

50376  
1990  
39

70849

50376  
N° d'ordre: 510  
1990  
39

# THESE

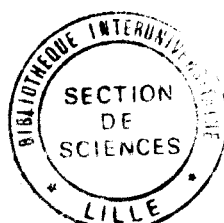
présentée à

l'Université des Sciences et Techniques de Lille  
Flandres-Artois

pour obtenir le grade de

**Docteur en Electronique**

par



**Bruno BACCUS**

Ingénieur ISEN

**SIMULATION BIDIMENSIONNELLE**

**DE TECHNOLOGIES SILICIUM**

**L'APPROCHE MULTICOUCHE**

Soutenue le 22 Mars 1990, devant la commission d'examen :

Président : MM. SALMER  
Rapporteurs : PONCET  
FAUQUEMBERGUE  
Examineurs : COLLARD  
POLAK  
LOMBARDI  
MARTINEZ  
DECARPIGNY

*à Carole,*

*à Nicolas,*

*à mes parents,*

*à la mémoire de mon grand-père*

Cette thèse a été préparée  
à l'INSTITUT SUPERIEUR D'ELECTRONIQUE DU NORD  
au sein du laboratoire associé 253 au C.N.R.S.

*J'exprime ma reconnaissance à Monsieur le Professeur Salmer qui a accepté la présidence de ce jury. De plus je le remercie pour sa compréhension eu égard aux particularités de ce manuscrit.*

*Je remercie sincèrement Messieurs A. Poncet et R. Fauquembergue pour avoir accepté la lourde tâche d'être rapporteur et pour leur intérêt porté à ce travail.*

*Messieurs A. Martinez, S. Polak, J.N. Decarpigny et C. Lombardi m'ont fait l'honneur de participer à ce jury. Je tiens à leur exprimer ici ma profonde gratitude, espérant que nous pourrions poursuivre des travaux communs dans d'autres occasions.*

*Je tiens à remercier tout particulièrement Monsieur D. Collard pour son soutien constant durant ce travail et pour les nombreuses discussions que nous avons eu à ce propos. Que notre collaboration continue d'être fructueuse pour le futur d'IMPACT 4 !*

*J'exprime ma profonde gratitude à Monsieur M. Lannoo pour l'accueil qu'il m'a réservé au sein du laboratoire. Sa passion pour la recherche et son enthousiasme sont les meilleurs encouragements.*

*Je tiens à remercier très sincèrement Monsieur D. Morel qui a largement contribué aux possibilités du code lors des premières approches. J'espère vivement que cet apport technique et amical pourra se développer.*

*Enfin, Monsieur E. Dubois a suivi le déroulement de ce travail, avec bien souvent des conseils avisés. Merci pour sa compréhension et sa bonne humeur.*

## SOMMAIRE

	Pages
<b><u>CHAPITRE I : INTRODUCTION</u></b> .....	1
I. L'évolution des technologies silicium .....	2
II. L'évolution des simulateurs de procédé .....	5
III. Le programme IMPACT 4 .....	7
Références (chapitre I) .....	8
<b><u>RESUME</u></b> .....	10
Références .....	30
<b><u>CHAPITRE II : STRUCTURE GENERALE ET OUTILS DE DEVELOPPEMENT</u></b> .....	31
I. Définition du problème .....	32
II. La méthode des éléments finis et organisation des données ...	32
2.1 Choix de la méthode des éléments finis .....	32
2.2 Structure de données .....	33
III. Génération de maillage .....	37
3.1 Généralités .....	37
3.2 Stratégie .....	38
3.3 Génération de maillage .....	39
Références (chapitre II) .....	43

<b>CHAPITRE III : ETAPES ELEMENTAIRES</b> .....	44
<b>I. Etapes de dépôt et de gravure</b> .....	45
1.1 Modèles .....	45
1.1.1 modèles de gravure .....	46
1.1.2 modèles de dépôt .....	50
1.2 Algorithmes .....	53
1.2.1 l'algorithme du "string model" et les problèmes géométriques associés .....	53
1.2.2 relations avec la génération de maillage et la structure de données .....	55
<b>II. Implantation ionique</b> .....	58
2.1 Distributions verticales .....	59
2.1.1 la méthode de Monte Carlo et les calculs analytiques .....	59
2.1.2 distribution Gaussienne .....	60
2.1.3 distribution bi-Gaussienne .....	60
2.1.4 distribution Pearson IV .....	61
2.1.5 paramètres par défaut .....	63
2.2 Le problème multicouche .....	65
2.3 Calculs bi-dimensionnels .....	68
2.3.1 le problème numérique .....	68
2.3.1.1 formulation .....	68
2.3.1.2 aspects pratiques .....	71
2.3.2 aspects de modélisation .....	74
<b>III. Diffusion</b> .....	79
3.1 Les équations de diffusion .....	80
3.1.1 niveaux de modèles .....	80
3.1.2 des lois de Fick à la formulation générale .....	81
3.1.3 conditions aux frontières .....	83

	Pages
3.2 Modèles .....	85
3.2.1 silicium .....	85
3.2.1.1 coefficient de diffusion .....	85
3.2.1.2 diffusion du phosphore .....	88
3.2.1.3 phénomènes de clustering .....	91
3.2.2 les autres matériaux: oxyde, nitrure et .....	92
polysilicium	
3.2.2.1 oxyde et nitrure de silicium .....	92
3.2.2.2 polysilicium .....	93
3.3 Résolution avec la méthode des éléments finis .....	94
3.3.1 application à la diffusion des dopants .....	95
3.3.2 aspects non-linéaires et non-stationnaires .....	98
3.3.2.1 la méthode de Newton-Raphson .....	99
3.3.2.2 le schéma implicite incomplet .....	101
3.4 Résolution du système linéaire .....	105
3.4.1 méthodes directes et itératives .....	105
3.4.2 structure de la matrice avec le problème .....	106
multicouche et mise en forme	
3.4.3 gradients conjugués et décomposition incomplète ...	109
 IV. Oxydation .....	 114
4.1 Introduction .....	114
4.1.1 le point de vue de la modélisation .....	114
4.1.1.1 modèles uni-dimensionnels .....	114
4.1.1.2 modèles bi-dimensionnels .....	116
4.1.2 le point de vue de la simulation .....	116
4.2 Simulation bi-dimensionnelle de l'oxydation .....	118
4.2.1 initialisation de la structure .....	120
4.2.2 diffusion des agents oxydant .....	121
4.2.2.1 équation et conditions aux frontières .....	121
4.2.2.2 calcul du coefficient de réaction .....	121
aux interfaces	





	Pages
3.3 Problème de diffusion couplée .....	170
<b>IV. Applications</b> .....	<b>174</b>
4.1 Transistor bipolaire isolé par tranchée .....	174
4.2 Simulation couplée de procédé et de dispositif .....	178
<b>Références (chapitre IV)</b> .....	<b>180</b>
<b><u>CHAPITRE V</u> : APPLICATIONS</b> .....	<b>183</b>
<b>I. Transistor LDD-MOS</b> .....	<b>184</b>
1.1 Introduction .....	184
1.2 Résultats de simulation .....	184
<b>II. Effets parasites intervenant dans un transistor bipolaire ...</b>	<b>191</b>
<b>à émetteur muré</b>	
2.1 Introduction .....	191
2.1.1 définition du problème .....	191
2.1.2 description du procédé .....	192
2.2 Résultats de simulation .....	192
2.2.1 formation de l'effet de pincement de base .....	193
2.2.2 Sensibilité à la sur-attaque de l'oxyde .....	195
2.2.3 Simulations de dispositif .....	198

<b>III. Application à une technologie bipolaire avancée</b> .....	200
3.1 Technologies bipolaires avancées .....	200
3.2 Description du dispositif .....	201
3.3 Connexion entre les bases extrinsèque et intrinsèque ....	206
3.4 Configuration à émetteur muré .....	209
<b>Références (chapitre V)</b> .....	211
<b><u>CHAPITRE VI</u> : CONCLUSIONS</b> .....	213
<b>Références (chapitre VI)</b> .....	216

## CHAPITRE I

### INTRODUCTION

#### Résumé

Dans un premier temps, l'évolution des technologies silicium est analysée et les besoins correspondants dans le domaine de la simulation de procédé sont définis. Dans un deuxième temps, l'évolution naturelle des programmes détermine une autre définition des besoins en simulation de procédé.

Ces remarques sont utilisées comme base pour la définition du but général de ce travail -le développement d'un simulateur de procédé multicouches-, et de ses caractéristiques telles que les possibilités géométriques, les étapes élémentaires prises en compte et enfin certaines procédures numériques.

## I. L'EVOLUTION DES TECHNOLOGIES SILICIUM

Les technologies silicium ont connu une expansion extraordinaire depuis l'invention du transistor bipolaire par Shockley, Bardeen et Brattain en 1947, d'une part, et l'invention du circuit intégré par J.S. Kilby en 1959, d'autre part [1]. La figure 1 montre la croissance rapide du nombre de composants par circuit, durant les 30 dernières années [2]. Depuis l'apparition de technologies SSI (Small-Scale-Integration) jusqu'aux technologies actuelles les plus performantes dénommées ULSI (Ultra Large Scale Integration), plus d'un million de transistors ont été intégrés par circuit, en particulier dans plusieurs microprocesseurs.

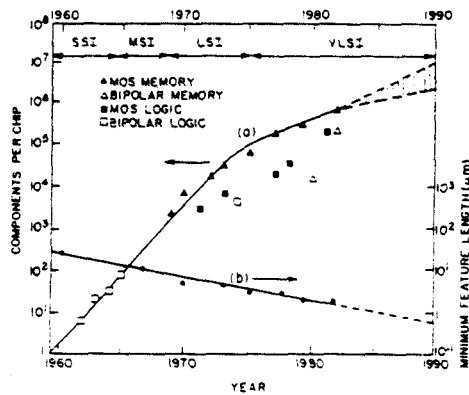


Fig. 1 (a) Croissance exponentielle du nombre de composants par circuit intégré.  
(b) Décroissance exponentielle des dimensions minimales [2].

L'une des raisons principales de cette évolution est la miniaturisation constante des composants élémentaires, c'est à dire les transistors MOS ou bipolaires. Chaque type de technologie silicium a subi ce phénomène, de manière spécialement prononcée pour les composants MOS. A ce titre, la figure 2 montre les étapes de fabrication principales d'un composant LDD-MOS standard. Les procédés actuels sont caractérisés par une longueur de grille d'un micromètre et par une épaisseur d'oxyde de grille d'environ 15 nanomètres. Les expérimentations les plus récentes ont débouchées sur des transistors à longueur de grille inférieure à 0.1 micromètre [4]. La croissance résultante dans la densité d'intégration est très bien illustrée par les mémoires DRAM, avec 4 à 16 Mbits par circuit. L'utilisation massive d'étapes technologiques complexes résulte en la conception de cellules élémentaires extrêmement compactes, comme illustré par l'exemple présenté en figure 3.

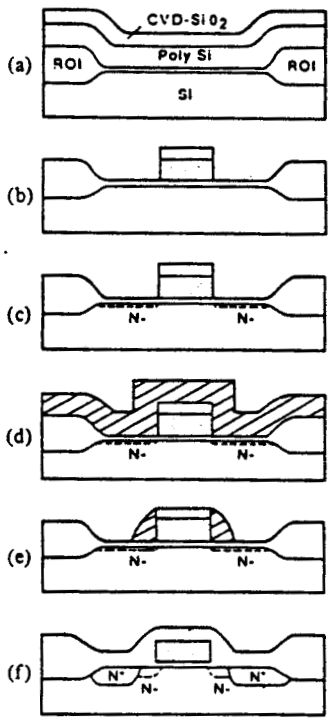
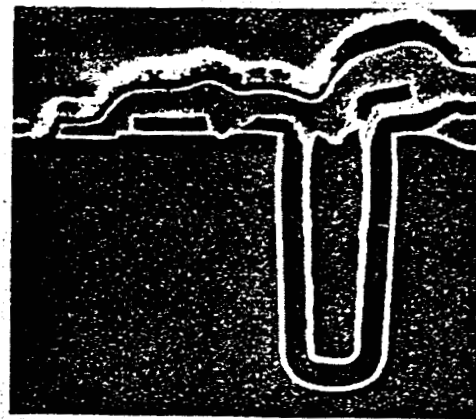
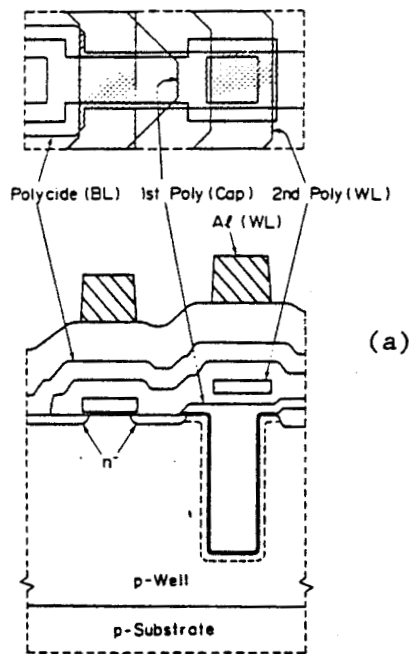


Fig. 2 Etapes de fabrication d'un transistor LDDMOS [3].

- (a) formation de l'oxyde de grille, et dépôts de polysilicium et de  $\text{SiO}_2$ .
- (b) Gravure sèche de type RIE.
- (c) Implantation ionique  $n^-$ .
- (d) dépôt d'oxyde.
- (e) gravure RIE pour la formation de l'espaceur.
- (f) Implantation ionique et recuit.



(b)

Fig. 3 (a) Dessin des masques et représentation schématique d'une cellule utilisant une tranchée, dans une mémoire 4 Mbit.  
(b) Photographie SEM de cette cellule réalisée expérimentalement. La tranchée a une profondeur de 3 micromètres [5].

Un autre aspect est relatif à la complexification croissante des procédés, pour les étapes élémentaires (telles que les recuits thermiques rapides), ou pour l'intégration complète d'une technologie (BiCMOS). La figure 4 décrit un procédé BiCMOS, autorisant une fréquence de coupure de 16 GHz. Cette exemple comporte quelques unes des caractéristiques des technologies bipolaires modernes telles que l'isolation par tranchée et les émetteurs polysilicium [7]. Ce dernier aspect est tout à fait représentatif de la tendance générale vers les structures multicouches.

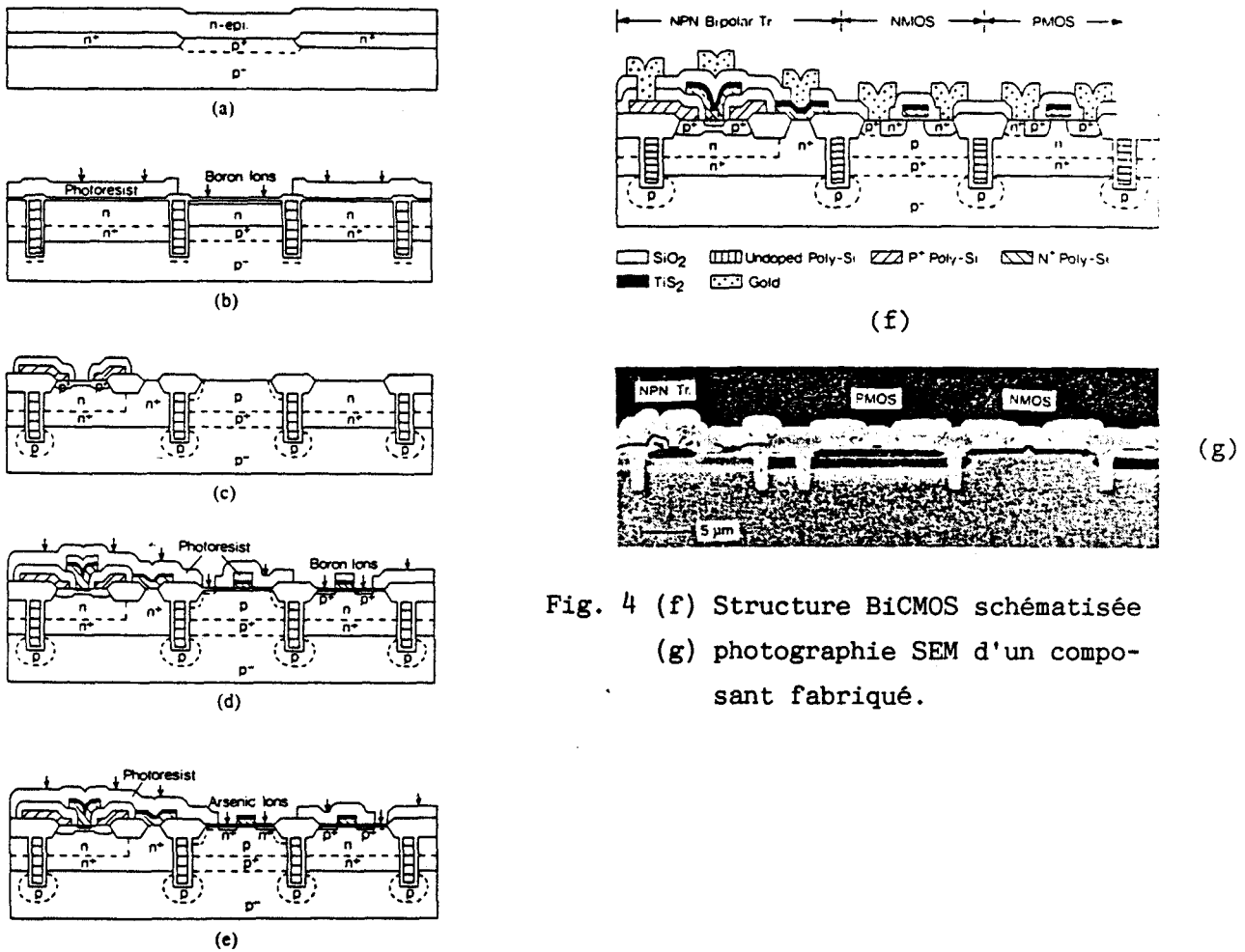


Fig. 4 (f) Structure BiCMOS schématisée  
(g) photographie SEM d'un composant fabriqué.

Fig. 4 Etapes de fabrication d'un procédé BiCMOS, avec des composants bipolaires auto-alignés, et emploi du polysilicium [6].

- (a) croissance de la couche epitaxiale.
- (b) implantation du puit P.
- (c) gravure RIE de la base polysilicium et de l'espaceur (oxyde).
- (d) implantation ionique des sources et drains des PMOS.
- (e) implantation ionique des sources et drains des NMOS.

De nombreuses expérimentations sont nécessaires afin de contrôler et d'optimiser de tels procédés. La minimisation des temps et coûts induits par ces investigations peut être facilitée par l'utilisation de simulateurs de procédé et de dispositif, suivant en cela l'emploi maintenant très répandu de conception assistée par ordinateur dans d'autres domaines comme la mécanique ou l'aérodynamique. Une vision idéaliste serait de considérer que la simulation est capable de prédire le comportement et les performances des dispositifs semiconducteurs. Cela n'est malheureusement pas vrai pour de nombreuses raisons parmi lesquelles on peut distinguer le manque de généralité des simulateurs et de la modélisation. La simulation peut cependant être un outil très utile, car cela autorise à la fois une compréhension accrue des phénomènes physiques et la possibilité de varier considérablement les paramètres et conditions des procédés. Dans le cas le plus général, les itérations sur des simulations couplées de procédé et de dispositif facilitent la détermination de paramètres technologiques comme l'implantation ionique ou les températures de recuit, en vue de performances électriques bien précises. Dans cette optique, l'état de l'art peut être évalué par les travaux où des variations de paramètres de procédé telles que la dispersion dans la valeur de la longueur de grille sont détectées à partir de la simulation [8].

En considérant ces points, nous pouvons définir les caractéristiques requises pour les simulateurs de procédé actuels:

- pas de spécialisation pour une technologie particulière.
- généralité géométrique incluant l'aspect multicouches et les phénomènes particuliers aux interfaces.
- au moins bi-dimensionnels (2D), ou tri-dimensionnels (3D).
- un haut niveau de modélisation et une architecture ouverte afin de suivre l'évolution rapide des technologies.

## II. L'EVOLUTION DES SIMULATEURS DE PROCÉDES

Les progrès en simulation de procédés peuvent être évalués de plusieurs manières, suivant que les aspects dimensionnels ou de modélisation sont considérés. Se concentrant sur le premier point, trois cas sont examinés ci-après. A ce stade de l'introduction, il faut noter que cet état de l'art est décrit au moment du début de ce travail (1986); pour cette raison, des résultats récents n'ont pas été considérés et seront reportés dans la conclusion.

Concernant les programmes unidimensionnels, depuis les premières versions de SUPREM jusqu'au simulateur multicouches SUPREM III [9], un haut niveau de maturité a été atteint, reflété par l'emploi très répandu de ces programmes en industrie. Un bon niveau de modélisation est inclus pour la diffusion des impuretés dans le silicium ou dans le polysilicium, avec des approches différentes pour SUPREM III et ASPREM [10].

Les composants étant essentiellement bidimensionnels par nature (technologies MOS) ou comportant des effets parasites 2D (technologies bipolaires), des profils de dopage 2D suffisamment précis sont de plus en plus nécessaires. Cependant, les simulateurs 2D ne sont pas toujours extrêmement mûrs, et ceci pour deux raisons principales. D'une part le niveau de modélisation devrait être accru de manière significative pour les technologies les plus avancées [11]. D'autre part, des progrès sont également nécessaires pour les aspects numériques ou de développement de logiciel [12]. Dans tous les cas, le compromis est difficile à obtenir par rapport à des ressources informatiques raisonnables (puissance de calcul et mémoire). De fait, la majorité des simulateurs ne traite que le silicium (SUPRA [13], IMPACT [14], les premières versions de TITAN [15]...), ou est restreinte à des systèmes bi-couches silicium/oxyde de silicium (FEDSS [16]). Quant aux programmes multicouches, ils sont dévolus à la simulation de topologie et ne sont pas utilisables pour des procédés complets (SAMPLE [17]).

Les programmes 3D peuvent être considérés comme faisant partie d'un domaine plus expérimental et moins bien établi, même si plusieurs applications ont été traitées [18], et des simulations couplées de procédé et de dispositif ont été effectuées [19]. En particulier, les aspects numériques et informatiques de tels développements sont encore des points très sensibles.

On peut donc déduire que deux améliorations doivent être examinées, concernant les trois aspects cités ci-dessus: le 2D multicouches ou des simulations 3D réalistes. Ce dernier point n'a pas été considéré à cause des coûts associés à un tel projet, en termes de ressources informatiques ou de développement de logiciel.



### III. LE PROGRAMME IMPACT 4

Résumant les évolutions des technologies silicium et des simulateurs de procédé, nous définissons maintenant le but général de ce travail et donc les caractéristiques principales du programme à développer.

Ce programme 2D doit être aussi général que possible, traitant les technologies MOS, bipolaires ou BiCMOS. Du point de vue topologique, non seulement le silicium et l'oxyde doivent être pris en compte, mais aussi d'autres matériaux comme le nitrure de silicium, le silicium polycristallin, la résine ou l'aluminium, par exemple. Le programme doit être également capable de simuler toutes les étapes élémentaires intervenant dans la fabrication des composants: dépôt et gravure, diffusion et oxydation, implantation ionique, et enfin epitaxie.

Ces points sont surtout relatifs aux aspects géométriques et structurels. Par ailleurs, ce programme doit être aussi un "laboratoire", afin d'investiger les effets ou l'importance de nouvelles modélisations. En ce sens, les modifications ultérieures devront être aisément incorporables, ceci définissant des besoins bien précis au niveau de l'organisation logicielle, ainsi que des méthodes numériques fiables (solveurs linéaires, génération de maillage).

Dans le chapitre II, la structure de données d'IMPACT 4 et les outils de développement sont présentés, ainsi que les procédures de génération de maillage. Le chapitre III décrit les étapes élémentaires et la manière dont elles sont modélisées et simulées: dépôt/gravure, implantation ionique, diffusion en ambiance neutre, oxydation. Le chapitre IV présente une approche originale afin de résoudre un problème très délicat: le maillage adaptatif. Les applications sont détaillées dans le chapitre V pour une technologie MOS standart, des transistors bipolaires incluant une technologie complètement implantée et une structure utilisant le silicium polycristallin. Enfin, le chapitre VI examine l'évolution des autres simulateurs de procédé et les suites de ce travail.

REFERENCES BIBLIOGRAPHIQUES :

- [1] J.S. Kilby, "Invention of the integrated circuits", IEEE Trans. Electron Devices, ED-23, 648, 1976.
- [2] S.M. Sze, "VLSI Technology", MC Graw Hill, 1983.
- [3] P.J. Tsang, S. Ogura, W.W. Walker, J.F. Shepard and D.L. Critchlow, "Fabrication of high-performance LDDFET's with oxide sidewall-spacer technology", IEEE Trans. Electron Devices, ED-29, 590-596, 1982.
- [4] D.P. Kern, "Sub 0.1  $\mu\text{m}$  silicon MOSFETs", in Proc. ESSDERC conf., 633-642, September 1989.
- [5] T. Furuyama, "An experimental 4-Mbit CMOS DRAM", IEEE Journal of Solid-state circuits, SC-21, 605-611, October 1986.
- [6] T. Yamaguchi and T.H. Yuzuriha, "Process integration and device performance of a submicrometer BiCMOS with 16 GHz Ft double poly-bipolar devices", IEEE Trans. Electron Devices, ED-36, 890-896, May 1989.
- [7] D. D. Tang, P.M Solomon, T.H. Ning, D.D. Isaac and R.E. Burger, "1.25  $\mu\text{m}$  Deep-groove-isolated self-aligned bipolar circuits", IEEE Journal of Solid-state circuits, SC-17, 925-931, October 1982.
- [8] S. Tazawa, T. Takeda, K. Yokoyama, M. Tomizawa and A. Yoshii, "Application of two-dimensional process/device simulation for evaluating MOSFET fabrication processes", Solid-State Electronics, vol. 30, No 4, 375-381, 1987.
- [9] C.P. Ho, J.D. Plummer, S.E. Hansen and R.W. Dutton, "VLSI process modeling - SUPREM III", IEEE Trans. Electron Devices, ED-30, 1438-1453, 1983.
- [10] K. Sakamoto, K. Nishi, T. Yamaji, T. Miyoshi and S. Ushio, "Complete process modeling for VLSI multilayer structures", J. Electrochem. Soc., vol. 132, 2457-2463, 1985.
- [11] J.P. Krusius, " Process modeling of submicron complementary metal oxide semiconductor very large scale integrated circuits", J. Vac. Sci. Technology, A4 (3), 905-911, May/June 1986.
- [12] S.J. Polak, R.J.G. Goosens and N.E.B. Cown, "Requirements for device and process modeling of submicron devices", in Proc. ESSDERC conf., 119-130, September 1989.
- [13] D. Chin, M.R. Kump, H.G. Lee and R.W. Dutton, "Process design using two-dimensional process and device simulators", IEEE Trans. Electron Devices, ED-29, 336-340, 1982.

- [14] D. Collard and K. Taniguchi, "IMPACT - a point-defect-based two dimensional process simulator: modeling the lateral oxidation enhanced diffusion of dopants in silicon", IEEE Trans. Electron Devices, ED-33, 1454-1462, 1986.
- [15] A. Gerodolle, S. Martin and A. Marrocco, "Finite element method applied to 2D MOS process simulation and defect diffusion: program TITAN", in NASECODE IV conf., Boole Press, June 1985.
- [16] L. Borucki, H.H. Hansen and K. Varahramyan, "FEDSS - a 2D semiconductor fabrication process simulator", IBM J. Res. Develop., vol. 29, no 3, 263-276, 1985.
- [17] W.G. Oldham et al, "A general simulator for VLSI lithography and etching processes: part II - Application to deposition and etching", IEEE Trans. Electron Devices, ED-27, 1455-1459, 1980
- [18] S. Onga and K. Taniguchi, " A three-dimensional process simulator and its application to submicron VLSI's", Dig. Tech. papers, 1985 Symposium VLSI Technology, 68-69, May 1985.
- [19] N. Shigyo and R. Dang, "Three dimensional device simulation using a mixed process/device simulator", in Process and device modeling, ed. by W.L. Engl, p. 301, 1986.

## RESUME

Dans cette partie sont résumées les principales caractéristiques du travail entrepris et les points originaux y sont particulièrement soulignés. Les aspects suivants sont successivement passés en revue:

- i) Définition de la structure de données et génération de maillage.
- ii) Simulation des étapes élémentaires suivantes: dépôt/gravure, implantation ionique, diffusion et oxydation. Pour ces différents points, l'accent est essentiellement mis sur la résolution numérique et en particulier sur la résolution du problème multicouche.
- iii) Présentation d'une approche originale du maillage adaptatif incluant une évaluation des erreurs de discrétisation spatiale, ainsi que des rafraichissements de la grille. Cette nouvelle méthode est comparée aux stratégies classiques de maillage.
- iv) Trois applications sont brièvement présentées et l'influence de paramètres technologiques sur les performances de 2 technologies bipolaires est analysée.

Les points ayant donné lieu à publications y seront également référencés.

## I. STRUCTURE GENERALE ET OUTILS DE DEVELOPPEMENT

### 1.1 Structure de données

Etant donnée la complexité des structures multicouches à prendre en compte, il est essentiel de définir une structure de données apte à supporter ces domaines de simulation et à faciliter l'écriture du code. Dans ce cadre, une structure originale a été incorporée où les différentes informations sont regroupées en entités et accédées à partir de macro instructions. Ceci a été déterminé en étroite liaison avec le choix de la méthode des éléments finis qui ne souffre d'aucune limitation en ce qui concerne le maillage de structures complexes.

### 1.2 Génération de maillage

La génération de maillage est bien sûr le point clé concernant la discrétisation spatiale du domaine de simulation. Dans le cas de structures très importantes (en technologie BiCMOS par exemple), une grande précision dans les résultats n'est obtenue qu'au prix d'un grand nombre de noeuds dans certaines régions (source/drain ou zone d'émetteur/base), tandis que d'autres régions n'ont pratiquement aucun gradient de concentrations à supporter (substrat ou puits). Pour ce faire, des éléments triangulaires à trois noeuds ont été utilisés, avec des fonctions de forme linéaires. La stratégie générale pour le silicium est de définir deux zones, la première, de densité nodale importante, est destinée aux zones de forts gradients tandis que la deuxième zone, plus lâche, est réservée au puit ou au substrat [1]. Un tel maillage est représenté en figure 1. La génération elle-même fait appel à un générateur de maillage issu de l'INRIA, employant la triangulation de Delaunay, une étape de régularisation étant ensuite appliquée. Les points intérieurs sont générés sous forme d'une grille sur-imposée sur le domaine à mailler, les points extérieurs aux frontières étant éliminés. La figure 2 montre le maillage obtenu à la fin de la simulation d'un transistor bipolaire auto-aligné, à émetteur polysilicium et isolation par tranchée [2]. Comme on peut le voir sur cette figure, des grilles régulières sont utilisées pour les matériaux autres que le silicium.

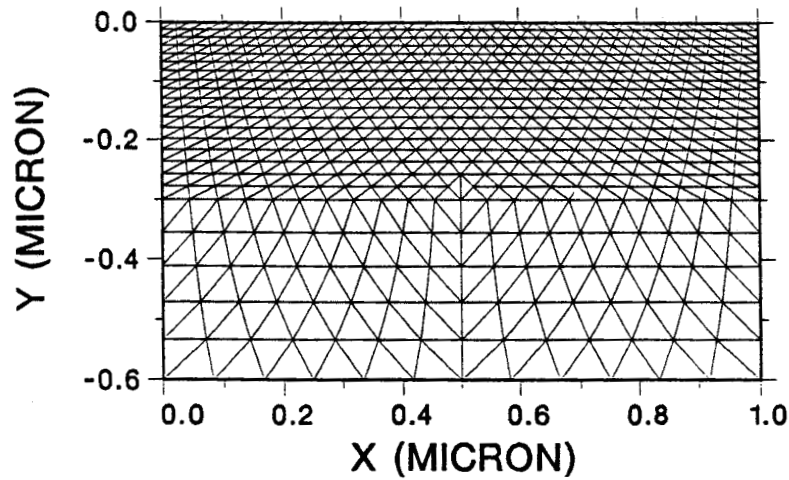


Fig. 1 Exemple de maillage initial g n r  pour le silicium.

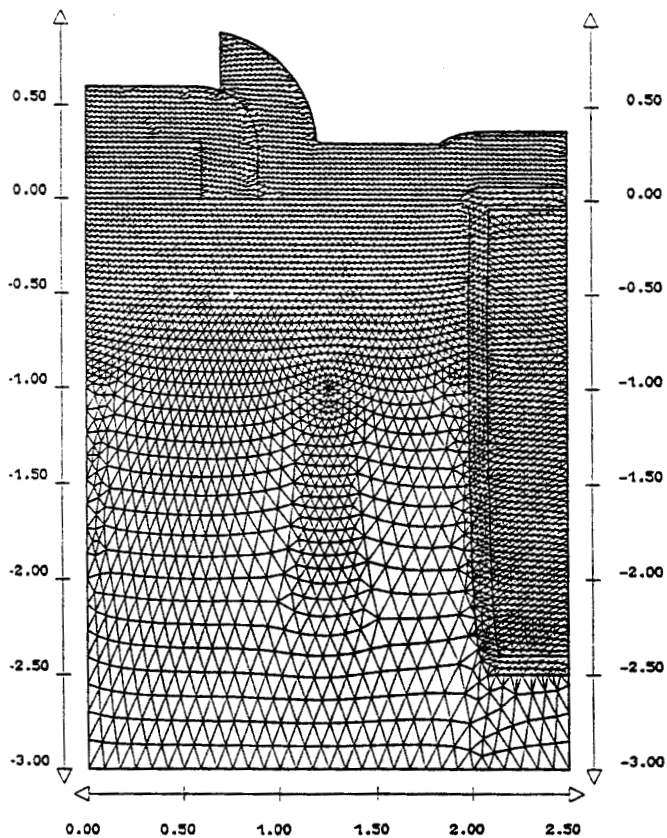


Fig. 2 Maillage   la fin d'une simulation compl te d'un transistor bipolaire, incluant plusieurs niveaux de polysilicium et une isolation par tranch e [2].

## II. LES ETAPES ELEMENTAIRES

### 2.1 Dépôt et gravure

#### 2.1.1 Modèles

Comme il n'existe pas de formulation générale des vitesses de dépôt ou de gravure en fonction d'un appareillage déterminé, des formulations analytiques sont utilisées. C'est ainsi que 4 composantes peuvent être combinées en ce qui concerne la gravure. En plus de modèles classiques tels qu'isotropique ou anisotropique, 2 composantes permettent de simuler finement les étapes de gravure sèche. Dépendant de l'angle solide, ces 2 composantes reproduisent précisément la formation de tranchées dans le silicium comme illustré dans la figure 3, où une comparaison est effectuée entre des résultats expérimentaux et la simulation avec IMPACT 4. Le dépôt permet quant à lui l'étude de la couverture de marche avec une composante isotropique ou un modèle de type source hémisphérique.

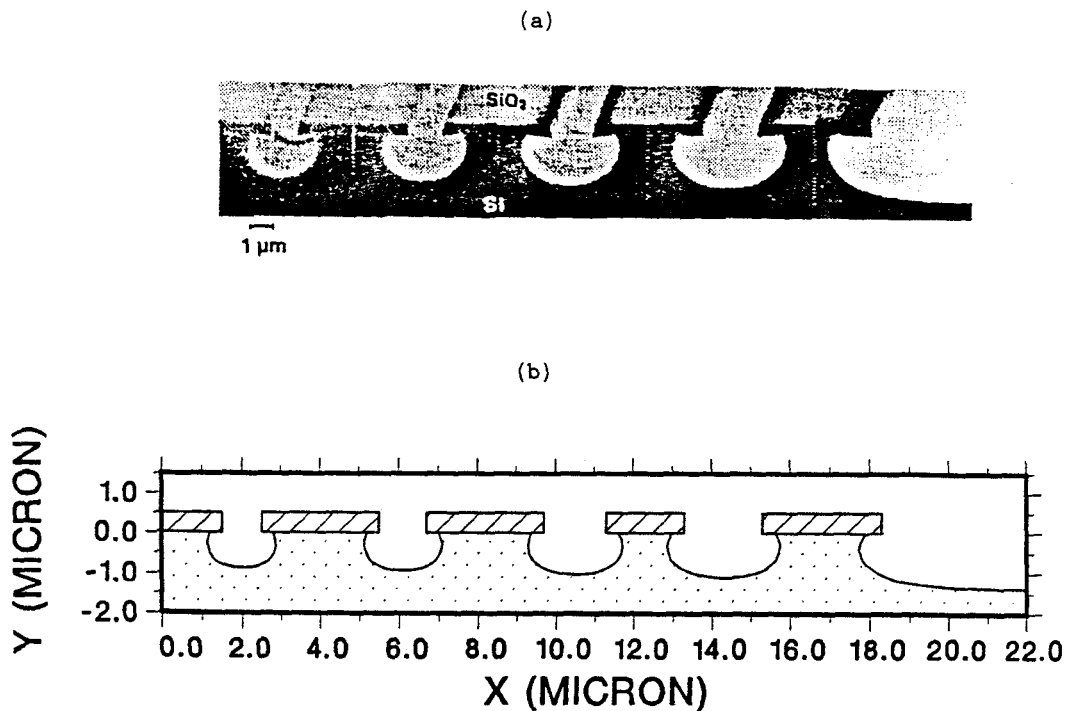


Fig. 3 Formation de tranchées dans le silicium par gravure sèche.

(a) photographie MEB d'un résultat expérimental.

(b) résultat de la simulation avec IMPACT 4.

### 2.1.2 Algorithmes

L'évolution des profils lors de ces étapes est calculée en utilisant la méthode classique dite du "string model". Hormis les problèmes usuels intervenant dans ce type de méthode, deux approches originales ont été développées. La première concerne le lien entre le string modèle et le maillage utilisé dans les étapes de diffusion. L'approche unifiée consiste à considérer les points du string modèle comme étant les sommets des triangles (du maillage élément fini) situés sur les frontières. Ceci évite toute interpolation de formes entre les différents modules et autorise une définition suffisamment précise des enveloppes de chaque couche.

Le deuxième point original concerne la gestion des domaines après ces étapes, considérant que des domaines en contact et de même matériaux sont automatiquement réunis en un seul. De la même manière, les étapes de gravure peuvent entraîner la formation de nouveaux domaines [3]. Ceci est illustré sur la figure 4.

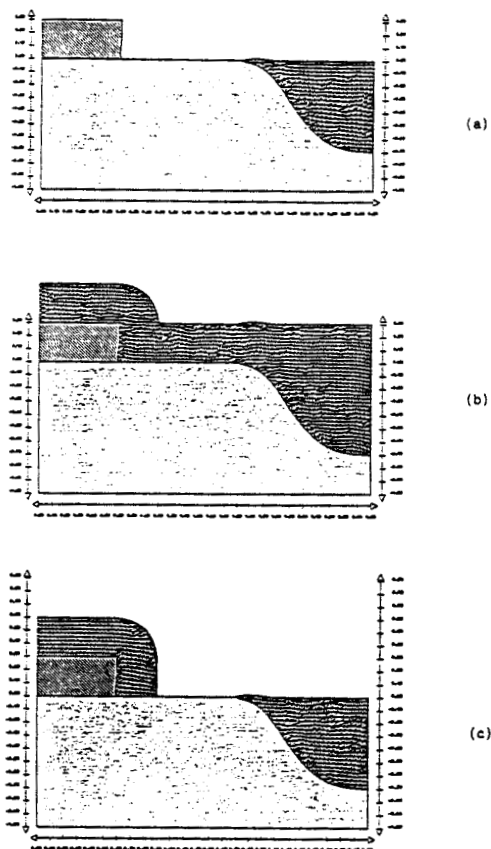


Fig. 4 Exemple d'évolution de domaines oxyde lors d'étapes de dépôt et de gravure, avec une couche de polysilicium et un oxyde de type enterré, comme structure initiale (a) [3].



