2.10) s'identifient à des composants MOS ou FOD et ne possèdent pas de collecteur faiblement dopé. L'élargissement de la base est peu perceptible et il n'y a pas de discontinuité dans le régime de fonctionnement du transistor.

Dans le cas d'un transistor NPN, pour une jonction collecteur-base fortement dissymétrique P^-N^+ , on a $N_a \ll N_d$ d'où $C_{eff} = N_a$. Afin de prendre en compte la densité de charge électrique due au passage du courant, il convient de modifier l'expression de C_{eff} :

$$C_{eff}^{mpn} = N_a + J_n/qv_{l_n} - J_p/qv_{l_p} \quad (2.31)$$

Cette dernière expression n'est valable que si l'on suppose la constance des densités de courants dans la région de collecteur. Pourtant, cette région est le siège de la multiplication des courants par avalanche. Toutefois, les coefficients d'ionisation des porteurs (Equ. 2.27) ne prennent une valeur significative que pour des valeurs du champ électrique proches et supérieures à la valeur de E_n^{crit} , soit pour des champs électriques de l'ordre de 10^6 V/cm . Ainsi, l'essentiel de la multiplication par avalanche se produit dans une zone réduite qui représente typiquement 10% [93] de l'extension totale de la charge d'espace (Fig. 2.12). Cette région de multiplication se situe près de la jonction métallurgique (ou plus précisément près du maximum du champ électrique). L'expression 2.31 de C_{eff} s'avère donc valable en première approximation car les courants J_n , J_p et les densités de porteurs libres n et p sont constants dans une grande partie (90%) de la région de collecteur.

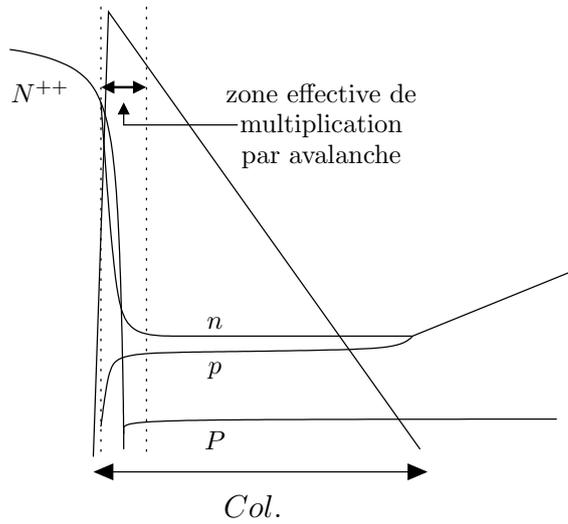


FIG. 2.12 – Profils des porteurs libres et du champ électrique dans la région de collecteur d'un transistor bipolaire NPN. La zone effective de multiplication par avalanche représente une partie limitée de la région de fort champ électrique (environ 10%).

De la même manière, la concentration effective dans un transistor bipolaire PNP sera donnée par :

$$C_{eff}^{mp} = N_d + J_p/qv_{lp} - J_n/qv_{ln} \quad (2.32)$$

Il découle des relations 2.31 et 2.32 que les conditions suivantes doivent être satisfaites pour assurer l'existence d'une charge d'espace de signe correct :

$$\frac{J_n}{qv_{ln}} < N_d + \frac{J_p}{qv_{lp}} \quad \text{et} \quad \frac{J_p}{qv_{lp}} < N_a + \frac{J_n}{qv_{ln}} \quad (2.33)$$

dans le cas d'un transistor bipolaire PNP et NPN respectivement. En d'autres termes et de manière approchée, il n'est pas possible de générer dans la région de multiplication par avalanche, un courant supérieur à celui qui entre dans la région, dans les régimes de fortes densités de courant. Nous verrons cependant que cela peut se produire dans un cas limite.

2.4 Étude unidimensionnelle fort courant des TBA

L'objectif n'est pas ici de construire un modèle précis des composants mais de mettre en évidence les phénomènes physiques clés sur lesquels repose leur fonctionnement. L'étude et l'optimisation des composants passent par l'utilisation d'outils de simulation numérique beaucoup plus évolués mais dont les résultats doivent toujours être critiqués. En outre, si ces outils donnent accès aux différentes grandeurs physiques en tout point du maillage de la structure, telle la densité de porteurs, le champ électrique, etc. . . , il est toujours nécessaire d'interpréter et de comprendre les phénomènes observés. Aussi, il est indispensable d'avoir une connaissance préalable du composant aussi poussée que possible pour comprendre, critiquer et valider les résultats obtenus.

Au moyen de l'assemblage des régions précédemment décrites, nous allons déterminer la caractéristique tension-courant d'un TBA à fortes densités de courant et lier les particularités observées aux phénomènes physiques sous-jacents.

2.4.1 Mise en équation du problème

Si l'on considère la figure 2.13, on constate que l'espace offert dans la profondeur physique de base W_b se voit partagé par la région électrique de base ($W_{b_{eff}}$) et la zone de charge d'espace de la région de collecteur (W_{av}). Cette dernière s'étend principalement du côté le moins dopé de la jonction collecteur-base et est le siège de la multiplication par avalanche. Nous allons exprimer $W_{b_{eff}}$ et W_{av} à partir des équations déterminées précédemment dans les différentes régions électriques.

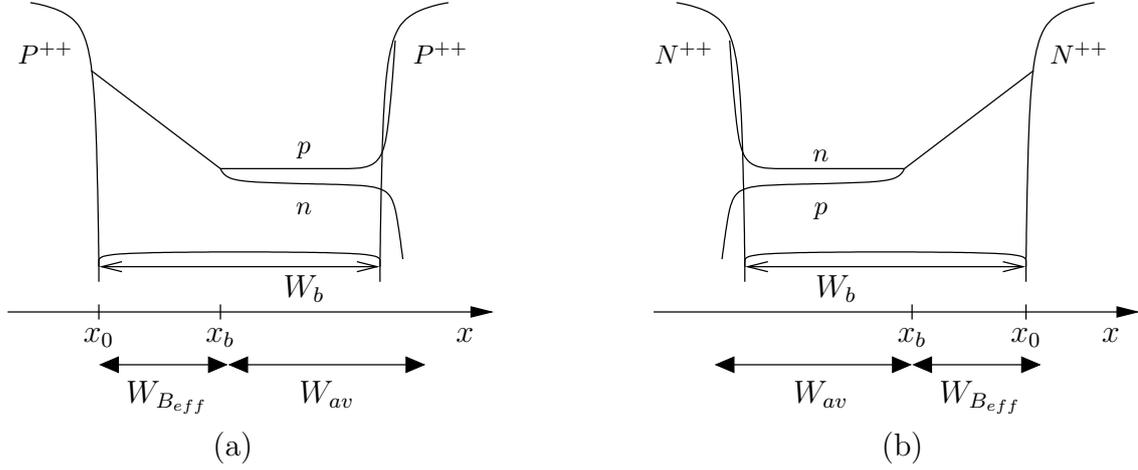


FIG. 2.13 – Profils des porteurs libres dans la base physique de largeur W_b : (a) dans le cas d'un transistor PNP et (b) dans le cas d'un transistor NPN. Les extensions des régions électriques de base ($W_{b_{eff}}$) et de collecteur (W_{av}) sont également représentées.

Calcul de l'extension de la région de base $W_{b_{eff}}$

La concentration de porteurs dans la base côté émetteur se déduit de l'équation (2.9) :

$$p_{x_0}^{pnp} = \sqrt{\frac{J_n}{qh_E}} \quad p_{x_0}^{npn} = \sqrt{\frac{J_p}{qh_E}} \quad (2.34)$$

celle du côté collecteur est approximativement donnée à partir des équations (2.22) :

$$p_{x_b}^{pnp} = \frac{J_p}{qv_{l_p}} \quad p_{x_b}^{npn} = \frac{J_n}{qv_{l_n}} \quad (2.35)$$

car il y a continuité de la concentration de porteurs et des densités de courant entre la région de base et de collecteur. Ceci nous permet de calculer l'extension de la région de base $W_{b_{eff}} = |x_0 - x_b|$ à partir de la relation (2.21) :

$$W_{b_{eff}} = \left| \frac{p_{x_0} - p_{x_b}}{\frac{\partial p}{\partial x}} \right| \quad (2.36)$$

avec $\frac{\partial p}{\partial x}$ le gradient des porteurs dans la base.

Calcul de l'extension de la zone de collecteur W_{av}

Les équations (2.7) et (2.25) permettent de donner une expression de la valeur de l'intégrale d'ionisation en fonction de la densité de courant générée et totale :

$$\mathfrak{S}_p^{npn} = \frac{J_n}{J_T} \quad \text{et} \quad \mathfrak{S}_n^{npn} = \frac{J_p}{J_T} \quad (2.37)$$

Connaissant les valeurs de \mathfrak{S}_n et \mathfrak{S}_p , on peut calculer la tension soutenue par la région de collecteur à partir des équations 2.29 et 2.30 :

$$V_{av}^{pnp} = \sqrt[4]{\frac{1}{2(A_n - A_p) \left(q \frac{C_{eff}^{pnp}}{\varepsilon_{Si}} \right)^3 \ln \left(1 - \frac{A_p - A_n}{A_p \mathfrak{S}_p^{pnp}} \right)}} \quad (2.38)$$

pour les transistors PNP, avec C_{eff}^{pnp} donné par l'équation (2.32), et

$$V_{av}^{npn} = \sqrt[4]{\frac{1}{2(A_p - A_n) \left(q \frac{C_{eff}^{npn}}{\varepsilon_{Si}} \right)^3 \ln \left(1 + \frac{A_p - A_n}{A_n \mathfrak{S}_n^{npn}} \right)}} \quad (2.39)$$

pour les transistors NPN avec C_{eff}^{npn} donnée par l'équation (2.31). Et enfin, calculer l'extension de la zone de charge d'espace qui est donnée pour une jonction plane abrupte par :

$$W_{av} = \sqrt{\frac{2\varepsilon_{Si}}{qC_{eff}} V_{av}} \quad (2.40)$$

la charge d'espace s'étendant principalement du côté le moins dopé, donc la base, pour une jonction fortement dissymétrique comme dans le cas considéré ici.

Obtention de l'équation

Nous avons démontré que les courants de trous et d'électrons sont constants dans la région de base et sur l'essentiel de la région de collecteur. Ils sont donc constants sur l'ensemble de la base physique, étant donnée leur continuité à la limite des deux régions (soit en x_b sur la figure 2.13). De plus, si l'on connaît le courant total, les valeurs de J_n et J_p dépendent seulement de la valeur de la dérivée du profil de porteurs dans la base (Equ.2.13). W_{av} (Equ.2.40) et $W_{b_{eff}}$ (Equ.2.36) sont donc des fonctions de J_T et $\frac{\partial p}{\partial x}$ seulement. La résolution de l'équation implicite :

$$W_{av} \left(\frac{\partial p}{\partial x}, J_T \right) + W_{b_{eff}} \left(\frac{\partial p}{\partial x}, J_T \right) - W_b = 0 \quad (2.41)$$

pour J_T et W_b fixés, permet de calculer la valeur de $\frac{\partial p}{\partial x}$, de laquelle on déduit la valeur de la tension dans la zone de collecteur (V_{av}) à partir de l'équation (2.38) pour un transistor PNP et (2.39) pour un NPN. Pour simplifier encore le problème, on ne prend en compte que la chute de potentiel dans la région de collecteur, où le champ électrique est fort, devant celle de la région de base où il est faible, et devant celle de la jonction émetteur-base. La tension aux bornes du transistor est alors égale à V_{av} . On construit les caractéristiques $V_{av}(J_T)$ en résolvant graphiquement ou numériquement l'équation (2.41), pour diverses valeurs du courant et pour différentes profondeurs de base physique.

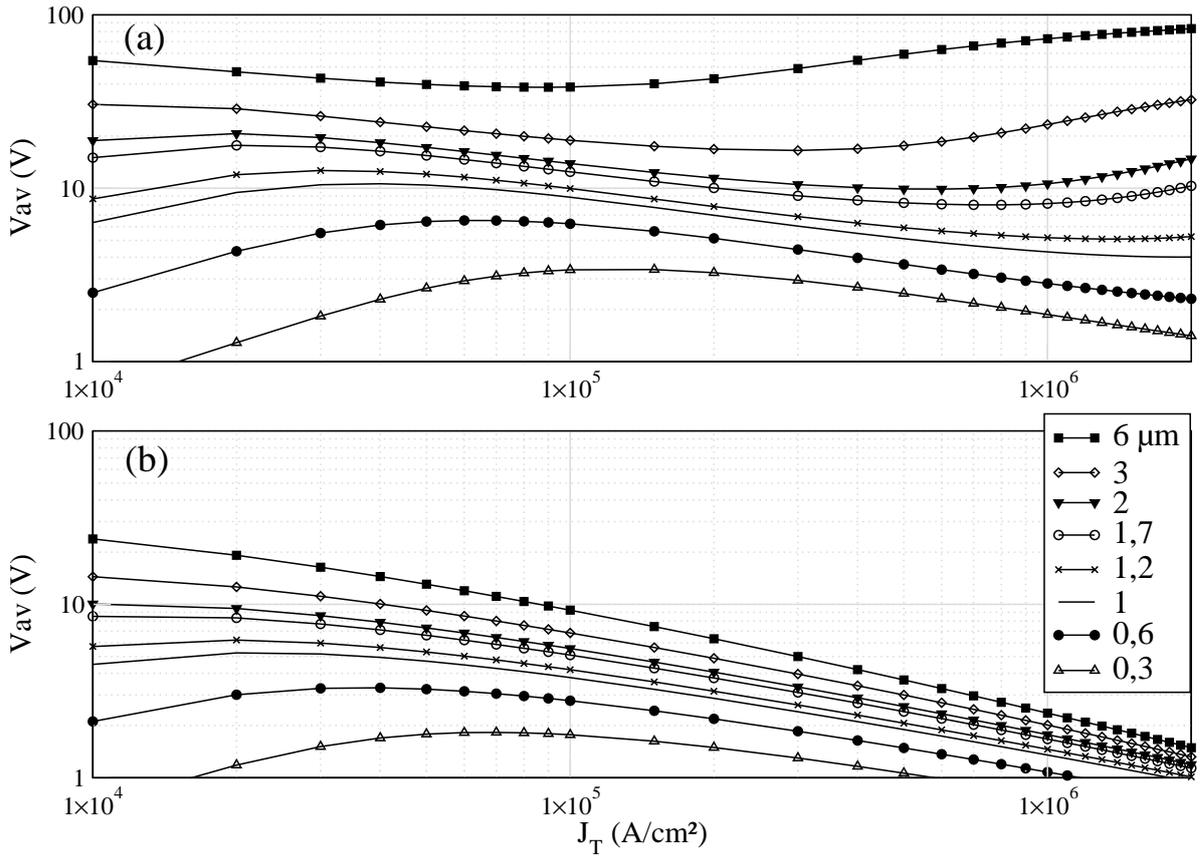


FIG. 2.14 – Tension aux bornes de la zone d'avalanche en fonction de la densité de courant pour différentes profondeurs de base physique, dans le cas d'un transistor PNP (a) et NPN (b).

2.4.2 Caractéristique tension-courant

Les caractéristiques sous fortes densités de courant sont présentées à la figure 2.14, pour chacun des deux types de composants NPN et PNP et pour différentes profondeurs de base physique. Pour une même profondeur de base, la tension de maintien d'un transistor PNP est systématiquement supérieure à celle d'un transistor NPN. Cette différence provient principalement de la plus faible valeur du facteur de multiplication du courant de trous dans le cas du transistor PNP (Fig.2.15) [68, 69]. On notera aussi l'accroissement de la tension de maintien avec l'augmentation de la profondeur de base, associée à la diminution du gain concomitante. Le modèle rend donc parfaitement compte du comportement habituel et bien connu des composants étudiés.

2.4.3 Effets spécifiques des fortes densités de courant

L'effet le plus remarquable est la rapide diminution de la tension pour les densités de courant croissantes, particulièrement au-delà de $1 \times 10^5 \text{ A/cm}^2$. Cette diminution est très marquée dans le cas des NPN, quelle que soit la profondeur de base. Seuls les transistors PNP possédant une base de faible profondeur suivent le même type de comportement. Pour des profondeurs plus importantes, cette chute est beaucoup moins marquée et l'on peut même constater une croissance de la tension d'autant plus importante et précoce que la profondeur de base est importante.

La méthode de résolution graphique de l'équation 2.41 permet de comprendre l'origine de ce phénomène. Elle repose sur le tracé de $W_{av}(\frac{\partial p}{\partial x}, J_T) + W_{b_{eff}}(\frac{\partial p}{\partial x}, J_T)$ à partir des équations 2.40

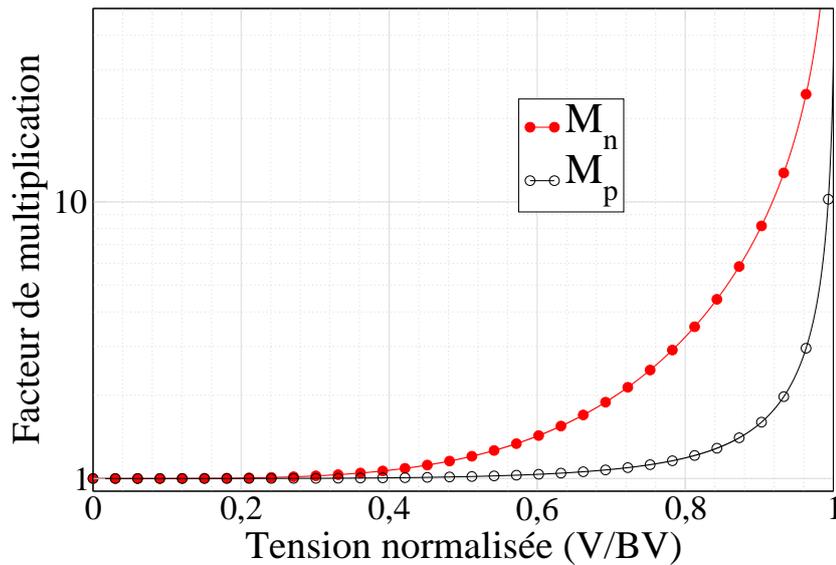


FIG. 2.15 – Comparaison des facteurs de multiplication d'un courant d'électrons M_n (Equ.2.25 et 2.29) et d'un courant de trous M_p (Equ.2.25 et 2.30) en fonction de la tension normalisée par rapport à la tension de claquage.

et 2.36, présenté dans la figure 2.16(a), associé à celui de la tension d'avalanche (Fig : 2.16(b)) calculée à partir des équations 2.38 et 2.39 dans le cas des transistors bipolaires PNP et NPN respectivement.

La méthode de résolution graphique est la suivante : pour une profondeur de base physique choisie, on détermine au moyen des courbes du haut (Fig : 2.16(a)) la valeur de $\partial p/\partial x$ associée à une densité de courant donnée, puis la tension correspondante en reportant la valeur de $\partial p/\partial x$ sur la courbe de même densité de courant dans la courbe du bas (Fig : 2.16(b)).

Cas des transistors NPN

Quelle que soit la valeur de la pente du profil des porteurs dans la base, la tension aux bornes de la zone de collecteur diminue très rapidement avec la densité de courant dans le cas des transistors NPN (Figure 2.16(b)). Cette diminution provient de l'augmentation de la concentration effective de la jonction collecteur-base, (C_{eff}) avec la densité de courant d'électrons (Equ.2.31) qui implique la réduction de la tension de claquage de la jonction, et par conséquent de la valeur de la tension de maintien du transistor. La diminution de C_{eff} que pourrait engendrer l'augmentation de la densité de courant de trous (Equ.2.31) se trouve limitée dans les transistors NPN où selon les conventions fixées précédemment, le gradient du profil de porteurs dans la base est positif. En effet, si l'on se reporte à la figure 2.11, on constate que le courant de trous pourra au maximum, pour $\partial p/\partial x = 0$, ne représenter qu'un quart du courant total, soit un tiers de la densité de courant d'électrons. Ainsi, même pour des profondeurs de base importantes, le comportement des transistors converge vers un régime de fonctionnement où le gradient des porteurs dans la base tend vers zéro (Fig.2.16(a)) et la tension de claquage décroît avec la densité de courant. Les profils des porteurs obtenus par simulation numérique unidimensionnelle sont présentés dans la figure 2.17. Ils illustrent et confirment ces explications pour une profondeur de base de $1,7 \mu\text{m}$.

Les résultats de simulations numériques permettent aussi de confirmer les hypothèses émises pour la simplification des équations. En particulier, la linéarité du profil des porteurs libres dans la région de base est vérifiée, bien que quelque peu masquée par l'échelle logarithmique

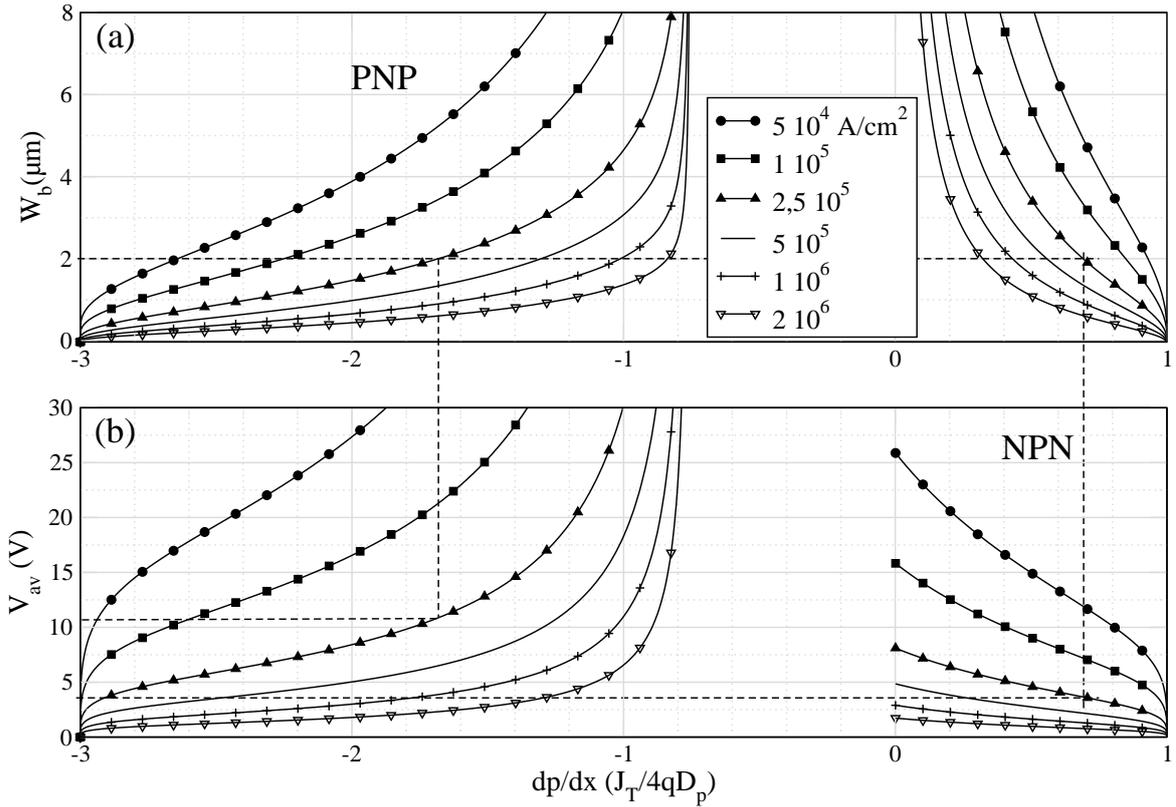


FIG. 2.16 – Résolution graphique de l'équation 2.41, dans le cas d'un transistor PNP (côté gauche) et d'un transistor NPN (côté droit). (a) Profondeur de base physique (Espace total occupé par les régions de base et de collecteur) et (b) valeur de la tension aux bornes de la région de collecteur en fonction de la pente du profil de porteurs dans la base, pour différentes densités de courant.

sur la figure 2.17(a). On peut également observer le profil constant de la densité des porteurs sur l'essentiel de la région de collecteur, car une majeure partie de la génération par avalanche est réalisée autour du maximum de champ électrique. Elle se traduit en particulier par le profil linéaire du champ électrique. La principale limitation du modèle semi-analytique provient de l'approximation des coefficients de multiplication par une fonction en E^7 qui ne prend pas en compte leur saturation à fort champ électrique (Equ.2.27), et conduit à des valeurs de tension extrêmement faibles à très forte densité. Cependant, ceci ne remet pas en cause le bien fondé des explications avancées qui se veulent plus qualitatives que quantitatives et privilégient une analyse des phénomènes physiques fondamentaux associés aux caractéristiques électriques du composant.

Cas des transistors PNP

La chute de tension à forte densité de courant dans les transistors PNP (Fig.2.16(b)) est elle aussi associée à l'augmentation de C_{eff} due à la charge amenée par le courant de trous dans la zone de charge d'espace du collecteur (Equ.2.32). Mais dans ce type de transistor, le comportement de la base permet, contrairement au cas des transistors NPN, l'existence de densités de courant d'électrons égales ou supérieures à celle des trous (Fig.2.11). La charge électrique des électrons associée à leur densité de courant tend alors à diminuer la valeur de C_{eff} (Equ.2.32) ce qui implique l'augmentation de la tension de claquage de la jonction collecteur-

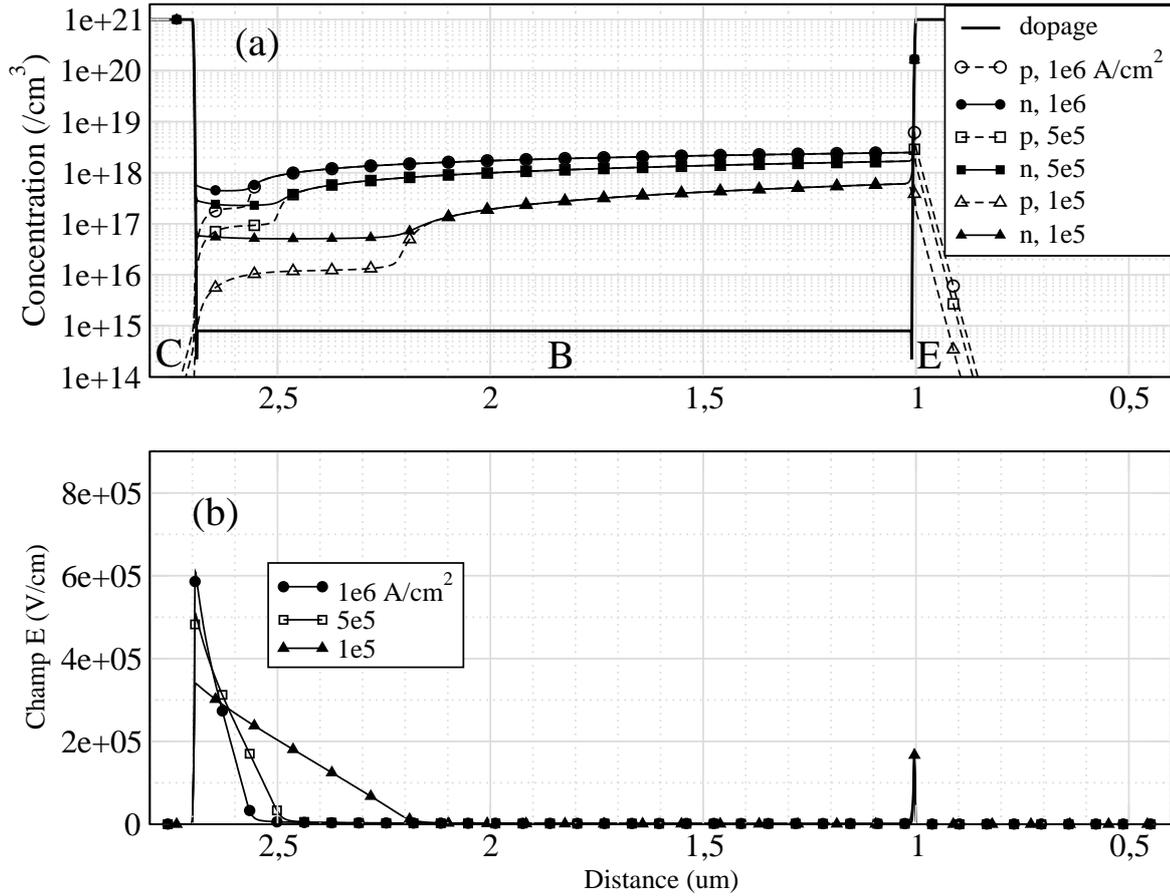


FIG. 2.17 – Profils de la concentration des électrons n et des trous p (a) et répartition du champ électrique (b) dans un transistor bipolaire NPN, obtenus par simulation numérique unidimensionnelle pour différentes valeurs de densités de courant.

base. La tension augmente donc très rapidement lorsque les densités de courant d'électrons et de trous sont proches, c'est-à-dire pour un gradient du profil de porteurs dans la base proche de (-1) sur la figure 2.16(b). On remarque qu'il existe des régimes de fonctionnement où la valeur du gradient est supérieure à (-1) (Fig 2.16(b)) et où la densité de courant d'électrons est donc supérieure à celle du courant de trous (Fig.2.11). Ceci est possible car la vitesse limite de dérive des électrons est supérieure à celle des trous. Ainsi, si l'on se reporte aux équations 2.22, la densité d'électrons est plus faible que celle des trous pour des densités de courant égales. La limite du régime de fonctionnement, qui correspond à l'égalité des densités d'électrons et de trous est donnée en terme de densité de courant par la relation :

$$J_n = J_p \frac{v_{lp}}{v_{ln}} \quad (2.42)$$

On peut constater sur la figure 2.16 que l'augmentation de la tension est d'autant plus tardive vis-à-vis de la densité de courant que la profondeur de base physique est petite. Pour de faibles profondeurs de base, le comportement d'un transistor PNP est similaire à celui du NPN, avec toutefois une tension légèrement supérieure.

Dans les technologies les plus récentes ($< 0,15 \mu\text{m}$), l'emploi de substrats très dopés dégrade les performances des transistors NPN réalisés grâce au NMOS, car leur résistance de base se trouve extrêmement réduite. La résistance de base des transistors PNP basés sur les PMOS,

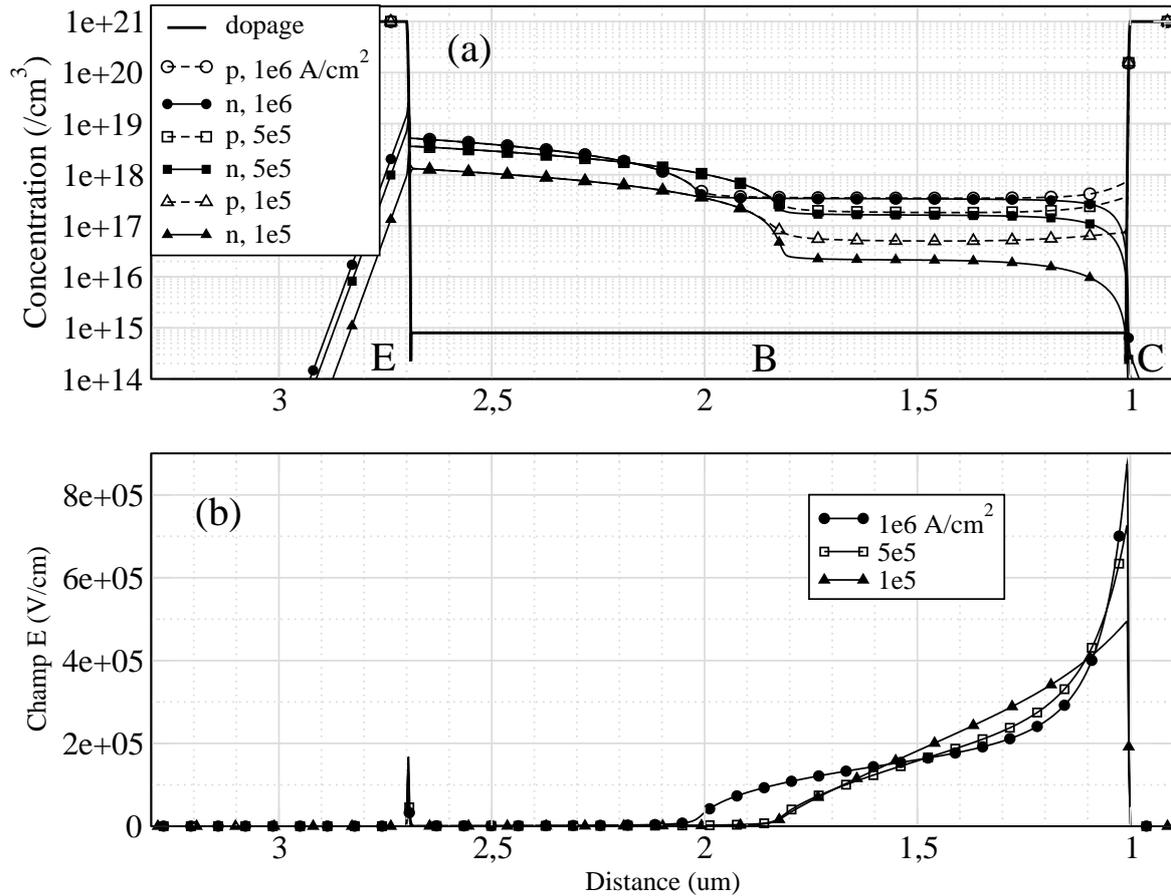


FIG. 2.18 – Profils de la concentration des électrons n et des trous p (a) et répartition du champ électrique (b) dans un transistor bipolaire PNP obtenus par simulation numérique unidimensionnelle pour différentes valeurs de densités de courant.

n'est pas affectée car ils sont réalisés dans un puits N qui «isole» le composant du substrat. Le comportement des transistors PNP, comparable à forte densité de courant à celui des transistors NPN, leur permet finalement de présenter de meilleures performances que ces derniers [70].

Dans le cas de bases profondes, le comportement des transistors PNP est par contre très différent de celui des NPN. Ce n'est pas une chute, mais une augmentation de la tension qui est observée dans les transistors PNP. L'énergie dissipée dans la région de collecteur est alors très importante et la robustesse du transistor PNP est très inférieure à celle du NPN. Les résultats de simulations numériques présentés dans la figure 2.18 pour un transistor PNP de profondeur de base de $1,7 \mu\text{m}$ illustrent parfaitement le comportement précédemment décrit. Encore une fois, les hypothèses émises pour la simplification des équations sont justifiées. La répartition du champ électrique est cependant assez déformée lorsque les concentrations des électrons et trous dans la zone de charge d'espace sont proches, mais cela ne traduit pas un défaut majeur de l'approche.

2.5 Phénomène de focalisation du courant

L'étude unidimensionnelle qui vient d'être menée, montre la rapide décroissance de la tension aux bornes de la zone de collecteur dans un TBA lorsque la densité de courant augmente. Parti-

culièrement marquée dans les transistors NPN, nous allons voir que cette chute de tension est la cause de la focalisation du courant observée dans ces composants. L'utilisation de la technique de microscopie à émission lumineuse permet d'observer la non-uniformité de la distribution du courant le long du collecteur d'un transistor [59]. L'observation la plus remarquable est que la portion de collecteur utilisée est proportionnelle au courant total dans le composant, jusqu'à l'étalement du courant sur la totalité de la longueur du collecteur [94].

2.5.1 Étude théorique

Un transistor bipolaire réel peut être décrit de manière segmentée, comme la mise en parallèle du composant élémentaire décrit précédemment. Étant donnée la rapide décroissance de la

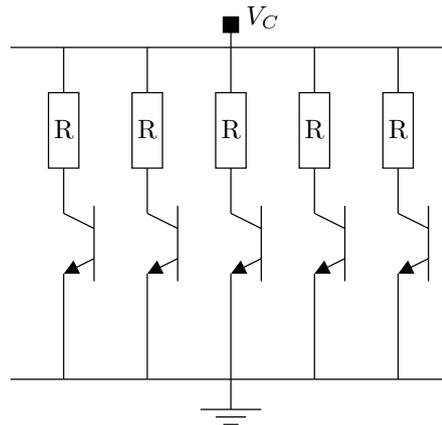


FIG. 2.19 – Modélisation d'un TBA par la mise en parallèle de transistors élémentaires.

tension dans un transistor élémentaire pour les densités de courant croissantes, il paraît évident que le moindre déséquilibre de la densité de courant entre les composants conduit à la concentration de la totalité du courant dans un seul. Ce phénomène est d'autant plus important que nous avons affaire à deux phénomènes intimement couplés : la multiplication par avalanche du courant d'électrons dans la région de collecteur fournit un courant de trous permettant la réinjection d'un courant d'électrons par effet bipolaire (gain du transistor) dans la région de collecteur.

Le modèle proposé ne prend cependant pas en compte les chutes de potentiel dans les résistances d'accès à la région de collecteur et d'émetteur. Pourtant, même de faibles valeurs de ces résistances vont entraîner des chutes de potentiel non négligeables. Leur prise en compte, limitée à celle du collecteur, comme présenté dans la figure 2.19, permet d'expliquer clairement le phénomène d'étalement du courant couramment observé.

Considérons le cas d'un transistor possédant une base de $1,7 \mu\text{m}$. La tension aux bornes d'un segment est représentée sur la figure 2.20 pour différentes valeurs de la résistance de collecteur R . Cette tension est obtenue en ajoutant la tension calculée aux bornes de la zone de collecteur à la chute de tension dans la résistance de collecteur R . La résistance de collecteur modifie peu la tension pour les faibles densités de courant. Par contre, à forte densité de courant, c'est elle qui va déterminer la tension totale. La tension présente une valeur minimum V_{MIN} pour une densité de courant donnée J_{MAX} . Cette tension est d'autant plus basse que la densité de courant maximum associée est grande et que la valeur de la résistance est faible.

Lorsque le courant total croît dans le transistor, il se localise alors dans un seul transistor élémentaire mais de manière limitée, grâce à l'augmentation de la tension due à la résistance de collecteur. En effet, lorsque la densité de courant dans un transistor élémentaire dépasse J_{MAX} , le courant supplémentaire sera conduit par le transistor voisin et ainsi de suite, jusqu'à utilisation de l'ensemble des transistors élémentaires disponibles. La résistance de collecteur, appelée aussi

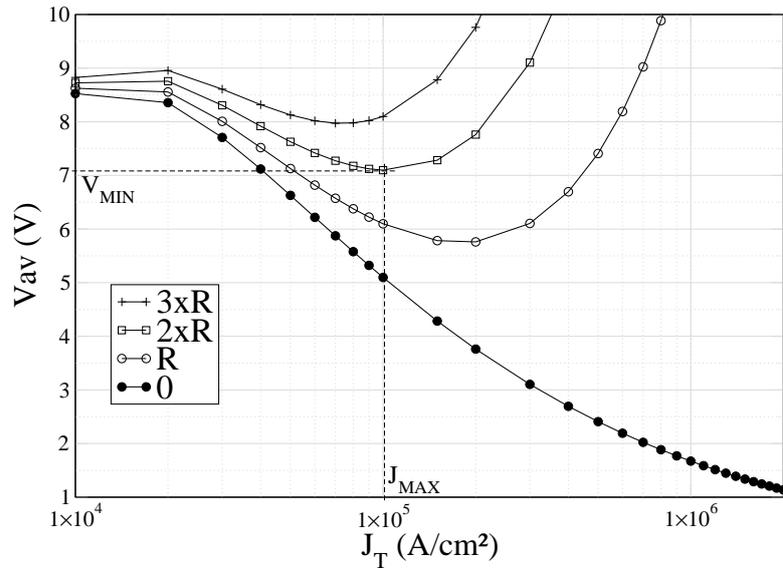


FIG. 2.20 – Compensation de la chute de potentiel par une résistance de collecteur.

résistance de ballast permet de contrôler la valeur de J_{MAX} et d'éviter ainsi une trop importante focalisation du courant qui conduirait rapidement à un second claquage thermique et à une faible robustesse du composant vis-à-vis des ESD.

2.5.2 Vérification expérimentale

Des observations expérimentales réalisées sur des TBA permet de vérifier le modèle proposé. Ces TBA sont réalisés à partir d'un transistor NMOS d'une technologie $1,7 \mu m$.

Le composant utilisé possède deux doigts de grilles et deux drains placés de part et d'autre d'une source, comme schématisé sur la figure 2.21. Afin de tester l'influence de la résistance de ballast, des composants réalisés avec différentes distances entre les prises de contacts des collecteurs et la grille (DCG), toutes dimensions identiques par ailleurs, ont été fabriqués. La

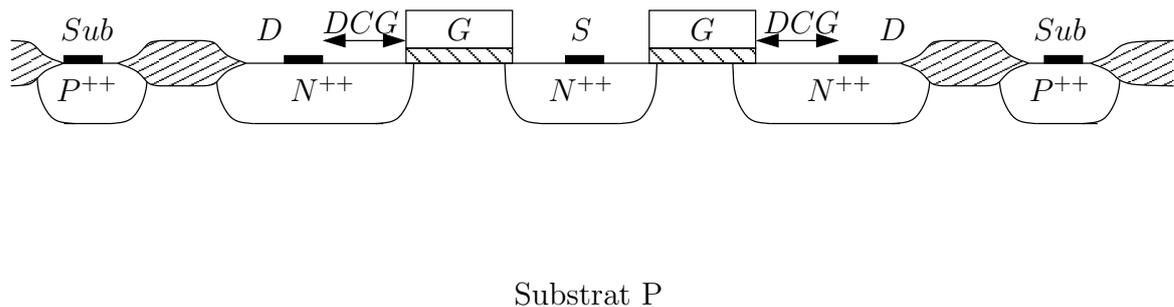


FIG. 2.21 – Vue en coupe du transistor NMOS utilisé.

figure 2.22 permet de comparer les caractéristiques TLP de trois transistors pour des valeurs de DCG de $2,3$, $5,2$ et $8,2 \mu m$ à faibles valeurs du courant, pour lesquelles il n'est pas distribué sur la totalité du doigt. La résistance après le déclenchement est extrêmement faible car le courant se déploie progressivement sur la longueur du doigt, sous une tension quasiment constante. Comme prévu par notre modèle, cette tension est d'autant plus faible que la résistance de ballast est faible.

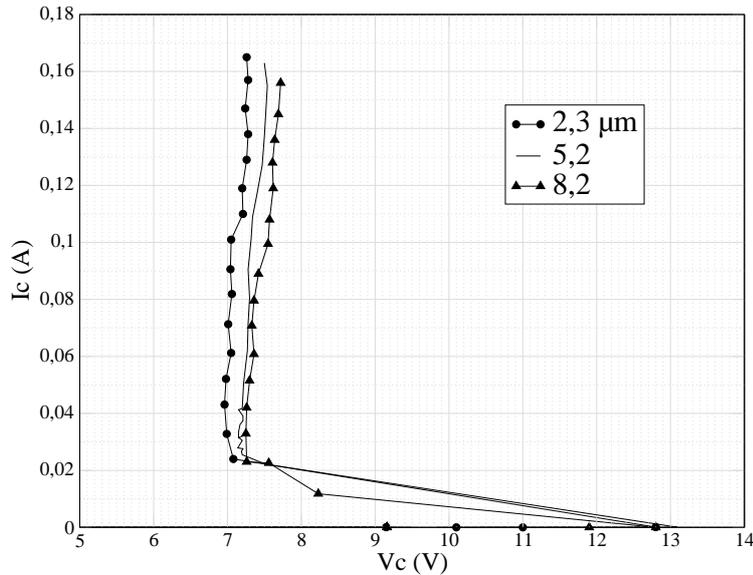


FIG. 2.22 – Caractéristique TLP pour différentes valeurs de la distance entre bord de grille et contact de collecteur.

Enfin les images d'EMMI présentées dans la figure 2.23 sur les composants ayant une distance DCG de 5,2 et 8,2 μm illustrent la dépendance de la densité de courant maximale avec la résistance de ballast.

Pour des courants identiques, l'étalement du courant le long du doigt est plus important dans le cas du transistor le plus ballasté. Ceci témoigne de la plus faible valeur de J_{MAX} pour une résistance de collecteur plus élevée, comme prévue par notre modèle. Lorsque le courant s'est étendu sur toute la longueur doigt, le second doigt prend le relais [61]. L'observation du composant possédant une distance DCG de 2,3 μm n'est pas possible en régime statique car la densité de courant J_{MAX} est trop importante et l'échauffement trop localisé conduit rapidement au second claquage thermique.

Nous devons enfin mentionner l'impact négatif sur la robustesse ESD, de l'emploi de siliciure qui permettent de réduire la résistance des drains et des sources [63, 95]. La dégradation de la robustesse découle directement de l'affaiblissement de la résistance de ballast avec les conséquences que l'on vient de présenter ici.

2.6 Destruction du composant : aspect thermique

2.6.1 Chemin vers la destruction du composant

Si l'ajout d'une résistance de ballast permet de limiter la densité de courant maximale en début de décharge et d'éviter une destruction prématurée, son rôle est aussi très important dans la stabilisation thermique du composant à plus fortes densités de courant. Il est important de souligner ici que nous avons affaire à un système physique qui est très éloigné de son équilibre thermodynamique. Même si à première vue le courant se distribue de manière uniforme dans le composant, il existe au cours du temps de fortes fluctuations de la densité de courant. La robustesse du dispositif réside dans sa capacité à limiter l'amplitude de ces fluctuations. En effet, si cette amplitude est trop importante, les élévations locales de température vont précipiter la destruction du composant en initialisant précocement le phénomène de second claquage thermique. La résistance de ballast permet de limiter cette amplitude maximum. S'il apparaît une augmentation locale de la densité de courant due à un échauffement, elle est compensée par la

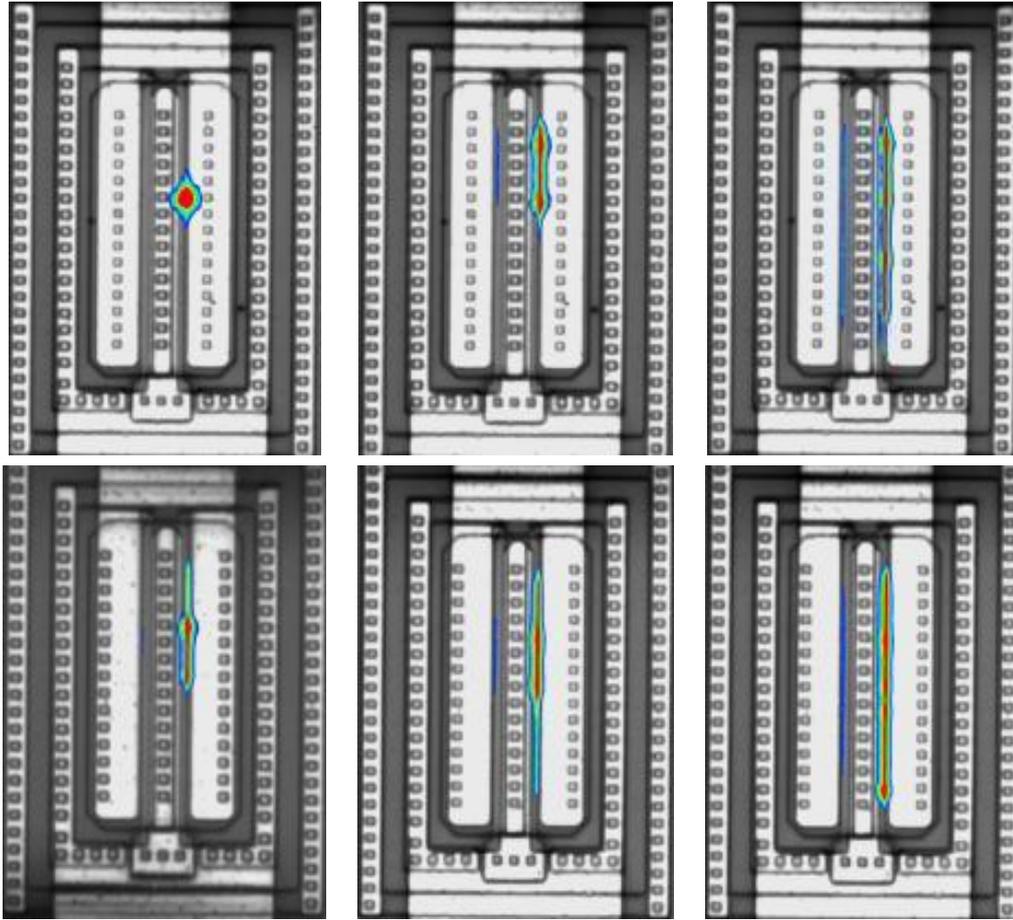


FIG. 2.23 – Images EMMI du transistor de distance DCG $5,2 \mu\text{m}$ (rangée du haut) et de celui de distance DCG de $8,2 \mu\text{m}$ (rangé du bas) pour des valeurs de courant statiques de 50, 70 et 100 mA de gauche à droite.

chute de tension dans la résistance de ballast qui dépoliarise cette région et concourt ainsi à la stabilisation thermique de l'ensemble.

Le mécanisme de dégradation des transistors NMOS a été étudié par Pascal SALOME dans [96] à partir des résultats d'analyse de défaillance. Nous allons rapidement résumer les résultats présentés dans sa thèse de doctorat [58]. Il apparaît que les défauts sont initialisés à la jonction collecteur-base et découlent du second claquage thermique. Ces défauts initiaux se traduisent par une légère augmentation du courant de fuite du composant et ne semblent pas être catastrophiques vis-à-vis de la fiabilité [97]. En particulier, ils ne sont pas un site privilégié pour un nouveau défaut lors de stress suivants. Ces défauts sont qualifiés de «soft breakdown». Pour des stress plus longs ou plus intenses, les défauts initialisés à la jonction collecteur-base entraînent la focalisation du courant d'émetteur. Il apparaît alors un nouveau défaut dans la jonction émetteur-base. Lorsque les deux défauts se rejoignent, ils forment un filament qui court-circuite le drain et la source. Cette défaillance, qui se traduit par un courant de fuite important, est qualifiée de «hard breakdown».

L'utilisation du modèle semi-analytique permet d'étudier l'impact de la température sur les différentes régions électriques indépendamment. Si l'on reprend le modèle d'un TBA par la mise en parallèle de transistors élémentaires associés à leurs résistances de ballast (Fig. 2.19), nous pouvons étudier l'effet de l'échauffement de la région d'émetteur ou de collecteur du transistor de l'une des sections.

Une augmentation de la température à la jonction émetteur-base correspond à une diminution

du paramètre de recombinaison d'émetteur h_E [98]. La figure 2.24 montre l'impact de cette diminution sur la tension aux bornes d'une section en fonction de la densité de courant. Le paramètre de recombinaison d'émetteur est de $1,5 \cdot 10^{-14}$ pour la courbe originale. Pour une augmentation de température de 200 K, ce paramètre est réduit à $1 \cdot 10^{-14}$ (courbe " h_E " Fig. 2.24). Pour une tension donnée de 8 V, l'échauffement entraîne un accroissement de la densité de courant (Fig. 2.24). L'emballement thermique peut alors entraîner une focalisation du courant dans cette section.

Pour une augmentation de température de 200 K dans la région de collecteur, les coefficients de multiplication par avalanche sont réduits de moitié environ. Dans ces conditions, on constate une diminution de la densité de courant (courbe "avalanche" Fig. 2.24) dans la section échauffée, qui tend à limiter l'emballement thermique et la focalisation du courant dans la section.

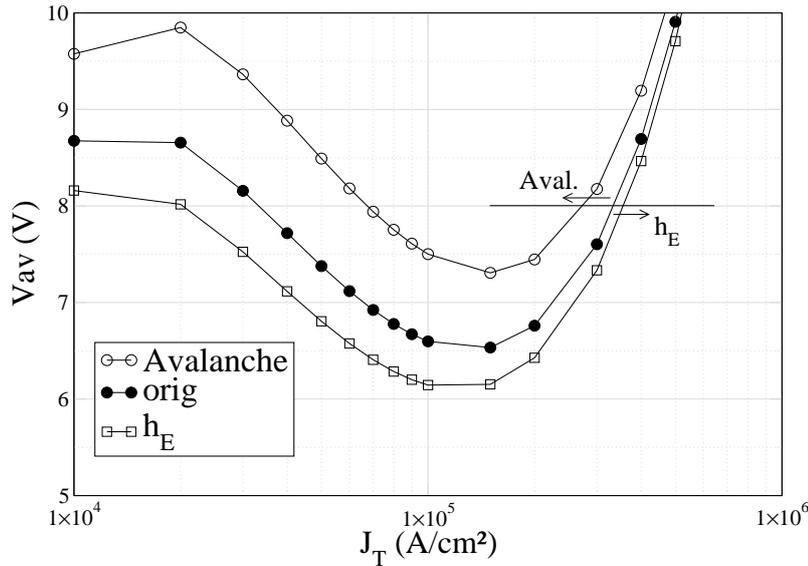


FIG. 2.24 – Effet de la température de 200 K sur la tension aux bornes de la zone d'avalanche en fonction de la densité de courant. Courbe originale (orig) ($T=300$ K). Effet de la diminution du paramètre de recombinaison d'émetteur (h_E), et de la diminution des paramètres d'avalanche A_n et A_p (avalanche), pour une augmentation de la température de 200 K dans ces régions.

2.6.2 Proposition d'un nouveau type de composant

Le couplage thermique de la zone d'avalanche qui est très chaude, avec la jonction émetteur-base, apparaît comme un phénomène qui peut précipiter la destruction du composant. De plus, on peut intuitivement comprendre que le fort couplage électrique des deux phénomènes de multiplication que sont l'effet bipolaire et la multiplication par avalanche favorise la concentration spatiale du courant. La région de collecteur, qui est la région la plus chaude du composant lors d'un stress ESD, possède la propriété d'éviter l'emballement thermique, contrairement à la région d'émetteur qui va le favoriser. Un dispositif dans lequel les couplages électriques et thermique entre la jonction émetteur-base et la région de collecteur seraient fortement diminués, doit posséder une forte robustesse aux ESD. Pour obtenir un découplage important, il est nécessaire d'éloigner significativement la jonction émetteur-base de la région de multiplication par avalanche, ce qui revient à augmenter la profondeur de base effective du TBA.

D'après les résultats obtenus au moyen du modèle semi-analytique, seuls les transistors NPN sont de bons candidats pour réaliser ce type de composant. Leur tension de maintien n'augmente pas rapidement lorsque leur profondeur de base est agrandie contrairement au transistor bipolaire

PNP. L'étude et la mise au point de transistors NPN possédant une base effective profonde est l'objet du chapitre suivant.

2.7 Conclusion et perspectives

Après avoir présenté les différents types de composants permettant la réalisation des protections ESD basées sur le TBA, nous avons décrit l'état de l'art de leur principe de fonctionnement. Les aspects liés aux fortes densités de courant lors d'une décharge électrostatique ne sont que trop rarement étudiés dans la littérature.

En s'appuyant sur les connaissances issues des transistors bipolaires de puissance et des transistors bipolaires haute fréquence, une étude approfondie du comportement des TBA à forte densité de courant a pu être menée. L'approche régionale unidimensionnelle utilisée a été adaptée au cas des TBA, en particulier pour la région de collecteur du transistor qui est le siège de la génération par avalanche.

Les résultats obtenus par cette approche semi-analytique, ont permis de mettre en évidence les phénomènes physiques spécifiques à ces composants. La différence de comportement des transistors bipolaires NPN et PNP a été mise en évidence. Une nouvelle explication de l'origine de la focalisation du courant dans ces dispositifs en a été déduite.

Enfin, après avoir abordé les aspects thermiques, à l'origine de la destruction du composant, nous avons pu proposer le principe d'un nouveau type de TBA NPN dont la profondeur de base effective serait importante. L'étude de ces composants potentiellement capables d'une excellente robustesse aux décharges électrostatiques est menée dans le chapitre suivant.

Pour finir, nous souhaitons souligner que l'étude du système physique que constitue le TBA dans les régimes de fortes densités de courant, met en jeu la physique des systèmes très éloignés de leur équilibre thermodynamique. Il y a sans doute beaucoup de progrès potentiels à réaliser en s'appuyant sur les outils développés pour l'étude des phénomènes non linéaires [99], comme les transitions de phases ou le chaos, qui sont des aspects très modernes et actuels de la physique.

Le modèle unidimensionnel développé s'inspire des modèles distribués de composants bipolaires de puissance étudiés au LAAS. Cette dernière approche de modélisation est particulièrement prometteuse pour les composants de protection ESD. Son extension à ce domaine pourrait permettre de fournir des modèles de composant ESD qui prennent en compte les effets des fortes densités de courant voire même les aspects thermiques. Les modèles distribués reposent sur la résolution de l'équation du profil des porteurs dans la région de base. Le développement, sans doute difficile, d'une méthode de résolution des profils de porteurs dans la région de collecteur permettrait d'étudier et de modéliser précisément les aspects dynamiques du fonctionnement des TBA. L'aspect dynamique devient en effet primordial pour les stress ESD de type CDM.

Chapitre 3

Optimisation de transistors bipolaires autopolarisés

L'étude menée au chapitre précédent nous a conduit à proposer un nouveau type de composant bipolaire NPN dont la profondeur de base serait grande pour assurer un découplage important entre les régions d'émetteur et de collecteur. Ce type de composant ne peut pas être basé sur l'utilisation de transistors NMOS ou de transistors bipolaires latéraux (FOD). En effet, si la tension de maintien des transistors NPN dépend peu de leur profondeur de base, ceci n'est vrai que pour de très fortes densités de courant. Ces composants seraient particulièrement difficiles à déclencher et présenteraient avant d'atteindre leur repliement, un régime de fonctionnement à relativement haute tension et fort courant qui leur serait fatal. Nous allons montrer dans ce chapitre comment contourner ce problème en utilisant des transistors bipolaires NPN possédant un collecteur faiblement dopé.

Une étude approfondie du fonctionnement de ce transistor lors d'une décharge électrostatique va être menée. Pour cela, nous utiliserons intensivement la simulation physique, après avoir pris conscience des limitations qu'elle comporte dans le cas particulier des décharges électrostatiques. Une étude précise devrait faire appel à la simulation tridimensionnelle, mais les lourds moyens informatiques qui seraient nécessaires la rendent très difficilement réalisable. Toutefois, des simplifications permettent par une double étude en deux dimensions, d'obtenir une description plus accessible et plus efficace du problème, au prix cependant de certaines limitations. La compréhension fine du fonctionnement permettra d'étudier et de proposer un ensemble de règles pour obtenir des composants optimisés et extrêmement robustes. Finalement, des résultats expérimentaux obtenus dans trois technologies différentes viendront illustrer et valider les conclusions tirées des simulations.

3.1 Utilisation des simulateurs électrothermiques

Les structures de protection ESD ne sont pas réutilisables directement d'une technologie à l'autre. Un dispositif peut être très performant dans une technologie et totalement inefficace dans une autre. Le développement de structures de protection dans une nouvelle technologie requiert donc un certain nombre d'itérations en termes de conception, réalisation et caractérisation. Les tests ESD expérimentaux (HBM, TLP) donnent des informations précieuses sur les caractéristiques d'un composant mais aident peu la compréhension des mécanismes physiques mis en jeu pendant une décharge. Le recours à la simulation pour prédire le comportement des structures de protection et optimiser leurs performances apparaît donc indispensable. Idéalement, elle devrait permettre de faire l'économie de plusieurs réalisations technologiques et de nombreux motifs de caractérisation, ce qui peut représenter un gain de temps et d'argent considérable dans le développement d'une technologie.

3.1.1 «GIGO law»

L'étude et l'optimisation des composants de protection ESD passe par l'utilisation de la simulation numérique. Il convient d'insister sur le fait que tirer profit des logiciels de ce type demande infiniment plus d'investissement de la part de l'utilisateur qu'un banal logiciel de traitement de texte. Il est primordial de retenir la loi fondamentale selon laquelle tous paramètres d'entrée erronés entraînent des mauvais résultats (GIGO law : Garbage In, Garbage Out) qui peuvent mener à une interprétation très éloignée de la réalité.

En premier lieu, la description de la structure donnée au simulateur doit être aussi proche que possible de la structure réelle. En particulier, les profils de dopages utilisés doivent être conformes aux profils réels, obtenus par exemple par des mesures par sonde ionique (SIMS (Secondary Ion Mass Spectroscopy)).

Le choix des modèles physiques et la réalisation d'un maillage efficace sont aussi des aspects particulièrement critiques. Ces deux étapes sont cruciales pour obtenir des résultats précis et fiables et reposent en grande partie sur une connaissance préalable aussi poussée que possible du fonctionnement du composant simulé. À ce titre, l'enseignement tiré de l'approche semi-analytique du chapitre précédent constitue un bon outil pour critiquer et améliorer les résultats obtenus par simulation.

3.1.2 Limitation des modèles physiques

Les simulateurs physiques permettent de résoudre de manière couplée les équations physiques des semi-conducteurs (équation de Poisson et équations de continuité des porteurs) et l'équation de la chaleur [90]. L'utilisation de ces simulations électrothermiques est indispensable pour comprendre le fonctionnement des composants de protection ESD, dans lesquels l'élévation de la température, localisée dans la région de collecteur, est très importante lors d'une décharge électrostatique de forte intensité.

Les simulateurs électrothermiques ont été initialement développés pour l'étude des composants classiques de la microélectronique. Les conditions rencontrées lors d'une décharge électrostatique s'éloignent de ces régimes de fonctionnement classiques par deux aspects. D'une part, la température dans le composant est très élevée (jusqu'à atteindre la limite de fusion du silicium), et d'autre part, les densités de courants mises en jeu sont très grandes. Or, les modèles physiques ne sont pour la plupart validés que pour des températures inférieures à 600 K [100, 101]. En conséquence de la très forte injection, les modèles de mobilité et de génération-recombinaison sont utilisés dans une région proche de leur limite de validité. Les paramètres des modèles thermiques, comme la capacité calorifique, le coefficient de diffusion thermique du silicium le sont également, mais dans une moindre mesure.

Dans les TBA, le champ électrique, fortement influencé par la densité des porteurs libres, atteint des valeurs très importantes. Les modèles de multiplication par avalanche ont été développés à l'origine pour obtenir les tensions de claquages de composant [92], ou pour les régimes de faible multiplication, comme dans le collecteur d'un transistor bipolaire pour les régimes normaux de fonctionnement. Leur validité pour des conditions à la fois de forte avalanche et d'intense champ électrique est particulièrement discutable. De plus, la zone de plus haute température se situe précisément dans la région où les phénomènes d'avalanche sont importants. Une description plus précise de la physique ayant trait à cette région pourrait reposer sur l'utilisation des modèles dits hydrodynamiques [102], dans lesquels la température des électrons et des trous ne sont plus considérées égale à celle du réseau cristallin. Malheureusement, les paramètres de ces modèles sont mal connus dans des conditions classiques et à plus forte raison dans les conditions extrêmes relatives aux décharges électrostatiques. L'ajout de deux équations supplémentaires, pour le calcul de la température des porteurs, pénalise de plus la convergence et la rapidité des calculs [17].

3.1.3 Approche et solutions retenues

Les simulations présentées dans ce document ont été réalisées avec le logiciel Dessis de la société ISE [101].

Afin de simuler précisément la réponse du composant à un stress ESD, nous avons mis en œuvre la simulation mixte en utilisant le schéma électrique équivalent du modèle HBM.

L'utilisation de profils de dopages analytiques, dont les paramètres sont ajustés sur des simulations soigneusement calibrées du procédé de fabrication, assure une description précise du composant. L'emploi de profils analytiques autorise une grande souplesse pour l'étude des modifications de géométrie. Elle permet d'éviter l'utilisation d'une étape de simulation technologique de la totalité du composant, qui serait lourde en temps de calcul.

Le choix des modèles physiques a été réalisé à partir d'un critère de stabilité des simulations et de pertinence vis-à-vis des conditions physiques rencontrées dans le fonctionnement du composant. Pour l'ensemble des modèles, la dépendance avec la température a été activée :

- *Concentration intrinsèque* : L'utilisation de l'approximation de Boltzmann sur la statistique de Fermi-Dirac est légitime pour les conditions physiques rencontrées lors d'un stress ESD. La température est supérieure à 300 K et la forte densité de porteurs dans la zone active du composant reste largement inférieure à 10^{19} Atome/cm³. L'utilisation de la statistique de Fermi-Dirac pénaliserait de plus la convergence et la vitesse des calculs [17]. Le modèle de rétrécissement de la bande interdite (Band Gap Narrowing) à fort dopage a également été utilisé.
- *Mobilité des porteurs* : les modèles de dépendance de la mobilité avec le dopage, de saturation avec le champ électrique et de dégradation avec le niveau d'injection (carrier-carrier scattering) ont été utilisés.
- *Génération-Recombinaison* : les modèles SRH (Shockley Read Hall) et Auger pour les recombinaisons ont été choisis. Le terme de recombinaison Auger est particulièrement important dans les régions de forte densité de porteurs et pour modéliser correctement l'efficacité d'injection de l'émetteur des transistors bipolaires.

Le modèle de génération par avalanche de van Overstraeten-de Man a été sélectionné pour la bonne convergence constatée dans les simulations et les tensions de maintien obtenues qui sont en bon accord avec l'expérience.

Le simulateur offre la possibilité de calculer les courants associés aux effets thermoélectriques. Plusieurs tests nous ont montré que les courants dus au couplage thermoélectrique ne semblent pas jouer un rôle fondamental dans les simulations réalisées, malgré le fort gradient de température présent dans le composant. Le calcul des termes supplémentaires introduits dans les équations fondamentales dégrade la convergence et implique une forte augmentation du temps de simulation qui nous a conduit à les laisser de côté.

L'importante étape du calibrage des simulations doit être particulièrement soignée. Le calibrage des profils de dopages est primordial. Comme nous le verrons, la tension de maintien des TBA dépend directement des profils de dopage de collecteur.

Le bon accord avec l'expérience des caractéristiques statiques du transistor, comme les caractéristiques $I_C(V_{CE})$, obtenues par simulation doit être vérifié. Un calibrage précis des modèles, basé sur ces caractéristiques, est superflu car l'ordre de grandeur des densités de courant lors d'un stress ESD est plusieurs fois supérieur. Le domaine d'utilisation des modèles physiques pour des caractéristiques statiques et lors d'un stress ESD est donc totalement différent.

Conclusion

La somme de paramètres incertains ou inadaptés aux conditions extrêmes rencontrées dans un composant subissant un stress ESD, contraint à beaucoup de prudence quant aux résultats obtenus par simulation. La prédiction de la robustesse d'un composant apparaît même à l'heure actuelle comme illusoire. Cependant, en limitant leur utilisation dans des conditions peu éloignées des domaines de validité des modèles, nous pouvons obtenir des informations fiables sur le

fonctionnement du composant et en faire, pour le moins, ressortir les tendances. Les simulateurs électrothermiques ont souvent été utilisés avec succès pour comprendre le fonctionnement des composants de protection ESD [17, 60, 94, 103].

3.2 Cas d'un composant bipolaire NPN vertical

Nous allons analyser en détail le fonctionnement d'un transistor bipolaire NPN vertical. Les liens entre les caractéristiques électriques du composant et sa structure physique et géométrique seront mis en évidence, en particulier pour les régimes de fortes densités de courant.

3.2.1 Description de la structure

La figure 3.1(b) représente une coupe transversale du transistor choisi pour notre étude, tandis que la figure 3.1(a) en donne le schéma électrique équivalent. Le transistor est utilisé dans la configuration la plus courante pour une structure de protection ESD, c'est-à-dire avec la base et l'émetteur court-circuités à la masse. Il est réalisé dans une technologie $1,8 \mu\text{m}$ dite de puissance intelligente ou smart power qui permet de réunir sur une même puce les fonctions de contrôle et de puissance. Elle utilise une épitaxie N sur un substrat P. L'isolation entre les différents caissons est assurée par des diffusions profondes P^+ .

Le transistor npn vertical de la figure 3.1 est celui qui est proposé par la bibliothèque standard de la technologie. Il possède une très faible résistance de collecteur grâce à une couche enterrée très dopée et une diffusion profonde N^+ . Deux modifications sont réalisées vis-à-vis de la structure standard : une résistance de ballast est introduite dans l'émetteur en éloignant le contact du bord de la diffusion, et un seul des contacts de base est relié à la masse.

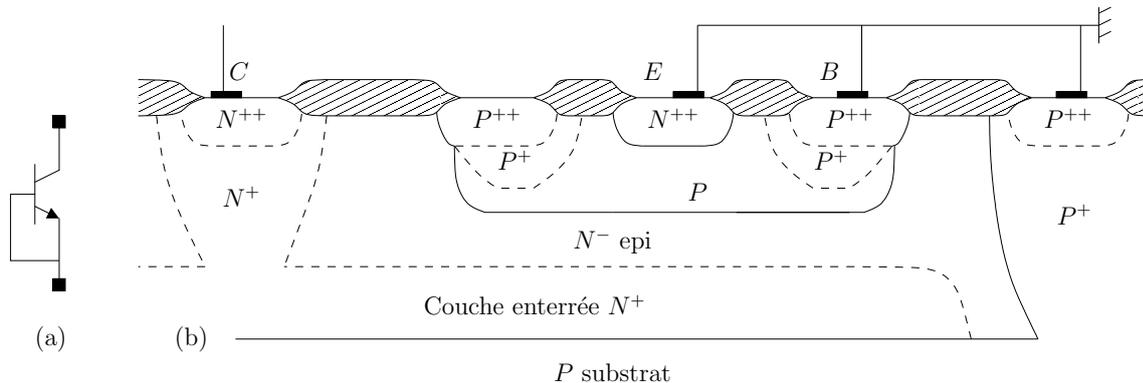


FIG. 3.1 – Schéma électrique (a) et coupe technologique (b) du transistor bipolaire NPN vertical.

3.2.2 Modes de fonctionnement lors d'une décharge électrostatique

Lorsqu'un stress ESD HBM négatif est appliqué sur le collecteur, le composant conduit le courant au travers de sa diode collecteur-base polarisée en direct. Sa robustesse est très grande dans ce mode de fonctionnement car la puissance dissipée dans la structure est faible et son échauffement limité. Ce cas ne requiert pas de précaution particulière lors de la conception.

Dans le cas d'un stress ESD positif, le transistor bipolaire NPN se déclenche lorsque la tension sur le collecteur dépasse la tension de claquage de la jonction collecteur-base. Le courant d'avalanche qui parcourt la résistance interne de base permet la mise en conduction directe de la jonction émetteur-base. L'effet bipolaire, conjugué à la multiplication par avalanche à la jonction collecteur-base, conduit alors au phénomène de repliement (ou snapback) de la tension de collecteur. Le transistor fonctionne alors en régime autopolarisé (cf. chapitre précédent).

Notons que par construction la tension de claquage collecteur-base correspond à celle de la jonction P^{++}/N_{epi}^{-} .

3.2.3 Caractéristiques électriques

En fonction des variations géométriques et des options pour ajuster la tension de déclenchement, cette structure permet d'obtenir d'excellentes performances ESD allant de 3,5 kV (soit 26 V par micron de longueur du doigt d'émetteur) à plus de 10 kV (75 V/ μm) pour des stress HBM positifs [60].

La caractéristique TLP du transistor de la figure 3.1 est présentée dans la figure 3.2. Deux aspects remarquables se dégagent de l'observation des caractéristiques électriques. D'une part, la tension de repliement du composant ($V_{t1}=70\text{V}$) est très supérieure à la tension de claquage de la jonction collecteur-base ($BV_{CB}=57\text{V}$). D'autre part, la tension de maintien ($V_H=25\text{V}$) est très inférieure à la tension de claquage en base ouverte ($BV_{CE0}=32\text{V}$). Nous allons montrer que

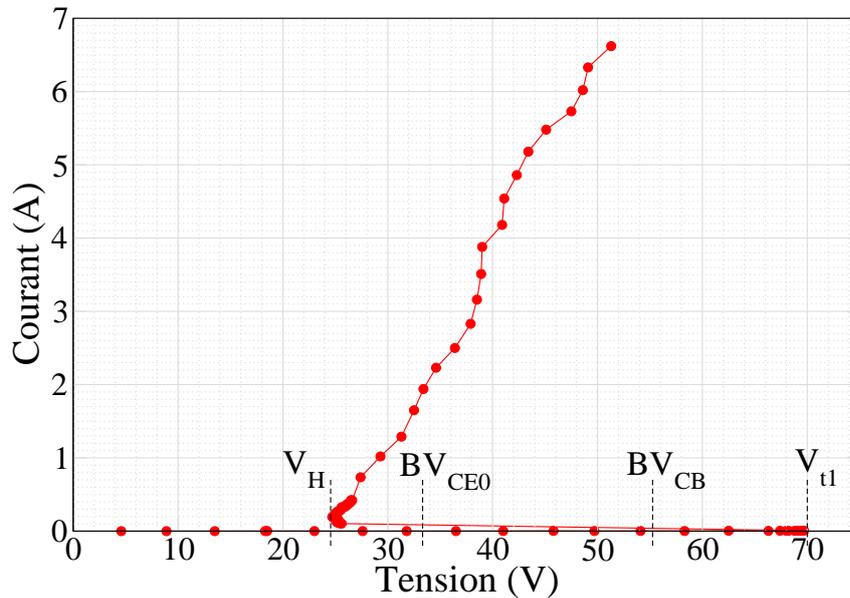


FIG. 3.2 – Caractéristique TLP du composant NPN vertical. La tension de claquage de la jonction collecteur-base BV_{CB} et la tension de claquage en base ouverte BV_{CE0} , mesurées à $10\ \mu\text{A}$, sont également reportées

l'origine de ces écarts provient des niveaux de courants auxquels sont mesurées ces différentes grandeurs. En effet, les tensions BV_{CB} et BV_{CE0} sont mesurées pour de faibles niveaux de courant ($10\ \mu\text{A}$), contrairement aux tensions V_{t1} et V_H qui sont typiquement obtenues pour des courants de l'ordre du mA à plusieurs dizaines ou centaines de mA.

L'étude approfondie du fonctionnement du transistor lors d'un stress ESD, au moyen de la simulation physique bidimensionnelle, va nous permettre de lier la valeur de ces tensions aux caractéristiques physiques et à la géométrie du composant.

3.3 Étude de son fonctionnement par la simulation physique bidimensionnelle

La structure du transistor NPN, comme présentée dans la figure 3.1, a été transcrite dans le simulateur. Les conditions de simulations ont été décrites dans la section 3.1. L'observation de

l'évolution des grandeurs physiques comme le champ électrique, la densité de courant, le taux de génération par impact et la température, durant une décharge HBM permet d'analyser en détail les phénomènes.

3.3.1 Surtension avant le repliement

Le premier aspect remarquable provient de la forte valeur de V_{t_1} vis-à-vis de la tension de claquage BV_{CB} . Ceci laisse supposer à priori, que le transistor bipolaire n'est pas déclenché avant d'atteindre la tension V_{t_1} et que le courant circule dans la diode collecteur-base polarisée en inverse.

Pourtant, pour des tensions de collecteur comprises entre BV_{CB} et V_{t_1} , les résultats de simulation indiquent que le courant circule majoritairement au travers de l'émetteur du transistor et non de son contact de base. Pour une tension de collecteur (V_C) d'un volt inférieure à V_{t_1} , l'observation de la densité de courant dans le composant (Fig. 3.3(a)) montre que le courant se répartit sur toute la largeur de l'émetteur avec des densités de l'ordre de $5.10^3 A/cm^2$. (La densité de courant est plus importante sur le bord d'émetteur opposé au contact de base et diminue lorsque l'on s'approche de ce contact. Cette non-uniformité est due à la chute de potentiel dans la résistance de base, et sera traitée plus en détail par la suite.)

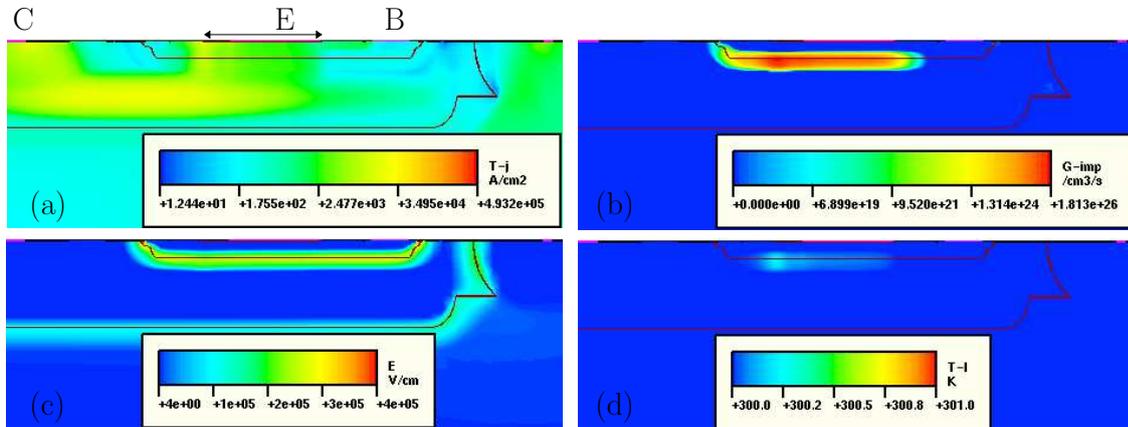


FIG. 3.3 – Densité de courant (a), taux de génération par impact (b), champ électrique (c) et température (d) avant le repliement de la tension de collecteur ($V_C = V_{t_1} - 1$ Volt), obtenus par simulation d'une coupe bidimensionnelle du transistor bipolaire NPN vertical soumis à une décharge HBM de 2 kV.

Le transistor bipolaire est donc déclenché et se trouve en régime autopolarisé comme en témoigne l'important taux de génération par avalanche à la jonction collecteur-base sous la diffusion d'émetteur (Fig 3.3(b)). Toutefois, sa tension de maintien est importante car sa jonction collecteur-base est constituée par la jonction P/N_{epi}^- dont la tension de claquage BV_{CB}^{NPN} est importante⁴.

Le transistor a été déclenché par la diode collecteur-base formée par la jonction P^{++}/N_{epi}^- et dont la tension de claquage BV_{CB} est inférieure à BV_{CB}^{NPN} . En dépit du fort champ électrique qui règne à la jonction P^{++}/N_{epi}^- avant le repliement (Fig. 3.3(c)), le taux de génération y est faible (Fig. 3.3(b)) ainsi que la densité de courant (Fig. 3.3(a)). Malgré la tension élevée à ses bornes, le courant circulant dans cette diode est limité car sa résistance série est relativement importante, en particulier du côté de l'épitaxie N. Si, par construction, la valeur de cette résistance vient à être réduite, l'augmentation du courant et l'échauffement trop important de cette

⁴La tension de maintien d'un transistor autopolarisé dépend directement de la tension de claquage de sa jonction collecteur-base (cf. chapitre précédent).

diode pourrait entraîner la dégradation prématurée du composant. Cet aspect particulier sera expérimentalement mis en évidence dans la suite sur un autre type de composant.

Si la conduction du transistor bipolaire au-delà de la tension BV_{CB} et avant le repliement de la tension de collecteur vient d'être démontrée, il n'en reste pas moins que sa caractéristique présente une résistance dynamique très importante dans ce régime de fonctionnement à moyen courant. L'origine de ce phénomène découle de l'effet des fortes densités de courant dans la zone de charge d'espace de la jonction collecteur-base. Dans cette zone, les porteurs transitent à leur vitesse limite (v_l) et la charge électrique qu'ils transportent peut devenir non négligeable vis-à-vis de celle des impuretés dopantes ionisées. L'effet des charges mobiles devient important, en premier lieu, dans la région de collecteur la plus faiblement dopée. Si l'on néglige le courant de trous⁵ dans cette région, la charge due au courant d'électrons vient se soustraire à celle des dopants ionisés. La neutralisation sera totale pour une densité de courant critique $J_n^{crit} = N_D q v_l$ où N_D est la valeur du dopage de la région faiblement dopée de collecteur (Fig 3.4(b)). Pour le composant considéré ici, la valeur de cette densité critique de courant est d'environ 6.10^3 A/cm^2 , et correspond bien aux densités de courants observées dans la figure 3.3.

Les caractéristiques électriques de la jonction collecteur-base sont donc profondément modifiées en fonction de la densité du courant. En particulier, la valeur de sa tension de claquage doit être reconsidérée, d'autant plus qu'elle détermine la valeur de la tension de maintien du transistor. Toute augmentation ou diminution de la valeur de la tension de claquage entraîne une variation dans le même sens de la tension de maintien. L'effet de la densité des porteurs libres tendant à diminuer le dopage de la zone faiblement dopée (Fig 3.4(a)(b)), la jonction présente une tension de claquage croissante avec la densité de courant. La croissance concomitante de la tension de maintien explique ainsi l'importante valeur de la résistance dynamique observée sur la caractéristique TLP.

Lorsque la densité de charge des porteurs mobiles est du même ordre de grandeur que le dopage N_D , la région peu dopée tend à devenir intrinsèque, c'est-à-dire que la densité nette de charge électrique dans cette région devient faible (Fig 3.4(c)(d)(e)). Dans ces conditions, l'extension de la charge d'espace se trouve rapidement limitée par la couche enterrée fortement dopée. La jonction collecteur-base est alors assimilable à une diode PIN. En outre, la tension de claquage des diodes PIN ne dépend que de la distance entre ses régions plus fortement dopées, et est donnée par la relation [104] :

$$BV_{pin} = \left(\frac{W_i^6}{1,8.10^{-35}} \right)^{\frac{1}{7}} \quad (3.1)$$

avec W_i (Fig 3.4(d)) la longueur de la zone intrinsèque en cm.

L'augmentation de la tension aux bornes du composant présente donc une limite maximale lorsque le régime de diode PIN est atteint. La résistance dynamique va alors fortement diminuer⁶ et la tension aux bornes du composant va atteindre sa valeur maximale qui dépend directement de la tension de claquage de la diode PIN et donc de la distance W_i .

Le régime de diode PIN correspond à des valeurs de densité de courant de l'ordre de la densité critique. Il marque la fin du régime de fonctionnement que nous désignerons par la suite comme le régime de moyen courant. Au delà, le transistor bascule dans le régime de fonctionnement que nous qualifierons de fort courant. L'entrée dans ce dernier régime s'accompagne d'une chute brutale de la tension, qui correspond au repliement de la caractéristique TLP. Dans le même temps, on assiste au phénomène d'élargissement de la région de base ainsi qu'à une importante modification des caractéristiques de la jonction collecteur-base (Fig 3.4(f)) qui est déplacée à la frontière des régions N^- et N^{++} du collecteur.

⁵Ceci est toujours valable en première approximation pour les transistors NPN dans lesquels le courant de trous représente au maximum un tiers du courant total (cf. chapitre précédent).

⁶La diminution de la résistance dynamique n'est pas visible sur la caractéristique TLP du transistor NPN étudié ici, mais sera mise en évidence sur d'autres structures par la suite.

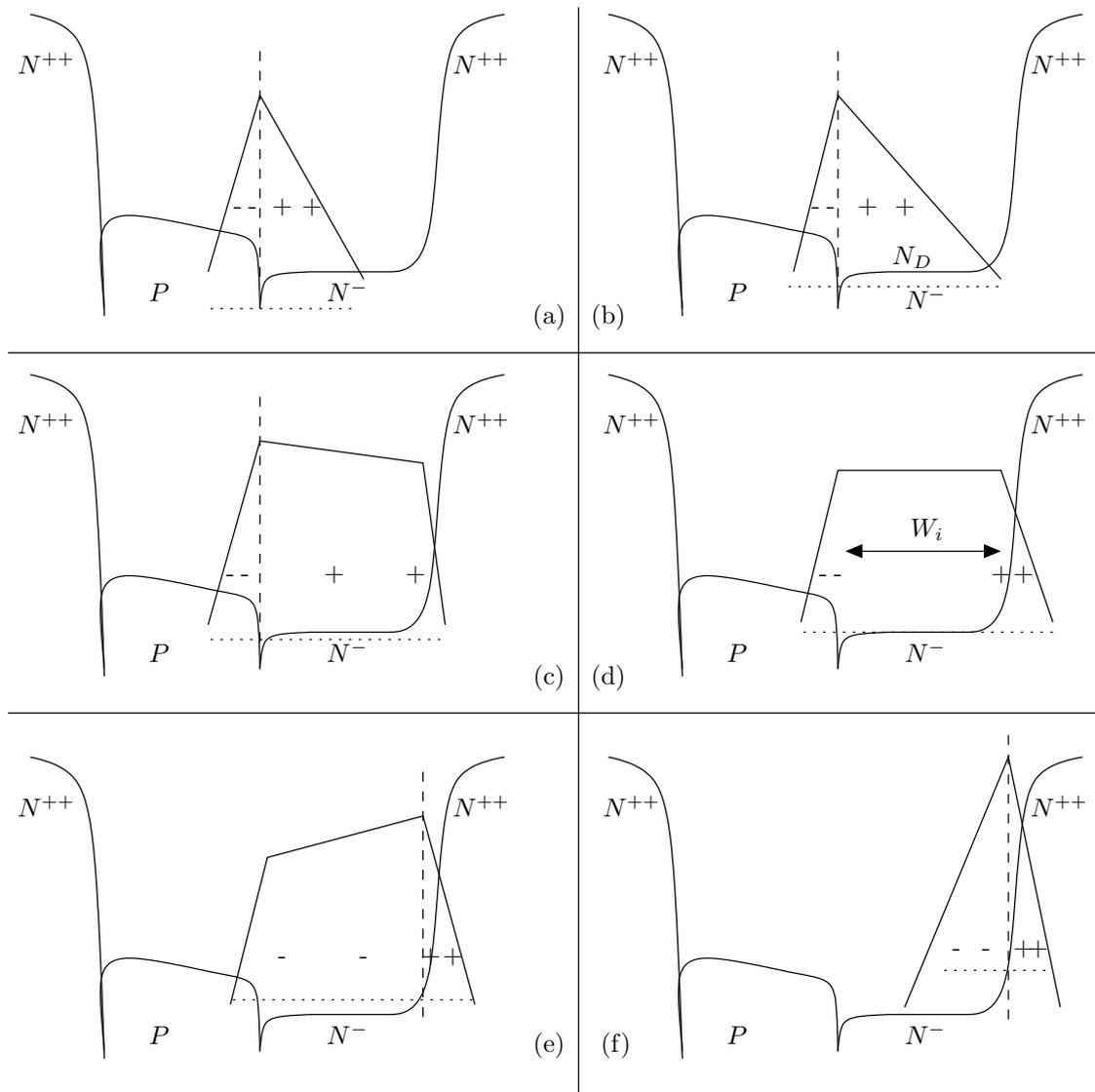


FIG. 3.4 – Évolution schématique du profil du champ électrique dans la jonction collecteur-base, pour des densités de courant croissantes. La densité de charge due au courant d'électron est représentée en traits pointillés courts dans la zone de charge d'espace. La position de la jonction (lieu de changement de signe de la charge d'espace) est représentée en traits pointillés longs.

3.3.2 Repliement de la tension, élargissement de la région de base

Pour des densités de courants très supérieures à la densité critique, le maximum de champ électrique, initialement situé à la jonction métallurgique, se déplace à la frontière de la couche enterrée et de l'épitaxie (Fig. 3.4(f) et Fig. 3.5(c)). Ce phénomène est identique à l'effet Kirk [82] avec lequel il partage la cause fondamentale. Il existe cependant une différence importante car la jonction collecteur-base est ici le siège d'une importante multiplication par avalanche, ce qui n'est pas le cas dans le phénomène décrit par Kirk. En outre, l'effet Kirk est une limite dans le fonctionnement d'un transistor bipolaire. Dans le cas présent, il s'agit de son régime de fonctionnement normal. La figure 3.6 représente les densités de porteurs et le champ électrique selon une coupe verticale dans la région de plus forte densité de courant, pour différentes valeurs du courant dans le composant. Pour de faible (FD) et moyenne (MD) densités de courant, on reconnaît le comportement décrit dans la section précédente. À forte densité de courant (HD), le déplacement du maximum de champ électrique s'accompagne d'un élargissement de la base

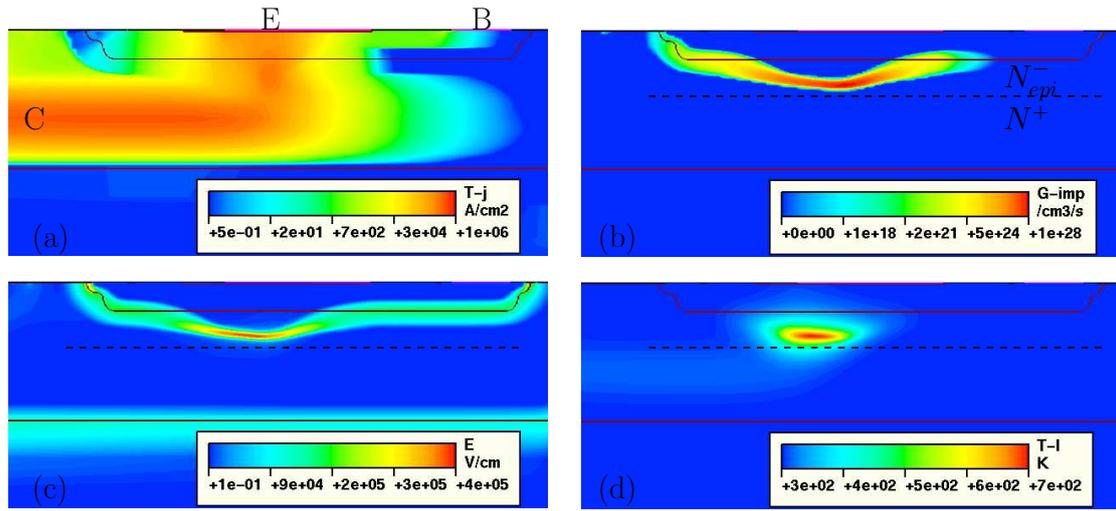


FIG. 3.5 – Densité de courant (a), taux de génération par impact (b), champ électrique (c) et température (d) au pic de courant d'une décharge HBM de 2kV obtenus par simulation d'une coupe bidimensionnelle du transistor bipolaire NPN vertical.

physique dont la profondeur $W_{b_{eff}}^{HD}$ devient supérieure à la profondeur physique W_b alors qu'elle était inférieure $W_{b_{eff}}^{F-MD}$ dans les régimes de faible et moyenne densité de courant (Fig. 3.6).

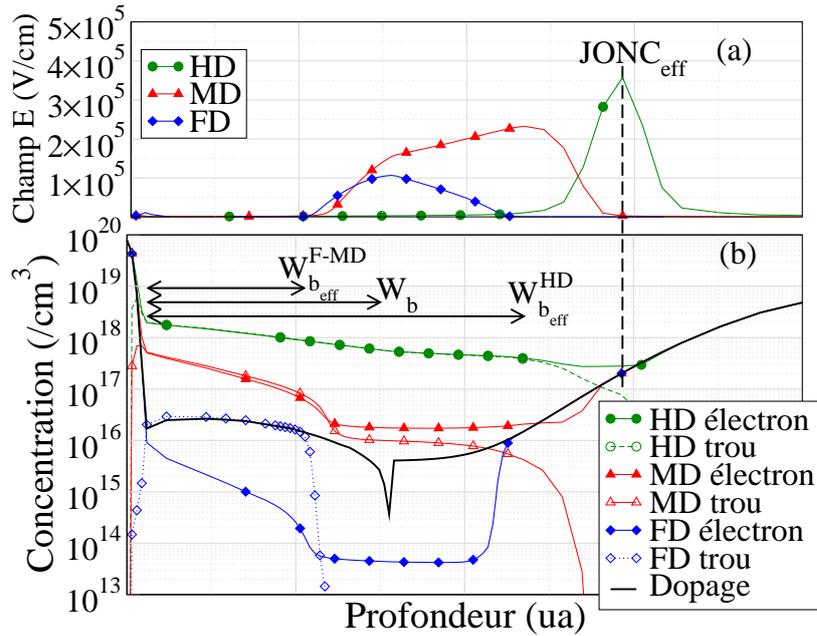


FIG. 3.6 – Coupe des densités de porteurs (a) et du champ électrique (b) pour différentes densités de courant dans le composant : haute densité HD, moyenne densité MD et faible densité FD.

De la même manière qu'une jonction C-B métallurgique, nous pouvons définir à forte densité de courant une jonction C-B effective située à la frontière des régions faiblement et fortement dopées de collecteur et dont la position correspond au maximum de champ électrique ($JONC_{eff}$ Fig. 3.6).

Le fonctionnement du composant à fort courant s'identifie alors à celui décrit au chapitre précédent par l'approche semi-analytique. En effet, la densité des porteurs libres est si grande

que les régions où le dopage n'a pas de valeur très importante ($< 5 \cdot 10^{16}/\text{cm}^3$) ne joue qu'un rôle négligeable dans le fonctionnement du composant (Fig. 3.6(b)). Comme nous l'avons vu au chapitre précédent, la tension de claquage de la jonction collecteur-base effective est alors déterminée par une concentration effective qui est du même ordre de grandeur que la densité des porteurs libres. Par conséquent, la tension de claquage de cette jonction est maintenant beaucoup plus faible que celle de la jonction métallurgique P/N_{epi}^- (BV_{CB}^{NPN}) et la tension de maintien correspondante est plus faible que la tension statique BV_{CE0} .

«Strong snapback»

L'origine du «strong snapback» ou sur-repliement observé par Mergens dans les composants LDMOS [105], apparaît ici clairement. En effet, la tension de maintien dépend de la tension de claquage de la jonction effective collecteur-base. Comme nous allons le voir, cette tension de claquage est généralement basse et dépend du profil de dopage entre les régions fortement et faiblement dopées du collecteur. Ainsi, la tension de claquage de la jonction métallurgique collecteur-base BV_{CB} qui dépend du dopage de la base et de la région N^- de collecteur ne détermine pas la tension de maintien. La valeur de BV_{CB} peut être très importante alors que la tension de maintien sera particulièrement faible. Dans ces conditions, on observe un repliement de la caractéristique TLP qui peut paraître extraordinairement important («extraordinary strong snapback» [105]).

3.3.3 Phénomènes de focalisation du courant

Deux phénomènes principaux sont à l'origine de la focalisation du courant : d'une part, la non-uniformité du courant dans l'émetteur, entraînée par la chute de potentiel dans la résistance intrinsèque de base, et d'autre part, le comportement de la région de collecteur à très forte densité de courant.

Influence de la résistance de base

Pour des niveaux de courant moyen, dans la partie de la courbe TLP correspondant à la surtension avant le repliement (Fig. 3.3(a)), la densité de courant n'est pas uniforme sur la largeur de l'émetteur et se concentre à son extrémité opposée au contact de base. Contrairement au fonctionnement habituel d'un transistor bipolaire NPN, le courant de base sort du contact car il est délivré par la multiplication par avalanche dans la région de collecteur. Ce courant s'écoule de gauche à droite dans la résistance intrinsèque de base R_B^{int} (Fig. 3.7), et provoque une variation du potentiel dans la base sur la largeur de l'émetteur. La polarisation de la diode émetteur-base en direct est donc plus importante du côté opposé au contact de base et la densité de courant plus grande. De la même manière, pour des densités de courant élevées (Fig. 3.7) l'essentiel du courant s'écoule sur une partie seulement de la largeur de l'émetteur située à l'opposé du contact de base.

L'introduction d'une résistance de ballast dans l'émetteur, réalisée en éloignant le contact du bord de la diffusion, permet de compenser en partie l'effet focalisant de la résistance de base. Elle permet aussi d'atténuer l'apparition d'une région de plus forte densité de courant à l'extrémité de l'émetteur. En effet, la direction du courant étant verticale, le flanc de la diffusion impliquerait une plus forte densité de courant en bord d'émetteur.

L'effet focalisant induit par la résistance de base est connu de longue date dans les transistors bipolaires [106]. L'étude analytique de ce phénomène est particulièrement ardue. Les solutions analytiques proposées dans la littérature font appel à trop de simplifications et d'hypothèses pour être une aide efficace à la conception des dispositifs étudiés ici. On se limitera donc à l'emploi de la simulation numérique, associée à la connaissance de règles simples et tendances fondamentales. Retenons en particulier, que plus la résistivité de la diffusion de base intrinsèque est faible (donc son dopage important), moins la focalisation qu'elle entraîne est importante [107].

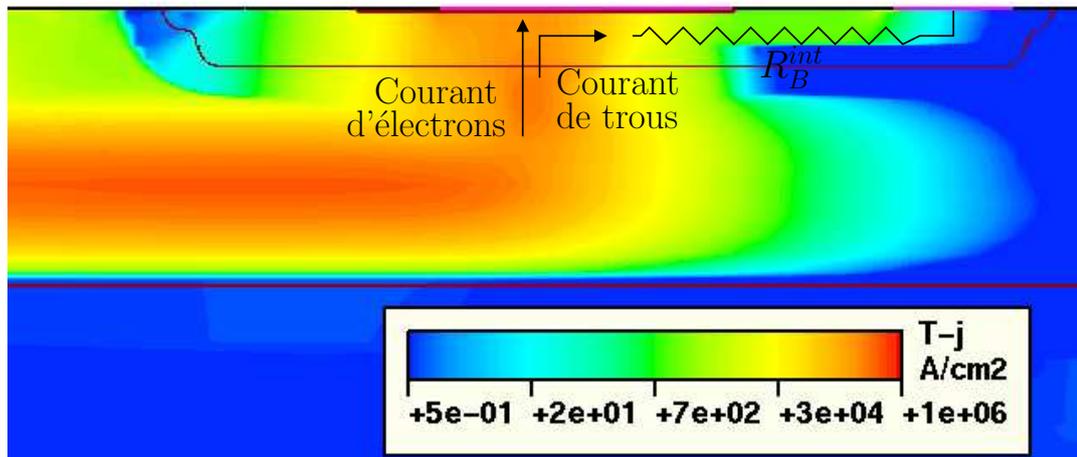


FIG. 3.7 – Focalisation du courant due à la résistance de base intrinsèque du transistor bipolaire NPN vertical. Densité de courant au pic de courant d'une décharge HBM de 2kV, obtenue par simulation d'une coupe bidimensionnelle.

Influence de la région de collecteur

La répartition du courant à moyen niveau (Fig. 3.3(a)) et à fort niveau (Fig. 3.5(a)), montre une différence dans le degré de focalisation qui apparaît beaucoup plus important à fort courant.

Dans le régime de moyens courants, la non-uniformité du courant provient seulement des effets de la résistance intrinsèque de base. Dans ce régime de fonctionnement, nous avons montré qu'une augmentation de la densité de courant dans la jonction collecteur-base implique une croissance de sa tension de claquage. Ce phénomène tend à uniformiser la densité du courant et compense en partie la focalisation due au courant de base.

Le régime de fort courant a été étudié en détail au chapitre précédent. Nous avons montré que la chute de tension aux bornes de la région de collecteur pour une densité de courant croissante, associée à la diminution de la tension de claquage de la jonction effective collecteur-base, entraîne la focalisation du courant. Seules les résistances séries du composant, en particulier celle du collecteur, permettent de limiter cet effet. Ce phénomène de focalisation vient donc s'ajouter, dans le régime de fort courant, à celui provoqué par la résistance de base. La focalisation due au courant de base est d'ailleurs secondaire dans ce régime, mais détermine cependant la position de la région de forte densité de courant à l'extrémité de l'émetteur opposée au contact de base.

Nature électrique des focalisations

L'origine des focalisations est de nature purement électrique. Avant le repliement de la tension de collecteur, l'échauffement du composant est négligeable. L'élévation de température juste avant le repliement est de l'ordre de quelques dixièmes de degré dans la figure 3.3(c). Les aspects thermiques importants se manifestent plus tard, après la focalisation du courant associée à la formation de la jonction effective.

Conclusions

La distribution du courant est relativement uniforme avant le repliement et devient fortement non uniforme ensuite. Avant le repliement, dans le régime de moyen courant, seul l'effet de la résistance de base tend à focaliser le courant, le comportement de la région de collecteur tendant à l'uniformiser. Dans ce régime, la tension de claquage de la jonction C-B du transistor augmente avec la densité de courant. Cette augmentation est à l'origine de la forte résistance dynamique observée sur la caractéristique TLP.

Après le repliement, pour le régime de fort courant, les effets focalisants de la résistance de base et de la région de collecteur se conjuguent. Le courant est fortement non-uniforme dans le composant et le maximum du champ électrique est déplacé de la jonction C-B métallurgique à une jonction effective située à la frontière des régions N^- et N^{++} du collecteur. Dans ces conditions de fonctionnement, le comportement du transistor peut être décrit par le modèle semi-analytique développé au chapitre précédent.

Contrairement aux intuitions habituelles, la focalisation du courant est ici bénéfique car elle s'accompagne d'une importante diminution de la tension de maintien qui permet de dissiper une énergie globalement plus faible dans le composant.

La faible dimension des composants de protection (de l'ordre de 10^4 cm^2) vis-à-vis de l'importante valeur d'un courant ESD (plusieurs Ampères), entraîne inévitablement de très fortes densités de courant. Les phénomènes de focalisation à fort courant ne peuvent donc être occultés. L'étude semi-analytique menée au chapitre précédent pour les très fortes densités de courant revêt donc un aspect fondamental pour comprendre le fonctionnement d'un transistor bipolaire lors d'une décharge électrostatique. Le tableau 3.1 résume le comportement d'un TBA NPN possédant une région de collecteur faiblement dopée, dans les différents régimes de fonctionnement.

	Faible courant	Moyen courant	Fort courant
Densité de courant dans le composant	$J \ll J_{crit}$	$J \approx J_{crit}$	$J \gg J_{crit}$
Ordre de grandeur du courant	μA	mA	A
Tension aux bornes du composant	BV_{CB}	$\approx BV_{CB}$ + forte résistance série	V_H
Position de la jonction C-B	Jonction métallurgique	Jonction métallurgique	Jonction délocalisée (N^-/N^{++})
Variation de la tension de claquage de la jonction C-B avec la densité de courant	Nulle	Augmentation	Diminution
Uniformité du courant dans le composant	Uniforme	Quasi-uniforme	Non uniforme

TAB. 3.1 – Comportement du transistor bipolaire NPN à collecteur faiblement dopé (N^-/N^+) lors d'une décharge électrostatique dans les différents régimes de fonctionnement.

3.3.4 Comportement selon la longueur de l'émetteur

Afin d'optimiser un transistor bipolaire pour réaliser une structure de protection ESD, il est important d'étudier l'uniformité du courant sur la longueur de l'émetteur.

La structure étudiée

Pour étudier la distribution du courant, une coupe longitudinale de la structure a été transcrite dans le simulateur électrique (Fig. 3.8(e)). Aucun contact n'est pris dans la base, ainsi laissée flottante. La validité de cette approche repose sur deux approximations :

1. À fort courant, le courant sortant du contact de base est négligeable par rapport au courant total. (Ceci a été abordé au chapitre précédent).
2. Ce courant, qui traverse la résistance intrinsèque de base, circule en grande partie orthogonalement au plan de la coupe, il ne peut donc pas entraîner d'importante variation de potentiel dans la région de base du plan de coupe.

Ces approximations raisonnables, évitent l'utilisation d'une simulation tridimensionnelle. Dans les composants étudiés ici, les gradients des grandeurs physiques sont très importants. Le nombre de points de maillage garantissant une discrétisation correcte est donc très grand. Les ressources informatiques nécessaires pour mener correctement à bien une simulation 3D de ce type sont donc considérables.

La structure simulée est présentée dans la figure 3.8(e). Afin, de simplifier sa géométrie, les résistances série d'émetteur et de collecteur ont été estimées et introduites par le biais d'une résistance distribuée de contact. La simplicité de la géométrie permet de réaliser un maillage fin et régulier de la structure, en particulier à la frontière des zones faiblement et fortement dopées du collecteur. Le facteur d'aire utilisé est égal à la largeur de la zone focalisée observée dans la coupe selon la largeur de l'émetteur.

Résultats de simulation

La simulation électrothermique est réalisée pour un stress de 500 V HBM appliqué sur le collecteur (émetteur relié à la masse). L'évolution de la tension de collecteur et du maximum de la température dans la structure sont représentés dans la figure 3.9. L'état interne du composant, en termes de champ électrique, densité de courant et de température, est décrit dans la figure 3.8 pour différents instants au cours de la décharge.

Initialisation d'une zone de forte densité de courant

Malgré l'absence de toute résistance de base, le courant est très rapidement focalisé au-dessous de l'une des extrémités de l'émetteur (Fig. 3.8(a)). Cette focalisation résulte des propriétés de la région de collecteur à fort courant (la tension de claquage de la jonction C-B effective diminue avec la densité de courant), de la même manière que dans la coupe selon la largeur de l'émetteur.

Malgré la présence d'une résistance de ballast, les bords d'émetteur sont le siège de plus fortes densités de courant. Ils sont donc des lieux préférentiels pour la transition vers le régime de fort courant.

Très rapidement, la formation d'une seule zone focalisée de courant apparaît dans le composant. La symétrie de la structure est pourtant parfaite dans cette simulation, même en termes de maillage. Dans ce cas, le déséquilibre entre les deux extrémités d'émetteur, est attribuable aux erreurs de calcul numérique. Ces erreurs sont rapidement amplifiées par le couplage électrique et spatial des deux phénomènes de multiplication constitués d'une part, par la multiplication par avalanche à la jonction collecteur-base et l'effet bipolaire (efficacité d'injection de la jonction émetteur-base), d'autre part.

Nous avons pu relever l'extrême sensibilité du lieu d'initialisation en fonction des variations de maillage, des conditions de polarisation ou de légers déséquilibres introduits dans la structure (soit au travers du dopage, soit en ajoutant sur la région de base un contact de petite taille lié à la masse par une résistance de forte valeur ($M\Omega$)). Tout cela témoigne du caractère fortement chaotique que peut avoir ce système. Ce comportement découle directement du couplage étroit qu'il existe entre les deux phénomènes de multiplication et du phénomène d'emballement qu'ils entraînent. Cependant, ce caractère peut être partiellement contrôlé en introduisant volontairement une dissymétrie dans la structure.

Dans la littérature, l'extrême sensibilité du lieu d'initialisation a été rapportée expérimentalement dans les transistors bipolaires NPN aussi bien verticaux que latéraux [40, 108, 109, 110]. Un

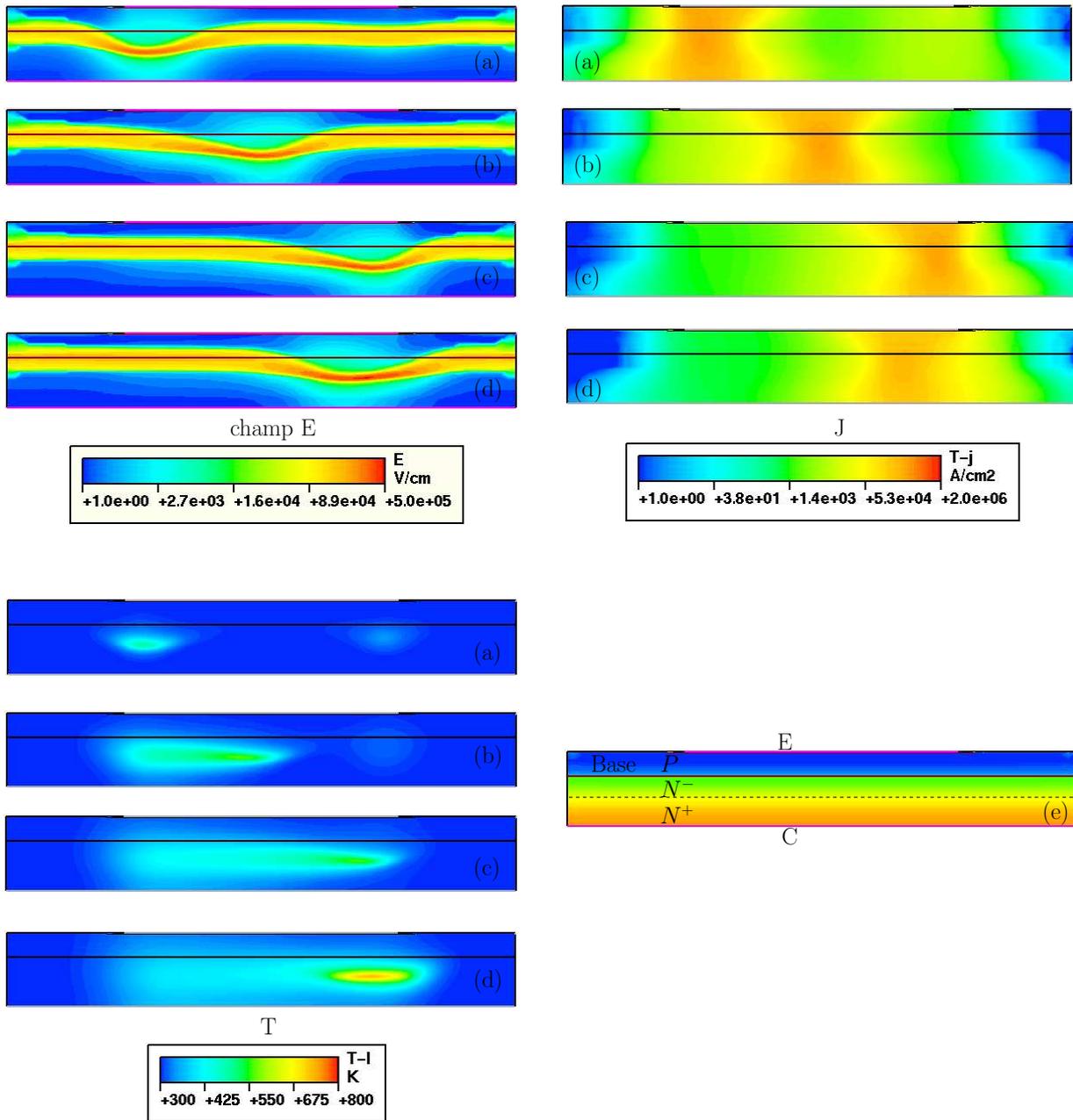


FIG. 3.8 – Simulation bidimensionnelle d'un stress de 500 V HBM dans une coupe selon la longueur du doigt d'émetteur. Distributions du champ électrique, de la densité de courant et de la température à 10 ns (a), 30 ns (b), 50 ns (c) et 70 ns (d) (voir figure 3.9). (e) Représentation des dopages dans la coupe.

caractère totalement aléatoire a été observé pour une structure parfaitement symétrique [110]. L'initialisation dans une région préférentielle a été observée par C. Russ et ses coauteurs qui l'ont attribuée, dans leur cas, à une dissymétrie dans la courbure de jonction [40].

Déplacement de la zone de forte densité de courant

Une puissance thermique importante est dissipée dans la zone focalisée, mais le déplacement de cette zone sur toute la longueur du doigt permet d'homogénéiser la dissipation de puissance et la température. Ce déplacement peut être expliqué par la diminution du phénomène d'ava-

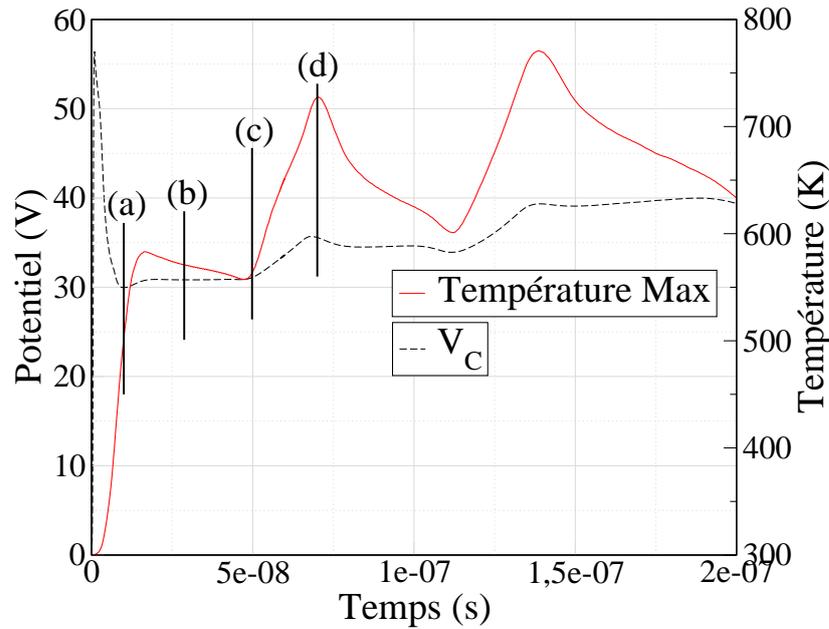


FIG. 3.9 – Évolutions de la tension de collecteur et du maximum de la température au cours du temps, obtenues par simulation électrothermique bidimensionnelle d'un stress de 500 V HBM dans une coupe selon la longueur du doigt d'émetteur. Les distributions du champ électrique, de la densité de courant et de la température à 10 ns (a), 30 ns (b), 50 ns (c) et 70 ns (d) sont représentées sur la figure 3.8.

lanche avec la température. Les coefficients d'ionisation sont en effet des fonctions rapidement décroissantes de la température (Fig. 3.10).

Le fort champ électrique et la forte densité de courant qui règnent dans la zone focalisée entraînent une augmentation locale de la température. La diminution des coefficients d'ionisation qui en résulte provoque le déplacement du maximum de densité de courant en bordure de la zone et conduit à son mouvement apparent vers des régions plus froides.

Lorsque la zone focalisée atteint le bord d'émetteur, l'élévation de température devient importante jusqu'à l'inversion de son sens de mouvement. Les extrémités de la structure apparaissent donc comme des régions où le risque d'initialisation d'un second claquage thermique est grand. Elles devront faire l'objet de précautions particulières pour optimiser les performances d'un composant.

Remarques sur la validité et les limitations des simulations

La plupart des paramètres physiques utilisés dans les simulateurs ne sont validés que pour des températures allant jusqu'à 600K. Nous avons donc volontairement limité les simulations à un stress ESD de relativement faible valeur qui garantit des élévations de températures compatibles avec le simulateur.

Il n'existe pas de méthode simple pour vérifier expérimentalement le mécanisme de déplacement de la zone de forte densité de courant. L'uniformité du courant est généralement observée dans ce type de composant soit par la technique d'interférométrie laser [114], soit par la microscopie à émission lumineuse [115]. Ceci n'est pas en contradiction avec les résultats de simulation présentés ici, car ces deux techniques ne donnent pas une image directe de la densité de courant et réalise des moyennes dans le temps et/ou sur un grand nombre d'impulsions. Des résultats expérimentaux obtenus par la technique d'imagerie thermique par interférométrie laser, ont été présentés par Pogany et ses coauteurs [108]. Les auteurs ont défini une vitesse d'étalement de la

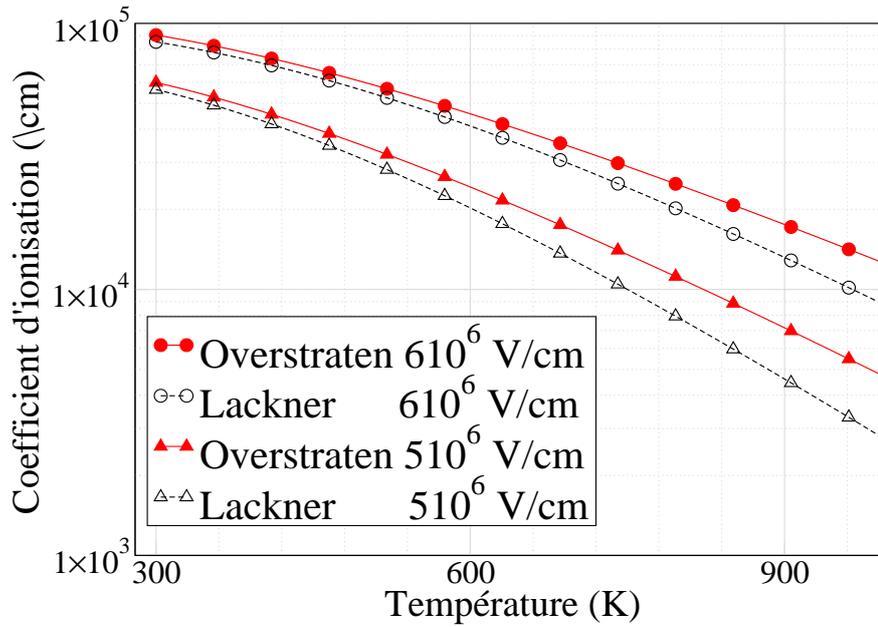


FIG. 3.10 – Variation du coefficient de multiplication par avalanche des électrons en fonction de la température pour deux valeurs de champ électrique, selon le modèle de Van Overstraeten de Man [111, 112] et le modèle de Lackner [113].

température qui est de l'ordre de $1\mu\text{m}/\text{ns}$. La vitesse d'étalement de la température obtenue dans nos simulations est du même ordre de grandeur. Cela permet d'accorder une certaine validité à nos simulations.

Une version de la technique d'interférométrie laser permettant d'obtenir une image de la répartition de température dans les composants à un instant donné d'une seule impulsion de décharge a été développée récemment [116]. Le mouvement de la région de plus haute température pour des impulsions de relativement faible courant (comme dans le cas de la simulation présentée) a pu être mis en évidence [117]. Ces résultats récents, confirment donc la validité des résultats de simulation présentés ici.

Si l'on combine les résultats des simulations 2D réalisées dans les deux plans (largeur et longueur d'émetteur), il apparaît évident que la focalisation du courant est fondamentalement tridimensionnelle. Dans les plans de coupe selon la largeur ou la longueur du composant, la simulation représente une bonne approximation d'un point de vue électrique car les lignes de courant sont essentiellement verticales et sont donc contenues dans ces plans. En revanche, l'énergie thermique est dissipée dans toutes les directions de l'espace. Limiter le flux thermique à un seul plan conduit donc à largement surestimer la température en particulier dans la zone focalisée. Nous touchons ici à l'aspect limitant le plus l'utilisation de simulations 2D dans cette étude. Cela restreint en particulier les investigations aux cas de décharges de faible intensité. Toutefois, les observations réalisées donnent de précieux éléments pour comprendre le comportement réel du composant.