## 2.2 Implantation ionique

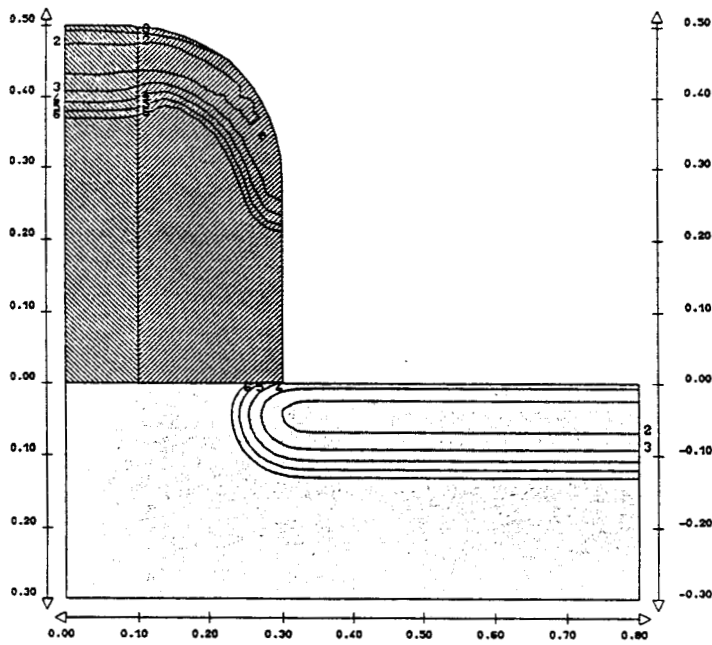
### 2.2.1 Modèles

Bien que des méthodes numériques telles que la méthode de Monte Carlo permettent des résultats très intéressants, les contraintes liées au temps de calcul entraînent l'utilisation de formulations analytiques pour les distributions verticales d'implantation (Gaussienne, Pearson IV...). Les paramètres sont extraits de nombreux résultats expérimentaux avec en particulier une tabulation pour les implantations de  $\text{BF}_2$ .

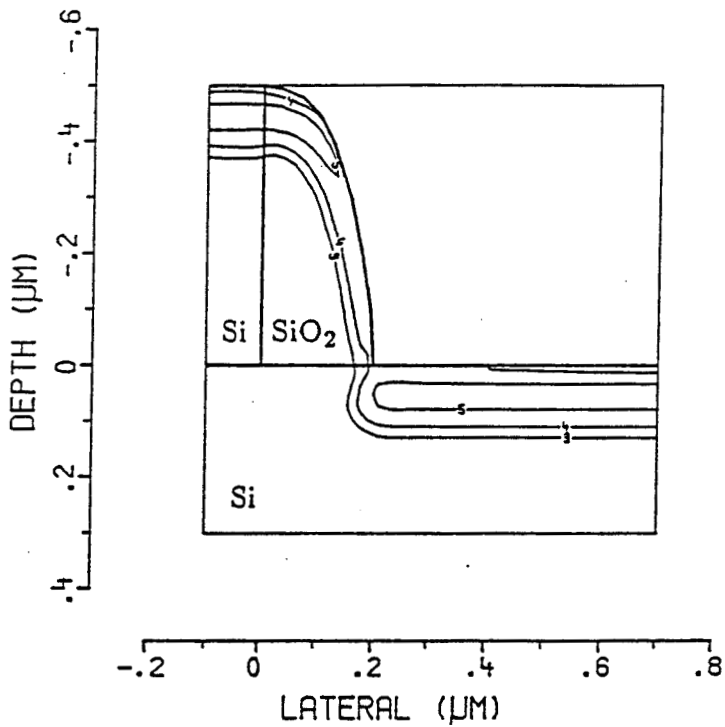
### 2.2.2 Méthodes numériques

Le problème multicouche est résolu en employant une méthode basée sur le calcul des doses implantées dans chaque couche ("dose matching"). La faible différence de densité entre les matériaux usuels en technologie silicium autorise des calculs précis avec cette méthode. Dans cette étude, la méthode a été étendue au cas bidimensionnel, où le choix des différents paramètres (déviations latérales en particulier) permet des résultats tout à fait satisfaisants. Ce dernier point est illustré avec la figure 5 qui montre une comparaison entre ces calculs et un résultat obtenu avec la méthode de Monte Carlo. Il convient de noter que l'efficacité de notre méthode en terme de temps de calcul a été grandement augmentée par la détermination de tabulations avant l'étape d'implantation elle-même.

De plus, le faisceau d'implantation peut être incliné par rapport au substrat, afin d'étudier les phénomènes d'ombrage. Enfin, la déviation latérale utilise un modèle très fin, déduit de simulations de Monte Carlo, où ce paramètre dépend bien sûr de l'énergie d'implantation, mais aussi de la profondeur. L'impact de ces modélisations sur les profils de dopage source/drain de transistors MOS a été évalué.



(a)



(b)

Fig. 5 Comparaison entre des calculs analytiques (a) et la méthode de Monte Carlo (b), pour une implantation d'arsenic sur une structure de type LDD-MOS.

## 2.3 Diffusion

### 2.3.1 Modèles

Afin de se focaliser sur les phénomènes intervenant aux interfaces (ségrégation des impuretés), nous utilisons un niveau de modélisation largement validé par de nombreux résultats expérimentaux et d'autres simulateurs de procédé. Dans ce cas, on suppose que la diffusion se produit essentiellement sous forme lacunaire, la contribution de chaque type de lacunes étant incorporée sous la forme du produit d'un coefficient de diffusion et d'un terme dépendant de la concentration en électrons libres. Les phénomènes de formation d'agrégat ainsi qu'une modélisation plus détaillée de la diffusion du phosphore sont également inclus. Enfin, dans une première étape, le polysilicium est considéré comme ayant les mêmes propriétés que le silicium; une équation de diffusion à coefficients constants régit la redistribution des dopants dans l'oxyde et le nitrure de silicium.

### 2.3.2 Résolution

La méthode des éléments finis est utilisée pour la résolution des équations de diffusion. Une technique de résidus pondérés et la méthode de Galerkin permettent d'obtenir un ensemble d'équations couplées dont les inconnues sont les concentrations en chaque point du maillage. La résolution des problèmes non-linéaires et non-stationnaires emploie un schéma implicite incomplet qui se révèle plus efficace qu'une méthode de Newton-Raphson classique (figure 6).

L'une des difficultés majeures introduites par l'aspect multicouches consiste en la résolution du système linéaire. La généralité des structures ne permet pas un numérotage optimal des noeuds, aussi une méthode directe semble difficilement applicable. Pour cette raison, une solution itérative à été utilisée: la méthode des gradients conjugués avec préconditionnement LDU. Cependant, l'introduction des flux de ségrégation rend la matrice non diagonale dominante et non symétrique. Une mise en forme de cette matrice, par le biais des coefficients de ségrégation de chaque matériau, permet de retrouver un meilleur conditionnement de la matrice [4]. La figure 7 montre ainsi l'évolution de la norme du vecteur résiduel en fonction du nombre d'itérations, pour plusieurs valeurs de pas de temps.

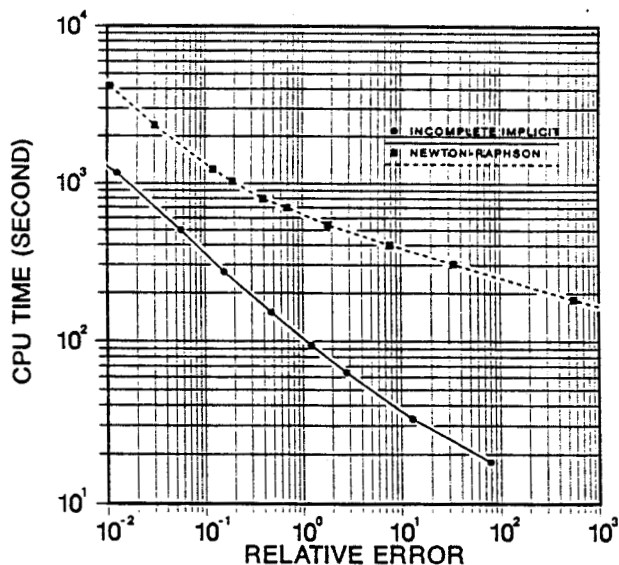


Fig. 6 Evolution du temps de calcul en fonction de l'erreur relative pour les methodes implicite incomplète et de Newton-Raphson.

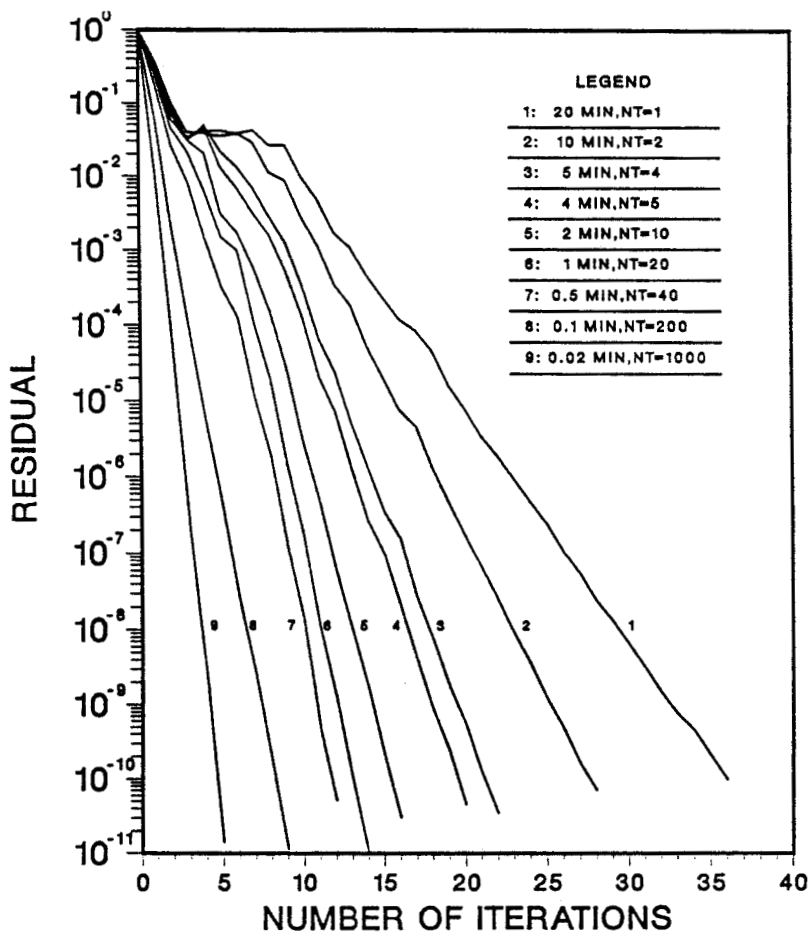


Fig. 7 Norme du vecteur résiduel en fonction du nombre d'itérations. Les courbes sont déterminées pour différentes valeurs du pas de temps (de 20 minutes à 1,2 seconde), sur une diffusion classique de type source/drain.

## 2.4 Oxydation

### 2.4.1 Modèles

La modélisation de Deal et Grove est bien sûr utilisée, en conjonction avec des termes additionnels permettant de prendre en compte les vitesses d'oxydation très rapides lors d'oxydations sèches (modèle de Massoud, Plummer et Irene). De même, l'effet du dopage substrat est inclus dans la détermination des paramètres. Enfin, nous avons fait le choix d'une modélisation de l'oxydation bidimensionnelle sous forme d'un fluide visqueux incompressible, négligeant les effets non-linéaires. Ce choix a plusieurs intérêts: la possibilité d'oxyder n'importe quelle structure (comparé à des formulations analytiques); un niveau de modélisation suffisant pour des températures plus élevées que 960°C avec des effets non-linéaires peu prononcés; la possibilité d'extensions futures vers des formulations plus physiques, prenant en compte l'effet des contraintes.

### 2.4.2 Résolution

Les points marquants de cette simulation sont les suivants:

- i) Initialisation de la structure par un dépôt sélectif d'oxyde, pour des simulations incluant à la fois le silicium et plusieurs couches d'oxyde, de nitrure et de polysilicium.
- ii) Diffusion des agents oxydants (équation de Laplace). La détermination fine du coefficient de réaction à l'interface Si/SiO<sub>2</sub> permet de calculer des épaisseurs d'oxyde en 2D aussi précisément que pour des calculs plus simples sur des structures 1D.
- iii) Résolution de l'écoulement du fluide visqueux incompressible (équations de Navier-Stockes). Pour cela, la méthode des éléments finis est appliquée. Il est à noter que le même type d'éléments triangulaires -et donc de maillage- est utilisé pour la diffusion des agents oxydants (et des dopants), et pour la résolution de l'équation de Navier-Stockes [5]. La condition d'incompressibilité est obtenue quant à elle par des itérations sur la pression et la vitesse (algorithme de Chorin).

iv) Mise à jour de la structure et remaillage dans tous les domaines (sauf ceux de matériaux nitrure).

La figure 8 montre un exemple d'application sur une structure complexe incluant une tranchée silicium. La structure initiale est reportée en figure 8a, tandis que la figure 8b montre le résultat final après oxydation humide.

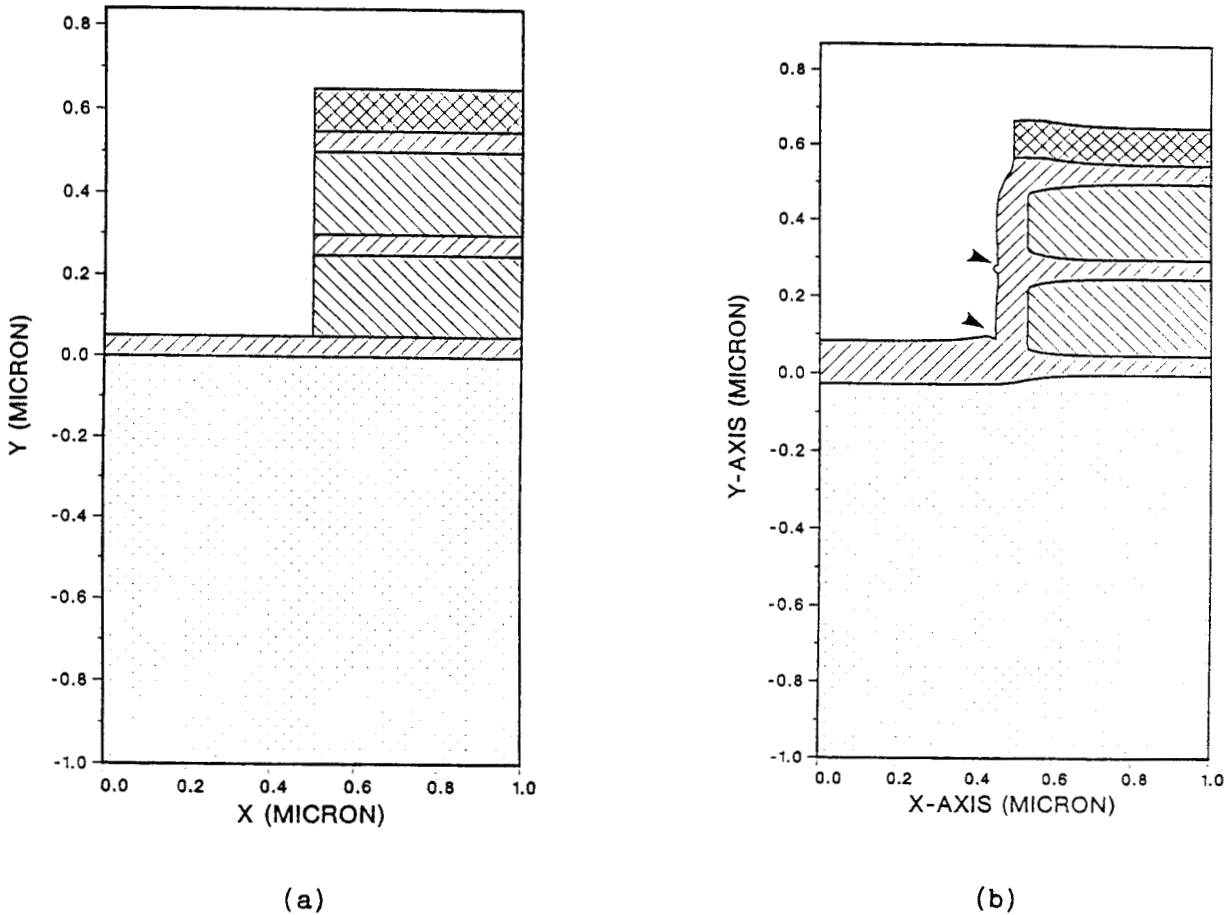


Fig. 8(a) Structure initiale contenant:

une couche de nitrure (haut de la structure), 2 couches de polysilicium, 3 couches d'oxyde et le substrat de silicium.

(b) Résultat final après oxydation humide (1000°C, 5 mn). L'effet d'écoulement de type fluide visqueux incompressible est mis en évidence par les 2 flèches.

### 2.4.3 Couplage avec la diffusion des impuretés

Des procédures ont été développées afin de reproduire correctement d'un point de vue numérique la diffusion des impuretés pendant les étapes d'oxydation. Pour ce faire, les procédures d'interpolation ont été particulièrement étudiées et permettent d'obtenir une bonne conservation de la dose (confortée par des flux additionnels aux interfaces silicium et polysilicium/oxyde), tout en garantissant la stabilité numérique. A titre d'exemple, la figure 9 montre une comparaison entre un calcul effectué avec IMPACT 4 et SUPREM II. Il a d'autre part été montré que ce schéma est compatible avec les conditions temporelles liées à la résolution de l'équation de diffusion (schéma implicite incomplet).

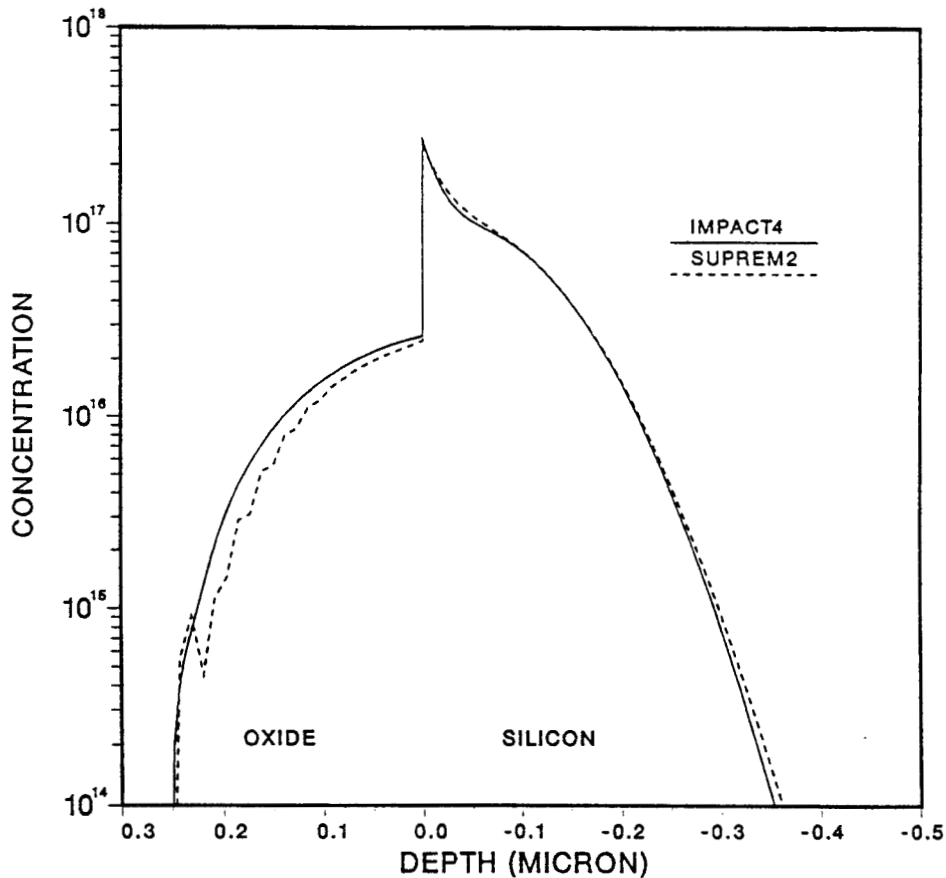


Fig. 9 Comparaison entre les résultats de simulation d'IMPACT 4 et de SUPREM II, dans le cas d'une oxydation humide à 1000°C sur un substrat dopé phosphore.

### III. MAILLAGE ADAPTATIF [6,7]

#### 3.1 Motivations

Bien que les méthodes présentées ci-dessus soient suffisantes dans les cas les plus courants, elles souffrent cependant de limitations dans le cas de très grandes structures (technologie BiCMOS, implantation dans le fond d'une structure d'isolation par tranchée). Par ailleurs, aucune estimation de l'erreur due à la discrétisation spatiale n'est effectuée. Ceci a motivé une étude sur la possibilité d'un maillage adaptatif pour la simulation de procédé multicouches.

#### 3.2 Techniques numériques et génération de maillage

Afin de satisfaire les besoins spécifiques introduits par la simulation de procédé (et surtout son aspect non-stationnaire), des procédures de raffinement ont été mises au point, basées essentiellement sur la limitation de la connectivité associée à chaque noeud. L'utilisation de ces procédures, ainsi que la triangulation de Delaunay et la régularisation, permettent d'obtenir une bonne transition entre les zones fortement raffinées et les zones plus lâches.

Le critère choisi pour le raffinement d'un triangle donné réside en fait en la limitation du gradient maximal de concentration admissible sur chaque élément ( $\alpha$  par la suite). Ce paramètre  $\alpha$  dépend de plus du niveau de concentration. Par ailleurs, une méthode de rafraîchissement automatique de la grille a été développée, ce qui permet de minimiser le nombre de noeuds nécessaires à une bonne prise en compte des profils de diffusion. L'évolution de  $\alpha$  durant les étapes de diffusion permet de déterminer l'occurrence de tels rafraîchissements. La figure 10 montre une structure de type tranchée avec le maillage initial (10a), le maillage après raffinement (10b), et les isovaleurs de bore après implantation d'anti-canal (10c).



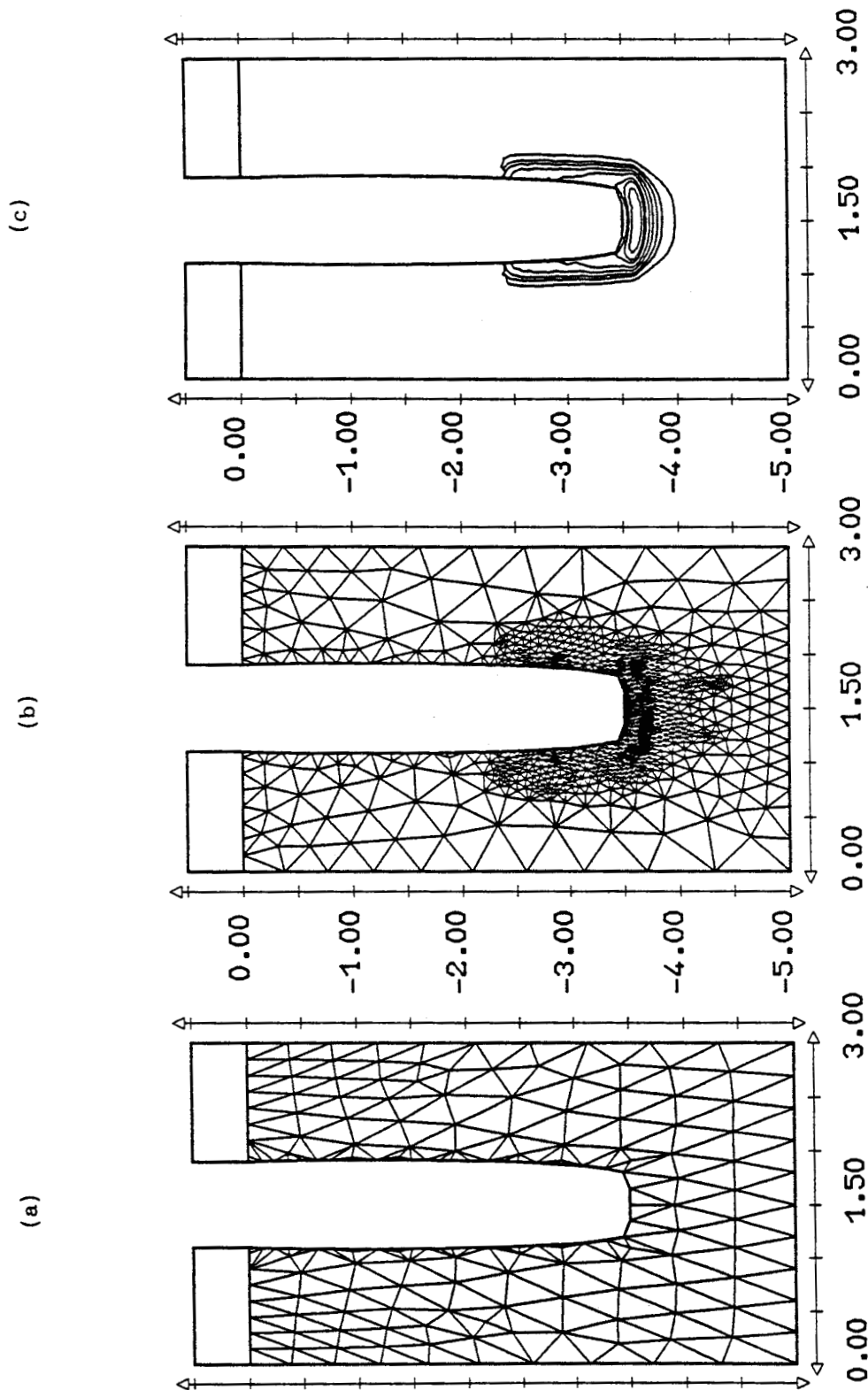


Fig. 10 Implantation ionique de bore dans une structure comportant une tranchée silicium. 6 itérations sont nécessaires entre le maillage initial (a) et le maillage raffiné (b). Les isovaleurs de bore sont également représentées (c).

### 3.3 Exemples test (structures bipolaires)

L'erreur liée à la discrétisation spatiale est estimée en fin de simulation par le calcul analytique de la tension de perçage de structures test bipolaires, dans le cas de la diffusion d'une seule impureté, ou pour deux impuretés couplées [8]. Ceci a permis de déterminer qu'en dessous d'une taille de maille de 50 Å, le maillage n'a pratiquement plus d'influence sur le résultat final. Par ailleurs, il a été montré que l'utilisation de nos procédures permet d'obtenir des résultats très précis en 2D (même avec un maillage initial de l'ordre de 0.5  $\mu\text{m}$ ), avec en plus une réduction significative du temps de calcul nécessaire à l'obtention d'une précision donnée (figure 11). Enfin, il a été déduit qu'une valeur de 5 pour  $\alpha$  permet d'obtenir des résultats très précis.

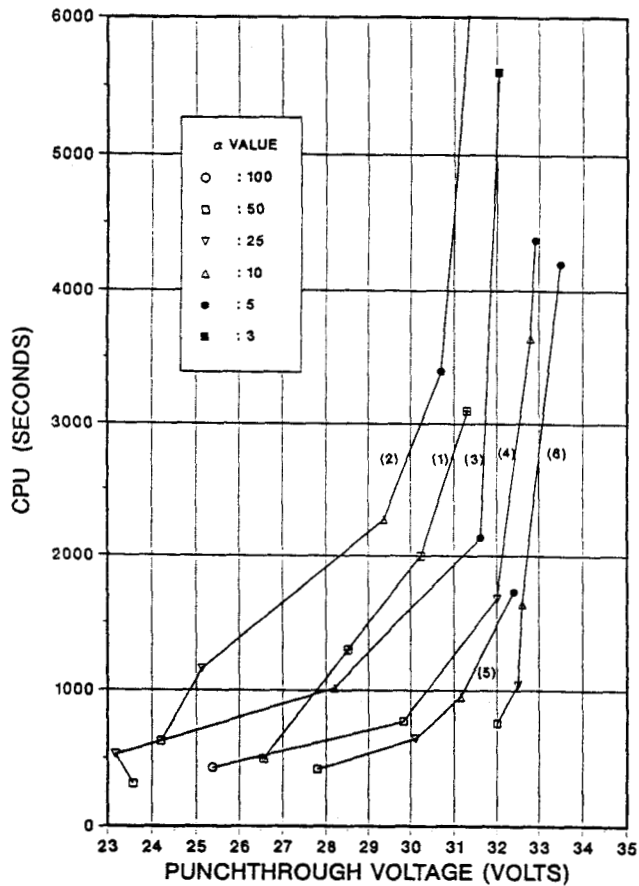


Fig. 11 Temps de calcul nécessaire pour obtenir une certaine précision (spécifiée par la tension de perçage) dans le cas de diffusion de l'arsenic pour la formation d'un émetteur. Le résultat exact est 34,5 Volts. La courbe 1 représente une stratégie de maillage fixe (méthode usuelle) et la courbe 6 la combinaison la plus efficace des différentes procédures utilisées pour le maillage adaptatif.

Ces procédures ont par ailleurs été confrontées à des cas plus complexes soit d'un point de vue structurel (transistor bipolaire à émetteur polysilicium et isolation par tranchée), ou soit par des approches couplées simulation de procédé/simulation de dispositif [9]. Ainsi, la figure 12 montre l'effet du raffinement de maillage sur le gain d'un transistor bipolaire, et une comparaison avec une caractéristique expérimentale. Ce type de résultat montre que des calculs aussi précis peuvent être obtenus en 2D avec des procédures de maillage adaptatif, qu'en 1D, où il n'y a pratiquement aucune limitation en ce qui concerne la discrétisation spatiale.

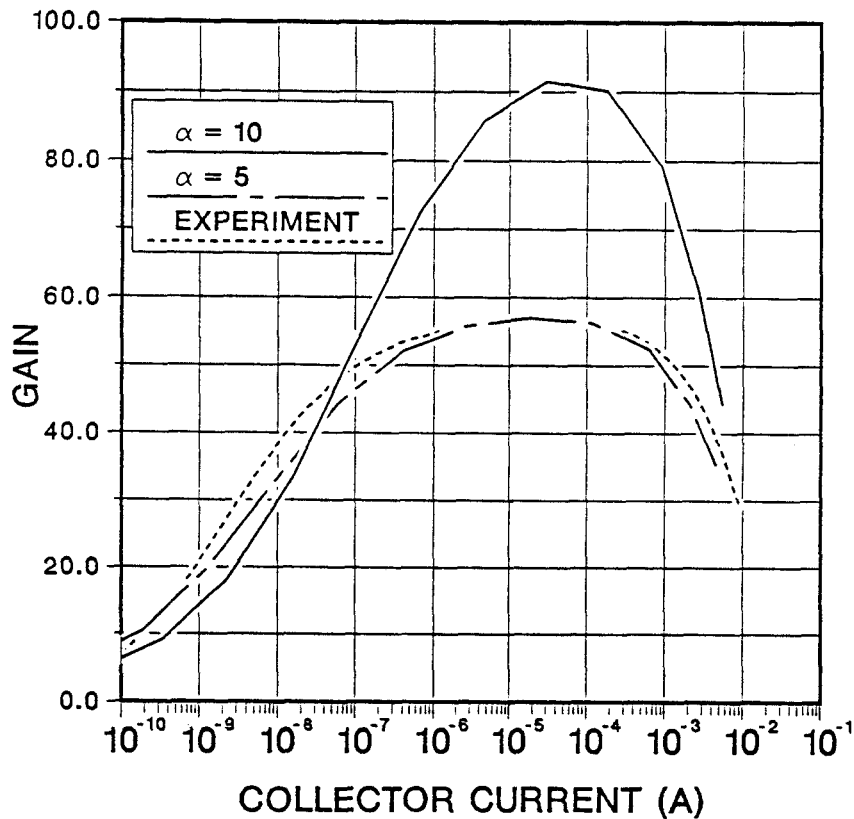


Fig. 12 Gain en fonction du courant collecteur d'un transistor bipolaire NPN, pour différentes valeurs de  $\alpha$ . Une comparaison avec des résultats expérimentaux est également représentée.

#### IV. APPLICATIONS

##### 4.1 Transistor LDD-MOS

Un transistor LDD-MOS a été simulé, incluant la formation de la grille par attaque sèche, et de l'espaceur par des étapes successives de dépôt et de gravure d'oxyde. L'étape finale de re-oxydation a enfin été calculée avec le niveau de modélisation décrit en 2.4.1. A titre d'exemple, la figure 13 montre les isovaleurs d'arsenic à la fin du procédé, où l'on peut voir que tous les effets de ségrégation sont bien respectés. De même, les isovaleurs d'arsenic dans l'oxyde sont le reflet du déplacement bidimensionnel de l'oxyde lors de l'étape d'oxydation, conjugué avec les différents phénomènes de ségrégation (incorporant par exemple cette impureté dans l'oxyde de grille).

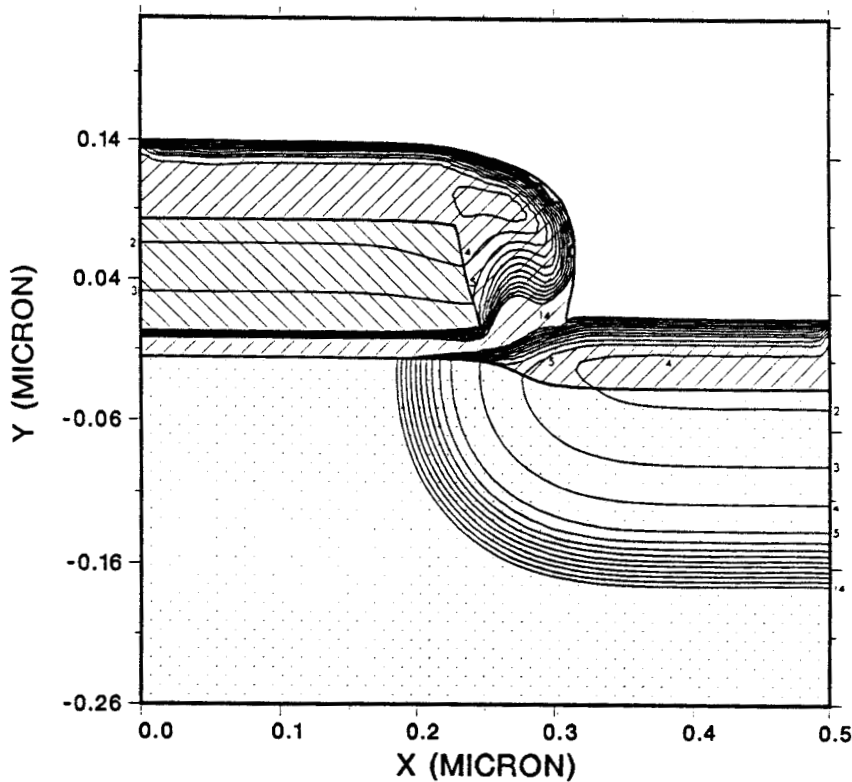


Fig. 13 Isovaleurs d'arsenic à la fin de la simulation d'un transistor LDD-MOS.

1:  $3 \cdot 10^{21}$ , 2:  $1 \cdot 10^{21}$ , 3:  $3 \cdot 10^{20}$ , 4:  $1 \cdot 10^{20}$ ...  
14:  $1 \cdot 10^{15}$  at/cm<sup>3</sup>.

#### 4.2 Effets parasites intervenant dans un transistor bipolaire à émetteur muré

Une étude systématique effectuée sur une technologie bipolaire à émetteur muré a permis de déterminer les conditions technologiques responsables d'un fonctionnement parasite en perçage, ainsi que l'endroit exact où apparait ce courant [10,11]. Il a été ainsi montré qu'une zone de rétrécissement de la largeur de base peut apparaitre dans la zone du bec d'oiseau, par la suite d'une modification des conditions d'implantation (contrôle insuffisant d'une étape de gravure). L'évolution des jonctions émetteur/base pour 3 valeurs différentes de sur-gravure est ainsi reportée sur la figure 14. Des simulations de dispositif utilisant ces résultats ont confirmé que la zone de pincement de base n'intervient pas le long de l'interface silicium/oxyde mais bien dans le substrat de silicium [12]. Cette étude a donc permis d'expliquer les dispersions observées sur une technologie réelle.

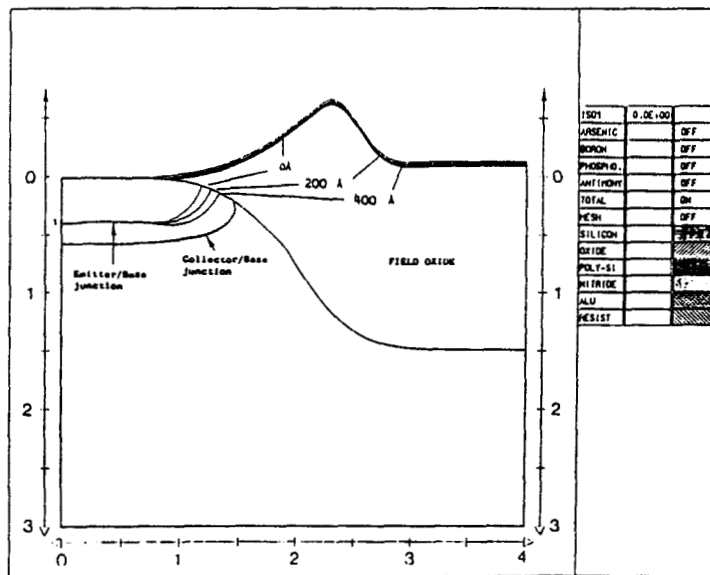


Fig. 14 Positions des jonctions dans le transistor bipolaire à émetteur muré suivant 3 valeurs de la sur-gravure d'oxyde (intervenant avant l'implantation ionique d'arsenic pour l'émetteur).

### 4.3 Application à une technologie bipolaire avancée [2]

Les possibilités du programme ont par ailleurs été testées sur un transistor bipolaire auto-aligné à émetteur et base extrinsèque en polysilicium, isolé par tranchée. L'oxyde natif situé à l'interface polysilicium/silicium a été pris en compte par une condition de ségrégation pour la diffusion des dopants. Un résultat typique est montré sur la figure 15, où l'on peut reconnaître les 3 zones de dopage de type P<sup>+</sup>: l'anti-canal dans le fond de la tranchée, la base extrinsèque (obtenue par diffusion depuis le polysilicium), et enfin la base intrinsèque (d'abord créée par une implantation auto-alignée de BF<sub>2</sub> puis un recuit neutre).