Seule la simulation 3D pourrait permettre une meilleure description du phénomène. Cependant, on se heurte à deux grandes difficultés pour l'utiliser efficacement :

- Un grand nombre de points de maillage est nécessaire pour discrétiser correctement les grandeurs physiques qui présentent ici de très forts gradients.
- La dynamique rapide du phénomène de déplacement induit une discrétisation temporelle très fine vis-à-vis de la durée totale d'une décharge.

À l'heure actuelle, les moyens informatiques nécessaires pour réaliser une simulation de ce type sont encore trop coûteux pour espérer obtenir des résultats dans un temps raisonnable. Cette

approche ne peut pas être retenue, du moins tant que les modèles physiques n'auront pas été validés sinon adaptés aux régimes de très forte injection et haute température.

3.3.5 Extrapolations aux très forts courants

Fort des éléments de compréhension apportés par les simulations 2D dans les deux plans de coupe, nous pouvons risquer une extrapolation du fonctionnement du transistor aux plus forts niveaux de courant.

Sous l'effet de la limitation de la densité de courant maximum entraînée par la résistance série de collecteur, la taille de la zone focalisée va progressivement s'étendre, pour des niveaux de courant croissants, selon la longueur de l'émetteur (comme observé dans le transistor MOS au chapitre précédent). Cette extension sera limitée par la taille de l'émetteur, à condition qu'aucun phénomène de second claquage thermique ne se produise entre temps. Au-delà de cette limite, le courant supplémentaire entraînera une augmentation de la densité de courant maximale. Durant l'extension, la tension de collecteur varie peu car la surface offerte au passage du courant s'accroît et la résistance dynamique de la structure apparaît très faible. Ensuite, le composant présente une résistance dynamique qui correspond aux résistances série de collecteur et d'émetteur. La densité de courant et la température au sein de la structure seront toujours fortement fluctuantes. La destruction du composant sera provoquée par l'initialisation locale d'un second claquage thermique dans la région de fort champ électrique et haute température située à la frontière des régions faiblement et fortement dopées de collecteur. Les analyses de défaillance réalisées grâce à la technique FIB (Focus Ion Beam) confirme que le défaut se situe bien à cette frontière [60].

Dans la dimension selon la largeur d'émetteur, l'extension de la zone focalisée est limitée par l'effet de la résistance de base. Toutefois, on remarque sur les figures 3.5 (a) et (d), un décalage entre la zone de plus forte densité de courant et la zone de plus haute température qui témoigne du déplacement de la zone focalisée sous l'effet de la température. Mais ce déplacement est ralenti sinon bloqué par l'effet de la résistance de base. Notons que dans les conditions de très forte densité de courant, la principale différence entre les simulations selon la longueur et la largeur d'émetteur réside dans la présence ou non d'un contact liant localement la région P de base au potentiel de masse. La robustesse d'un composant dont la base serait laissée flottante est donc potentiellement supérieure. Toutefois, les contraintes de conception demandent de garantir la passivité du composant de protection lors du fonctionnement normal du circuit. Les risques de perturbations, lors de transitoires rapides par exemple, sont trop élevés pour permettre l'utilisation d'un transistor à base flottante ou quasi-flottante (en la liant à la masse par une résistance de forte valeur).

3.4 Augmentation de la profondeur de la région de base effective

Les transistors bipolaires NPN possédant un collecteur faiblement dopé rendent possible la conception de composant dont la profondeur de base effective ($W_{b_{eff}}$) à fort courant peut être bien supérieure à la base physique (W_b). $W_{b_{eff}}$ peut être ajusté en modifiant la profondeur W_i de la région faiblement dopée de collecteur (Fig. 3.11). Hors problème de réalisation, nous allons montrer dans quelles mesures le découplage thermique entre la région chaude de collecteur et la jonction E-B peut être obtenu. Puis, nous étudierons l'impact de l'élargissement de la profondeur effective de base sur la distribution du courant dans la structure.

3.4.1 Découplage thermique

Nous avons montré au chapitre précédent que l'impact de la température à la jonction émetteur-base conduit à une augmentation du courant dans la zone échauffée. L'échauffement de la jonction E-B peut donc induire une focalisation du courant.

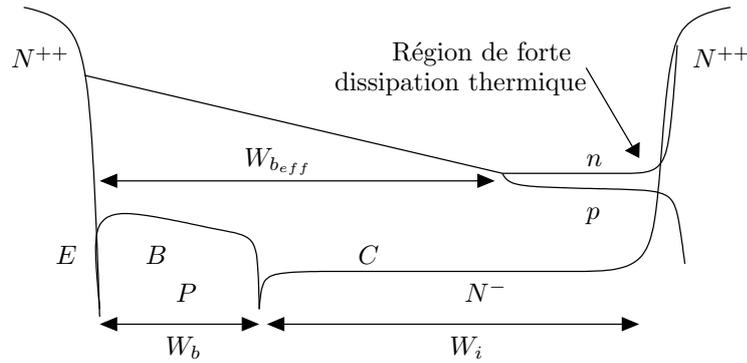


FIG. 3.11 – Profils des porteurs libres (électrons n et trous p) à forte densité de courant dans un transistor NPN à collecteur faiblement dopé.

La région de transition entre le collecteur faiblement et fortement dopé, siège d'un fort champ électrique, dissipe une importante quantité de chaleur qui va diffuser dans le milieu environnant. Compte tenu du coefficient de diffusion thermique du silicium $D = 0,09 \mu\text{m}^2/\text{ns}$, \sqrt{Dt} donne la distance parcourue par le flux de chaleur en fonction du temps. Pour une impulsion de 200 ns, la chaleur aura donc diffusé d'une distance de l'ordre de $4 \mu\text{m}$. Au-delà d'une distance de $10 \mu\text{m}$ de la source de chaleur, l'élévation de température devient totalement négligeable durant la durée de l'impulsion [118]. Ceci nous donne un ordre de grandeur de la distance $W_{b_{eff}}$ nécessaire pour assurer un bon découplage thermique entre la région chaude du collecteur et la jonction E-B au cours d'un transitoire de courant ESD.

3.4.2 Découplage électrique

Afin de limiter l'étude aux seuls aspects électriques, les simulations physiques suivantes sont réalisées dans des conditions isothermes. Pour réduire les temps de calculs, seule une moitié de la longueur d'émetteur est simulée, les conditions de simulation restant par ailleurs identiques aux précédentes.

La répartition de la densité de courant obtenue au pic du courant de décharge est représentée sur la figure 3.12(a) pour la structure originale (fort couplage) et dans la figure 3.12(b) pour une profondeur de base effective ($W_{b_{eff}}$) élargie par l'augmentation de l'épaisseur de la couche épitaxiée. Aucun mouvement de la zone de forte densité de courant n'est observé, ce qui corrobore l'origine thermique de son mouvement.

L'impact de l'élargissement de la base effective apparaît clairement sur la figure 3.12. La focalisation du courant à la jonction effective collecteur-base ($JONC_{eff}$) est peu modifiée. Comme nous l'avons vu, cette focalisation dépend principalement de la résistance série de collecteur. En revanche, l'uniformité de la distribution du courant sous l'émetteur est largement améliorée. Les coupes des densités de courant et du potentiel sous l'émetteur (Fig. 3.13) témoignent de l'uniformisation du courant.

La constriction du courant sous l'émetteur est donc l'image de la focalisation du courant à la jonction effective. La séparation spatiale du lieu de multiplication par avalanche et du phénomène d'injection à la jonction E-B permet de bénéficier de la décroissance de l'efficacité d'injection (du gain du bipolaire) avec l'augmentation de la densité de courant. Ce dernier phénomène tend à uniformiser le courant dans la jonction E-B de la même manière que la résistance de ballast d'émetteur. De plus, le découplage électrique permettra à la région de forte densité de courant de se déplacer plus facilement sous l'effet de la température.

La tension de maintien n'est pas significativement augmentée par l'élargissement de la région de base (Fig. 3.14). En effet, la chute de potentiel dans la longueur supplémentaire de la région de base, qui est en régime de très forte injection, est particulièrement faible.

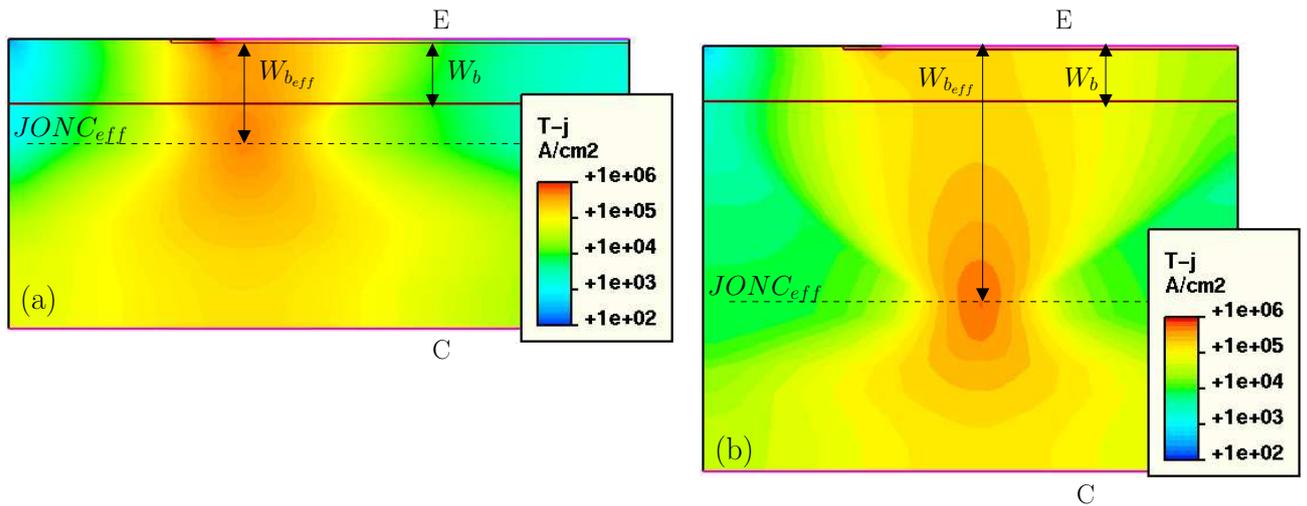


FIG. 3.12 – Simulation électrique bidimensionnelle de l'influence du découplage entre l'émetteur et la région de génération par avalanche. La densité de courant est représentée dans le cas d'un fort couplage (a), et d'un faible couplage (b).

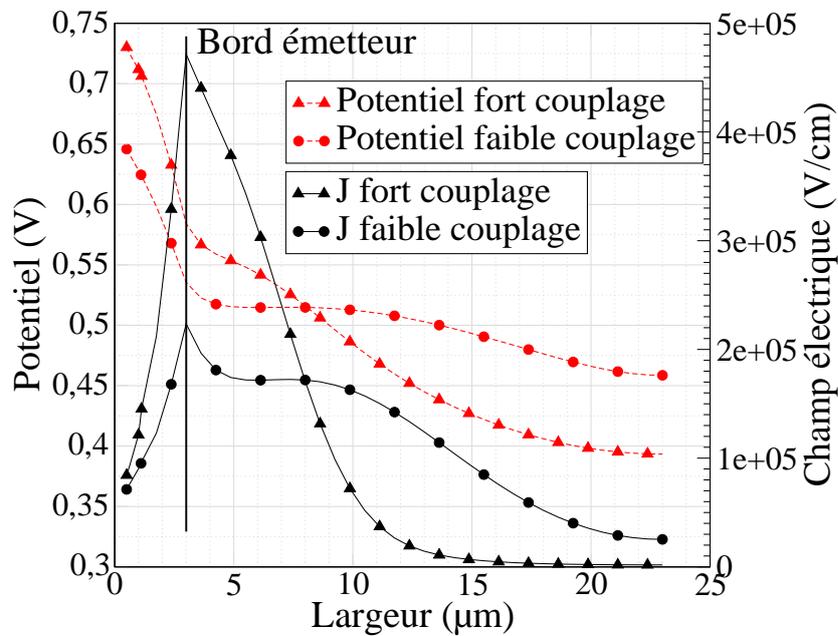


FIG. 3.13 – Coupe de la tension et de la densité de courant dans la base sous l'émetteur dans la figure 3.12. Dans le cas d'un fort et d'un faible couplage entre émetteur et la région de génération par avalanche.

L'effet positif de l'élargissement de la base effective sur la distribution du courant vient d'être démontré. En revanche, il entraîne une importante augmentation de la tension de repliement V_{i1} , qui passe de 65 V à 200 V au regard des résultats de simulation sur la structure simplifiée (Fig. 3.14). Cette augmentation découle de l'approfondissement de la région faiblement dopée de collecteur (W_i) qui, comme nous l'avons montré à la section 3.3.1, influe directement sur la tension maximale avant le repliement. Cet aspect ne limitera pas la robustesse intrinsèque du transistor car il correspond à un régime de relativement faible densité de courant où le courant est pratiquement uniforme dans le composant. En revanche, la diode de déclenchement (jonction

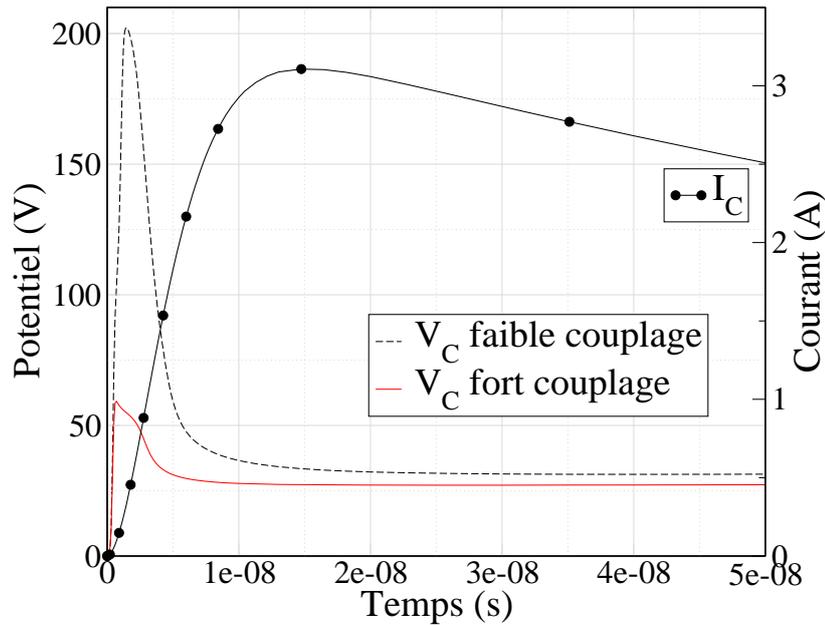


FIG. 3.14 – Variation de la tension de collecteur en fonction du temps, simulée pour un stress HBM de 2kV, dans le cas d'un fort et d'un faible couplage entre l'émetteur et la région d'avant-lanche.

P^{++}/N^{-} dans la coupe selon la largeur de l'émetteur Fig. 3.1) pourrait être dégradée. De plus, la présence de la surtension limite les possibilités d'utilisation du transistor. Il faudra donc envisager des techniques pour la réduire.

3.5 Influence du profil de dopage de collecteur sur la tension de maintien

La tension de maintien du composant est définie par les caractéristiques de la jonction effective de collecteur. Le caractère graduel du profil de dopage de la couche enterrée va donc influencer sur son comportement.

Dans la figure 3.15, les résultats de simulation obtenus pour une structure possédant un profil de diffusion de collecteur abrupt sont comparés à ceux de la structure précédente. La tension de maintien décroît de 10 V et le mouvement de la zone de forte densité de courant sous l'effet de la température n'est pas affecté.

L'importante diminution de la tension de maintien tient dans la modification des caractéristiques de la charge d'espace dans la région de fort champ électrique du collecteur. Les figures 3.16 et 3.17 donnent les profils du champ électrique, des porteurs libres et du dopage dans cette région, dans le cas d'un profil de collecteur abrupt et d'un profil graduel, respectivement.

Pour un profil de dopage abrupt (Fig. 3.16), la charge d'espace de signe négatif (à gauche de la jonction effective) provient principalement de la densité des électrons. La charge positive (à droite de la jonction) provient des impuretés dopantes ionisées (type N). Les propriétés électriques de cette jonction sont donc proches de celle d'une jonction abrupte qui a été traité au chapitre précédent dans le cadre du modèle semi-analytique des TBA.

Pour un profil de dopage graduel (Fig. 3.17) les charges provenant des dopants et des électrons libres sont du même ordre de grandeur. La morphologie de la jonction obtenue à fort courant peut toutefois être assimilée à celle d'une jonction linéaire. Si l'on néglige la densité de charge des trous et que l'on considère simplement celles des électrons et des donneurs ionisés, on peut

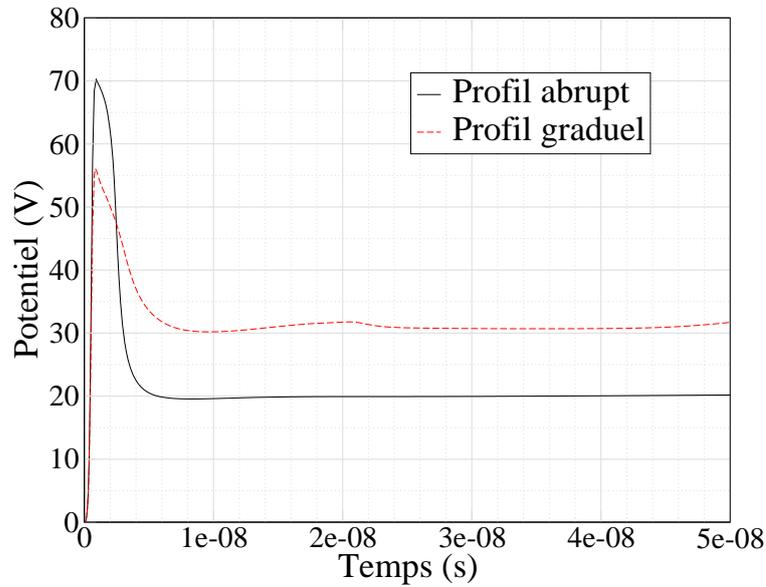


FIG. 3.15 – Variation de la tension de collecteur en fonction du temps, simulée pour un stress HBM de 2kV, dans le cas de profils graduel et abrupt.

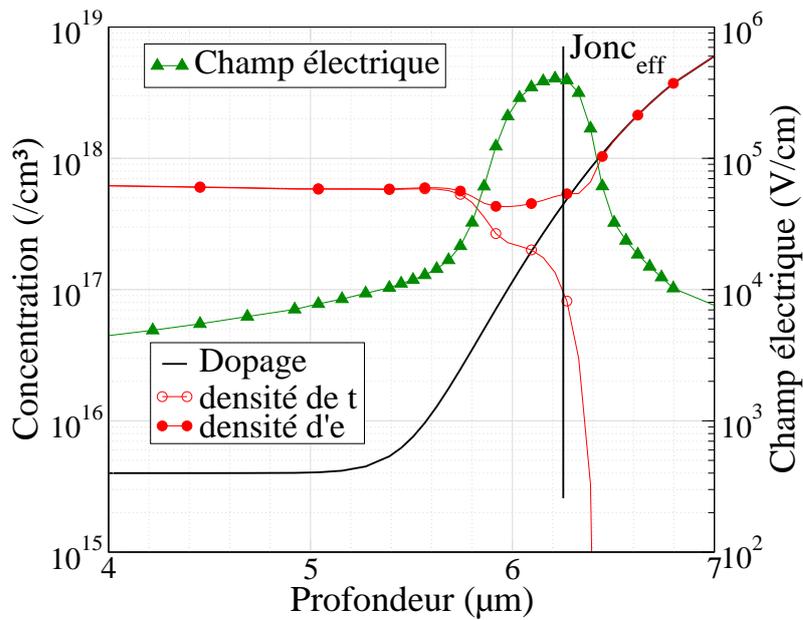


FIG. 3.16 – Coupe du champ électrique et profils des porteurs dans la région de forte densité de courant du collecteur, pour un profil de dopage abrupt. La jonction collecteur-base effective est représentée ($Jonc_{eff}$).

approcher le profil de charge d'espace autour de la jonction délocalisée par un profil linéaire (Fig. 3.18). En effet, si autour de la jonction délocalisée, on approxime le profil des électrons et du dopage par des profils linéaires, la soustraction de ces profils donnent le profil de charge d'espace qui est donc lui aussi linéaire. La pente de ce dernier est sensiblement égale à la dérivée du profil de dopage car la pente du profil des électrons est beaucoup plus faible. La tension de claquage d'une jonction linéaire est donnée par :

$$BV_{lin} = 9,2 \cdot 10^9 a^{-\frac{2}{5}} \quad (3.2)$$

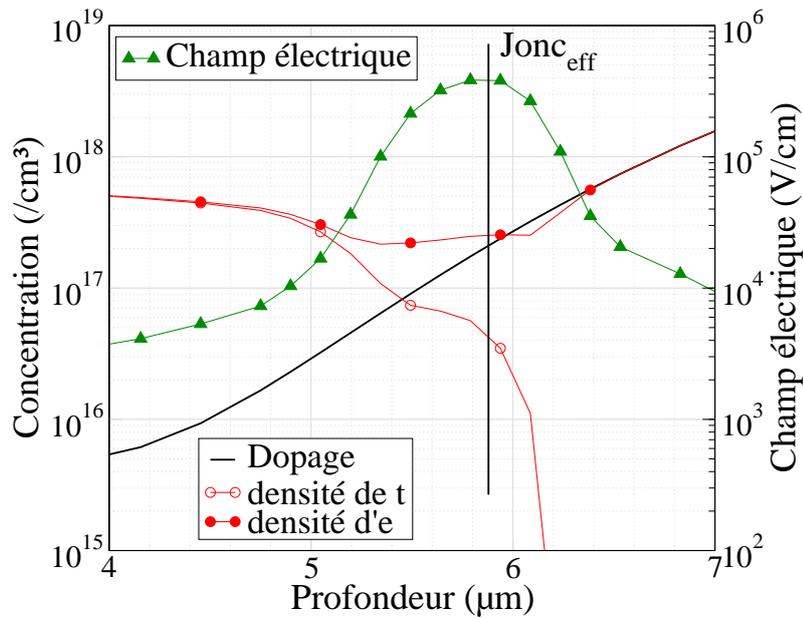


FIG. 3.17 – Coupe du champ électrique et profils des porteurs dans la région de forte densité de courant du collecteur, pour un profil de dopage graduel. La jonction collecteur-base effective est représentée ($Jonc_{eff}$).

avec a la valeur de la dérivée du profil de porteurs. Cette expression montre que plus le profil est graduel (valeur de a faible), plus la tension de claquage est grande.

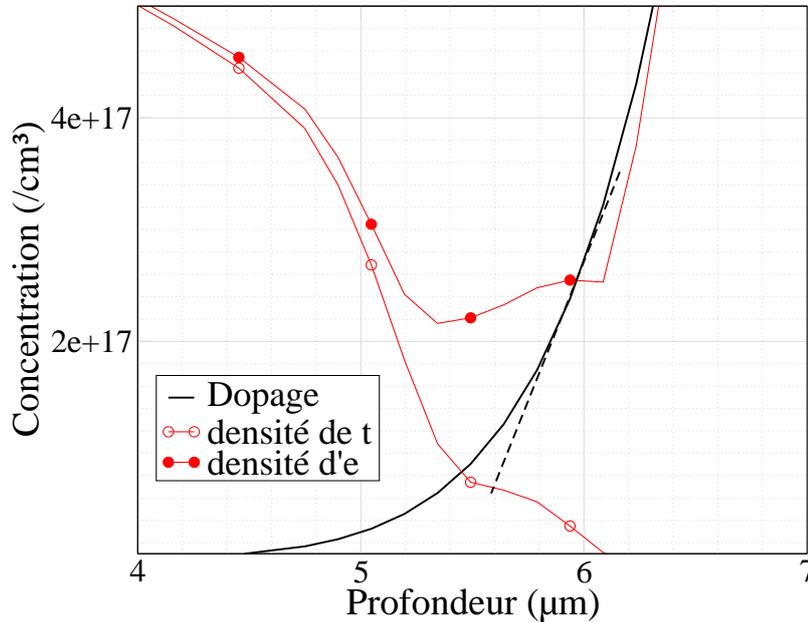


FIG. 3.18 – Approximation d'une jonction délocalisée graduelle par une jonction linéaire. Profils des porteurs dans la région de forte densité de courant du collecteur, pour un profil de dopage graduel (identique à la figure 3.17 excepté l'échelle linéaire pour la concentration).

L'utilisation astucieuse des différentes gradualités des profils de dopage disponibles dans une technologie permet ainsi d'obtenir une gamme variée de tensions de maintien. Cela a été mis

en évidence par De Heyn et ses coauteurs [119]. Nous apportons ici une explication claire et approfondie à leurs observations.

3.6 Bilan : Éléments d'optimisation pour la conception

Le fonctionnement des TBA lorsqu'ils conduisent une décharge électrostatique apparaît particulièrement singulier comparé à celui des transistors bipolaires classiques. L'étude menée a permis de mettre en évidence les phénomènes physiques clés.

Élargissement de la région de base

D'un point de vue électrique, la région de collecteur est à l'origine d'un important effet focalisant sur le courant, principalement limité par la résistance de collecteur. Inversement, la jonction émetteur-base possède un effet défocalisant grâce à la diminution de son efficacité d'injection avec la densité de courant (diminution du gain du bipolaire). D'un point de vue thermique, une élévation de température locale dans la jonction émetteur-base favorise l'augmentation de la densité de courant dans cette région. À l'inverse, grâce à l'effet de la diminution du phénomène d'avalanche avec la température, la région de collecteur possède la propriété de se s'auto-stabiliser thermiquement. La décroissance du phénomène d'avalanche apparaît comme particulièrement bénéfique. Elle permet d'éviter l'emballement thermique dans les points chauds en régulant la densité de courant des régions les plus chaudes vers les plus froides.

L'élargissement de la région de base apparaît comme une excellente solution pour stabiliser à la fois électriquement et thermiquement le composant et retarder l'initialisation locale d'un second claquage thermique à l'origine de la défaillance. L'éloignement spatial de la région chaude du collecteur et de la jonction émetteur-base supprime les risques de focalisation d'origine thermique et permet de profiter de l'effet défocalisant de la jonction E-B. L'auto-stabilisation thermique de la région de collecteur, qui est le siège d'une importante dissipation de chaleur, se voit par la même occasion favorisée. Les résultats expérimentaux présentés dans la suite confirment une amélioration importante de la robustesse avec l'élargissement de la région de base.

L'obtention d'un régime de fonctionnement avec une importante profondeur de région de base passe par l'utilisation de transistors bipolaires NPN possédant un collecteur faiblement dopé. L'élargissement de la région de base à fort courant n'entraîne qu'une faible augmentation de la tension de maintien. En revanche, la tension de repliement est fortement accrue et limite les applications possibles de ces dispositifs. Le développement de techniques permettant de diminuer cette tension est nécessaire. Plusieurs solutions ont été vérifiées expérimentalement et sont présentées dans la suite.

Choix du profil de dopage de collecteur

L'emploi d'un profil de dopage abrupt permet de diminuer la tension de maintien. La puissance totale dissipée dans la structure étant d'autant plus faible que la tension de maintien est basse, il est préférable d'utiliser les profils de dopages les plus abrupts pour obtenir une robustesse importante. Il faut veiller cependant à choisir une tension compatible avec la tension d'alimentation du circuit à protéger, pour respecter les marges de conception.

Réduction des zones de plus fortes densités de courant et de température

Afin de garantir de bonnes performances, il faut éliminer l'existence de régions de plus fortes densités de courant ou de températures.

La base du transistor devra être choisie aussi dopée que possible afin de limiter l'effet de focalisation de sa résistance intrinsèque. Des résistances de ballast doivent être introduites à la fois dans le collecteur et l'émetteur, d'une part pour éviter l'apparition d'une plus forte densité de courant en bordure de diffusion d'émetteur et d'autre part, pour limiter la densité maximale

de courant due à la réduction de la tension de claquage de la jonction collecteur-base à forte densité de courant.

3.7 Résultats expérimentaux

Les éléments d'optimisation énoncés précédemment ont été appliqués et vérifiés dans une première technologie, puis transposés dans deux autres. L'un des principaux objectifs est de réaliser des composants extrêmement robustes et de taille réduite [120]. Plusieurs variations de géométries ont été réalisées et vont permettre d'illustrer et de confirmer les comportements déduits des simulations. L'étude de plusieurs solutions de déclenchement sera menée pour réduire la surtension avant le repliement. Enfin, les résultats d'analyse de défaillance par la technique de stimulation photoélectrique laser permettront de confirmer et d'étudier le mécanisme de défaillance des composants réalisés.

3.7.1 Description des composants réalisés

L'étude expérimentale a été menée dans trois technologies aux caractéristiques sensiblement différentes.

Technologie 1

La première, qui sera désignée par la suite par technologie 1, est une technologie dite «CMOS analogique» 1,2 μm , sans siliciures ni drains faiblement dopés (LDD). Cette technologie utilise un substrat P. Elle comprend entre 12 et 20 niveaux de masque qui permettent de réaliser une large variété de composants destinée à des applications de basse ou haute tension (6 V ou 12 V) :

- composant actifs :
 - transistor MOS (canal N et P) basse ou haute tension
 - transistor LDMOS
 - transistors bipolaires (NPN et PNP)
- composant passifs :
 - résistances en polysilicium et diffusées
 - polysilicium hautement résistif
 - capacités

Ainsi, le nombre de masques utilisés varie en fonction des options nécessaires à la réalisation des applications (régulateur de tension de type LDO «Low Drop Out», référence de tension de type bandgap, amplificateur opérationnel)

Technologie 2

La seconde technologie utilisée est une technologie CMOS 1 μm , sans siliciures ni LDD. Cette technologie sur substrat P est dédiée aux applications de conversion de l'énergie et intègre des composants LDMOS hautes tensions jusqu'à 700 V :

- composant actifs :
 - transistor MOS (canal N et P) basse tension (9 V)
 - transistor LDMOS moyenne et haute tensions (20 V, 45 V, 200 V, 400 V, 700 V)
 - transistors bipolaires (NPN et PNP)
- composant passifs :
 - résistances en polysilicium et diffusées

Les applications, principalement basées sur les techniques de découpage, vont du convertisseur haute tension alternative (85 Vac, 256 Vac) vers basse tension continue (5 Vdc, 48 Vdc), à la conversion continu-continu basse tension.

Technologie 3

La technologie 3 est une technologie CMOS NWell $0,6 \mu\text{m}$ possédant des LDD mais pas de siliciures. Basée sur un substrat P, cette technologie très simple permet la réalisation de circuits aussi bien analogiques que digitaux avec des tensions d'alimentation de 3 V à 5 V.

Soulignons enfin que pour toutes ces technologies, l'absence de couche enterrée ne permet pas la réalisation de transistors bipolaires verticaux.

3.7.2 Coupes technologiques

La profondeur effective de base, définie précédemment, est figée pour un transistor bipolaire vertical. L'impact de son accroissement sur la robustesse aux décharges électrostatique ne pourrait être étudié qu'en augmentant l'épaisseur de la couche épitaxiée, ce qui n'est pas réalisable dans une technologie figée. La structure utilisée dans ces travaux permet de contourner ce problème tout en reposant sur les mêmes principes de fonctionnement.

Technologie 1

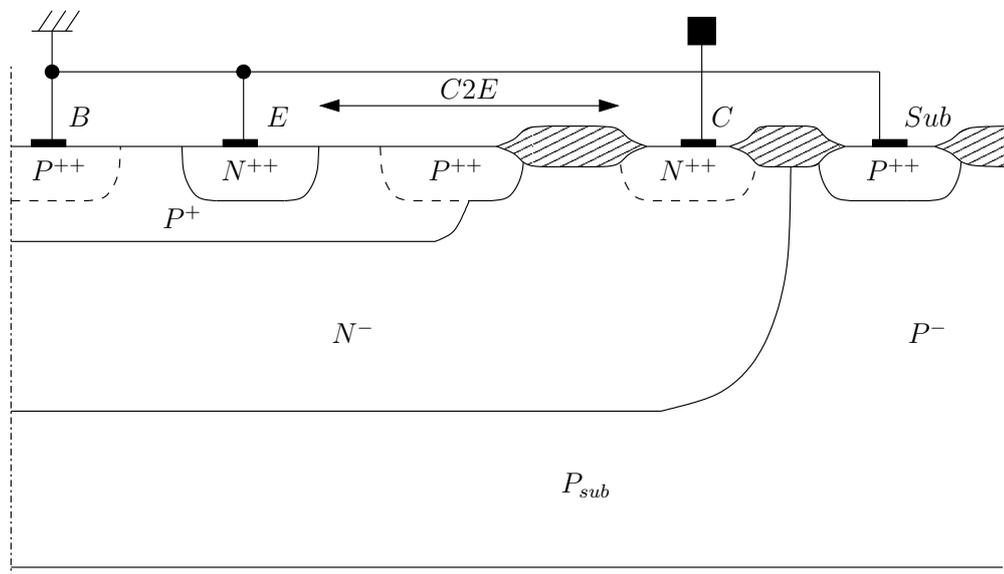


FIG. 3.19 – Coupe technologique du transistor NPN (technologie 1).

La coupe technologique du transistor bipolaire NPN étudié dans la technologie 1 est représentée sur la figure 3.19. La structure est symétrique par rapport à son contact central de base et seule une moitié du composant est donc dessinée. La base du transistor est constituée d'une couche P^+ assez fortement dopée. Cette diffusion est aussi utilisée pour la réalisation des sources et drains des PMOS hautes tensions de la technologie. Le collecteur est constitué d'une diffusion N^- dans laquelle une diffusion N^{++} permet de prendre un contact.

Ce type de transistor bipolaire, dépourvu de couche enterrée, présente cependant un mode de fonctionnement identique au transistor vertical étudié auparavant. Sous l'effet des fortes densités de courant, le maximum de champ électrique, initialement situé à la jonction métallurgique collecteur-base P^+/N^- , est repoussé à la frontière des couches N^- et N^{++} du collecteur. Le caractère très abrupt du profil de dopage à cette frontière implique une tension de maintien relativement faible de 8 V.

Les résultats de simulation électrique bidimensionnelle permettent de confirmer et d'étudier plus en détail leur mode de fonctionnement. À fort courant, l'important taux de génération

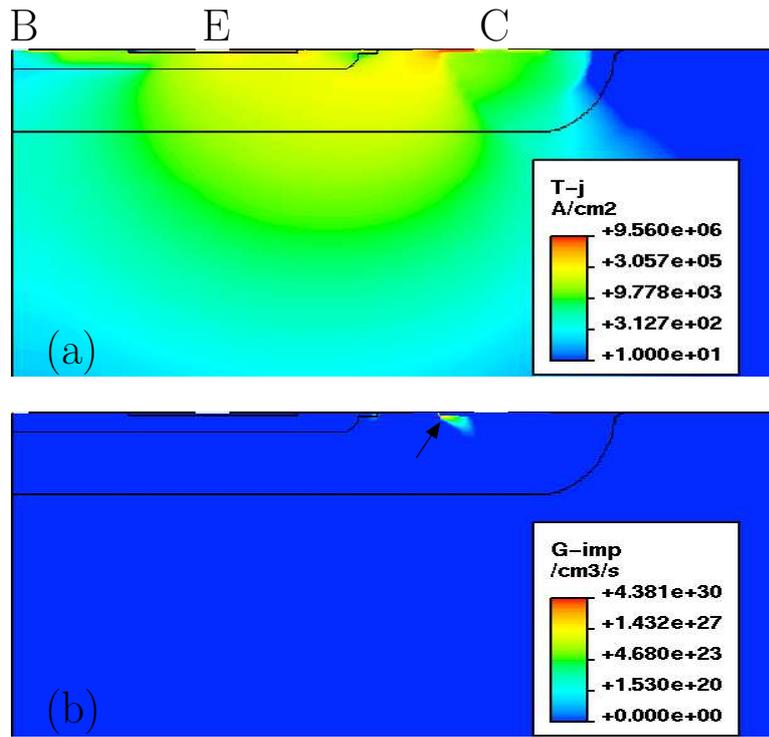


FIG. 3.20 – Densité de courant (a) et taux de génération par avalanche (b) dans le transistor NPN (technologie 1) après son repliement, au pic de courant d’une décharge HBM de 2kV.

par avalanche à la frontière N^-/N^{++} témoigne de la formation d’une jonction collecteur-base effective (Fig. 3.20(b)). Le courant injecté par l’émetteur circule assez profondément dans le composant avant de remonter à la jonction C-B effective (Fig. 3.20(a)). Dans la base effective du transistor bipolaire, située entre la jonction E-B et la jonction effective C-B, le chemin suivi par le courant correspond à la formation d’un transistor de plus fort gain possible. Rappelons que le gain d’un transistor dépend de la profondeur de la base et de son dopage. Le gain est d’autant plus grand que la profondeur et le dopage de la base sont faibles. Le courant emprunte donc un chemin qui réalise le meilleur compromis entre une courte distance et un dopage réduit tout au long du chemin. La surface du composant étant plus dopée (présence de la diffusion P^+), il est plus favorable que le courant emprunte un chemin «en profondeur» correspondant à des régions plus faiblement dopées, au détriment d’une distance plus importante.

La profondeur effective de base du transistor bipolaire ainsi formé à fort courant, peut être ajustée en fonction du dessin de la structure en modifiant la distance C2E (Fig. 3.19) entre les diffusions N^{++} d’émetteur et de prise de contact du collecteur.

Technologie 2

Le même type de transistor bipolaire NPN a été étudié dans la technologie 2. Sa coupe technologique est représentée sur la figure 3.21. On notera simplement l’ajout d’une prise de contact P^{++} , située à droite de l’émetteur, dans la diffusion P de base. Cette connexion permettra l’ajout d’un élément de déclenchement externe, telle la diode représentée sur la figure. En outre, les dopages de la diffusion de base (P) et de collecteur (N^-) sont inférieurs à ceux de la technologie 1. Ces dopages plus faibles sont nécessaires pour tenir les hautes tensions dans les applications visées par cette technologie.

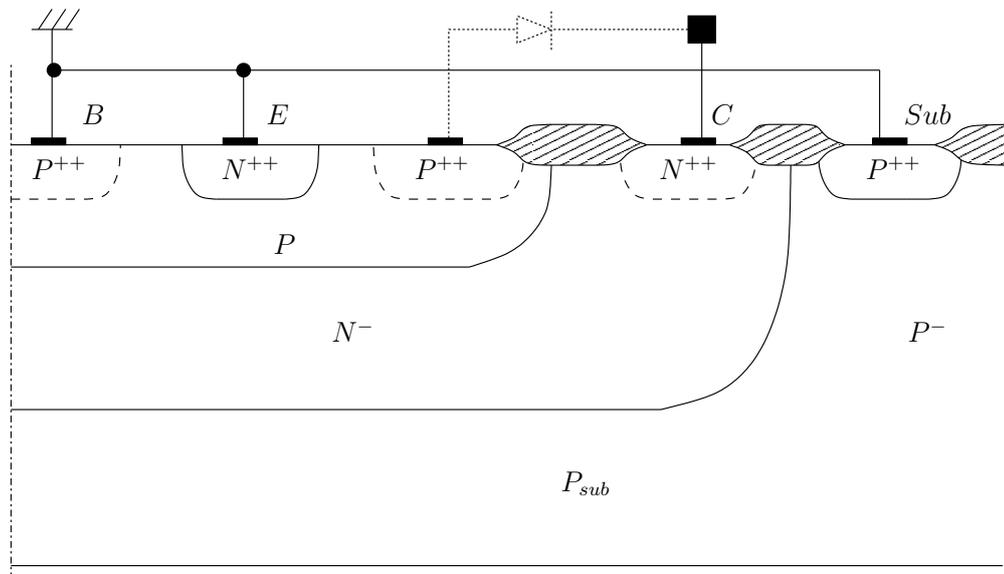


FIG. 3.21 – Coupe technologique du transistor NPN (technologie 2).

Technologie 3

La coupe technologique du transistor bipolaire NPN réalisé dans la technologie 3 est représentée sur la figure 3.22. Les diffusions disponibles dans cette technologie ne permettent pas la construction d'un transistor totalement identique aux précédents. Aucune diffusion P n'est disponible pour réaliser la base du transistor bipolaire. Toutefois, le dopage en surface du substrat P permettant la réalisation des transistors NMOS est relativement important dans cette technologie destinée aux basses tensions. Ainsi, le dopage P sous l'émetteur est relativement important et permet de limiter la focalisation due à la résistance de base. En outre, la concentration de dopant diminue rapidement de la surface vers le substrat P_{sub} . Aux régimes de forte densité de courant, ceci favorise le passage du courant dans la profondeur du composant, comme en témoigne la simulation de la figure 3.23.

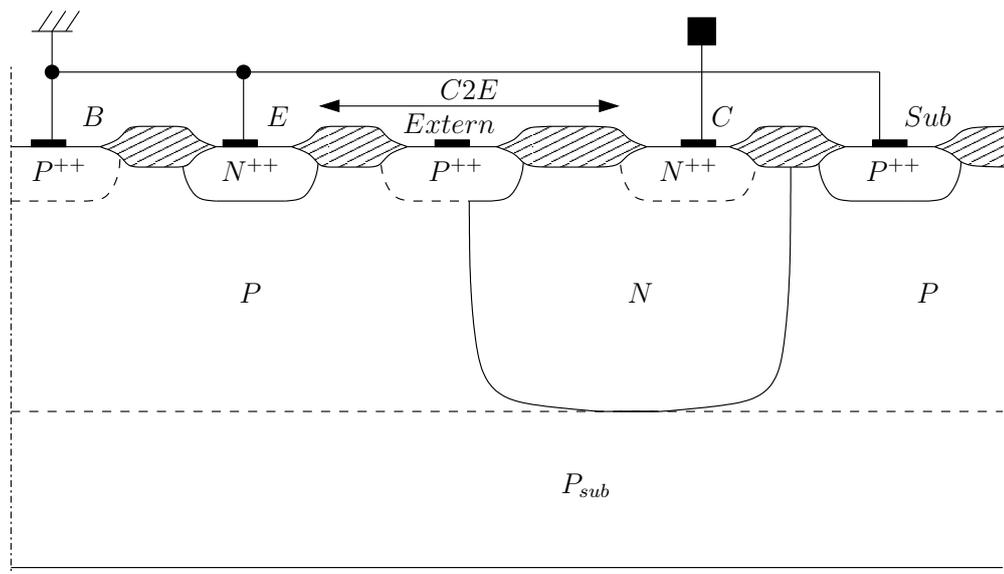


FIG. 3.22 – Coupe technologique du transistor NPN (technologie 3).

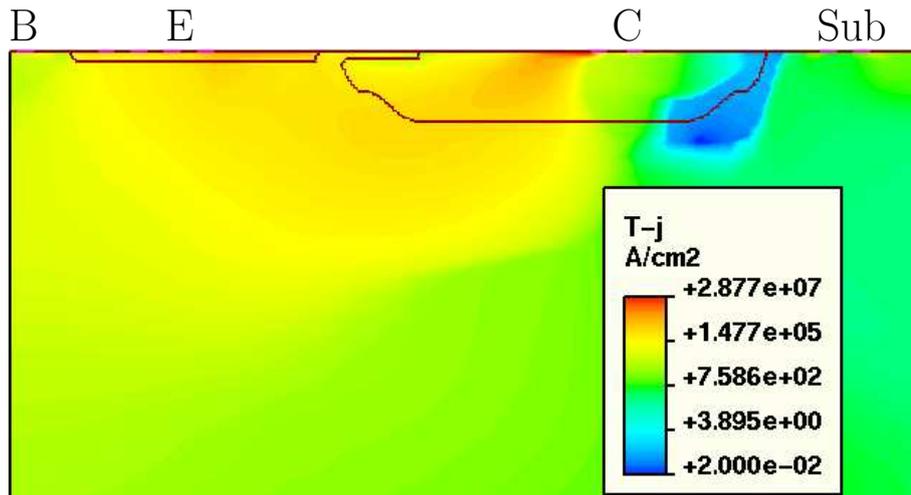


FIG. 3.23 – Densités de courant dans la TBA (technologie 3) au pic de courant d'une décharge HBM de 2kV.

3.7.3 Critère de défaillance

Pour déterminer la robustesse d'un composant aux décharges électrostatiques, la méthode classique consiste à lui faire subir un stress d'amplitude connue au moyen de testeurs respectant les normes HBM, CDM, etc. . . , puis à vérifier son intégrité physique et fonctionnelle.

Le choix du critère de défaillance du composant doit être l'objet d'une attention particulière. Dans la majorité des cas, ce critère repose sur le suivi de l'évolution de caractéristiques électriques avant et après un stress. Deux approches peuvent alors être retenues pour définir le niveau de robustesse d'un dispositif aux ESD. L'une, purement fonctionnelle, qui garantit qu'après avoir subi un stress d'intensité donnée, le composant continue à respecter les spécifications pour lesquelles il a été conçu. L'autre, plus stricte, qui garantit qu'aucune de ses caractéristiques ne sera modifiée après le stress. Cette dernière approche n'est pas toujours envisageable car elle impose la mise en œuvre d'appareillages garantissant une grande précision de mesure. Pour illustrer la différence entre ces deux critères de défaillance, imaginons le cas d'un circuit inverseur CMOS dont le courant de consommation au repos est spécifié comme inférieur à 1 μA sous 10 V à 25°C. Avant tout stress, le courant mesuré est de 0,1 μA . Après un stress HBM de 5kV, ce courant passe à 0,4 μA . Le critère de non-évolution des caractéristiques du composant n'est pas respecté car le courant au repos est multiplié par quatre. Le critère de fonctionnalité est en revanche parfaitement respecté. Des mesures supplémentaires pourraient faire apparaître que le critère de non-évolution n'est respecté que pour des stress HBM inférieurs à 2 kV. . .

Pour traiter efficacement ce problème, nous pouvons distinguer deux types de test : d'une part, le test d'application réalisé sur des circuits complets, et d'autre part, le test de composant de protection seul. Si le test de fonctionnalité est bien adapté au test d'application, en particulier pour sa plus grande facilité de mise en œuvre, il est déconseillé pour la caractérisation des structures de protections. En effet, divers travaux ont démontré que l'évolution des caractéristiques électriques, en particulier l'augmentation du courant de fuite, correspond à une «défiabilisation» du composant [121, 122]. Il est donc important de fournir des structures de protection exemptes de ce type de comportement, afin d'assurer une excellente fiabilité aux produits qui l'utiliseront. De plus, il est aisé de détecter une évolution du courant de fuite pour une structure de protection seule, alors que cette évolution peut être masquée dans le courant de fuite global d'un circuit complet. Fournir des composants de protection ESD extrêmement fiables permet ainsi de limiter les risques de défiabilisation des applications, et de légitimer l'utilisation d'un critère de fonctionnalité pour mesurer la robustesse d'un produit.

Pour l'ensemble des composants présentés par la suite, les résultats de robustesse aux dé-

charges électrostatiques sont donnés en utilisant le critère de non-modification des caractéristiques statiques quel que soit le type de stress appliqué au composant. Pour un transistor bipolaire, la caractéristique statique utilisée provient de la mesure du courant de fuite de sa jonction collecteur-base polarisée en inverse jusqu'à sa tension de claquage, en limitant le courant à 1 ou 10 μA maximum. Une méthode commode pour définir mathématiquement le critère de défaillance, consiste à calculer l'intégrale de la courbe représentant le logarithme du courant de fuite en fonction de la tension. Toute augmentation de cette aire traduit une augmentation du courant de fuite et constitue le critère de défaillance. L'utilisation du logarithme permet d'équilibrer, dans la caractéristique statique, l'importance relative des faibles niveaux de courant de fuite face aux plus forts niveaux associés au claquage du composant.

3.7.4 Influence de la distance collecteur-émetteur

En technologie 1

Les composants ont d'abord été développés dans la technologie 1. L'obtention de structures optimisées lors d'un premier lot de test a permis de réaliser un second lot afin de mettre en évidence les aspects et paramètres critiques pour la conception.

Le paramètre le plus critique, mis en évidence aux travers des simulations, est la profondeur effective de base dans le régime de fort courant. Dans la structure choisie (Fig. 3.19), ce paramètre est facilement ajustable au travers de la distance collecteur-émetteur (C2E).

Les transistors possédant une distance C2E optimale supportent des décharges HBM de forte intensité, d'au moins 12 kV (Tab. 3.2). Lorsque cette distance est réduite de 0,75 μm leur robustesse est diminuée et varie dans une plus large gamme comprise entre 9 et 12 kV. De la même manière, pour une réduction supplémentaire de 0,75 μm , les niveaux de stress HBM entraînant une modification de la caractéristique statique sont compris entre 4 et 7 kV.

C2E	optimal	opt - 0,75 μm	opt - 1,5 μm
HBM (kV)	12–13	9–12	4–7

TAB. 3.2 – Performance HBM en fonction de l'éloignement (C2E) des diffusions fortement dopées du collecteur et de l'émetteur (technologie 1).

Ce comportement confirme le rôle crucial de ce paramètre sur la robustesse des composants et souligne l'importance d'assurer un découplage efficace, à la fois thermique et électrique, des phénomènes d'injection (à la jonction E-B) et de multiplication par avalanche (à la jonction C-B effective).

L'accroissement de la dispersion des résultats avec la diminution de la distance C2E, témoigne du caractère aléatoire de l'occurrence d'une défaillance. Comme cela a pu être supposé à partir des résultats de simulations, l'initialisation d'un second claquage thermique apparaît d'autant moins probable que la profondeur de base effective est importante. Accroître la profondeur effective de base permet donc d'augmenter la robustesse en favorisant les effets qui participent à la stabilisation thermique du composant.

Les caractéristiques TLP de ces composants pour des courants relativement faibles, autour de la région de repliement de la caractéristique, sont superposées sur la figure 3.24.

La tension de claquage collecteur-base mesurée en statique à 16 V correspond à la tension de claquage de la diode P^{++}/N^{-} . Cette diode permet de déclencher le transistor bipolaire et conduit une partie du courant total dans la partie de la caractéristique où le courant est faible (< 150 mA) et où la tension s'élève rapidement. L'élévation de la tension se produit jusqu'à ce que le transistor bipolaire entre dans le régime de diode PIN précédemment décrit. L'entrée dans ce régime de fonctionnement, qui correspond à l'arrêt de l'accroissement de la tension de claquage de la jonction collecteur-base, est particulièrement visible sur la caractéristique du

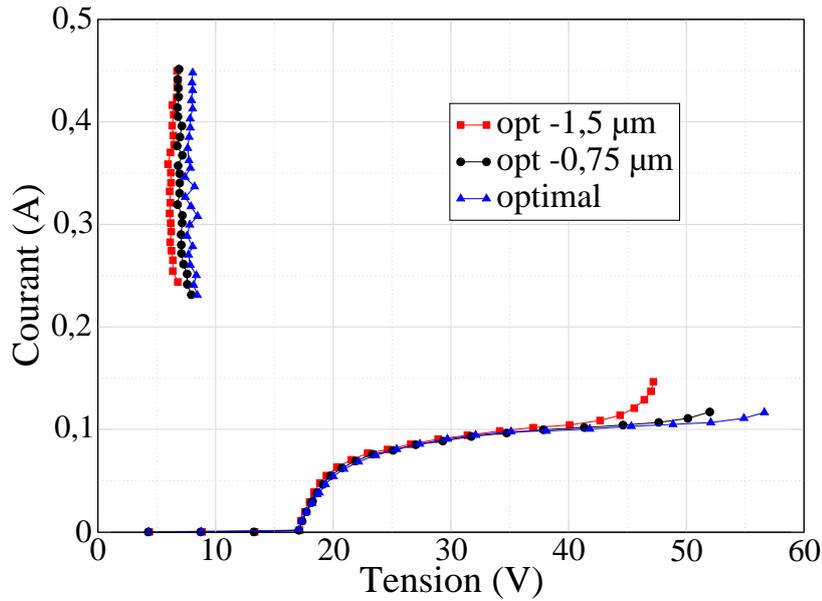


FIG. 3.24 – Influence de la distance collecteur-émetteur sur la caractéristique TLP bas courant du transistor NPN (technologie 1).

composant possédant une distance C2E réduite de $1,5 \mu\text{m}$ par rapport à la valeur optimale. La dépendance de la tension maximale avec la profondeur de la zone intrinsèque de la diode PIN, c'est-à-dire la profondeur de la région N^- de collecteur et donc la distance C2E, est nettement visible. La réduction de cette profondeur correspond bien avec une diminution de la tension de repliement (V_{t1}).

Enfin, la tension de maintien (V_H) de l'ordre de 7 V, varie peu avec la distance C2E. L'augmentation de la profondeur de base (effective) de $1,5 \mu\text{m}$ correspond à un accroissement de tension de maintien inférieur à 2 V.

En technologie 3

Grâce à l'expérience acquise dans la technologie 1, le transfert du savoir-faire dans cette technologie très différente a permis d'obtenir, dès le premier lot de test, des composants robustes à des stress de 5 kV HBM. Il convient d'insister ici, sur l'importante différence qui existe entre les technologies 1 et 3, ainsi que sur la nécessité d'adopter une structure de composant sensiblement différente. Les composants optimisés obtenus dans un second lot, présentent d'excellentes performances, équivalentes à celles qui sont obtenues dans la technologie 1.

L'influence de la distance C2E sur la robustesse HBM, résumée dans le tableau 3.3, est semblable à celle obtenue pour les composants de la technologie 1. L'augmentation de la robustesse

C2E	optimal	opt - $0,7 \mu\text{m}$	opt - $1,4 \mu\text{m}$
HBM (kV)	12–13	8–12	5–10

TAB. 3.3 – Performances HBM (minimum-maximum) en fonction de l'éloignement (C2E) des diffusions fortement dopées du collecteur et de l'émetteur (technologie 3).

avec l'élargissement de la base effective et le caractère aléatoire de l'apparition d'un second claquage thermique sont également observés. Ces résultats s'ajoutent au précédent pour confirmer la validité du modèle de fonctionnement construit à partir de la simulation numérique.

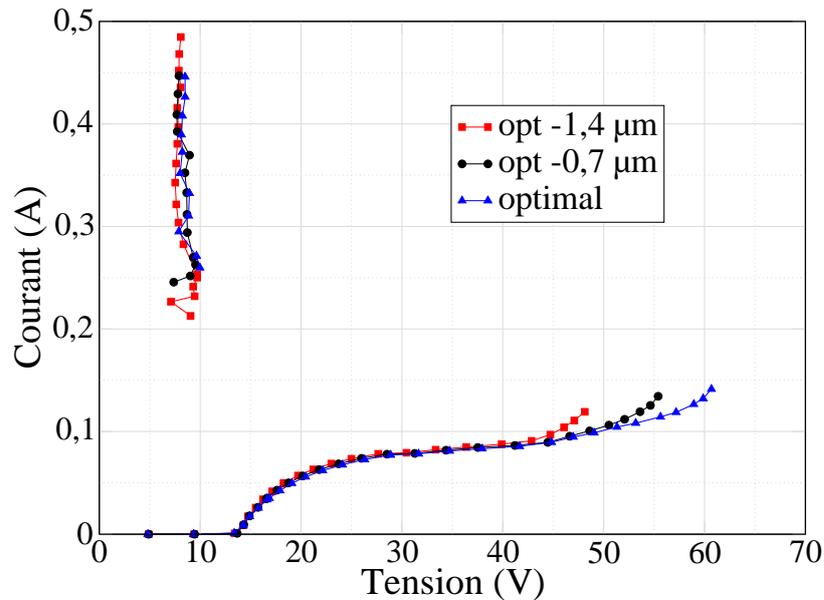


FIG. 3.25 – Influence de la distance collecteur-émetteur sur la caractéristique TLP bas courant du transistor NPN (technologie 3).

Les caractéristiques TLP à bas courant, données dans la figure 3.25, présentent les mêmes aspects que celles des composants issus de la technologie 1. La tension de maintien est peu dépendante de la distance collecteur-émetteur. Cette dépendance est ici masquée par les incertitudes de mesures du banc de test TLP. La précision sur la tension de maintien est en particulier limitée par l'utilisation d'un calibre de tension élevé sur l'oscilloscope, due à la présence de la surtension avant le repliement.

La principale différence entre les caractéristiques TLP des technologies 1 et 3 réside dans la forme de la courbe lors de l'augmentation de la tension avant le repliement. En effet, cette partie de la caractéristique dépend de la diode collecteur-base. Cette diode est du type $P^+|N^-|N^{++}$ dans la technologie 1 et du type $P|N|N^{++}$ dans la technologie 3. Toutefois, l'augmentation de la tension de repliement avec l'agrandissement de la distance C2E apparaît clairement.

3.7.5 Influence de la géométrie

L'impact des variations de géométrie a été étudié dans la technologie 1. Deux aspects vont être présentés :

- L'évolution de la robustesse HBM et des caractéristiques TLP avec la taille totale de la structure, au travers de la longueur de son émetteur
- L'effet de quelques variations dans la géométrie des extrémités des diffusions de collecteur et d'émetteur.

Variation de longueur d'émetteur

La robustesse des composants à un stress HBM est reportée dans le tableau 3.4, pour trois longueurs d'émetteur.

Il existe une bonne proportionnalité de la robustesse au stress HBM avec la taille du composant. Le composant de référence, identique à celui de la section précédente, possède une longueur d'émetteur de $34 \mu\text{m}$. La robustesse maximum des composants dont la longueur d'émetteur est de $45 \mu\text{m}$ n'a pas pu être déterminée car elle dépasse les possibilités du testeur utilisé qui peut

L_E (μm)	23	34	45
HBM (kV)	8–9	12–13	>16

TAB. 3.4 – Performance HBM (minimum-maximum) en fonction de la longueur L_E d'émetteur du transistor bipolaire NPN (technologie 1).

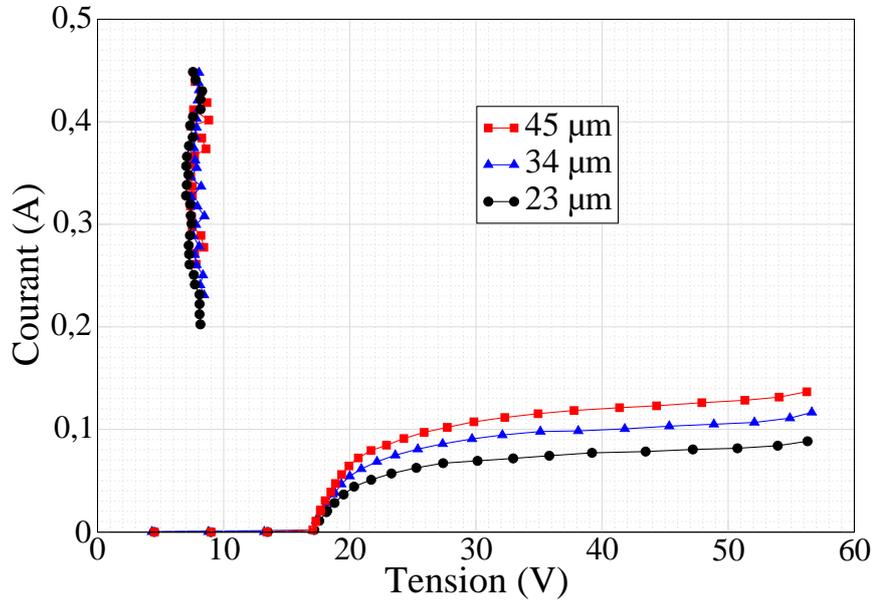


FIG. 3.26 – Influence de la longueur d'émetteur du composant sur la caractéristique TLP bas courant du transistor NPN (technologie 1).

fournir un stress maximal de 16 kV HBM.