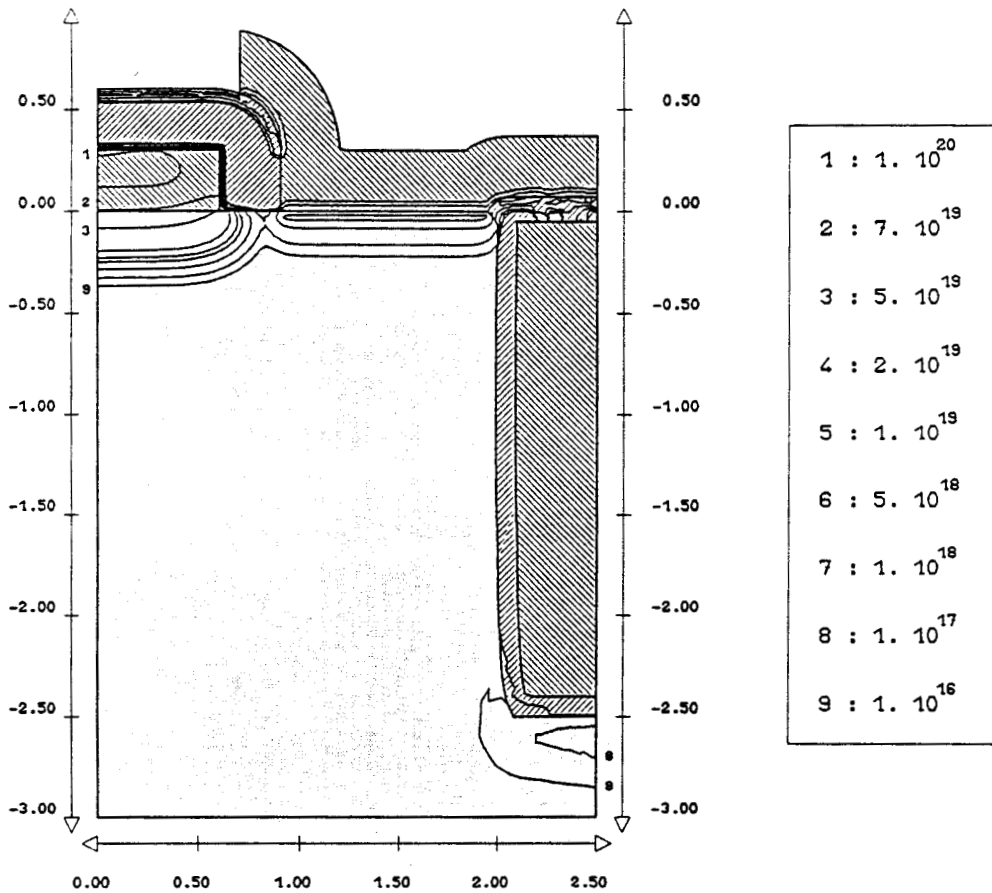


Fig. 15 Isovaleurs de bore à la fin de la simulation du transistor bipolaire, incluant 2 niveaux de polysilicium, un espaceur en oxyde, une tranchée dans le silicium isolée par un oxyde fin et remplie par du polysilicium.

Ce type de simulation permet d'étudier le lien entre les bases extrinsèque et intrinsèque -facteur déterminant pour une bonne tenue en tension de claquage et de perçage du dispositif- en fonction de la largeur de la couche d'oxyde entre les deux niveaux de polysilicium. La figure 16 montre l'évolution des profils de dopage le long de la surface du silicium, pour trois largeurs différentes de l'espaceur d'oxyde, montrant qu'une largeur d'environ 0.2  $\mu\text{m}$  est nécessaire pour un lien correct entre les deux bases. C'est là un exemple typique où l'utilisation d'un simulateur de procédé multicouches se justifie pleinement et permet d'obtenir des résultats utiles pour l'optimisation de procédés technologiques avancés.

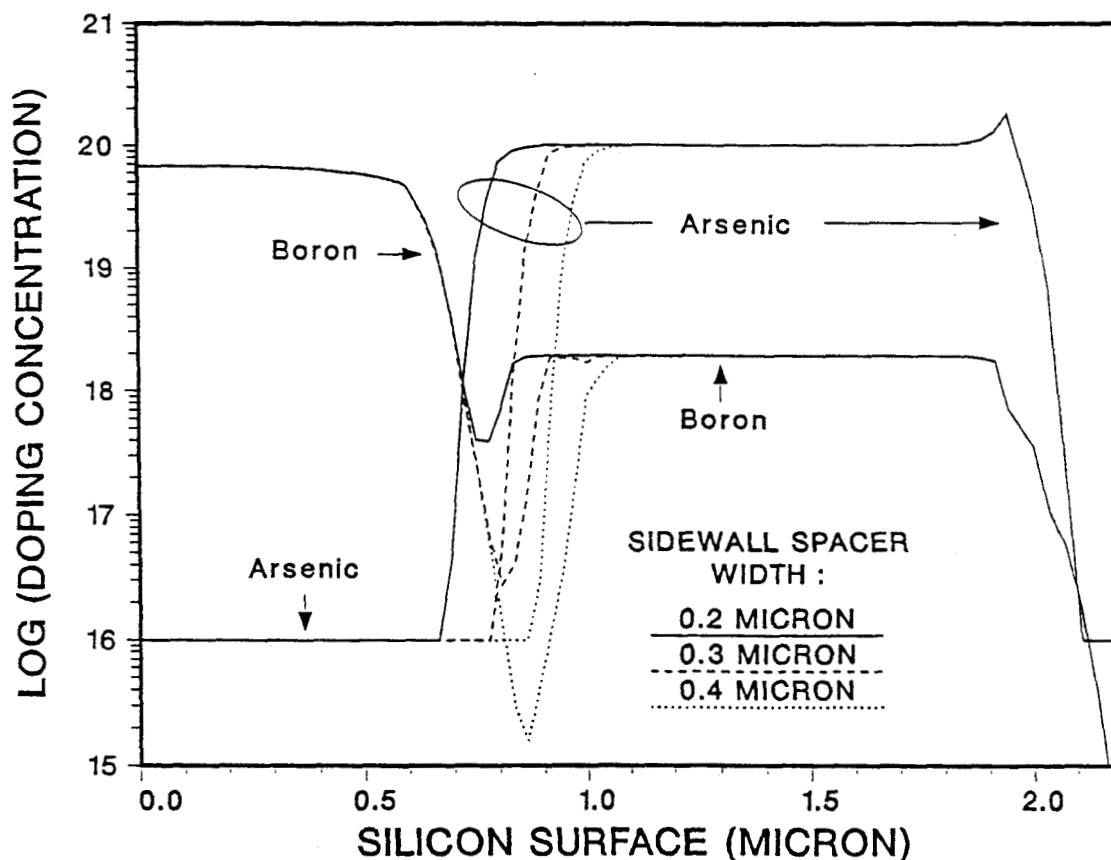


Fig. 16 Concentration en arsenic et bore le long de la surface du silicium, dans le cas de 3 largeurs de l'espaceur d'oxyde. Le lien entre les 2 bases est clairement insuffisant pour une largeur de l'ordre de 0.4  $\mu\text{m}$ .

REFERENCES BIBLIOGRAPHIQUES :

- [1] D. Collard, B. Baccus, E. Dubois and D. Morel, "IMPACT 1-2-3: an integrated 2D process/device simulator for MOS technology", in lecture notes of the software forum, NASECODE VI conf., 16-31, July 1989.
- [2] B. Baccus, E. Dubois, D. Collard and D. Morel, "Efficient two-dimensional multilayer process simulation of advanced bipolar devices", Solid-State Electronics, Vol. 32, No 11, 1013-1023, November 1989.
- [3] B. Baccus, D. Collard, E. Dubois and D. Morel, "Two-dimensional multilayer process simulation of bipolar devices using a multilayer process simulator: IMPACT 4", BCTM conf., Minneapolis, 164-167, September 1988.
- [4] B. Baccus, D. Collard, E. Dubois and D. Morel, "IMPACT 4 - a general two-dimensional multilayer process simulator", in proc. SISDEP III conf., 255-266, Eds G. Baccarani and M. Rudan, Bologna, September 1988.
- [5] D. Collard, B. Baccus and E. Dubois, "Two-dimensional multilayer process simulation", VLSI process/device modeling workshop, 15-16, Osaka, May 26-27, 1989.
- [6] B. Baccus, D. Collard and E. Dubois, "Adaptive mesh refinement for multilayer process simulation", accepté NUPAD III conf., June 3-4 1990, Hawaii.
- [7] B. Baccus, D. Collard and E. Dubois, "Adaptive mesh refinement for multilayer process simulation using the finite element method", soumis IEEE Trans. CAD.
- [8] B. Baccus, D. Collard and E. Dubois, "Dynamic mesh refinement for multilayer process simulation", in lecture notes of the software forum, NASECODE VI conf., 108-113, July 1989.
- [9] E. Dubois, J.L. Coppée, B. Baccus and D. Collard, "A study of the electrical performances of isolation structures", accepté IEEE Trans. Electron Devices.
- [10] B. Baccus, D. Collard et E. Dubois, "Simulation bidimensionnelle de technologie bipolaire", Journées GCIS, contribution orale, Mai 1987.
- [11] B. Baccus and D. Collard, "Two-dimensional study of Subilo-N technology", research contract for RTC PHILIPS Composants, ISEN, November 1988.
- [12] E. Dubois, B. Baccus and D. Collard, "Analysis of the punchthrough effect in walled emitter bipolar transistors", in proc. ESSDERC conf., 789-792, September 1989.



## CHAPTER II

### GENERAL STRUCTURE AND DEVELOPMENT TOOLS

#### Abstract

In this chapter, some of the problems related to the data structure of the process simulator IMPACT<sup>4</sup> are discussed, and the use of a data base is justified. The Finite Element Method is also briefly presented, considering some of the implications in mesh generation. A regular or progressive grid structure is defined, inducing slight refinement, determined *a priori*.

## I. PROBLEM DEFINITION

The development of a simulation program is a long and tedious task, referring to the software and programming aspects. The architecture and data structure are key points: they facilitate the writing itself, but also the possible evolutions of the program, for new models and material incorporations. An other aspect deals with the interface of the program with other simulators for coupled process/device approaches. Finally an easy 'debugging' is possible only if a modular structure has been previously defined. Hence, the first purpose of this chapter is to present briefly the solutions that have been used in data structure, and to define the advantages and major limitations.

The final goal of such a program is the optimisation of technologies, but also the test and definition of new models. However, in order to analyse the simulations results with complete confidence, we must be sure of the precision and reliability of the numerical procedures. The techniques used in the resolution of the diffusion equation will be detailed in the next chapter, but here we will introduce some problems related to mesh generation. A first solution is presented here, that is often acceptable in usual simulations.

## II. FINITE ELEMENT METHOD AND DATA ORGANIZATION

### 2.1 Choice of the Finite Element Method

One of the basic problem to be solved is the diffusion equation. We have then to choose a numerical method, compatible with the overall definition of the program. Two solutions can be examined: rectangular Finite Difference Method (FDM) and Finite Element Method (FEM). The first one has been extensively used in the past, due to the ease in programmation, the small memory consumption and also the possibility to use

efficient linear solver, exploiting the matrix structure. Although some process simulators use advanced version of this method as done in COMPOSITE [1], it does not seem to be adequate in the most general case. They are severe restrictions associated, such as the generality in the silicon shapes (trench and overhang for example). Finally the extension to the multilayer case has been shown to be very difficult [2].

On the other hand, FEM has strong advantages. The choice of element type can be done upon the special characteristics to be treated (e.g. degree of non linear formulation). Particularly, triangles allow an easy definition of complex shapes, without limitations. Finally, local mesh refinements can be defined and allow good compromise between precision and number of points. From the numerical point of view, dose conservation is guaranteed even with a very coarse mesh, it is a very important point in our case. One of the drawback is the matrix structure, often sparse and not regular compared to FDM solution. It has some consequences on the computer ressources, but the compromise with the above cited advantages leads to the choice of FEM.

## 2.2 Data structure

The data structure is now defined, in connection with the choice of FEM and the multilayer characteristic. Several general data structures for process/device simulation have been proposed, among them one of the most sophisticated is the DAMSEL code [3]. Here we have chosen a more specialized data base, defined primary for FEM programs and their graphic post-processors [4].

Generally speaking (Fig. 1), the data used during calculations are classified into several categories: domains (layers), elements (in the present study, only 3-noded triangular elements have been used but other types can be supported), nodes, material properties, impurity models, characteristics of the current step... All these entities are accessed through the data base as shown in fig. 2. Each domain is characterized by its material, the list of the boundary nodes, internal nodes and elements if the domain is meshed. The entities are then created only when they are needed to optimize the memory requirements. It is important to point out that there is no hierarchy in the access to entities and hence no practical limitation occurs in data manipulation.

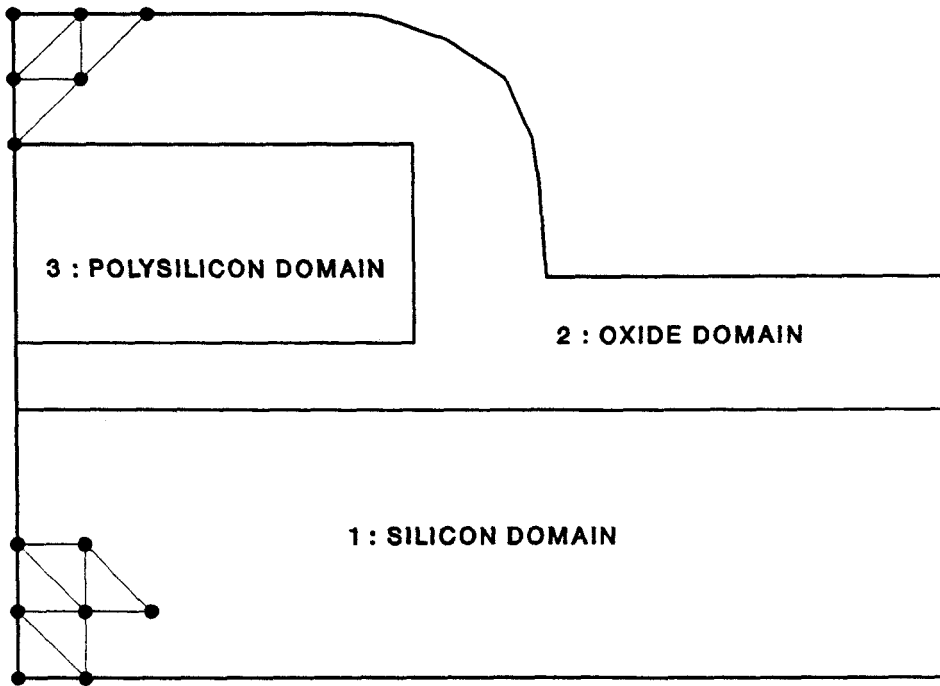


Fig. 1 Typical example of a simulation containing 3 domains (layers) with materials silicon, oxide and polysilicon. The domains 1 and 3 are meshed with triangular elements (3 nodes per element).

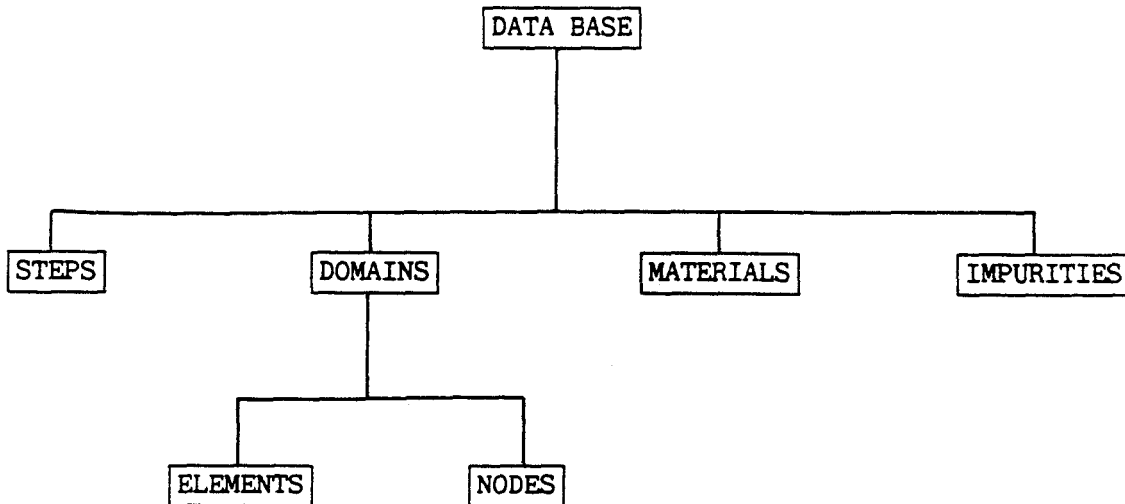


Fig. 2 General entities for 2D multilayer problems. A hierarchical structure is presented for the nodes, elements and domains in order to clarify the structure, although nodes or elements can be accessed directly.

The informations about each individual of an entity (node for example) are obtained by the following Fortran call :

Read mode : CALL GETNOD(idom,inod,iret)

Write mode : CALL PUTNOD(idom,inod,iret)

where idom is the number of the concerned domain (1 for silicon, for example), inod the node number in this domain, and iret the flag indicating the existence of this individual. The informations are transfered through COMMON statements:

```
COMMON /NODCOM/ NUMNOD,IBNNOD(2),ISTNOD,CORNOD(2),CAS,CBO,CPH,CSB,  
              COX
```

NUMNOD : number associated by the user

IBNNOD : boundary conditions

IBNNOD(1) :

= -2 sides or bottom

= -1 air

= 0 internal

= 1 the neighbor domain is 1 (silicon)

= 2 the neighbor domain is 2 (oxide or polysilicon...)

= -3 triple point (point common to 3 domains)

IBNNOD(2) :

if IBNNOD(1) > 0, boundary node number in domain  
number IBNNOD(1)

ISTNOD : status (active or inactive)

CORNOD : coordinates

CAS : arsenic concentration

CBO : boron concentration

CPH : phosphorus concentration

CSB : antimony concentration

COX : oxidant concentration

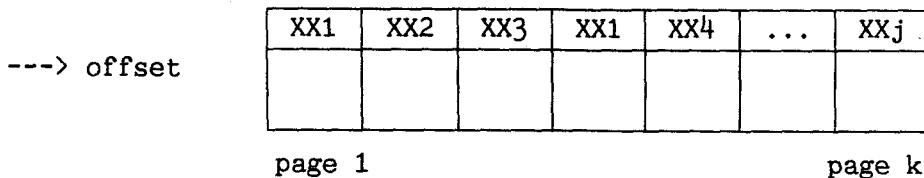
In a general way, the informations about an individual of the entity XXi are accessed by :

```
CALL GETXXi(num1,num2,iret)
CALL PUTXXi(num1,num2,iret)
```

All of the informations related to the individuals are stored in two files, the first one being a directory (or a table of adresses) organized as follow:

Entity type	number	location
XX1	-	page/offset
XX2	-	page/offset
XX3	-	page/offset

where 'number' is a combination of num1, num2 and a value related to the entity type. The page and offset informations refer to the second file that is divided into several pages, each one devoted to an entity:



In this case, the entity XX1 is represented by a large number of individuals, so that page 4 is also filled with individuals of this entity. A certain number of pages are kept in core memory and the choice of these pages is determined from the priorities defined for each entity and from the most frequently used ones.

The efficiency of these procedures has been evaluated in two ways during this work. Probably the most important aspect is the ease of development, induced by the use of this data structure. This is specially true for extensive data manipulation (topology simulation). Secondly, it has been shown that the system is very efficient when the number of entities is very important -greater than 20,000- [5].

### III. MESH GENERATION

#### 3.1 Generalities

Mesh generation is one of the major point in the discretization scheme. Very large structures (BiCMOS) must be simulated, with strong and localized impurity gradients (source/drain, emitter and base). In order to quantify this problem when simulating such structures, high precision should be obtained within a very small area (1000-5000 Å) in a large simulation domain (several microns). On the other hand, the generated meshes can greatly affect the confidence in the study of physical process models. In this part, we will present the standard procedure used for mesh generation in IMPACT<sup>4</sup>. Sufficient precision can be reached, with a limited interaction with the user.

It is well known that triangular elements are appropriate for describing complex (arbitrarily shaped) structures. Hence, this element type is chosen in a natural manner. In preliminary works, 6-noded triangles (quadratic shape function) were used for the diffusion problem [6], but this imposed severe requirements on the mesh. Especially, local mesh refinement are absolutely needed in order to avoid oscillations or numerical instabilities (fig. 3). This experience indicates that it can not be extended to a general multilayer problem. For this reason, 3-noded (or so-called P1) elements with linear shape function are used in this study.

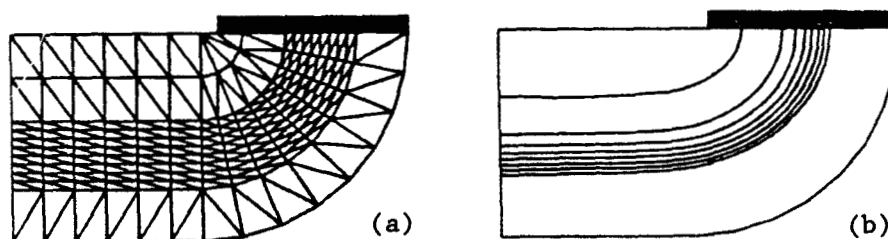


Fig. 3 Two-dimensional predeposition simulation. (a) Adapted mesh using P2 elements (quadratic shape function). (b) isoconcentrations [6].

### 3.2 Strategy

Taking into account the remarks of the first paragraph, we can see that regular meshes can not be used, in order to limit the number of nodes and element in the simulation domain. Here we propose a simple approach, that has been proved to be efficient in a large number of cases [7-8]. Referring to fig. 4, the strategy is as follow :

- 2 zones can be defined, each one is divided into lateral 'bands'. Inside each zone, the triangles are nearly similar.
- In each zone, the number of elements per band can be decreased from a band to the next one in the depth direction (by a factor included between 0 and 4). In Fig. 4, A factor of 1 is chosen in the first (upper) zone and a factor of 2 is chosen in the second (lower) zone.
- The first zone contains the highest triangle density and must cover the doping area where strong concentration gradients are believed to occur (source/drain, emitter). The second zone is for the deep quasi-uniform doping profiles (wells in CMOS).

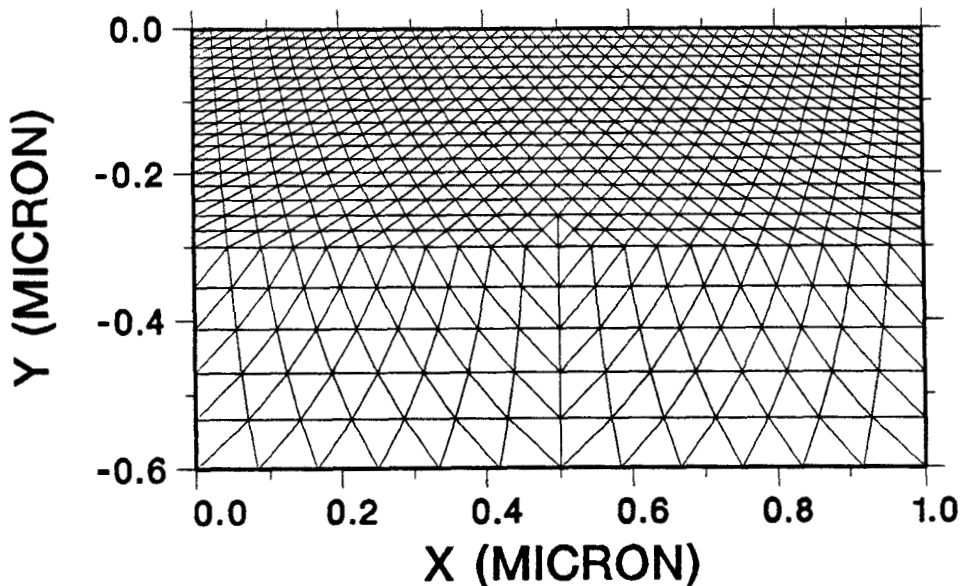


Fig. 4 Example of a typical mesh generated for the silicon domain.



This method is applied for the silicon substrate. The major limitations of this method are: i) an interaction with the user, ii) some knowledge of the final extension of the doping profile should be obtained (thus it can be defined as an *a priori* method), iii) in the case of very large domains, it is difficult to limit the total number of nodes. However, almost regular meshes are generated, allowing a good stability in the calculated doping profiles.

### 3.3 Mesh generation

The mesh generator used in this study originates from the MSHPTS software from the INRIA library [9]. The employed Delaunay triangulation is the one which provides as near a set of equilateral triangles as is possible with the given points. Details on this mesh generator will not be given here, but its integration in the program is briefly presented.

For this purpose, let us examine a complex structure to be meshed (Fig. 5 to 7). The boundary nodes are defined (they have been obtained from the first grid generation and the subsequent etching, deposition or oxidation steps). Here, for the sake of simplicity, an uniform spacing between these nodes has been chosen. The following procedure is then applied:

- i) A regular grid of points is surimposed on the domain (Fig. 5)
- ii) The mesh generator is called, determining the triangulation plotted in fig.6. As the nodes outside the boundaries are not included in the mesh, they can be removed from the list of nodes. During this operation, the internal nodes are not moved, giving rise to triangles with bad aspect ratio (this is obvious in Fig. 6 for some elements near the boundary).
- iii) A regularization step (also defined sometimes as point smoothing) is then used: each internal node is moved to the geometrical center of its topological neighbours (the boundary nodes are not concerned by this operation). This iterative method increases the triangles quality as demonstrated in Fig. 7.

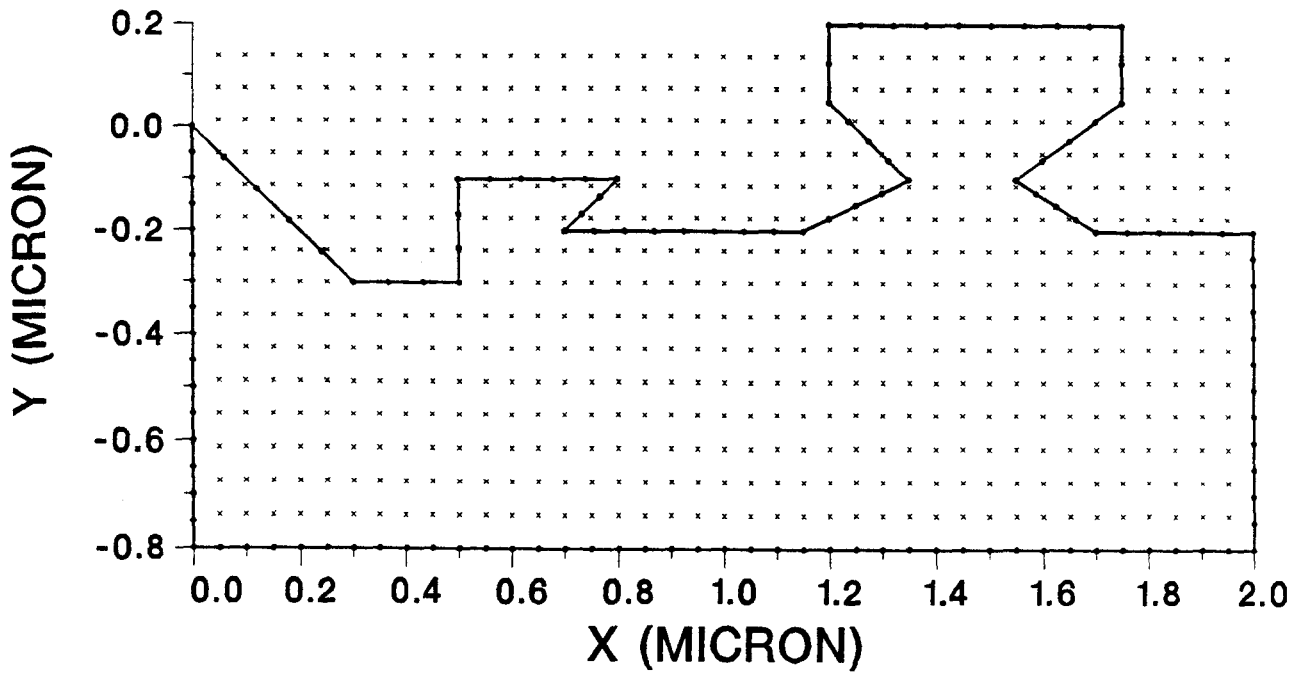


Fig. 5 Test example illustrating the mesh generation process. Boundary nodes and internal nodes (from a regular grid) are defined.

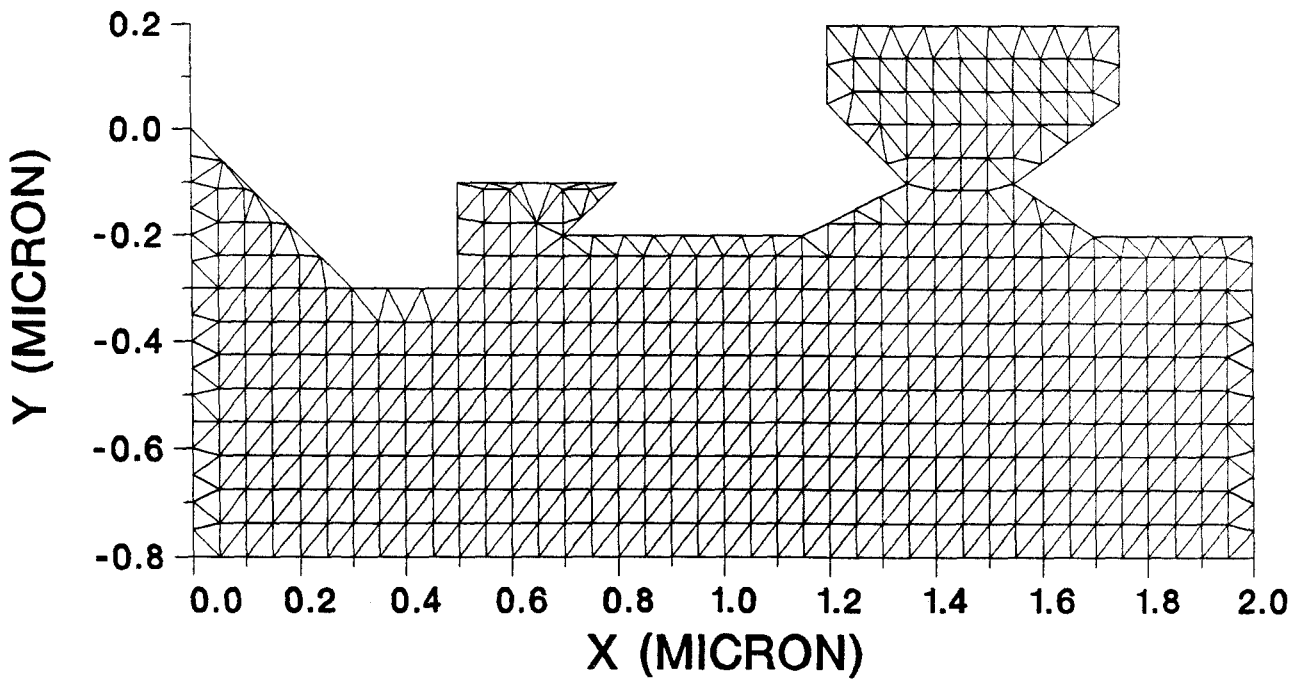


Fig 6 Resulting structure after mesh generation. The internal nodes outside or too close to the boundaries have been removed.

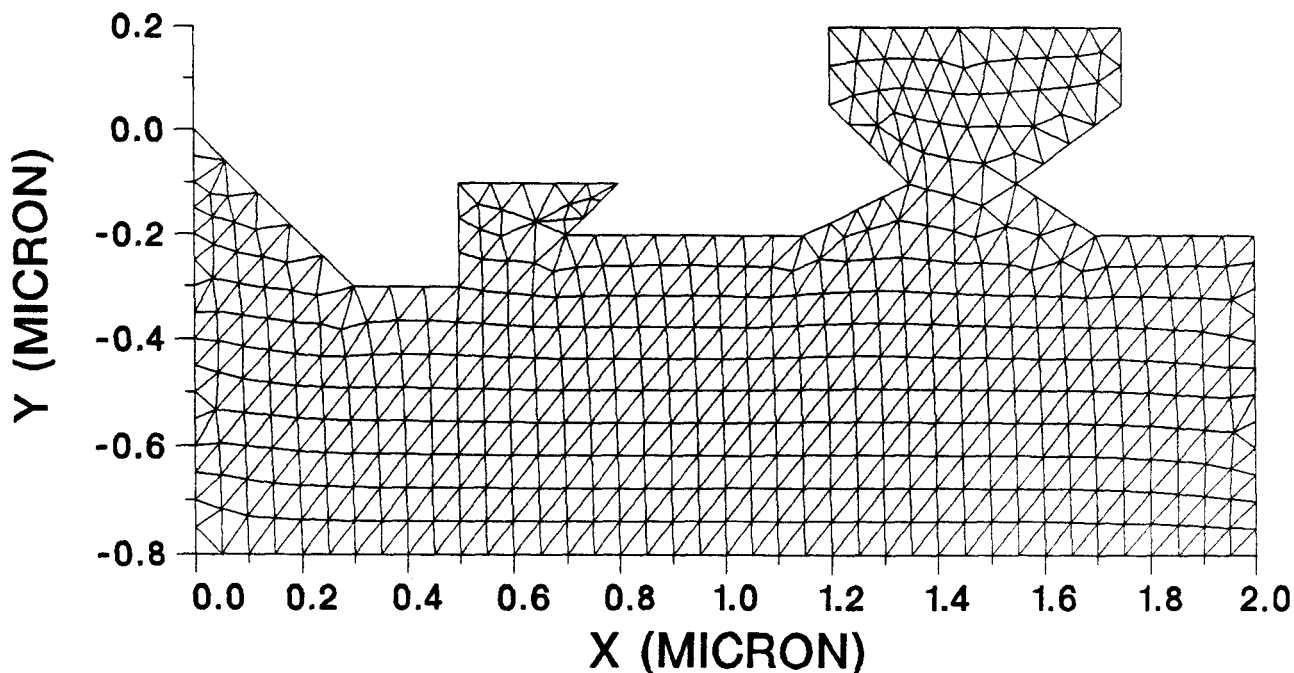


Fig. 7 Resulting structure after regularization.

The procedure is generalized for every layer, except that the user can define a progressive grid in the silicon domain for the internal nodes as shown in Fig. 4, whereas a regular grid is always used automatically for the other layers (the node spacing is calculated from the mean distance between boundary nodes). Finally, each interface between layers is determined by double nodes (with the same coordinates), one belonging to the first layer, the other to the second layer. This enables the treatment of impurity segregation. As a conclusion, Fig. 8 shows a realistic application at the end of a complete simulation of an advanced self-aligned bipolar transistor [10]. It features a trench isolation structure, filled with polysilicon, two other polysilicon layers for the extrinsic base and emitter formation, respectively. A progressive grid was defined at the beginning of the simulation for silicon, and the other layers have been meshed with regular grids.

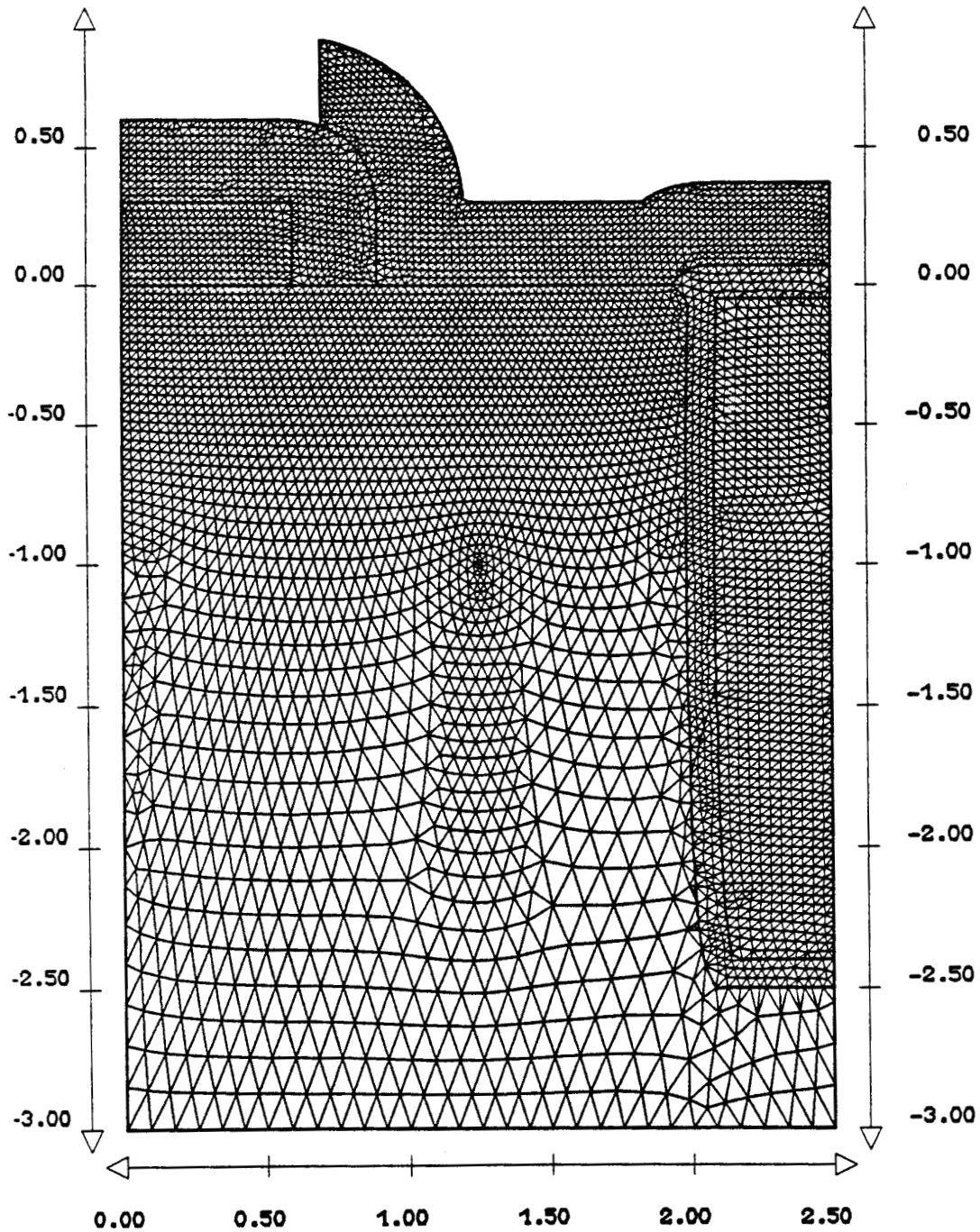


Fig. 8 Example of meshes generated at the end of a complete simulation in a self-aligned polysilicon bipolar transistor [10].

REFERENCES :

- [1] J. Lorenz, J. Pelka, H. Ryssel, A. Sachs, A. Seidl and M. Svoboda, "COMPOSITE - A complete modeling program of silicon technology", IEEE Trans. Electron Devices, ED-32, No 10, 1977-1986, October 1985.
- [2] K. Kato, N. Shigyo, T. Wada, S. Onga, M. Konaka and K. Taniguchi, "A supervised simulation system for process and device designs based on geometrical data interface", IEEE Trans. Electron Devices, ED-34, 2049-2058, 1987.
- [3] C.H. Corbex, A.F. Gerodolle, S.P. Martin and A.R. Poncet, "Data structuring for process and device simulations", IEEE Trans. on Computer Aided Design, vol 7, No 4, 489-500, April 1988.
- [4] D. Morel and D. Delannoy, "Visuel and Paging libraries", ISEN , 1987.
- [5] D. Morel, private communication.
- [6] D. Collard and J.N. Decarpigny, "Two-dimensional process simulation using a quadratic finite element discretization", COMPEL, vol. 3, No 1, 17-33, March 1984.
- [7] D. Collard, B. Baccus and E. Dubois, "IMPACT user's guide", ISEN, 1987.
- [8] B. Baccus, D. Collard, E. Dubois and D. Morel, "IMPACT<sup>4</sup> user's guide", ISEN, January 1990.
- [9] F. Hecht and A. Marrocco, internal report on MSHPTS, INRIA, Rocquencourt, February 1986.
- [10] B. Baccus, E. Dubois, D. Collard and D. Morel, "Efficient two dimensional multilayer process simulation of advanced bipolar devices", Solid-State Electronics, vol. 32, No 11, 1013-1023, 1989.

## CHAPTER III

### PROCESS STEPS

#### Abstract

Each process step is reviewed in this chapter: etching/deposition, ion implantation, diffusion, and oxidation. Within each part, the related models are first presented, followed by the algorithmic or numerical methods involved in the simulation of these steps.

Original points concerning the algorithmic aspects are detailed for the parts concerning the simulation of etching/deposition steps and oxidation. In particular, the relation between mesh generation and layer management in the data structure is discussed. Original procedures, solving the dopant redistribution at interfaces, are detailed, with a special emphasis on boundary fluxes.