Les caractéristiques TLP de ces composants, représentées sur la figure 3.26, donnent des informations sur l'uniformité du courant au sein des composants dans les différents régimes de fonctionnement. Le niveau du courant de repliement (I_{t1}) est proportionnel à la taille du dispositif. Ce comportement confirme l'uniformité du courant dans le transistor pour le régime de moyen courant. Rappelons que le repliement du composant correspond à la transition du maximum de champ électrique de la jonction métallurgique à la jonction effective située à la transition N/N^{++} du dopage de collecteur. Nous avons montré que le courant est relativement uniforme avant le repliement et que la transition du maximum de champ électrique se produit pour une densité de courant supérieure à la densité de courant critique. Le courant total nécessaire au repliement est donc proportionnel à la taille du composant. En revanche, la tension de repliement, qui est proportionnelle à la profondeur de la zone faiblement dopée de collecteur, est indépendante de la taille du composant et égale à 57 V.

En outre, la tension de maintien est identique quelle que soit l'aire du transistor. Ceci témoigne de la non-uniformité du courant après le repliement. Le degré de focalisation du courant est en effet indépendant de la taille du composant.

Influence des extrémités de diffusion

L'étude menée par simulation montre que les extrémités des doigts d'émetteur et de collecteur représentent des régions favorables à la formation de régions plus chaudes. Afin d'étudier cet aspect, plusieurs variations de dessin des masques de diffusion de collecteur et d'émetteur ont été étudiées.

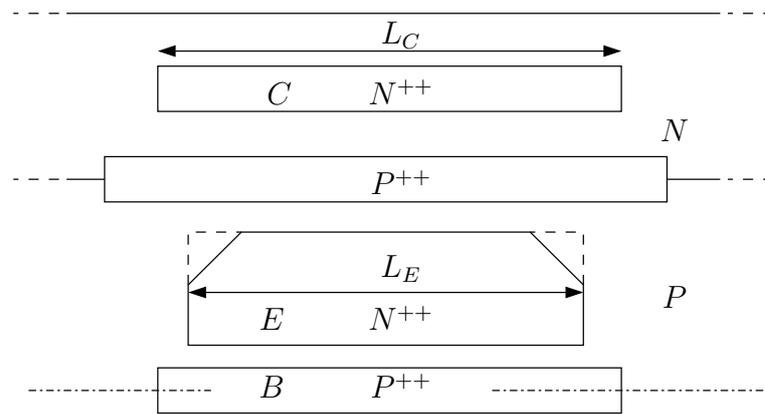


FIG. 3.27 – Vue de dessus schématique, des diffusions du transistor NPN.

La figure 3.27 donne une vue de dessus schématique, des diffusions utilisées pour réaliser le transistor.

Le transistor optimisé possède des extrémités d'émetteur aux coins cassés (Fig. 3.27). L'utilisation d'un émetteur rectangulaire, représenté en pointillés sur la figure, réduit les performances du composant. Dans ce cas, la tenue aux décharges HBM varie dans une plage de 8 à 12 kV. Les extrémités de l'émetteur doivent donc faire l'objet d'attentions particulières pour obtenir de bonnes performances.

Les longueurs de collecteur L_C et d'émetteur L_E sont représentées sur la figure 3.27. La robustesse HBM de composants possédant un collecteur de longueur identique à l'émetteur $L_C = L_E$ et de composants pour lesquels $L_C > L_E$ a été mesurée. Pour des stress HBM séparés dans le temps d'au moins une minute, la robustesse est identique dans les composants où L_C est identique à L_E et où L_C est supérieur à L_E . En revanche, si le stress est répété trois fois avec un intervalle d'une seconde entre chaque stress, la robustesse du composant possédant un collecteur plus long que son émetteur est réduite dans une plage variant de 10 à 12 kV HBM, alors que celle des composants où $L_C = L_E$ reste inchangée (12-13 kV). Il semble que dans ce dernier cas, la thermalisation du transistor n'est pas totale entre deux stress. La subsistance de régions plus chaudes entraîne un déséquilibre dans le composant qui peut précipiter sa dégradation.

Toute interprétation de ces derniers résultats est particulièrement difficile. Le comportement du composant de ce point de vue est difficile à simuler. Le choix d'une solution optimale pour le dessin des extrémités d'émetteur et de collecteur reste encore empirique et ne peut se baser que sur une succession de réalisations et de caractérisations expérimentales reposant sur l'intuition et l'expérience du concepteur. Une solution particulièrement intéressante consiste en l'adoption de géométries à symétrie circulaire qui font disparaître ce problème. Des travaux plus ou moins récents ont montré l'amélioration des performances avec ce type de géométrie [123, 124, 125].

3.7.6 Comparaison des performances avec les composants standards

Afin d'illustrer les excellentes performances obtenues avec les transistors bipolaires optimisés, il est intéressant de comparer leurs performances avec celles de composants de protections classiques.

La performance d'un composant de protection ESD basé sur un TBA, peut être donnée en termes de tension HBM par unité de longueur d'émetteur (longueur cumulée de tous les doigts) ou, en tenue de tension HBM par unité de surface occupée. Pour être équitable, la comparaison doit reposer sur des composants dont les tensions de maintien sont proches. Dans ce cadre, les protections classiques, réalisées à partir des transistors NMOS constituent une excellente référence.

Les performances des transistors bipolaires NPN optimisés et des composants de protection

basés sur les transistors NMOS sont résumées dans les tableaux suivants. Le tableau 3.5 donne la robustesse HBM des composants par unité de longueur d'émetteur dans les trois technologies étudiées. La robustesse HBM par unité de surface occupée par le composant est donnée dans le

Technologie	NMOS	NPN
1 (CMOS 1,2 μm)	10,6	175
2 (CMOS HT 1 μm)	13	200
3 (CMOS 0,6 μm)	6	162

TAB. 3.5 – Performances HBM des composants par unité de longueur d'émetteur ($V/\mu\text{m}$).

tableau 3.6. Ces chiffres sont calculés à partir de mesures réalisées sur des transistors multi-doigts dont la robustesse était d'au moins 5 kV HBM.

Technologie	NMOS	NPN
1 (CMOS 1,2 μm)	0,65	2,4
2 (CMOS HT 1 μm)	0,7	1,5
3 (CMOS 0,6 μm)	0,4	2,5

TAB. 3.6 – Performances HBM des composants par unité de surface utilisée ($V/\mu\text{m}^2$).

Les performances des transistors NPN sont excellentes vis-à-vis de celles des composants NMOS. Ces derniers possèdent cependant de bonnes performances. Ces résultats sont comparables aux performances de composants de protection réalisés au moyen de thyristors, qui restent cependant plus performants, mais sont rarement utilisés par crainte des problèmes de latch-up.

Le transistor bipolaire vertical, qui a servi de base à l'étude approfondie menée à partir des simulations, possède des performances HBM de 75 $V/\mu\text{m}$ et 0,96 $V/\mu\text{m}^2$. Sa tension de maintien plus grande que celle des transistors NPN optimisés, explique en partie ses performances inférieures. La meilleure dissipation thermique, assurée par la position de la région de haute température dans la profondeur du substrat, devrait permettre une meilleure robustesse. Cependant la profondeur de base effective, qui ne peut être modifiée, ne doit pas être optimale dans ce composant.

Il faut modérer ces bons résultats par l'existence de la surtension avant le repliement, qui limite le domaine d'utilisation des transistors NPN optimisés. Des techniques pour supprimer cet inconvénient peuvent cependant être proposées. Elles ont fait l'objet d'une étude préalable qui est présentée dans la section 3.7.7 suivante.

L'impact de certains paramètres technologiques sur les performances peut être mis en évidence à partir de ces données. La diminution de la profondeur de la diffusion N^{++} entraîne une diminution des performances, particulièrement remarquable sur la tenue HBM par unité de longueur d'émetteur. La profondeur de cette diffusion est en effet la plus importante dans la technologie 2 et la plus faible dans la technologie 3. Les transistors NMOS y sont plus sensibles car la circulation du courant dans ces structures est essentiellement latérale et se situe en surface. La chute des performances est beaucoup moins marquée sur les transistors NPN où le courant circule plus en profondeur.

Cette diminution de performance apparaît moins clairement dans les performances HBM par unité de surface. Si l'on compare les chiffres dans les technologies 1 et 3 pour les transistors NMOS, la diminution de performance est partiellement compensée par la réduction des dimensions technologiques. Cette compensation permet de maintenir les performances des transistors NPN par unité de surface. Le cas de la technologie 2 est particulier car la haute tenue en tension des applications visées implique l'utilisation de règles de dessins plus strictes sur l'éloignement des différentes diffusions. Cet aspect entraîne une augmentation de la surface du transistor NPN

réalisé dans cette technologie et pénalise sa performance HBM par unité de surface.

3.7.7 Techniques de déclenchement

Les performances des composants développés sont excellentes en termes de robustesse ESD par rapport à leur surface. La caractéristique TLP complète du composant optimisé en techno-

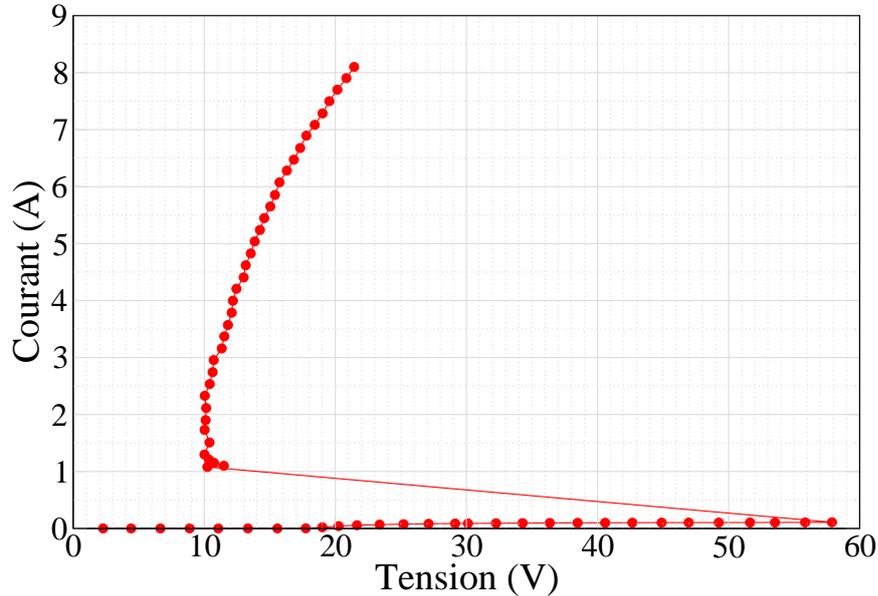


FIG. 3.28 – Caractéristique TLP fort courant du transistor NPN (technologie 1).

logie 1 est représentée dans la figure 3.28. Sa faible tension de maintien et sa faible résistance à l'état passant de l'ordre de $2\ \Omega$ permettent de l'utiliser pour protéger efficacement un circuit contre des décharges électrostatiques très intenses. Mais la présence d'une surtension avant le repliement limite son domaine d'application. Ce composant a été utilisé avec succès comme étage primaire d'une protection à deux étages, et retenu comme stratégie de protection pour les entrées dans un circuit commercial.

Afin d'étendre leur domaine d'utilisation, plusieurs techniques de déclenchement ont été testées sur les transistors des technologies 2 et 3.

Nous avons vu que le repliement de la caractéristique est obtenu au-delà d'une certaine densité de courant critique dans le transistor bipolaire. Il est donc nécessaire d'utiliser des techniques de déclenchement qui vont permettre de polariser le transistor jusqu'à ces densités de courant, tout en limitant la tension sur le collecteur. Pour cela, il faut substituer le courant de trous fourni par la multiplication par avalanche dans la jonction collecteur-base, par un courant d'origine différente. Ce courant doit polariser le transistor bipolaire d'une façon similaire au courant issu de la multiplication par avalanche. Cette polarisation doit être maintenue jusqu'à obtenir la densité de courant critique dans le composant qui va induire son repliement.

Déclenchement par transistor PMOS en technologie 3

Une première solution consiste à injecter du courant au moyen d'un élément de déclenchement externe. Pour ce faire, il est judicieux d'utiliser la diffusion P^{++} située à droite de l'émetteur dans les figures 3.21 et 3.22 pour les technologies 2 et 3, respectivement. Le courant de trous initialement fourni par la jonction C-B peut être substitué par un courant injecté à l'aide de ce contact. Cette disposition spatiale permet de conserver le sens de circulation du courant de

trous dans la structure, ce qui ne serait pas le cas si l'on utilisait le contact de base du transistor bipolaire. En effet, quelle que soit la méthode de déclenchement, le courant de base circulera toujours, après le repliement, de la région de collecteur (où il est créé) vers le contact de base.

L'injection d'un courant peut être réalisé au moyen de tout dispositif placé entre le contact de collecteur et le contact P^{++} (noté Extern sur la figure 3.22). Les contraintes dans le choix de ce dispositif portent sur sa tension de déclenchement, la tension à ses bornes lorsqu'il est déclenché et une robustesse suffisante pour conduire le courant nécessaire pour atteindre le repliement du transistor bipolaire. La tension de déclenchement de ce dispositif et la tension à ses bornes lorsqu'il conduit du courant, doivent être suffisamment grandes vis-à-vis de la tension de maintien du transistor bipolaire pour garantir que seul le TBA conduira le courant après son déclenchement. Par exemple, l'utilisation d'un transistor NMOS est à proscrire car sa tension de maintien est trop proche de celle du transistor NPN. Nous avons pu vérifier expérimentalement que la robustesse de cet ensemble (NPN+NMOS) possède la robustesse du transistor NMOS seul.

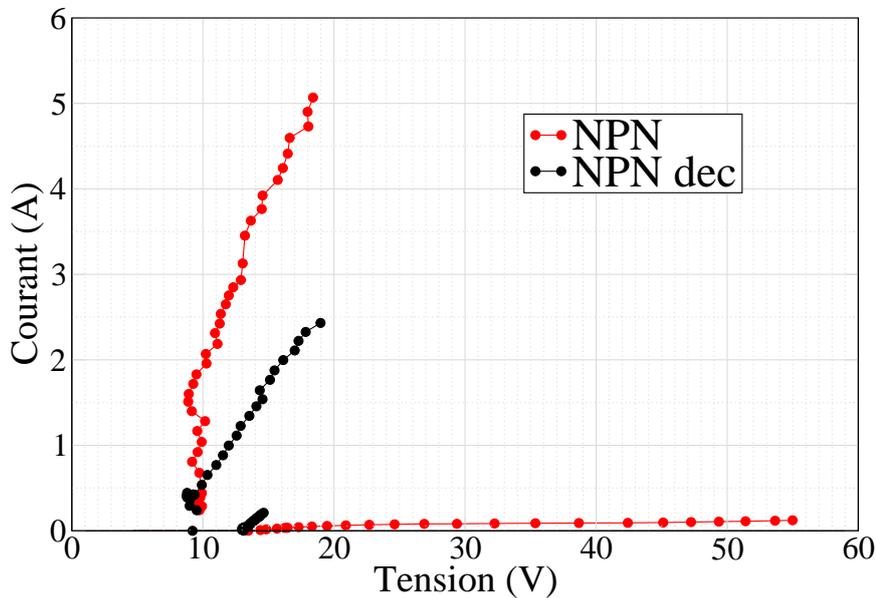


FIG. 3.29 – Caractéristique TLP fort courant du transistor NPN seul et déclenché par un transistor PMOS (NPN dec) (technologie 3).

Les caractéristiques TLP de transistors NPN réalisés en technologie 3, sans dispositif de déclenchement et déclenché par un transistor PMOS avec grille à la masse, sont représentées sur la figure 3.29. La surtension avant le repliement est bien éliminée et le déclenchement du transistor bipolaire est effectif. En revanche, la résistance à l'état passant est doublée dans la structure déclenchée. Ce problème provient du déclenchement d'une moitié seulement du transistor bipolaire (qui possède deux diffusions de collecteur). En effet, le dispositif de déclenchement n'a été connecté que sur l'une des deux diffusions P^{++} , créant ainsi une dissymétrie trop importante pour permettre le déclenchement de l'autre doigt de collecteur. La robustesse du dispositif aux décharges HBM est elle aussi diminuée de moitié. Cette solution avait été initialement retenue afin de réduire le courant nécessaire au déclenchement, l'objectif de cette première structure de test étant de démontrer qu'il est possible de déclencher ces transistors au moyen de dispositifs extérieurs.

Étude de diverses techniques de déclenchement en technologie 2

Diverses approches permettant de réduire la tension de déclenchement ont été testées dans la technologie 2. Les relativement faibles dopages utilisés dans cette technologie impliquent que le niveau du courant critique correspondant au repliement est particulièrement faible. Pour la même raison, la tension de déclenchement est particulièrement élevée. La caractéristique TLP du transistor NPN réalisé dans cette technologie est présentée dans la figure 3.30. La tension de déclenchement est d'environ 130 V et le courant de déclenchement de l'ordre du milliampère. La tension de maintien de 8 V est du même ordre que celle obtenue dans les autres technologies. La tension de claquage collecteur-base statique est de 60 V. La robustesse de composant est de

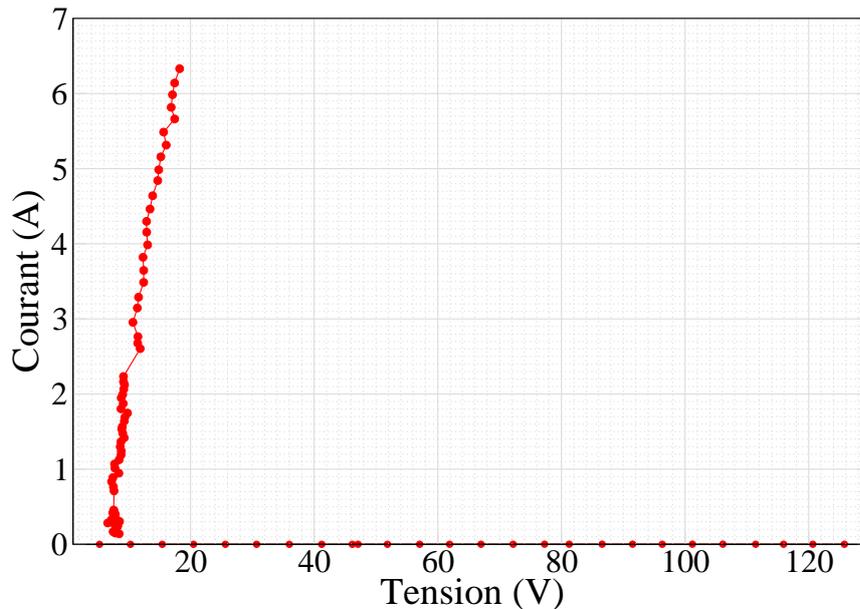


FIG. 3.30 – Caractéristique TLP fort courant du transistor NPN (technologie 2).

14 kV HBM. Trois solutions de déclenchement ont été étudiées pour réduire la surtension.

La première est identique à la technique utilisée dans la technologie 3. Une diode externe au composant permet de fournir le courant de base jusqu'au repliement du transistor bipolaire. Cette diode possède une tension de claquage statique de 28 V. La tension de repliement du transistor bipolaire est alors réduite à 50 V et la robustesse du composant est conservée.

La seconde solution étudiée est basée sur l'utilisation de la diode interne formée par la diffusion P^{++} dans la diffusion de collecteur N^{-} .

La coupe technologique de ce composant est représentée sur la figure 3.31. La distance entre la diffusion P^{++} et la diffusion N^{++} de collecteur a été réduite par rapport au composant de référence. La tension de claquage statique est alors de 44 V et la robustesse du composant ainsi formé est extrêmement faible, inférieure à 2 kV HBM.

Enfin, une troisième solution de déclenchement, plus originale, a été testée. La coupe technologique de ce composant est représentée sur la figure 3.32. Une diffusion de type P^{++} est accolée à la diffusion N^{++} de collecteur. Ces diffusions forment une diode Zener dont la tension de claquage est particulièrement faible. Le fonctionnement de cette structure repose sur le principe des anneaux de garde qui sont utilisés habituellement comme technique pour augmenter la tenue en tension des jonctions cylindriques. Pour une polarisation inverse du composant, la diffusion P^{++} est flottante jusqu'à ce que la région dépeuplée de la jonction collecteur-base l'atteigne, pour une tension dite de perçage. Au-delà de cette tension, le potentiel de la diffusion P^{++} est fixé et la diode Zener se trouve polarisée en inverse. Lorsque la tension de claquage de la Zener est

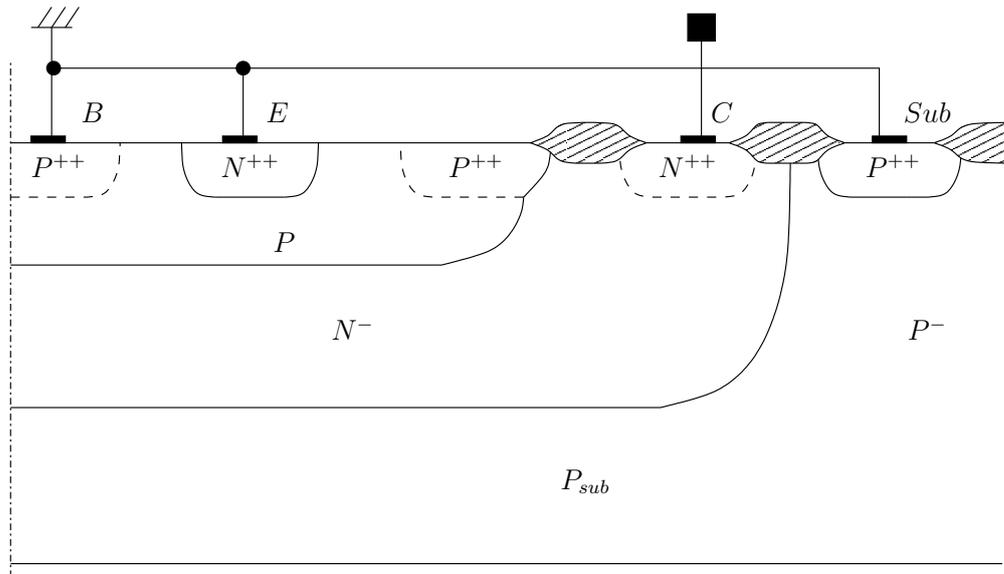
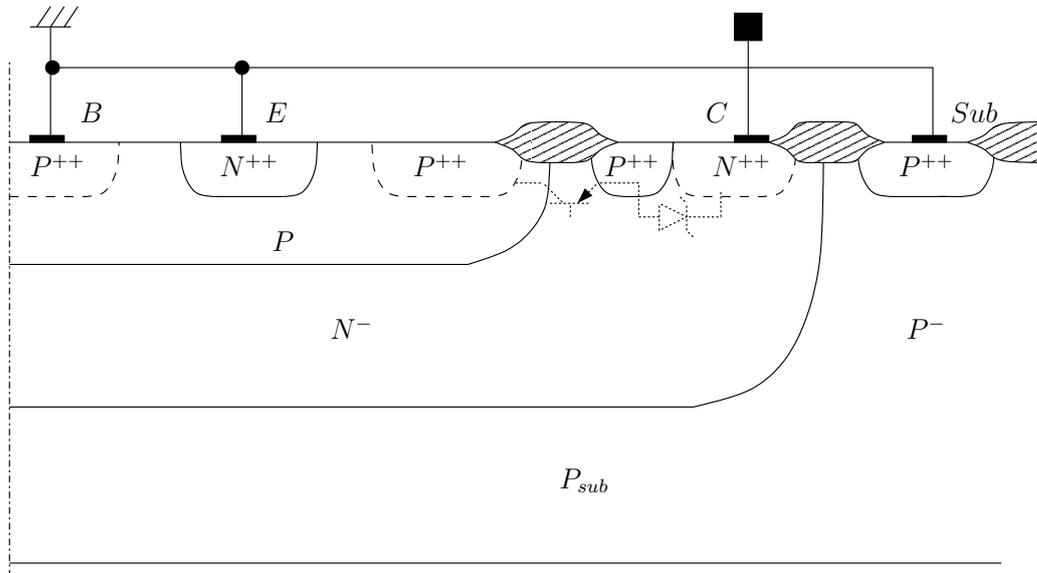


FIG. 3.31 – Coupe technologique du transistor NPN déclenché par diode interne (technologie 2).


 FIG. 3.32 – Coupe technologique du transistor NPN déclenché par diffusion P^{++} flottante (technologie 2).

atteinte, le courant circule dans le composant et permet de déclencher le transistor bipolaire. La tension de mise en route est donc la somme de la tension de perçage et de claquage de la diode Zener. La principale différence avec la technique de déclenchement utilisant la diode P^{++}/N^- réside dans le lieu de la génération par avalanche, qui se situe ici à la jonction P^{++}/N^{++} . Ainsi, une zone de fort champ électrique est présente à proximité de la jonction N^{++}/N^- et favorise le passage du transistor bipolaire dans son régime de fort courant. Dans cette configuration, la tension de claquage statique du composant est de 25 V et sa tension de repliement de 72 V. Ses performances HBM sont améliorées vis-à-vis de celle du composant optimisé précédent. Ce composant n'a pu être dégradé par le niveau de stress maximal de 16,4 kV HBM que peut fournir le testeur utilisé.

Sur les trois techniques de déclenchement étudiées deux s'avèrent donc efficaces. Les caracté-

ristiques TLP obtenues pour des tensions et courants proches du déclenchement sont représentées sur la figure 3.33. L'évolution des caractéristiques statiques après l'application d'une décharge

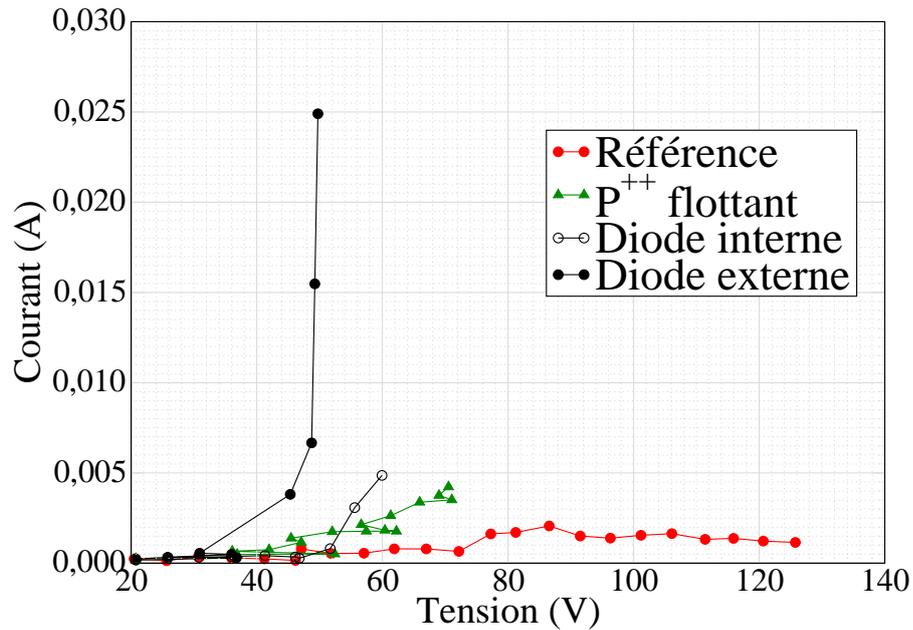


FIG. 3.33 – Caractéristiques TLP des composants NPN dans leurs régions de déclenchement, pour diverses techniques de déclenchement (technologie 2).

HBM qui dégrade le composant est présentée sur la figure 3.34. La signature électrique de la défaillance correspond toujours à une diminution de la tension de claquage pour les dispositifs les plus robustes. En revanche, une élévation du courant de fuite est observée pour le composant peu robuste déclenché par la diode interne. Ces signatures de défaillance donnent un indice sur la localisation du défaut dans la structure.

L'élévation du courant de fuite correspond à l'apparition d'un défaut court-circuitant la jonction métallurgique collecteur-base. La faible robustesse du composant déclenché par la diode interne peut donc être attribuée à la dégradation de cette diode associée à la formation d'un filament court-circuitant sa jonction métallurgique.

La diminution de la tension de claquage, observée dans tous les autres cas, ne peut être attribuée à un filament court-circuitant la jonction. Le fonctionnement du composant bipolaire à forte densité de courant s'accompagne de la formation d'une région de haute température à la jonction effective. La diminution de la tension de claquage peut donc être attribuée à la formation d'un filament de silicium fondu initialisé dans cette région et qui s'est propagé vers la jonction métallurgique collecteur-base sans l'avoir transpercée. Ainsi, aucune modification du courant de fuite du composant n'est observée tant que le champ électrique et la charge d'espace de la jonction collecteur-base n'ont atteint le défaut. Cette interprétation va être confirmée par l'analyse de défaillance.

3.7.8 Analyse de défaillance

Dans la technologie 3, des transistors NPN ont été dessinés pour permettre leur analyse par la face avant. Pour cela, la disposition des métallisations est adaptée pour laisser découvertes les surfaces de silicium susceptibles de présenter un intérêt pour l'analyse. Ces composants issus du premier lot de test ne sont pas totalement optimisés, en particulier en termes de distance collecteur-émetteur, et possède une robustesse légèrement inférieure à 5 kV HBM. Cependant,

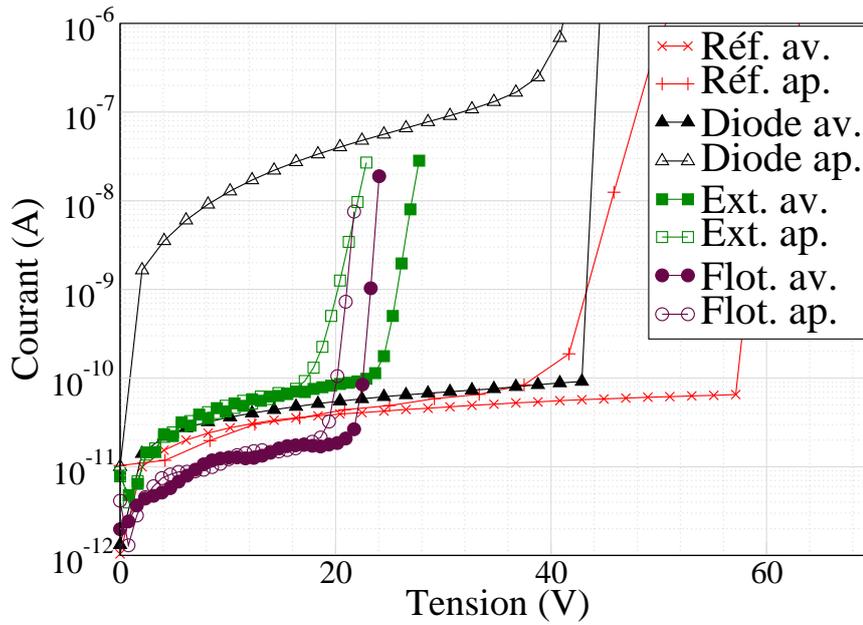


FIG. 3.34 – Mesures de fuites avant (av.) et après (ap.) l’application d’un stress destructeur pour les transistors NPN selon différentes techniques de déclenchement (technologie 2). Réf. : composant de référence, Diode : déclenchement par la diode interne, Ext. : déclenchement par une diode externe, Flot. : déclenchement par région P^{++} flottante.

le mécanisme de défaillance qui va être mis en évidence est représentatif du mode de défaillance de composants plus robustes car leurs signatures de défaillance électrique sont identiques.

Les évolutions des caractéristiques statiques de deux composants NPN1 et NPN2, après plusieurs stress HBM d’intensité croissante, sont présentées sur les figures 3.35 (a) et (b) respectivement.

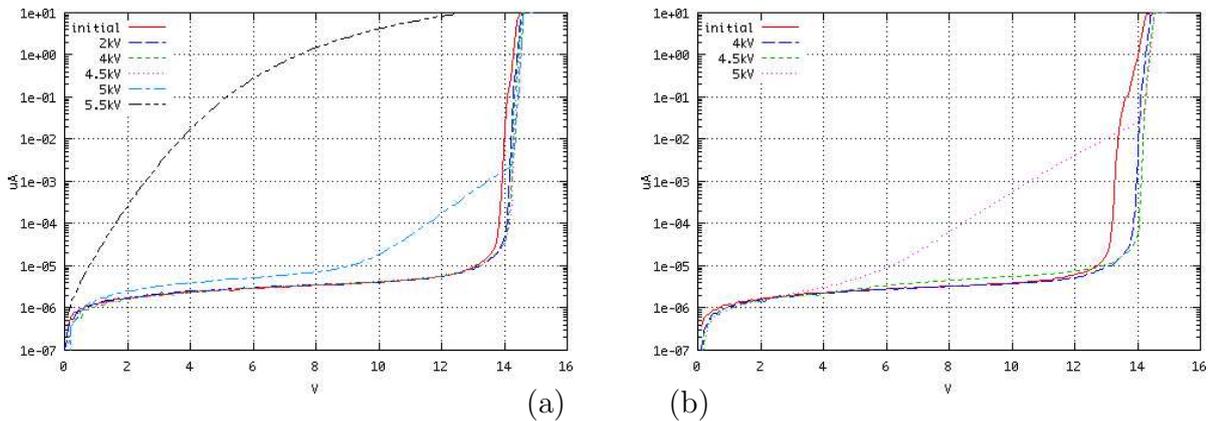


FIG. 3.35 – Évolution de la caractéristique statique en inverse pour des niveaux croissants de stress HBM, de deux composants (technologie 3) NPN1 (a) et NPN2(b).

Les techniques d’analyse NBOBIC et OBIC permettent une étude particulièrement fine de la nature du défaut présent dans le silicium. Leur principe de fonctionnement a été présenté dans le premier chapitre. Rappelons simplement que la présence d’un filament fondu dans le silicium se traduit par une diminution systématique du signal NBOBIC dans la région du défaut. Par contre, la variation du signal OBIC dépend des caractéristiques du défaut. Si le défaut ne court-

circuite pas la jonction, le signal OBIC diminue dans la région du défaut et se comporte donc comme le signal NBOBIC. Si le défaut court-circuite la jonction, une augmentation du signal est observée.

Les images obtenues au moyen de ces techniques sur les composants NPN1 et NPN2, sont présentées dans la figure 3.36. Les images OBIC sont obtenues pour une tension de 5 V sur le collecteur du transistor. Seules les techniques OBIC ont permis la détection de défauts dans ces deux composants. La photo-émission et la technique TLS ne donnent pas de résultats car les courants de fuite mis en jeu sont trop faibles [126].

La caractéristique tension-courant statique du composant NPN1 après le dernier stress HBM, traduit la présence d'un défaut court-circuitant la jonction collecteur-base. L'image OBIC du composant (Fig. 3.36(a)), dans la région grisée représentée sur la figure 3.36(e), montre une élévation locale du signal OBIC, traduisant la présence de ce type de défaut. La réduction du signal NBOBIC (Fig. 3.36(b)), dans la même région, confirme sa présence.

La dégradation du transistor NPN2 se traduit par un décalage de la tension de claquage, ou du moins l'accroissement significatif du courant de fuite pour une tension de l'ordre de 6 V. Malgré le très faible courant de fuite, dû à la faible taille du défaut présent, l'image OBIC (Fig. 3.36(c)) laisse apparaître assez clairement une diminution du signal sur la gauche du doigt N^{++} de collecteur. L'image NBOBIC (Fig. 3.36(d)) confirme la présence d'un défaut dans cette région. Ceci permet d'affirmer que le défaut ne court-circuite pas la jonction collecteur-base.

Le mécanisme de défaillance correspond donc bien à un filament de silicium fondu, dont la formation débute dans la région de forte température située à la jonction effective (N^{++}/N^{-}), et qui se propage vers la jonction métallurgique collecteur-base. Après refroidissement, ce filament constitue une région polycristalline qui selon l'amplitude du stress appliqué, peut venir court-circuiter la jonction collecteur-base ou pas.

Ce mode de défaillance est tel qu'il peut y avoir formation de plusieurs filaments avant de conduire à un court-circuit. Cela permet donc de retarder la défaillance catastrophique et participe probablement à améliorer la très grande robustesse de ce type de composant.

3.8 Cas d'un TBA PNP

Le modèle semi-analytique élaboré dans le chapitre 2 a permis de montrer que le comportement des TBA NPN et PNP sont très différents pour des profondeurs de base importantes ($>2\mu\text{m}$). La tension aux bornes de la région de collecteur d'un TBA NPN décroît rapidement avec la densité de courant, alors que celle d'un PNP augmente. La réalisation d'un TBA PNP équivalent aux transistors NPN présentés dans ce chapitre est donc fondamentalement impossible. Afin de confirmer cet aspect, un transistor PNP a été dessiné dans la technologie 3, en adaptant le dessin de masque du meilleur transistor NPN réalisé.

La robustesse HBM de ce transistor est très faible 0,5 kV. Sa caractéristique TLP est représentée dans la figure 3.37. Aucun repliement de la tension de collecteur n'est observé et sa résistance dynamique est très élevée.

La faible tension obtenue pour le dernier point de la caractéristique TLP témoigne de l'apparition d'un second claquage thermique dans le composant. La forte augmentation du courant de fuite observée dans la caractéristique statique (de la dizaine de pico-ampères au micro-ampère), témoigne de la dégradation du composant après cet événement.

3.9 Conclusion

Dans ce chapitre, une étude approfondie des TBA de type NPN possédant une région de collecteur faiblement dopé a été menée. Après avoir bien mesuré leurs limitations, les outils de simulation physique bidimensionnel ont permis une analyse approfondie du fonctionnement de ces composants.

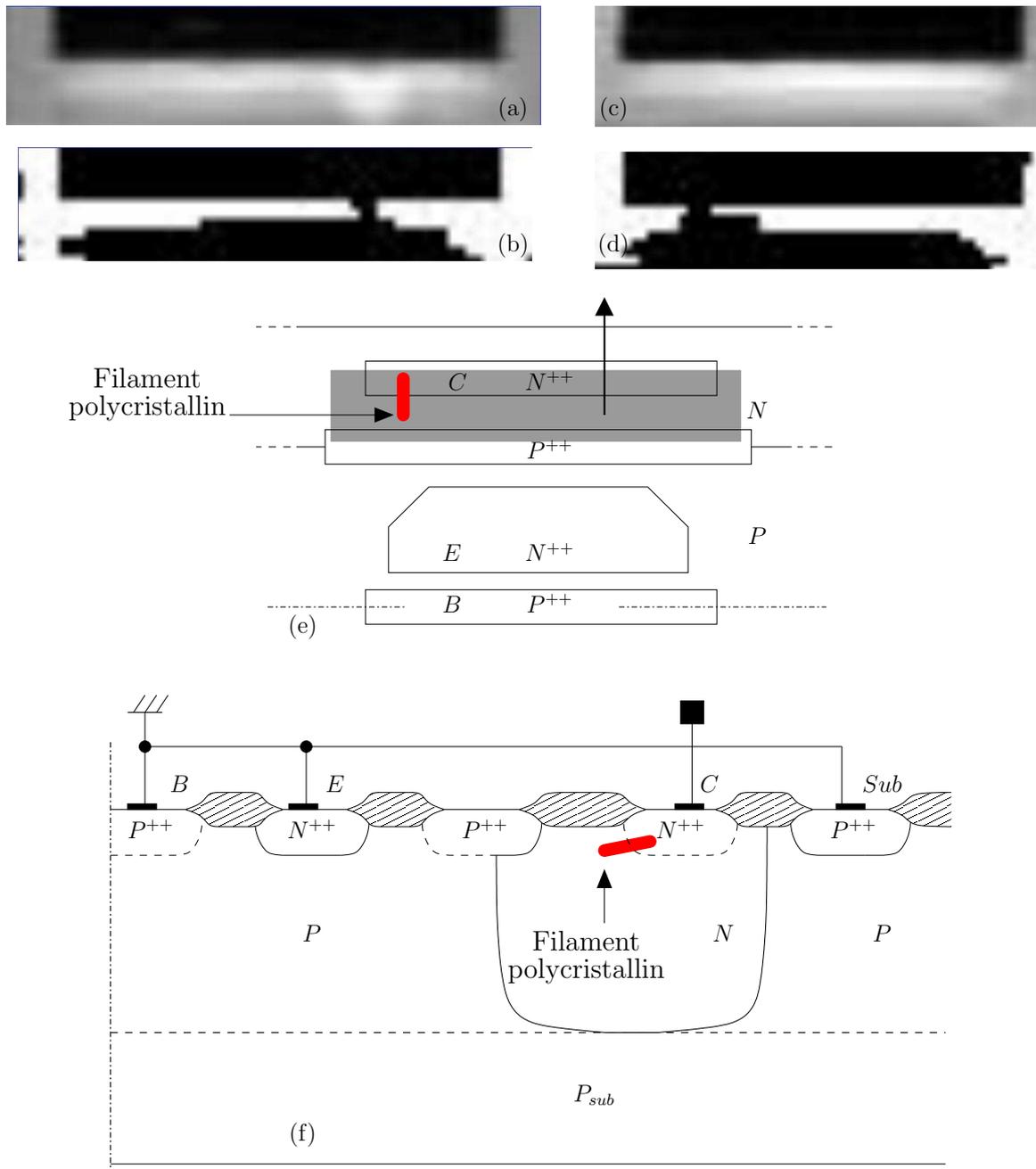


FIG. 3.36 – Analyse de défaillance des transistors NPN1 et NPN2 par les techniques d’analyse OBIC (NPN1(a) NPN2(c)) et NBOBIC (NPN1(b) NPN2(d)). La région analysée est grisée dans la représentation schématique de la vue de dessus du composant (e). Position du filament polycristallin représenté sur la coupe technologique du composant (f).

L’étude d’un composant bipolaire NPN vertical a permis de mettre en évidence trois régimes de fonctionnement distincts en fonction de l’ordre de grandeur des densités de courant dans le transistor. Ces régimes sont liés à l’évolution des caractéristiques électriques de la jonction collecteur-base qui sont modifiées par la charge électrique des porteurs libres. L’ensemble des grandeurs électriques caractéristiques des TBA ont pu être liées aux paramètres géométriques et technologiques, comme par exemple :

- la tension de repliement qui dépend principalement de la profondeur de la couche faiblement dopée de collecteur.

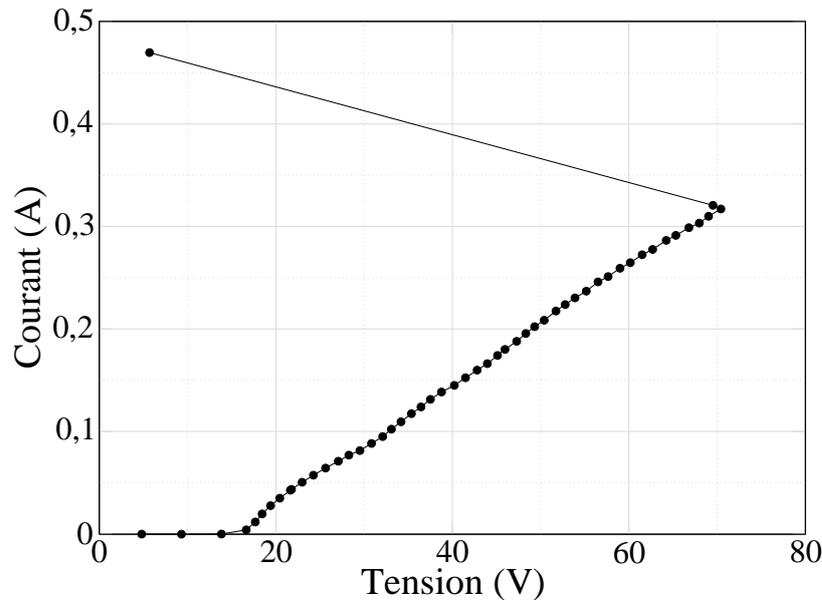


FIG. 3.37 – Caractéristiques TLP du composants PNP (technologie 3).

- le courant de repliement qui est proportionnel à la taille du composant et qui dépend du dopage de la couche faiblement dopée de collecteur.
- la tension de maintien qui est fonction du profil de dopage à la frontière des couches faiblement et fortement dopées du collecteur.

Les origines de la focalisation du courant dans les TBA ont été mises en évidence. De nature purement électrique, la focalisation du courant est principalement due, dans le régime de fort courant, à la formation d'une jonction collecteur-base effective dont la tension de claquage décroît avec la densité de courant. Cet aspect de la focalisation a été traité plus en détail dans le chapitre 2. Dans une moindre mesure, la résistance intrinsèque de base est aussi à l'origine d'une non-uniformité du courant.

La cause de la défaillance d'un TBA est associée à l'initialisation d'un second claquage thermique. Afin de retarder l'apparition de ce phénomène destructeur, nous avons vu comment éviter la formation de régions plus chaudes dans le composant. Nous avons montré au moyen de la simulation physique comment l'augmentation de la profondeur de base effective assure un meilleur découplage à la fois thermiquement et électriquement des régions électriques d'émetteur et de collecteur. Ce découplage permet de favoriser les effets qui réduisent la densité de courant dans les régions les plus chaudes et de limiter les phénomènes d'emballement thermique. Pour des décharges de faible intensité, nous avons montré le mouvement de la zone focalisée de courant sous l'effet de la température. Ce mouvement est difficilement observable expérimentalement mais Pogany et ses collaborateurs ont réussi à le mettre en évidence très récemment pour des composants similaires, grâce aux techniques d'interférométrie laser.

Un ensemble de recommandations pour la réalisation de composants de protection performants a découlé de cette étude approfondie des TBA NPN possédant une région de collecteur faiblement dopée.

Les résultats expérimentaux obtenus dans trois technologies viennent confirmer et illustrer le bien fondé de notre étude théorique. Une limite importante pour l'utilisation des composants développés provient de leur tension de repliement très élevée. Plusieurs techniques de déclenchement ont été proposées et vérifiées expérimentalement. L'impossibilité de réaliser ce type de composant à base effective très profonde avec des transistors bipolaires PNP a également été démontrée expérimentalement.

Enfin, le mode de défaillance des composants a pu être mis en évidence par les techniques d'analyse de défaillance très fines, OBIC et NBOBIC, basées sur la génération photoélectrique par faisceau laser.

Chapitre 4

Modélisation de composants bipolaires autopolarisés

Le nombre de cycles de conception des circuits intégrés associé aux problèmes de protection contre les décharges électrostatique s'accroît avec la complexité des circuits et la réduction des dimensions technologiques. Il est donc primordial de prendre en compte et vérifier, dès le début de la phase de conception d'un circuit, le bon fonctionnement de la stratégie de protection choisie. D'une part, les éléments parasites amenés par les composants de protection ne sont pas toujours négligeables, et d'autre part, l'absence d'interaction entre la stratégie de protection et la partie fonctionnelle du circuit, lors du fonctionnement normal aussi bien que lors d'une décharge électrostatique, doit être vérifiée. Pour cela, des outils de conception spécifiques doivent être introduits dans l'environnement de conception des circuits intégrés. La modélisation compacte de type SPICE des composants de protection pour les régimes de fort courant est l'un des éléments nécessaires pour atteindre cet objectif.

Le second chapitre a fait apparaître les spécificités des phénomènes de fort courant dans les TBA. Sans changer l'approche retenue au sein du laboratoire, qui consiste à baser le modèle des composants ESD sur un modèle SPICE classique auquel sont ajoutés des éléments spécifiques, nous l'enrichirons des effets des fortes densités de courant mis en évidence dans les chapitres précédents.

4.1 État de l'art de la modélisation de type SPICE des protections ESD

Les éléments utilisés comme dispositifs de protection contre les ESD sont des composants classiques (diodes, transistors bipolaires, transistors MOS, ...) dont le dessin technologique est adapté aux régimes de fort courant dans lesquels ils opèrent lors d'une décharge. Toutefois, les modèles SPICE standards ne sont pas adaptés pour décrire ces régimes de fonctionnement extrêmes. Les modèles doivent être étendus pour décrire les phénomènes physiques caractéristiques de leur fonctionnement comme le claquage par avalanche des jonctions, les caractéristiques à résistance négative (ou repliement), la modulation de conductivité, etc ...

C'est en premier lieu pour les technologies CMOS que les décharges électrostatiques ont posé un problème majeur. Aussi, le composant de protection qui a été le plus étudié et modélisé est le transistor bipolaire NPN latéral parasite des transistors NMOS. La modélisation des transistors NPN verticaux est par contre très peu abordée dans la littérature car il n'a été utilisé que plus récemment avec l'avènement des technologies BiCMOS et de puissance intelligente.

4.1.1 Transistors bipolaires NPN

Le fonctionnement des transistors bipolaires NPN utilisés comme protection ESD a été décrit dans la section 2.1.2. Rappelons que le fonctionnement le plus complexe de ce dispositif correspond au cas où une décharge positive est appliquée sur le collecteur, l'émetteur et la base court-circuités étant portés au potentiel de référence nul. Le transistor est initialement bloqué. La tension de collecteur s'accroît jusqu'à atteindre la tension de claquage par avalanche de la jonction collecteur-base. Le courant circulant dans la résistance interne (ou intrinsèque) de base permet alors de polariser la jonction base-émetteur en direct, et de déclencher le transistor bipolaire.

Dans l'autre configuration de fonctionnement, pour une décharge négative, la diode collecteur-base du composant est polarisée en direct. Ce cas peut être facilement traité avec des modèles classiques. En revanche, le mode de fonctionnement en inverse ne peut être décrit directement par les modèles SPICE standards [127] du transistor bipolaire (modèles d'Ebers-Moll ou de Gummel-Poon), qui n'incluent pas les phénomènes physiques liés au claquage par avalanche des jonctions. Le modèle VBIC [128] prend en compte ces mécanismes à la jonction collecteur-base, mais est limité au seul cas de faible multiplication, ce qui n'est pas suffisant pour les composants ESD.

Les modèles standards doivent donc être étendus au régime de fort courant propre aux décharges électrostatiques. Ce faisant, deux approches sont envisageables. La première consiste à développer un modèle complet et spécifique [59]. L'inconvénient majeur de cette méthode est qu'elle nécessite une extraction complète des paramètres électriques du nouveau modèle. La seconde approche, plus couramment utilisée, consiste à adapter et enrichir les modèles existants (« macromodélisation »). Ainsi, des éléments de modèle propres aux forts niveaux de tension et de courant sont ajoutés au modèle classique. L'avantage de cette technique est qu'elle limite le nombre global de paramètres électriques à extraire aux paramètres spécifiques aux régimes de claquage par avalanche et de fort courant.

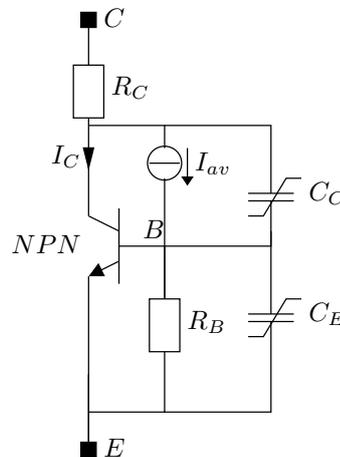


FIG. 4.1 – Modèle compact d'un transistor NPN autopolarisé.

Un schéma électrique du modèle compact du transistor NPN autopolarisé est présenté sur la figure 4.1 [103]. Divers éléments sont associés au modèle standard du transistor bipolaire représenté par l'élément NPN. La source de courant I_{av} permet de décrire les phénomènes d'avalanche dans la jonction collecteur-base. La résistance R_B placée entre la base et l'émetteur représente la résistance interne de base qui sert à polariser la jonction base-émetteur en direct et ainsi à déclencher le transistor. Enfin, les capacités non-linéaires des jonctions collecteur-base (C_C) et émetteur-base (C_E) sont particulièrement importantes pour décrire la réponse dynamique du transistor.

La valeur du courant d'avalanche I_{av} est calculée en fonction du courant I_C dans le collecteur du transistor au moyen de l'expression :

$$I_{av} = (M - 1)I_C \quad (4.1)$$

où M est le facteur de multiplication par avalanche de la jonction collecteur-base, donné par la formule de Miller [67] :

$$M = \frac{1}{1 - \left(\frac{V_{CB}}{BV_{CB}}\right)^m} \quad (4.2)$$

avec $2 < m < 6$ un nombre qui dépend des caractéristiques de la jonction, déterminé de manière empirique. Ainsi, dès que la tension aux bornes de la jonction collecteur-base (V_{CB}) atteint sa tension de claquage BV_{CB} , la source I_{av} fournit le courant pour déclencher la transistor bipolaire et le maintenir en fonctionnement.

La valeur de la résistance de base R_B est généralement considérée comme constante malgré la modulation de conductivité dans la base. Elle est obtenue à partir de l'expression :

$$R_B = \frac{V_{BE}}{I_{t1}} \quad (4.3)$$

où la tension aux bornes de la jonction émetteur-base V_{BE} est d'environ $0,5 V$ au point de repliement du transistor [40], et I_{t1} est le courant de repliement du composant déterminé à partir de la caractéristique TLP.

Lorsque le transistor NPN est un composant standard de la technologie, il est intéressant de réutiliser directement son modèle électrique fourni dans la bibliothèque 4.1 [115, 129, 130]. Le nombre de paramètres électriques à extraire est ainsi limité. En revanche, certains paramètres de ce modèle tels que les capacités non-linéaires des jonctions devront être supprimés pour éviter les redondances avec celles (C_C et C_E) placées en externe. Dans le cas contraire [103], les paramètres d'un modèle d'Ebers-Moll ou de Gummel-Poon classiques devront aussi être extraits.

La tension de maintien V_H dépend fortement du facteur de multiplication par avalanche du courant de collecteur M et du gain du transistor bipolaire β . Différentes méthodes sont utilisées pour calibrer V_H en ajustant les facteurs M [103, 129] et β [103].

La résistance à l'état passant après le déclenchement est principalement liée aux résistances de collecteur R_C . Sa valeur est directement extraite de la caractéristique TLP. De plus, selon le modèle choisi pour le transistor NPN, certains effets comme la dépendance entre le gain du transistor et le courant de collecteur ne sont pas systématiquement inclus. Ainsi, pour calibrer la résistance à l'état passant du modèle à fort niveau, le gain du transistor dans cette région est également ajusté [103].

4.1.2 Transistors NMOS

Lorsqu'il est utilisé comme structure de protection contre les ESD, le transistor NMOS a généralement sa grille, sa source et son substrat court-circuités à la masse. Reposant sur l'action du transistor bipolaire latéral parasite, le comportement du composant le plus critique à modéliser correspond à une décharge électrostatique positive appliquée sur son drain. Le fonctionnement et le modèle de cette structure sont semblables à ceux d'un transistor NPN. Pour une décharge positive, la tension de drain monte jusqu'à atteindre la tension de claquage de la jonction drain-substrat. Le courant résultant permet de polariser localement la jonction source-substrat en direct au travers de la résistance de substrat. Le transistor bipolaire NPN parasite est alors déclenché, le drain jouant le rôle de collecteur, le substrat de base et la source d'émetteur.

L'une des spécificités des transistors NMOS provient de la possibilité d'utiliser l'action du transistor MOS pour réduire la tension de déclenchement du dispositif. Pour cela, il faut appliquer une tension de grille supérieure à la tension de seuil du transistor NMOS, le plus couramment en utilisant un couplage capacitif (section 2.1.3). Le courant circulant alors dans le canal

du transistor MOS est multiplié par avalanche à la jonction drain-substrat, fournissant ainsi le courant de substrat (base) qui permet de déclencher le transistor NPN à des niveaux de tension plus faibles que la tension de claquage drain-substrat.

Le schéma électrique typique du macromodèle couramment utilisé pour décrire le comportement des transistors NMOS au régime de fort courant et de l'ESD est présenté dans la figure 4.2 [131, 132, 133, 61]. Le modèle est en tous points semblable à celui des transistors NPN autopolarisés avec en parallèle un transistor NMOS.

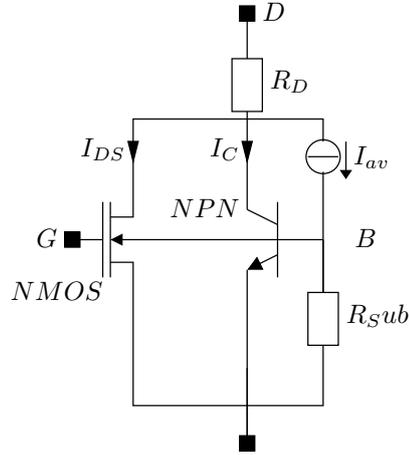


FIG. 4.2 – Modèle compact d'un transistor NMOS incluant son transistor bipolaire NPN parasite.

L'expression du courant d'avalanche I_{av} dépend maintenant à la fois du courant du MOS (I_{DS}) et du courant dans le transistor bipolaire (I_C). Il est donné par l'expression :

$$I_{av} = (M - 1)(I_{DS} + I_C) \quad (4.4)$$

Plusieurs expressions du facteur de multiplication par avalanche M sont rencontrées dans les publications. La plus classique est basée sur la formule empirique de Miller (Equ : 4.5). Afin de prendre en compte la variation de la tension de maintien en fonction de la polarisation de grille [134], la tension de saturation $V_{Dsat} = f(V_G)$ du transistor NMOS est soustraite à la tension appliquée sur le drain, ce qui conduit à l'expression suivante pour le facteur de multiplication du courant [133, 135] :

$$M = \frac{1}{1 - \left(\frac{V_D - V_{Dsat}}{BV_{DSS}} \right)^m} \quad (4.5)$$

Une seconde formulation a été établie par Amerasekera [134]. Nous avons vu au chapitre 2 que le facteur de multiplication M est lié aux intégrales d'ionisation \mathfrak{S} par la relation :

$$M = \frac{1}{1 - \mathfrak{S}} \quad (4.6)$$

Si l'on considère les coefficients d'ionisation des électrons et des trous égaux ($\alpha_n = \alpha_p = \alpha$), les intégrales d'ionisation des électrons et des trous (Equ : 2.26) sont alors identiques et se simplifient en :

$$\mathfrak{S}_n = \mathfrak{S}_p = \mathfrak{S} = \int \alpha dx \quad (4.7)$$

Les coefficients d'ionisation étant de la forme :

$$\alpha = \alpha^\infty \exp\left(-\frac{E^{crit}}{E}\right) \quad (4.8)$$

le facteur de multiplication s'exprime alors de la façon suivante :

$$M = \frac{1}{1 - K_1 \exp\left(\frac{-K_2}{V_D}\right)} \quad (4.9)$$

où K_1 et K_2 sont deux coefficients pouvant être extraits de la mesure statique de $\ln(1 - 1/M)$ en fonction de $1/V_D$ dans la zone de claquage par avalanche de la jonction drain-substrat, avant le déclenchement du transistor bipolaire.

Plusieurs approches ont été proposées afin de modéliser la résistance de substrat R_{SUB} . La plus simple consiste à choisir une résistance de valeur constante. Elle est calculée de façon similaire à la résistance de base du transistor bipolaire présenté à la section précédente, pour ajuster le point de déclenchement du transistor NPN [133, 134]. La modélisation à l'aide de lignes de transmission, qui permet de tenir compte de la géométrie de la structure et de l'épaisseur du substrat, est présentée dans [136, 137]. La modulation de conductivité dans la base implique une réduction de la valeur de la résistance de base à fort niveau de courant. Afin de prendre en compte cet aspect, la résistance de base peut être modélisée par une fonction mathématique [138].

Les paramètres du modèle du transistor bipolaire NPN, généralement basé sur le modèle d'Ebers-Moll [133], sont extraits des caractéristiques statiques et TLP. La tension de maintien dépend des paramètres M et β qui sont finalement ajustés lors du calibrage des simulations. Pour les forts courants, la résistance dynamique du composant est principalement due à la résistance de ballast dans le drain et aux résistances des prises de contacts. Ce dernier aspect est pris en compte par la résistance série R_D dans le drain.

4.2 Approche retenue et justification

Le principal objectif de la modélisation compacte des protections ESD est de permettre la simulation globale d'un circuit avec ses protections ESD associées. Deux types de simulation sont envisageables :

- l'étude du fonctionnement normal du circuit
- l'analyse de l'efficacité du circuit de protection ESD

Dans le premier cas, les modèles doivent rendre compte des éléments parasites que représentent les composants de protection ESD, afin de vérifier qu'ils ne perturbent pas le fonctionnement normal. Dans le second cas, les modèles doivent permettre d'analyser la réponse du circuit lors d'un stress ESD, donc à fort niveau de courant, pour vérifier qu'il n'existe aucune surtension ou courant trop important dans le circuit à protéger.

Les modèles classiques de composants sont parfaitement adaptés au premier type de simulation, qui correspond à des niveaux de tension et de courant classiques. De plus, les protections ESD sont généralement réalisées à partir des composants standards de la technologie. Des modèles aux paramètres éprouvés et parfaitement calibrés sont donc disponibles dans les bibliothèques de composant. Il est donc particulièrement intéressant de construire le modèle des composants de protection ESD à partir de ces derniers, en ajoutant des éléments de circuits supplémentaires pour prendre en compte le régime de fonctionnement à fort courant lors d'une décharge électrostatique. L'extraction des paramètres est ainsi réduite aux régimes de fort courant, ce qui évite un travail redondant pour l'extraction des paramètres classiques.

Au delà de cet argument purement pratique, nous pouvons différencier deux régimes de fonctionnement pour un TBA polarisé en inverse au cours d'une décharge électrostatique de type HBM [139]. En début de décharge, la tension varie très rapidement aux bornes du composant jusqu'à son déclenchement puis son repliement. Dans le même temps, le courant total dans la structure reste à des niveaux suffisamment faibles pour que les modèles classiques du composant restent valides. Après le déclenchement, les niveaux de courant sont très grands, mais la tension aux bornes du composant varie peu et relativement lentement par rapport au début de la décharge. Nous pouvons donc distinguer un premier régime de faible courant avec de très fortes et

rapides variations de tension, qui précède un second régime de fort courant où les variations de tension sont beaucoup plus faibles et plus lentes.

Nous pouvons donc profiter des modèles classiques pour décrire le comportement du composant dans le premier régime où les niveaux de courant sont suffisamment faibles. Le comportement dans le second régime sera essentiellement décrit par les éléments de circuits supplémentaires qui devront essentiellement rendre compte des effets des très fortes densités de courant.

Les modèles précédemment développés au laboratoire [61, 115] sont basés sur cette approche. Nous ne souhaitons pas ici en redémontrer l'efficacité, mais montrer comment les effets spécifiques des fortes densités de courant peuvent être intégrés dans ce type de modélisation.

4.3 Modélisation des TBA à collecteur faiblement dopé

Nous nous intéresserons plus particulièrement ici aux effets des fortes densité de courant. Seul l'aspect statique, déjà complexe, sera abordé. Les aspects dynamiques seront traités plus en détail dans la section 4.6.

4.3.1 Présentation générale du modèle

La méthodologie utilisée pour développer le modèle repose sur l'ajout d'éléments de circuit (source de courant, résistance, capacités non linéaires, ...) au modèle standard du transistor bipolaire.

La figure 4.3(a) représente la structure physique unidimensionnelle d'un TBA avec son schéma électrique équivalent.

Le schéma électrique utilisé pour la simulation est représenté dans la figure 4.3(b). Une source de courant I_{av} est placée entre le collecteur et la base du transistor. Le courant issu de la multiplication par avalanche dans la région électrique de collecteur est calculé en fonction de la tension collecteur-base et du courant I_C^{int} circulant dans le collecteur du transistor. Cette source permet également de simuler le claquage par avalanche de la jonction. Le courant d'avalanche se partage entre la base du transistor bipolaire et la résistance de base R_B^{int} .

Les capacités non linéaires des jonctions collecteur-base, émetteur-base et collecteur-substrat, qui ne sont pas représentées ici, doivent être supprimées du modèle du transistor et remplacées par des diodes supplémentaires [61]. En effet, la valeur du courant I_C^{int} circulant dans le collecteur du transistor est utilisée pour calculer le courant issu de la multiplication par avalanche. Or, les courants capacitifs dans la jonction collecteur-base ne sont pas multipliés par avalanche car ils ne correspondent pas à un déplacement physique de porteurs dans cette région, mais à une variation temporelle de champ électrique⁷. Idéalement, il est donc nécessaire de les différencier des courants de porteurs pour calculer le courant d'avalanche et obtenir une représentation correcte du fonctionnement dynamique du transistor.

Le courant d'avalanche fourni par la source est donné par la relation utilisé par Dutton [129] :

$$I_{av} = k_1(M - 1)I_C^{int} \quad (4.10)$$

avec M le facteur de multiplication par avalanche et k_1 un facteur empirique qui permettra d'ajuster la tension de repliement.

⁷Souvent appelés, par abus de langage, courant de déplacement, ces pseudo-courants permettent d'assurer la continuité et la conservation du courant. Ils sont donnés, dans les équations de Maxwell, par la relation $J_{dep} = \epsilon \frac{\partial E}{\partial t}$, en tout point de l'espace.

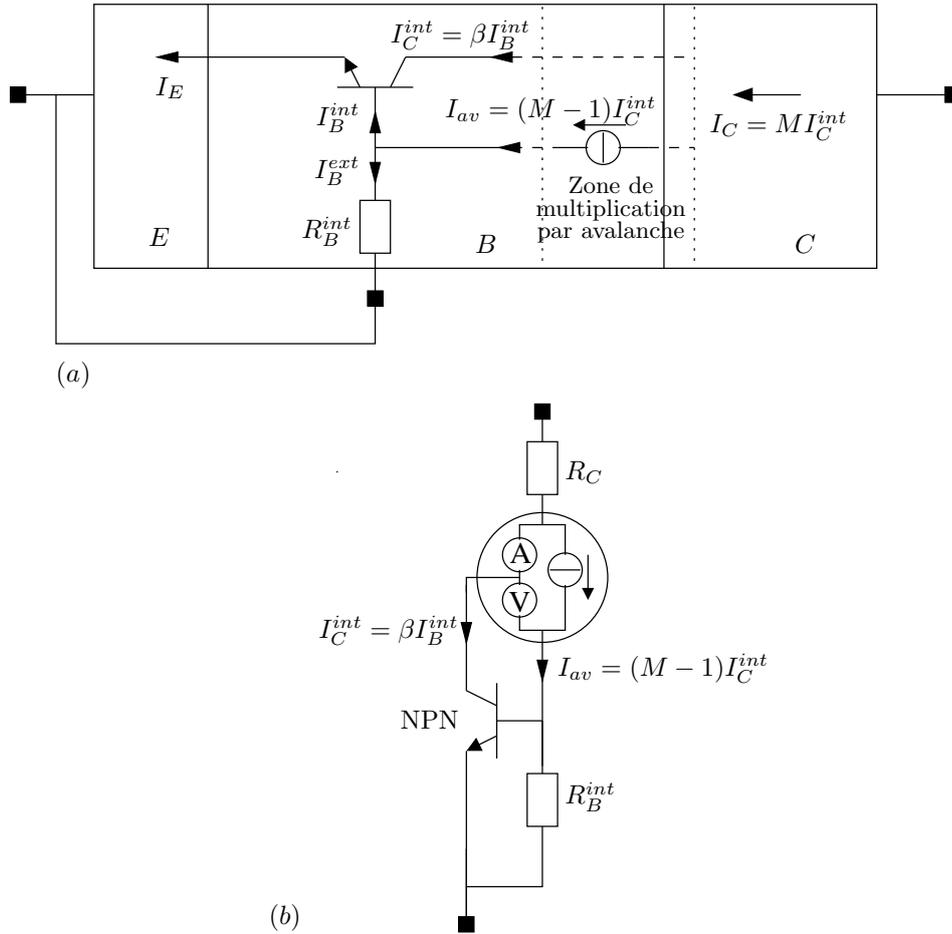


FIG. 4.3 – Modélisation du comportement de la région électrique de collecteur. Modèle unidimensionnel du transistor NPN autopolarisé (a) et son schéma électrique équivalent pour la simulation (b).

4.3.2 Méthode de calcul du facteur de multiplication

Dans un transistor NPN, le courant de collecteur correspond à un courant d'électrons. Après le déclenchement du transistor, ce courant est multiplié par avalanche dans la jonction collecteur-base et fournit le courant de base nécessaire pour maintenir la jonction émetteur-base polarisée en direct. Il est donc important de calculer aussi précisément que possible la valeur du facteur de multiplication par avalanche, d'autant plus qu'il détermine directement la tension de maintien du composant. Pour cela, l'expression du facteur de multiplication est calculée à partir de l'intégrale d'ionisation des électrons \mathfrak{S}_n dans une jonction plane abrupte. Le facteur de multiplication par avalanche du courant de collecteur I_C^{int} est alors donné par :

$$M = \frac{1}{1 - \mathfrak{S}_n} \quad (4.11)$$

Le principal inconvénient de cette formulation découle de la singularité de l'expression lorsque la tension de claquage est atteinte. Dans ces conditions, la valeur de \mathfrak{S}_n tend vers 1 et le dénominateur de l'expression tend vers 0. La variation de M est donc extrêmement rapide pour les tensions proches de la tension de claquage et implique des problèmes de convergence. L'aspect le plus pénalisant pour la stabilité des calculs provient cependant du changement de signe de l'expression 4.11, lorsque \mathfrak{S}_n devient⁸ supérieur à 1. La solution que nous proposons

⁸Ceci peut se produire durant la simulation si un pas de calcul en tension trop important est utilisé.

permet de résoudre ces problèmes sans pénaliser ni le temps de calcul, ni la précision pour le régime de fonctionnement autour de la tension de maintien après le repliement de la tension de collecteur.

Afin, d'éliminer le changement de signe pour $\mathfrak{S}_n > 1$, nous pouvons approximer l'équation 4.11 au moyen de son développement limité :

$$M \approx \sum_{j=0}^{Nb} (\mathfrak{S}_n)^j \quad (4.12)$$

avec Nb le nombre de termes. On remarque que ce développement correspond à la somme des Nb termes d'une suite géométrique de raison \mathfrak{S}_n , ce qui conduit à :

$$M \approx \frac{1 - \mathfrak{S}_n^{Nb+1}}{1 - \mathfrak{S}_n} \quad (4.13)$$

Cette dernière expression présente d'excellentes propriétés pour assurer la convergence et la précision des calculs. Sa valeur est strictement positive quel que soit \mathfrak{S}_n , même supérieur à 1. Elle n'est pas définie pour $\mathfrak{S}_n = 1$, mais les limites à gauche et à droite de ce point sont identiques et ont pour valeur⁹ $(Nb + 1)$. Ainsi, la fonction est presque continue autour du point singulier, ce qui assure l'absence de tout problème de convergence au voisinage de ce point. La probabilité de tomber sur le point singulier lors des calculs numériques (arithmétique à virgule flottante) est infiniment faible. Toutefois, nous pouvons prévenir cette éventualité par un test conditionnel donnant la valeur $(Nb + 1)$ à M si \mathfrak{S}_n est exactement égal à 1,0. La figure 4.4 donne les représentations de $(M - 1)$ obtenues à partir des fonctions 4.11 et 4.13, en fonction de la tension appliquée normalisée par rapport à la tension de claquage. L'expression utilisée pour calculer \mathfrak{S}_n est issue de la résolution analytique de l'intégrale d'ionisation qui sera donnée dans la suite. L'accord entre les deux expressions est parfait pour les valeurs de $(M - 1)$ inférieures

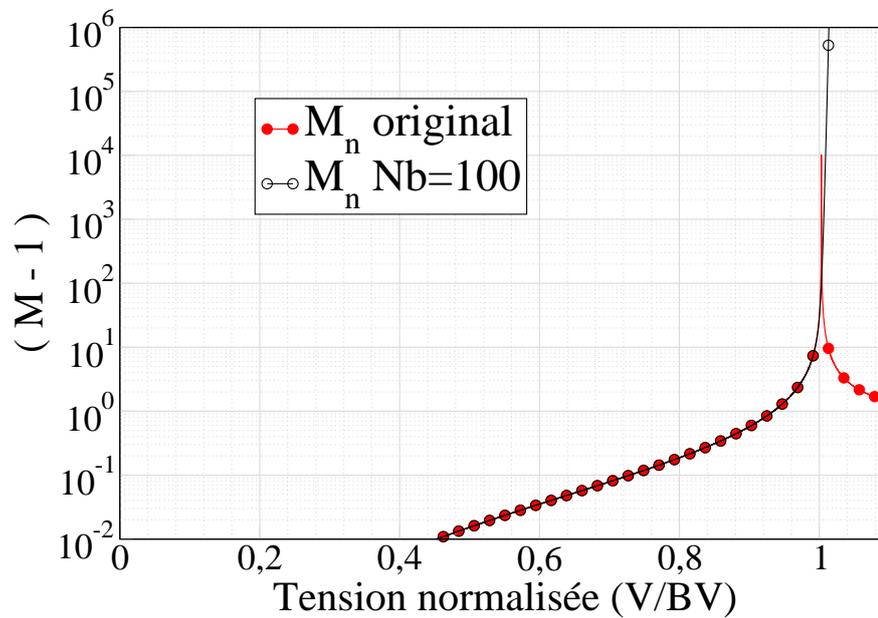


FIG. 4.4 – Variation de $(M - 1)$ en fonction de la tension appliquée normalisée par rapport à la tension de claquage. La courbe "original" est issue de l'équation 4.11 alors que la seconde courbe est obtenue à partir de l'équation 4.13 pour un nombre de terme $Nb = 100$.

⁹La somme des termes d'une suite géométrique de raison 1 est égale à $(Nb + 1)$.

à 10. L'utilisation de l'équation 4.13 n'a donc pas d'influence dans la région correspondant aux régimes où le composant est replié.

Le choix du nombre de termes (Nb) repose sur un compromis entre la précision et la bonne convergence des calculs au moment du claquage par avalanche. Plus Nb est élevé, plus la précision est grande mais plus la pente de la courbe au voisinage de la tension de claquage est importante et pénalise la convergence. Une valeur de 100 pour Nb assure un excellent compromis.

L'expression utilisée pour le calcul de l'intégrale d'ionisation \mathfrak{S}_n est particulièrement important pour déterminer avec précision la tension de maintien du transistor. Nous avons utilisé l'équation 2.29 établie au chapitre 2 et que nous rappelons ici :

$$\mathfrak{S}_n = \frac{A_p}{A_p - A_n} \left[\exp \left(2(A_p - A_n) \left(\frac{qC_{eff}}{\varepsilon_{Si}} \right)^3 V_{CB}^4 \right) - 1 \right] \quad (4.14)$$

avec V_{CB} la tension entre le collecteur et la base du modèle SPICE du transistor. Cette équation, obtenue par une approche purement analytique et physique, constitue la plus précise des expressions de l'intégrale d'ionisation des électrons dans une jonction plane abrupte. Bien que le claquage des jonctions soit généralement initialisé dans une jonction cylindrique, la jonction collecteur-base qui est le siège de la génération par avalanche à fort courant peut être assimilée à une jonction plane [61].

4.3.3 Variation du coefficient de multiplication avec la densité de courant

Nous avons démontré que la concentration effective C_{eff} de la jonction collecteur-base, est une fonction du courant qui circule au travers du collecteur. La prise en compte de la variation de C_{eff} va permettre de modéliser la variation de la tension de claquage de la jonction collecteur-base en fonction du courant dans le composant.

L'effet de la densité des porteurs libres dans la jonction collecteur-base devient significatif lorsque la densité de courant atteint des valeurs proches de la densité de courant critique. La densité n d'électrons dans la région de fort champ électrique du collecteur est donnée par la relation :

$$n = \frac{I_C^{int}}{S q v_{ln}} \quad (4.15)$$

avec S la section totale offerte au passage du courant, q la charge élémentaire et v_{ln} la vitesse limite des électrons à fort champ électrique. En effet, le courant d'électrons à l'entrée de cette région du côté de la base, correspond au courant circulant dans le transistor NPN du macromodèle (I_C^{int} , Fig. 4.3). La densité d'électrons n'est cependant pas constante sur l'ensemble de la région. Elle varie d'une quantité approximativement égale à celle des trous générés par avalanche. Or, nous avons montré que la densité du courant de trous, ne peut dépasser le tiers du courant total circulant dans le composant. Aussi, nous avons choisi d'approximer la densité d'électrons par une densité constante sur toute la région électrique de collecteur. Ce choix permet de calculer facilement l'impact de la densité de porteurs sur la tension de claquage et ne pénalise pas trop la précision des calculs car l'aspect essentiel est d'obtenir l'ordre de grandeur de la densité d'électrons en fonction la densité de courant.

Trois régimes de fonctionnement, correspondants au calcul de trois tensions de claquage, peuvent être distingués en fonction de la densité de porteurs libres. Ces trois régimes sont représentés sur la figure 4.5 en termes de distribution du champ électrique en fonction de la densité de porteurs libres.

1. Figure 4.5(a). Pour une densité d'électrons plus faible que le dopage N_d de la région faiblement dopée de collecteur, la tension de claquage de la jonction collecteur-base est une fonction croissante de la densité de porteurs. La densité des électrons vient en effet se soustraire à la densité de charge d'espace dans la région faiblement dopée. Tout se passe comme si le dopage de cette région diminuait, ce qui correspond à une augmentation de la tension de claquage.

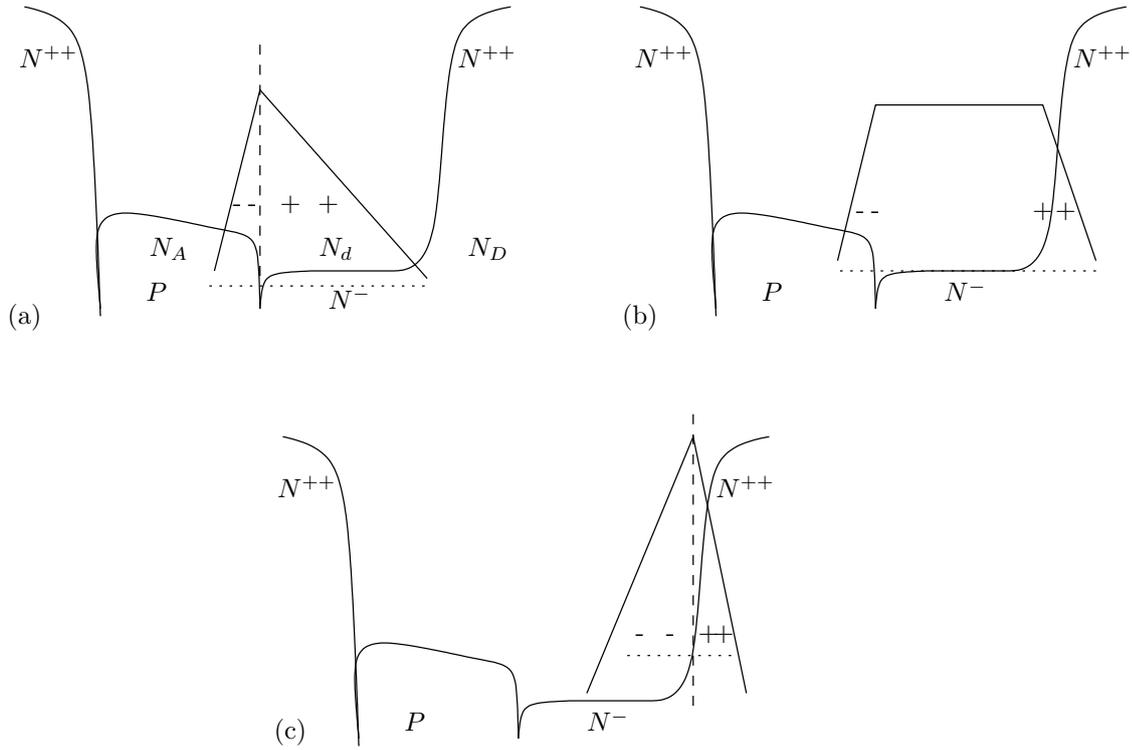


FIG. 4.5 – Représentation des profils de champs électriques pour les trois régimes de calcul de la tension de claquage de la jonction collecteur-base. La densité de charge (négative) due au courant d'électron est représentée en traits pointillés courts dans la zone de charge d'espace. La position de la jonction (lieu de changement de signe de la charge d'espace) est représentée en traits pointillés longs.

2. Figure 4.5(b). L'extension de la charge d'espace dans le collecteur faiblement dopé est limitée par la couche fortement dopée. L'accroissement de la tension de claquage est donc limité. Au maximum, cette tension est égale à la tension de claquage de la diode PIN (PN^-N^{++}). Ce régime sera atteint pour des densités d'électrons de l'ordre de N_d .
3. Figure 4.5(c). Lorsque la densité des électrons est supérieure à N_d , la nature de la jonction collecteur-base est modifiée car elle détermine maintenant le dopage effectif de la jonction. La tension de claquage de la jonction effective formée à la frontière des régions faiblement (N_d) et fortement (N_D) dopées de collecteur est alors une fonction décroissante de la densité de porteurs libres.

Pour le premier régime de fonctionnement (Fig. 4.5(a)), la concentration effective de la jonction collecteur-base est donnée par la relation :

$$C_{eff}^{(1)} = \frac{(N_A + n)(N_d - n)}{N_A + N_d} \quad (4.16)$$

avec N_A le dopage de la diffusion de base et N_d le dopage de la région faiblement dopée du collecteur. Cette fonction est décroissante pour n croissant et devient négative pour $n > N_d$.

La tension de claquage maximale que peut atteindre la jonction collecteur-base correspond à la tension de claquage de la diode PIN (BV_{PIN}). La valeur de la concentration effective correspondant à ce second régime est obtenue en inversant la formule de Gharbi [140] :

$$C_{eff}^{(2)} = \left(\frac{5,65 \cdot 10^{13}}{BV_{PIN}} \right)^{\frac{4}{3}} \quad (4.17)$$

Enfin, dans le troisième régime (Fig. 4.5(c)), la concentration effective est donnée par :

$$C_{eff}^{(3)} = \frac{(n - N_d)(N_D - n)}{N_D - N_d} \quad (4.18)$$

qui est une fonction croissante pour n croissant et négative tant que $n < N_d$ et où N_D est le dopage de la région fortement dopée du collecteur. On peut également, dans le cas du transistor à collecteur graduel, calculer la tension de claquage de la jonction linéaire associée aux forts courants (cf. section 3.5). Pour cela, quelques manipulations algébriques permettent à partir de l'expression analytique du profil de dopage gaussien, de calculer la valeur de la pente du profil de dopage au point où la valeur du dopage est égal à la densité de porteurs libres n . L'expression 3.2 donnée dans le chapitre précédent permet alors de calculer la tension de claquage en fonction de la pente. Enfin, l'utilisation de la formule de Gharbi donne une valeur pour C_{eff} qui sera utilisée pour calculer l'intégrale d'ionisation.

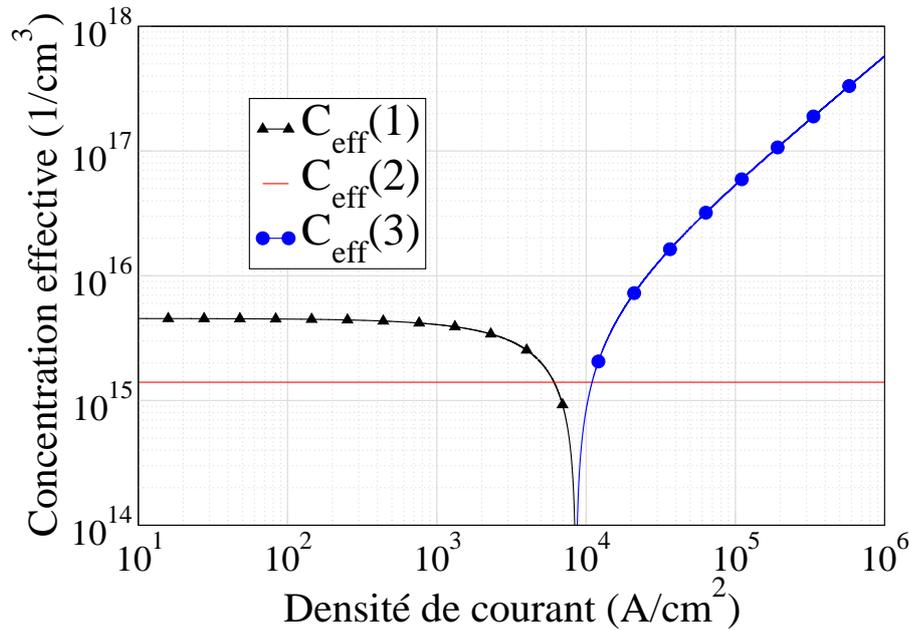


FIG. 4.6 – Variation des trois concentrations effectives ($C_{eff}^{(1)}$, $C_{eff}^{(2)}$, $C_{eff}^{(3)}$) en fonction de la densité de courant ($J = nqv_{in}$), pour $N_A = 5 \cdot 10^{16}$, $N_d = 5 \cdot 10^{15}$, $N_D = 1 \cdot 10^{20}$ et $BV_{PIN} = 150 \text{ V}$.

L'algorithme permettant de sélectionner, parmi les trois valeurs disponibles, la concentration effective à utiliser pour le calcul de l'intégrale d'ionisation est particulièrement simple. Il consiste simplement à donner à C_{eff} la valeur la plus grande des trois calculées $\text{MAX}(C_{eff}^{(1)}, C_{eff}^{(2)}, C_{eff}^{(3)})$. La figure 4.6 donne la variation de ces concentrations en fonction de la densité de courant qui est directement proportionnelle à n ($n = \frac{J}{q v_{in}}$). Pour n faible, à faible courant, la plus grande valeur

de C_{eff} correspond à $C_{eff}^{(1)}$. Lorsque n augmente, pour un courant croissant dans la structure, la valeur de $C_{eff}^{(1)}$ décroît jusqu'à devenir inférieure à $C_{eff}^{(2)}$ qui est indépendante de n . $C_{eff}^{(3)}$ ne prend des valeurs importantes que lorsque n est grand vis-à-vis de N_d , soit en terme de densité de courant $J > J_{crit} = N_d q v_{in}$, et ne sera donc finalement utilisée qu'aux fortes densités de courant.

Cette approche permet donc de prendre en compte les modifications des caractéristiques électriques de la jonction collecteur-base d'une façon relativement simple et efficace. Elle est étroitement liée aux caractéristiques physiques du composant et offre donc la possibilité d'analyser l'impact des modifications de technologie et de dessin du composant. La section 4.5 donne

un exemple d'utilisation de cette approche pour modéliser le composant optimisé développé dans la technologie 3.

4.4 Remarques pour la modélisation de transistors NMOS

De la même manière que pour les transistors NPN à collecteur faiblement dopé, la modélisation rigoureuse des transistors NPN parasites associés aux transistor NMOS doit prendre en compte la variation de la tension de claquage de la jonction drain-substrat (ou collecteur-base) avec le niveau de courant.

Le schéma type de macromodèle est représenté dans la figure 4.7(a). Il est similaire à celui

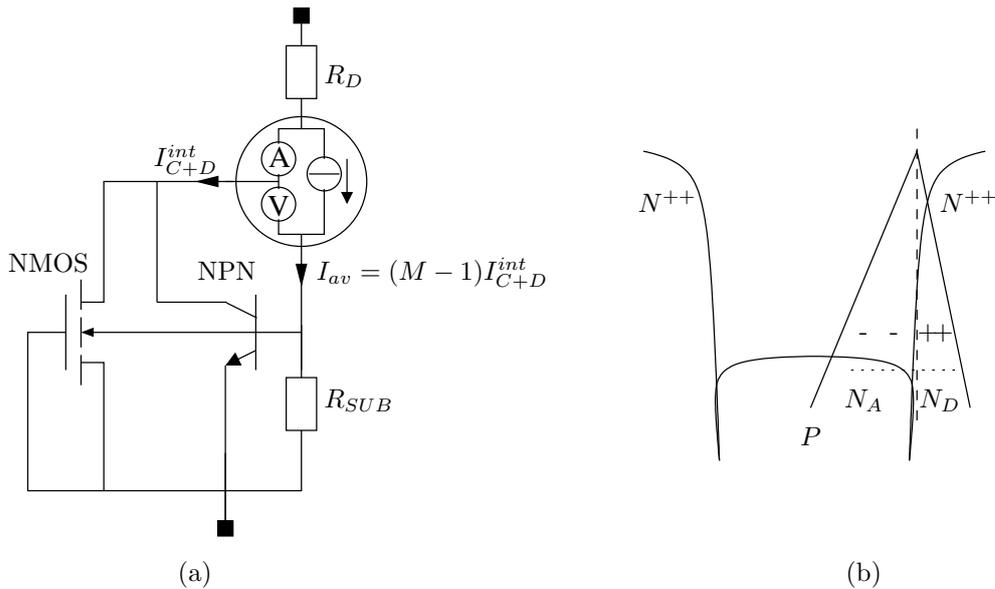


FIG. 4.7 – (a) Schéma électrique équivalent du transistor NMOS utilisé pour la modélisation. (b) Représentation de la répartition de champ électrique dans un transistor NMOS, pour le calcul de la tension de claquage de la jonction collecteur-base (drain-substrat). La densité de charge (négative) due au courant d'électron est représentée en traits pointillés courts dans la zone de charge d'espace. La position de la jonction est représentée en traits pointillés longs.

utilisé pour le transistor NPN à l'exception de l'élément NMOS dont le modèle est issu de la bibliothèque de la technologie et permet de prendre en compte l'effet du transistor MOS.

Le calcul du facteur de multiplication par avalanche est réalisé de la même manière que dans le cas des transistors à collecteur faiblement dopé. Seule l'expression de C_{eff} , qui dans ce cas est unique, doit être adaptée pour les transistors NMOS.

Sa valeur se déduit directement de la figure 4.7(b) :

$$C_{eff} = \frac{(N_A + n)(N_D - n)}{N_A + N_D} \quad (4.19)$$

L'extraction des paramètres utiles au modèle est sensiblement identique à celle d'un transistor bipolaire à collecteur faiblement dopé présenté dans la section suivante. La valeur de la résistance de base est en revanche déduite de la caractéristique TLP au moyen de la relation 4.3.

4.5 Modélisation du TBA de la technologie 3

La coupe technologique du transistor NPN issu de la technologie 3 est rappelée dans la figure 4.8. Le modèle présenté à la section 4.3.1 est utilisé pour la modélisation. La prise en

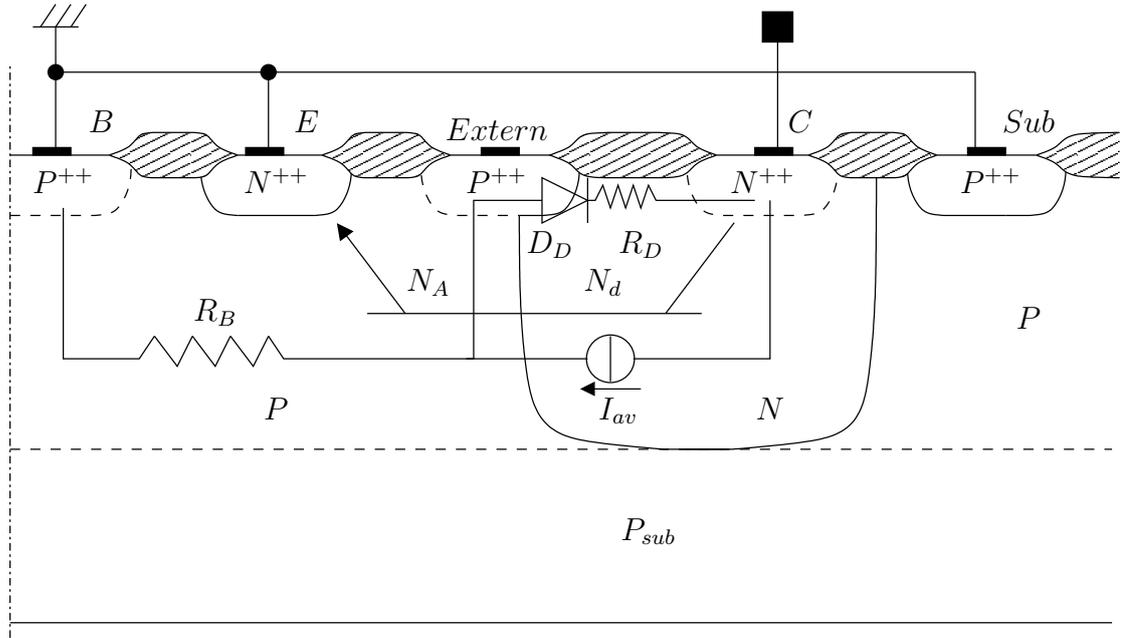


FIG. 4.8 – Coupe technologique du transistor NPN réalisé dans la technologie 3. Les principaux composants à modéliser sont représentés.

compte de la diode D_D qui détermine la tension de claquage statique du composant, implique une légère modification.

4.5.1 Modélisation de la diode D_D

Le courant qui circule dans la diode D_D au delà de sa tension de claquage BV_D est déterminé par sa résistance série R_D (Fig. 4.8). Cette résistance faiblement dopée possède une caractéristique non linéaire qui conduit à une saturation du courant à une valeur critique I_{crit} au delà d'une certaine tension appliquée. Nous avons adapté l'expression proposée dans [141] pour ce type de résistance faiblement dopée, afin de modéliser le courant circulant dans la diode en fonction de la tension V à ses bornes :

$$I_D = \frac{V - BV_D}{R_d \sqrt{1 + \left(\frac{V - BV_D}{R_d I_{crit}} \right)^2}} \quad (4.20)$$

pour $V > BV_D$ et $I_D \approx 0$ pour $V < BV_D$. Sa caractéristique électrique $I_D(V)$ est linéaire pour les valeurs de V proche de BV_D . Dans cette région, la valeur de la résistance est R_d .

Le courant issu de la multiplication par avalanche dans le collecteur et celui circulant dans la diode sont ajoutés pour donner I_{av} dans le modèle. Dans les résultats de simulations que nous présentons ici, les calculs sont réalisés par un module VerilogA inséré en tant que source de courant dans le macromodèle.

4.5.2 Extraction des paramètres statiques

Le modèle du transistor bipolaire utilisé ici n'est pas disponible dans la bibliothèque de la technologie. En effet, ce transistor n'est pas un composant standard de la technologie. Nous en profitons pour décrire la méthode d'extraction de l'ensemble des paramètres statiques. La caractérisation est réalisée à partir de composants dont toutes les broches (base, émetteur,

collecteur, substrat et Extern¹⁰ (Fig. 4.8)) sont indépendantes. Les mesures sont effectuées grâce à un analyseur de paramètres de type HP4142.

Paramètres de la diode D_D

La caractéristique de la diode polarisée en inverse est obtenue en appliquant une tension entre les broches de collecteur et Extern. La courbe donnant le courant dans la diode en fonction de la tension est présentée sur la figure 4.9.

La tension de claquage BV_D de la diode, sa résistance série R_d à faible tension et le courant maximum I_{crit} sont extraits directement de cette courbe.

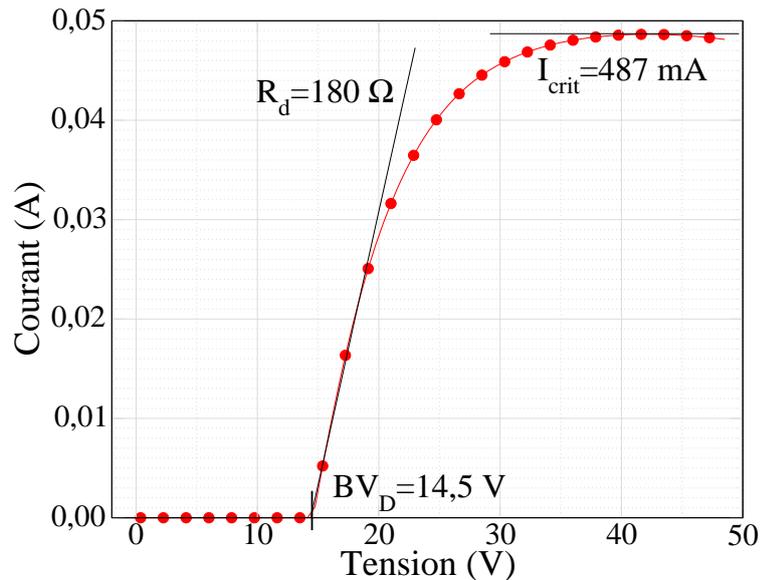


FIG. 4.9 – Caractéristique statique en inverse de la diode D_D associée au transistor NPN (technologie 3) et extraction des paramètres électriques du modèle.

Résistance de base

La valeur de la résistance de base est extraite de la mesure du courant en fonction de la tension appliquée entre la base et le contact Extern. Cette caractéristique est représentée sur la figure 4.10.

La valeur de la résistance correspond à l'inverse de la pente de la courbe obtenue.

Paramètres du transistor

Les paramètres nécessaires à la modélisation du transistor NPN sont les suivants :

- I_S , le courant de saturation du transistor
- I_{SE} , le courant de saturation de la jonction émetteur-base
- n_{EL} , le facteur de non-idéalité de la jonction émetteur-base
- β_{max} , le gain en courant
- IL , la valeur du courant de collecteur à partir de laquelle le gain en courant du transistor atteint sa valeur de plateau
- IKF , la valeur du courant de collecteur à partir de laquelle le gain en courant du transistor chute en régime de forte injection

¹⁰Extern est le contact sur la diffusion P^{++} qui permet le déclenchement du transistor par un dispositif externe

Les tracés du courant de collecteur I_C et du courant de base I_B du transistor, en fonction de la tension base-émetteur V_{BE} , donnent accès à ces paramètres électriques élémentaires pour la simulation [127].

Le tracé du rapport I_C sur I_B , qui donne le gain du transistor, en fonction du courant de collecteur permet d'extraire les paramètres IL et IKF (Fig. 4.11).

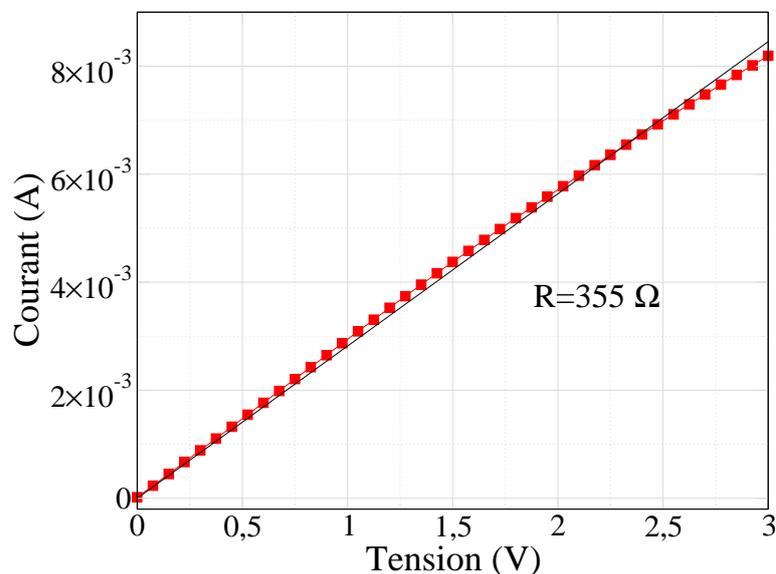


FIG. 4.10 – Caractéristique statique de la résistance de base du transistor NPN (technologie 3). Extraction de la valeur de la résistance.

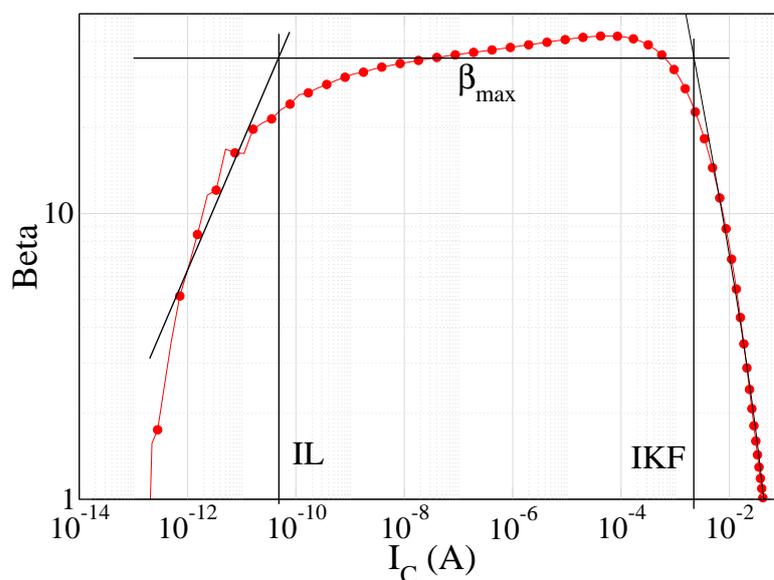


FIG. 4.11 – Gain en courant du transistor bipolaire en fonction du courant de collecteur (technologie 3). Extraction des paramètres de fort et faible courant relatifs à β .

Résistance de collecteur

À forts niveaux de courant, l'augmentation de la tension aux bornes du composant pour des courants croissants est principalement liée aux résistances de collecteur et d'émetteur. Elles correspondent aux résistances de diffusion et des prises de contact. La valeur de la résistance à l'état passant du transistor (R_{ON}) est extraite de la caractéristique TLP (Fig. 4.12) et donne la valeur de la résistance de collecteur R_C du macromodèle.

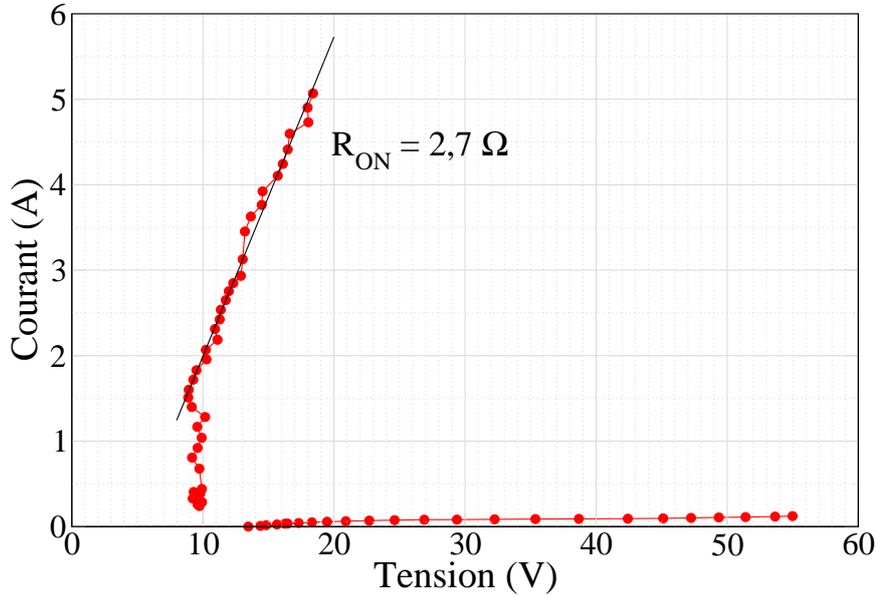


FIG. 4.12 – Caractéristique TLP fort courant du transistor NPN (technologie 3). Extraction de la résistance de collecteur (R_C) à fort courant.

4.5.3 Résultats de simulation, limitation du modèle

La plupart des paramètres du modèle ont été extraits à partir des caractéristiques expérimentales. Seules les valeurs du dopage de la base N_A , de la région de collecteur faiblement dopée N_d et de la section S du passage du courant dans le composant doivent être choisies en connaissance des paramètres physiques du composant. La valeur de la région fortement dopée de collecteur, qui ne joue pas de rôle important, est choisie égale à 10^{20} At/cm^3 . Enfin, la tension de claquage de la diode PIN peut être estimée à partir de la profondeur de la région faiblement dopée de collecteur par la formule 3.1. Toutefois, l'expression utilisée pour l'intégrale d'ionisation n'étant pas adaptée¹¹ au cas d'une diode PIN, ce paramètre pourra être réajusté pour retrouver la tension maximale de la caractéristique TLP avant le repliement.

Résultats

Le modèle a été implémenté et testé avec le simulateur SPICE SpectreS qui permet l'utilisation du langage VerilogA pour décrire la source de courant d'avalanche. Les résultats des simulations statiques réalisées avec ce macromodèle après un calibrage des paramètres (en particulier N_A , N_D , S et BV_{PIN}), dans une limite de 10% des valeurs extraites expérimentalement, permet d'obtenir une excellente corrélation de la simulation avec la mesure TLP avant le repliement de la tension de collecteur (Simu1, Fig. 4.13). La tension de maintien est en revanche

¹¹L'expression de l'intégrale d'ionisation est obtenue dans l'hypothèse d'une jonction plane abrupte sans limitation de charge d'espace.

largement surestimée, de plus de 10 V. La limitation du modèle provient de la prise en compte de la chute du gain du transistor bipolaire NPN à forte injection. En effet, à fort courant la valeur du gain calculée par le modèle varie comme l'inverse du courant de collecteur et tend donc vers zéro. Or, nous avons montré dans le chapitre 2 que le rapport des courants d'électrons et de trous dans la base du transistor est proche de 2. Pour les régimes de très fort courant spécifiques aux TBA durant un ESD, le gain du transistor bipolaire est donc de l'ordre de 2. Le résultat de la simulation statique, après modification du modèle du transistor bipolaire de façon à ce que le gain du transistor soit égal à 2 quelque soit le courant, est représenté par la courbe Simu2 dans la figure 4.13. L'accord avec la mesure est alors parfait à fort courant mais mauvais à plus faible courant. L'ajustement des paramètres de la diode D_D et de BV_{PIN} permet cependant d'obtenir une bonne représentation de la réponse du composant (Simu3, Fig. 4.13) mais s'éloigne d'une représentation physique précise.

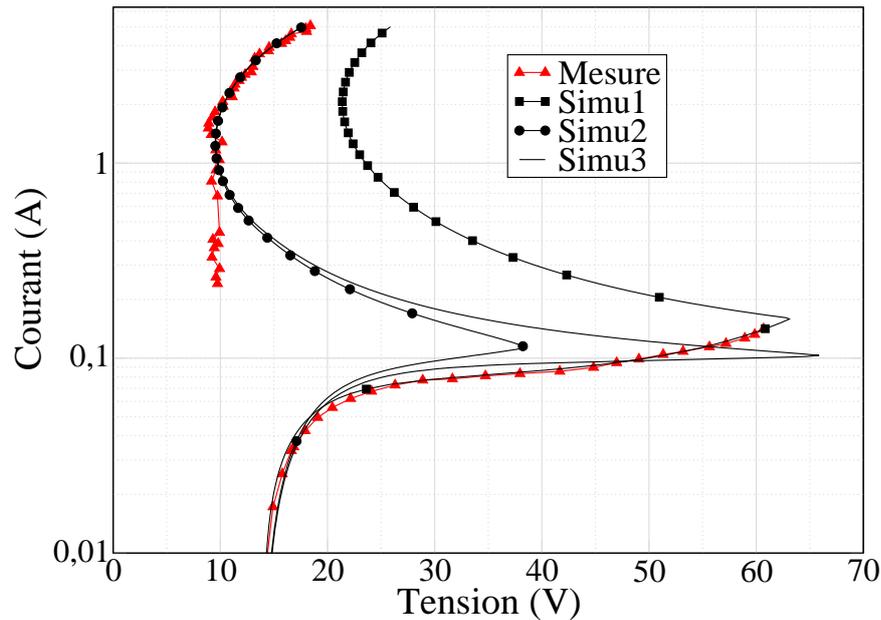


FIG. 4.13 – Comparaison des résultats de simulation et des résultats expérimentaux pour la caractéristique quasi-statique TLP du transistor NPN optimisé (technologie 3). La courbe Simu1 est obtenu en prenant en compte la chute du gain et la variation du facteur de multiplication à fort courant. Simu2 est obtenu dans les même conditions que Simu1 en éliminant la chute du gain à fort courant et en le prenant égal à 2. Pour Simu3 les conditions sont identiques à Simu2 ($\beta=2$, M variable), seuls les paramètres sont modifiés pour ajuster la caractéristique simulée aux mesures.

L'utilisation du paramètre empirique k_1 (Equ. 4.10) pour compenser cet effet n'est pas possible ici. Il influe en effet à la fois à faible et fort niveaux de courant. Pour l'ensemble des résultats de simulation présentés, sa valeur est égale à 1.

Limitations

Une solution consisterait à modifier le code informatique du modèle du transistor bipolaire de façon à limiter la chute du gain à fort courant à une valeur minimale de 2. L'ensemble des paramètres du modèle pourrait alors être conservé et cette solution n'aurait pas d'impact sur les caractéristiques à faible niveau de courant. La modélisation du composant parasite que représente la structure ne serait pas non plus affectée. Toutefois, ceci n'est réalisable que si le code source du modèle est disponible et demande un travail assez lourd pour l'intégration dans

l'environnement de simulation.

La formulation de la chute du gain à fort courant comme une fonction de $\frac{1}{I_C}$ est bien connue. Nous avons démontré que le gain du transistor est proche de 2 dans un TBA pour de forte densité de courant. L'évolution du gain entre ces deux régimes reste à étudier.

Enfin, on remarque sur la figure 4.13 que la simulation ne correspond pas à la mesure pour des courants supérieurs à 200 mA et inférieurs à 1 A. Dans cette plage, la tension varie peu aux bornes du composant réel. Ce régime correspond à l'étalement progressif du courant dans la totalité du composant. Or, nous avons choisi une section S de passage du courant constante dans le modèle du transistor. Afin, de prendre en compte la non uniformité du courant, il serait nécessaire de faire varier S en fonction du courant total. Rappelons que la non-uniformité du courant n'est effective qu'au moment du repliement de la tension et jusqu'à un certain niveau où le courant est finalement redistribué dans tout le composant. C'est donc seulement dans une plage de courant restreinte qu'il serait nécessaire de recalculer S de manière à ce que la densité de courant dans le composant soit égale à la densité maximum (définie dans la section 2.5 du chapitre 2). Comme nous l'avons vu, cette dernière dépend principalement de la résistance de ballast dans le collecteur du transistor bipolaire. Toutefois, hormis la précision du modèle, l'intérêt de prendre en compte ce phénomène est limité. Le régime considéré n'est pas critique pour analyser le bon fonctionnement d'une stratégie de protection ESD. De plus, si cet effet est modélisé, la chute de la tension au moment du repliement sera beaucoup plus abrupte et rapide et défavorisera la bonne convergence des simulations.

4.6 Étude de la dynamique

Une décharge électrostatique est une impulsion de courant de forte intensité et d'une durée totale extrêmement courte. Dans le cas d'un stress ESD de type CDM, les aspects dynamiques revêtent une importance primordiale, le temps de réponse des dispositifs de protection étant proche de l'ordre de grandeur de la durée de l'impulsion de courant. De plus, l'amplitude du pic de courant lors de ce type de décharge est plus important que pour un stress de type HBM. L'association des effets de très forte densité de courant et de la dynamique extrêmement rapide laisse supposer que la modélisation des composants pour les stress de type CDM sera difficile avec les modèles compacts classiques. Aussi, nous limiterons ici notre réflexion au cas plus classique des stress de type HBM.

Comme nous l'avons fait remarquer à la section 4.2, la réponse en inverse des TBA peut être décomposée en deux parties : l'une correspondant au déclenchement du composant, où le courant est relativement faible et la variation de tension aux bornes du composant est très rapide, l'autre où la tension varie beaucoup plus lentement et le courant est très grand dans le composant. Les aspects dynamiques sont donc beaucoup plus importants en début d'impulsion HBM, et vont jouer un rôle primordial lors du déclenchement du composant.

4.6.1 Dynamique du déclenchement

Dans les premiers instants d'une décharge ESD, le courant circulant dans le transistor est de nature purement capacitive [71, 17]. Il est principalement lié à la charge de la capacité collecteur-base dans un transistor bipolaire, et aux capacités drain-substrat et drain-grille dans un transistor MOS. La condition de déclenchement du transistor bipolaire correspond à une tension émetteur-base V_{BE} supérieure à sa tension de seuil V_s en direct $V_{BE} > V_s$.

Cette condition est remplie dès que le courant dans la résistance de base est suffisant. Ce courant peut avoir deux origines différentes :

- le courant d'avalanche dans la jonction collecteur-base (polarisée en inverse)
- ou le courant capacitif de la jonction collecteur-base

Le courant d'avalanche sera prépondérant lors de phénomènes lents alors que le courant capacitif pourra devenir important pour des variations très rapides. Les courants capacitifs peuvent donc entraîner le déclenchement du transistor avant le claquage par avalanche de la jonction collecteur-base. Ce type de mise en route de la protection est souvent désigné comme un déclenchement par dv/dt [59]. Cet aspect peut fortement influencer les mesures de robustesse de certains composants de protection en fonction de l'équipement utilisé pour le test. En effet, chaque testeur peut générer des impulsions ESD dont les temps de montée sont sensiblement différents [142]. Le déclenchement plus précoce du transistor, lors de transitoires plus rapides, peut permettre dans certains cas de limiter la tension maximale aux bornes du composant et améliorer sa robustesse. Une étude plus détaillée du déclenchement par dv/dt est menée dans la section suivante.

Dès qu'une tension suffisante est atteinte pour polariser la diode émetteur-base en direct, le déclenchement du transistor bipolaire n'est cependant pas instantané. Son temps de réponse est de l'ordre du temps de transit τ_B des porteurs dans la base. Pour un transistor NPN, il est donné par la relation :

$$\tau_B = \frac{W_B^2}{m \frac{k_B T}{q} \mu_n} \quad (4.21)$$

où μ_n est la mobilité des électrons dans la base et m un paramètre qui dépend du niveau d'injection, avec $m = 2$ à faible injection et $m = 4$ à forte injection. Le temps de transit dans la base est donc un paramètre très important pour la modélisation de type SPICE. Le déclenchement du composant se situant dans des régimes intermédiaires de courant, sa valeur peut être choisie entre les valeurs de faible et de forte injection, et calculée grâce à la formule 4.21.

Pour les transistors NMOS dont la grille est liée à la masse, le fonctionnement du composant est identique. En revanche, le déclenchement du transistor bipolaire parasite peut être contrôlé par l'action du transistor NMOS. Pour permettre ce fonctionnement, le contact de grille doit être lié à la masse par une résistance. La tension de grille se voit ainsi couplée à la tension de drain par la capacité drain-grille intrinsèque et parfois par une capacité supplémentaire ajoutée entre drain et grille. Cette méthode de déclenchement du transistor est appelée couplage de grille. En début de décharge, l'élévation de la tension de grille permet de mettre en fonctionnement le transistor NMOS. Le courant d'électrons dans son canal va alors être multiplié par avalanche à la jonction drain-substrat. Le courant de trous généré par l'avalanche circule dans la résistance de base et permet de déclencher le transistor bipolaire avant le claquage par avalanche de la jonction drain-substrat. Une étude plus détaillée du couplage de grille est présentée dans la section 4.6.3.

4.6.2 Déclenchement par dv/dt

Le schéma électrique du macromodèle complet d'un transistor NPN est donné dans la figure 4.14(a). Les capacités non linéaires collecteur-base C_{CB} , émetteur-base C_{EB} et collecteur-substrat C_{CS} sont ajoutées dans le macromodèle statique pour prendre en compte les courants capacitifs dans le composant. Comme nous l'avons vu dans la section 4.3.1, ces capacités, en particulier C_{CB} , doivent être supprimées dans le modèle du transistor bipolaire car les courants capacitifs ne sont pas multipliés par avalanche.

Les conditions de déclenchement du transistor peuvent être étudiées à l'aide du schéma simplifié présenté dans la figure 4.14(b). En effet, avant le déclenchement, les courants ne peuvent emprunter que des chemins capacitifs, le transistor NPN étant initialement bloqué. Deux chemins principaux sont disponibles, l'un au travers de la capacité C_{CS} , l'autre étant représenté dans la figure 4.14(b). Le premier n'influe pas directement sur le déclenchement du transistor. En revanche, le second détermine le comportement du composant.

Afin de simplifier le problème analytique, nous allons considérer un courant I_{ESD} , circulant dans la branche de circuit (Fig. 4.14(b)), dont la variation est linéaire dans le temps :

$$I_{ESD} = at \quad (4.22)$$

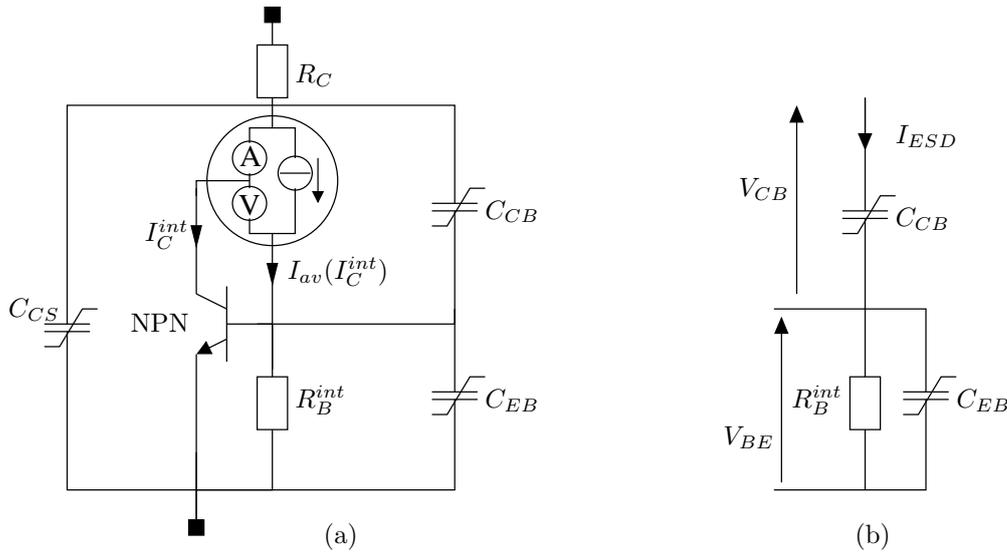


FIG. 4.14 – Schéma du macromodèle complet d'un TBA NPN (a). Circuit simplifié pour l'étude de la dynamique du déclenchement (b).

avec a en A/s la pente de la rampe de courant. Pour une décharge HBM typique de 2 kV, l'ordre de grandeur de a est d'environ $\frac{1,3}{10 \cdot 10^{-9}} = 1,3 \cdot 10^8$ A/s. Nous allons également considérer les capacités comme linéaires et constantes.