## I. ETCHING/DEPOSITION STEPS

In parallel with the increasing need to use 2D process simulators for diffusion, oxidation and ion implantation, extensive works have been performed to understand and to analyse the line-edge profiles resulting from a sequence of the following process steps: photolithography, etching and deposition. Specialized simulators were developed, among them the well-known SAMPLE program was first dedicated to photolithography [1], and then updated to a more general simulator including deposition and etching steps [2-3]. This program has been also used as part of more general simulators such as COMPOSITE [4] or SPIRIT-I [5]. However, the current trend is to replace SAMPLE by home-made topography routines in order to generalize the geometrical capabilities of etching and deposition steps simulation. This is also the case for the present study.

The models will be first described, followed by the geometrical and algorithmic problems. Moreover, the link between the string model algorithm and the mesh generation is discussed in order to minimize shapes interpolations, allowing the simulation of complete processes, with sufficient accuracy.

### 1.1 Models

Although various models have been proposed in the past, there is no general formulations for the modeling of etching and deposition steps. In particular, the reactors or systems used during experimental procedures may have their own characteristics, restricting the generality of such models. For this reason, basic components (isotropic, anisotropic and so forth) are defined and mixed for each process step. In this part, etching models will be first reviewed, followed by deposition models.

### 1.1.1 Etching models

Most of the etching processes can be classified experimentally as isotropic or anisotropic. In the case of wet etching (as example oxide etching using HF solution), the process is purely isotropic: each surface point experiences the same etching rate, in the direction of the perpendicular bisector of the adjacent points (Fig. 1a). If the duration of this step is not accurately controlled, undercut may occur. In the case of dry etching, a combination of isotropic and anisotropic components can describe the step. Fig. 1b indicates the surface point motion, parallel to the incident beam. In some cases, these components can be deduced from experiments as performed in [6].

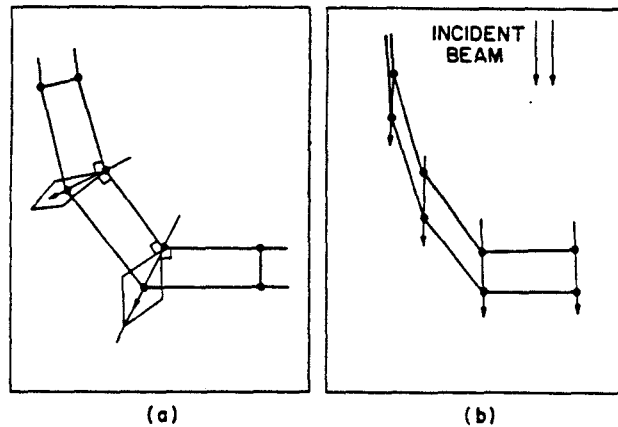


Fig. 1 Isotropic (a) and anisotropic (b) etching processes.

Referring to Fig. 2a, these two components are simply implemented in the program, the lateral ( $V_x$ ) and vertical ( $V_y$ ) rates of displacement are given below:

- 1) Anisotropic component with rate  $R_1$  :

$$\begin{aligned} V_{x1} &= 0. \\ V_{y1} &= R_1 \end{aligned} \tag{3.1}$$

- 2) Isotropic component with rate  $R_2$  :

$$\begin{aligned} V_{x2} &= R_2 \cdot \sin(\Phi) \\ V_{y2} &= -R_2 \cdot \cos(\Phi) \end{aligned} \tag{3.2}$$



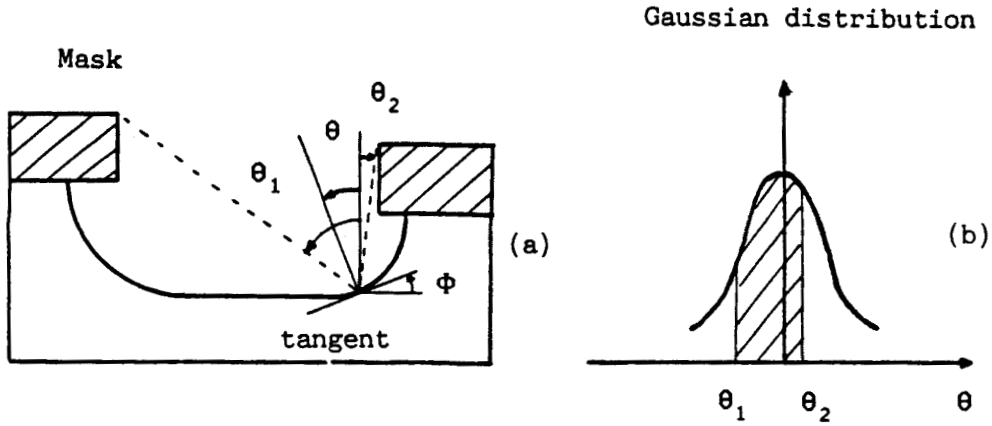


Fig. 2 (a) Description of the angles used in the different etching components  
 (b) Illustration of the Gaussian distribution for equation (3.3)

A combination of these two basic components allows to adjust the degree of anisotropy. In the case of RIE (Reactive Ion Etching), this method is not sufficient. Following the approach of Yamamoto [7], silicon grooves can be accurately modeled, when adding two anisotropic components :

3) Anisotropic component for ions with rate  $R_3$ :

$$V_{x3} = R_3 \cdot \int_{\theta_1}^{\theta_2} \sin(\theta) \cdot P(\theta) \cdot d\theta$$

$$\text{with } P(\theta) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{\theta^2}{2\sigma^2}\right)$$

$$V_{y3} = R_3 \cdot \int_{\theta_1}^{\theta_2} \cos(\theta) \cdot P(\theta) \cdot d\theta \quad (3.3)$$

4) Anisotropic component for radicals with rate  $R_4$  :

$$V_{x4} = R_4 \cdot \frac{\cos(\theta_1) - \cos(\theta_2)}{2}$$

$$V_{y4} = R_4 \cdot \frac{\sin(\theta_2) - \sin(\theta_1)}{2}$$

(3.4)

As can be seen, these two components are related to a probability distribution  $P(\theta)$  for the concerned etching species.  $P(\theta)$  is uniform for radicals and gaussian for ions (Fig. 2b), with a dependence on the possible incident angle, between  $\theta_1$  and  $\theta_2$ . By combining these two anisotropic components and the isotropic one, very interesting results can be obtained, since the etched profiles depend on the heights and shapes of the mask (see for example Fig. 3).

Calculating condition	Symmetric mask	Asymmetric mask
Mask thickness thick 1.0 $\mu\text{m}$ thin 0.5 $\mu\text{m}$ Pattern width 0.8 $\mu\text{m}$		
Isotropic		
Ion	$\sigma = 1^\circ$	
	$\sigma = 30^\circ$	
Radical		

Fig. 3 Simulated profiles with various etching components [7].

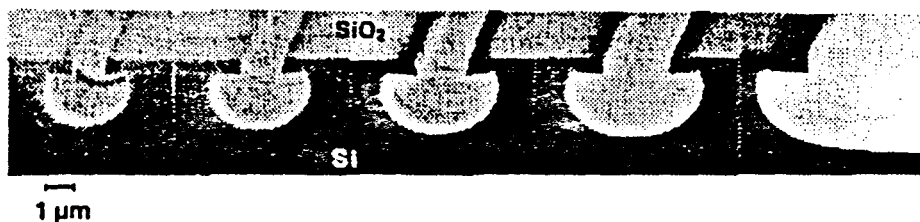
The correct implementation of this model is checked by the following example: formation of silicon grooves using several  $\text{SiO}_2$  mask widths [7]. The etching gas is  $\text{SF}_6$  with a pressure and rf power of 0.04 Pa and 160 W, respectively.

The extracted etching rates for silicon are [7]:

- Isotropic :  $R_2 = 0.03 \mu\text{m}/\text{min}$
- Anisotropic ions :  $R_3 = 0.01 \mu\text{m}/\text{min}$
- Anisotropic radicals :  $R_4 = 0.1 \mu\text{m}/\text{min}$

The standard deviation for the Gaussian distribution for ions is 10 degrees. Fig. 4a shows the experimental SEM result and Fig. 4b the simulated profile with IMPACT<sup>4</sup>. These two results agree well each other, in particular the groove depth dependence on the mask width is well simulated as reported in Fig. 5 where the groove depth has been normalized using the  $2 \mu\text{m}$  mask pattern size.

(a)



(b)

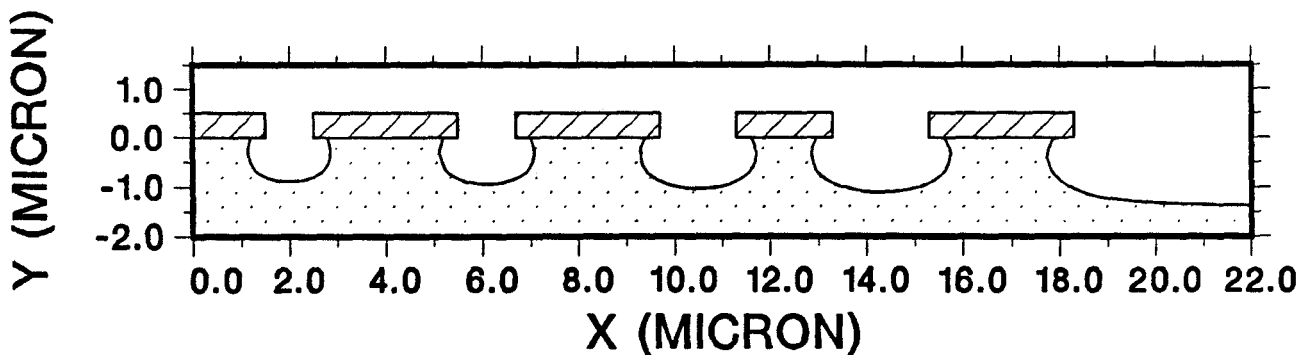


Fig. 4 Silicon groove formation.

- (a) SEM photograph of experimental results.
- (b) Calculated result with IMPACT<sup>4</sup>.

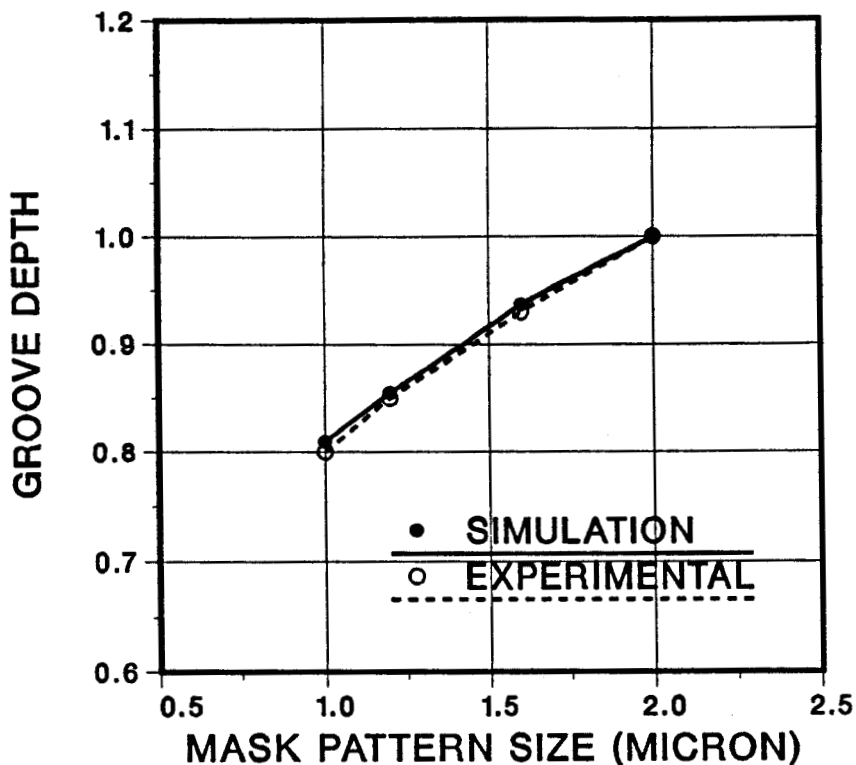


Fig. 5 Relationship between the mask pattern size and the groove depth (normalized using the depth from the 2 $\mu$ m mask size).

Although these possibilities are very useful for the simulation of trench formation, there is no relation between the pressure, power or gas type, and the individual components R2, R3 and R4. A progress will be made in the future by determining each species included in the plasma [8] or by performing Monte Carlo simulations, as was calculated for ion assisted plasma etching [9].

#### 1.1.2 Deposition models

A good way to introduce the deposition models is to analyse the step coverage of deposited films [10]. Three general types can be distinguished and are illustrated in Fig. 6. Fig. 6a shows a perfect conformal step coverage. In this case, the reactants adsorb on surface and then rapidly

migrate along the surface before reacting. An example is given by the use of TEOS deposition,  $\text{Si}(\text{OC}_2\text{H}_5)_4$ , at low pressure with an usual temperature of  $700^\circ\text{C}$ . This can be easily modeled by an isotropic component. The second case is illustrated in Fig. 6b, where the mean-free path of the gas is much larger than the dimension of the step, but with a reduced migration rate. In this case, the resulting profile depends on the possible arrival angle of molecules. A typical example is the deposition from silane and oxygen at reduced pressure. Finally, the third type is shown in Fig. 6c where no surface migration and a short mean-free path occur. This can be again the result of deposition of silane and oxygen, but performed at atmospheric pressure.

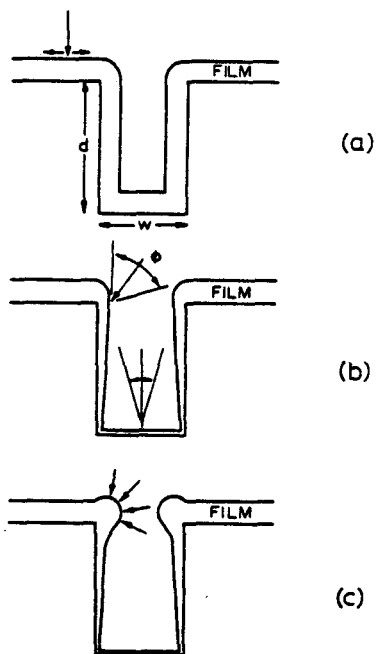


Fig. 6 Shapes of deposited films, illustrating three types of step coverage efficiency [10].

The first and second types have been included in the program. The first type is simply modeled by an isotropic component in a similar way than equation (3.2). The second type is often related to a hemispherical source and is simulated as follow [11]:

$$\begin{aligned} \Delta X &= \text{Rate} \cdot \frac{\cos(\theta_1) - \cos(\theta_2)}{2} \\ \Delta Y &= \text{Rate} \cdot \frac{\sin(\theta_2) - \sin(\theta_1)}{2} \end{aligned} \tag{3.5}$$

where  $\theta_1$  and  $\theta_2$  are the lower and upper bounds of the incident angles of the vapor stream (in a similar way than in the anisotropic components for RIE). Referring to equation (3.4), we can also deduce that an uniform probability distribution is implicitly assumed. This modelisation allows to reproduce shadowing effects and may cause the formation of 'cracks' in the vicinity of concave corners. This is the case for oxide deposition or Aluminium sputtering. Fig. 7 displays an example of oxide deposition using this model, resulting in a poor step coverage: lower film thickness against vertical wall and in the concave corner. For a planetary system, more sophisticated formulations have been proposed, depending on the geometrical configurations of the evaporator [12].

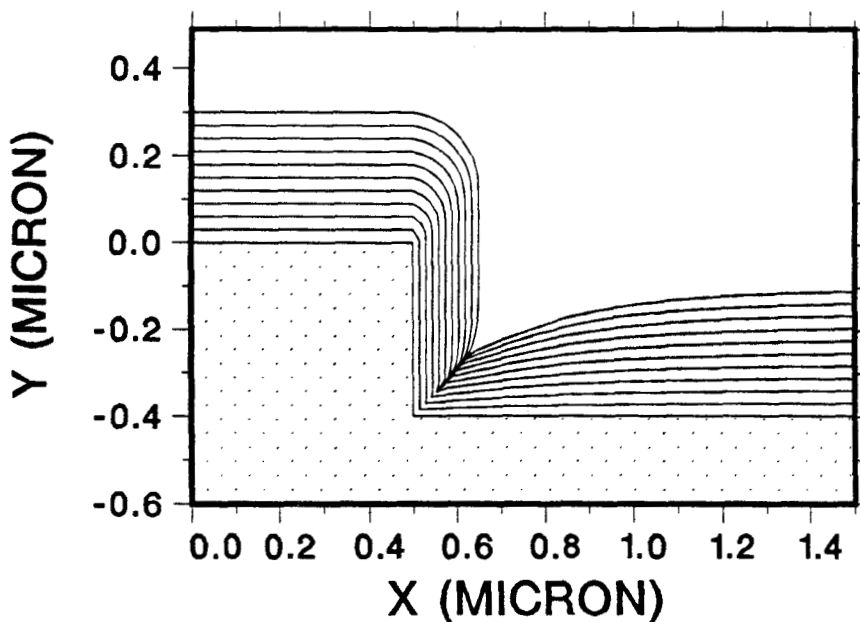


Fig. 7 Thin film deposition using a hemispherical source model (IMPACT4 simulation)

## 1.2 Algorithms

### 1.2.1 The string model algorithm and the related geometrical problems

The surface evolution is calculated in a natural manner, by using the so-called string model algorithm [13], in which this surface is approximated by a series of points joined by straight line segments. Although other methods have been proposed, such as the cell or ray models, this algorithm is always used in the simulators. This very simple discretization requires a large amount of programming in order to solve all the topological problems that arise. Some of them are listed below but will not be detailed since they have been already described in the literature [7-13].

The first point is to minimize the errors in shape definition, induced by the discretization of the surfaces. The common way is to consider an uniform distance between points, which implies an uniform distribution of the position errors. Once a calculation of the new surface has been performed, points are added or deleted in order to fill this criterion. In a parallel manner, the time discretization is determined so that the maximum displacement is less than half the mean distance between points. This allows a good precision in the calculated profiles as was demonstrated in Fig. 4 and 5.

An other classical problem arises from loop formation in concave corners, as illustrated in Fig. 8. In this case the loop is deleted and the intersection point is added in the list of nodes.

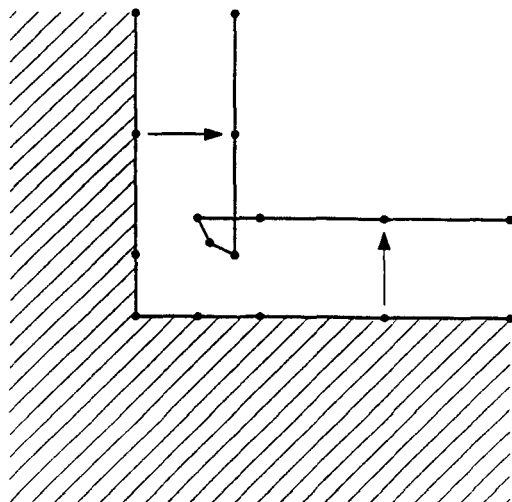


Fig. 8 Example of loop formation in concave corner during isotropic deposition

Beyond these usual problems, more critical situations are encountered. For the sake of simplicity, two typical cases are described in the following. The first one is depicted in Fig. 9. Assuming a perfect mask (nitride for example), an isotropic etching is performed. This can be the result of a small oxide etching after LOCOS oxidation. Here, the point at the boundary between nitride, oxide and air must be carefully treated, since the isotropic component forces this point to follow the interface with nitride. In the first versions of SAMPLE, the problem was not important because such interfaces were planar. In our case, special treatments have been developed.

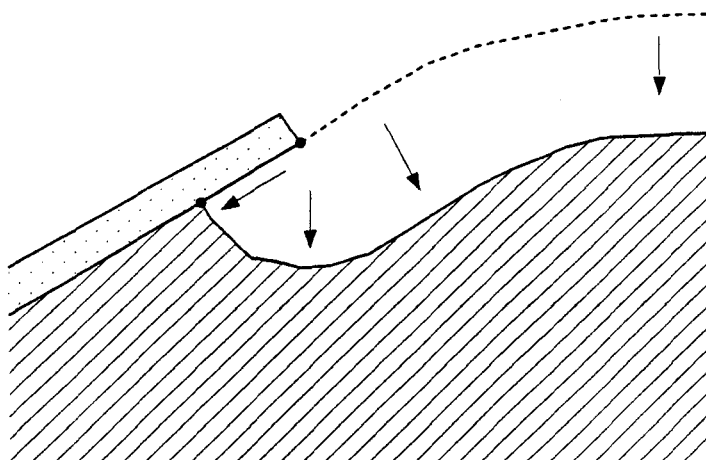


Fig. 9 Boundary condition arising during isotropic etching with a perfect mask.

The other example is reported in Fig. 10, where an oxide layer is anisotropically etched for the formation of a sidewall spacer (MOS technology). Again a problem arises at the boundary between materials. In particular, several points must be introduced i) at the sharp polysilicon corner, ii) at the polysilicon/oxide interface at the location specified by the arrow. In a general manner, the point location must be corrected when a boundary has been crossed, for materials with different etching rates. Although all of these algorithms increase dramatically the CPU cost, they are absolutely necessary in order to maintain a correct shape definition.



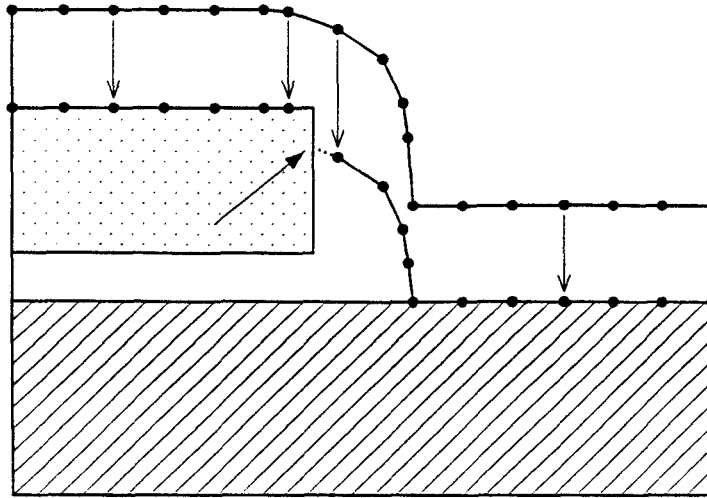


Fig. 10 Other boundary problem, appearing during anisotropic etching, when string points cross underlying domains.

### 1.2.2 Relations with mesh generation and data structure

The original aspect here is to couple the discretization used during topology calculation with the discretization needed for mesh generation (FEM method). Implicitly, the points used in the string model algorithm can describe with a sufficient accuracy the outer boundaries of the layers. The solution is to use the string points as boundary nodes for mesh generation. By numerous test structures, it has been found that the distance between points required in each discretization is compatible. This is a very important characteristic since the interpolations of shapes and interfaces between parts of the program are avoided during a complete simulation. Fig. 11 shows the etching of an oxide layer under resist masks. Selectivity with resist is supposed to be perfect, but not with the underlying silicon substrate. In order to illustrate the above discussion, the mesh generated within each layer is also displayed (for simplicity, the number of nodes has been severely limited). 'Double' nodes are used at the interfaces between layers, for the treatment of impurity segregation (section 3.1.1).

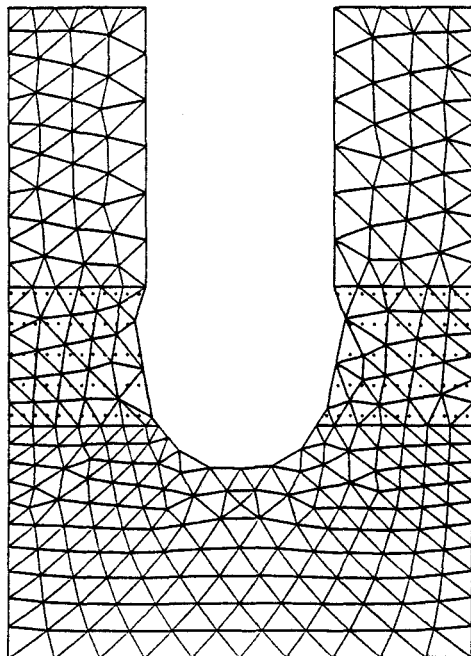
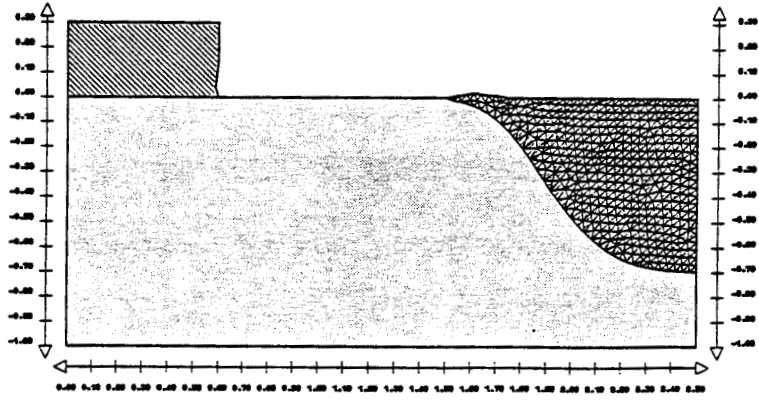
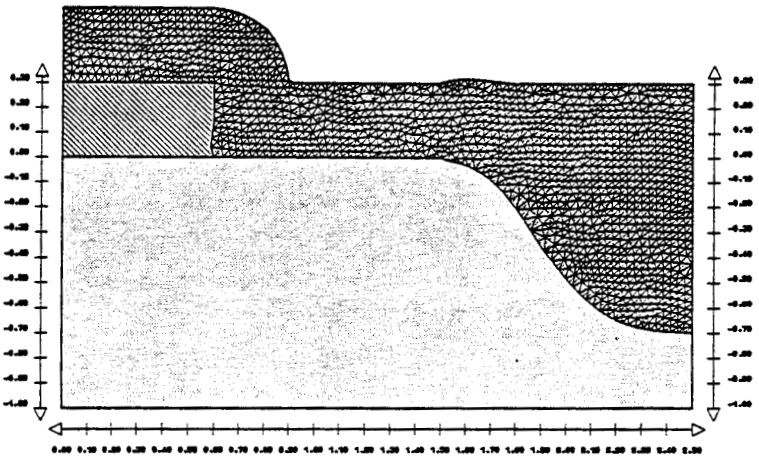


Fig. 11 Simple example of mesh generation after an etching process.

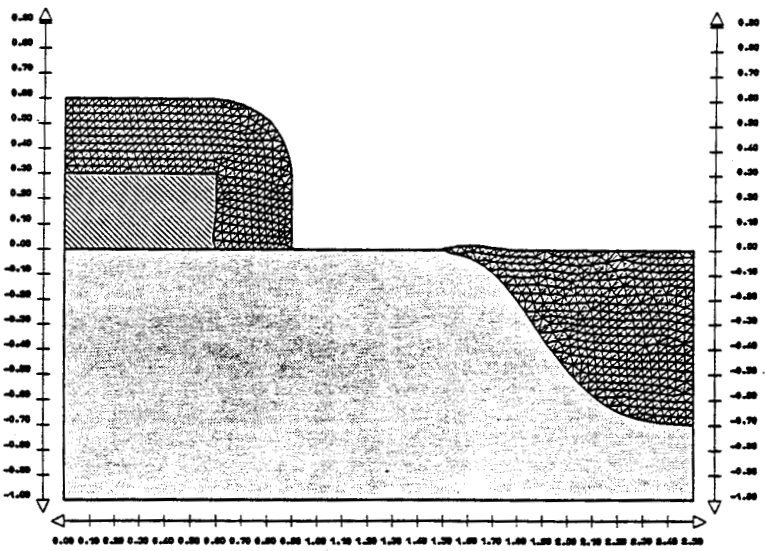
An other original aspect concerns some choices related to data structure. For this purpose, an example provided from a complete simulation of a self-aligned polysilicon emitter bipolar transistor is given in Fig. 12 [14]. Fig. 12a shows the fully recessed oxide (FOX) isolation and the first polysilicon layer for the extrinsic base contact. Oxide deposition is then performed for the isolation between the two polysilicon layers (the second poly-layer will act as a dopant source for emitter) and for the sidewall spacer definition (Fig. 12b). It is interesting to notice that the two oxide layers (FOX and deposited) are automatically merged and are recognized as a single domain stored in the data base. In the same manner, two separate layers can be merged to form a single domain, one of the domains being deleted in the data base. Fig. 12c shows the resulting profile after anisotropic etching of silicon oxide to open the polysilicon emitter contact. In this case the initial oxide layer has been split into two different domains (hence the data base contains four domains or layers). Mesh generation is always automatically performed in each of the etched layers with dopant interpolations if needed.



(a)



(b)



(c)

Fig. 12 Example of interactions between mesh generation, domains management in the data base and etching-deposition steps.

## II. ION IMPLANTATION

Ion implantation is now the mostly used doping technique for VLSI processing. Among others, the following interesting characteristics are listed below [15]:

- reproducible and homogeneous doping concentrations by a tight control of the implanted species and the dose.
- Efficient use of masking layers (e.g.  $\text{SiO}_2, \text{Si}_3\text{N}_4$ ) and realization of self-aligned structures.
- low temperature process.
- Possibility of focused ion beam.

So, a good modeling of ion implantation is needed in order to obtain accurate initial doping profiles before diffusion steps. Two approaches are commonly used:

- i) Numerical techniques such as Monte Carlo calculations or solution of the Boltzmann transport equations allow physical insight and accurate results, at the expense of large CPU times.
- ii) Analytical formulations are usually implemented in process simulators. In this case, the resulting concentrations are the product of the dose and of a vertical distribution.

In order to describe the modeling approach implemented in IMPACT<sup>4</sup>, the sub-chapter organization is as follows: in a first part, the vertical distributions are described, in the second part the 1D multilayer problem is examined, the choice of the dose matching algorithm being justified. Finally, the two dimensional problem is analysed and some interesting features are presented: tilted implants and the modeling of the depth dependent lateral standard deviation.

## 2.1 Vertical distributions

### 2.1.1 Monte Carlo versus analytical calculations

In order to obtain the depth distribution of implanted profiles, the most accurate results are issued from Monte Carlo calculations (as example the TRIM program [16]), or from the solution of Boltzmann transport equations [17]. Although this latter method offers a computation speed advantage over the Monte Carlo (MC) method, the rigorous approach of MC method and its intrinsic 3D characteristic make it more attractive. Three major applications of MC method are currently reported:

- i) Algorithms have been developed, aimed to minimize CPU consumption, such as the works of Hobler *et al* [18]: use of auxiliary grid and 'superposition' method have allowed the simulation of realistic 2D structures and an ideal 3D trench. An other 3D approach has been also reported [19].
- ii) The determination of spatial moments of the distribution functions to be used in the analytical models [20-21]. Since 2D measurements of doping profiles are not available or with a limited confidence in their validity, MC calculations allows to verify the results of analytical methods.
- iii) From a physical point of view, the damage induced by the ion implantation can be estimated [22]. Channeling is also accounted for.

MC method is a powerful tool in order to understand the underlying physical phenomena but it requires too large CPU time for a systematic use needed in process optimization. For this reason, distribution functions (analytical calculations) are widely used, whose parameters may be obtained from MC calculations, experimental measurements, or theories. The LSS theory has been first proposed by Lindhard, Scharff and Schiott [23]. From this formulation, tabulated values have been determined such as reported in the well-known book of Gibbons [24]. These parameters used in distribution functions are defined mathematically from a probability density function  $f(x)$  [25]:

$$\text{Mean value or projected range} : R_p = \int_{-\infty}^{+\infty} x \cdot f(x) \cdot dx \quad (3.6)$$

$$\text{Standard deviation} : \sigma_p = \sqrt{\int_{-\infty}^{+\infty} (x-R_p)^2 \cdot f(x) \cdot dx} \quad (3.7)$$

$$\text{Skewness} : \gamma = \frac{\int_{-\infty}^{+\infty} (x-R_p)^3 \cdot f(x) \cdot dx}{\sigma_p^3} \quad (3.8)$$

$$\text{Kurtosis} : \beta = \frac{\int_{-\infty}^{+\infty} (x-R_p)^4 \cdot f(x) \cdot dx}{\sigma_p^4} \quad (3.9)$$

Three distributions are currently used: Gaussian, joined-half Gaussian and Pearson IV. We now describe these distributions and the way they are used in IMPACT4.

### 2.1.2 Gaussian distribution

This distribution uses only  $R_p$  and  $\sigma_p$  ( $\gamma=0$  and  $\beta=3$ ) and will be used for all the implanted species in silicon oxide, silicon nitride, and resist:

$$f(x) = \frac{N_{\square}}{\sqrt{2\pi} \cdot \sigma_p} \cdot \exp\left(-\frac{(x-R_p)^2}{2 \cdot \sigma_p^2}\right) \quad (3.10)$$

where  $N_{\square}$  is the implantation dose (at/cm<sup>2</sup>).

### 2.1.3 Joined-half-Gaussian distribution

In practice, asymmetrical distributions are often used for arsenic, phosphorus and antimony implantation in silicon (and polysilicon in our case). The joined-half-Gaussian distribution is then defined [26]:

$$f(x) = \begin{cases} x < R_m & \frac{2 \cdot N_{\square}}{\sqrt{2\pi}(\sigma_1 + \sigma_2)} \cdot \exp\left(-\frac{(x-R_m)^2}{2 \cdot \sigma_1^2}\right) \\ x > R_m & \frac{2 \cdot N_{\square}}{\sqrt{2\pi}(\sigma_1 + \sigma_2)} \cdot \exp\left(-\frac{(x-R_m)^2}{2 \cdot \sigma_2^2}\right) \end{cases} \quad (3.11)$$

The modal projected range  $R_m$  and the 'one-sided' standard deviations  $\sigma_1$  and  $\sigma_2$  are deduced from the relations (3.13) to (3.15), providing that condition (3.12) is satisfied:

$$|\gamma| < \frac{4-\pi}{\pi-2} \cdot \sqrt{\frac{2}{\pi-2}} \cong 0.99527 \quad (3.12)$$

$$R_p = R_m + \sqrt{\frac{2}{\pi}} \cdot (\sigma_2 - \sigma_1) \quad (3.13)$$

$$\sigma_p = \sqrt{(\sigma_1^2 - \sigma_1 \cdot \sigma_2 + \sigma_2^2) - \frac{2}{\pi} \cdot (\sigma_2 - \sigma_1)^2} \quad (3.14)$$

$$\gamma = \frac{\sqrt{\frac{2}{\pi}} \cdot (\sigma_2 - \sigma_1) \cdot \left( \left( \frac{4}{\pi} - 1 \right) \cdot (\sigma_1^2 + \sigma_2^2) + \left( 3 - \frac{8}{\pi} \right) \cdot \sigma_1 \cdot \sigma_2 \right)}{\sigma_p^3} \quad (3.15)$$

As  $R_p$ ,  $\sigma_p$  and  $\gamma$  are usually extracted from tabulations, a cubic equation is established and solved analytically, from equations (3.13) to (3.15). This allows the calculations of  $\sigma_1$ ,  $\sigma_2$  and  $R_m$ .

#### 2.1.4 Pearson IV distribution

However, in order to reproduce experimental profiles exhibiting the channeling effect (due to the fact that the silicon substrate is not amorphous), the kurtosis  $\beta$  must be used. In this case, Pearson IV distribution is generally chosen and can, to some extent, modelize this effect. Pearson IV was first introduced by Hofker [27]. In fact, the Pearson distributions (I to VII) are based on a differential equation:

$$\frac{df}{dy}(y) = \frac{y-a}{b_0+b_1 \cdot y+b_2 \cdot y^2} \cdot f(y) \quad , \text{ with } y=x-R_p \quad (3.16)$$

where a, b<sub>0</sub>, b<sub>1</sub> and b<sub>2</sub> are calculated from equations (3.17) to (3.20):

$$a = - \frac{\gamma \cdot \sigma_p \cdot (\beta+3)}{10 \cdot \beta - 12 \cdot \gamma^2 - 18} \quad (3.17)$$

$$b_0 = - \frac{\sigma_p^2 \cdot (4\beta - 3\gamma^2)}{10 \cdot \beta - 12 \cdot \gamma^2 - 18} \quad (3.18)$$

$$b_1 = - \frac{\gamma \cdot \sigma_p \cdot (\beta+3)}{10 \cdot \beta - 12 \cdot \gamma^2 - 18} \quad (3.19)$$

$$b_2 = - \frac{2\beta - 3\gamma^2 - 6}{10 \cdot \beta - 12 \cdot \gamma^2 - 18} \quad (3.20)$$

Depending on the value of  $\gamma$  and  $\beta$ , the Pearson type I to VII can be obtained. However, due to the particularities of implanted profiles, only the type IV is currently used. In this case the following relations hold:

$$\left\{ \begin{array}{l} 0 < \gamma^2 < 32 \\ \beta > \frac{39 \cdot \gamma^2 + 48 + 6 \cdot (\gamma^2 + 4)^{3/2}}{32 - \gamma^2} = \beta_{\min} \end{array} \right. \quad (3.21)$$

Using this condition and solving equation (3.16) leads to the expression of Pearson type IV distribution:

$$f(x) = K \cdot \left( - \left( b_0 + b_1 \cdot (x-R_p) + b_2 \cdot (x-R_p)^2 \right) \right)^{\left( \frac{1}{2 \cdot b_2} \right)} \cdot \exp \left( - \frac{b_1/b_2 + 2 \cdot a}{\sqrt{4 \cdot b_2 \cdot b_0 - b_1^2}} \cdot \operatorname{atan} \left( \frac{2 \cdot b_2 \cdot (x-R_p) + b_1}{\sqrt{4 \cdot b_2 \cdot b_0 - b_1^2}} \right) \right) \quad (3.22)$$



### 2.1.5 Defaulted parameters

For the materials: silicon oxide, silicon nitride and resist, Gaussian distributions are used for all impurities, with LSS parameters.

For silicon and polysilicon, identical distributions and parameters are used (we can consider that this approach is valid if the polysilicon grain size is greater than 1000Å and smaller than the polysilicon layer thickness [28]). The vertical distribution is defaulted according to the implanted element and each parameter can be calculated using LSS theory, expressions from Ryssel [29] (RY) or data from Simard-Normandin and Slaby [30] (SN). It should be noticed that there is no general expression for  $\beta$  and only empirical formulations are available.

*Arsenic, Phosphorus and Antimony :*

defaulted model : Joined-half Gaussian.

MODEL	$R_p$	$\sigma_p$	$\gamma$	$\beta$
Gaussian	LSS	LSS	0	3
Joined-half-Gaussian	LSS	LSS	LSS	-
Pearson IV	LSS	LSS	LSS	RY

$$\beta = \beta_{\min}$$

*Boron :*

defaulted model: Pearson IV

MODEL	$R_p$	$\sigma_p$	$\gamma$	$\beta$
Gaussian	LSS	LSS	0	3
Joined-half-Gaussian	LSS	LSS	LSS	-
Pearson IV, Energy $\leq$ 30 KeV	SN	SN	SN	SN
Pearson IV, Energy $\geq$ 30 KeV	LSS	LSS	RY <sup>1</sup>	RY <sup>2</sup>

Rem. 1 :  $\gamma = - \frac{\text{Energy}}{300}$

2 :  $\beta = 8 + 25(\beta_{\min} - 3)$

$BF_2$  :

defaulted model : Pearson IV

Energy  $\leq$  120 KeV : the parameters are calculated using SN data. The lateral standard deviation is the same as for boron with an effective energy of Energy/5.

Energy  $\geq$  120 KeV : The models are the same as for boron with an effective energy of Energy/5.

As an example, Fig. 13 shows a comparison between the experimental profile of a low energy boron implant and several analytical distribution profiles [30].