La tension aux bornes de la jonction collecteur-base V_{CB} se déduit ainsi facilement, sachant que le courant dans une capacité est donné par $i = C \frac{dv}{dt}$. L'intégration de cette équation en considérant que la capacité est déchargée à $t=0$ conduit à l'expression :

$$V_{CB} = \frac{a}{2C_{CB}} t^2 \quad (4.23)$$

Dans la partie basse de la branche de circuit considérée (Fig. 4.14(b)), le courant I_{ESD} se partage entre la résistance R_B^{int} et la capacité C_{EB} . L'expression de la tension V_{BE} peut être déterminée au moyen des transformations de Laplace.

$$V_{BE} = aR_B^{int} [t - \tau_e (1 - e^{-\frac{t}{\tau_e}})] \quad (4.24)$$

avec $\tau_e = R_B^{int} C_{EB}$.

Cette expression difficile à manipuler peut se simplifier dans les deux cas extrêmes où $t \gg \tau_e$ et $t \ll \tau_e$. Dans le premier cas, l'expression 4.24 se simplifie en :

$$V_{BE}^{(1)} = aR_B^{int} t \quad (4.25)$$

Ce qui correspond au cas où tout le courant I_{ESD} circule uniquement dans la résistance de base. Dans le second cas, l'utilisation du développement limité au second ordre de l'exponentielle permet de simplifier l'expression 4.24 :

$$V_{BE}^{(2)} = \frac{a}{2C_{EB}} t^2 \quad (4.26)$$

Ce qui correspond au cas où tout le courant circule uniquement dans la capacité E-B.

Le temps t_{dyn} d'établissement de la tension de seuil V_s de la jonction E-B sous l'effet des seuls courants capacitifs se déduit directement des équations 4.25 et 4.26 pour chacun des cas limites :

$$t_{dyn}^{(1)} = \frac{V_s}{aR_B^{int}} \quad \text{avec} \quad t_{dyn}^{(1)} \gg \tau_e \quad \text{et} \quad t_{dyn}^{(2)} = \sqrt{\frac{2C_{EB}}{a} V_s} \quad \text{avec} \quad t_{dyn}^{(2)} \ll \tau_e \quad (4.27)$$

Le premier cas de figure correspond à une résistance de base faible et une capacité E-B grande, et inversement dans le second cas.

Quel que soit le cas de figure, le déclenchement dynamique ne sera effectif qu'à condition que la tension collecteur-base n'ait pas atteint la tension de claquage BV_{CB} de cette jonction. Le temps d'établissement t_{BV} de cette tension se déduit de l'équation 4.23 :

$$t_{BV} = \sqrt{\frac{2C_{DB}}{a} BV_{CB}} \quad (4.28)$$

La condition pour observer un déclenchement dynamique est $t_{dyn} < t_{BV}$. Ceci se traduit dans les relations :

$$a > \left(\frac{V_s}{R_B^{int}} \right)^2 \frac{1}{2C_{CB}BV_{CB}} \quad (4.29)$$

et

$$\frac{C_{EB}}{C_{CB}} < \frac{BV_{CB}}{V_s} \quad (4.30)$$

pour le premier et second cas, respectivement.

La sensibilité au déclenchement dynamique dépend donc étroitement du composant étudié. Pour les transistors dans lesquels la résistance de base est faible vis-à-vis de l'impédance de la capacité E-B (premier cas), le déclenchement dynamique se produit seulement si la pente du transitoire en courant est supérieure à un seuil (Equ. 4.29). Dans le cas inverse (résistance de base forte), le déclenchement dynamique est indépendant du temps de montée du courant. En effet, le pont diviseur de tension constitué par les capacités E-B et C-B détermine seul la condition de déclenchement (Equ. 4.30). Si la capacité C_{EB} est suffisamment faible vis-à-vis de C_{CB} , la tension V_{BE} s'élève plus rapidement que la tension V_{CB} et atteint la tension de seuil V_s avant que V_{CB} n'ait atteint la tension de claquage.

Dans les deux cas, l'augmentation de la capacité collecteur-base favorise le déclenchement dynamique car elle retarde le déclenchement par le claquage de la jonction collecteur-base. D'une manière générale, plus l'impédance entre collecteur et base est faible vis-à-vis de l'impédance entre base et émetteur, plus le déclenchement dynamique est favorisé.

Le déclenchement dynamique n'est pas un problème simple à traiter. Le recours aux outils de simulation est ici indispensable pour étudier précisément le comportement d'un composant et prendre en compte la non-linéarité des capacités de jonction. L'analyse de ces phénomènes par la simulation physique bidimensionnelle est souvent difficile à interpréter car on ne peut pas discriminer l'origine des courants dans les électrodes du composant. En outre, les courants capacitifs sont distribués dans tout le composant. La modélisation de type SPICE doit donc prendre en compte les phénomènes distribués.

Prise en compte des effets distribués

Dans les transistors bipolaires verticaux, la répartition du courant, dans le cas d'un déclenchement lié au courant d'avalanche, n'est pas identique à celle liée aux courants capacitifs. La figure 4.15 représente une coupe d'un transistor bipolaire NPN vertical avec la répartition de ces courants. Alors que le courant d'avalanche I_{BV} est créé dans la partie cylindrique de la jonction collecteur-base, les courants capacitifs se distribuent sur l'ensemble de la jonction avec une densité de courant J_c relativement uniforme. Le courant I_{BV}^{seuil} nécessaire pour déclencher le transistor est donné par la tension de seuil de la jonction (V_s) divisé par la résistance totale R_B de la base (du point Y au point X Fig. 4.15).

Le calcul du courant capacitif nécessaire pour le déclenchement est plus complexe en raison des effets distribués. Le courant de base circule de la droite vers la gauche sur la figure 4.15. Ce

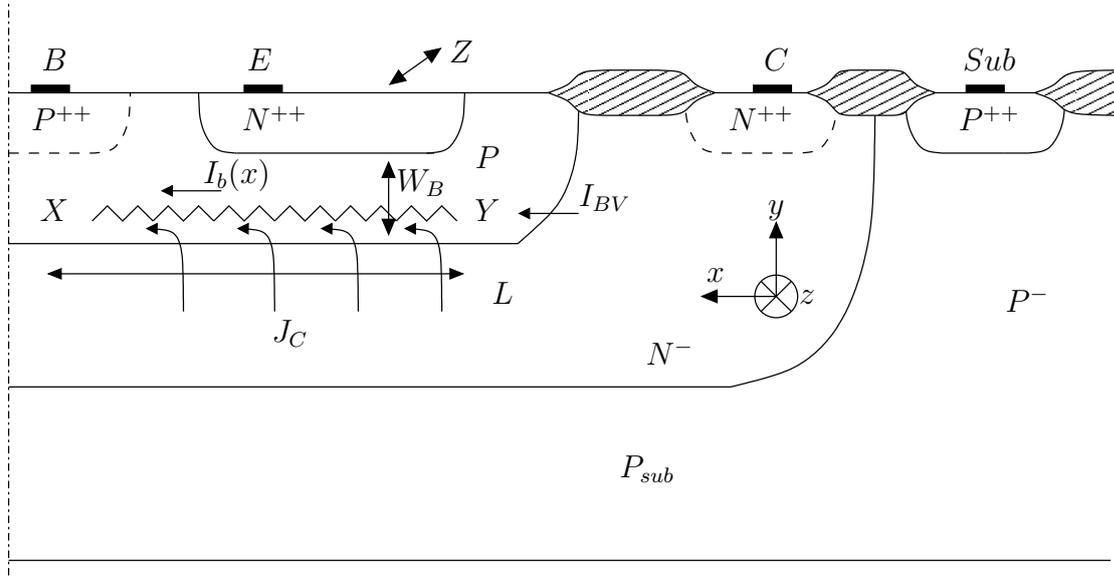


FIG. 4.15 – Représentation schématique de la distribution du courant capacitif et du courant d'avalanche, avant le déclenchement, dans la coupe technologique d'un transistor bipolaire vertical.

courant est invariant selon l'axe z . Si l'on note Z la longueur du doigt d'émetteur, la variation du courant dans la base selon l'axe x est :

$$dI_B = J_c Z dx \quad (4.31)$$

Si on considère la densité de courant capacitif dans le collecteur uniforme et si l'on place l'origine de l'axe x au point Y le courant dans la base est donné par :

$$I_B(x) = \int_0^x J_c Z dx' = J_c Z x \quad (4.32)$$

$$(4.33)$$

sachant que $I_B(Y) = 0$. La variation de tension dans la base est alors :

$$dV_B(x) = I_B(x) \frac{\rho_B}{W_B Z} dx \quad (4.34)$$

avec W_B la profondeur de la base et ρ_B sa résistivité. La tension dans la base au point Y est donc donnée par :

$$V_B(Y) = \int_X^Y dV_B(x) = \int_X^Y J_c Z \frac{\rho_B}{W_B Z} x dx = J_c \frac{\rho_B}{W_B} \frac{L^2}{2} \quad (4.35)$$

La condition de déclenchement est atteinte lorsque $V_B(Y)$ dépasse la tension de seuil V_s de la jonction émetteur-base. La densité de courant au seuil de déclenchement est donc :

$$J_{c_{seuil}} = \frac{2V_s}{\frac{\rho_B}{W_B} L^2} \quad (4.36)$$

Ce qui donne, en multipliant par la surface de la capacité collecteur-base (LZ), le courant de seuil du déclenchement :

$$I_{c_{seuil}} = \frac{2V_s}{R_B} \quad (4.37)$$

Cette dernière relation montre qu'il faut un courant capacitif deux fois supérieur au courant issu de l'avalanche à la jonction cylindrique collecteur-base, pour déclencher le transistor bipolaire. Ce résultat peut également être interprété comme si la résistance de base était réduite de moitié.

Afin, de prendre en compte cet aspect, le schéma du macromodèle doit être modifié (Fig. 4.16). La résistance de base est partagée en deux résistances dont la somme des valeurs reste égale à R_B . Les conditions de déclenchement par le courant d'avalanche restent donc inchangées. En revanche, la valeur de la résistance de base (xR_B) pour le déclenchement dynamique peut être facilement ajusté avec la valeur de x (avec $1 > x > 0,5$).

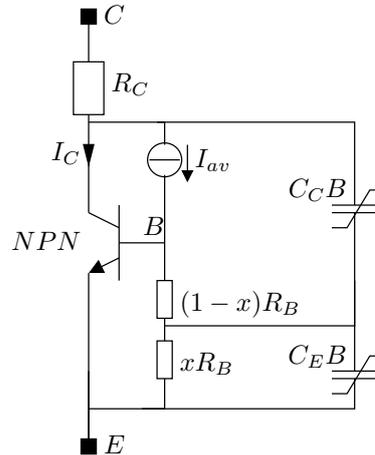


FIG. 4.16 – Modèle compact d'un transistor NPN autopolarisé avec ajustement de la résistance de base pour le déclenchement par dv/dt .

L'information sur la dynamique est difficile à obtenir expérimentalement car beaucoup d'éléments parasites jouent un rôle important lors des mesures. Les inductances et les capacités parasites liées par exemple aux fils d'interconnexions, au plot de contact, au boîtier, etc. . influent sur l'aspect du signal obtenu. Il faut minimiser ces éléments parasites ou les connaître avec précision pour obtenir une information fiable sur l'évolution de la tension et du courant dans le composant. Ceci représente un travail particulièrement difficile car un transitoire de courant ESD est associé à de très hautes fréquences et des phénomènes fortement non-linéaires.

4.6.3 Méthode d'optimisation du couplage de grille

Dans les transistors NMOS, le couplage de grille est utilisé pour diminuer la tension maximum aux bornes du composant, et éviter ainsi l'apparition de surtensions avant le déclenchement du composant. À notre connaissance, il n'existe aucune méthodologie pour choisir la valeur de la capacité de couplage et la valeur de la résistance entre la grille et la masse. Jusqu'à présent, ce choix semble reposer sur une approche très empirique fondée sur une série de réalisations et de tests. Nous proposons ici une méthode pour optimiser l'utilisation du couplage de grille afin de minimiser la surtension lors du déclenchement d'un transistor de protection NMOS.

Motivations

La suppression ou la réduction de la surtension lors du déclenchement d'un transistor de protection est un aspect important pour protéger efficacement les oxydes de grilles des transistors du cœur du circuit. L'oxyde de grille du transistor NMOS de protection est également soumis à cette surtension car sa grille est directement liée à la masse et son drain est au potentiel le plus élevé. Cet oxyde mince peut donc également être dégradé pendant la surtension avant le déclenchement.

Dans les régions proches de la jonction drain-substrat, le champ électrique dans l'oxyde de grille est élevé (Fig. 4.17). De plus, la jonction drain-substrat est le siège d'une forte génération par avalanche qui donne naissance à un grand nombre de porteurs chauds à proximité de l'oxyde de grille. Cette configuration est particulièrement défavorable et peut entraîner une dégradation

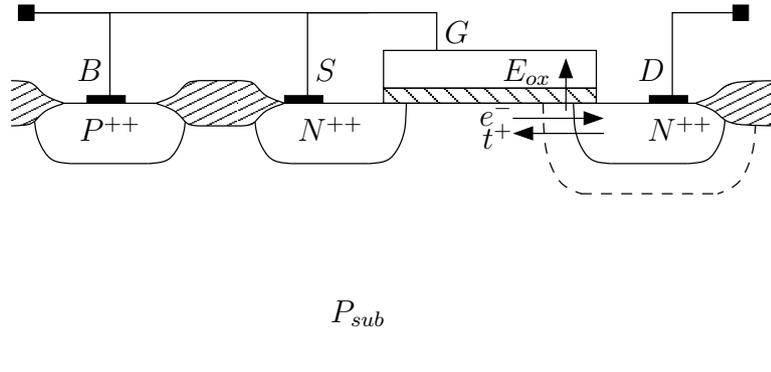


FIG. 4.17 – Coupe technologique schématique d'un transistor NMOS. Avant le déclenchement, la tension entre drain et grille est associée à un champ électrique E_{ox} important dans l'oxyde de grille qui est situé à proximité de la région de multiplication par avalanche des électrons e^- et des trous t^+ dans la jonction drain-substrat.

prématurée de l'oxyde de grille. Toutefois, le piégeage de charges dans l'oxyde n'est pas ici un aspect critique car une modification des caractéristiques électriques typiques du transistor NMOS est acceptable tant qu'elle n'affecte pas le courant de fuite de la protection.

L'utilisation du couplage de grille permet de diminuer le champ électrique dans l'oxyde. L'élévation de la tension de grille entraîne une diminution de la tension entre grille et drain V_{GD} . L'optimisation du couplage de grille doit donc également être menée dans l'objectif de minimiser V_{GD} au cours d'une décharge HBM.

Étude du déclenchement d'un transistor NMOS

La figure 4.18(a) représente le schéma équivalent d'un transistor NMOS de protection dans lequel les capacités et résistances sont séparées des modèles statiques du transistor MOS et du transistor bipolaire parasite.

Les transistors NMOS et NPN sont bloqués au début d'une décharge. L'étude de la réponse du composant en début de décharge peut donc être menée à partir du schéma de la figure 4.18(b) où les transistors MOS et bipolaire sont supprimés. Afin de simplifier ce schéma complexe, sans perte de généralité, nous pouvons négliger la résistance de la source et du drain car elles sont de faible valeur et les courants sont relativement faibles dans le régime étudié. En outre, la transformation du réseau en triangle, formé par les capacités $(C_{GB} + C_G^{ext})$, C_{GD} et C_{DB} , en étoile permet de se ramener au schéma présenté dans la figure 4.18(c). Le numérateur des expressions donnant les capacités équivalentes est :

$$CC = C_{GB}(C_{GD} + C_G^{ext}) + C_{GB}C_{DB} + (C_{GD} + C_G^{ext})C_{DB} \quad (4.38)$$

On peut identifier dans ce dernier schéma deux branches, Br^B et Br^G , similaires à celle étudiée pour le déclenchement par dv/dt d'un transistor bipolaire dans la section 4.6.2.

Comme dans le cas du transistor NPN simple, le déclenchement du transistor bipolaire parasite peut être lié soit au claquage par avalanche de la jonction drain-substrat soit à un courant capacitif suffisant dans la résistance de base. Une nouvelle possibilité de déclenchement s'ajoute ici, la mise en route du transistor NMOS par le couplage de grille. Trois conditions de déclenchement sont donc possibles (Fig. 4.18)(c) :

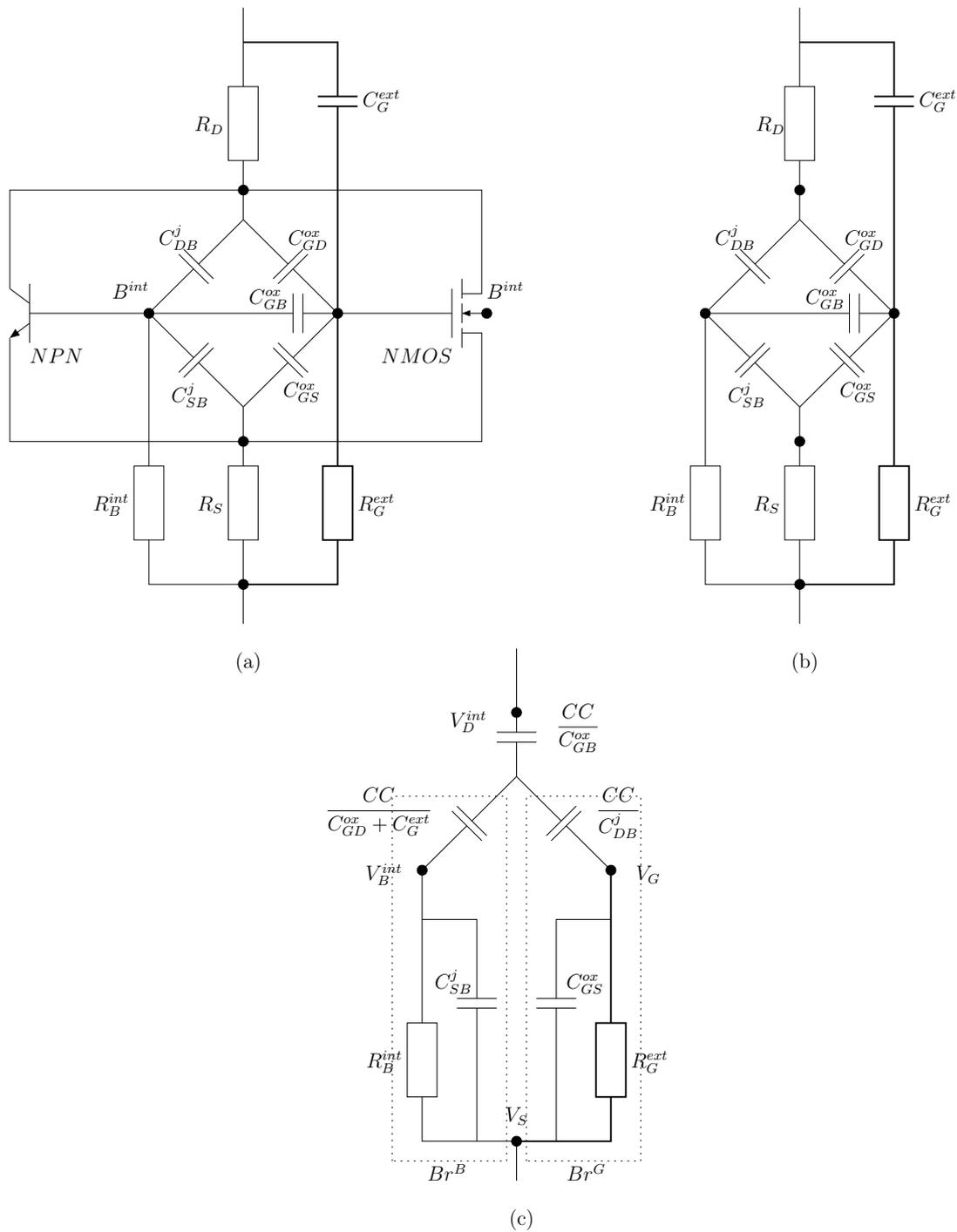


FIG. 4.18 – Schéma équivalent d'un transistor MOS. Les exposants j et ox donnent la nature des capacités (jonction ou oxyde respectivement). La capacité C_{GB}^{ext} et la résistance R_G^{ext} sont externes au transistor et placées sur la puce à proximité immédiate du transistor NMOS.

- $V_D^{int} - V_B^{int} > BV_{DB}$
- $V_B^{int} > V_s$
- $V_G > V_t$

où BV_{DB} est la tension de claquage de la diode drain-substrat, V_s la tension de seuil de la diode émetteur-substrat et V_t la tension de seuil du transistor NMOS.

Le courant d'une décharge se partage entre les deux branches de circuit. Il empruntera

principalement la branche offrant l'impédance la plus faible. Si la capacité C_{GD}^{ext} n'est pas utilisée (déconnectée), la branche Br^B est la plus favorable au passage du courant car les capacités dans cette branche ont de plus fortes valeurs, résultant donc en une plus faible impédance. En effet, les valeurs des capacités de jonction C_{DB} et C_{SB} sont d'au moins un ordre de grandeur supérieures aux capacités de recouvrement de l'oxyde de grille sur le drain et la source C_{GD} et C_{GS} . De plus, la valeur de la résistance de grille externe doit au moins être de l'ordre de grandeur de la résistance de base, afin que la tension sur la grille soit maintenue durant un temps suffisamment long et que l'augmentation de la tension de grille soit suffisante.

L'utilisation d'une simple résistance de grille sans capacité externe ajoutée C_G^{ext} ne peut donc permettre d'obtenir une diminution de la tension de déclenchement, quel que soit la valeur de R_G^{ext} . Il apparaît donc indispensable de rééquilibrer le courant dans les deux branches en ajoutant une capacité externe entre grille et drain. De cette façon, l'équilibre entre les deux branches est retrouvé dès que la capacité ajoutée en externe C_G^{ext} est de l'ordre de grandeur de la capacité drain-substrat C_{DB} . L'ajout de la capacité externe est également favorable pour le déclenchement dynamique car, au travers de l'augmentation de l'ensemble des capacités liées du terme CC (Equ. 4.38), l'élévation de la tension entre drain et substrat va être réduite et retarder le claquage par avalanche de la jonction drain-substrat.

Pour que le déclenchement par couplage de grille soit possible, il faut donc que la capacité ajoutée en externe soit du même ordre de grandeur ou supérieure à la capacité drain-substrat. Pour la même raison, la résistance de grille doit être de plus faible valeur que la résistance de base afin de diminuer l'impédance de la branche Br^G .

La vitesse à laquelle le transistor MOS doit être coupé après le déclenchement du transistor bipolaire, permet d'affiner le choix de la résistance et de la capacité à utiliser. Le courant dans le canal du transistor MOS doit être coupé avant l'apparition du régime de très fort courant. Pour cela, la constante de temps du circuit RC de grille $\tau_G = R_G^{ext} C_G^{ext}$ doit être inférieure à la nano-seconde. Cette dernière contrainte fixe donc le produit $R_G^{ext} C_G^{ext}$ et lie les valeurs de R_G^{ext} et C_G^{ext} . Les valeurs optimales sont donc obtenues pour la plus faible valeur de C_G^{ext} telle que les deux conditions $C_G^{ext} > C_{DB}$ et $R_G^{ext} < R_B$ soient satisfaites.

Application

Une étude du déclenchement des transistors de protection NMOS réalisés dans la technologie 3 (décrite dans le chapitre précédent) a permis de valider cette méthode. Les résultats de simulations SPICE d'un transistor NMOS soumis à un stress HBM de 2 kV pour différentes configurations de couplage de grille sont présentés dans la figure 4.19. Ils confirment les résultats du développement théorique précédent. Si aucune capacité n'est ajoutée en externe entre drain et grille, le pic de tension avant le déclenchement persiste. La capacité de couplage est très faible dans ce cas et même pour une forte valeur de la résistance de grille (ici 140 k Ω), la tension V_G augmente peu en début de décharge et ne permet pas de déclencher le transistor NMOS.

L'ajout d'une capacité externe et l'application des règles établies précédemment, permettent d'éliminer totalement la surtension. De plus, la tension maximale aux bornes de l'oxyde de grille, donnée par $V_D^{int} - V_G$ est également très fortement diminuée. La polarisation de la grille est rapidement coupée avant d'atteindre le régime de fort courant. Seul le TBA parasite assure donc la conduction des fortes valeurs de courant. Remarquons également la nette réduction de la vitesse du transitoire de tension sur le drain due à l'augmentation de la capacité équivalente entre drain et substrat.

Finalement, l'utilisation des simulations SPICE permet d'affiner le choix de la capacité et de la résistance de couplage de grille pour optimiser la réponse du composant.

Ces solutions ont été testées en pratique sur les transistor NMOS de la technologie 3. Les versions des composants dont le déclenchement par couplage de grille a été optimisé, voient leur robustesse ESD doublée vis-à-vis de composants dont la grille est reliée directement ou par une simple résistance à la masse, sans capacité externe supplémentaire.

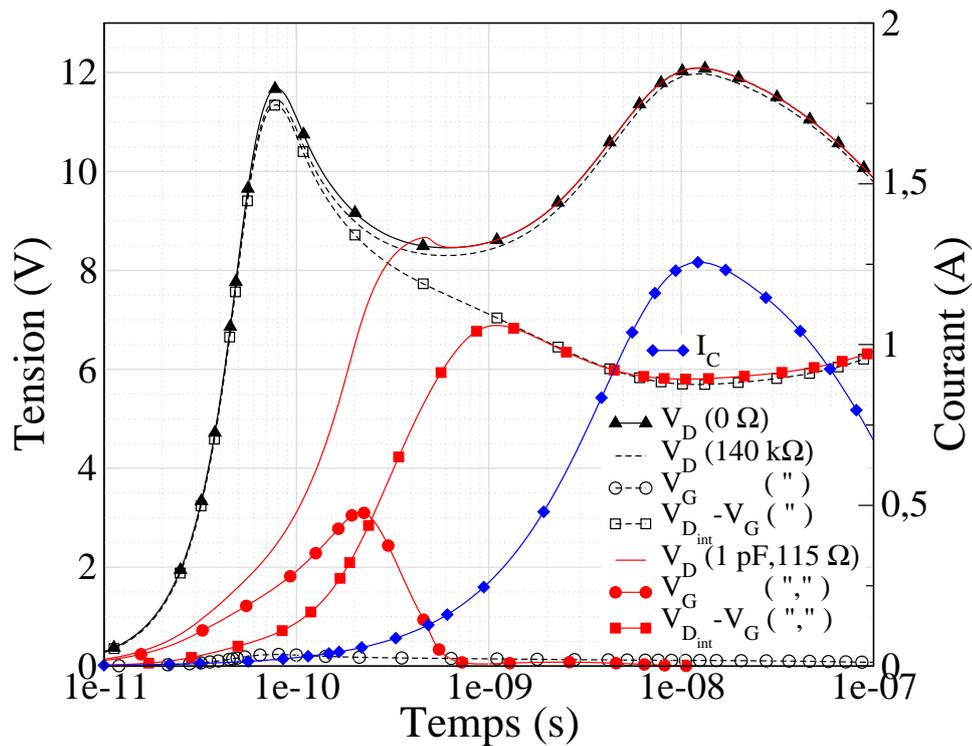


FIG. 4.19 – Réponse du composant NMOS à un stress HBM de 2 kV, simulé par le modèle SPICE pour différentes configurations de déclenchement par couplage de grille. Si une résistance ou une capacité de couplage sont utilisées, leurs valeurs sont indiquées dans la légende.

4.7 Conclusion

Après un résumé des différentes méthodes et techniques de modélisation utilisées pour la simulation compacte des composants de protection ESD basés sur les TBA, l'approche sur laquelle nous avons développé nos modèles a été justifiée.

La minimisation du nombre de paramètres à extraire, par la réutilisation des modèles présents dans les bibliothèques d'une technologie, permet de développer rapidement les modèles des composants de protection. En s'appuyant sur les modèles performants déjà développés au sein du laboratoire, nous avons cherché à les enrichir des phénomènes associés aux fortes densités de courant mis en évidence dans les deux chapitres précédents. L'approche modulaire, retenue pour son efficacité vis-à-vis de l'extraction des paramètres, permet d'introduire facilement ces effets mais implique certaines limitations. Les modèles développés présentent l'avantage d'être étroitement liés à la structure interne du composant et offre ainsi la possibilité de mieux comprendre et analyser l'impact de certains paramètres structurels.

La dynamique du déclenchement des transistors bipolaires et MOS est un sujet généralement peu abordé. Malgré la relative complexité des phénomènes, une analyse simplifiée a permis de mettre en évidence les paramètres clés du déclenchement des composants. L'effet distribué des courants capacitifs dans le composant a été mis en évidence. Une amélioration du macromodèle a été proposée pour leur prise en compte.

Enfin, à partir de ces études sur la dynamique du déclenchement, nous avons présenté une méthode d'optimisation du déclenchement par couplage de la grille des transistors de protection NMOS. Basée sur une étude de la distribution des courants dans les différentes capacités du transistor MOS, cette méthode a permis d'améliorer significativement les performances des composants.

Chapitre 5

Étude des stratégies de protections

L'efficacité de la protection des circuits intégrés contre les décharges électrostatiques repose sur le développement de composants de protection robustes et performants. Dans les circuits modernes, le développement de solutions de protection ne peut cependant pas se limiter à la seule optimisation d'une bibliothèque de composants élémentaires. Le choix et la disposition de ces protections sur le circuit est une étape cruciale pour obtenir une protection efficace. La complexité croissante des fonctions assurées par les circuits et la demande d'une très bonne robustesse aux ESD, rendent cette tâche de plus en plus difficile. L'obtention d'un circuit qui répond aux critères de robustesse ESD requis dans le cahier des charges d'un circuit, augmente parfois le nombre de cycles de conception. Il est donc indispensable de disposer d'outils performants pour déterminer la stratégie de protection la mieux adaptée à chaque cas.

Le développement de modèles compacts des composants, adaptés aux régimes de fort courant ESD, permet l'analyse globale du circuit et de sa protection [143]. L'étude d'un circuit simple va permettre de démontrer l'efficacité et les limites de cette approche. Nous verrons que les règles classiques de conception ne sont pas suffisantes pour garantir les performances de la protection. En effet, l'interaction entre les composants de protection et de leurs anneaux de garde contre le phénomène de latch-up lors d'une décharge électrostatique sera mise en évidence. L'origine d'un cas critique lors de décharges MM sera également expliquée. Enfin, de nouvelles règles de conception seront définies pour résoudre ces problèmes et garantir les performances des stratégies de protection proposées aux concepteurs de circuits intégrés.

5.1 Stratégies de protection des circuits

La protection d'un circuit contre les décharges électrostatiques consiste à offrir des chemins spécifiques pour les courants des décharges. L'impédance des chemins doit être suffisamment faible pour limiter les tensions et protéger la partie active du circuit. Pour cela, un réseau de composants de protection intégrés doit être constitué afin de détecter et conduire les décharges quelles que soient les broches d'entrée et de sortie du courant.

Dans le cadre d'une technologie CMOS, la constitution de ce réseau repose sur quelques schémas de base, définis en fonction du type de broche à protéger.

5.1.1 Protection d'entrée

La protection des étages d'entrée d'un circuit CMOS est très critique. Les oxydes de grille des transistors NMOS et PMOS sont particulièrement vulnérables aux ESD. Ils sont généralement protégés par un circuit de protection à deux étages, constitué par les éléments P de 1 à 4 sur le schéma de la figure 5.1. Ces éléments sont souvent des diodes ou des transistors MOS en technologie CMOS [144] et des transistors bipolaires ou des thyristors dans les technologies BiCMOS [145, 146]

Dans le schéma de protection à deux étages, le rôle du premier étage, ou protection primaire, est de conduire la majeure partie du courant de décharge. Le second, qui est séparé du premier par une résistance R_E , permet de garantir une limitation de la tension sur les oxydes de grille à une valeur suffisamment basse. La tension de déclenchement de la protection primaire pourra donc être supérieure à la tension de claquage des oxydes, laissant ainsi plus de liberté dans le choix des composants de protection utilisés.

En fonction des contraintes de conception et du type de protection globale adopté, le schéma général d'une protection d'entrée présenté dans la figure 5.1 peut être simplifié. La protection la plus rudimentaire consiste à utiliser un seul composant de protection P_1 capable de conduire le courant d'une décharge quel que soit son sens de circulation. L'utilisation d'un transistor de protection NMOS à grille couplée ou non permet par exemple de réaliser ce type de protection [144].

Une solution typique consiste à réaliser une «protection en π ». Dans ce cas, le schéma est simplement constitué par les protections P_1 et P_2 séparées par une résistance. La protection P_1 est, par exemple, réalisée à l'aide d'un transistor NPN latéral de type FOD robuste, alors que la protection du second étage est un transistor NMOS de plus petite taille mais dont la tension de déclenchement est plus faible que celle du premier étage. Lors d'une décharge HBM circulant de la broche d'entrée vers la broche de masse, le courant va d'abord circuler dans la protection P_2 au travers de la résistance R_E . L'élévation de tension aux bornes de cette dernière va permettre de déclencher le transistor NPN P_1 qui va finalement drainer l'essentiel du courant de la décharge. Dans le cas plus favorable d'une décharge de signe opposé, circulant de la masse (VSS) vers l'entrée, les transistors P_1 et P_2 se comportent comme des diodes en direct. Ils offrent alors un chemin particulièrement robuste et limitent la tension appliquée sur l'oxyde de grille à une valeur très basse.

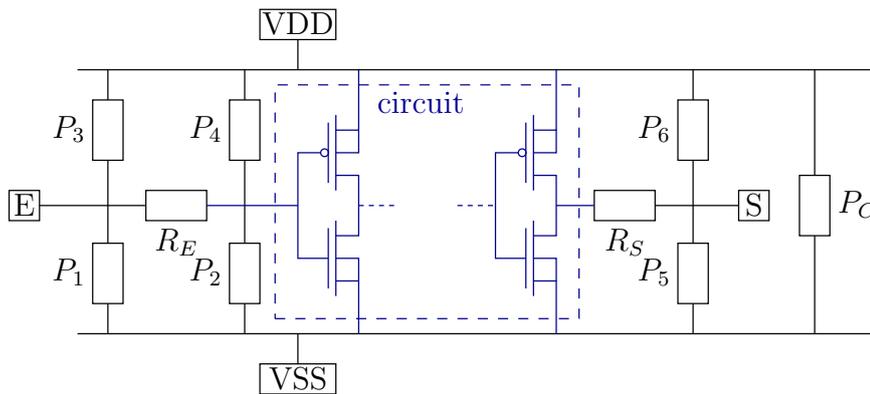


FIG. 5.1 – Représentation schématique d'un réseau de protection ESD pour un circuit en technologie CMOS.

Remarquons finalement que le schéma général de la protection d'entrée de la figure 5.1, s'apparente à une double structure en π . La seconde structure en π est formée par les protections P_3 et P_4 et fonctionne de façon identique à la première.

5.1.2 Protection de sortie

Un schéma identique à celui utilisé pour la protection d'entrée peut être retenu pour une sortie. Les contraintes sont cependant très différentes dans ce cas. L'étage de sortie doit être généralement capable de fournir un courant relativement important. L'utilisation d'une résistance pour séparer une structure à deux étages est donc souvent à proscrire. Toutefois, les contraintes sont moins importantes dans ce cas. En effet, les jonctions drain-substrat des transistors de sortie sont des éléments moins sensibles aux ESD que les oxydes de grille. L'utilisation d'un simple

étage de protection est généralement suffisante. Il est cependant fortement recommandé de placer une résistance R_S de faible valeur entre l'étage de sortie du circuit et l'étage de protection ESD (Fig. 5.1). Cette précaution permet de limiter la part du courant de décharge susceptible de circuler dans le circuit.

En outre, la taille relativement importante des transistors de sortie d'un circuit permet souvent une auto-protection de l'étage de sortie. Dans ce cas, des règles de dessins spécifiques aux ESD leur sont appliquées lors de la conception. Néanmoins, le fonctionnement de ces transistors lors d'une décharge peut, dans le cas de transistors MOS, entraîner une dégradation de l'oxyde de grille et une modification de leurs caractéristiques électriques inacceptables pour des applications analogiques. Dans ce cas, l'utilisation d'un étage de protection spécifique est préférable.

5.1.3 Protection du bus d'alimentation

Le dispositif de protection placé entre les lignes d'alimentation du circuit est souvent désigné par le terme de protection centrale P_C (ou power clamp) (Fig. 5.1). Le premier rôle de cet élément de protection est de conduire les courants de décharge entre les lignes d'alimentation (VDD et VSS).

Le dispositif de protection couramment utilisé en technologie CMOS est un transistor NMOS ou un transistor bipolaire latéral de type FOD, dont le déclenchement est assuré par le claquage de la jonction drain-substrat ou collecteur-base [147]. et dont le fonctionnement lors d'une décharge est du type TBA. L'utilisation d'un transistor NMOS dont la tension de déclenchement est abaissée grâce au couplage de grille, est également courante.

Les contraintes de conception sont moins importantes que pour les protections d'entrée ou de sortie. En effet, la capacité parasite de la protection et son courant de fuite sont moins critiques. La taille de la protection centrale peut donc être bien supérieure à celles disposées sur une entrée ou une sortie.

Il est ainsi possible d'utiliser une protection constituée par un transistor MOS de taille importante, commandé par un circuit de détection d'ESD [148]. Le circuit de détection commande la grille du transistor NMOS. Le transistor n'est donc pas déclenché par le claquage par avalanche de la jonction drain-substrat, et le courant n'est plus conduit par le transistor NPN parasite mais par le transistor MOS dont la taille importante permet de supporter le passage du courant ESD. La protection centrale devient alors une protection active. Ce type de solution permet de réaliser des protections efficaces dans les technologies fortement submicroniques pour lesquelles les marges de conception sont de plus en plus faibles. De plus, la modélisation compacte de son fonctionnement est plus aisée car elle ne met pas en jeu les phénomènes de claquage par avalanche et de repliement. Plus facile à optimiser et à transposer dans d'autres technologies, le principal inconvénient reste l'espace important occupé par la protection [149].

Le rôle de la protection centrale n'est généralement pas limité à conduire des décharges entre la masse et l'alimentation. Plusieurs chemins de conduction sont souvent possibles pour une configuration de décharge donnée (d'une broche à une autre) et la protection P_C constitue fréquemment un élément de ces différents chemins. Le degré de sollicitation de la protection centrale dépend fortement de la stratégie de protection globale adoptée.

5.1.4 Stratégie de protection globale

L'ensemble des protections ESD placées sur les plots d'entrée, de sortie et d'alimentations d'un circuit forment un réseau de protections. Quelques soient les broches d'entrée et de sortie du courant, ce réseau doit offrir un chemin à la décharge électrostatique. La protection d'un circuit doit donc être analysée d'une manière globale.

Le choix de la stratégie de protection repose principalement sur les contraintes suivantes :

- les spécifications ESD requises

- les spécifications électriques des broches d’entrée ou de sortie (excursion en tension, courant de fuite maximal, capacité et résistance maximales, fréquence de fonctionnement, ...)
- la surface de silicium disponible sur la puce

Pour satisfaire à ces dernières, deux grands types d’approche peuvent être distingués, les stratégies centralisée et distribuée. D’une manière générale, la stratégie globale retenue pour un circuit pourra être mixte.

Une stratégie de protection distribuée consiste à optimiser chacune des protections d’entrée ou de sortie et d’alimentation indépendamment. Un chemin robuste devra être prévu pour les décharges positive et négative entre une entrée/sortie et les deux lignes d’alimentation (VDD et VSS) ainsi qu’entre les lignes d’alimentation. De cette façon, plusieurs chemins de décharge parallèles sont créés pour les décharges entre deux broches d’entrée/sortie. La redondance introduite permet, dans une certaine mesure, de garantir la robustesse indépendamment de la résistance des pistes de métal. Cependant, les capacités et courants de fuite parasites introduits par les protections ne sont pas acceptables pour certaines applications.

Le principe d’une protection centralisée est présenté sur la figure 5.2. Les protections sur les entrées et sorties sont réalisées par des diodes qui vont diriger le courant de décharge vers la protection centrale P_C . Quelque soit la décharge, les diodes ne doivent pas être polarisées en inverse car elles sont peu robustes dans ce régime de fonctionnement. Leur mode de fonctionnement est donc unidirectionnel contrairement à la protection centrale qui doit être bidirectionnelle. Dans

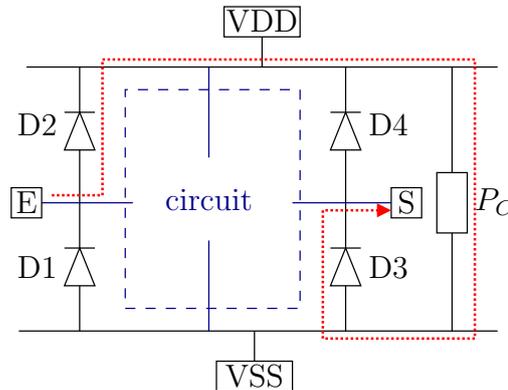


FIG. 5.2 – Représentation schématique d’une stratégie de protection centralisée.

ce type de protection, le chemin de décharge est unique pour un couple de broches donné. Le tableau 5.1 résume ces chemins pour les 12 configurations de décharge qu’offre ce circuit simple. Comme représenté sur la figure 5.2, un courant de décharge qui se propage de l’entrée vers la sortie va circuler au travers de la diode D2 puis dans la protection centrale polarisée en inverse pour sortir enfin après son passage dans la diode D3.

+ ↑	E	VSS	S	VDD
E		D2- P_C	D2- P_C -D3	D2
VDD	P_C -D1	P_C	P_C -D3	
S	D4- P_C -D1	D4- P_C		D4
VSS	D1		D3	P_C

TAB. 5.1 – Chemins de décharge entre chaque couple de broches. Le courant de décharge circule d’une broche choisie dans la première colonne vers une broche choisie dans la première ligne.

Le tableau 5.1 montre que la protection centrale est sollicitée dans une majorité des chemins de décharge, d’où son nom.

Une protection centralisée possède de nombreux avantages. Les diodes qui ne fonctionnent qu'en polarisation directe sont particulièrement robustes même si elles sont de taille réduite. Elles n'introduisent donc qu'une capacité et des courants de fuite réduits sur les entrées et sorties. La place occupée par le réseau de protection est également limitée.

Une stratégie de protection centralisée associée à l'utilisation d'une protection centrale active permet d'éviter les phénomènes d'avalanche quelque soit le chemin de décharge. La stratégie de protection est alors facile à simuler et à optimiser [150].

L'adoption de ce type de protection est cependant contraignante. Pour une configuration donnée, un seul chemin est offert au passage du courant. Il est donc indispensable de s'assurer que les chutes de potentiel dans les résistances des pistes métalliques, en particulier des lignes d'alimentation, sont suffisamment basses. Pour cela, les pistes métalliques doivent être correctement dimensionnées. La protection centrale peut également être dupliquée ou distribuée sur la longueur de la ligne d'alimentation [151, 152].

Si un plot d'entrée ou de sortie peut être polarisé à des tensions très supérieures à la tension d'alimentation, il n'est pas possible d'utiliser une diode pour le lier à l'alimentation. Ceci limite le domaine d'application des stratégies de protection centralisées. Des solutions spécifiques doivent souvent être étudiées pour des broches particulières. D'une manière générale, le réseau de protection ESD d'un circuit peut être constitué par une approche mixte entre protection centralisée et distribuée qui permet de bénéficier des avantages de chacune d'elles en fonction des contraintes de conception.

5.2 Étude d'une stratégie de protection

Alors que l'optimisation des composants de protection est largement étudiée et publiée, l'étude des stratégies de protection est un sujet moins abordé dans les publications. Nous proposons ici de démontrer que l'utilisation de la simulation de type SPICE permet une analyse approfondie du fonctionnement de la stratégie de protection d'un circuit. Nous allons montrer par ailleurs que cette approche associée à la seule optimisation des composants de protection, n'est pas suffisante pour garantir la robustesse d'un circuit.

5.2.1 Description du circuit de test

Le circuit d'évaluation, dont une vue schématique est donnée dans la figure 5.3, a été utilisé pour tester une stratégie de protection. La technologie utilisée pour réaliser le circuit a été décrite dans la section 3.7.1 (technologie 1). Il s'agit d'une technologie CMOS 1,2 μm analogique. La tension d'alimentation VDD de ce circuit est de 6 V. Il est composé d'un circuit interne très simple et d'un étage de sortie qui réalisent une simple fonction d'inversion. Une double structure en π est utilisée pour protéger l'étage d'entrée. Elle est constituée par deux transistors NMOS (M1 et M3), une diode D2 et un transistor PMOS M4. La protection centrale est un transistor NMOS M2 avec un couplage capacitif de grille. L'étage de sortie est auto-protégé grâce aux transistors NMOS et PMOS qui suivent des règles de dessin spécifiques pour leur permettre de conduire les décharges sans être détruits (résistance de ballast dans le drain, ...).

Pour un stress de type HBM, la robustesse des transistors NMOS M2 et M1 est de 7 kV et 6 kV, respectivement. Une robustesse de 4 kV est visée pour l'ensemble du circuit.

Vérification de la stratégie de protection par simulation

Afin de vérifier le bon fonctionnement de la stratégie de protection, les différentes configurations de test HBM ont été simulées. Dans chaque cas, les différences de potentiel et les niveaux de courant critiques ont été contrôlés à l'intérieur du circuit. Les tensions maximales admissibles sur les oxydes de grille et la tension de claquage drain-source des transistors NMOS sont de 33 V

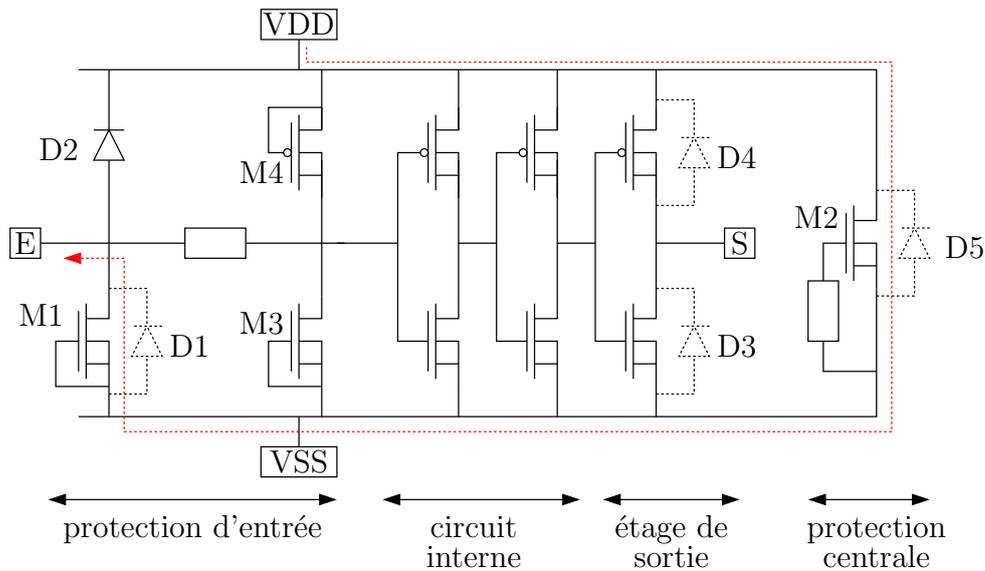


FIG. 5.3 – Schéma électrique du circuit de test.

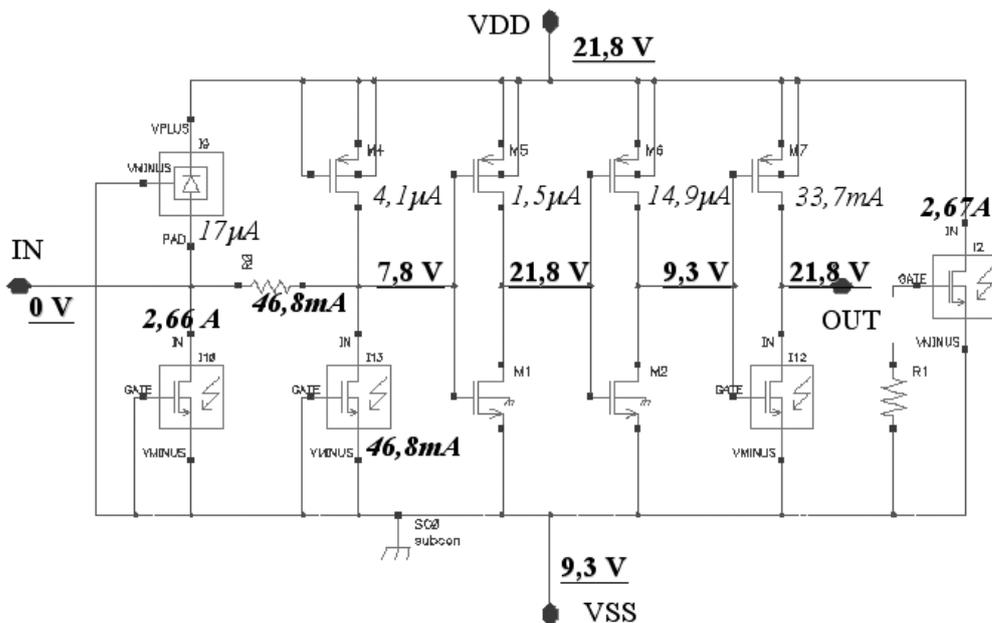


FIG. 5.4 – Résultats de simulation SPICE pour un stress HBM positif de 4 kV appliqué entre VDD et E. Les courants sont donnés au pic de la décharge et les tensions au moment où la tension est maximale sur l'ensemble du chemin.

et 15 V, respectivement. Les niveaux de courants dans le circuit interne ne doivent pas dépasser des valeurs compatibles avec les transistors qui le constituent.

La figure 5.4 représente le résultat de la simulation d'un stress de 4 kV HBM entre les broches VDD et E du circuit. Les valeurs de tension reportées sur la figure correspondent à l'instant où

la tension est la plus forte sur l'ensemble du chemin. Les valeurs de courant correspondent au moment du pic de courant de la décharge HBM. L'ensemble des simulations a permis de vérifier que la stratégie est efficace dans chacune des 12 configurations de décharge HBM.

5.2.2 Confrontation des résultats expérimentaux et simulés

La caractérisation de la robustesse HBM du circuit a été réalisée en utilisant la procédure de test suivante :

- Le stress HBM positif est appliqué sur un couple de broches dont l'une sert de référence.
- Pour chaque configuration, trois composants sont testés avec des pas de tension HBM de 500 V.
- Entre chaque stress, la fonctionnalité du circuit est vérifiée pour un courant de 100 μA fournie par l'étage de sortie.
- De plus, les courants de fuite de l'alimentation et de l'étage d'entrée sont contrôlés en l'absence de charge sur la sortie, à la fois pour les états logiques haut et bas.

Le critère de défaillance est atteint lorsqu'un courant de fuite est supérieur à 1 μA . Les résultats des caractérisations sont résumés dans le tableau 5.2. La disposition particulière du tableau permet de reprendre les symétries du réseau de protection et de grouper les cas les plus critiques.

kV + \uparrow	E	VSS	S	VDD
E		6	5	13
VDD	3	7	5	
S	6,5	7,5		15
VSS	14		16	16

TAB. 5.2 – Résultats des tests HBM entre chaque broches. Une broche de la première colonne est stressée positivement par rapport à une broche de la première ligne.

+ \uparrow	E	VSS	S	VDD
E		M1 ou D2,M2	D2,M2,D3 ou M1,D3	D2
VDD	M2,D1	M2	M2,D3	
S	D4,M2,D1	D4,M2		D4
VSS	D1		D3	D5

TAB. 5.3 – Chemin de décharge prévu par la simulation SPICE dans chaque configuration de stress.

La simulation SPICE permet de déterminer le chemin de décharge et les composants actifs pour chaque configuration de stress HBM. Ces chemins sont résumés dans le tableau 5.3. Les chemins de décharge qui ne mettent en jeu que des diodes polarisées en direct sont groupés dans la dernière colonne et la dernière ligne du tableau. Ils ne constituent pas des cas critiques pour la robustesse du circuit.

La robustesse des chemins composés devrait être proche de celle de l'élément le plus fragile. La robustesse des diodes, toujours polarisées en direct, étant très élevée, le minimum de robustesse attendue devrait donc correspondre à la robustesse des transistors M2 et M1 de 7 kV et 6 kV, respectivement. Or, plusieurs configurations de décharge présentent des tenues bien inférieures. Et plus particulièrement, la stratégie de protection vérifiée par simulation SPICE ne répond pas aux attentes d'une tenue globale de 4 kV dans le cas d'un stress VDD/E.

Afin de vérifier la validité des résultats de simulation, nous les avons confrontés à plusieurs résultats expérimentaux. La configuration de décharge de l'entrée vers la sortie est d'un intérêt

particulier car deux chemins de décharge sont possibles et mettent en jeu deux ou trois composants de protection. Pour cette configuration, les caractéristiques TLP expérimentales et simulées (Fig. 5.5) sont en bon accord. Remarquons que les résistances des pistes de métal doivent être extraites avec précision pour obtenir une caractéristique correcte. Les deux décrochages successifs de la caractéristique témoignent du déclenchement des deux chemins possibles pour la décharge. D'après les informations tirées des simulations, le courant circule au travers des composants D2, M2 et D3 pour la première partie de la courbe TLP, puis par les composants M1 et D3 pour la seconde partie (Fig. 5.5). Le chemin D2,M2,D3 est actif dans un premier temps car le transistor M2 est déclenché par couplage capacitif, contrairement au transistor M1 dont la grille est liée à la masse. Pour des niveaux de courant plus importants, sous l'effet des chutes de potentiel dans la diode D2 et les pistes métalliques, la tension aux bornes du transistor M1 devient suffisante pour permettre son déclenchement. Le courant emprunte alors le chemin M1,D3.

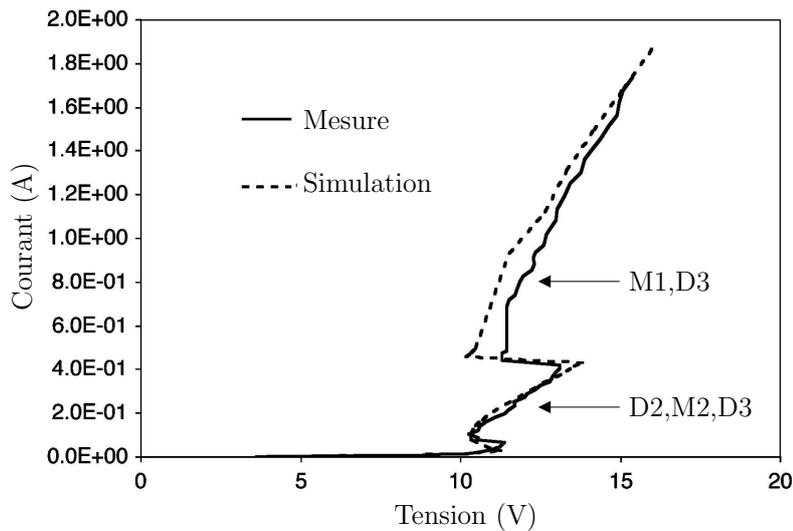


FIG. 5.5 – Caractéristiques TLP expérimentale et simulée de la broche d'entrée par rapport à celle de sortie.

L'utilisation de la microscopie d'émission lumineuse (photoémission) couplée au banc de mesure TLP permet de déterminer expérimentalement le chemin suivi par une décharge [41]. La confrontation du chemin prédit par simulation à celui déterminé expérimentalement permet de valider les résultats de simulation.

La vue du circuit de test étudié est donné dans la figure 5.6. Les différents composants sont repérés sur l'image.

Des images de photoémission ont été réalisées pour des courants TLP de 370 mA et 1,7 A appliqués entre l'entrée et la sortie. La longueur de l'impulsion TLP est de 120 ns et la fréquence de répétition de 10 Hz. La durée totale de l'exposition est de 5 min. L'image de la figure 5.7(a) montre que seul le transistor M2 est actif pour un courant de 370 mA. Le courant emprunte donc le chemin par les composants D2,M2 et D3. Les diodes D2 et D3, qui sont polarisées en direct, émettent moins de photons que le transistor M2 polarisé en inverse, et dans des longueurs d'onde pour lesquelles la caméra utilisée est moins sensible. Leur fonctionnement n'est donc pas visible sur l'image de photoémission. Les courants obtenus par simulation (Fig. 5.7(b)) montrent effectivement que seul le transistor M2 est actif dans ces conditions. De la même façon, l'image de photoémission montre que seul le transistor MOS M1 est actif pour un courant de 1,7 A (Fig. 5.8(a)), ce qui correspond parfaitement avec les résultats obtenus par simulation dans la figure 5.8(b).

Les résultats de simulation ne peuvent pas être remis en cause. La simulation s'avère donc être un outil performant pour prédire les chemins de décharge.

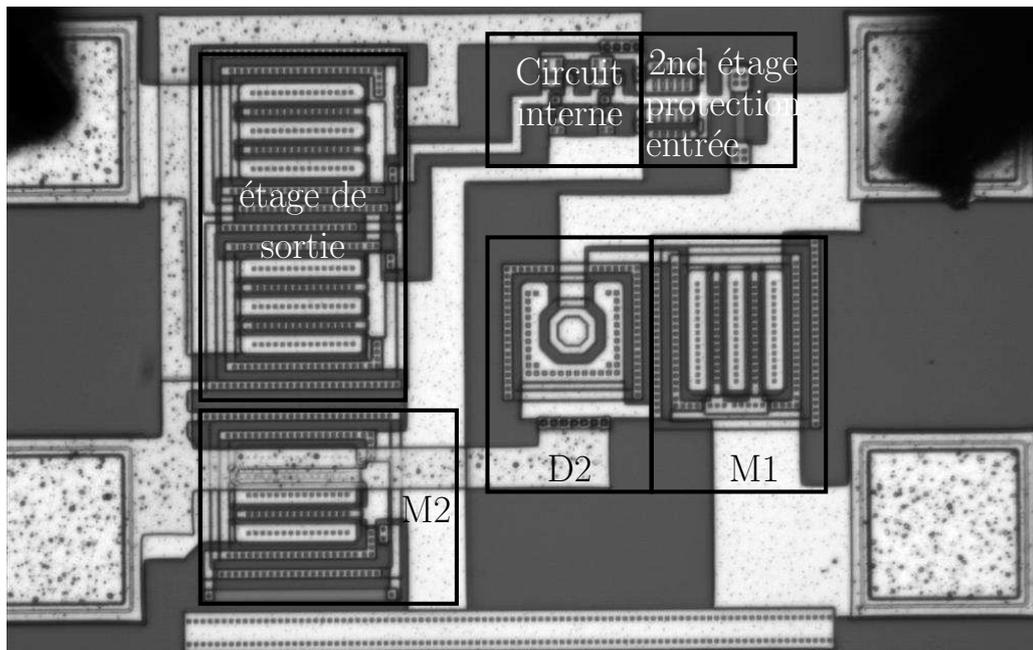


FIG. 5.6 – Vue du circuit de test.