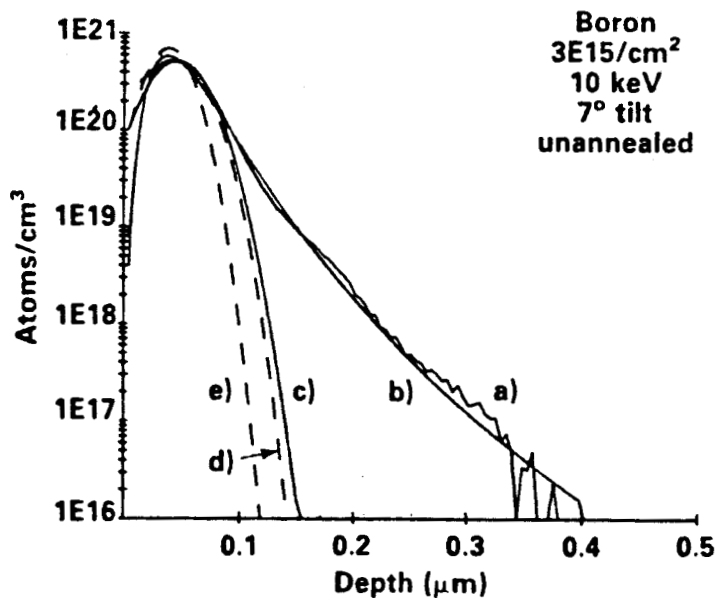


Fig. 13 a) experimental boron profile.  
b) best-fit Pearson IV (Simard-Normandin and Slaby, implemented in IMPACT<sup>4</sup>).  
c) best-fit Gaussian.  
d) Pearson IV profile in SUPREM III.  
e) LSS Gaussian profile.

## 2.2 The multilayer problem

The modeling of the multilayer problem (i.e. when implant is performed through several thin layers) is of major interest. This is the general characteristic of the program IMPACT<sup>4</sup> but also of actual technologies. As silicon oxide and nitride layers are often used as masks, their masking capability should be carefully checked (with respect to the ion implantation conditions: film thickness, energy). Moreover, thin surface layers may also reduce the channeling effect or allow to adjust the depth distribution of implanted ions. In the following the numerical techniques developed to model this problem are reviewed.

Let us assume an implantation into a mask/bulk structure, where the mask and bulk are referred as material 1 and 2 respectively, and  $t$  is the mask thickness. Up to now, five different formulations have been proposed, discussed with qualitative arguments [31] or in a quantitative manner by comparisons with MC simulations [32]. From these studies, we deduce that only three formulations can be used with a reasonable accuracy:

$$\text{Model 1 : } \begin{cases} C_1(x) & x < t \\ \alpha \cdot C_2 \left( x-t, \left( 1-\frac{R_{p2}}{R_{p1}} \right) \right) & x > t \end{cases} \quad (3.23)$$

with  $\alpha$  is such that the dose ( $N_D$ ) conservation is achieved :

$$\int_{-\infty}^{+\infty} C(x) \cdot dx = N_D$$

$$\text{Model 2 : } \begin{cases} C_1(x) & x < t \\ C_2(x-(t-t')) & x > t \end{cases} \quad (3.24)$$

with a value of  $t'$  to be further detailed.

$$\text{Model 3 : } \begin{cases} \frac{R_{p2}}{R_{p1}} \cdot C_2 \left( \frac{R_{p2}}{R_{p1}} \cdot x \right) & x < t \\ C_2 \left( x-t, \left( 1-\frac{R_{p2}}{R_{p1}} \right) \right) & x > t \end{cases} \quad (3.25)$$

where  $C_i$  and  $R_{p_i}$  refer to the vertical distribution and projected range in material  $i$ , respectively. In silicon technology, the material densities are almost the same. In this case, these 3 models give satisfactory results and often lead to similar results (it should be noted that all of these models fail in case of GaAs technology). Here, model 2 is chosen since i) it is formulated in a natural way and dose conservation is intrinsically achieved, ii) it has been compared favorably with experiments in a 3-layer structure [28]. Using Fig. 14, this algorithm (sometimes referenced as "dose matching model") is now described. The profile  $C_1(x)$  in material 1 is integrated to obtain  $N_{\square 1}$ :

$$N_{\square 1} = \int_0^t C_1(x) \cdot dx \quad (3.26)$$

Then the profile  $C_2(x)$  is integrated in material 2, until a thickness  $t'$  such that:

$$N'_{\square 2} = \int_0^{t'} C_2(x) \cdot dx = N_{\square 1} \quad (3.27)$$

The value of  $t'$  is used to shift the profile in material 2 (Fig. 14c).

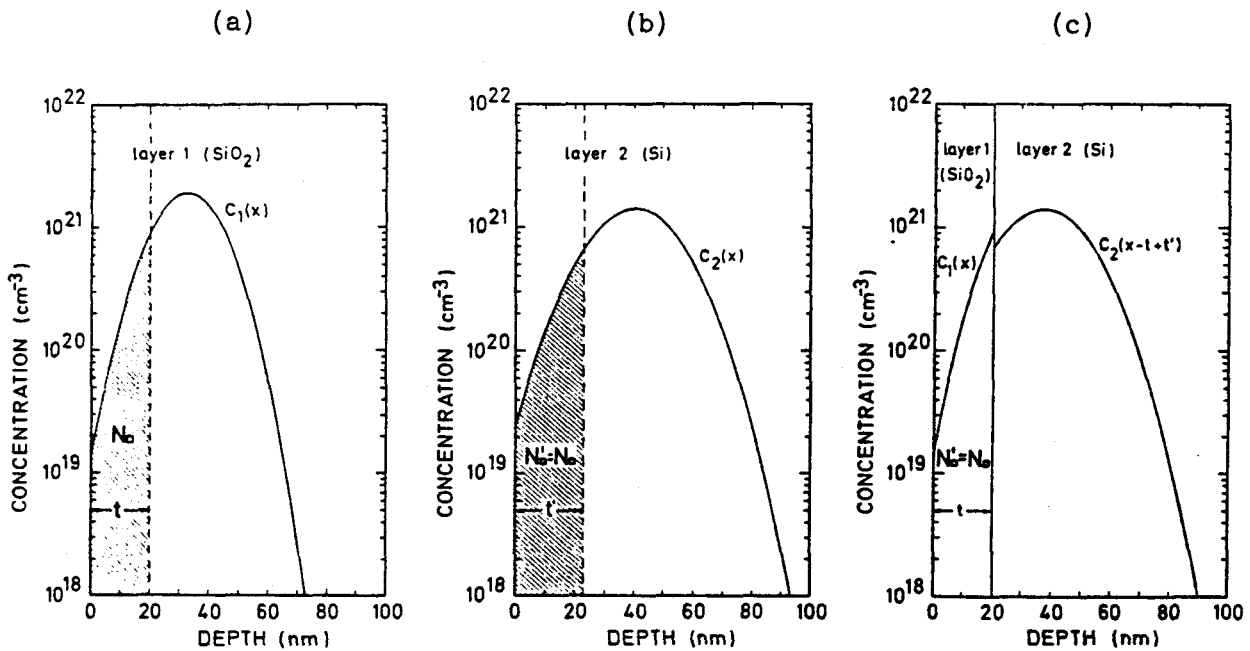


Fig. 14 Schematic representation of the dose matching model [31]

This model can be extended to the general case: for every layer  $j$ , a thickness  $t'_{j-1}$  is calculated:

$$N'_{\square j} = \int_0^{t'_{j-1}} C_j(x) \cdot dx = \sum_{\nu=1}^{j-1} N_{\square \nu} = \int_0^{t'_{j-1}} C(x) \cdot dx \quad (3.28)$$

The profile  $C(x)$  is given by:

$$C(x) = \begin{cases} C_1(x) & 0 \leq x \leq t_1 \\ C_j(x - t_{j-1} + t'_{j-1}) & t_{j-1} \leq x \leq t_j \end{cases} \quad (3.29)$$

An example of a 1D multilayer calculation is given in Fig. 15, for the implantation of arsenic (energy: 60 KeV, dose:  $5 \cdot 10^{15}$  at/cm<sup>2</sup>) thru 0.02  $\mu\text{m}$  of silicon oxide and 0.02  $\mu\text{m}$  of silicon nitride. We may remark that the profiles are discontinuous at the interfaces between materials, probably in a more pronounced manner than using the MC calculations [25]. However, this characteristic is always obtained with the available numerical formulations.

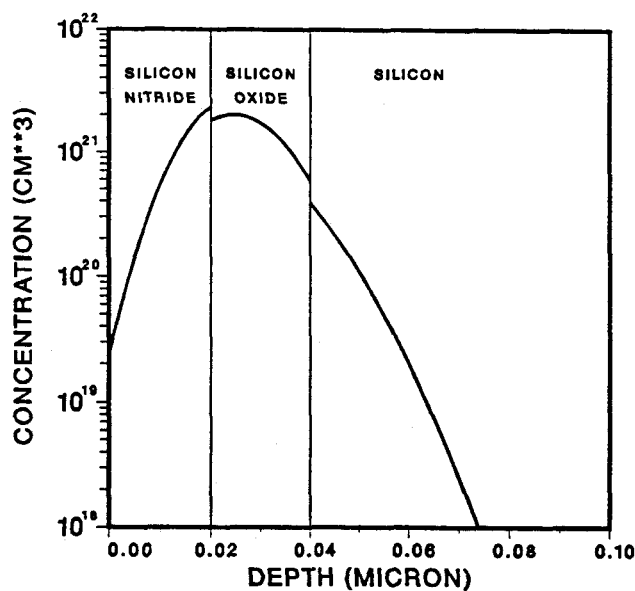


Fig. 15 Application of the dose matching model on a 3 layers structure. The energy is 60 keV and the dose is  $5 \cdot 10^{15}$  at/cm<sup>2</sup>.

## 2.3 Two-dimensional calculations

### 2.3.1 Numerical problem

#### 2.3.1.1 formulation

We now examine the numerical problems encountered in the 2D problem. The first works aimed to determine the doping profiles in silicon with a non constant mask thickness. The method is to perform a convolution of the previously described vertical profiles with a Gaussian distribution in the lateral direction:

$$C(x,y) = \frac{1}{\sqrt{2 \cdot \pi \cdot \sigma_{py}^2}} \cdot \int_{-\infty}^{+\infty} C(x; t_{\text{mask}}(y')) \cdot \exp\left(-\frac{(y-y')^2}{2 \cdot \sigma_{py}^2}\right) \cdot dy' \quad (3.30)$$

where  $\sigma_{py}$  is the lateral standard deviation (usually obtained from the tabulated values of Gibbons [24]) and  $t_{\text{mask}}(y')$  determines the mask thickness at the abscissa  $y'$ . If we assume a Gaussian profile for  $C(x)$  and the same stopping power for mask and substrate, the widely used formulation of Runge [33] is retrieved:

$$C(x,y) = \frac{N_{\square}}{2\pi \cdot \sigma_p \cdot \sigma_{py}} \cdot \int_{-\infty}^{+\infty} \exp\left(-\frac{(y-y')^2}{2 \cdot \sigma_{py}^2} - \frac{(x - (R_p - t_{\text{mask}}(y')))^2}{2 \cdot \sigma_p^2}\right) \cdot dy' \quad (3.31)$$

In our case,  $C(x; t_{\text{mask}}(y'))$  is calculated using the dose matching model, so that the accuracy of 1D calculations is also obtained in quasi 1D part of 2D simulations (it seems that there is no report about the extension in 2D of this dose matching model). The problem is to evaluate (3.30) in the general case. For this purpose, it is worthwhile to remark that this expression is always evaluated numerically unless the mask is infinitely high. In this case, if the width of the window is  $2a$ , (3.31) is analytically integrated to give:

$$C(x,y) = \frac{N_D}{\sqrt{2 \cdot \pi} \cdot \sigma_p} \cdot \exp\left(-\frac{(x-R_p)^2}{2 \cdot \sigma_p^2}\right) \cdot A \quad (3.32)$$

$$\text{with : } A = \frac{1}{2} \cdot \left[ \operatorname{erfc}\left(\frac{y-a}{\sqrt{2} \cdot \sigma_{py}}\right) - \operatorname{erfc}\left(\frac{y+a}{\sqrt{2} \cdot \sigma_{py}}\right) \right]$$

$$\text{and : } \operatorname{erfc}(y) = \frac{2}{\sqrt{\pi}} \cdot \int_y^{\infty} e^{-t^2} \cdot dt$$

The idea is then to use (3.32) to calculate (3.30). Hence, the structure is divided into many slices, between  $y_k$  and  $y_{k+1}$  as represented in Fig. 16 [34]. The contribution of this segment can be calculated with equation (3.32), when we suppose the surface to be flat between these two points. The final profile is obtained by the superposition of all contributions:

$$C(x,y) = \frac{1}{2} \cdot \sum_k C(x; t_{\text{mask}}(y_k)) \cdot \left[ \operatorname{erfc}\left(\frac{y-y_{k+1}}{\sqrt{2} \cdot \sigma_{py}}\right) - \operatorname{erfc}\left(\frac{y-y_k}{\sqrt{2} \cdot \sigma_{py}}\right) \right] \quad (3.33)$$

As a first example, Fig. 17 shows a boron implantation into a silicon trench structure with two oxide masks. 2D effects are obvious in these masking layers, whereas boron is implanted only in the bottom of the trench, due to the concave structure and re-entrant angles of the silicon shape.

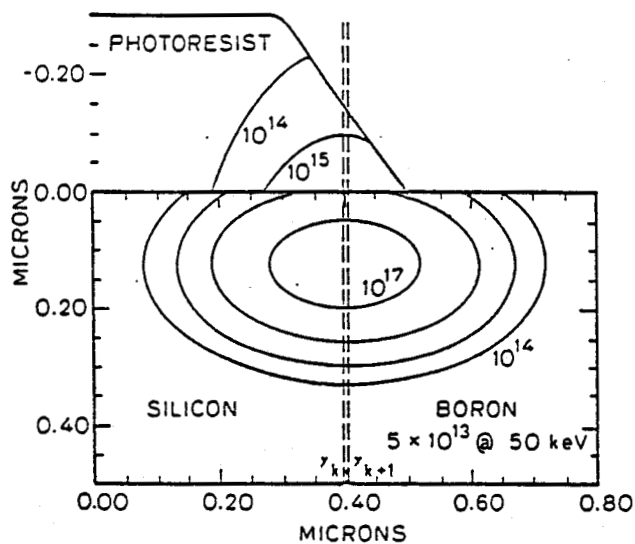


Fig. 16 Example of boron implantation profile from a single segment through a photoresist mask edge [34].

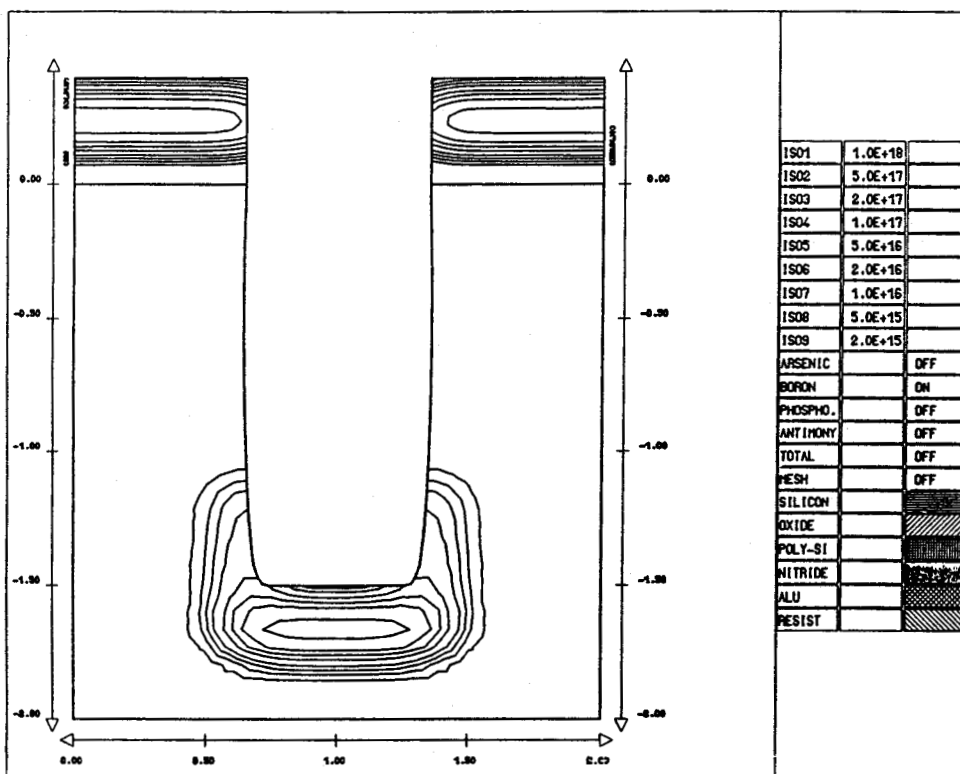


Fig. 17 Boron implantation into a silicon trench with two oxide masks.  
Energy : 50 keV. Dose:  $1.10^{13}$  at/cm<sup>2</sup>.



### 2.3.1.2 practical aspects

Equation (3.33) must be evaluated for each point of coordinates  $(x_i, y_i)$  of the simulation domain. We have adopted the following procedure (Fig. 18):

- i) The vertical distribution is calculated using the dose matching algorithm. The parameters involved in this calculation depend on the location of the integration segment  $[y_k, y_{k+1}]$  and of the materials lying on a vertical under this segment.
- ii) The value of  $\sigma_{py}$  depends only on the material that contains the point  $(x_i, y_i)$ .

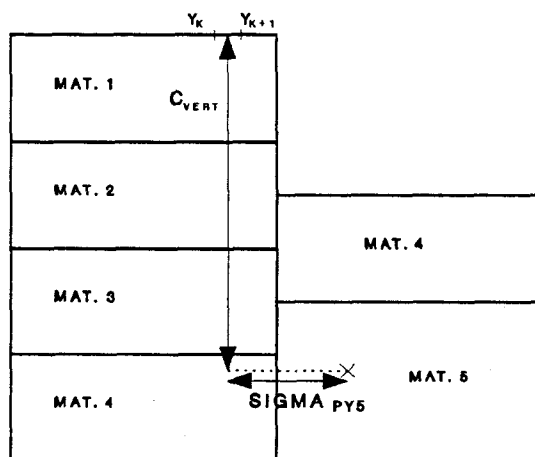


Fig. 18 Explanation of the choices in the value of the parameters for equation (3.33).

To illustrate this choice, Fig. 19a shows the result of an arsenic implant into a MOS structure with a polysilicon gate (from 0 to  $0.1 \mu m$ ) and a sidewall spacer oxide. Here we focus on the boundary vicinity between these 2 materials. The isoconcentrations shapes agree well with MC calculations in this region (Fig. 19b) and validate the pseudo-2D technique described above.

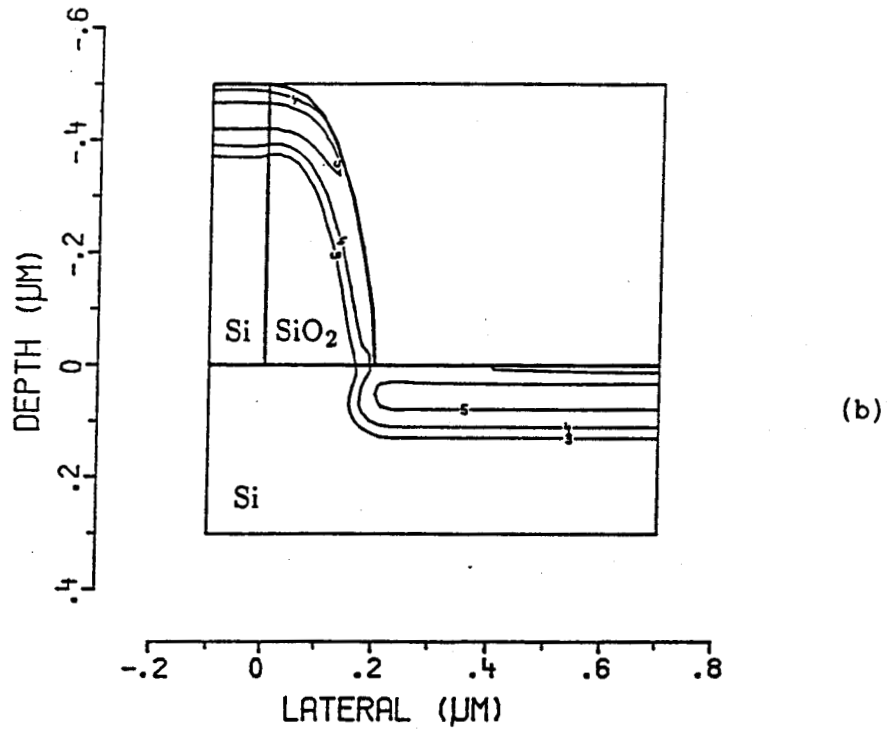
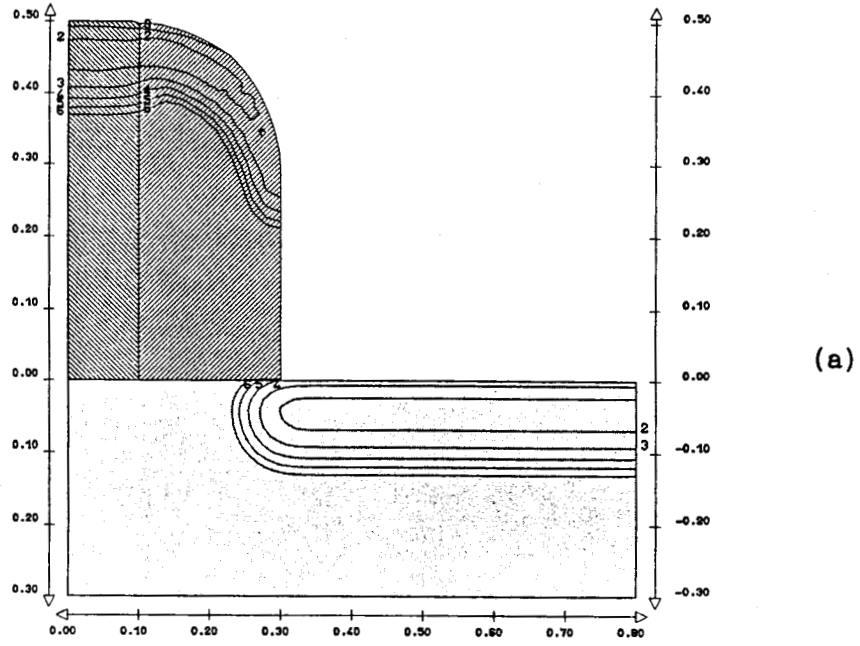


Fig. 19 Comparison between analytical and Monte Carlo calculations.  
80 keV and  $10^{15}$  at/cm<sup>2</sup> arsenic implant on a LDD structure.  
a) IMPACT4 analytical results. The contour lines represent:  
1:  $10^{21}$ , 2:  $10^{20}$ , 3:  $10^{19}$ , 4:  $10^{18}$ , 5:  $10^{17}$ , 6:  $10^{16}$  at/cm<sup>3</sup>.  
b) Monte Carlo result (contour lines represent the logarithm  
of dopant concentration divided by the dose) [35].

As shown in Fig. 17, the program takes also into account re-entrant angles. In this case, when calculating the vertical contribution, the depth  $x$  used as argument in  $C$  is minored by the "voids" (this means that implicitly we suppose that the ions do not change their direction when leaving the upper part of layers).

Since solving directly (3.33) requires large CPU time, several techniques have been developed that allows a reduction factor of 5 to 10 in computing time:

- i) The surface points  $[y_k, y_{k+1}]$  are in fact the nodes used in mesh generation. This a reasonable assumption if we compare the value of  $\sigma_{py}$  (often greater than  $100\text{\AA}$ ) with the usual inter nodes distance in surface (always the smallest of the simulation domain). This first choice minimizes the number of points.
- ii) (3.33) is evaluated for each  $[y_k, y_{k+1}]$  segment lying within a  $\pm 6.\sigma_{py}$  interval around the lateral coordinate of the point where (3.33) is to be calculated. Outside this interval, the difference between the 2 erfc functions is negligible.
- iii) For each vertical distribution, the maximum concentration is calculated. As example for a Gaussian distribution:
$$C_{\max} = \frac{N_{\square}}{\sqrt{2\pi}.\sigma_p} \quad (3.34)$$
If the value of  $C$  at the considered point is less than  $10^{-9} .C_{\max}$ , the contribution of the segment  $[y_k, y_{k+1}]$  is neglected and the erfc functions are not evaluated.
- iv) A first tabulation is determined by a pre-processor: the thickness given as input, this tabulation returns the corresponding integrated dose.

A second tabulation is then determined, and will be used extensively during the 2D calculations. For each segment  $[y_k, y_{k+1}]$ , along a vertical line, the ordinates of each crossing with underlying materials are determined, and the corresponding ordinates  $t'$  are calculated with the 1D dose matching algorithm. This means that during the 2D calculations, we have (for each point of the mesh), to calculate the vertical contribution of all the segments as stated in ii), but all the parameters of the corresponding distribution function are already computed.

### 2.3.2 Modeling aspects

In the above discussion, it has been assumed that the lateral standard deviation is constant. In recent investigations [21,36], it has been clearly shown that this parameter is depth dependent. As example, Fig. 20 displays the value of the lateral standard deviation (normalized with the standard deviation) as a function of energy and depth ( $Z/R_p$ ). These results have been obtained from extensive MC simulations [21]. Following the approach of Hobler et al, we use tabulated values of  $\sigma_{py}$  in silicon. This allows a good modeling of the lateral spread of ions, which is obviously of prime importance due to the continuous decrease in minimal device dimensions. Moreover, the increasing use of Rapid Thermal Annealing (RTA) will enforce the importance of initial conditions (e.g. the doping profile after ion implantation).

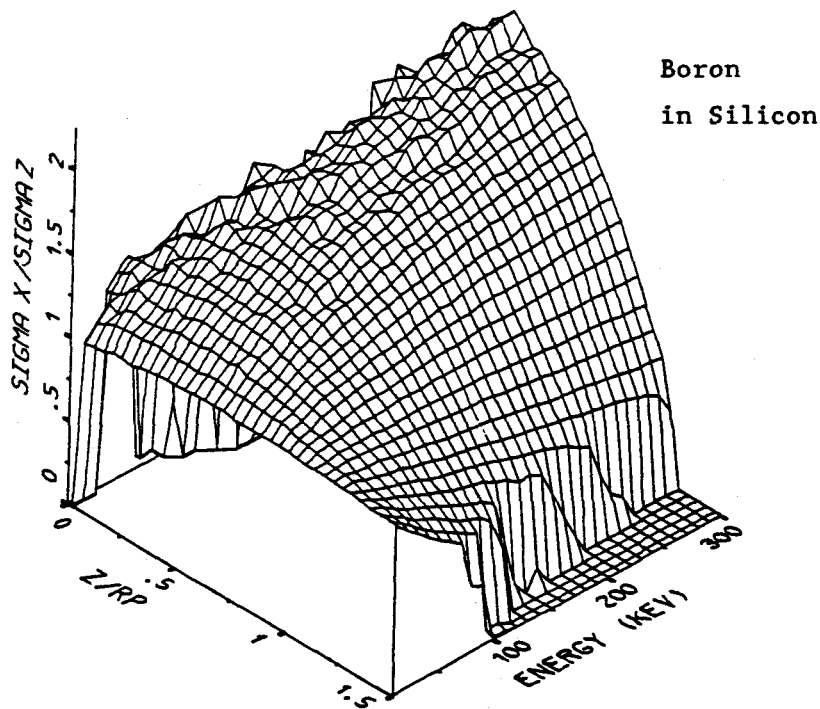


Fig. 20 Evolution of the lateral standard deviation as a function of depth and energy [21].

More sophisticated modelisations of the lateral distribution have been proposed including Pearson VII [21] or mixed moments [37]. However, from the MC results presented in [21], it is believed that the above approach is sufficient.

An other aspect concerns the simulation of tilted implants. They are used in modern technology, either to minimize the channeling effect or even to produce uniform doping inside trench structures for DRAM [38]. Here a simple technique consists in rotating the complete physical simulation domain (hence the ion beam is always normal in the coordinate system). Good description of the doping profile near the edge of the simulation domain is achieved since it is artificially extended of  $6.\sigma_{py}$  at both sides.

Fig. 21 illustrates the influence of the modeling level we have implemented in IMPACT4. Here we are interested in the source/drain arsenic ion implantation in a MOS structure. With the standard level of modeling, the resulting profile is plotted in Fig. 21a. When introducing the dependence of the lateral standard deviation with depth, less arsenic come to rest near the surface -under the oxide spacer- as depicted in Fig. 21b. Finally Fig. 21c and 21d show the influence of tilted implants with an angle of  $7^\circ$  and  $-7^\circ$ , respectively.

The advantages of the overall method are as follows: i) the integrated dose in the domain corresponds to the implanted one, ii) any type of vertical distribution can be used, iii) the calculation is performed independently of the domains and nodes ordering. On the other hand, a detailed inspection shows that this numerical procedure is still a "pseudo 2D" calculation as was implicitly described in Fig. 18. It seems that there is no analytical solution with intrinsic 2D characteristics.

Fig. 22 displays a comparison between an analytical calculation (Fig. 22a) and a MC simulation (Fig. 22b). In this latter figure, the additional concentration originates from ions which enter the mask near the edge, leave it laterally and re-enter the substrate [18]. Such results can only be obtained with MC methods.

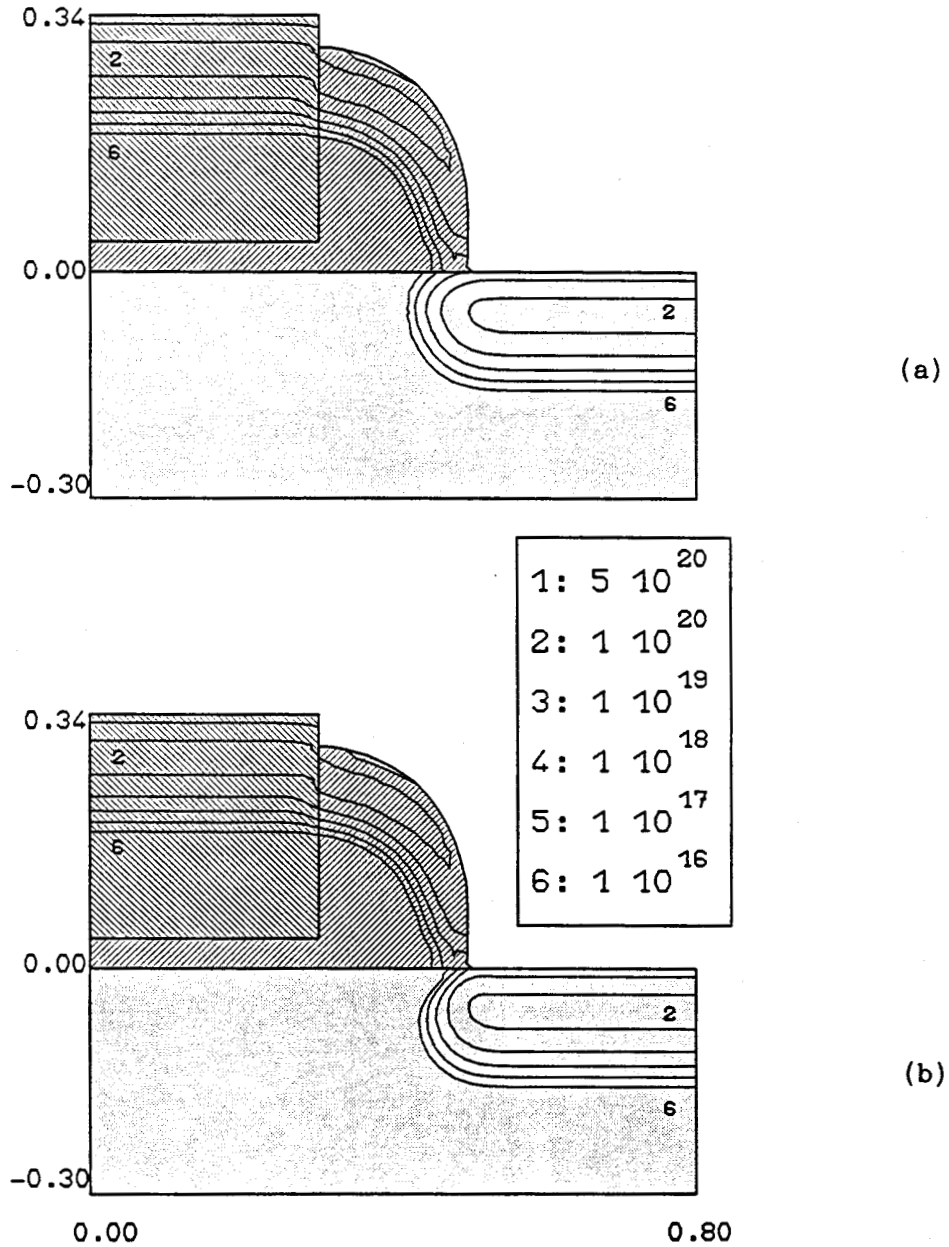


Fig. 21 Source/drain ion implantation in a LDD-MOS structure (arsenic, 100 keV and  $10^{15}$ at/cm<sup>2</sup>).

- (a): In this figure, the standard level of modeling is used: joined-half Gaussian distribution for silicon and polysilicon, Gaussian for the oxide.
- (b): The depth dependent lateral standard deviation is used. The effect of the low values of  $\sigma_{py}$  for small depths is obvious in silicon.

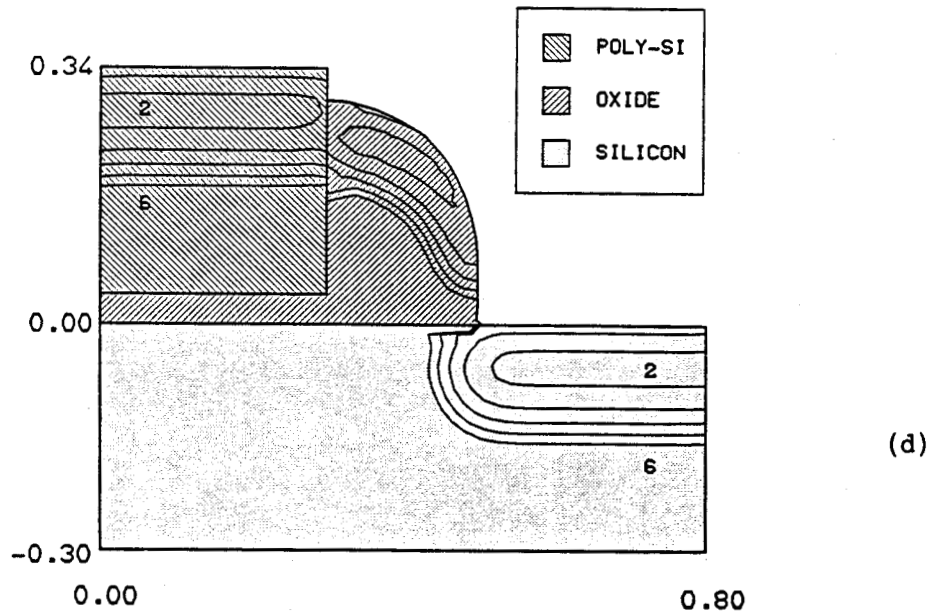
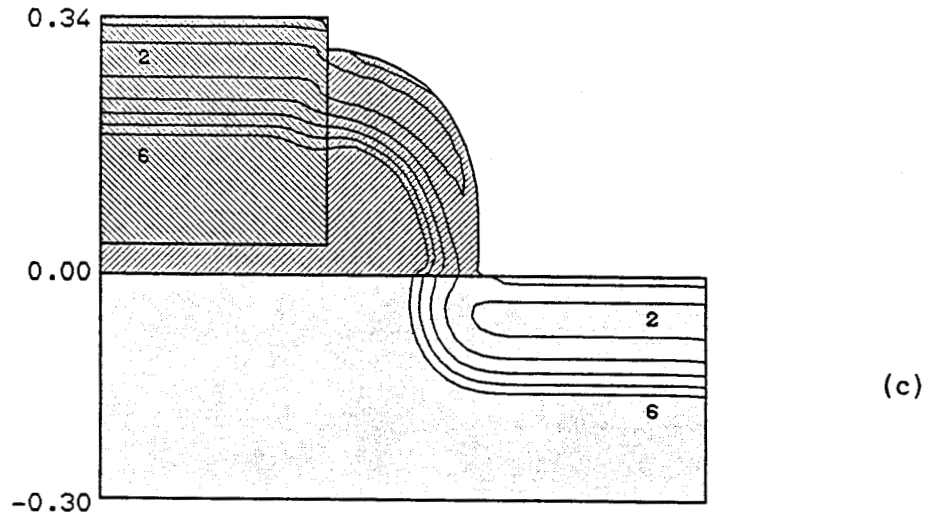


Fig. 21 (c) Same example with a tilt angle of  $7^\circ$ , resulting in large arsenic concentration in silicon, under the oxide spacer.

(d) Same example with a tilt angle of  $-7^\circ$ . In this case, the arsenic concentration is strongly affected under the spacer.

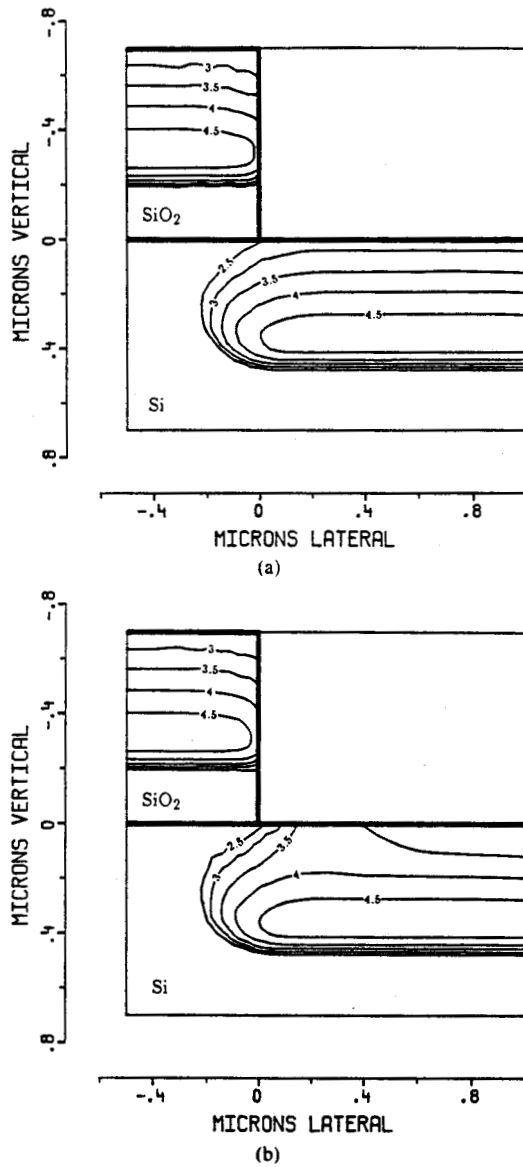


Fig. 22 Typical example of the validity limit of analytical models in the case of boron implantation (100 keV). The contour lines represent the logarithm of the dopant concentration divided by the dose [18].  
(a) Analytical results.  
(b) Full Monte Carlo simulation.



### III. DIFFUSION

Diffusion is the mechanism responsible for the redistribution of dopants in silicon. The impurities can be introduced by ion implantation or by high concentration gas. By monitoring the processing conditions (e.g. dopant type, temperature and duration of the step), desired shapes of doping profiles may be obtained, in terms of surface concentration (threshold voltage adjustment in MOS technology), dose (base Gummel number in bipolar transistor), concentration gradient (LDDMOS), or junction depth.

However, diffusion is probably the processing step where the most important progress in modeling are still expected. Although significant progress have been achieved during the last few years, controversy remains about the modeling, the equations and their related parameters (as example transient phenomena occuring during Rapid Thermal Annealing -RTA-). In this work we have chosen to focus on interfacial phenomena and the multilayer aspect. This means that standard and well established models have been preferred to unmatore or not fully tested ones. Simulation of diffusion step is here presented in 4 parts:

- i) Formulation of the diffusion equations and the related assumptions.
- ii) The modeling level incorporated in the program: expression of the diffusion coefficient, clustering effects and so forth.
- iii) Resolution by the Finite Element Method and temporal discretization.
- iv) Resolution of the linear system.

### 3.1 The diffusion equations

#### 3.1.1 Modeling levels

Due to the complexity in modeling this processing step, one should be careful to the sophistication level incorporated in process simulators. Three levels can be described (when focusing on the silicon material):

- i) Severe simplifications (as example constant diffusion coefficient, but this is not always necessary) allow to solve analytically the diffusion equation in 1D or even in 2D geometries [39,40]. Although it offers significant reduction of CPU time and also can be extended to more advanced modeling levels [41], we can not think to incorporate such techniques in a general purpose simulator.
- ii) The inclusions of the following models : concentration dependent diffusion coefficients taking into account the impurity interaction with vacancies, field enhancement factor, clustering phenomenon and Oxidation Enhanced Diffusion (OED), have proved to have a wide range of validity and reasonable values of fitted parameters. This modelisation level has been incorporated in the program and will be detailed in the following sections.
- iii) The most advanced level but also the most controversial includes the full interaction of dopants and point defects. In this case, one has to solve diffusion equations for:
  - i) the impurities with explicit effects of interstitial and vacancies,
  - ii) the points defects with appropriate boundary conditions.

Several formulations have been proposed with various levels of complexity [41-46]. These are often very recent works (apart from the last reference) reflecting the degree of comprehension of these problems. A complete and efficient treatment of the most advanced level of modeling is beyond the scope of this work (see for example [47]).

### 3.1.2 From Fick's laws to the general formulation

The first Fick's law relates the flux of impurity  $i$  to the concentration gradient:

$$\vec{J}_i = -D_i \cdot \vec{\nabla} C_i \quad (3.35)$$

where  $C_i$  is the impurity concentration and  $D_i$  the diffusion coefficient. In addition to the diffusion term, the effect of the internal electric field (conduction term) is incorporated in the expression of  $\vec{J}_i$ :

$$\vec{J}_i = -D_i \cdot \vec{\nabla} C_i - Z_i \cdot \mu_i \cdot C_i \cdot \vec{E} \quad (3.36)$$

where  $Z_i$  : is the charge state of the impurity (+1 for singly ionized acceptors, -1 for singly ionized donors).

$\mu_i$  : is the mobility ( $\text{cm}^2 \text{v}^{-1} \text{s}^{-1}$ ).

$\vec{E}$  : is the local electric field.

By applying Einstein's relation, the mobility is replaced by the diffusion coefficient (assuming the validity of Boltzmann statistics):

$$D_i = \frac{kT}{q} \cdot \mu_i \quad (3.37)$$

On the other hand, the second Fick's law, assuming the mass conservation, is simply expressed as:

$$\frac{\partial C_i}{\partial t} = - \text{div } \vec{J}_i \quad (3.38)$$

By substituting (3.36) into (3.38), we obtain the classical diffusion equation, expressed here in 1D for simplicity:

$$\frac{\partial C_i}{\partial t} = \frac{\partial}{\partial x} \left[ D_i \cdot \left( \frac{\partial C_i}{\partial x} + Z_i \cdot \frac{q}{kT} \cdot C_i \cdot E \right) \right] \quad (3.39)$$

We now focus on the expression of the electric field. For this purpose, we can use the Poisson equation:

$$\text{div}(\vec{\nabla}\psi) = \frac{q}{\epsilon} \cdot (n-p-C) \quad (3.40)$$

where  $C$  is the total net concentration of ionized impurities,  $\psi$  is the built-in potential,  $n$  and  $p$  are the electrons and holes concentrations, respectively. Two alternatives must be discussed: whether i) (3.40) is solved exactly or ii) by assuming the local charge neutrality. In this latter case, the left hand term is neglected. An extensive comparison between these two methods has been performed in [48] and has shown that the doping profiles computed with these two approaches are very similar, although the charge neutrality assumption is not strictly valid in the case where several species diffuse during the same step. If the diffusion time is greater than a given transient mechanism (approximately of the order of 100s), - that is usually the case - quite the same results are obtained. For this reason, we do not need to solve Poisson equation.

A second level of simplification is further introduced if we assume the validity of Boltzmann statistics for the calculation of  $n$  and  $p$  in (3.40). In fact, in the case of high doping levels, the semiconductor becomes degenerate and the impurities are not fully ionized.  $C$ ,  $n$  and  $p$  should then be calculated using Fermi-Dirac statistics as proposed in [49]. However, it complicates the formulation, with small implications on the final results since the computed electric field obtained with Fermi-Dirac statistics is only slightly higher than that computed with the assumptions of complete ionized impurities and validity of Boltzmann statistics. These simplifications are always used in process simulators.

Within the frame described above, (3.40) is rewritten as:

$$n_1 \cdot e^{\left(\frac{q\psi}{kT}\right)} - n_1 \cdot e^{\left(\frac{q\psi}{kT}\right)} - C = 0 \quad (3.41)$$

The electric field is then calculated explicitly:

$$E = - \frac{\partial\psi}{\partial x} = - \frac{kT}{q} \cdot \frac{1}{2 \cdot n_1 \cdot \sqrt{1 + \left(\frac{C}{2n_1}\right)^2}} \cdot \frac{\partial C}{\partial x} \quad (3.42)$$

Combining (3.39) and (3.42) form:

$$\frac{\partial C_i}{\partial t} = \frac{\partial}{\partial x} \left[ D_i \cdot \left( \frac{\partial C_i}{\partial x} - Z_i \cdot \frac{C_i}{2n_i \sqrt{1 + \left(\frac{C}{2n_i}\right)^2}} \cdot \frac{\partial C}{\partial x} \right) \right] \quad (3.43)$$

In the case of a single diffusing specy, it is convenient to introduce the factor h:

$$h = 1 + Z_i \cdot \frac{C_i}{2n_i \sqrt{1 + \left(\frac{C}{2n_i}\right)^2}} \quad (3.44)$$

h varies between 1 ( $C_i \ll n_i$ ) and 2 ( $C_i \gg n_i$ ) and is usually called the "electric field" enhancement factor.

### 3.1.3 Boundary conditions

In a general manner, the boundary conditions apply at the interfaces between two materials. If one of the material is a gas (without dopant in high concentration), an evaporation condition is usually assumed:

$$J_i = -h_e \cdot C_i \quad (3.45)$$

where  $h_e$  is the mass transport coefficient ( $\text{cm.s}^{-1}$ ),  $C_i$  the surface concentration of impurity i, and  $J_i$  the flux defined positive from the gas to the material. If dopant is present in this gas, a predeposition process will occur (this was the first technique to introduce dopants in silicon):

$$J_i = h_p \cdot (C_{is} - C_i) \quad (3.46)$$

The flux  $J_i$  now depends on the surface concentration  $C_{is}$ , imposed by the processing conditions. As the value of  $h_p$  is usually very large (about  $10^{-6} \text{ cm.s}^{-1}$ ), a Dirichlet condition is often applied:  $C_i = C_{is}$ . There is an upper bound to this formulation since there is a maximum concentration above which no more impurities can be introduced in silicon. This concentration is called the solid solubility of the impurity and depends on the type of dopants and on the temperature.

Simpler conditions are assumed at the left and right sides of the domain (and also in the bottom), that is there is no outward normal flux of dopants through these boundaries:

$$\frac{\partial C_i}{\partial n} = 0 \quad (3.47)$$

Finally, considering an interface between two materials handled by the program, the segregation phenomenon must be treated (unequal distribution of dopants occur between the two materials). The flux of dopants from region 1 to region 2 is given by:

$$J_{12} = - h. ( C_1 - m_{12} \cdot C_2 ) \quad (3.48)$$

where  $C_1$  and  $C_2$  are the interfacial concentrations in regions 1 and 2, respectively.  $m_{12}$  is the equilibrium segregation coefficient:

$$m_{12} = \left. \frac{C_1}{C_2} \right|_{eq.} \quad (3.49)$$

It must be noted here that the diffusion equation, described in (3.43) or in an other form depending on the material properties, has to be solved in the two materials. The flux (3.48) is used as a link between these two layers. It has been reported that the oxidation of arsenic or phosphorus-doped silicon results in significant pile-up of dopants at the silicon interface. An empirical formulation has been proposed [50]:

$$C_I = 2.10^{-7} \cdot C_B \cdot \exp\left(\frac{0.26}{kT}\right) \quad (3.50)$$

where  $C_I$  is the pile-up dose ( $\text{cm}^{-2}$ ) and  $C_B$  is the concentration in silicon away from this pile-up region. However, there is no satisfactory modelisation of this phenomenon, not included in this work. The numerical implications of equation (3.48) are reported in details in section 3.4 and the formulation of segregation under oxidizing ambient is detailed in section 4.3.

### 3.2 Modeling

#### 3.2.1 Silicon

##### 3.2.1.1 Diffusion coefficient

Impurities diffuse in silicon via vacancy and interstitial mechanisms as illustrated in Fig. 23. In the level of modeling we have chosen, the preference is the vacancy diffusion mechanism. It has been used for some times [50] and well adjusted parameters allow good results. In this formulation, since we assume the validity of Boltzmann statistics, a relation can be established between vacancy concentrations and the diffusion coefficient. In the extrinsic condition, the dopant concentration is greater than  $n_i$  and  $n_i$  is also greater than the concentration of point defects, so that the Fermi level depends only on the ionized dopant concentration. As an example, the acceptor type vacancy concentration can be expressed as:

$$\frac{C_V^-}{(C_V^-)^i} = \exp\left(\frac{E_F - E_i}{kT}\right) \quad (3.51)$$

where :  $C_V^-$  is the acceptor vacancy concentration in extrinsic silicon  
 $(C_V^-)^i$  is the acceptor vacancy concentration in intrinsic silicon  
 $E_F$  is the Fermi level of the extrinsic silicon (eV)  
 $E_i$  is the intrinsic Fermi level

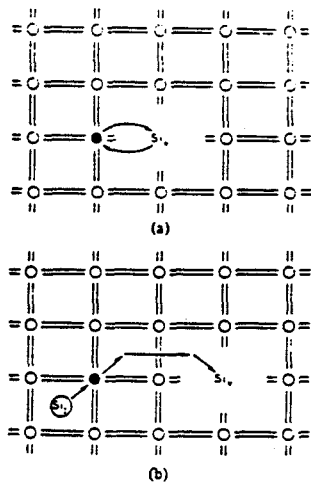


Fig. 23 Vacancy and interstitial-assisted diffusion mechanisms [50].

- a) Idealized vacancy mechanism
- b) Watkins replacement mechanism (a silicon interstitial displaces a substitutional impurity, which travels interstitially to a vacant lattice site).

Since  $n/n_i$  can be also expressed as a function of the Fermi level, we obtain:

$$\frac{C_v^-}{(C_v^-)^i} = \frac{n}{n_i} \quad (3.52)$$

On the other hand, if we assume that the diffusion coefficient is proportional to the vacancy concentration, we can write

$$\frac{D}{D_i} = \frac{n}{n_i} \quad (3.53)$$

where  $D$  and  $D_i$  are the coefficient diffusion in extrinsic and intrinsic silicon, respectively. Generalizing this formulation to the other charge states of vacancies leads to the widely used equation for  $D$ :

$$D = D^x + D^- \left( \frac{n}{n_i} \right) + D^= \left( \frac{n}{n_i} \right)^2 + D^+ \left( \frac{n_i}{n} \right) \quad (3.54)$$

where  $D^*$  is the intrinsic diffusivity associated with particular charge states of point defects (we assume  $D^x$  to be independent of the Fermi level).  $D^*$  is usually expressed in an Arrhenius form:

$$D^* = D_o^* \cdot \exp \left( - \frac{E_a^*}{kT} \right) \quad (3.55)$$

These values have been determined experimentally. In fact, for each specy, only specific vacancy charge states play a role in (3.54). Arsenic and antimony diffuse with neutral and singly negative charged vacancies. Boron diffuses with neutral and singly positive charged vacancies. Phosphorus diffuses with neutral and negative charged vacancies, but (3.54) is insufficient to describe correctly the observed experimental profiles of phosphorus, as reported in 3.2.1.2.



The following table summarizes the numerical values for the prefactor  $D_0^*$  ( $\text{cm}^2\text{s}^{-1}$ ) and the activation energy  $E_a^*$  (eV) [51-52]:

Element	$D_0^*$	$E_a^*$	$D_0^-$	$E_a^-$	$D_0^=$	$E_a^=$	$D_0^+$	$E_a^+$
B	0.037	3.46					0.76	3.46
P	3.85	3.66	4.44	4.00	44.2	4.37		
Sb	0.213	3.65	15.0	4.08				
As	0.066	3.44	12.0	4.05				

As example Fig. 24 shows the evolution of arsenic diffusivity with the donor concentration.

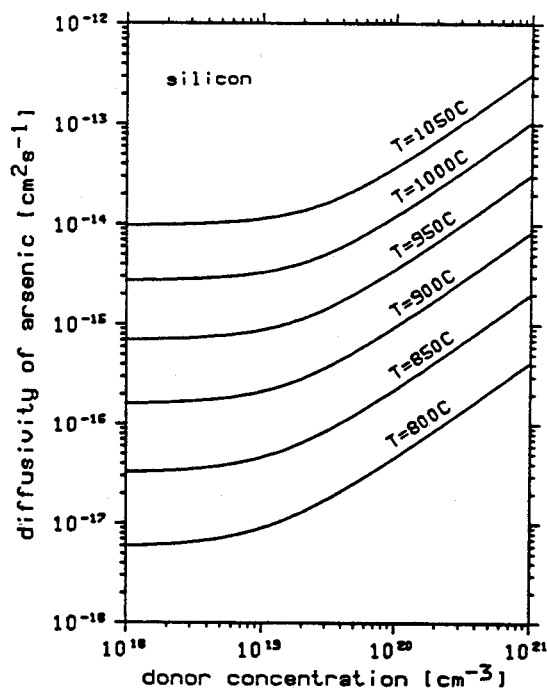


Fig. 24 Diffusivity of arsenic versus donor concentration in silicon [25].

3.2.1.2 Phosphorus diffusion

Phosphorus exhibits the most complex behaviour concerning diffusion.

"Anomalous" effects are often observed in high concentrations profiles:

- i) "kink" formation
- ii) enhanced tail diffusion
- iii) enhancement of the other impurities diffusion.

This latter aspect was very detrimental in earlier bipolar technologies where phosphorus was used as dopant for emitter. In this case, the enhanced diffusion of boron modifies the base-collector junction. Fair proposed a model for phosphorus diffusion, that revealed to explain in a quantitative manner these anomalous effects [52]. This model is now almost always used in the process simulators. It is briefly described in the following.

Three regions can be determined (Fig. 25):

- i) the high concentration region
- ii) the kink region
- iii) the tail region (low concentrations)

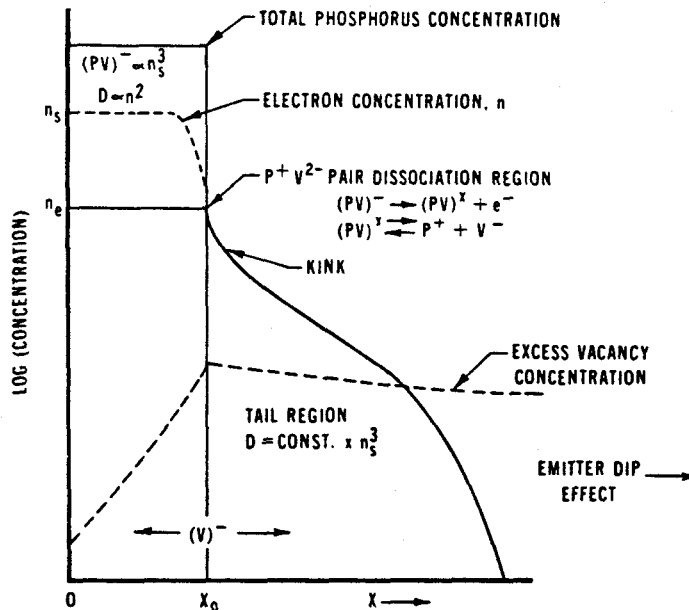


Fig. 25 Idealized phosphorus profile and the corresponding three regions accordingly to Fair-Tsai theory [52].

In the first region, some of the phosphorus ions ( $P^+$ ) pair with doubly charged vacancies ( $V^{2-}$ ) to form  $(PV)^-$  complex. It is believed that the concentration of  $(PV)^-$  depends on the surface concentration  $n_s$ . In this high concentration region, the diffusion coefficient is given by:

$$D_{HC} = D^x + D^- \cdot \left(\frac{n}{n_i}\right)^2 \quad (3.56)$$

In the kink region, when the electron concentration reaches a given value  $n_e$ , dissociation of  $(PV)^-$  occurs, giving rise to an excess in vacancy concentration in the tail region. This dissociation level  $n_e$  is given by:

$$n_e = 4.65 \cdot 10^{21} \exp\left(-\frac{0.39}{kT}\right) \text{ cm}^3 \quad (3.57)$$

The diffusion in the kink region where  $n_{\text{kink}} < n < n_e$  is expressed as:

$$D_{\text{kink}} = \left( D^x + D^- \cdot \left(\frac{n_e}{n_i}\right)^2 \right) \cdot \left(\frac{n_e}{n}\right)^2 \quad (3.58)$$

The diffusion in the tail region ( $n < n_{\text{kink}}$ ) is expressed as:

$$D_{\text{tail}} = D^x + D^- \cdot \frac{n_s^3}{n_e^2 n_i} \cdot \left[ 1 + \exp\left(\frac{0.3}{kT}\right) \right] \quad (3.59)$$

$n_{\text{kink}}$  is defined as the concentration where  $D_{\text{kink}} = D_{\text{tail}}$ . As  $D_{\text{tail}}$  is proportional to  $n_s^3$ , this latter value must be calculated at each time step, but it seems also to be a very large (and ambiguous) dependence. Finally, an empirical relation between the total phosphorus concentration and the free electron concentration reads:

$$C = n + 2.04 \cdot 10^{-41} \cdot n^3 \quad (3.60)$$

In the case of very high surface concentration (approximately above  $5 \cdot 10^{20}$  at/cm<sup>3</sup>), this approach is not sufficient since it predicts a too large enhancement of diffusivity in the tail region. To correct the model, the effect of band-gap narrowing, induced by lattice strains, has been included [53]. This band-gap narrowing is given by:

$$\Delta E_g = -1.3 \cdot 10^{22} (C_s - C_z) \text{ eV} \tag{3.61}$$

where :  $C_z = \frac{2.43 \cdot 10^{26}}{T_c^2} \text{ cm}^{-3}$  ( $T_c$  in celsius degrees)

The final equations of the diffusion coefficient for phosphorus are:

$$D_{HC} = \left( D^x + D^- \cdot \left( \frac{n}{n_i} \right)^2 \cdot \exp \left( \frac{\Delta E_g}{kT} \right) \right) \tag{3.62}$$

$$D_{kink} = \left( D^x + D^- \cdot \left( \frac{n_e}{n_i} \right)^2 \cdot \exp \left( \frac{\Delta E_g}{kT} \right) \right) \cdot \left( \frac{n_e}{n} \right)^2 \tag{3.63}$$

$$D_{tail} = D^x + D^- \cdot \frac{n_s^3}{n_e^2 n_i} \cdot \left[ 1 + \exp \left( \frac{0.3}{kT} \right) \right] \cdot \exp \left( \frac{3\Delta E_g}{kT} \right) \tag{3.64}$$

Fig. 26 shows the effect of this additional term. Although this overall modelisation has proved his capability in reproducing experimental results, recent investigations seem to demonstrate that interstitials should play also an important role in phosphorus diffusion. There is still a need for developing a more physical model [47].

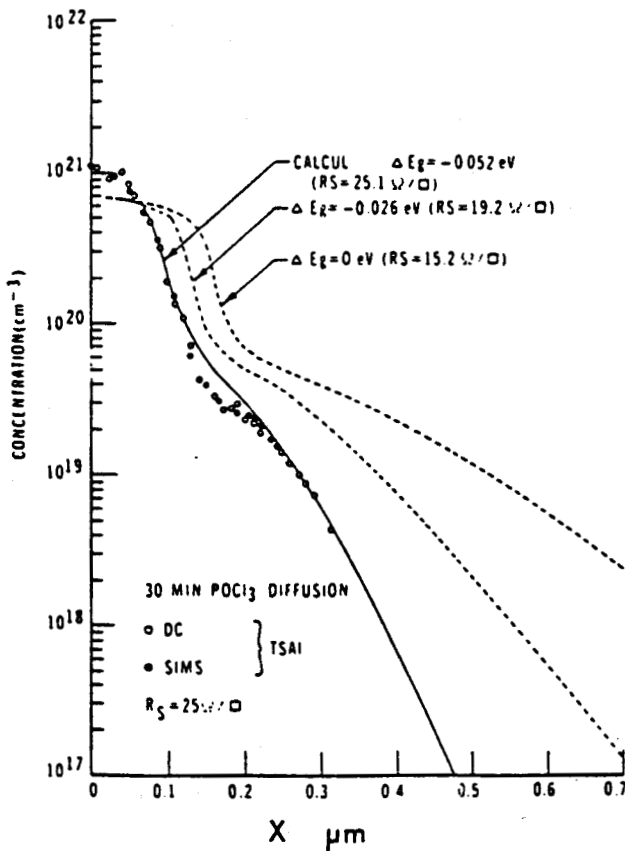


Fig. 26 Influence of band-gap narrowing on simulated phosphorus profiles [53].

### 3.2.1.3 Clustering phenomenon

It is now well-known that above a given concentration, the dopants are electrically inactive. This phenomenon has been explained by precipitation or clustering formation. In the following we will describe briefly a classical model of arsenic clustering, proposed by Guerrero in [54], determining a relation between the arsenic concentration  $C$  and the active concentration  $N$ . Clustering (or declustering) is described by the following reaction equation:



where  $m=3$ , number of atoms per cluster. In fact it is generally assumed that the effect of dynamic clustering and declustering is negligible. By using mass action law, we can express a relation between the free electron concentration and the active concentration:

$$n = \frac{N}{1 - (m-1) \cdot kN^m} \quad (3.66)$$

where  $k$  is the equilibrium constant. (3.66) induces an upper limit for  $N$ :

$$N_{\max} = \frac{1}{[(m-1)k]^{1/m}} \quad (3.67)$$

$N_{\max}$  is defined here as the solubility limit of arsenic. As  $k$  depends on temperature,  $N_{\max}$  is approximated by:

$$N_{\max} = 1.89 \cdot 10^{22} \exp\left(-\frac{0.453}{kT}\right) \text{ cm}^3 \quad (3.68)$$

Finally, the relation between total and active concentrations is given by:

$$C = N \cdot \left[ 1 + \frac{\frac{m}{m-1} \left(\frac{N}{N_{\max}}\right)^m}{\left(1 - \frac{N}{N_{\max}}\right)^m} \right] \quad (3.69)$$

This relation is very useful for two reasons:

- i) It can predict the fraction of dopants that are electrically active.
- ii) As the clusters are immobile, it influences the final diffusion profile. This means that (3.36) must be modified to take into account clustering effects. If only arsenic diffuses [55]:

$$\frac{\partial C}{\partial t} = \frac{\partial}{\partial x} \left[ D \cdot \frac{\partial C}{\partial x} \left( \frac{\partial N}{\partial C} + \frac{N}{\sqrt{4n_i^2 + \left( N + \frac{m-1}{m} \cdot (C-N) \right)^2}} \cdot \left( \frac{m-1}{m} + \frac{1}{m} \cdot \frac{\partial N}{\partial C} \right) \right) \right] \quad (3.70)$$

Concerning boron, we have used the experimental results of Ryssel [56-57]. In this case, we suppose 12 atoms per cluster and the following relation holds:

$$C = N + m \cdot N_{\max}^{1-m} \cdot N^m \quad (3.71)$$

### 3.2.2 The other materials: silicon oxide, silicon nitride and polysilicon

Throughout the section 3.2.1, we have seen the difficulties in establishing the physical background for the comprehension of diffusion processes in silicon. This situation is also encountered in the other materials, especially in polysilicon.

#### 3.2.2.1 silicon oxide and silicon nitride

The formulation of diffusion in silicon oxide is greatly simplified compared to that in silicon. This means that we suppose a constant diffusivity and the use of Fick's laws in their simplest form. In this case, the diffusion equation is silicon oxide and in silicon nitride is given by:

$$\frac{\partial C}{\partial t} = D \frac{\partial^2 C}{\partial x^2} \quad (3.72)$$

Numerical values for  $D$  are the same than those used in the SUPREM III program. In any case, as these values are strongly reduced compared to silicon, we can consider that for these materials the most important phenomena are i) segregation effects at the interfaces, ii) initial distributions when using ion implantation. However, the knowledge about the diffusion of impurities in silicon oxide is far from complete and recent investigations have tried to establish a more physical comprehension [58].

### 3.2.2.2 Polysilicon

Polysilicon is a material of major importance in silicon technology. It is used extensively for gate formation in MOS technology but also, since a few years, in bipolar technology in order to enhance the emitter efficiency. Polysilicon is often grown by using low pressure chemical vapor deposition (LPCVD) with possible *in situ* doping. Unfortunately, polysilicon has extremely complex behaviours. As example, depending on temperature and deposition rates, amorphous or polycrystalline structures may be obtained. On the other hand, a good modeling of grain growth and grain boundaries is a very difficult task (it is believed that grain boundaries allow rapid diffusion path for dopants). A model is implemented in SUPREM III which simulates with various success diffusion in polysilicon [59]. An other approach has been also proposed in ASPREM [60]. However an important progress has been performed in 2D with the effects of dopant segregation, grain growth and parallel diffusion of the dopants along the grain boundaries and inside grains [61-62]. This kind of work allows to extract macro-models, applicable in general purpose simulators, in particular it should be helpful when considering polysilicon as a diffusion source.

Here, we have chosen as a first step to implement the same models in polysilicon than in silicon. If we consider very large grains, this assumption is realistic. For practical purposes, it is often convenient to multiply by a given factor (usually 100) the diffusion coefficients in silicon to obtain those in polysilicon. An other aspect that has been treated in this work deals with the polysilicon/silicon interface where segregation-like conditions have been used. This point will be reported in more details in the application chapter.

### 3.3 Resolution with the Finite Element Method

In the previous sections we have focused our attention on the formulation of the diffusion equation in the general case, and on a brief review of the modeling aspects. This results in a set of coupled non-linear partial differential equations (PDE) that has obviously no analytical solution in the general case. The discretization in space and time of these PDE gives a finite number of unknowns, the concentrations at each point of the discretization grid. From a practical point of view, three steps are performed to solve the diffusion problem:

- i) Discretization of the differential equations in space. It is assumed that the simulation domain has been divided in an appropriate manner (mesh definition).
- ii) Resolution of the non linear and non-stationary problems. In our case, this first problem is well reflected by the evolution of the diffusion coefficient with dopant concentration (Fig. 24).
- iii) Resolution of the linear system of equations, by direct or iterative methods.

The overall resolution of such problems is very consuming both in terms of memory and number of operations (CPU time). This explains that carefull choices must be decided in order to achieve interesting performances and sometimes to avoid numerical non-convergence. In this part, the application of the Finite Element Method (FEM) to the diffusion of dopants is described and the temporal aspect is also treated. As the ultimate problem is the resolution of the linear system, it is detailed specially in section 3.4.

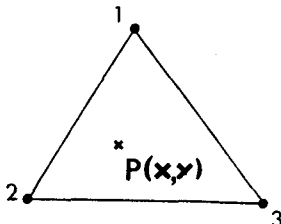


### 3.3.1 Application of FEM to the diffusion of dopants

FEM has been first employed in the field of mechanics [63] but its potentiality has allowed its extension towards other problems and in particular, in the field of process and device simulation. We have already justified the choice of FEM over Finite Difference Method in chapter II (2.1). The application of FEM to the diffusion of dopants in silicon has been addressed by previous works in our laboratory. It has been extensively reported in [55] and here we present only the major features.

As already stated, the simulation domain is divided in triangular elements. The value of a variable at any point is calculated using a linear combination of the values at each node of the triangle containing this particular point.

For our present problem, the unknowns are the concentration at these nodes. In a triangle 1 depicted below, the concentration at the point  $P(x,y)$  is expressed as:

$$C(x,y) = \sum_{i=1}^3 N_{i1}(x,y) \cdot C_i \quad (3.73)$$


The diagram shows a triangle with three nodes labeled 1, 2, and 3. Node 1 is at the top vertex, node 2 is at the bottom-left vertex, and node 3 is at the bottom-right vertex. A point labeled  $P(x,y)$  is marked with an asterisk inside the triangle.

where  $C_i$  is the concentration at the node  $i$  and  $N_{i1}$  is the shape function of node  $i$  in element 1. Here, we have chosen 3-noded triangular elements with linear shape functions (triangular P1 elements). The 3 shape functions depend only on the position of the point in the element and are non-zero only in the element 1. P1 elements can be expressed and manipulated very simply (compared to higher order elements).

Considering as  $x_i$  and  $y_i$  the coordinates of node  $i$  in the general system of coordinates, these functions  $N_{i1}$  are defined as [63]:

$$N_{11}(x,y) = \frac{[x_2 y_3 - x_3 y_2 + x(y_2 - y_3) + y(x_3 - x_2)]}{2.S_1} \quad (3.74a)$$

$$N_{21}(x,y) = \frac{[x_3y_1 - x_1y_3 + x(y_3 - y_1) + y(x_1 - x_3)]}{2.S_1} \quad (3.74b)$$

$$N_{31}(x,y) = \frac{[x_1y_2 - x_2y_1 + x(y_1 - y_2) + y(x_2 - x_1)]}{2.S_1} \quad (3.74c)$$

where  $S_1$  is the triangle area :

$$S_1 = \frac{[x_1y_2 - x_2y_1 + x_2y_3 - x_3y_2 + x_3y_1 - x_1y_3]}{2} \quad (3.74d)$$

For the sake of simplicity, we can re-write (3.73) as:

$$C = \sum_{il} N_{i1} . C_i \quad (3.75)$$

where  $il$  is the set of nodes contained in element 1.

Due to the particular problem we have to solve that includes strong non-linearities, a variational formulation is not always available, so we use a weighted residual approach. If we suppose two diffusing species (U and V), the differential equation (3.43) that governs the diffusion of a dopant U can be expressed as:

$$\frac{\partial U}{\partial t} = \frac{\partial J_x}{\partial x} + \frac{\partial J_y}{\partial y} \quad (3.76)$$

$$\text{with : } J_v = D_U \frac{\partial U}{\partial v} + D_V \frac{\partial V}{\partial v} \quad (v=x \text{ or } v=y)$$

where  $D_U$  and  $D_V$  are the respective contributions to the diffusivity of U. (3.76) can be rewritten in a simpler form:

$$H(U) = 0 \quad (3.77)$$

The method is in fact to compute the unknown vector  $\widehat{U}$  (concentration at each node of the mesh), in order to minimize the spatial error  $R(x,y)$ :

$$H(\widehat{U}) = R(x,y) \quad (3.78)$$

The weighted residual approach consists in calculating the  $N$  (total number of unknowns) values  $U_i$  that fullfil the following condition:

$$\int_{\Omega} w_j(x,y) \cdot R(x,y) \cdot dx \cdot dy = 0 \quad (3.79)$$

where  $w_j$  is the weighting function. By introducing (3.76) we obtain:

$$\int_{\Omega} w_j \cdot \left( \frac{\partial J_x}{\partial x} + \frac{\partial J_y}{\partial y} \right) \cdot d\Omega - \int_{\Omega} w_j \cdot \frac{\partial U}{\partial t} \cdot d\Omega = 0 \quad (3.80)$$

The Galerkin procedure states that this function  $w_j$  is in fact the union of the shape functions  $N_{j1}$  associated with node  $j$  (that is each element containing this node). Equation (3.80) is further modified by i) applying Green transformation, ii) replacing the integrals over the domain by the sum of the integral over each element, and iii) using the shape functions [55]. The  $j$ th equation is then as follows:

$$\begin{aligned} F_j = & - \sum_{elj} \int_{\Omega_l} \left[ N_{j1x} \cdot \left( D_U \cdot \sum_{i1} N_{i1x} \cdot U_i + D_V \cdot \sum_{i1} N_{i1x} \cdot V_i \right) + \right. \\ & \left. N_{j1y} \cdot \left( D_U \cdot \sum_{i1} N_{i1y} \cdot U_i + D_V \cdot \sum_{i1} N_{i1y} \cdot V_i \right) \right] \cdot d\Omega_l \\ & + \sum_{elj} \int_{\Gamma_lj} N_{j1} \cdot F \cdot d\Gamma_l - \sum_{elj} \int_{\Omega_l} N_{j1} \cdot \sum_{i1} N_{i1} \cdot \frac{\partial U_i}{\partial t} \cdot d\Omega_l = 0 \end{aligned} \quad (3.81)$$

where :  $elj$  is the set of triangles  $l$  containing the node  $j$   
 $N_{j1x}$  is the spatial derivate of  $N_{j1}$  with respect to  $x$   
 $\Gamma_lj$  are the sides of elements  $l$  containing node  $j$ , and located on the domain boundary  
 $F$  is the boundary flux

Actually , all the terms contained in (3.81) are computed from the geometrical data and are used to build a system of N coupled equations of this form. The two points we have now to detail are i) the treatment (and the discretization) of the temporal term  $\frac{\partial U}{\partial t}$ , and ii) the resolution of the non-linear problem.

### 3.3.2 Non-linear and temporal aspects

The set of equations (3.81) can be expressed synthetically as:

$$K(U,V).U + C.\frac{\partial U}{\partial t} = 0 \quad (3.82)$$

where U is the vector of nodal concentration values, K(U,V) and C are the stiffness and "capacity" matrix, respectively. Focusing on the time dependent problem, we will investigate the possible temporal discretization schemes. In a general manner, this problem can be solved from time t to time t+Δt by setting:

$$[\alpha.U_{t+\Delta t} + (1-\alpha).U_t].K(U_{t+\alpha\Delta t},V_{t+\alpha\Delta t}) + \frac{1}{\Delta t}.[U_{t+\Delta t} - U_t].C = 0 \quad (3.83)$$

The parameter α, taken to be in the interval [0,1], determines the method to be used:

α	Method
0	Forward differences;forward Euler;explicit
½	Trapezoidal rule;midpoint rule;Crank-Nicholson
1	Backward differences;backward Euler;implicit

(3.84)

If α=0, the solution is immediate since:

$$K(U_t,V_t).U_t + \frac{1}{\Delta t}.[U_{t+\Delta t} - U_t].C = 0 \quad (3.85)$$

However, this method is nowadays rarely used, as stability imposes severe time step restriction and hence this method is conditionnaly stable.

In this work, we prefer to use the implicit method which is unconditionnaly stable (the limit on the value of the time step is determined by the accuracy). in this case, (3.83) reduces to:

$$K(U_{t+\Delta t}, V_{t+\Delta t}) \cdot U_{t+\Delta t} + \frac{1}{\Delta t} \cdot [U_{t+\Delta t} - U_t] \cdot C = 0 \tag{3.86}$$

This formulation complicates the resolution, due to the non-linear aspects. In the following we will compare two solutions:

- i) Use of Newton-Raphson method to solve (3.86)
- ii) Use of the incomplete implicit scheme.

### 3.3.2.1 The Newton-Raphson method

The Newton-Raphson is a well-known technique for solving non linear problems. Assuming that  $U^{k-1}$  is the solution at the (k-1)th iteration, the solution vector at the next iteration must be chosen in order to minimize the error (or residual) F. By applying the method of Newton-Raphson we obtain:

$$-\Delta U^k \cdot \left[ \frac{\partial F}{\partial U} \right]_{U^{k-1}} = F(U^{k-1}) \tag{3.87}$$

where  $\partial F / \partial U$  (or G in the following) is usually called the tangential or jacobian matrix. A term  $G_{j,i}$  is then expressed as:

$$G_{j,i}^k = \frac{\partial F_j^k}{\partial U_i^k} \tag{3.88}$$

where  $F_j^k$  is the jth row of F, at the kth iteration. We can express the terms of the jacobian and of the left-hand side (B):

$$G_{ji}^k = - \sum_{elij} \int_{\Omega} D_U^{k-1} \cdot [N_{j1x} \cdot N_{i1x} + N_{j1y} \cdot N_{i1y}] \cdot d\Omega \quad (3.89)$$

$$+ \sum_{elij} \int_{\Gamma_{lij}} N_{j1} \cdot \frac{\partial F^k}{\partial U_i^k} \cdot d\Gamma - \frac{1}{\Delta t} \cdot \sum_{elij} \int_{\Omega} N_{j1} \cdot N_{i1} \cdot d\Omega$$

and:

$$B_j^k = \sum_{elj} \int_{\Omega} \left[ N_{j1x} \cdot \left( D_U^{k-1} \cdot \sum_{ml} N_{m1x} \cdot U_m^{k-1} + D_V^{k-1} \cdot \sum_{im} N_{m1x} \cdot V_m^{k-1} \right) \right.$$

$$\left. + N_{j1y} \cdot \left( D_U^{k-1} \cdot \sum_{ml} N_{m1y} \cdot U_m^{k-1} + D_m^{k-1} \cdot \sum_{im} N_{m1y} \cdot V_m^{k-1} \right) \right] \cdot d\Omega$$

$$- \sum_{elj} \int_{\Gamma_{lj}} N_{j1} \cdot F^{k-1} \cdot d\Gamma + \frac{1}{\Delta t} \cdot \sum_{elj} \int_{\Omega} N_{j1} \cdot \sum_{ml} N_{m1} \cdot (U_m^{k-1} - U_m^0) \cdot d\Omega \quad (3.90)$$

Convergence of this method is achieved in our case usually in about 3 to 6 iterations, when the following criterion is fulfilled:

$$\max \left( \frac{C^k - C^{k-1}}{C^k} \right) \leq 1\% \quad \text{with : } C^k \geq 10^{10} \text{ at/cm}^3 \quad (3.91)$$

At this point we must notice that we make use of a mass-lumping technique that enforces the diagonal of the jacobian matrix. Considering the temporal term in (3.89), we report all the offdiagonal terms on the diagonal element (e.g.  $G_{jj}$ ). This means that the temporal term is re-written as:

$$- \frac{1}{\Delta t} \cdot \sum_{elj} \int_{\Omega_1} N_{j1} \cdot d\Omega_1 \quad (3.92)$$

3.3.2.2 The incomplete implicit scheme

The idea is here to linearize directly (3.86). For this purpose, the stiffness matrix is evaluated at time  $t$  [64-65]:

$$K(U_t, V_t) \cdot U_{t+\Delta t} + \frac{1}{\Delta t} \cdot [U_{t+\Delta t} - U_t] \cdot C = 0 \quad (3.93)$$

This means that a single loop of the Newton-Raphson method is performed. A study of the number of temporal loops needed to obtain a certain accuracy has been performed in [64]. As example, Fig. 27 shows the evolution of the junction depth versus the number of loops, for several degrees of non-linearity, in the case of a predeposition step. This figure can be used as abacus in order to select a time step for a given accuracy.