5.2.3 Analyse de défaillance

Pour certains chemins de décharge VDD/E, VDD/S et E/S, les résultats de simulation ne permettent pas de prédire la robustesse du circuit étudié. La simulation montre qu'aucune surtension ou courant destructifs ne pourraient conduire à la destruction du circuit interne. Pourtant, la robustesse du circuit est nettement inférieure aux attentes, en particulier dans le cas du stress VDD/E où il ne peut supporter de décharge supérieure à 3 kV HBM.

Afin de comprendre l'origine du problème rencontré, une localisation de défaillance a été réalisée sur les circuits utilisés pour caractériser les configurations qui présentent une faible robustesse. Pour chacun de ces cas, la fonctionnalité du circuit défaillant n'est pas affectée. Seul le courant de fuite de l'alimentation possède une valeur anormalement élevée, variant de $2 \mu\text{A}$ à $20 \mu\text{A}$ en fonction des circuits alors qu'il était inférieur au nano-Ampère avant l'application des stress HBM.

Dans aucun des cas, l'utilisation de la technique de photoémission, en tant que technique d'analyse de défaillance, n'a permis de localiser des défauts, quelque soit l'importance du courant de fuite. Seules les techniques de localisation OBIRCH et SEI ont permis de déterminer le lieu de la défaillance. Les images OBIRCH obtenues dans le cas des chemins VDD/E et E/S sont présentées dans la figure 5.9.

La localisation du défaut est particulièrement précise grâce à cette technique d'analyse. Comme le montre la figure 5.9(a), le défaut observé sur le circuit qui a subi un stress HBM entre les broches VDD et E, se situe dans l'anneau de protection contre le latch-up du transistor de protection M1. De façon similaire, le défaut créé lors d'un stress HBM entre les broches E et S du circuit se situe dans l'anneau de latch-up du transistor NMOS de l'étage de sortie.

Ceci laisse supposer que le transistor NPN parasite formé par le drain du transistor (émetteur), le substrat (base) et l'anneau $N^{++}N$ de protection contre le latch-up (collecteur), est un élément qui joue un rôle important lors d'une décharge.

La jonction émetteur-base de ce transistor correspond à la diode drain-substrat D1 du transistor NMOS M1 (Fig. 5.3). Dans la figure 5.11, cette diode est remplacée par un transistor NPN pour représenter le transistor parasite NPN1. De la même manière, le transistor parasite NPN3 remplace la diode D3.

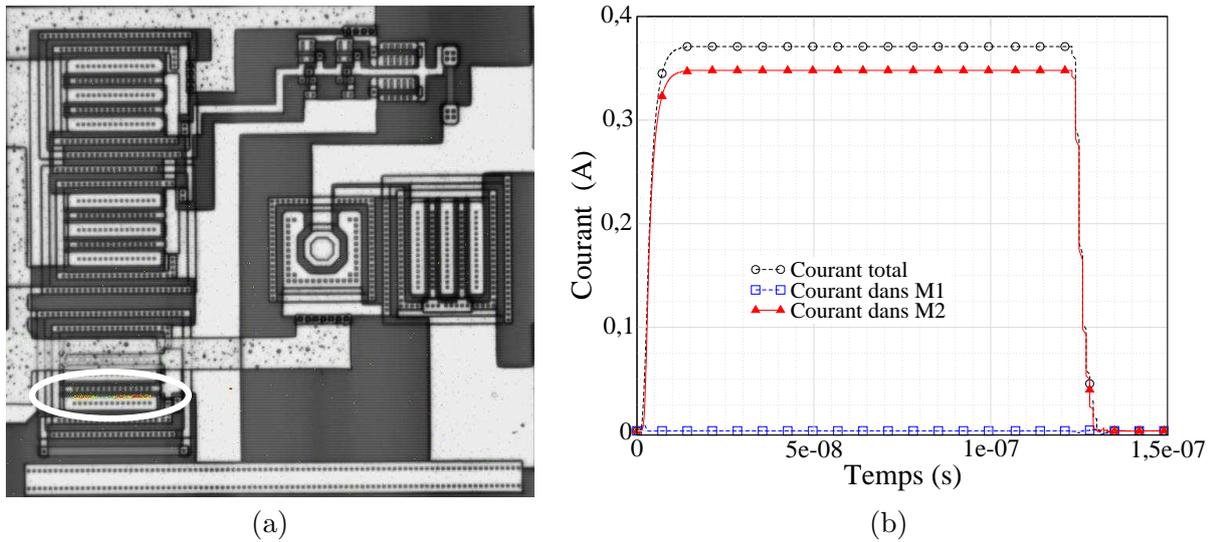


FIG. 5.7 – (a) Image en photoémission pour un courant TLP de 370 mA entre l'entrée et la sortie. (b) Courant dans les transistors M1 et M2 au cours d'une impulsion TLP de 370 mA.

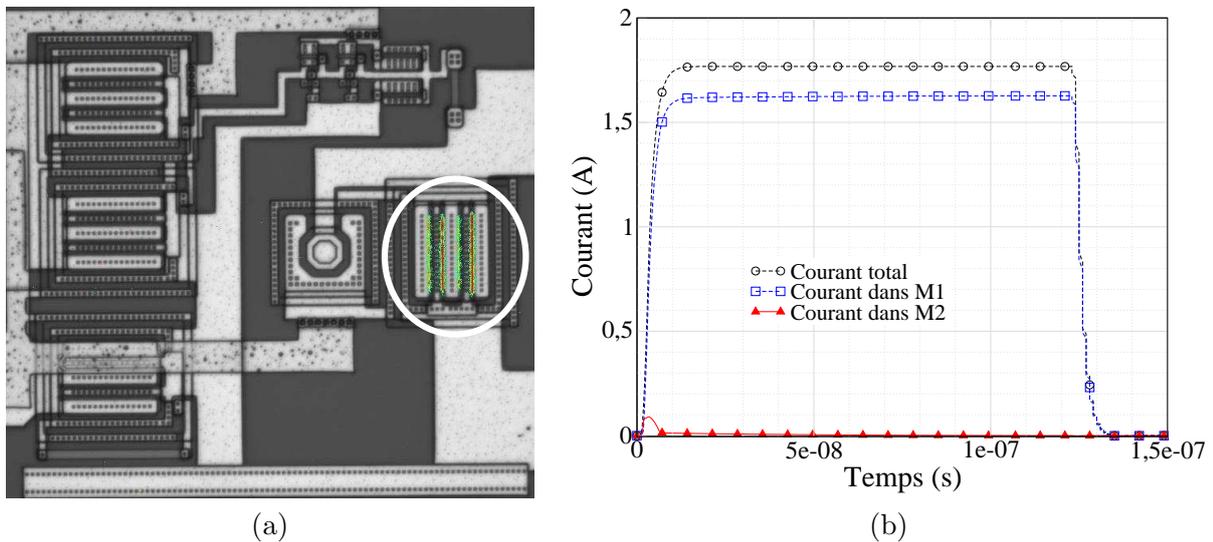


FIG. 5.8 – (a) Image en photoémission pour un courant TLP de 1,7 A entre l'entrée et la sortie. (b) Courant dans les transistors M1 et M2 au cours d'une impulsion TLP de 1,7 A.

Ce transistor dont le collecteur est relié à la ligne d'alimentation VDD, offre donc un chemin supplémentaire lors d'une décharge de VDD vers E (Fig. 5.11). De la même façon, le transistor bipolaire parasite NPN3 associé au transistor NMOS de l'étage de sortie, offre un chemin supplémentaire dans le cas des décharges VDD/S et E/S.

L'utilisation de la simulation physique bidimensionnelle en mode mixte permet de vérifier que le transistor bipolaire parasite NPN1 conduit une part importante du courant lors d'une décharge VDD/E, malgré son très faible gain en courant. Pour cela, une description bidimensionnelle est utilisée pour chacun des transistors M1 et M2. Les simulations physiques de ces deux composants sont couplées au travers d'un circuit SPICE qui permet également d'introduire le circuit équivalent du modèle HBM pour générer le courant décharge. Lors de la décharge VDD/E, l'essentiel du courant circule dans le transistor de la protection centrale qui est alors polarisé en inverse. Sa tension de maintien, de l'ordre de 7 V, définit la tension entre VDD et VSS qui

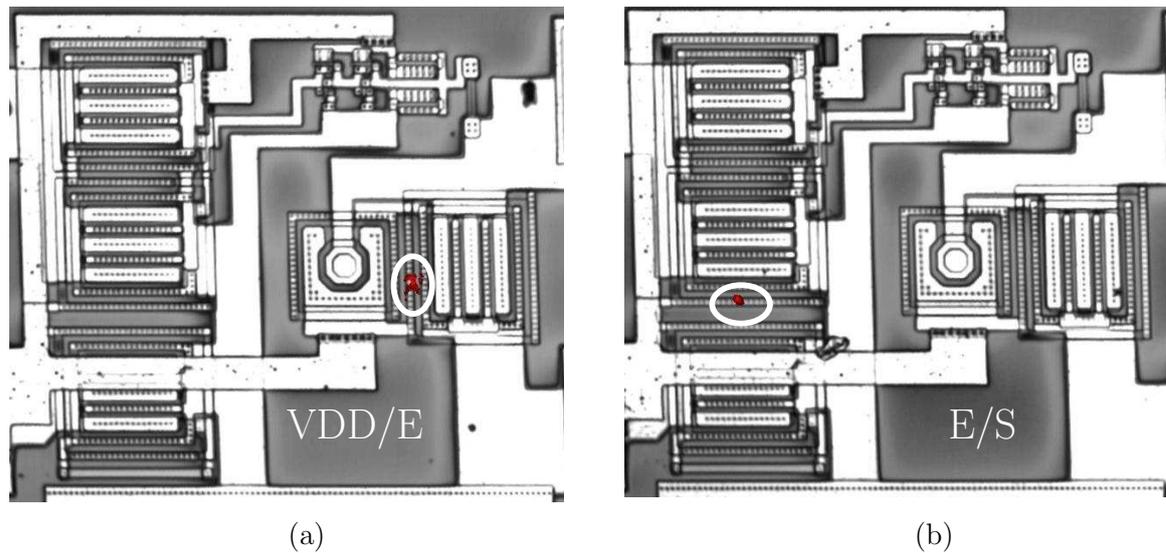


FIG. 5.9 – Localisation de la défaillance par la technique OBIRCH pour des circuits ayant subi un stress HBM entre les broches VDD et E (a) et entre l'entrée E et la sortie S (b). Les lieux de défaillances sont cerclés de blanc.

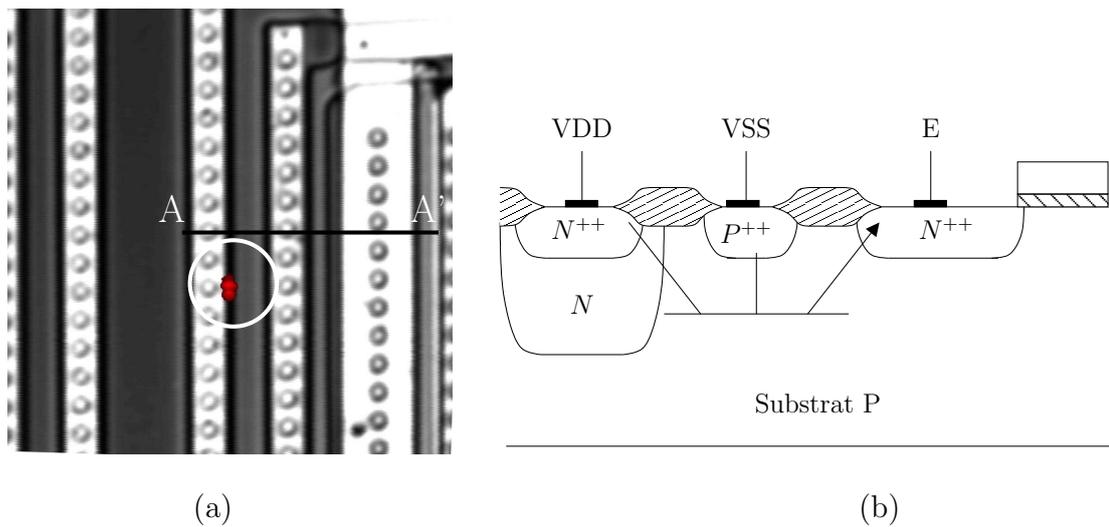


FIG. 5.10 – (a) Image OBIRCH de la localisation du défaut pour le stress E/S. (b) Coupe A-A' schématique du transistor bipolaire latéral parasite.

correspond également à la tension appliquée sur la jonction collecteur-base du transistor NPN1 alors polarisé en inverse. Dans ces conditions, le courant circulant dans le collecteur du transistor parasite représente environ 15% du courant total au cours de la simulation d'une décharge de 4 kV HBM.

L'utilisation de la photoémission couplée au banc de test TLP permet de confirmer le fonctionnement du transistor parasite pour les fortes valeurs du courant dans le cas d'un courant TLP circulant de l'entrée du circuit vers sa sortie (stress E/S) (Fig. 5.12). Dans cette image, la formation d'un point de focalisation du courant au cours des 5 minutes d'acquisition, apparaît clairement. Ce point est situé dans l'anneau de protection contre le latch-up du transistor NMOS de sortie. Sa position correspond avec le lieu de défaillance mis en évidence par l'imagerie OBIRCH.

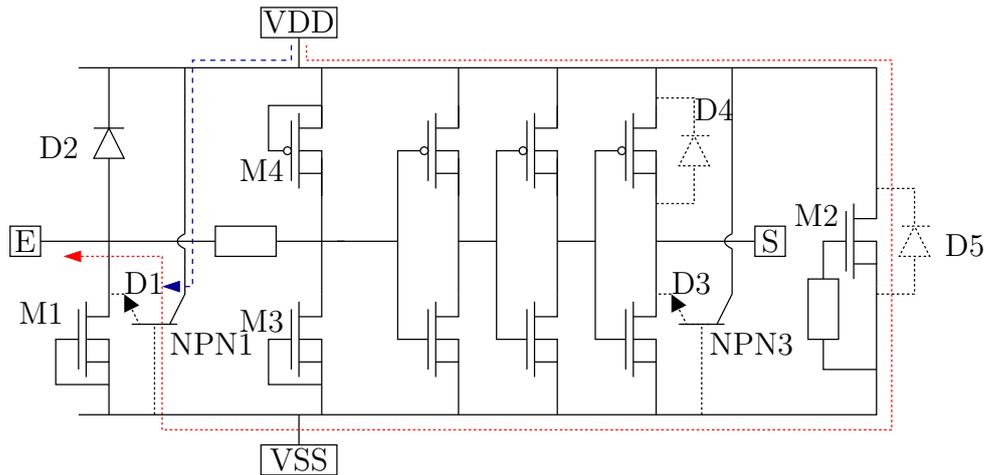


FIG. 5.11 – Schéma électrique du circuit de test dans lequel les diode D1 et D3 sont remplacées par les transistors parasites NPN1 et NPN3 qu'elles forment avec leur anneau de protection contre le latch-up.

5.3 Étude et optimisation des circuits de protection élémentaires

L'analyse précédente met en évidence que la seule conception de composants de protection robustes n'est pas suffisante pour assurer la protection efficace d'un circuit. Il apparaît en particulier nécessaire de prendre en compte les anneaux de protection contre le latch-up d'un composant de protection ESD dès sa conception, afin de garantir les performances des stratégies de protection globales.

En outre, les structures de protection ESD jouent un rôle important vis-à-vis du problème de latch-up. Au cours d'un test de latch-up, les protections ESD sont polarisées en direct, et injectent une grande quantité de porteurs dans le substrat. Les courants associés peuvent entraîner le déclenchement de thyristors parasites dans la partie du circuit la plus proche du composant de protection. Le déclenchement d'un thyristor parasite entraîne un important courant de consommation qui n'est limité que par l'alimentation utilisée. Ce phénomène constitue le problème du latch-up. La forte valeur du courant peut conduire à la destruction du thyristor incriminé et donc du circuit.

Afin, de limiter et de contrôler l'injection de porteurs dans le substrat, les composants de protection doivent être entourés d'anneaux de protection généralement connectés aux lignes d'alimentation [153]. Le rôle de ces anneaux est de collecter l'essentiel des porteurs injectés par le composant.

Comme nous venons de le montrer dans la section précédente, ces anneaux de protection contre le latch-up peuvent interagir avec la stratégie de protection ESD d'un circuit et dégrader fortement ses performances. L'objectif de l'étude suivante est de comprendre le comportement des transistors bipolaires parasites mis en cause dans la dégradation des performances ESD. L'utilisation de la simulation physique va permettre d'étudier les phénomènes de forte densité de courant dans ces transistors. Finalement, de nouvelles règles de conception seront proposées et validées expérimentalement.

5.3.1 Résultats sur les circuits de test

Pour simplifier l'étude, nous avons utilisé un réseau de protection simplifié, qui est présenté dans la figure 5.13. Ce circuit simple constitue le réseau de protection élémentaire de toute

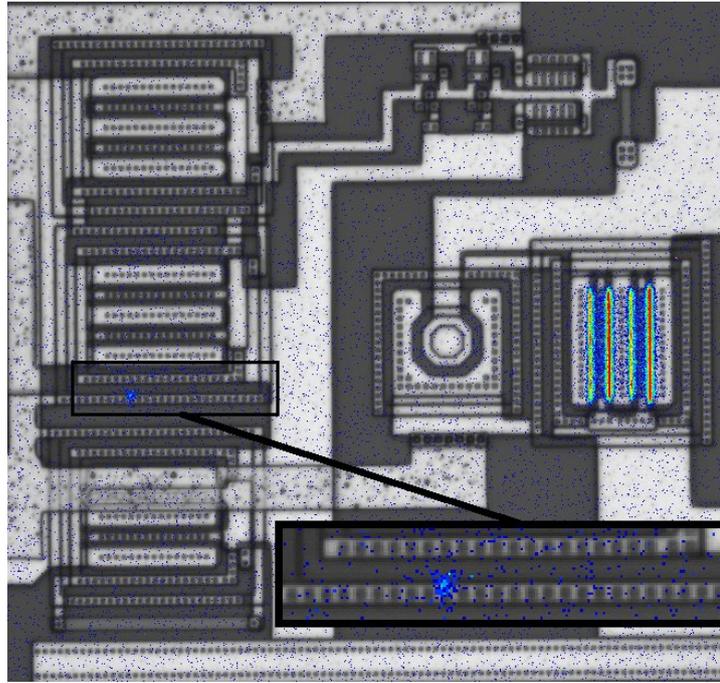


FIG. 5.12 – Image en photo-émission pour un courant de 2,3 A circulant entre les broches d'entrée et de sortie du circuit. Fréquence de répétition 10 Hz, durée totale d'exposition 5 minutes.

stratégie de protection ESD, dans le cas d'une alimentation unique. En effet, tout chemin de

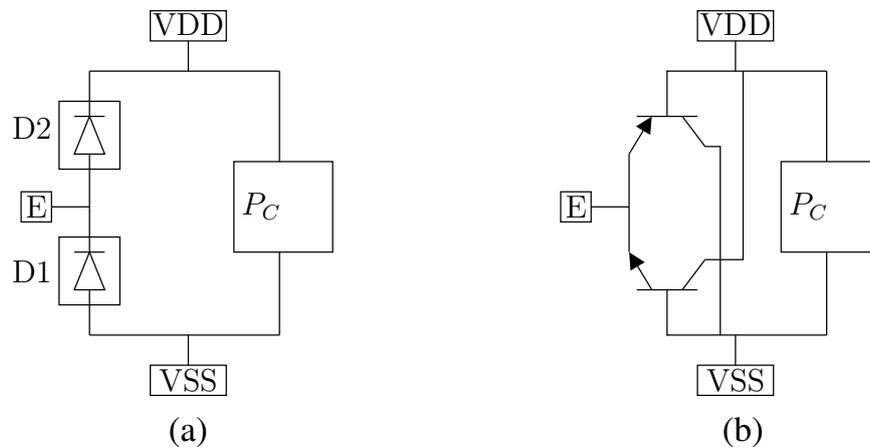


FIG. 5.13 – Réseau de protection élémentaire (a) et schéma équivalent pour la prise en compte des anneaux de protection contre la latch-up des diodes.

décharge d'un réseau de protection complexe peut être décomposé aux moyens des six chemins de décharges associés au réseau élémentaire.

La technologie CMOS 0,6 μm décrite dans la section 3.7.1 est utilisée pour la réalisation des circuits de test. Les diodes D1 et D2 sont formées par les jonctions N^{++}/P et P^{++}/N (Fig. 5.14), respectivement.

L'anneau de protection contre le latch-up de la diode D1 est constitué par une diffusion $N^{++}N$ qui entoure la diode, alors que celui de la diode D2 est constitué par une diffusion P^{++} (Fig. 5.14). Ces anneaux sont respectivement connectés aux lignes d'alimentation VDD et VSS pour les diodes D1 et D2. La protection centrale P_C est réalisée au moyen d'un transistor NMOS

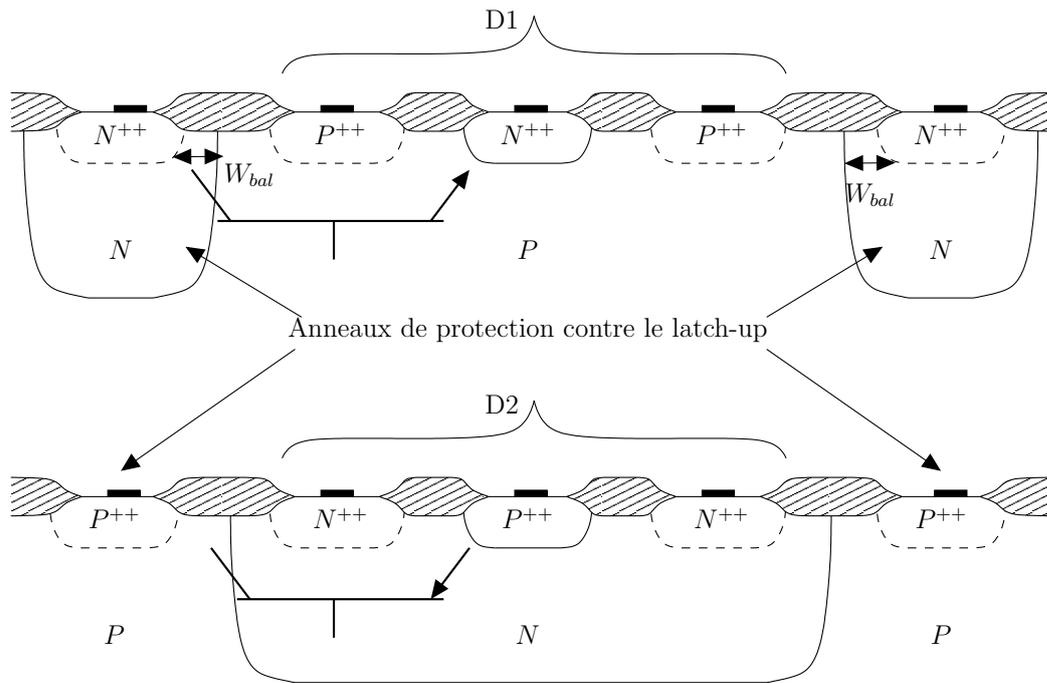


FIG. 5.14 – Représentation schématique de la coupe technologique des diodes D1 et D2 avec leurs anneaux de protection contre le latch-up associés.

à grille à la masse dont la robustesse HBM est de 2 kV.

Au cours de tests de latch-up, un circuit est alimenté et les diodes ESD D1 et D2 sont successivement polarisées en direct. Le rôle des anneaux de latch-up est de collecter les porteurs injectés par les diodes et de limiter ainsi l'injection de porteurs dans le substrat du circuit. Les diodes et leurs anneaux de garde forment des transistors bipolaires parasites (Fig. 5.13(b)). Ce transistor bipolaire est du type NPN latéral pour la diode D1 et de type PNP vertical pour la diode D2 (Fig. 5.14). Le gain de ces transistors est très faible car la distance entre leur collecteur et leur émetteur est très importante (plusieurs microns).

Les résultats des caractérisations HBM de chacun des six chemins de décharge possibles dans le réseau de protection élémentaire sont résumés dans le tableau 5.4.

kV + \uparrow	E	VSS	VDD
E		2	7
VDD	1	2	
VSS	8		>10

TAB. 5.4 – Résultats des tests HBM en kV entre chaque broche. Une broche de la première colonne est stressée positivement par rapport à une broche de la première ligne.

Pour des décharges positives de VSS vers E, VSS vers VDD et de E vers VDD, la robustesse est très importante car seules des diodes polarisées en direct sont mises en jeu dans le chemin de décharge. Les diodes D1 et D2 sont respectivement utilisées dans le cas des stress VSS/E et E/VDD. Pour le stress VSS/VDD, la diode intrinsèque drain-substrat du transistor NMOS de protection centrale est polarisée en direct. Au cours d'un stress de VDD vers VSS, seule la protection centrale est déclenchée et peut supporter des décharges d'au moins 2 kV. Le chemin de décharge dans le cas d'un stress VDD/E est composé de la protection centrale polarisée en inverse et de la diode D1 polarisée en direct. La robustesse de la protection centrale étant de 2 kV et de 8 kV pour la diode D1, on peut supposer que la robustesse de ce chemin est de 2 kV.

Or, la mesure donne une valeur de seulement 1 kV. Finalement, pour une décharge entre les broches E et VSS, le chemin de décharge est constitué par la diode D2 et la protection centrale. Dans ce cas, une robustesse de 2 kV est obtenue comme on pouvait le supposer.

D'après les résultats obtenus dans la section précédente, nous pouvons supposer que la faible robustesse obtenue pour un stress VDD/E est liée à la dégradation de l'anneau de latch-up de la diode.

5.3.2 Analyse par la simulation physique

Afin d'étudier le comportement du transistor NPN parasite associé à l'anneau de protection des diodes au cours d'une décharge HBM, nous avons utilisé l'outil de simulation physique bidimensionnelle. Il permet de mettre en évidence le fonctionnement du transistor NPN parasite et d'expliquer pourquoi le transistor PNP associé à la diode D2 ne dégrade pas de la même manière la robustesse du circuit.

L'ensemble du réseau de protection élémentaire est simulé au moyen de la simulation en mode mixte. Les diodes D1, D2 avec leurs anneaux de protection contre le latch-up et la protection centrale sont décrites par trois structures bidimensionnelles distinctes. Ces trois composants sont connectés au travers du simulateur SPICE qui permet également d'utiliser le circuit équivalent du modèle HBM comme générateur de décharge. De cette façon, le comportement du réseau de protection élémentaire pour un stress HBM est décrit avec une grande précision.

Deux configurations de stress revêtent un intérêt plus particulier : les stress E/VSS et VDD/E. Ces deux cas ont été simulés pour des décharges HBM de 1 kV et 2 kV. Les résultats des simulations sont présentés dans la figure 5.15(a) pour la décharge VDD/E et dans la figure 5.15(b) pour la configuration E/VSS. Dans chacun des cas, les résultats obtenus pour les stress de 1 et 2 kV HBM sont superposés.

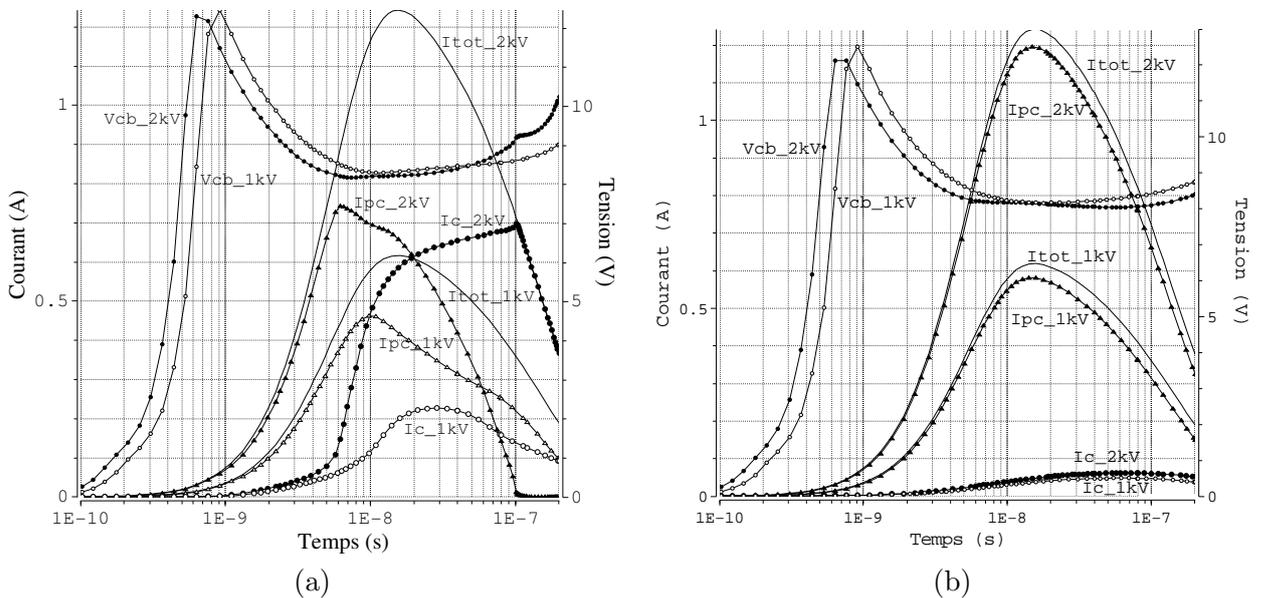


FIG. 5.15 – Répartition des courants dans le réseau de protection élémentaire obtenue par simulation en mode mixte pour des stress de 1 et 2 kV HBM appliqués entre les broches VDD et E (a) et les broches E et VSS (b). ($V_{cb_}$) tension entre collecteur et base du transistor NPN parasite, ($I_{tot_}$) courant total de la décharge. ($I_{pc_}$) et ($I_{c_}$) les courants circulant dans la protection centrale et le collecteur du transistor NPN (a) ou PNP (b), respectivement.

Au cours de ces décharges, les transistors NPN et PNP parasites associés aux diodes D1 et D2, se trouvent polarisés dans une configuration similaire au montage en base commune classique.

Dans le cas du stress VDD/E, la jonction émetteur-base du transistor NPN est polarisée en direct. La tension appliquée entre son collecteur et sa base est définie par la tension V_{PC} aux bornes de la protection centrale (Fig. 5.16(a)). Selon les résultats de simulation, cette tension est de l'ordre de 8 V après le repliement de la protection centrale lorsque le courant est maximum dans la diode (Fig 5.15).

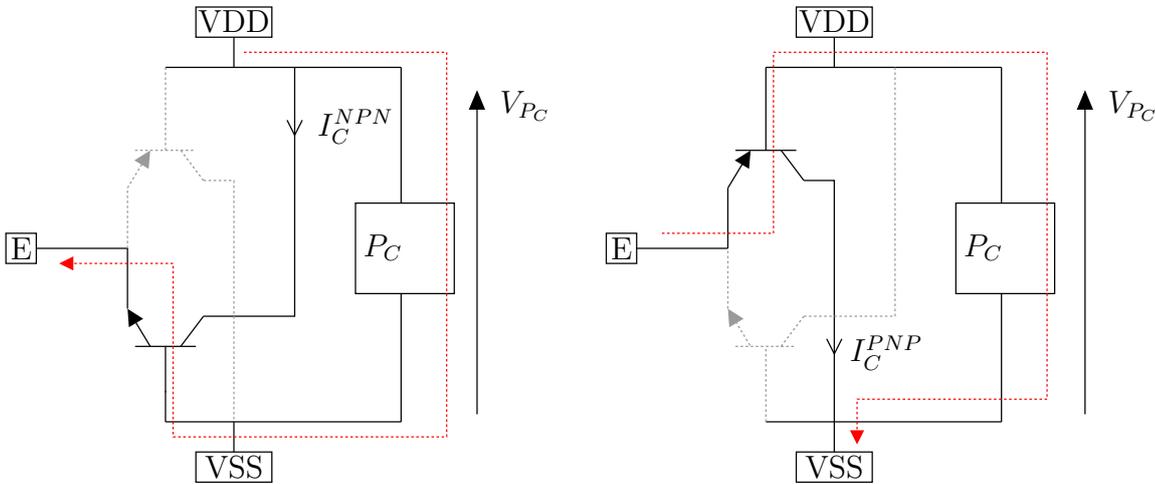


FIG. 5.16 – Représentation schématique du circuit utilisé pour la simulation physique bidimensionnelle en mode mixte du réseau de protection élémentaire soumis à un stress HBM. Chemin principal du courant de décharge dans le cas d'un stress VDD/E (a) et d'un stress E/VSS (b).

Pour un stress HBM de 1 kV, la majeure partie du courant de la décharge circule dans la protection centrale et environ 15% du courant circule dans le collecteur du transistor NPN parasite (Fig 5.15). De la même manière, le transistor PNP parasite associé à la diode D2 se trouve polarisé dans une configuration en base commune au cours d'une décharge entre E et VSS. Dans ce cas, seulement 5% du courant total de la décharge circule dans son collecteur (Fig 5.15)(b).

Le réseau de protection fonctionne correctement pour des décharges de 1 kV HBM. En revanche, pour une décharge de 2 kV, lors d'un stress de VDD vers E, le transistor bipolaire NPN parasite entre dans un mode de fonctionnement où il conduit la totalité du courant de la décharge. En pratique, son déclenchement va entraîner sa destruction. Par contre, aucun déclenchement du transistor bipolaire PNP parasite n'est constaté pour un stress de 2 kV HBM appliqué entre E et VSS.

Durant la décharge électrostatique, les jonctions collecteur-base des transistors NPN et PNP se trouvent polarisées en inverse (dans le cas d'un stress VDD/E pour le NPN et E/VSS pour le PNP). L'importante densité du courant qui circule dans la jonction émetteur-base entraîne un niveau d'injection très élevé dans la base. Dans ces conditions, malgré leur faible gain, ces transistors peuvent donc conduire un courant significatif. En effet, la densité de porteurs minoritaires dans leur base¹² devient importante avec la forte injection et les porteurs sont happés par le champ électrique de la jonction collecteur-base polarisée en inverse.

La différence de comportement du transistor NPN parasite vis-à-vis du transistor parasite PNP peut être attribuée à la différence fondamentale qu'il existe entre les facteurs de multiplication des courants d'électrons et de trous dans une jonction polarisée en inverse. Comme nous l'avons vu dans la section 2.4.2, le facteur de multiplication du courant d'électrons croît beaucoup plus rapidement que celui des trous avec la tension inverse. Dans une jonction plane abrupte, pour une valeur de tension inverse égale à 80% de la tension de claquage, le facteur

¹²Les porteurs minoritaires sont des électrons dans la base du transistor NPN et des trous pour le transistor PNP

de multiplication d'un courant d'électrons est de 3 alors que celui des trous reste proche de 1 (Fig. 2.15). Cette tendance n'est pas limitée au cas de jonctions abruptes et constitue une propriété générale quelque soit le profil de jonction [68].

Le transistor bipolaire NPN parasite peut donc se retrouver rapidement dans les régimes autopolarisés décrits dans les chapitres précédents. Le mécanisme de son déclenchement est d'ailleurs identique à celui utilisé pour diminuer la tension de repliement des transistors de protection développés précédemment. Le transistor PNP n'a en revanche que très peu de chance de se déclencher. De plus, nous avons montré que la tension de maintien d'un transistor PNP augmente très rapidement lorsque sa base est suffisamment profonde, même pour de très fortes densités de courant (section 2.4.3).

Remarquons finalement que le transistor bipolaire PNP parasite étant vertical, il possède, de par sa nature, une résistance de ballast naturelle dans son collecteur. La possibilité d'une focalisation du courant dans ce composant est ainsi extrêmement limitée.

5.3.3 Règles de dessin et validation

L'une des solutions pour résoudre le problème de la réduction de la robustesse du réseau élémentaire de protection lors d'un stress VDD/E, est d'éviter la mise en marche du transistor NPN parasite. Le déclenchement de ce transistor dépend fortement du niveau d'injection et donc de la densité de courant dans la diode émetteur-base. L'utilisation d'une diode de taille plus importante permettra de réduire la densité de courant mais entraînera une augmentation concomitante de sa capacité parasite, qui n'est pas toujours acceptable vis-à-vis du circuit à protéger.

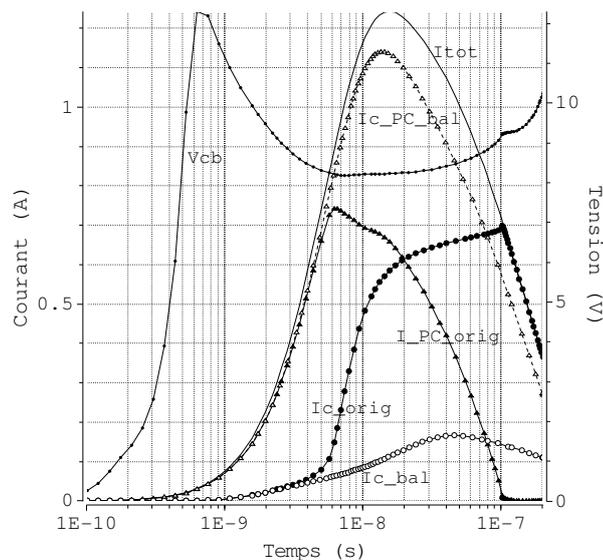


FIG. 5.17 – Répartition des courants dans le réseau de protection élémentaire obtenue par simulation en mode mixte pour un stress de 2 kV HBM appliqué entre les broches VDD et E. (V_{cb}) tension entre collecteur et base du transistor NPN parasite, (I_{tot}) courant total de la décharge. (I_{pc}) et (I_c) les courants circulant respectivement dans la protection centrale et le collecteur du transistor NPN, à la fois dans le cas de l'anneau de latch-up original ($_{-orig}$) et ballasté ($_{-bal}$).

L'élargissement de la base du transistor, réalisé en éloignant l'anneau de protection contre le latch-up, n'aura qu'un effet limité car le niveau d'injection dans la base est très élevé. Rappelons qu'au cours d'une décharge HBM de 2 kV, le pic de courant est de 1,3 A et entraîne des densités de courant dans le composant de protection de l'ordre de 10^5 A/cm². En revanche, l'ajout d'une

résistance de ballast dans le collecteur est une solution plus efficace. Cette résistance est obtenue en agrandissant la distance W_{bal} définie dans la figure 5.14. La chute de potentiel dans cette résistance, qui dépend directement du courant qui y circule, va réduire la tension appliquée sur la jonction collecteur-base et par conséquent le facteur de multiplication par avalanche. De cette façon, le courant total nécessaire pour déclencher le transistor bipolaire peut être repoussé à des valeurs bien supérieures à celle de destruction de la protection centrale.

La figure 5.17 présente les résultats de la simulation du réseau de protection élémentaire pour un stress VDD/E de 2 kV HBM dans le cas de l'anneau de protection contre le latch-up original et sa version ballastée. L'accroissement de $1\mu\text{m}$ de la distance W_{bal} permet de limiter le courant dans le transistor bipolaire parasite à un niveau qui ne permet pas son déclenchement et son passage dans un mode de fonctionnement autopolarisé.

Comme nous l'avons vu, il n'est pas nécessaire de porter d'attention particulière au transistor bipolaire PNP parasite, du moins dans la technologie utilisée ici. Dans des technologies plus avancées, les risques seront donc plus importants. En effet, pour des profondeurs de base plus faibles, nous avons mis en évidence que le comportement des transistor PNP tend vers celui des transistors NPN.

Le transistor bipolaire PNP parasite peut donc être plus facilement utilisé pour offrir un chemin de décharge parallèle dans le cas de décharges E/VSS [154, 151]. L'utilisation du transistor NPN dans le même objectif est plus difficile, comme nous venons de le montrer.

Grâce à l'introduction de la résistance de ballast dans le collecteur du transistor bipolaire NPN parasite, la robustesse attendue de 2 kV HBM pour l'ensemble du réseau de protection élémentaire a été atteinte expérimentalement.

Il convient cependant d'estimer l'impact de ce changement sur l'efficacité de l'anneau de protection contre le phénomène de latch-up. Pour cela, des simulations ont été réalisées au moyen de la simulation physique bidimensionnelle. La structure utilisée est représentée dans la figure 5.18. Un anneau de test supplémentaire constitué par une diffusion $N^{++}N$ est ajouté à proximité de la diode D1. Cet anneau, connecté à la ligne d'alimentation VDD, est situé à une distance de $25\mu\text{m}$ de l'injecteur constitué par la diffusion N^{++} de cathode de la diode D1. Cette

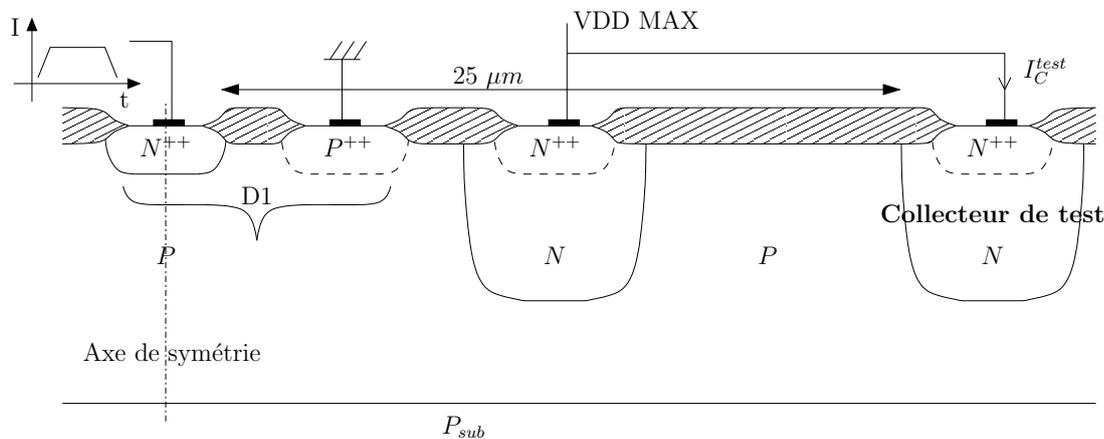


FIG. 5.18 – Représentation schématique de la diode D1 et de son anneau de protection contre le latch-up ainsi que le collecteur de test utilisé pour la simulation d'un test de latch-up.

structure permet de comparer le courant collecté par l'anneau de test dans le cas de l'utilisation de l'anneau de protection contre le latch-up original et de l'anneau ballasté. Pour réaliser le test, la tension d'alimentation est portée à la plus haute valeur autorisée par la technologie. Une impulsion de courant est appliquée sur la broche d'entrée. Son amplitude est de 100 mA, sa durée de 500 ms, et ses temps de montée et de descente de $5\mu\text{s}$. Une moitié seulement de la structure est simulée car elle possède un axe de symétrie. La diode D1 se trouve ainsi polarisée en direct et le courant injecté correspond à celui défini dans la norme de test de latch-up JEDEC [155].

Le courant collecté par l’anneau de test est de 2,5 mA avec l’anneau de protection original et de 1,6 mA avec sa version ballasté. L’ajout d’une résistance de ballast par l’accroissement de la distance W_{bal} a donc un effet positif sur l’efficacité de l’anneau de protection. En effet, la largeur de l’anneau étant augmentée, sa surface de collection est accrue. Cette modification n’a donc que des effets bénéfiques en termes de robustesse ESD et d’immunité au latch-up. La surface supplémentaire demandée n’est pas significative et la capacité parasite de la diode de protection n’est pas modifiée.

L’espace occupé et la dissymétrie du comportement des transistors parasites NPN et PNP peuvent être mis à profit pour réaliser une structure optimisée en termes de robustesse, d’espace et d’immunité au latch-up. En effet, les deux diodes de protection utilisées sur les entrée/sortie sont généralement très proches sur la puce. L’espace créé pour réaliser la résistance de ballast dans le collecteur du transistors parasite NPN peut être utilisé pour intégrer la diode P^{++}/N comme cela est représenté dans la figure 5.19.

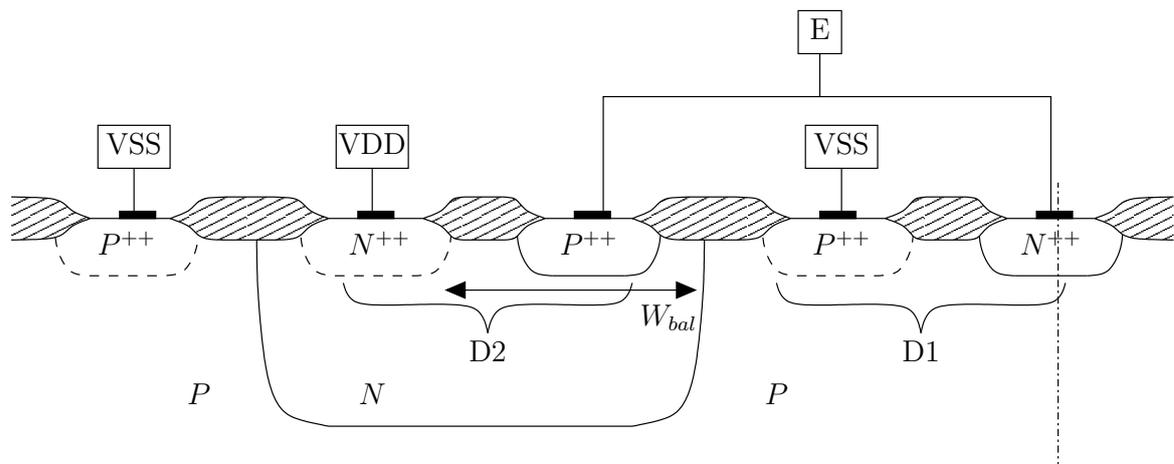


FIG. 5.19 – Coupe schématique de la structure intégrant les deux diodes D1 et D2.

L’intégration des deux diodes permet donc de ne consommer aucun espace supplémentaire. L’immunité au latch-up est améliorée, la robustesse HBM garantie, la capacité et le courant de fuite parasites des diodes minimisés vis-à-vis des performances obtenues.

5.4 Cas particulier pour le stress MM

5.4.1 Résultats expérimentaux

Une version optimisée du réseau de protection élémentaire qui vient d’être étudié a été testée pour réaliser la protection complète d’un circuit de conversion DC–DC possédant 6 entrées/sorties.

Pour l’ensemble du circuit, la robustesse de 2 kV HBM a été obtenue. Des tests complémentaires réalisés avec le modèle de décharge MM montre que ce circuit possède une robustesse supérieure à 200 V MM excepté pour deux configurations de test. Ce comportement inattendu a permis de mettre en évidence un nouveau phénomène qui limite la performance du réseau de protection global.

Après extraction des résistances des bus d’alimentation, le schéma partiel (trois des six entrées/sorties) du réseau de protection est représenté dans la figure 5.20. Les transistors bipolaires parasites des diodes de protection, qui sont hors de cause ici, ne sont pas représentés afin de simplifier le schéma.

Les configurations défailtantes correspondent à des stress MM négatifs sur les broches E2 et E3, la broche VSS servant de référence. La robustesse dans le cas du stress E3/VSS est de

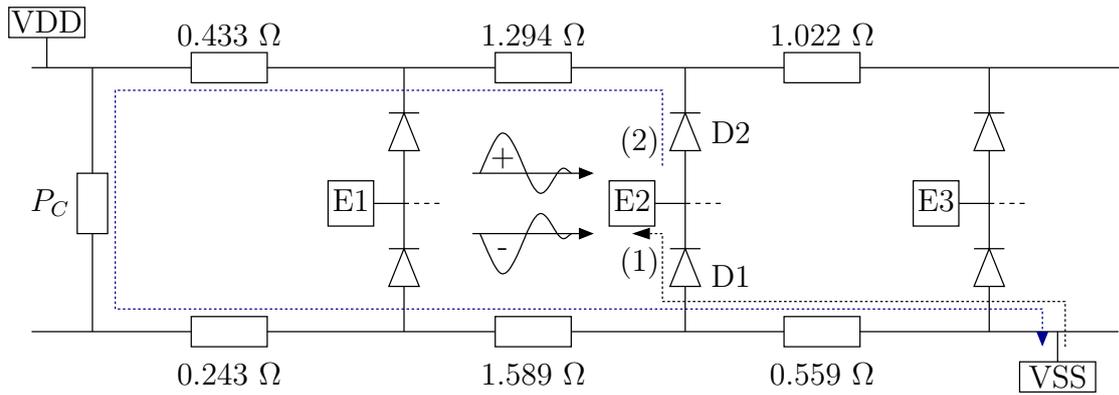


FIG. 5.20 – Schéma équivalent du réseau de protection ESD du circuit. Chemins de décharge lors d'un stress MM entre les broches E2 et VSS.

-170 V MM et de -200 V MM pour le stress E2/VSS. Pour chacun des deux cas, la robustesse pour un stress positif est supérieure à 200 V MM. Les analyses de défaillance montrent que la diode D1 est dégradée. De même, la diode connectée entre E3 et VSS est détériorée lors du stress MM négatif E3/VSS.

5.4.2 Origine du problème

Nous pouvons supposer deux origines possibles de la dégradation de ces diodes. La diode peut être détruite soit lorsqu'elle conduit en direct soit lorsque la tension inverse qui lui est appliquée dépasse sa tension de claquage. La première hypothèse ne peut pas être retenue car la robustesse de la diode en direct pour un stress HBM est de 8 kV. Le courant maximum et l'énergie dissipée lors d'un stress MM de 200 V sont bien inférieurs à ceux du stress HBM maximum qu'elle peut supporter. La seconde hypothèse n'est pas plus valide. Au cours d'une décharge MM, le courant change de sens. Pour le stress négatif considéré ici, la diode se trouve polarisée en inverse au cours de la seconde alternance. Le courant circule alors en empruntant le chemin (2) (Fig. 5.20). Le potentiel inverse appliqué sur la diode D1 correspond à la somme des chutes de potentiel sur ce chemin. Cette seconde alternance possède une amplitude égale aux deux tiers de la première. Or, au cours de la première alternance d'une décharge MM positive entre E2 et VSS le courant emprunte ce même chemin. Le maximum de tension appliqué sur la diode D1 est donc bien supérieur lors d'une décharge positive de 200 V MM qui ne conduit pourtant à aucune dégradation du circuit.

En réalité, la cause de la défaillance provient du phénomène d'avalanche dynamique lié au recouvrement inverse de la diode [156, 157]. Ce phénomène est plus particulièrement connu dans les diodes utilisées pour l'électronique de puissance. Il limite la tenue en tension des diodes lors de leur passage de l'état passant à l'état bloqué. Si ce passage est rapide, la tension inverse maximum que peut supporter une diode est inférieure à sa tension de claquage statique.

Dans le cas du stress MM négatif entre E2 et VSS, la diode D1 est polarisée en direct au cours de la première alternance de la décharge. Le très fort courant qui circule entraîne un très fort niveau d'injection de porteurs dans la diode (Fig. 5.21(a)). Au cours de la seconde alternance, la diode se trouve brutalement polarisée en inverse. Une grande quantité d'électrons et de trous, qui n'ont pas eu le temps de disparaître par recombinaisons, sont stockés dans le substrat. L'établissement de la tension inverse dans la diode correspond à l'apparition d'un champ électrique dans sa jonction. Les électrons libres présents dans le substrat vont être happés par ce champ (Fig. 5.21(a)), créant ainsi un courant d'électrons dans la région de fort champ électrique. Le facteur de multiplication du courant d'électrons étant significatif, bien avant d'atteindre la tension de claquage de la jonction, le courant de conduction qui circule dans la diode est très

important. Sous l'effet des fortes densités de courant, la diode polarisée en inverse présente alors

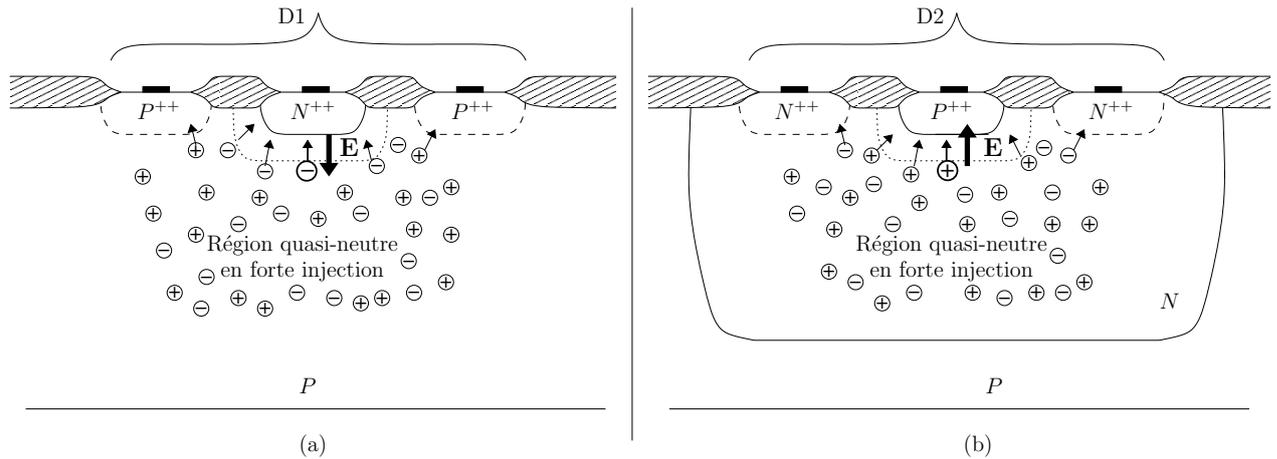


FIG. 5.21 – État interne d'un diode lors du recouvrement inverse, après une forte injection de courant direct, dans la diode D1 de type N^{++}/P (a) et dans la diode D2 de type P^{++}/N (b).

une caractéristique à résistance dynamique négative [158]. Une grande partie du courant de décharge va alors emprunter la diode polarisée en inverse car elle constitue un chemin de plus faible impédance. La formation d'un filament de courant, associé à la caractéristique de résistance négative, conduit rapidement à la destruction thermique de la diode [159]. Ce phénomène explique donc l'origine de la plus faible robustesse observée dans ce cas. La focalisation du courant dans la diode, n'est pas d'origine thermique mais purement électrique [160]. Comme nous avons déjà pu le montrer pour les TBA dans les chapitres précédents, les phénomènes physiques associés aux très fortes densités de courant jouent un rôle primordial. Les aspects thermiques ne jouent qu'un rôle secondaire, bien qu'important, dans la physique des composants soumis aux ESD.

Remarquons enfin que lors d'un stress MM positif entre E2 et VDD, la diode D2 se trouve polarisée dans les mêmes conditions. Pourtant, aucune diminution de la robustesse n'est observée. En effet, dans ce cas, ce sont les trous stockés dans la diode qui sont happés par le champ électrique (Fig. 5.21(b)). Le facteur de multiplication du courant de trous n'étant important que pour des tensions proches de la tension de claquage statique, la multiplication par avalanche dans la jonction est peu importante et aucun phénomène d'emballement ne se produit.

5.4.3 Solution et recommandations

La dégradation de la robustesse dans le cas de stress MM négatifs entre une entrée ou une sortie et la broche VSS est évidemment d'autant plus importante que la tension inverse appliquée sur la diode est importante. Pour cette raison, la valeur des résistances des lignes d'alimentation joue un rôle primordial. Le cas de plus faible robustesse (-170 V MM) correspond à la broche E3 (Fig. 5.20). pour laquelle la résistance totale des lignes d'alimentation est maximale. La robustesse s'accroît pour la broche E2 (-200 V MM) et ne pose plus de problème pour la broche E1. Dans ce cas particulier, la modification des pistes métalliques de façon à réduire les résistances des lignes d'alimentation a permis d'obtenir la robustesse attendue.

Le problème qui vient d'être exposé montre une nouvelle fois que l'optimisation des protections intégrées contre les décharges électrostatique doit être menée à la fois du point de vue des composants de protection et du réseau de protection qu'ils vont former. Pour étudier ce dernier aspect, l'optimisation de réseaux élémentaires est un outil très utile. Il permet à la fois l'étude des problèmes liés aux anneaux de protection contre le latch-up et peut également être utilisé pour optimiser le réseau de protection vis-à-vis du recouvrement inverse des diodes lors d'une décharge MM. Pour cela, il apparaît très important d'exprimer les performances du réseau de

protection élémentaire en terme de robustesse en fonction de la valeur des résistances des lignes d'alimentation.

Le réseau élémentaire ainsi optimisé et dont les performances sont définies vis-à-vis des conditions d'utilisation, permettra de garantir le succès de la stratégie de protection utilisée pour un circuit, dès les premiers essais expérimentaux. Associé à la simulation de type SPICE, qui permet de vérifier l'absence d'interaction du réseau de protection ESD avec le circuit à protéger, cette approche a l'ambition de garantir les performances des solutions de protection fournies aux concepteurs de circuits intégrés.

5.5 Conclusion

Après un rapide état de l'art des stratégies de protection contre les décharges électrostatiques utilisées pour les circuits intégrés, l'étude d'un réseau de protections d'un circuit simple a été présentée. L'utilisation des macromodèles des composants de protection dans les simulations de type SPICE a permis de vérifier l'efficacité de la stratégie de protection utilisée. Les chemins prévus par simulation ont été confrontés avec succès à ceux déterminés expérimentalement grâce au couplage de l'analyse par photoémission et du banc de mesure TLP.

Cependant, nous avons mis en évidence que le développement de composants de protection optimisés et l'utilisation de la simulation SPICE ne sont pas suffisants pour garantir la robustesse ESD d'un circuit. En effet, l'interaction des composants de protection au sein du réseau de protection peut provoquer une diminution significative des performances.

Une faiblesse associée aux anneaux de protection contre le latch-up a pu être mise en évidence grâce à la performance de la localisation de défaillance par la technique de stimulation thermique laser OBIRCH. Les anneaux de protection contre le latch-up placés autour de composants de protection forment des transistors bipolaires parasites de type NPN ou PNP. Dans certaines configurations de décharge, ces composants peuvent être activés et détruits. Nous avons pu mettre en évidence et expliquer le comportement sensiblement différent des transistors bipolaires parasites de type NPN et PNP. La solution proposée grâce à cette étude permet de résoudre ce problème en modifiant simplement la géométrie de l'anneau de latch-up. L'utilisation astucieuse des résultats de cette étude a conduit à l'intégration des deux diodes de protection des entrées/sorties au sein d'une seule structure. Elle présente l'avantage de ne pas modifier les caractéristiques du composant de protection (capacité parasite, courant de fuite), de garantir la robustesse HBM et d'améliorer l'immunité au latch-up, sans demander de surface supplémentaire.