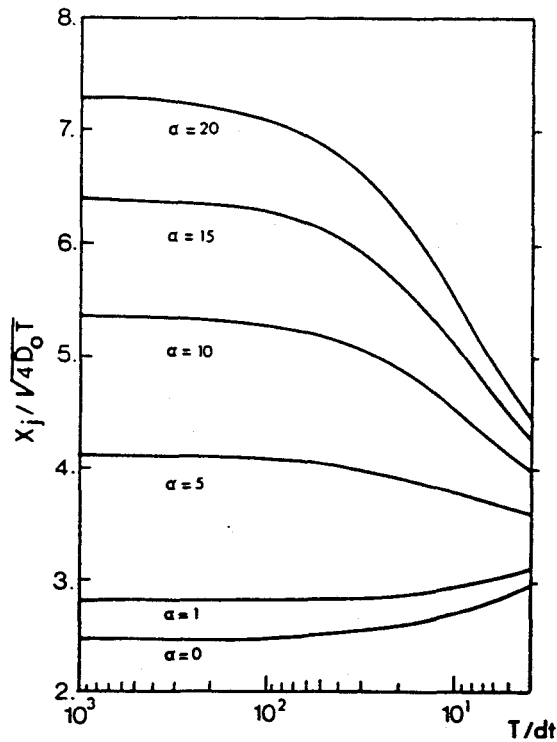


Fig. 27 Junction depth variation versus the number of loops for several non-linearity degrees ( $\alpha = \frac{C_{max}}{2 \cdot n_1}$ ) [64].

Choosing between i) the Newton-Raphson method and ii) the incomplete implicit scheme depends in fact on the relative importance of the non-linear and non-stationary aspects. For this purpose we use a test example, consisting in a diffusion step after an arsenic ion implantation. A constant time step value is used in the two schemes, that means the total diffusion time divided by the number of temporal loops. Fig. 28 shows the different behaviours of the two methods for several time steps. For the incomplete implicit scheme we can notice that in the case of small numbers of time steps (below 10), the non-linear aspect is not well taken into account, resulting in an artificial reduced gradient of concentration. An other behaviour is shown in Fig. 28b for the Newton-Raphson method, but in this case the accuracy is limited by the number of time steps, and large errors concerning the junction depth occur.

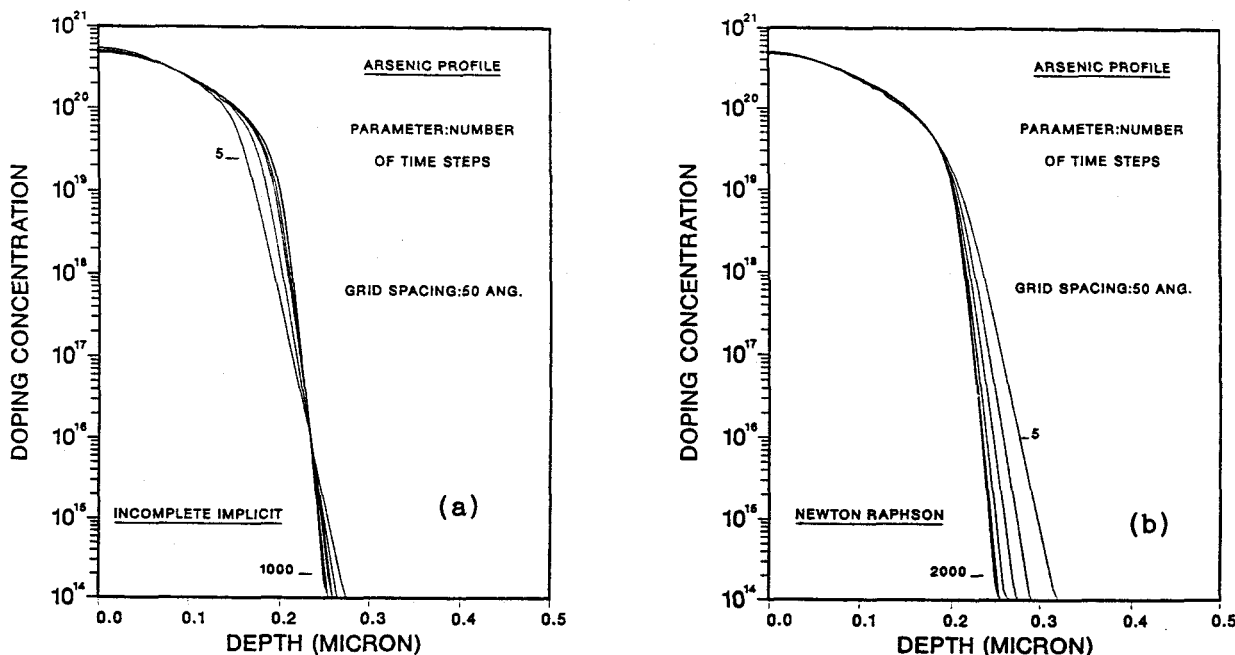


Fig. 28 Arsenic doping profile evolution with the number of time steps as a parameter (1000°C, 20 minutes).

- (a) incomplete implicit scheme: 5,10,20,30,100,1000 time steps
- (b) the Newton-Raphson method : 5,10,20,50,200,2000 time steps



The methods are evaluated by two approaches. In the first approach, a relative error is calculated:

$$\text{Err} = \frac{1}{N} \cdot \sum_{i=1}^N \frac{|C_i - \tilde{C}_i|}{\tilde{C}_i} \quad (3.94)$$

where  $N$  is the number of points,  $C$  is the concentration resulting from the conditions we want to evaluate and  $\tilde{C}$  is the best available solution (incomplete implicit scheme with 1000 time steps or Newton-Raphson method with 2000 time steps, since the two methods converge to the same solution).

In the second approach the error on the junction depth (for  $10^{15}$  at/cm<sup>3</sup>) is determined:

$$\text{Err}_{X_j} = \frac{X_j - \tilde{X}_j}{\tilde{X}_j} \quad (3.95)$$

Again,  $\tilde{X}_j$  is the junction depth deduced from the best available results. We measure the performances of each scheme by plotting the CPU time needed in order to reach a given accuracy. Fig. 29 and 30 show the results for the first and second approaches, respectively. It is easily deduced that the incomplete implicit scheme offers better features in terms of compromise between precision and CPU time. This shows that i) the temporal aspect is the major problem and that ii) we need a certain number of temporal loops in order to reach a reasonable accuracy.

It should be noted here that we have used the pure Newton-Raphson method. It could be improved with more sophisticated strategies, but it is believed that it should not reach the efficiency of the incomplete implicit scheme, since the non-stationary aspect predominates. It seems also that there is, up to now, no need for more sophisticated schemes in the choice of the time step. Such kind of scheme has been proposed for example in [66]. Finally, concerning the influence of the grid spacing and node density, a complete study is presented in chapter IV.

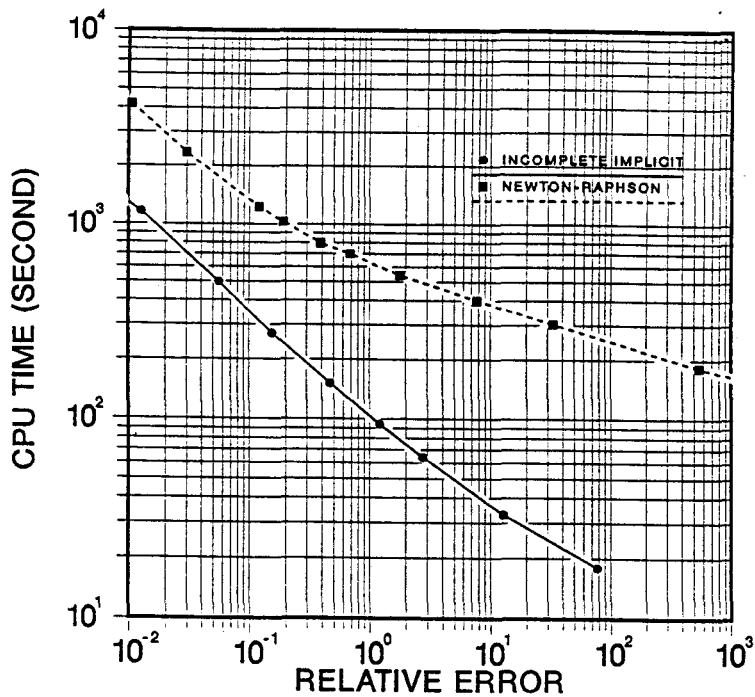


Fig. 29 Relative error versus CPU time for the two temporal schemes.

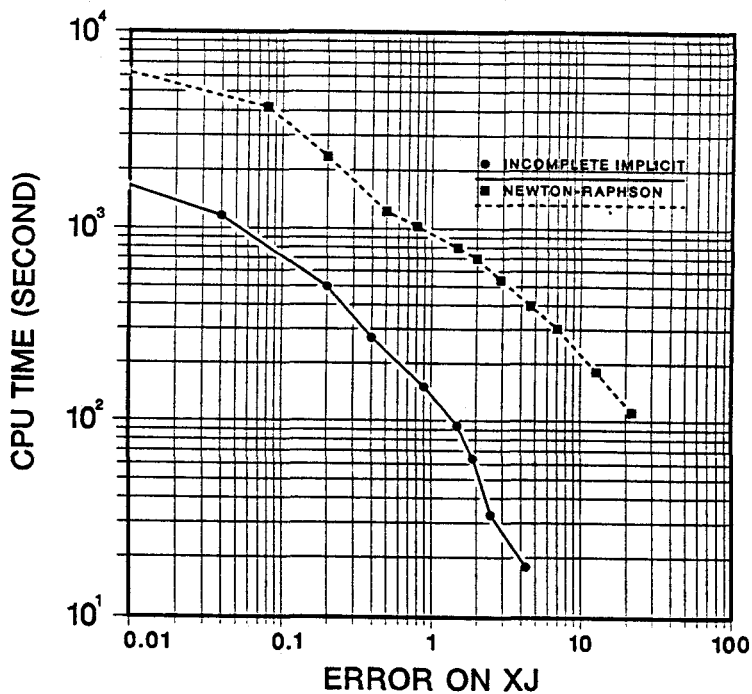


Fig. 30 Same figure for the accuracy on the junction depth (expressed in % at 10<sup>15</sup> at/cm<sup>3</sup>).

### 3.4 Resolution of the linear system

#### 3.4.1 Direct and iterative methods

Once the system of equations has been linearized, we have to solve a matrix system of the form (3.87). The jacobian matrix is sparse (i.e., contains many zeros), as each node has a reduced number of neighbors (less than 11 in general with the type of triangulation we have used). It has important implications for the choice of the linear solver. If we use meshes with regular shapes, or if the diffusion equation is solved only in silicon, it is possible to minimize the matrix bandwidth by a proper node ordering. As example the Cuthil Mc Kee algorithm is often used [64]. In this case direct methods are very efficient in terms of CPU time and memory consumption (as example the Gaussian elimination).

In our case this solution is not optimal since a node (and domain) ordering can not be very efficient, due to the general multilayer configuration. This means that direct methods are no more competitive (apart from the matrix assembly, a major portion of calculation is devoted to this resolution step and the performance of the linear solver determines the overall algorithm speed).

This means that one has to examine iterative methods. Successive-Over-Relaxation (SOR) method has been tested for diffusion in silicon only, and even in this simple case, a rapid decrease of the residual norm was only obtained when the jacobian is strictly diagonal dominant that requires very short time steps. Also, SOR implies large CPU ressources in the case of long time diffusion (well and LOCOS formations). It is possible to improve the performance of this method, but more sophisticated temporal schemes should be included [67].

In the following we present a solution based on conjugate gradient method with incomplete LDU decomposition. Before a presentation of this method, it is worthwhile to precise the matrix structure.

### 3.4.2 Matrix structure in the multilayer problem and unknown scaling

The diffusion equation for each specy is solved in a sequential decoupled manner. As example, if we study a coupled diffusion of arsenic and boron, (3.87) is solved for arsenic, then for boron (using the arsenic concentration deduced from the first resolution), and again for arsenic... This strategy is very efficient as it reduces drastically the memory size and increses the implicit degree, without any loss in precision for the interacting species [14,68]. In this case, the matrix is symmetric if the diffusion equation is solved in a single domain.

In the multilayer case, the matrix becomes asymmetric. Considering a simple test example (Fig. 31) with three domains, we now detailed the jacobian matrix. The points of interest are situated at the interfaces between several layers (2 for points A, B and C, and 3 for the point D). In the following we assume that the node ordering determining the matrix structure is as follows:

$$1, \dots, N_1, N_1+1, \dots, N_1+N_2, N_1+N_2+1, \dots, N_1+N_2+N_3 \quad (3.96)$$

where  $N_1$ ,  $N_2$  and  $N_3$  are the number of nodes in domain 1, 2 and 3 respectively. Considering now two nodes  $i$  and  $j$  at the interface between two domains, we recall the expression of the segregation fluxes:

$$\begin{aligned} F_{ij} &= h.(C_i - m_{ij}.C_j) & \text{with: } m_{ij} &= \left. \frac{C_i}{C_j} \right|_{\text{eq.}} \\ F_{ji} &= h.(m_{ij}.C_j - C_i) \end{aligned} \quad (3.97)$$

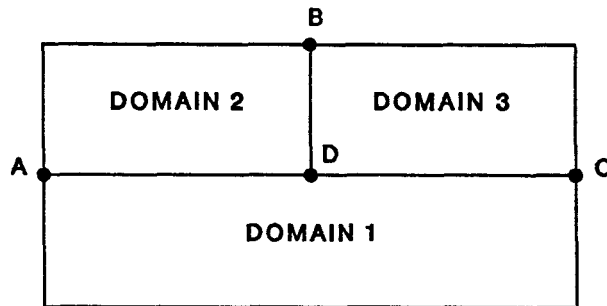


Fig. 31 Simple test example with 3 domains.

Using the Newton-Raphson method, the following terms will be incorporated in the jacobian matrix:

$$\frac{\partial F_{ij}}{\partial C_i} = h \quad ; \quad \frac{\partial F_{ij}}{\partial C_j} = -h \cdot m_{ij} \quad ; \quad \frac{\partial F_{ji}}{\partial C_i} = -h \quad ; \quad \frac{\partial F_{ji}}{\partial C_j} = h \cdot m_{ij} \quad (3.98)$$

With these above considerations it is possible to represent the matrix system (Fig. 32). If for example the material 1 is silicon and material 2 is oxide, a redistribution of arsenic implies a value for  $m_{12}$  of about 10 to 100, depending on the modeling. In this case, the matrix is not diagonal dominant. It has of course severe implications on the rate of convergence of an iterative method, in the extreme case convergence may not be reached, depending on the particular equations and algorithms.

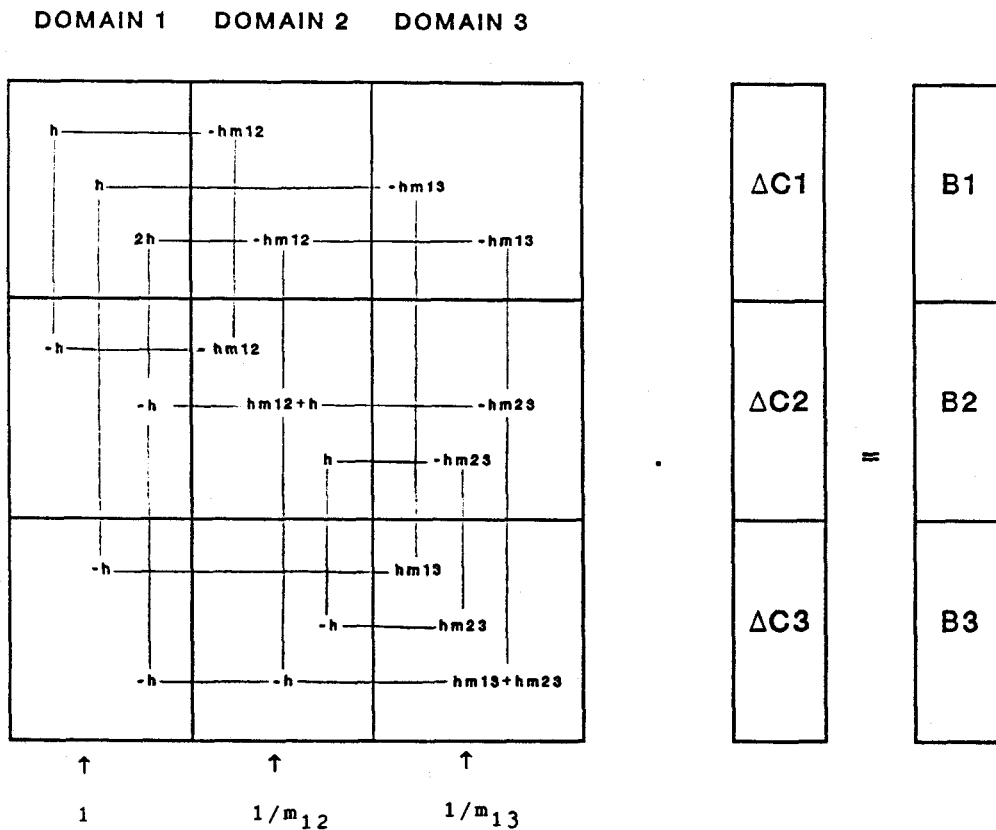


Fig. 32 Matrix structure showing the incorporation of the boundary flux terms in the jacobian matrix.

In order to accelerate the convergence rate, a scaling technique is introduced [14,69]. Due to the practical applications of the program, the first domain is always silicon. Hence, the columns of the jacobian matrix are divided by the value of the segregation coefficient (1 for the first set of columns,  $m_{12}$  in the second set of columns and  $m_{13}$  in the third set, as illustrated in Fig. 32). Considering the particular case of the triple node D, and calling D1, D2 and D3 the node numbers in domains 1,2 and 3 respectively, the initial value of the diagonal term in the jacobian matrix is:

$$G_{D3} = \frac{\partial F_{D3D1}}{\partial C_{D3}} + \frac{\partial F_{D3D2}}{\partial C_{D3}} = h \cdot m_{13} + h \cdot m_{23} \quad (3.99)$$

Assuming that relation (3.100) is satisfied, the term  $G_{D3}^{sc}$  after scaling is then given by (3.101):

$$m_{ij} = m_{ik} \cdot m_{kj} \quad (3.100)$$

$$G_{D3}^{sc} = \frac{G_{D3}}{m_{13}} = h \cdot \left( 1 + \frac{m_{23}}{m_{13}} \right) = h \cdot (1 + m_{21}) = h \cdot \left( 1 + \frac{1}{m_{12}} \right) \quad (3.101)$$

The final structure of the matrix system is given in Fig. 33, where it can be seen that the jacobian matrix is now diagonal dominant.

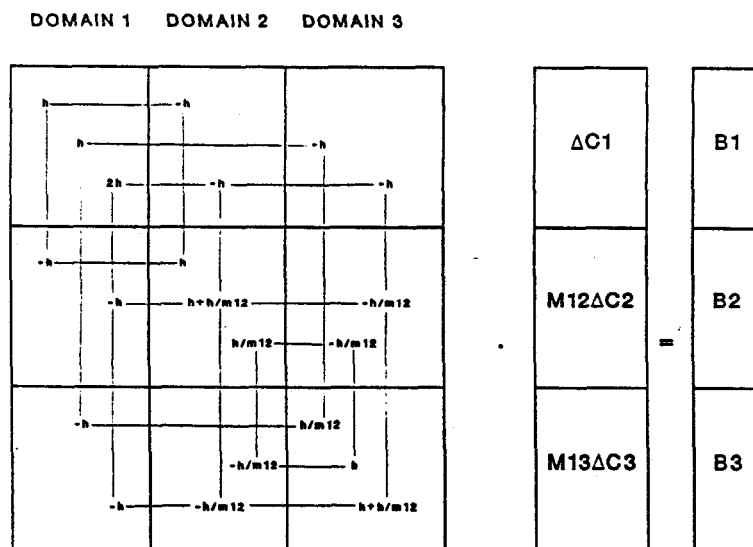


Fig. 33 Final structure of the matrix system after scaling.  
(the vector of unknowns has been multiplied by  $m_{11}$ ).

### 3.4.3 Conjugate gradient and incomplete decomposition

Once the scaling has been performed, the matrix system is solved by a conjugate gradient method in conjunction with an incomplete decomposition. Rewriting simply the linear system to be solved as  $Ax=b$ , the idea behind this kind of method is to approximate  $A^{-1}$  by a matrix  $\tilde{A}^{-1}$ . An iterative method is then applied on the system  $\tilde{A}^{-1} \cdot \tilde{A}x = \tilde{A}^{-1}b$ . As  $\tilde{A}^{-1} \cdot \tilde{A}$  resembles to the identity matrix (all eigenvalues should be close to 1), fast convergence is achieved when using conjugate gradient.

The approximate matrix is usually deduced from incomplete decomposition [70]. This leads to the now well-known ICCG methods (incomplete Choleski and conjugate gradients). In our case, we prefer a simpler approach where an incomplete LDU decomposition is performed. It was proposed in the case of slightly asymmetric matrices, with a band structure resulting from finite difference device simulation [71].

If we call  $a_i$  the diagonal elements of  $A$ ,  $b_i$  the off-diagonal elements, the matrix  $A$  is approximated by LDU where  $L$  and  $U$  are lower and upper triangle matrices, and  $D$  a diagonal matrix. In this method, the off diagonal elements of  $L$  and  $U$  are equal to those of  $A$ . The diagonal elements of  $L$  and  $U$  are denoted as  $\tilde{a}_i$ . The following recurrent formulation defines the value of  $\tilde{a}_i$ , and the value of  $\tilde{d}_i$  (diagonal elements of  $D$ ):

$$\tilde{a}_i = \frac{1}{\tilde{d}_i} = a_i - \sum_{j=1}^{i-1} b_j^2 \cdot \tilde{d}_j \quad (3.102)$$

One other interest of this decomposition method is that only the values of  $\tilde{d}_i$  must be stored, since  $L$  and  $U$  are directly deduced from  $A$  (jacobian matrix). A standard conjugate gradient method is then applied using this LDU decomposition:

$$\left. \begin{aligned} x_0 &= (\text{LDU})^{-1}b \\ r_0 &= b - Ax_0 \\ p_0 &= (\text{LDU})^{-1}r_0 \end{aligned} \right\} \text{Initialisation}$$

(3.103)

$$\left. \begin{aligned} \alpha_k &= \frac{(r_k, (\text{LDU})^{-1}r_k)}{(p_k, Ap_k)} \\ x_{k+1} &= x_k + \alpha_k p_k \\ r_{k+1} &= r_k - \alpha_k Ap_k \\ \beta_k &= \frac{(r_{k+1}, (\text{LDU})^{-1}r_{k+1})}{(r_k, (\text{LDU})^{-1}r_k)} \\ p_{k+1} &= (\text{LDU})^{-1}r_{k+1} + \beta_k p_k \end{aligned} \right\} \text{Iterations}$$

Fig. 34 shows the evolution of the residual vector norm with the iteration number, for different values of time steps. It corresponds to the first loop of the annealing test depicted in fig. 28a. When the time step value is increased, the diagonal terms are decreased (see for example (3.92)). This means that the number of iterations is affected by larger time steps, but not in a dramatic manner in terms of convergence and overall performance as demonstrated in Fig. 34.

Investigations have been performed using an ICCG method, but it was found that no significant improvements occurred, but with a more important need in memory consumption. It should be noted that the use of such methods is not current in the field of process simulation. An other scheme has been also proposed in [72], where the asymmetric system (as the diffusion for all the species is considered in the same matrix system) is solved by incomplete LDU factorization and conjugate gradient squared method (CGS).



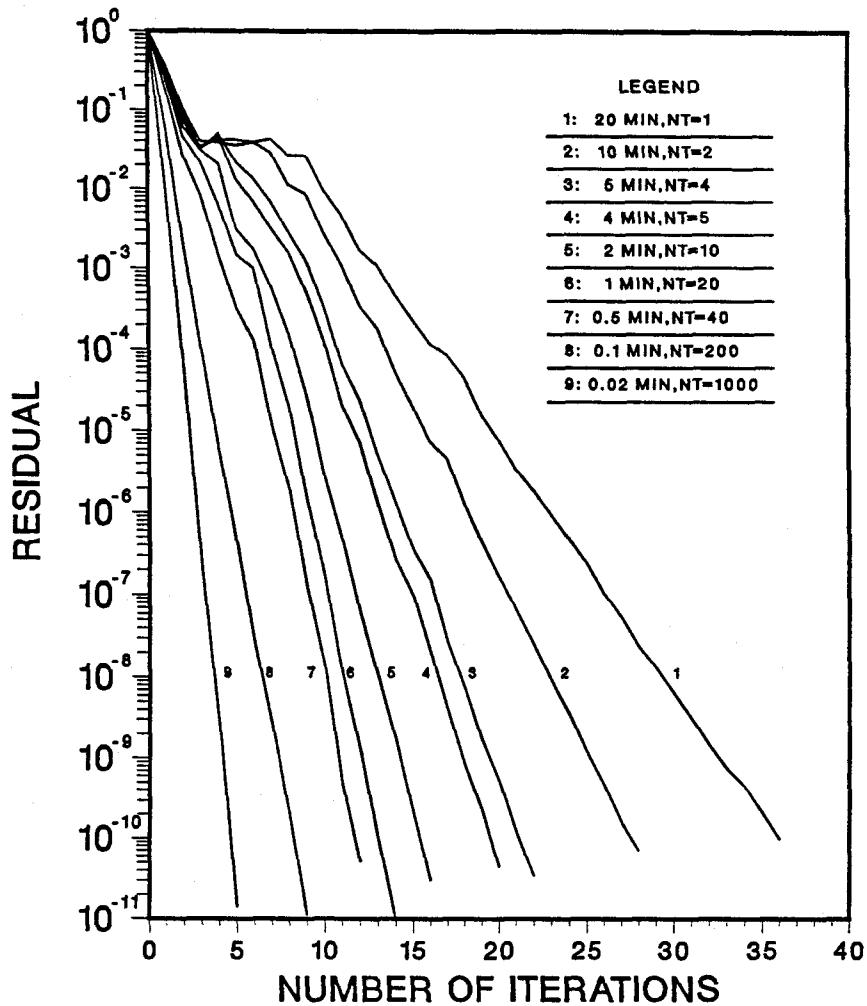


Fig. 34 Evolution of the residual vector norm  $\left\| \frac{r_k}{r_0} \right\|$  with the number of iterations for several values of the time step.

Finally, three applications are briefly presented. Fig. 35 shows a comparison between a result from SUPREM II and the present formulation, in the case of a phosphorus profile at the Si/SiO<sub>2</sub> interface, after an annealing in neutral ambient. The good agreement validates the treatment proposed above. Fig. 36 shows the arsenic contours in a MOS structure after annealing at 1000°C for 30mn (the initial structure has been already presented in Fig. 21a after ion implantation). This simulation performed with 30 temporal loops requires 20 mn on an IBM 4341 (1 mips), with a mesh containing 1057 nodes. 2D segregation effects are obvious at the Si/SiO<sub>2</sub> sidewall spacer interface.

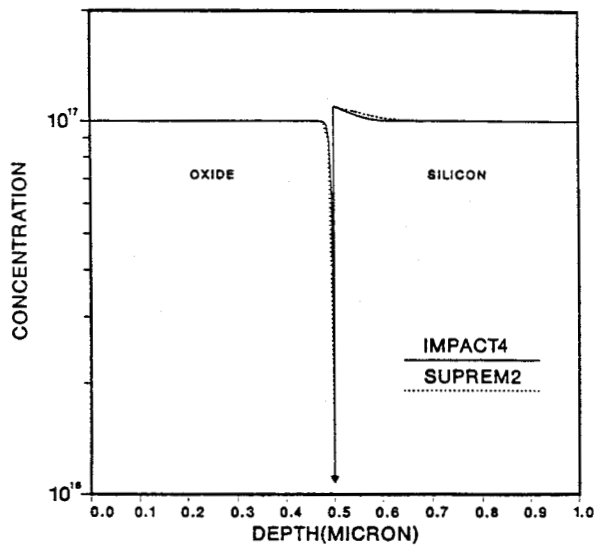


Fig. 35 Comparison between SUPREM II and IMPACT 4 for phosphorus redistribution at the silicon/silicon oxide interface. The two materials are initially doped at a constant level of  $10^{17}$  at/cm<sup>3</sup>, and are then annealed at 1000°C for 30 minutes.

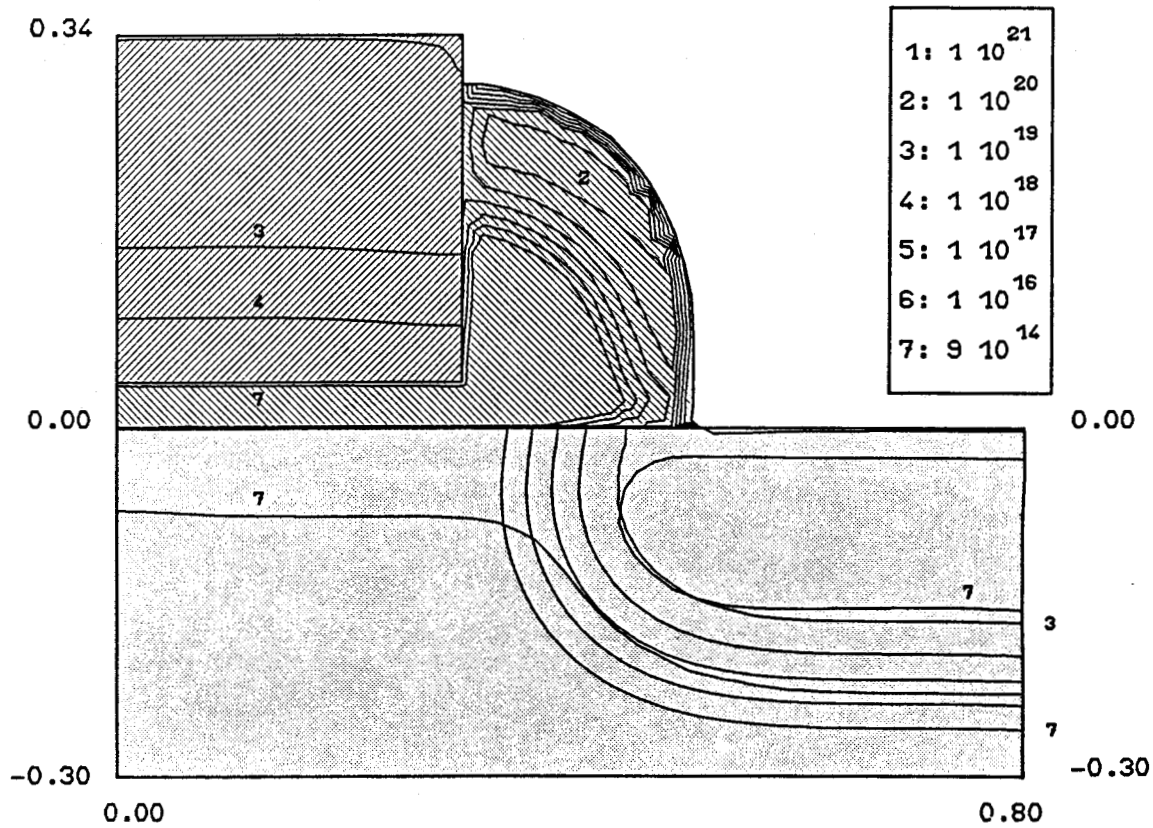


Fig. 36 Arsenic contours in an LDD-MOS transistor after annealing of an advanced bipolar transistor [14].

The last example displays the boron contours at the end of the simulation of a self-aligned polysilicon emitter bipolar transistor (Fig. 37). The extrinsic base is formed by diffusion from doped polysilicon, whereas the intrinsic base has been introduced by  $\text{BF}_2$  ion implantation [14,73]. This kind of simulation shows one of the most complex cases, containing 5 layers, several segregation conditions and a large number of unknowns (4015 nodes and 7402 elements). Other applications will be detailed in chapter V.

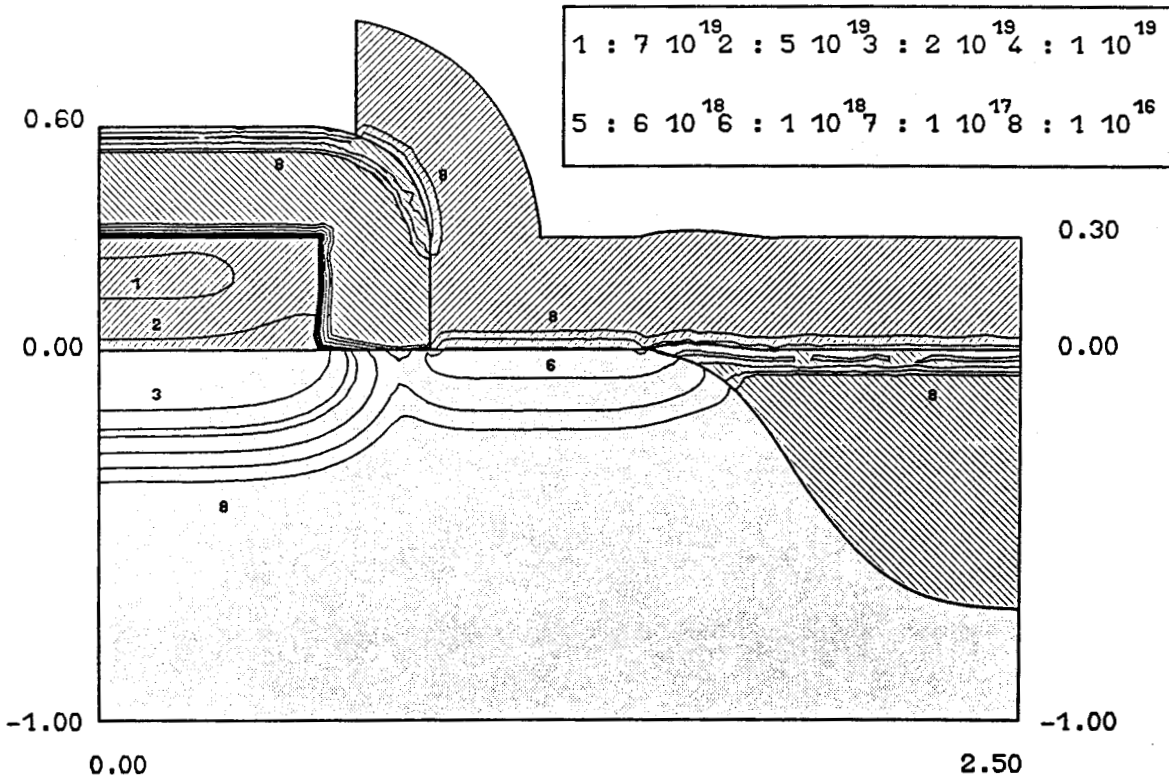


Fig. 37 Boron contours at the end of the complete simulation of an advanced bipolar transistor [14].

## IV. OXIDATION

### 4.1 Introduction

Thermally grown layers of silicon dioxide ( $\text{SiO}_2$ ) have been an important part of silicon technology since their first use in selective masking for diffusion, in passivation and in stabilization of silicon surfaces. They are also extensively used now for gate oxidation in MOS technology and for device isolation (as example in the LOCOS process).

In this introduction, the modeling and simulation points of view are discussed. The second section deals with the resolution of the 2D problem and the associated numerical procedures. Finally section 4.3 describes the algorithms involved in the dopant redistribution at interfaces during oxidation.

#### 4.1.1 The modeling point of view

##### 4.1.1.1 one-dimensional modeling

In response to the need for a predictive model (oxide thickness as a function of processing conditions), Deal and Grove developed their famous model in 1965 [74]. Despite the numerous discussions about some of its physical basis, this model has proved its capability in explaining or "fitting" experimental results. The model involves three processes: i) transport of oxidizing species from the ambient to the  $\text{SiO}_2$  surface, ii) diffusion through the oxide (Fick's laws) and iii) reaction with silicon at the  $\text{Si/SiO}_2$  interface. Assuming that the corresponding three fluxes are equal (static equilibrium is achieved), a linear-parabolic relation is derived:

$$\frac{X_0}{A/2} = \left( 1 + \frac{t+\tau}{A^2/4B} \right)^{1/2} - 1 \quad (3.104)$$

where :  $X_0$  is the the oxide thickness ( $\mu\text{m}$ )  
 :  $B/A$  is the linear rate constant ( $\mu\text{m}/\text{min}$ )  
 :  $B$  is the parabolic rate constant ( $\mu\text{m}^2/\text{min}$ )  
 :  $\tau$  is an "initial" time or a time offset (min), corresponding to an initial oxide thickness  $X_1$ :

$$\tau = \frac{X_1^2 + A.X_1}{B} \quad (3.105)$$

A and B depend on physico-chemical parameters:

$$A = 2.D_{\text{eff}} \cdot \left( \frac{1}{k} + \frac{1}{h} \right) \approx \frac{2.D_{\text{eff}}}{k} \quad (3.106a)$$

$$B = 2.D_{\text{eff}} \cdot \frac{C^*}{N_1} \quad (3.106b)$$

where  $D_{\text{eff}}$  is the diffusion coefficient of oxidizing species in the oxide,  $k$  is the surface reaction rate coefficient,  $C^*$  is the equilibrium bulk oxidant concentration in  $\text{SiO}_2$  (at the gas/oxide interface) and  $N_1$  represents the number of oxidant molecules included into an unit volume of oxide. The parabolic coefficient  $B$  depends mainly on diffusion through the oxide, whereas the linear coefficient  $B/A$  reflects the oxidation reaction at the  $\text{Si}/\text{SiO}_2$  interface.

One of the main weaknesses of this model is the lack of precision in the initial thin regime for dry oxidation ( $X_0 < 200\text{\AA}$ ). As this regime is now widely involved in modern MOS technology, a more detailed comprehension is of prime importance. Although various models and modifications of the Deal-Grove model have been proposed in the past [75-78], a comprehensive modeling of this problem is still expected [79]. We make use in the program of the results of an extensive study undertaken by Massoud, Plummer and Irene [80,81]. *In situ* ellipsometry allowed the obtention of very accurate measurements in the thin regime and the extraction of parameters in this region. The growth rate is then expressed as:

$$\frac{dX_0}{dt} = \frac{B}{(2X_0 + A)} + C_1 \cdot \exp\left(-\frac{X_0}{L_1}\right) + C_2 \cdot \exp\left(-\frac{X_0}{L_2}\right) \quad (3.107)$$

where the characteristics lengths  $L_1$  and  $L_2$  have a typical value of  $10\text{\AA}$  and  $70\text{\AA}$ , respectively. The first exponential contribution is expected to be the reflect of wafer cleaning and surface preparation procedures. The second exponential term should represent an intermediate phase. It is

believed that these two terms arise from additional fluxes, in parallel with the usual three fluxes of the Deal-Grove model. The same kind of parallel mechanism has been also proposed in [82]. Expression (3.107) furnishes the basis for very precise oxide thicknesses for the calculation of dry oxidation in the thin regime.

#### 4.1.1.2 two-dimensional modeling

Modern technologies always involve two-dimensional oxidations, from the simplest local oxidation of silicon (LOCOS) to more sophisticated schemes: fully-recessed oxides (as represented after in Fig. 41), SILO (sealed interface local oxidation), SWAMI (sidewall mask isolation), and finally, trench structures. In the case of very simple geometries (cylinders...), it is possible to get an idea of how stress develops: i) on convex surfaces, the previously grown oxide layer is pushed outward by the newly formed oxide and hence becomes tangentially stretched, ii) on concave surfaces, the outer layer becomes tangentially compressed [83].

In the general case, the solution of this problem is more difficult and the only possibility is to solve the diffusion of oxidant species and the equations governing the oxide flow. This latter phenomenon may be modeled by the Navier-Stokes equation [84,65] or by a viscoelastic behaviour [85-86]. To some extent, it allows to explain experimental results such as those observed on trench structures [87].

#### 4.1.2 The simulation point of view

When taking into account the special characteristics of IMPACT4, we have to choose a numerical strategy for 2D oxidation simulation. In the past, analytical solutions have been widely used. Once experimental oxide profiles are available, fitting procedures [88-89] allow to reproduce a variety of isolation structures. In this case, erfc functions are often used, either for practical purposes or by noting that this type of function may be the solution of oxidant diffusion in the oxide (and assuming vertical displacements) [90].

This solution has the main advantage to reduce considerably the CPU time, but the final shapes must be known prior to the simulation, and it is nearly impossible to predict isolation profiles for new technologies.