Nous avons également mis en évidence une configuration de stress MM dans laquelle le réseau de protection possède une robustesse moindre. L'origine de cette faiblesse correspond à un phénomène bien connu pour les diodes de l'électronique de puissance, associé aux fortes densités de courant et au mécanisme de recouvrement inverse des diodes. Lors de la commutation rapide d'un état passant, à très forte densité de courant, à l'état bloqué, le phénomène d'avalanche dynamique induit une diminution de la tenue en tension des diodes par rapport à leur tenue en tension statique. Cet effet est beaucoup plus marqué pour les diodes de type N^{++}/P que celles de type P^{++}/N . Dans la cas étudié, la diminution des résistances des lignes d'alimentation a permis de résoudre ce problème.

Afin de garantir les performances d'une stratégie de protection, la méthode de conception basée sur le développement de composants de protection et l'utilisation de la simulation SPICE doit être complétée par l'optimisation de réseaux de protections élémentaires. Ces réseaux permettent de caractériser et d'étudier les problèmes liés aux anneaux de protection de latch-up et à certaines configurations de stress MM. Un ensemble de réseaux de protections élémentaires, dont les performances sont définies en fonction de la résistance des lignes d'alimentation, est alors fourni aux concepteurs de circuit.

L'ambition de cette méthode est de garantir le succès d'une stratégie de protection choisie

par le concepteur dès les premiers essais expérimentaux. Elle contribue ainsi à réduire le nombre de cycles de conception liés aux problèmes de la protection contre les décharges électrostatiques.

Conclusion générale

Dans le contexte du besoin d'amélioration de la compétitivité et de l'augmentation des problèmes liés aux décharges électrostatiques, les fabricants de semiconducteur doivent se doter d'outils et de méthodologies performants pour la conception et l'optimisation des protections ESD intégrées sur les puces. Il est aujourd'hui nécessaire de prendre en compte au plus tôt la robustesse aux décharges électrostatiques dans le développement de nouvelles technologies et de nouveaux produits.

Les travaux présentés dans ce mémoire font suite aux études des thèses antérieures de C. Delage et G. Bertrand qui ont tour à tour traité de la conception de structures de protection et de leur modélisation [60, 61]. Ils viennent les compléter en amenant une réflexion plus approfondie sur les phénomènes physiques associés aux fortes densités de courant et en analysant plus complètement le fonctionnement des réseaux de protection ESD.

Reprenant les méthodologies proposées aux termes des thèses précédentes, en ajoutant une part plus importante aux possibilités offertes par les techniques de localisation de défaillance, ce mémoire s'attache plus particulièrement à l'étude et l'optimisation de structures de protection basées sur les transistors bipolaires NPN autopolarisés (TBA) et des réseaux de protection ESD des circuits.

Étant donnée la forte intensité du courant lors d'une décharge électrostatique et la relativement faible dimension des structures de protection, les densités de courant mises en jeu dans les composants sont bien plus importantes que dans les composants classiques. L'étude des TBA a donc été menée en s'appuyant sur les connaissances issues de la physique des transistors bipolaires de puissance et des transistors bipolaires hautes fréquences. Inspirée de celle utilisée pour les composants de puissance, une approche régionale unidimensionnelle adaptée au cas de TBA a permis d'étudier les phénomènes physiques spécifiques aux fortes densités de courant dans ces composants. La principale difficulté de la modélisation réside dans la description de la région électrique de collecteur qui est le siège de génération par avalanche. L'étude montre que dans les régimes de fort courant, la tension de claquage de la jonction collecteur-base des transistors dépend de la densité de courant. Elle explique également la différence de comportement des transistors bipolaires NPN et PNP en fonction de la profondeur de leur base. La diminution de la tension de claquage de la jonction collecteur-base des transistors NPN pour des densités de courant croissantes a conduit à une explication originale de la focalisation du courant observée dans les transistors NPN et démontré l'importance d'introduire une résistance de ballast dans leur collecteur. Le modèle semi-analytique a également permis d'étudier l'impact de l'échauffement des régions d'émetteur et de collecteur. Alors que la région de collecteur possède des propriétés favorables pour éviter l'apparition du second claquage thermique, la région d'émetteur a un comportement inverse. Il apparaît ainsi que l'éloignement spatial de la région chaude de collecteur de la région d'émetteur doit permettre d'obtenir des composants plus robustes aux décharges électrostatiques.

Après cette première approche, le principe d'un nouveau type de composant de type TBA NPN à collecteur graduel a pu être proposée, puis étudiée d'une manière plus approfondie au moyen de la simulation physique. Différents régimes de fonctionnement ont été mis en évidence en fonction de l'intensité du courant. Ces régimes sont liés à l'évolution des caractéristiques électriques de la jonction collecteur-base qui sont modifiées par la charge électrique des porteurs

libres due au passage du courant. Les valeurs des courants et des tensions caractéristiques et l'allure de la caractéristique TLP ont pu être associées aux paramètres géométriques et technologiques du composant. La tension et le courant de repliement sont respectivement liées à la profondeur et au dopage de la couche faiblement dopée du collecteur. La tension de maintien dépend des caractéristiques de la jonction effective formée à forte densité de courant à la frontière des régions faiblement et fortement dopées de collecteur. Nous avons montré que le profil de dopage de cette frontière détermine directement cette tension. Plus ce profil est abrupt, plus la tension de maintien sera basse. La formation de la jonction effective, dont la tension de claquage décroît pour des densités de courant croissantes, est également la cause de la focalisation du courant à forte densité de courant. À plus faible densité de courant, le gradient de potentiel dans la résistance intrinsèque de base est également dans une moindre mesure à l'origine d'une non-uniformité.

La robustesse de ces composants réside dans leur capacité à retarder l'apparition d'un second claquage thermique. Nous avons montré par simulation que malgré l'inévitable focalisation du courant, la diminution des coefficients de multiplication par avalanche avec la température permet d'uniformiser la dissipation thermique dans le composant. Lors d'une décharge de faible intensité, cette propriété provoque le déplacement de la région focalisée de courant dans les simulations réalisées. Ce mouvement difficilement observable expérimentalement étant donnée l'échelle de temps, a été mis en évidence très récemment pour des composants similaires par D. Pogany et son équipe grâce à des techniques d'interférométrie laser. Afin de retarder l'apparition d'un second claquage thermique, l'augmentation de la profondeur de base effective est apparue comme une excellente solution pour découpler à la fois thermiquement et électriquement les régions d'émetteur et de collecteur au cours d'une décharge. Les résultats de simulation ont confirmé cette hypothèse. Cette étude, basée sur l'utilisation de la simulation physique, a finalement permis de définir un ensemble de règles universelles de conception de TBA performants.

Ces règles ont été appliquées à la réalisation de composants pour trois technologies distinctes. Les résultats obtenus viennent confirmer et illustrer le bien fondé de l'étude théorique. L'utilisation des techniques de localisation de défauts reposant sur la stimulation photoélectrique laser de type OBIC ont permis de confirmer le mécanisme de défaillance. L'impossibilité de réaliser des structures équivalentes avec des transistors bipolaires PNP a également été confirmée expérimentalement. La principale limitation pour l'utilisation des composants réalisés est l'importante valeur de la tension de repliement.

La structure TBA optimisée a été utilisée avec succès dans une structure de protection d'entrée à deux étages dans un circuit commercial. Afin d'étendre le domaine d'utilisation de ces composants, des solutions de déclenchement pour diminuer la tension de repliement ont été également proposées et validées expérimentalement.

La prise en compte des effets des fortes densités de courant sur le comportement de la jonction collecteur-base a été abordée dans le cadre de la modélisation de type SPICE. Une nouvelle technique de calcul du facteur de multiplication du courant dans la région de collecteur est proposée. Elle permet de conserver une excellente précision des calculs et élimine les problèmes de convergence couramment observés. L'approche modulaire retenue pour son efficacité vis-à-vis de la réduction du travail d'extraction des paramètres, permet d'introduire les effets associés aux fortes densités de courant mais souffre de quelques limitations. Le modèle développé présente cependant l'avantage d'être étroitement lié à la structure physique du composant et offre donc la possibilité de mieux comprendre l'impact de certains paramètres. En dernier lieu, la modélisation des composants a été utilisée pour proposer une méthode d'optimisation de la technique de déclenchement par couplage de grille des transistors MOS. Cette méthode est développée après une analyse de la dynamique de déclenchement des transistors bipolaires et MOS. L'étude a également conduit à une amélioration du macromodèle pour tenir compte, au premier ordre, des effets distribués des courants capacitifs.

Enfin, l'efficacité de la simulation de type SPICE pour étudier les stratégies de protection ESD et prédire les chemins de décharges est démontrée en la confrontant aux résultats expé-

rimementaux. Toutefois, nous avons montré sur un cas pratique, que la robustesse attendue pour le circuit n'est pas toujours obtenue. La technique de localisation de défaillance OBIRCH a été mise à profit pour déterminer l'origine du problème qui provient des anneaux de protection des composants ESD contre le phénomène de latch-up. L'étude de réseaux de protection élémentaires par la simulation physique en mode mixte a permis d'expliquer et de proposer des solutions optimales en termes de robustesse ESD, d'immunité au latch-up, de surface, de réduction des capacités et des courants de fuite parasites. En outre, l'origine d'une faiblesse du réseau de protection lié à une configuration de stress MM a pu être expliquée.

L'optimisation de composant de protection et l'utilisation de la simulation SPICE ne sont pas suffisant pour garantir la robustesse de la protection globale d'un circuit aux ESD. Ces travaux montrent que la conception des structures de protection doit être complétée par l'étude de réseaux de protection élémentaires. Ils permettent d'étudier et de valider les protections en prenant en compte la protection contre latch-up et les spécificités du test MM. Afin de garantir les performances des protections ESD fournies aux concepteurs de circuit, un ensemble de réseaux de protection élémentaires doivent être proposés. La performance des réseaux devra être définie en fonction de contraintes d'utilisation comme la résistance maximale des lignes d'alimentation utilisées.

Nous proposons pour finir une méthodologie de développement de bibliothèque ESD dont la représentation schématique est donnée dans la figure 22

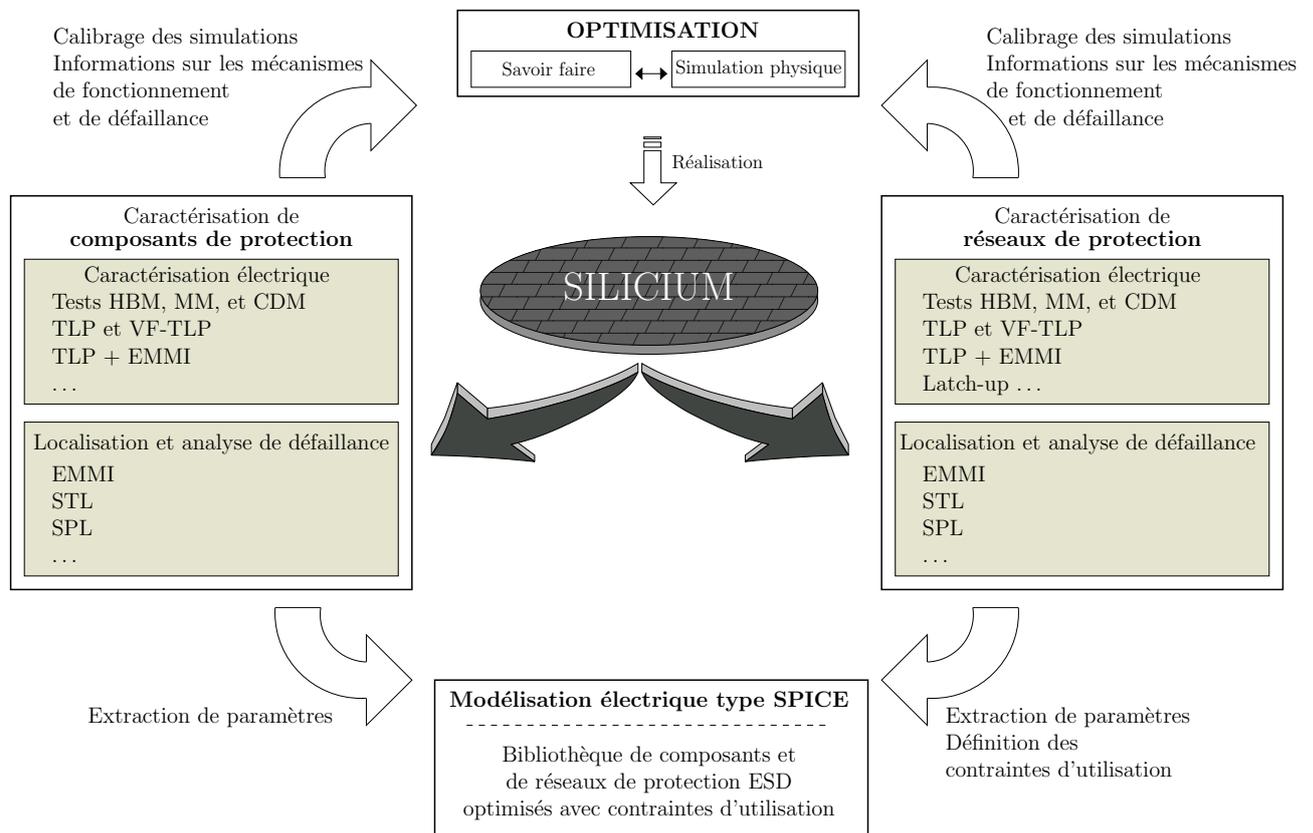


FIG. 22 – Méthodologie de développement de protections ESD intégrées.

Au terme de cette étude, diverses perspectives de travaux peuvent être envisagées :

- Les contraintes d'utilisation des réseaux de protection élémentaires restent à définir plus précisément.
- D'autres aspects comme la susceptibilité électromagnétique pourrai être considérés, pour évoluer vers une approche plus globale du traitement des agressions des circuits intégrés.

- L'utilisation de la modélisation distribuée basée sur une approche régionale semble très prometteuse pour décrire à la fois les effets des fortes densités de courant et le comportement dynamique. Elle apparaît très adaptée pour la modélisation des composants de protection ESD.
- Le développement d'outils et de méthodes d'extraction des paramètres dynamiques adaptés à l'échelle de temps et de courant des ESD permettrait de donner des informations fiables sur la dynamique des composants. Les acquisitions TLP ne sont pas pour le moment exploitables car les signaux sont déformés par des éléments parasites mal maîtrisés et caractérisés : boîtier, connexions, ...
- Des efforts importants seront encore nécessaires pour étendre le domaine de validité des modèles utilisés pour la simulation physique et permettre son utilisation pour réduire la part des tests expérimentaux.
- La physique des TBA met en jeu la physique des systèmes très éloignés de l'équilibre thermodynamique. Des progrès importants pourraient découler de l'utilisation des outils d'analyse développés pour l'étude des phénomènes non linéaires et du chaos.

Bibliographie

- [1] K. BOCK : ESD issues in compound semiconductor high-frequency devices and circuits. *Microelectronics Reliability*, 38:1781–1793, 1998.
- [2] C. RICHIER, P. SALOME, G. MABBOUX, I. ZAZA, A. JUGE et P. MORTINI : Investigation on different ESD protection strategies devoted to 3,3 V RF applications (2 Ghz) in a 0,18 μm CMOS process. Dans *EOS/ESD Symposium*, pages 251–259, 2000.
- [3] G. MENEGHESSO, A. CHINI, A. MASCHIETTO, E. ZANONI, P. MALBERTI et M. CIAPPA : Electrostatic discharge and electrical overstress on GaN/InGaN emitting diodes. Dans *EOS/ESD Symposium*, pages 249–254, septembre 2000.
- [4] A. BALDI, A. BRATOV, R. MAS et C. DOMÍNGUEZ : Electrostatic discharge sensitivity tests for ISFET sensors. *Sensors and Actuators B : Chemical*, 80(3):255–260, 2001.
- [5] J. A. WALRAVEN, J. M. SODEN, E. I. COLE JR., D. M. TANNER et R. E. ANDERSON : Human body model, machine model, and charged device model ESD testing of surface micromachined microelectromechanical systems (MEMS). Dans *EOS/ESD Symposium*, Portland, Oregon, 11 septembre 2001.
- [6] A. AMERASEKERA et C. DUVVURY : *ESD in silicon integrated circuits*. 1995.
- [7] J. E. VINSON et J. J. LIOU : Electrostatic discharge in semiconductor devices : an overview. *Proceedings of the IEEE*, 86(2):399–418, février 1998.
- [8] R. FEYNMAN, R. LEIGHTON et M. SANDS : *Les cours de physique de Feynman, électromagnétisme 1*, chapitre 9, pages 146–164.
- [9] ESD JOURNAL : <http://www.esdjournal.com/>.
- [10] C. DUVVURY et A. AMERASEKERA : ESD : A pervasive reliability concern for IC technologies. *Proceedings of the IEEE*, 81(5):690–702, avril 1993.
- [11] A.Z. WANG, H.G. FENG, K. GONG, R.Y. ZHAN et J. STINE : On-chip ESD protection design for integrated circuits : an overview for IC designers. *Microelectronics Journal*, 32:733–747, 2001.
- [12] J.C. LEE, G.D. CROFT, J.J. LIOU, W.R. YOUNG et J. BERNIER : Modeling and measurement approaches for electrostatic discharge in semiconductor devices and ICs : an overview. *Microelectronic Reliability*, 39:579–593, 1999.
- [13] JIUSHENG H., QIBIN D., FANG L., ZHENGXIN C. et PEIZHU L. : Electromagnetic field generated by transient electrostatic discharge (ESD) from person charged with low electrostatic voltage. Dans *EOS/ESD symposium*, pages 415–418, septembre 2001.
- [14] AUTOMOTIVE ELECTRONIC CONCIL : Human Body Model Electrostatic Discharge test, AEC-Q100-002-REV-C, 1998. 12p.
- [15] EIA/JEDEC STANDARD : Electrostatic Discharge (ESD), Sensitivity Testing Human Body Model (HBM), EIA/JEDEC-A114A, 1997. 9p.
- [16] ESD ASSOCIATION : Association Standard test Method for Electrostatic Discharge Sensitivity Testing – Human Body Model (HBM) Component level, ESD STM5.1-1998, 1998. 13p.

-
- [17] AGNÈS GUILHAUME : *Évaluation de la robustesse de circuits intégrés vis-à-vis des décharges électrostatiques*. Thèse de doctorat, Institut National des Sciences Appliquées de Lyon, 20 septembre 2002.
- [18] AUTOMOTIVE ELECTRONIC CONCIL : Machine Model Electrostatic Discharge test, AEC-Q100-003-REV-C, 1998. 11p.
- [19] EIA/JEDEC STANDARD : Electrostatic Discharge (ESD), Sensitivity Testing Machine Model (MM), EIA/JEDEC-A115A, 1997. 9p.
- [20] ESD ASSOCIATION : Association Standard test Method for Electrostatic Discharge Sensitivity Testing – Machine Model (MM) Component level, ESD STM5.2-1999, 1999. 11p.
- [21] EIA/JEDEC STANDARD : Field induced Charged-Device Model, Test method for Electrostatic Discharge withstand thresholds of microelectronics components, JESD22-C101, 1995. 7p.
- [22] ESD ASSOCIATION : Association Standard test Method for Electrostatic Discharge Sensitivity Testing – Charged Device Model (CDM) Component level, ESD STM5.3-1999, 1999. 12p.
- [23] H. GIESER et AL. : Survey on electrostatic susceptibility of integrated circuits. Dans *ESREF*, pages 447–456, 1994.
- [24] N. GUITARD, D. TRÉMOUILLES, S. ALVES, M. BAFLEUR, F. BEAUDOIN, P. PERDU et A. WISLEZ : ESD Induced Latent Defects In CMOS ICs And Reliability Impact. Dans *EOS/ESD Symposium*, 2004.
- [25] H. A. GIESER, P. EGGER, J. C. REINER et M. R. HERRMANN : A CDM Only reproducible field degradation and its reliability aspect. *Quality and reliability engineering international*, 10:341–350, 1994.
- [26] T. MALONEY et N. KHURANA : Transmission line pulsing techniques for circuits modelling of ESD phenomena. Dans *7th EOS/ESD Symposium*, pages 49–55, Minneapolis, MN, 1985.
- [27] N. MAURAN : Conception et réalisation d'un banc de caractérisation sous pointes pour mesures impulsionnelles haute énergie, 2 juillet 2003. Mémoire de diplôme ingénieur C.N.A.M.
- [28] D. G. PIERCE, W. SHILEY, B. D. MULCACHY, K. E. WARNER et M. WUNDER : Electrical overstress testing of 256K UVEPROM to rectangular and double exponential pulses. Dans *10th EOS/ESD Symposium*, pages 137–146, Anaheim, CA, octobre 1998.
- [29] A. AMERASEKERA et C. DUVVURY : The impact of technology scaling on ESD robustness and protection circuit design. Dans *16th EOS/ESD Symposium*, pages 237–245, Las Vegas, NV, 1994.
- [30] J. M. LUCHIES : *Electrostatic discharge in integrated circuits ; testing and protection*. Thèse de doctorat, Université de Twente, Pays-Bas, 1995.
- [31] C. MUSSHOF et AL : Risetime effects of HBM and square pulses on the failure thresholds of GGNMOS transistors. *Microelectronics Reliability*, 36(11/12):1743–1746, 1996.
- [32] C. RICHIER et AL : Study of the ESD behavior of different clamp configurations in a 0,35um CMOS technology. Dans *EOS/ESD Symposium*, pages 240–245, 1997.
- [33] W. STADLER et AL : Does the ESD-failure current obtained by transmission-line pulsing always correlate to human body model tests. Dans *EOS/ESD Symposium*, pages 366–372, 1997.
- [34] M. KELLY, G. SERVAIS, T. DIEP, D. LIN, S. TWEREFOR et G. SHAH : A comparison of electrostatic discharge models and failure signatures for CMOS integrated circuits devices. Dans *Proceedings of 17th EOS/ESD Symposium*, pages 175–185, Phoenix, AZ, 1995.
- [35] G. NOTERMANS, P. DE JONG et F. KUPPER : Pitfalls when correlating TLP, HBM and MM testing. Dans *20th EOS/ESD Symposium*, pages 170–176, Reno, NV, 1998.

-
- [36] H. GEISER et M. HAUNSCHILD : Very-fast transmission line pulsing of integrated structures and the charged device model. Dans *EOS/ESD Symposium*, pages 85–94, 1996.
- [37] J. KÖLZER, A. DALLMANN, G. DEBOY, J. OTTO et D. WEINMANN : Emission microscopy. *European Symposium on Reliability of Electron device Failure physics and analysis*, 2:625–649, octobre 1991.
- [38] C. LEROUX et D. BLACHIER : Light emission microscopy for reliability studies. *Microelectronics reliability*, 49:169–180, 1999.
- [39] M. CAVONE et AL : A method for the characterization and evaluation of ESD protection structures and networks. Dans *EOS/ESD Symposium*, pages 292–300, 1994.
- [40] CHRISTIAN RUSS, KARLHEINZ BOCK, MAHMOUD RASRAS, INGRID DE WOLF, GUIDO GROESENEKEN et HERMAN E. MAES : Non-uniform triggering of gg-nMOS_t investigated by combined emission microscopy and transmission line pulsing. Dans *EOS/ESD Symposium*, pages 177–186, 1998.
- [41] DAVID TRÉMOUILLES, NICOLAS GUITARD, MARISE BAFLEUR, NICOLAS NOLHIER et LIONEL LESCOUZERES : TLP and photo emission coupling, a powerful tool for study of ESD protection strategy. Dans *Workshop EOS/ESD/EMI*, 13 octobre 2002.
- [42] J. COLVIN : ESD failure analysis methodology. *Microelectronic Reliability*, 38:1705–1714, 1998.
- [43] F. BEAUDOIN, A. WISLEZ, M. BAFLEUR, D. LEWIS, R. DESPLATS, P. PERDU et D. TRÉMOUILLES : Laser beam based ESD defect localization in ICs. Dans *28th International Symposium on Testing and Failure Analysis (ISTFA '2002)*, Phoenix (USA), 3 novembre 2002.
- [44] T. BEAUCHÊNE, D. LEWIS, F. BEAUDOIN, V. POUGET, R. DESPLATS, P. FOUILLAT, P. PERDU, M. BAFLEUR et D. TRÉMOUILLES : Thermal laser stimulation and NB-OBIC techniques applied to ESD defect localization. Dans *XIIIth European Symposium on Reliability of Electron Devices, Failure Physics and Analysis (ESREF'2002)*, Bellaria (Italie), 7 octobre 2002.
- [45] K.S. WILLS, C. DUVVURY et O. ADAMS : Photoemission testing for ESD failures. Advantage and limitations. Dans *EOS/ESD Symposium*, pages 53–61, 1988.
- [46] M. HANNEMANN et A. AMERASEKERA : Photo emission as a tool for ESD failure localization and as a technique for studying ESD phenomena. Dans *European Symposium on Reliability of Electron device Failure physics and analysis*, pages 77–84, 1990.
- [47] F. BEAUDOIN : Thèse de doctorat, Université de Bordeaux 1, 2002.
- [48] T. BEAUCHÊNE : *Étude et localisation de défauts dans les circuits intégrés par stimulation photoélectrique laser*. Thèse de doctorat, Université Bordeaux 1, 2 avril 2004.
- [49] K. NIKAWA et S. TOZAKI : Testing and failure analysis. *ASM International*, pages 303–310, 1993.
- [50] K. NIKAWA, C. MATSUMOTO et S. INOUE : Verification and improvement of the optical beam induced resistance change (OBIRCH) method. Dans *ISTFA*, 1996.
- [51] K. NIKAWA et S. INOUE : LSI failure analysis using focused laser beam heating. *Microelectronic Reliability*, 37(12):1841–1847, 1997.
- [52] D. L. BARTON, K. BERNHARD-HÖFER et E. I. COLE JR. : FLIP-chip and "backside" techniques. *Microelectronics Reliability*, pages 721–730, 1999.
- [53] E. I. COLE JR. et AL. : Resistive interconnection localization. Dans *ISTFA*, 2001.
- [54] S. ITO et H. MONNA : Failure analysis of wafer using backside OBIC method. *Microelectronics Reliability*, 38:993–996, 1998.

- [55] T. WILSON et C.J.R. SHEPPARD : Observations of dislocations and junction irregularities in bipolar transistors using the OBIC mode of the scanning optical microscope. *Solid-State Electronics*, 29(11):1189–1194, 1986.
- [56] E. I. COLEJR., J. M. SODEN, J. L. RIFE, D. L. BARTON et C. L. HENDERSON : Novel failure analysis techniques using photon probing with a scanning optical microscope. Dans *IRPS*, pages 338–398, 1994.
- [57] ROMAIN DESPLATS, FELIX BEAUDOIN, PHILIPPE PERDU, P. POIRIER, DAVID TRÉMOUILLES, MARISE BAFLEUR et DEAN LEWIS : Backside localization of current leakage faults using thermal laser stimulation. *Microelectronics Reliability*, 41(9-10):1539–1544, octobre 2001. 12th European Symposium on Reliability of Electron Devices, Failure Physics and Analysis (ESREF'2001).
- [58] P. SALOMÉ : *Etude des décharges électrostatiques dans les circuits MOS submicroniques et optimisation de leur protections*. Thèse de doctorat, Institut National des Sciences Appliquées de Lyon, 22 avril 1998.
- [59] C. RUSS : ESD protection devices for CMOS technologies : Processing impact, modeling, and testing issues. Thèse de docteur-ingénieur de l'Université de Munich, Allemagne, 1999.
- [60] CHRISTELLE DELAGE : *Étude et conception de structures de protection contre les décharges électrostatiques en technologie BICMOS de puissance*. Thèse de doctorat, Institut National Polytechnique de Toulouse, 13 décembre 1999.
- [61] 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, 20 juillet 2001.
- [62] M.-D. KER, H.-S. HSU et J.-J. PENG : ESD implantation for subquarter-micron CMOS technology to enhance ESD robustness. *IEEE Transaction on Electron devices*, 50(10):2126–2134, octobre 2003.
- [63] T. POLGREEN et A. CHATTERJEE : Improving the ESD failure threshold of silicided nMOS output transistors by ensuring uniform current flow. *IEEE Transaction on Electron Devices*, 39(2):379–388, février 1992.
- [64] PH. LETURCQ et G. REY : *Physique des composants actifs à semiconducteurs*.
- [65] S. ROY MORRISON et R. BILETTE : Common Emitter Breakdown. *IEEE Transaction on Electron Devices*, pages 351–356, novembre 1963.
- [66] S.L. JANG : On the common-emitter breakdown voltage of bipolar junction transistors. *Solid-State Electronics*, 36(2):213–216, 1993.
- [67] S.L. MILLER : Ionization rates for holes and electrons in silicon. *Physical Review*, 105:1246–1249, 15 février 1957.
- [68] W. N. GRANT : Electron and hole ionization rates in epitaxial silicon at high electric fields. *Solid State Electronics*, 16:1189–1203, 1973. Version papier seulement.
- [69] P. PALESTRI, L. SELMI, G.A.M. HURKX, J.W. SLOTBOOM et E. SANGIORGI : Energy dependent electron and hole impact ionization in Si bipolar transistors. Dans *IEDM*, 98.
- [70] G. BOSELLI, C. DUVVURY et V. REDDY : Efficient pnp characteristic of pMOS in Sub-0.13 μm ESD protection circuits. Dans *24th Electrical Overstress/Electrostatic Discharge Symposium (EOS/ESD'02)*, 8 octobre 2002.
- [71] G. BOSELLI : *On high injection mechanisms in semiconductor devices under ESD conditions*. Thèse de doctorat, Université de Twente, Pays-Bas, 2001.
- [72] P. MAGNÉE : SiGe :C HBT device technology. What is the big deal? Dans *BCTM Short Course*, 7, av. du Colonel Roche, 28 septembre 2003. LAAS/CNRS.
- [73] S.M. SZE : *Physics of Semiconductor Devices 2nd Edition*. Wiley-Interscience publication, 1981.

-
- [74] A. S. GROVE : *Physics and Technology of Semiconductor Devices*. 67.
- [75] D.H. PONTIUS : *Current filamentation and second breakdown in thin film silicon-on-sapphire semiconductor devices*. Thèse de doctorat, Auburn University, 1972.
- [76] J. OETGEN AND R. JUNGBLUT AND U. KUHLMANN AND J. ARKENAU AND R. SITTING : Current filamentation in bipolar power devices during dynamic avalanche breakdown. *Solid-State Electronics*, 44:117–123, 2000.
- [77] H.A. SCHAFF et J.C. FRENCH : A survey of second breakdown. *IEEE Transaction on Electron Devices*, ED.13:613–618, 1966.
- [78] PH. LETURCQ : *Comportement électrique et thermique des transistors bipolaires aux forts niveaux de tension et de courant : application au phénomène de second claquage*. Thèse de doctorat, Faculté de Sciences de L’université de Toulouse, 10 octobre 1969.
- [79] J.M. REYNES : *Relations entre performances et paramètres structuraux des transistors bipolaires de puissance. Application à la conception des composants*. Thèse de doctorat, Institut National des Sciences Appliquées, 25 mars 1986.
- [80] H.B. GRUTCHFIELD et T.J. MOUTOUX : Current mode second breakdown in epitaxial planar transistors. *IEEE Transaction on Electron Devices*, ED-13(11):743–748, novembre 1966.
- [81] PHILIP L. HOWER et V. GOPALA KRISHNA REDDI : Avalanche injection and second breakdown in transistors. *IEEE Transactions on electron devices*, ED-17(4):320–335, avril 1970.
- [82] C. T. KIRK : A theory of transistor cutoff frequency (f_t) falloff at high current densities. *IRE Transaction on electron devices*, ED-9:164–174, mars 1962.
- [83] PH. LETURCQ : Physique des semi-conducteurs de puissance. *Techniques de l’ingénieur*, D3102, novembre 1999.
- [84] PH. LETURCQ : Composants semi-conducteurs de puissance bipolaires. Partie 1. *Techniques de l’ingénieur*, (D3106), février 2001.
- [85] PH. LETURCQ : Composants semi-conducteurs de puissance bipolaires. Partie 2. (D3107), mai 2001.
- [86] J.L. DEBRIE : *Modèle "distribué" de transistor IGBT pour simulation de circuits en électronique de puissance*. Thèse de doctorat, Institut National des Sciences Appliquées de Toulouse, 4 octobre 1996.
- [87] H. SCHLANGENOTTO et W. GERLACH : On the effective carrier lifetime in p-s-n rectifier at high injection levels. *Solid State Electronics*, 12:267–275, 1969.
- [88] F. BERZ, R. W. COOPER et S. FAGG : Recombination in the end regions of pin diodes. *Solid State Electronics*, 22:293–301, 1979.
- [89] A. MUNÓZ YAGÜE : Contribution à l’étude du comportement physique des dispositifs semi-conducteurs PIN et PNP. Thèse d’État, Université Paul Sabatier Toulouse, 19 décembre 1977.
- [90] S. SELBERHERR : *Analysis and simulation of semiconductor devices*. 1984.
- [91] H. TRANDUC, P. ROSSEL et J.L. SANCHEZ : Premier et second claquage dans les transistors MOS. *Revue de Physique Appliquée*, 19:859–878, 1984.
- [92] G. CHARITAT : Modélisation et réalisation de composants planar haute-tension. Thèse d’État, Université Paul Sabatier Toulouse, 28 septembre 1990.
- [93] JEAN URGELL : *Effets de multiplication par avalanche dans les jonctions au silicium*. Thèse de doctorat, Faculté des Sciences de l’Université de Toulouse, 10 mai 1969.
- [94] K. ESMARK : *Device simulation of ESD protection elements*, volume 128 de *Series in Microelectronics*. 2002.

- [95] KWANG-HOON OH, KAUSTAV BANERJEE, CHARVAKA DUVVURY et ROBERT W. DUTTON : Non-uniform conduction induced reverse channel length dependence of ESD reliability for silicided NMOS transistors. Dans *I.E.D.M.*, San Francisco, CA, décembre 2002.
- [96] P. SALOMÉ, C. LEROUX, D. MARIOLLE, D. LAFOND, J.P. CHANTE, PH. CREVEL et G. REIMBOLD : An attempt to explain thermally induced soft failures during low level ESD stresses : study of the differences between soft and hard NMOS failures. Dans *EOS/ESD Symposium*, pages 337–345, 1997.
- [97] J.M. LUCHIES et J.C. REINER : Properties of small leakage currents resulting from ESD stress and their latency aspects. Dans *European Symposium on Reliability of Electron devices Failure physics and analysis*, pages 217–222, Bordeaux France, 3 octobre 1995.
- [98] L. MUSSARD, P. TOUNSI, P. AUSTIN, G. BONNET, J-M. DORKEL et J. SAIZ : Power component models with thermally dependent parameters for circuit simulator. Dans *EPE European conference on Power Electronics and Applications*, 2003.
- [99] ECKEHARD SCHÖLL : *Nonequilibrium phase transition in semiconductor*. Springer-Verlag, 1987.
- [100] B. KRABBENBORG : *Modelling and simulation of electrothermal interaction in bipolar transistors*. Thèse de doctorat, Université de Twente, USA, 1994.
- [101] *Dessis ISE TCAD Release 7.0 manual*.
- [102] WOLFGANG QUADE, MASSIMO RUDAN et ECKEHARD SCHÖLL : Hydrodynamic simulation of impact-ionization effects in P-N junctions. *IEEE Transaction on computer-aided design*, 10(10):1287, octobre 1991. 1294.
- [103] A. D. STRICKER : *Technology computer aided design of ESD protection devices*, volume 122 de *Series in Microelectronics*. 2001.
- [104] V. ANANTHARAM et K.N. BHAT : Analytical solutions for the breakdown voltages of punched-trough diodes having having curved bondarities at the edges. *IEEE Transactions on Electronic Devices*, ED-27:939–945, mai 1980.
- [105] M. MERGENS, W. WILKENING et AL : Analysis and compact modeling of lateral DMOS power devices under ESD stress conditions. Dans *EOS/ESD Symposium*, pages 1–10, Orlando (USA), 1999.
- [106] G. REY et PH. LETURCQ : *Théorie approfondie du transistor bipolaire*. Masson et Cie, 1972.
- [107] PH. LETURCQ : Power bipolar transistor : a survey. Dans *Fourth Brazilian Workshop on Microelectronics*, pages 275–300, mars 1983.
- [108] D. POGANY, C. FÜRBOCK, M. LIZENGERGER, G. GROOS, K. ESMARK, P. KAMVAR, H. GOSSNER, M. STECHER et E. GORNIK : Study of trigger instabilities in smart power technology ESD protection devices using a laser interferometric thermal mapping technique. Dans *EOS/ESD Symposium*, pages 216–227, 2001.
- [109] M. LIZENGERGER, K. ESMARK, D. POGANY, C. FÜRBOCK, H. GOSSNER, E. GORNIK et W. FICHTNER : Study of triggering inhomogeneities in gg-nMOS ESD protection devices via thermal mapping using backside laser interferometry. *Microelectronics Reliability*, (40):1359–1364, 2000.
- [110] KWANG-HOON OH, CHARVAKA DUVVURY, CRAIG SALLING, KAUSTAV BANERJEE et ROBERT W. DUTTON : Non-uniform bipolar conduction in single finger NMOS transistors and implications for deep submicron ESD design.
- [111] A. G. CHYNOWETH : Ionization rates for electrons and holes in Silicon. *Physical Review*, 109(5):1537–1540, 1958.

-
- [112] R. V. OVERSTRAETEN et H. D. MAN : Measurement of the ionization rates in diffused silicon p-n junctions. *Solid State electronics*, 13:583–608, 1970.
- [113] T. LACKNER : Avalanche multiplication in semiconductors : A modification of Chynoweth's law. *Solid state electronics*, 34:33–42, 1991.
- [114] C. FÜRBOCK, K. ESMARK, M. LIZENGERGER, D. POGANY, G. GROOS, R. ZELSACHER, M. STECHER et E. GORNIK : Thermal and free carrier concentration mapping during ESD event in smart power ESD protection devices using an improved laser interferometric technique. *Microelectronics Reliability*, (40):1365–1370, 2000.
- [115] GÉRALDINE BERTRAND, CHRISTELLE DELAGE, MARISE BAFLEUR, NICOLAS NOLHIER, JEAN-MARIE DORKEL, QUANG NGUYEN, NICOLAS MAURAN, DAVID TRÉMOUILLES et PHILIPPE PERDU : Analysis and compact modeling of a vertical grounded-base NPN bipolar transistor used as an ESD protection in a smart power technology. *IEEE journal of Solid-State Circuits*, 36(9):1373–1381, septembre 2001.
- [116] D. POGANY, V. DUBEC, S. BYCHIKHIN, C. FRBCK, M. LITZENBERGER, G. GROOS, M. STECHER et E. GORNIK : Single-shot thermal energy mapping of semiconductor devices with the nanosecond resolution using holographic interferometry. *IEEE Electron Device Letter*, 23:606–608, octobre 2002.
- [117] D. POGANY, S. BYCHIKIN, E. GORNIK, M. DENISON, N. JENSEN, G. GROOS et M. STECHER : Moving current filaments in ESD protection devices and their relation to electrical characteristics. Dans *IRPS*, page 241, 2003.
- [118] G. KRIEGER : Thermal response of integrated circuit input devices to an electrostatic energy pulse. *IEEE Trans. Electron Devices*, ED-34(4):877–882, avril 1987.
- [119] V. DE HEYN, G. GROESENEKEN, B. KEPPENS, M. NATARAJAN, L. VACARESSE et G. GALLOPYN : Design and analysis of new protection structures for smart power technology with controlled trigger and holding voltage. Dans *IEEE International Reliability Physics Symposium*, pages 253–258, 2001.
- [120] DAVID TRÉMOUILLES, GÉRALDINE BERTRAND, MARISE BAFLEUR, NICOLAS NOLHIER et LIONEL LESCOUZERES : Design guidelines to achieve a very high ESD robustness in a self-biased NPN. Dans *24th Electrical Overstress/Electrostatic Discharge Symposium (EOS/ESD'02)*, pages 281–288, 8 octobre 2002.
- [121] B. KEPPENS, V. DE HEYN, N. MAHADEVA et G. GROESENEKEN : Contribution to standardization of transmission line pulse testing methodology. Dans *EOS/ESD Symposium*, pages 461–467, 2001.
- [122] H.A. GIESER, P. EGGER, J.C. REINER et M.R. HERRMANN : A CDM-only reproducible field degradatrion and its reliability aspect. *Quality and reliability enginneering international*, 10:341–350, 1994.
- [123] MASAKI MOMODOMI et FUMIO HORIGUCHI : A circular output protection device using bipolar action. Dans *IRPS*, pages 169–173, 1987.
- [124] MING-DOU KER et TAIN-SHUN WU : Novel octogonal device structure for output transistors in deep-submicron low-voltage CMOS technology. Dans *IEDM*, pages 889–892, 1996.
- [125] NILS JENSEN, GERHARD GROOS, MARIE DENISON, JAN KUZMIK, DIONYZ POGANY, ERICH GORNIK et MATTHIAS STECHER : Coupled Bipolar Transistors as Very Robust ESD Protection Devices for Automotive Applications. Dans *EOS/ESD symposium*, pages 313–326, 2003.
- [126] T. BEAUCHÊNE, D. TRÉMOUILLES, D. LEWIS, P. PERDU et P. FOUILLAT : Characterization of ESD induced defects using photovoltaic laser stimulation (PLS). Dans *XIVth European Symposium on Reliability of Electron Devices, Failure Physics and Aanalysis (ESREF'2003)*, 7 octobre 2003.

- [127] P. ANTOGNETTI et G. MASSOBRIO : *Semiconductor device modeling with SPICE*. 1988.
- [128] C. MCANDREW et AL : VBIC95 : the vertical bipolar inter-company model. *IEEE Journal of Solid-State Circuits*, 31:1476–1483, 1996.
- [129] R. W. DUTTON : Bipolar transistor modeling of avalanche generation for computer circuit simulation. *IEEE Transaction on Electron Devices*, ED-22(6):334–338, juin 1975.
- [130] K. OPALSKA et L. OPALSKI : A model of the BJT avalanche breakdown for circuit simulation. Dans *ECCTD*, pages 220–225, Budapest, septembre 1997.
- [131] A. AMERASEKERA et A. CHARTTERJEE : An investigation of BiCMOS ESD protection circuit elements and applications in submicron technologies. Dans *EOS/ESD symposium*, pages 265–276, 1992.
- [132] S. RAMASWAMY : *Modeling simulation and design guidelines for EOS/ESD protection circuits in CMOS technologies*. Thèse de doctorat, University of Illinois, Urbana-Champaign, octobre 1996.
- [133] M. MERGENS, W. WILKENING et AL. : Modular approach of high current MOS compact model for circuit-level ESD simulation including transient gate coupling behavior. Dans *IRPS*, pages 167–178, San Diego (USA), 1999.
- [134] A. AMERASEKERA et AL. : Modeling MOS snapback and parasitic bipolar action for circuit-level ESD and high current simulations. Dans *34th IEEE IRPS*, pages 318–326, 1996.
- [135] H. WOLF, HORST GIESER et W. STADLER : Bipolar model extension for MOS transistors considering gate coupling effects in the HBM ESD domain. *Microelectronics Reliability*, 39:1541–1549, 1999.
- [136] T. LI et AL : Substrate resistance modeling and circuit-level simulation of parasitic device coupling effects for CMOS I/O circuits under ESD stress. Dans *EOS/ESD Symposium*, pages 281–289, 1998.
- [137] P. SALOME, C. RICHER, S. ESSAIFI, C. LEROUX, I. ZAZA, A. JUGE et P. MORTINI : Extended SPICE-like model accounting for layout effects on snapback phenomenon during ESD events. *Microelectronics Reliability*, 39:833–838, 1999.
- [138] S. RAMASWAMY et AL : A unified substrate current model for weak and strong impact ionization in sub-0.25 μ NMOS devices. Dans *IEDM Proceedings*, pages 885–888, 1997.
- [139] DAVID TRÉMOUILLES, GÉRALDINE BERTRAND, MARISE BAFLEUR, FELIX BEAUDOIN, PHILIPPE PERDU, NICOLAS GUITARD et LIONEL LESCOUZERES : TCAD and SPICE modeling help solve ESD protection issues in analog CMOS technology. *Microelectronic Reliability*, 43(1):71–79, janvier 2003.
- [140] M. GHARBI : *La tenue en tension et le calibre en courant du transistor MOS vertical dans la gamme des moyennes tensions (300 à 1000 Volts)*. Thèse de doctorat, Université Paul Sabatier, 1985.
- [141] C.D. PARICK et R.M. PATRIKAR : A compact model for N-well resistor. *Solid-State Electronics*, 43:683–685, 1999.
- [142] J. BARTH et J. RICHER : Correlation considerations : real HBM to TLP and HBM testers. Dans *EOS/ESD 01*, pages 453–460, septembre 2001.
- [143] T. LI, D. SUH, S. RAMASWAMY, P. BENDIX, E. ROSENBAUM, A. KAPOOR et S. M KANG : Study of a CMOS I/O Protection Circuit using Circuit-Level Simulation. Dans *Int. Rel. Phys. Symp.*, pages 333–338, Denver, avril 1997.
- [144] R. FRIED, Y. BLECHER et S. FRIEDMAN : Structures for ESD protection in CMOS processes. *Microelectronics reliability*, 37(7):1111–1120, 1997.

-
- [145] A. AMERASEKERA et A. CHATTERJEE : An investigation of BiCMOS ESD protection circuits elements and applications in submicron technologies. Dans *EOS/ESD Symposium*, pages 265–276, 1992.
- [146] A. CHATTERJEE et AL. : Design and simulation of a 4kV ESD protection circuits for a 0.8 μm BiCMOS process. Dans *IEDM*, pages 913–916, 1991.
- [147] N. MAENE et AL. : On chip electrostatic discharge protections for inputs, outputs and supplies of CMOS circuits. Dans *EOS/ESD symposium*, pages 228–233, 1992.
- [148] R. MERRILL et E. ISSAQ : ESD design methodology. Dans *EOS/ESD Symposium*, pages 233–237, 1993.
- [149] S. DABRAL, R. ASLETT et T. MALONEY : Core clamps for low voltage technologies. Dans *EOS/ESD Symposium*, pages 141–149, 1994.
- [150] E.R. WORLEY, R. GUPTA, B. JONES, R. KJAR, C. NGUYEN et M. TENNYSON : Sub-micron chip ESD protection schemes wich avoid avalanching junctions. Dans *EOS/ESD Symposium*, pages 13–20, 1995.
- [151] CYNTHIA A. TORRES, JAMES W. MILLER, MICHAEL STOCKINGER, MATTHEW D. AKERS, MICHAEL G. KHAZHINSKY et JAMES C. WELDON : Modular, portable, and easily simulated ESD protection networks for advanced CMOS technologies. *Microelectronics Reliability*, 42:837–885, 2002.
- [152] M. KER : Whole-chip ESD protection design with efficient VDD-to-VSS ESD clamp circuits for submicron CMOS VLSI. *IEEE Transaction on Electron Devices*, 46(1), janvier 1999.
- [153] P. PAVAN, A. PELLESI, G. MENEGHESSO et E. ZANONI : Effects of ESD protection on latch-up sensitivity of CMOS 4-stripe structures. *Microelectronics Reliability*, 37(10/11):1561–1564, 1997.
- [154] L. SPONTON et AL : ESD protection structures for BCD5 smart power technologies. Dans *12th European Symposium on Reliability of Electron Devices Failure Physics (ESREF)*, pages 1683–1687, octobre 2001.
- [155] EIA/JEDEC STANDARD : IC Latch-up test EIA/JESD78, mars 1997.
- [156] JOSEF LUTZ et MARTIN DOMEIJ : Dynamic avalanche and reliability of high voltage diodes. *Microelectronics Reliability*, 43:529–536, 2003.
- [157] MARTIN DOMEIJ, BO BREITHOLTZ, JOSEF LUTZ et MIKAEL ÖSTLING : Dynamic avalanche in Si power diodes and impact ionization at the nn^+ junction. *Solid-state electronics*, 44:477–485, 2000.
- [158] HAROLD C. BOWERS : Space-charge-induced negative resistance in avalanche diodes. *IEEE Transaction on electron devices*, ED-15(6):343–350, juin 1996.
- [159] MARCEL W. MULLER et HENRY GUCKEL : Negative resistance and filamentary currents in avalanching silicon p^+i-n^+ . *IEEE Trasaction on electron devices*, ED-15(8):560–568, août 1968.
- [160] ALEX Q. HUANG, VICTOR TEMPLE, YIN LIU et YUANZHU LI : Analysis of the turn-off failure mechanism of silicon power diode. *Solid-state electronics*, 47:737–739, 2003.

Liste des publications

Revue :

- GÉRALDINE BERTRAND, CHRISTELLE DELAGE, MARISE BAFLEUR, NICOLAS NOLHIER, JEAN-MARIE DORKEL, QUANG NGUYEN, NICOLAS MAURAN, DAVID TRÉMOUILLES et PHILIPPE PERDU : Analysis and compact modeling of a vertical grounded-base NPN bipolar transistor used as an ESD protection in a smart power technology. *IEEE journal of Solid-State Circuits*, 36(9):1373–1381, septembre 2001.
- ROMAIN DESPLATS, FELIX BEAUDOIN, PHILIPPE PERDU, P. POIRIER, DAVID TRÉMOUILLES, MARISE BAFLEUR et DEAN LEWIS : Backside localization of current leakage faults using thermal laser stimulation. *Microelectronics Reliability*, 41(9-10):1539–1544, octobre 2001. 12th European Symposium on Reliability of Electron Devices, Failure Physics and Analysis (ESREF'2001).
- T. BEAUCHÊNE, D.LEWIS, F.BEAUDOIN, V.POUGET, R.DESPLATS, P.FOULLAT, P.PERDU, M.BAFLEUR et D.TRÉMOUILLES : Thermal laser stimulation and NB-OBIC techniques applied to ESD defect localization. *Microelectronic Reliability*, 43(3):439–444, mars 2003. XIIIth European Symposium on Reliability of Electron Devices, Failure Physics and Analysis (ESREF'2002).
- DAVID TRÉMOUILLES, GÉRALDINE BERTRAND, MARISE BAFLEUR, FELIX BEAUDOIN, PHILIPPE PERDU, NICOLAS GUITARD et LIONEL LESCOUZERES : TCAD and SPICE modeling help solve ESD protection issues in analog CMOS technology. *Microelectronic Reliability*, 43(1):71–79, janvier 2003.
- T. BEAUCHÊNE, D. TRÉMOUILLES, D. LEWIS, P. PERDU et P. FOULLAT : Characterization of ESD induced defects using photovoltaic laser stimulation (PLS). *Microelectronics Reliability*, 43(9-11):1577–1582, septembre 2003. XIVth European Symposium on Reliability of Electron Devices, Failure Physics and Analysis (ESREF'2003).
- D. TRÉMOUILLES, M. BAFLEUR, G. BERTRAND, N. NOLHIER, N. MAURAN et L. LESCOUZERES : Latch-up ring design guidelines to improve electrostatic discharges (ESD) protection scheme efficiency. *Journal of Solid-State Circuits*, 39, octobre 2004.

Conférences internationales :

- DAVID TRÉMOUILLES, GÉRALDINE BERTRAND, MARISE BAFLEUR, FELIX BEAUDOIN, PHILIPPE PERDU et LIONEL LESCOUZERES : TCAD and SPICE modeling help solve ESD protection issues in analog CMOS technology. Dans *23rd International Conference on Microelectronics (MIEL 2002)*, pages 749–752, Nis (Yougoslavie), 12 mai 2002.
- F. BEAUDOIN, A. WISLEZ, M. BAFLEUR, D. LEWIS, R.DESPLATS, P. PERDU et D. TRÉMOUILLES : Laser beam based ESD defect localization in ICs. Dans *28th International Symposium on Testing and Failure Analysis (ISTFA '2002)*, Phoenix (USA), 3 novembre 2002. ISTFA Best In Session Award.
- T. BEAUCHÊNE, D. LEWIS, D. TRÉMOUILLES, F. ESSELY, P. PERDU et P. FOULLAT : ESD defect localization and analysis using pulsed OBIC techniques. Dans *18th SYMPOSIUM ON MICROELECTRONICS TECHNOLOGY AND DEVICES CHIP IN SAMPA (SBMicro 2003)*, São Paulo (Bresil), 8 septembre 2003.

- DAVID TRÉMOUILLES, GÉRALDINE BERTRAND, MARISE BAFLEUR, NICOLAS NOLHIER et LIONEL LESCOUZERES : Design guidelines to achieve a very high ESD robustness in a self-biased NPN. Dans *24th Electrical Overstress/Electrostatic Discharge Symposium (EOS/ESD'02)*, pages 281–288, 8 octobre 2002.
- DAVID TRÉMOUILLES, MARISE BAFLEUR, GÉRALDINE BERTRAND, NICOLAS NOLHIER, NICOLAS MAURAN et LIONEL LESCOUZERES : Solving ESD protection latchup guard rings issue during electrostatic discharge (ESD) events. Dans *2003 Bipolar/BiCMOS Circuits and Technology Meeting (BCTM'2003)*, Toulouse (France), 28 septembre 2003.
- N. GUITARD, D. TRÉMOUILLES, S. ALVES, M. BAFLEUR, F. BEAUDOIN, P. PERDU et A. WISLEZ : ESD Induced Latent Defects In CMOS ICs And Reliability Impact. Dans *EOS/ESD Symposium*, 2004.
- N. GUITARD, D. TRÉMOUILLES, M. BAFLEUR, L. ESCOTTE, L. BARY, P. PERDU, G. SARRABAYROUSE, N. NOLHIER et R. REYNA-ROJAS : Low frequency noise measurements for ESD latent defect detection in high reliability applications. Dans *15th European Symposium on Reliability of Electron Devices, Failure Physics and Analysis (ESREF'2004)*, 2004.

Conférences et séminaires :

- DAVID TRÉMOUILLES : Conception et optimisation de circuits robustes aux décharges électrostatiques ; vers la fiabilité prédictive. Dans *Vèmes Journées Nationales du Réseau Doctoral de Microélectronique (JNRDM'2002)*, pages 194–195, Grenoble (France), 23 avril 2002.
- DAVID TRÉMOUILLES, NICOLAS GUITARD, MARISE BAFLEUR, NICOLAS NOLHIER et LIONEL LESCOUZERES : TLP and photo emission coupling, a powerful tool for study of ESD protection strategy. Dans *Workshop EOS/ESD/EMI*, 13 octobre 2002.
- T. BEAUCHÊNE, D. LEWIS, F. BEAUDOIN, H. LAPUYADE, P. PERDU, M. BAFLEUR et D. TRÉMOUILLES : Thermal laser stimulation and OBIC techniques applied to ESD defect localization. Dans *Workshop EOS/ESD/EMI*, 13 octobre 2002.
- MARISE BAFLEUR, DAVID TRÉMOUILLES, STÉPHANE ALVES, FELIX BEAUDOIN, PHILIPPE PERDU et ALAIN WISLEZ : Impact of CDM ESD stress on CMOS ICs reliability. Dans *Workshop ESD/EOS/EMI*, Toulouse (France), 13 octobre 2003.
- M. BAFLEUR, D. TRÉMOUILLES, S. ALVES, F. BEAUDOIN, P. PERDU et A. WISLEZ : Impact d'un stress ESD de type CDM sur la fiabilité d'un circuit intégré. Dans *8ème Atelier Analyse et Mécanismes de Défaillance des Composants pour l'Electronique*, Port d'Albret (France), 4 juin 2002.
- DAVID TRÉMOUILLES : Stratégies de protection intégrées contre les décharges électrostatiques. Comparaison des stress de types HBM et CDM. Dans *Séminaire Annuel de l'Ecole Doctorale Génie Electrique, Electronique, Télécommunications (GEET)*, 11 mars 2003.
- N. GUITARD, D. TRÉMOUILLES, M. BAFLEUR, L. ESCOTTE, L. BARY, P. PERDU, G. SARRABAYROUSE, N. NOLHIER et R. REYNA-ROJAS : Potentialities of low frequency noise measurement as esd latent defect detection for high reliability applications. Dans *2nd Workshop EOS/ESD/EMI*, pages 39–42, Toulouse (France), 13 mai 2004.
- F. ESSELY, D. TRÉMOUILLES, N. GUITARD, M. BAFLEUR, P. PERDU, A. TOUBOUL et D. LEWIS : OBIC techniques applied to study the impact of multiple ESD stresses. Dans *2nd Workshop EOS/ESD/EMI*, pages 35–37, Toulouse (France), 13 mai 2004.
- N. GUITARD, D. TRÉMOUILLES, M. BAFLEUR, L. ESCOTTE, L. BARY, P. PERDU, G. SARRABAYROUSE, N. NOLHIER et R. REYNA-ROJAS : Méthode de détection de défauts latents ESD dans les technologies CMOS basée sur des mesures de bruit basse fréquence. Dans *Atelier ANADEF 2004*, Port d'Albret (France), 8 juin 2004.

Résumé

Les travaux présentés dans ce mémoire visent à améliorer la méthodologie de conception et les performances des stratégies de protection contre les décharges électrostatiques (ESD) dans les circuits intégrés.

Pour cela, l'approche choisie est basée sur une analyse approfondie de la physique des composants soumis aux ESD et plus particulièrement, les effets des très fortes densités de courant.

L'étude, focalisée sur les transistors bipolaires autopolarisés, s'appuie sur la simulation physique 2D et l'utilisation des outils de localisation de défaillance basés sur les techniques de stimulation laser. L'analyse physique en résultant a permis d'une part, de définir des règles de dessin universelles pour l'obtention d'une robustesse ESD élevée et d'autre part, de proposer des macro-modèles de type SPICE originaux pour prendre en compte les effets des fortes densités de courant.

Enfin, après avoir mis en évidence plusieurs phénomènes limitant les performances des réseaux de protection, nous avons défini une méthodologie de conception améliorée permettant de les prendre en compte et de garantir la performance des solutions de protections fournies aux concepteurs de circuits.

Mots-clés: Décharges électrostatiques (ESD), Protection, Forte densité de courant, Transistor bipolaire autopolarisé, Latch-up, Simulation physique, Modélisation SPICE, Stimulation laser

Abstract

The research work presented in this thesis is aimed at improving the performance of electrostatic discharges (ESD) protection and the related design methodology for integrated circuits.

To achieve this goal, a thorough analysis of the physical mechanisms involved and more precisely of high current density effects has been carried out.

Such a study, focused on self-biased bipolar transistors, has been achieved with the help of 2D-numerical simulation and failure localization techniques based on laser stimulation. New universal design guidelines resulting in high performance ESD protections and original SPICE macro-models that take into account high current density effects, are proposed.

Finally, the physical mechanisms limiting the performance of ESD protection networks are studied. To cope with these issues, an improved design methodology that guarantees the robustness of ESD protections used by circuit designers, is proposed.

Keywords: Electrostatic discharge (ESD), Protection, High current density, Self biased bipolar transistor, Latch-up, Device simulation, SPICE modeling, Laser stimulation