The problem is more complex for multilayer simulations since arbitrarily shaped structures must be oxidized. Even in the case of a simple gate re-oxidation in MOS technology, it is very difficult to use analytical formula. For these reasons, we have decided to solve the diffusion of oxidant and the subsequent oxide flow by the resolution of Navier-Stockes equation. This seems to be a reasonable compromise between the general multilayer requirements, the physical concepts and the CPU demands. Using a viscoelastic modelisation would require too large CPU ressources and can be justified only in the case of a specialized simulator for 2D oxidation.

The viscous flow of oxide has been observed experimentally: the curvature of uniformly oxidized test wafers is proportional to the stress and it has been found that above a temperature of about 960°C, stress relax via viscous flow of oxide. This allowed to determine the variation of viscosity with temperature:

$$\mu(T) = \mu_0 \cdot \exp\left(\frac{E_\mu}{kT}\right) \quad (3.108)$$

with  $\mu_0 = 1.586 \cdot 10^{10}$  poises and  $E_\mu = 5.761$  eV [91]. However, we are aware of the domain of validity of such modeling. By using a simplified viscoelastic model, it is expected that the stress relax exponentially with time, from an initial  $\sigma_0$  value:

$$\sigma = \sigma_0 \cdot \exp\left(-\frac{t}{\tau}\right) \quad \text{and} \quad \tau = \frac{\mu}{E} \quad (3.109)$$

where  $E$  is Young's modulus for oxide ( $1.87 \cdot 10^{12}$  dyne/cm<sup>2</sup>). It is of major interest to plot the variation of the relaxation time  $\tau$  versus temperature (Fig. 38). On this figure is also indicated the time necessary to grow a given oxide thickness ( $t_0$ ). As  $\tau$  is greater than  $t_0$  under 960°C, stress are very important in this temperature region and hence a viscoelastic modeling would be more appropriate. On the other hand, it is possible to improve the quality of the viscous flow model by introducing non linear effects (influence of stress), as proposed in recent investigations [92]. Since such modeling of these complex phenomena requires a large amount of work, we consider that the presented resolution, based on viscous creeping flow (with constant parameters) is the necessary step for more elaborate models in the future.

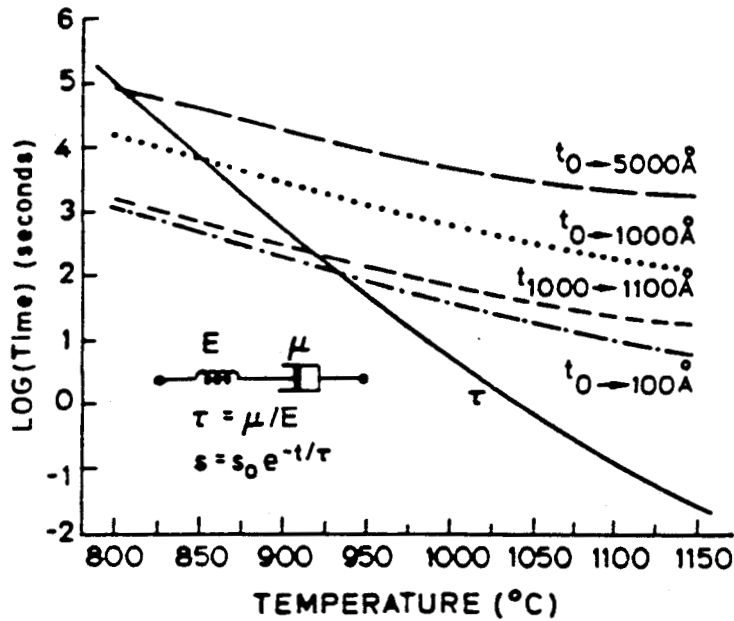


Fig. 38 Stress relaxation time and oxidation times as a function of temperature, with the grown oxide thickness as a reference [90].

#### 4.2 Two-dimensional oxidation simulation

As discussed previously, the oxidation rate at the Si/SiO<sub>2</sub> interface is determined by the available oxidant concentration. Then, a volume expansion occurs at this interface and the newly formed oxide pushes the old oxide. This oxide flow is modeled by the equations of an incompressible viscous fluid motion.

In order to clarify the organization of this section, the following flowchart describes the different steps of two-dimensional oxidation simulation.



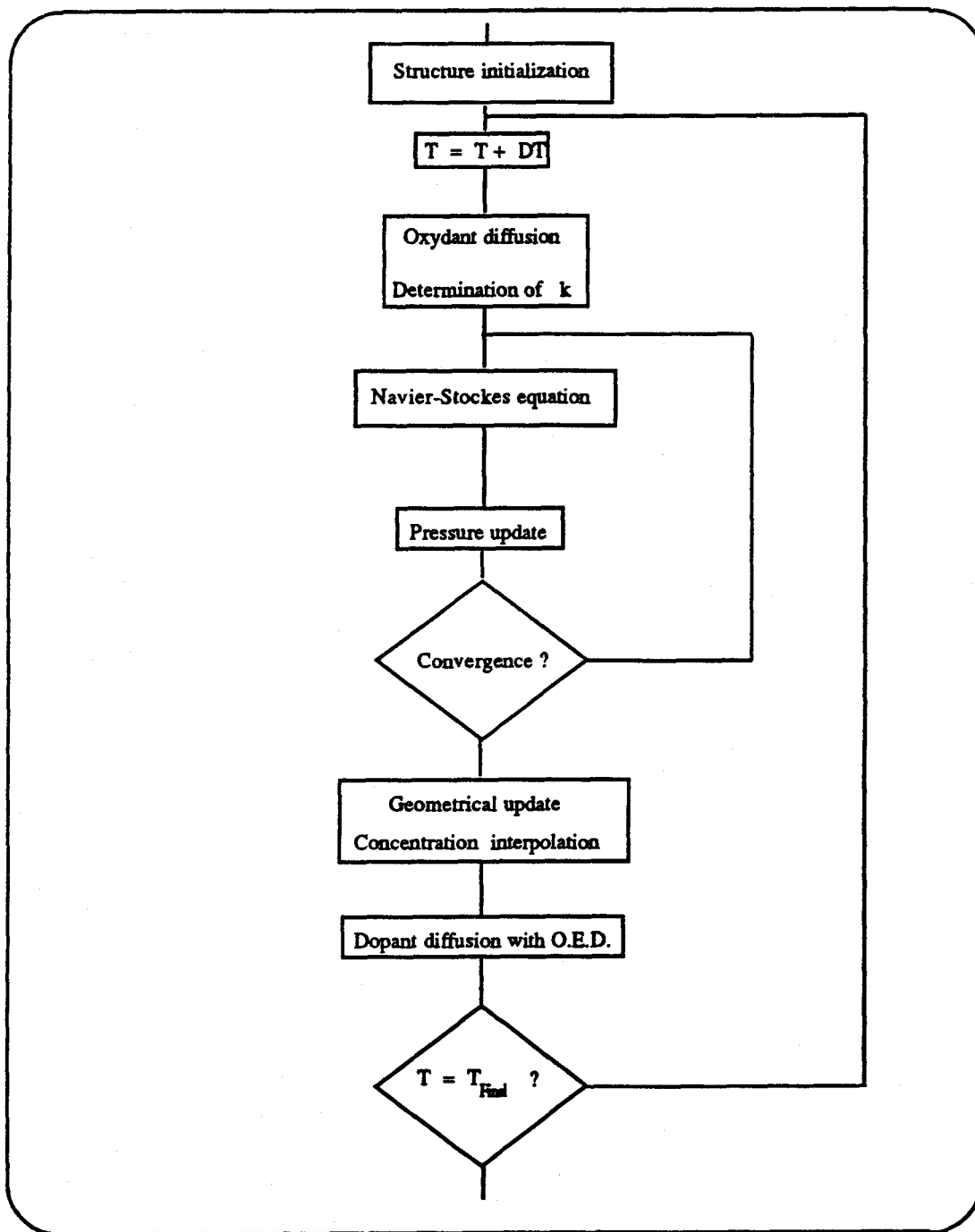


Fig. 39 Flowchart of the two-dimensional oxidation simulation.

#### 4.2.1 Initialization of the structure

It is interesting to explain rapidly how complex structures including nitride, polysilicon and oxide layers are initialized. For this purpose let us consider the structure represented in Fig. 40. The method is as follows:

- i) The nitride layer (here on top of the structure) is "virtually removed" from the data base.
- ii) A very thin oxide layer (5 Å) is deposited on the two polysilicon layers only (on the vertical edges at  $x=0.5 \mu\text{m}$ ). This means that no oxide is deposited in the regions where an oxide layer is in contact with air. On the other hand, this very thin oxide has no influence on the final oxide thicknesses.
- iii) This results in a single oxide layer, subsequently meshed and ready for the oxidation calculations.
- iv) The nitride layer is re-activated in the data base.

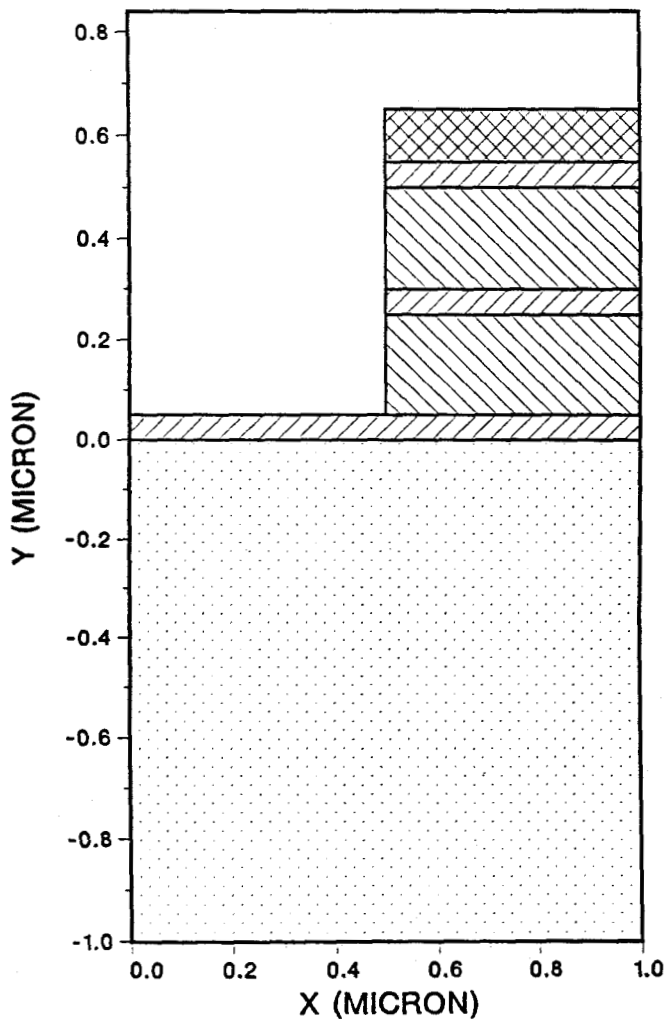


Fig. 40 Initial structure containing:

- a top nitride layer,
- 2 polysilicon layers,
- 3 oxide layers,
- the silicon substrate.

## 4.2.2 Oxidant diffusion

### 4.2.2.1 equation and the boundary conditions

Assuming the diffusion coefficient of oxidants to be independent of oxidant concentration (and in a more approximate manner of stress), the oxidant diffusion is simply expressed as:

$$D_{eff} \cdot \left( \frac{\partial^2 C}{\partial x^2} + \frac{\partial^2 C}{\partial y^2} \right) = \frac{\partial C}{\partial t} \quad (3.110)$$

where  $D_{eff}$  is the diffusion coefficient of oxidants. A study of the different terms of (3.110) shows that the temporal term can be neglected and thus (3.110) reduces to the Laplace equation [90]. The boundary conditions read:

$$\vec{F} \cdot \vec{n} = k \cdot C(x,y) \quad \text{at the Si/SiO}_2 \text{ interface} \quad (3.111a)$$

$$\vec{F} \cdot \vec{n} = -h \cdot (C^* - C(x,y)) \quad \text{at the gas/SiO}_2 \text{ interface} \quad (3.111b)$$

$$\vec{F} \cdot \vec{n} = 0 \quad \text{on the left and right sides} \quad (3.111c)$$

where  $k$  is the surface reaction coefficient and  $\vec{n}$  the normal unit vector.

### 4.2.2.2 Calculation of the surface reaction coefficient

It is important to describe precisely the value of the parameter  $k$  as it determines the local growth rates. First, it is calculated from the linear constant:

$$k = \frac{B}{A} \cdot \frac{N_1}{C^*} \quad (3.112)$$

In order to include the modeling of oxide growth in the thin regime [80,81], the linear constant is expressed in this case as:

$$\frac{B}{A} = \left( \frac{B}{A} \right)_{DG} + C_1 \cdot \exp\left(\frac{-X_0}{L_1}\right) + C_2 \cdot \exp\left(\frac{-X_0}{L_2}\right) \quad (3.113)$$

where  $\left(\frac{B}{A}\right)_{DG}$  is the value of the linear constant, deduced from the Deal and Grove modeling. This calculation allows the same accuracy in 2D oxidation simulations in dry ambient, than in simple 1D test examples.

The effect of substrate orientation is also accounted for [90]:

$$k(\vec{n}) = \left( k_{100} + \Delta \cdot \frac{k_{111} - k_{100}}{54.7} \right) \quad 0 \leq \Delta \leq 54.7 \quad (3.114a)$$

$$k(\vec{n}) = \left( k_{100} + \Delta \cdot \frac{k_{110} - k_{100}}{90} \right) \quad 54.7 \leq \Delta \leq 90 \quad (3.114b)$$

where  $\Delta$  is the angle between  $\vec{n}$  and the (100) direction. Finally, the effect of high substrate doping levels is also included. Following the works of Ho and Plummer [93-94], two multiplying factors are defined for the oxidation rate constants:

$$1 + \gamma \left( V_N^T - 1 \right) \quad \text{for } B/A \quad (3.115)$$

$$1 + \delta (C_T)^q \quad \text{for } B$$

where :  $\gamma$  ,  $\delta$  and  $q$  are determined experimentally.

$V_N^T$  is the normalized total vacancy concentration.

$C_T$  is the total vacancy concentration.

This latter effect becomes important for high doping levels (approximately above  $10^{19}$  at/cm<sup>3</sup> in phosphorus doped silicon, for example). The first multiplying factor is directly introduced in (3.112) and the second in the determination of  $D_{eff}$ .

Equation (3.110) and the related boundary conditions are solved using the numerical techniques proposed for the dopant diffusion, neglecting the temporal dependent term. Again we use P1 triangular elements for this purpose. As an example, Fig. 41 shows the contours of oxidant in a fully recessed oxide isolation structure, during wet oxidation. The masking effect of the nitride layer on the left side is obvious, as well as the consumption of oxidants at the Si/SiO<sub>2</sub> interface.

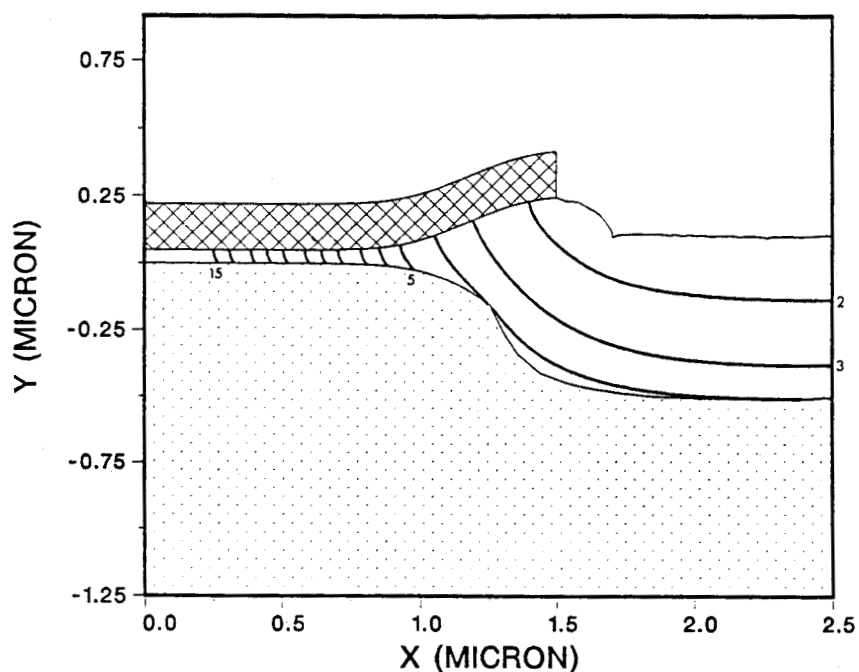


Fig. 41 Oxidant contours in a fully recessed oxide during wet oxidation. (1000°C). 1:  $5 \cdot 10^{19}$ , 2:  $2 \cdot 10^{19}$ , 3:  $10^{19}$ , 4:  $5 \cdot 10^{18}$  .....  
15 :  $10^{15}$  at/cm<sup>3</sup>.

#### 4.2.3 Resolution of the viscous fluid flow

##### 4.2.3.1 Formulation

The equations used in the viscous flow are in fact a simplified formulation of Navier-Stokes equations. These simplifications are discussed in details in [90] and are now widely used. Hence, it can be shown that the pressure gradient is balanced by the viscous deceleration, that is:

$$\mu \nabla^2 v_x = \frac{\partial P}{\partial x} \quad (3.116a)$$

$$\mu \nabla^2 v_y = \frac{\partial P}{\partial y} \quad (3.116b)$$

where  $V_x$  and  $V_y$  are the two components of velocity and  $P$  is the hydrodynamic pressure. Moreover, as compressibility of oxide is very low ( $2.7 \cdot 10^{-12}$  cm<sup>2</sup>/dyne), we assume this material to be incompressible. This is expressed simply as:

$$\rho \cdot \text{div } \vec{V} = - \frac{\partial \rho}{\partial t} \approx 0 \quad (3.117)$$

where  $\rho$  is the oxide density. The boundary conditions are as follows:

$$\vec{V} = -(1-\alpha) \cdot \frac{k \cdot C}{N_1} \cdot \vec{n} \quad \text{at the Si and poly / SiO}_2 \text{ interfaces} \quad (3.118a)$$

$$P = P_a - \frac{\gamma}{R} \quad \text{at the gas/SiO}_2 \text{ interface} \quad (3.118b)$$

where :  $\alpha$  is the ratio of consumed silicon to grown oxide (0.44)

$P_a$  is the ambient pressure

$\gamma$  is the surface tension coefficient ( $10^3$  dyne/cm)

$R(x,y)$  is the local curvature

Two conditions are applied at the SiO<sub>2</sub>/ silicon nitride interface for pressure and velocity:

$$P = P_a - \frac{\gamma}{R} \quad (3.119a)$$

$$\vec{V} \cdot \vec{l} = 0 \quad (3.119b)$$

where  $\vec{l}$  is the unit vector lateral to this interface. This latter relation reflects a strong bonding between the two materials. On the other hand, it should be noted that the effects of stress induced by the nitride bending are not taken into account.

#### 4.2.3.2 Iterative solution and resolution by FEM

One of the main difficulty arises from the incompressibility condition (3.117). Following the approach of Chin, velocity/pressure iterations are used (artificial-compressibility algorithm proposed by Chorin [95]). Combining the mass-continuity equation and the fact that the density of a compressible material changes with pressure, allows to express an iterative relation between pressure and velocity:

$$p^{n+1} = P^n - \omega \cdot \mu \cdot \text{div}(\vec{V}^n) \quad (3.120)$$

where  $\omega$  is a relaxation parameter optimised upon numerical requirements ( $\omega=0.6$ ) and  $n$  the iteration number. With this notation, (3.116) is rewritten as:

$$\mu \nabla^2 V_x^{n+1} = \frac{\partial P^{n+1}}{\partial x} \quad (3.121a)$$

$$\mu \nabla^2 V_y^{n+1} = \frac{\partial P^{n+1}}{\partial y} \quad (3.121b)$$

Convergence is achieved when the divergence of velocity is sufficiently small at any point of the mesh, that is for example:

$$\frac{|V_x^{n+1} - V_x^n|}{\max(V_x^n, V_y^n)} \leq \varepsilon \quad \text{and} \quad \frac{|V_y^{n+1} - V_y^n|}{\max(V_x^n, V_y^n)} \leq \varepsilon \quad (3.122)$$

with  $\varepsilon=10^{-2}$ . The set of equations (3.121) is then discretized with FEM method, in a similar way than the diffusion equation. Concerning the mesh we use again P1 triangular elements. Using P1 or P2 elements has a negligible influence on the final shapes, but 3-noded triangles allow a significant reduction in computer resources. However, in the future extensions, the accuracy on the calculation of stress will be critical and some formulations will have to be investigated. In the following we present briefly the resulting equation for (3.121a). For the sake of simplicity, we can introduce the expressions (3.123):

$$F_{xx} = \mu \cdot \frac{\partial V_x}{\partial x} \quad \text{and} \quad F_{xy} = \mu \cdot \frac{\partial V_x}{\partial y} \quad (3.123)$$

The Galerkin method applied to (3.121a) yields for the  $j$ th equation:

$$\begin{aligned} \sum_{elj} \int_{\Omega_l} (N_{j1x} \cdot F_{xx} + N_{j1y} \cdot F_{xy}) \cdot d\Omega_l - \sum_{elj} \int_{\Gamma_l} N_{j1} \cdot F \cdot d\Gamma_l \\ + \sum_{elj} \left( \frac{\partial P}{\partial x} \right)_1 \cdot \int_{\Omega_l} N_{j1} \cdot d\Omega_l = 0 \end{aligned} \quad (3.124)$$

It is obvious that we take here advantage of the work performed on the discretisation of the diffusion equation. Hence, the same notations are used as for equation (3.81). It should be noted that the pressure gradient is constant over each element, as we use P1 elements, allowing to extract the pressure derivative term out of the integral over each element.

In order to minimise the number of assembly operation, (3.121a) and (3.121b) are gathered in the same matrix, solved using the conjugate gradient method with preconditioning, as presented in section 3.4.3. One particular point results from the boundary conditions at the silicon nitride/silicon oxide interface. If we neglect the non-slip boundary condition (3.119b), equation (3.121a) and (3.121b) are not coupled. But if we introduce this special condition, the two equations are now coupled and the matrix becomes asymmetric. Effectively, as the two vectors  $\vec{V}$  and  $\vec{n}$  are now colinear:

$$V_x \cdot n_y - V_y \cdot n_x = 0 \quad (3.125)$$

This means that the the two fluxes  $F_x$  and  $F_y$  are imposed for the two components  $V_x$  and  $V_y$ , respectively:

$$F_x = h \cdot (V_y \cdot n_x - V_x \cdot n_y) \quad \text{and} \quad F_y = h \cdot (V_x \cdot n_y - V_y \cdot n_x) \quad (3.126)$$

The introduction of (3.126) in the system reflects the coupling between the two equations (3.121a) and (3.121b). The symmetry of the matrix can be recovered by multiplying the terms  $F_x$  and  $F_y$  by  $n_y$  and  $n_x$ , respectively [96].

#### 4.2.4 Geometrical updates

At each time step, after the resolution of the viscous fluid flow, the new layers boundaries are re-calculated. At the Silicon/oxide and polysilicon/oxide interfaces, the new locations of each point are calculated as follows:



$$\begin{aligned}x_{\text{new}} &= x_{\text{old}} - V_x \cdot \Delta t \cdot \frac{\alpha}{1-\alpha} \\y_{\text{new}} &= y_{\text{old}} - V_y \cdot \Delta t \cdot \frac{\alpha}{1-\alpha}\end{aligned}\tag{3.127}$$

where  $\Delta t$  is the value of the time step. For the nodes on the boundary with air, it is simply expressed as:

$$\begin{aligned}x_{\text{new}} &= x_{\text{old}} + V_x \cdot \Delta t \\y_{\text{new}} &= y_{\text{old}} + V_y \cdot \Delta t\end{aligned}\tag{3.128}$$

Once these new coordinates are calculated, it is possible to redefine the boundary of the oxide domain. A regular grid of point (as presented in chapter II, 3.3) is then sur-imposed on this layer, to define a new triangulation. Concerning the other layers, the new boundaries are deduced from the interfaces with the oxide layer and, in this case, the interior nodes of the old meshes can be used for the new triangulation, as the new boundaries are always included in the old silicon or polysilicon layers.

It should be noted that several solutions for the mesh re-generation in the oxide and silicon have been formulated in FEDSS [65] or SUPREM IV [97], in order to avoid a complete meshing procedure at each time step. Although these solutions offer significant reduction in CPU time, they are very difficult to manage [97] and moreover, it seems not realistic to generalize it to our general multilayer case.

#### 4.2.5 Examples

On the initial structure of Fig. 40, a small oxidation step is now performed. Fig. 42 depicts the final structure after this wet oxidation, in which the effects of the incompressible viscous flow modeling can be seen in the two regions specified by the arrows. Each layer has been updated with the strategy described above.

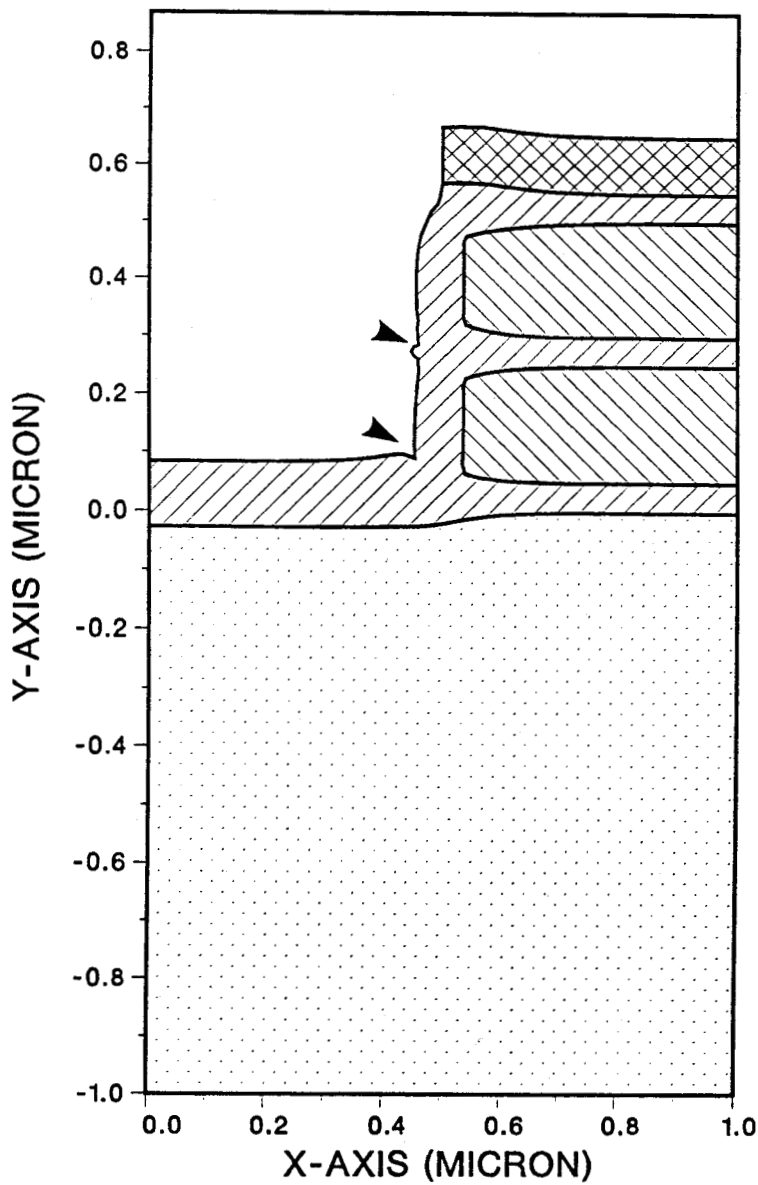


Fig. 42 Two-dimensional structure after wet oxidation (1000°C, 5 min).

An another complex test is represented in Fig. 43 and 44. Here the initial structure is a trench filled with polysilicon (Fig. 43). After patterning of a silicon nitride mask, a wet oxidation step is performed in order to obtain a smooth transition between the active part (under the nitride mask) and the trench isolation area, as shown in Fig. 44. Fig. 45 displays the evolution of each layer during this oxidation, showing the formation of the Bird's Beak and the "broadening" of the oxide layer at the junction between the trench and the thick oxide. This kind of technique is often used in advanced bipolar technologies. During these simulations, it is assumed that the nitride material can not be oxidized and that the growth rate of oxide on the polysilicon layer is the average value of growth rates between (100) and (111) directions in silicon.

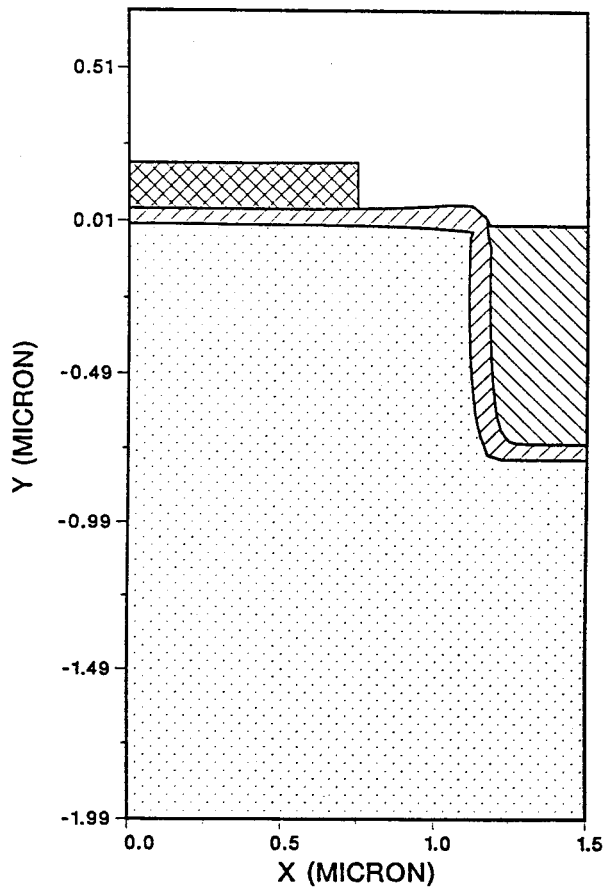


Fig. 43 Initial structure before oxidation. A silicon trench has been formed and oxidized, filled with polysilicon. A nitride masking layer protects the active area on the left part of the structure.

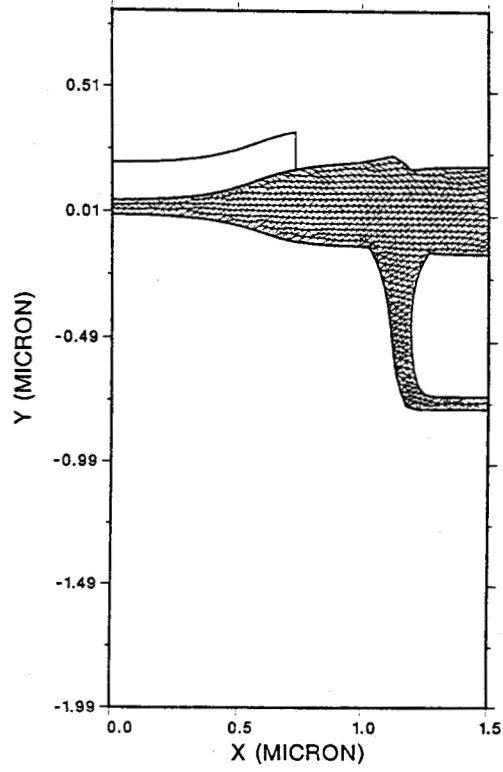


Fig. 44 Resulting structure after wet oxidation (950°C, 70 min) using 20 time steps. The final mesh is also plotted in the oxide domain.

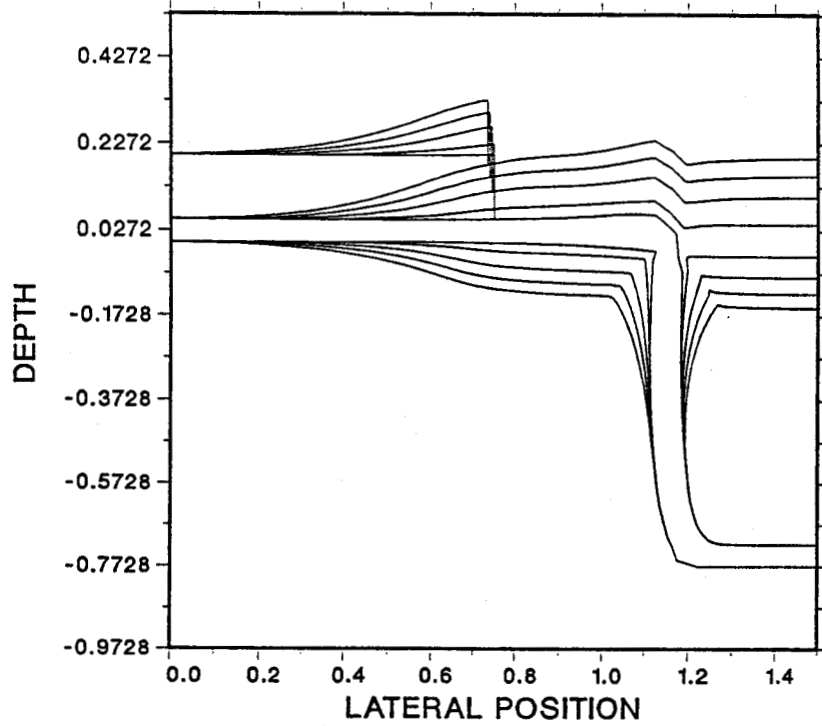


Fig. 45 Evolution of the oxide and nitride domains during this simulation.

### 4.3 Coupling between oxidation and diffusion

The resolutions and results presented in the preceding sections have been implemented in several dedicated simulators, where only the behaviour of oxide and nitride layers is studied. Less work has been undertaken concerning the coupling with dopant redistribution in silicon and polysilicon layers during oxidation. In this section, we detail some of the numerical algorithms aimed to solve this complex problem.

Analytical resolutions have been proposed in earlier investigations [98], but it is obvious that numerical procedures must be used in our general case. Several schemes have been reported in the past [99,65], in the case where dopant diffusion is solved on both sides of the interface (it is possible to solve this problem in a simpler manner when only the silicon domain is meshed, by using appropriate boundary conditions [55]). In this work, we have develop original procedures, that ensure:

- i) Dopant dose conservation during the oxidation and minimization of interpolations errors.
- ii) A correct description of the segregation phenomenon.
- iii) No numerical instabilities.
- iv) Compatibility with the temporal scheme used during the dopant diffusion.

In the following, we will describe the interpolation schemes, the additional fluxes incorporated in the diffusion equation, and finally an evaluation of the proposed method.

#### 4.3.1 Interpolations schemes for the oxide domain

In order to illustrate our problem, let us consider a simple 1D case (Fig. 46). After a time step, new oxide has been formed at the Si/SiO<sub>2</sub> interface (region (1) ), the old oxide is pushed and possibly deformed during the viscous flow -for 2D structures- (region (2) ), also some silicon has been consumed (region (3) ), as depicted in Fig. 46b.

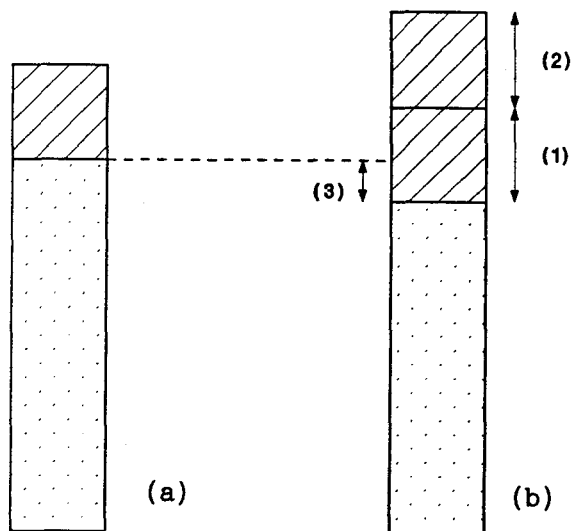


Fig. 46 Schematic 1D illustration of the oxidation of the initial silicon/oxide structure (a) to the resulting one (b).

The dopant concentrations are interpolated from the old meshes to the new ones in the silicon and polysilicon layers, since these new domains are always included in the initial ones. It is thus performed in a simple manner, and it has been found that in this case a linear-logarithmic interpolation gives the best results.

The situation is far more complex for the new oxide domain. Referring to Fig. 46b, interpolations are performed in three steps:

- i) interpolations on the old oxide that has been displaced (region 2).
- ii) interpolations on the initial silicon (and polysilicon) layers.
- iii) determination of dopant concentrations in the intermediate band (difference between regions (1) and (3)).

For the first step, there is no particular problem, since the new positions of the old oxide displaced are extracted from the solution of Navier-Stokes equation.

The choices in the second step are crucial, since it determines the numerical stability of the overall procedure. After numerous experiments, it has been found that the best way is to separate the nodes that are on the boundary with silicon and the other internal nodes. For the first group of nodes, the dopant concentration is given by:

$$C_{Ox} = \frac{C_{Si}}{m_{Si-Ox}} \quad (3.129)$$

where  $C_{Si}$  is the interpolated value on the old silicon mesh,  $m_{Si-Ox}$  is the segregation coefficient of dopants (different for each specy) and  $C_{Ox}$  is the new dopant concentration in the oxide. For the second group of nodes (internal nodes),  $C_{Ox}$  is determined from:

$$C_{Ox} = \alpha \cdot C_{Si} \quad (3.130)$$

The two choices (3.129) and (3.130) have been determined after numerical tests, but they can also be justified in the sense that (3.129) is a good initial condition for the subsequent diffusion resolution, and (3.130) minimizes the difference of total dose in the complete structure before oxidation, and after this interpolation procedure.

Concerning the third step (difference between regions (1) and (3)), it is more difficult to find reasonable values for dopant concentrations. In this case, for each point A belonging to this region, the dopant concentration is the one of a point already interpolated during the second step, and situated at the minimal distance of this point A.

### 4.3.2 Boundary fluxes

Using the interpolations procedures described above, it is obvious that the dopant dose is often not the same before oxidation and after these interpolations. For each impurity  $i$ , we calculate this difference of dose:

$$\Delta Q^i = Q_e^i - Q_s^i \quad (3.131)$$

where  $Q_e^i$  and  $Q_s^i$  are the total dose (in all of the layers) after interpolation and before oxidation, respectively. The idea is then to introduce a new boundary flux that can be expressed as [100]:

$$F_j^i = \frac{V_j \cdot C_j^i \cdot \Delta Q^i}{\sum_k V_k \cdot C_k^i \cdot L_k} \cdot \frac{1}{\Delta t} \quad (3.132)$$

where :  $j$  is the oxide node number

$k$  is an index for all the oxide nodes (in boundary with silicon or polysilicon layers).

$V_j$  is the oxide velocity (determined from oxidant diffusion see equation (3.118a)).

$C_j^i$  is the concentration of node  $j$  for impurity  $i$ .

$L_j$  is the length of the boundary at the node  $j$  (sum of the mid-distances of point  $j$  to the 2 neighbor boundary points of the oxide mesh).

This is in fact an additional flux with the static segregation condition. In the following, this method will be evaluated and comparisons with other schemes will be presented.



### 4.3.3 Results

This method has been tested in a large number of cases, including variations on the oxidation rate, the type of impurity (and hence the value of the segregation coefficient, greater than 1 for n-type and less than 1 for boron).

As an example, phosphorus is implanted in silicon at an energy of 120 keV and a dose of  $2 \cdot 10^{12}$  at/cm<sup>2</sup>. An oxidation step is then simulated, for 30 min at 1000°C. Fig. 47a and 47b show the evolution of the phosphorus surface concentration and dose in silicon, respectively. It can be seen that a very high precision is achieved when the number of time steps is approximately greater than 30. Fig. 48 provides a comparison between our result with 100 time steps and a calculation issued from SUPREM II, the same parameters being used in the two simulations. The excellent agreement validates our approach.

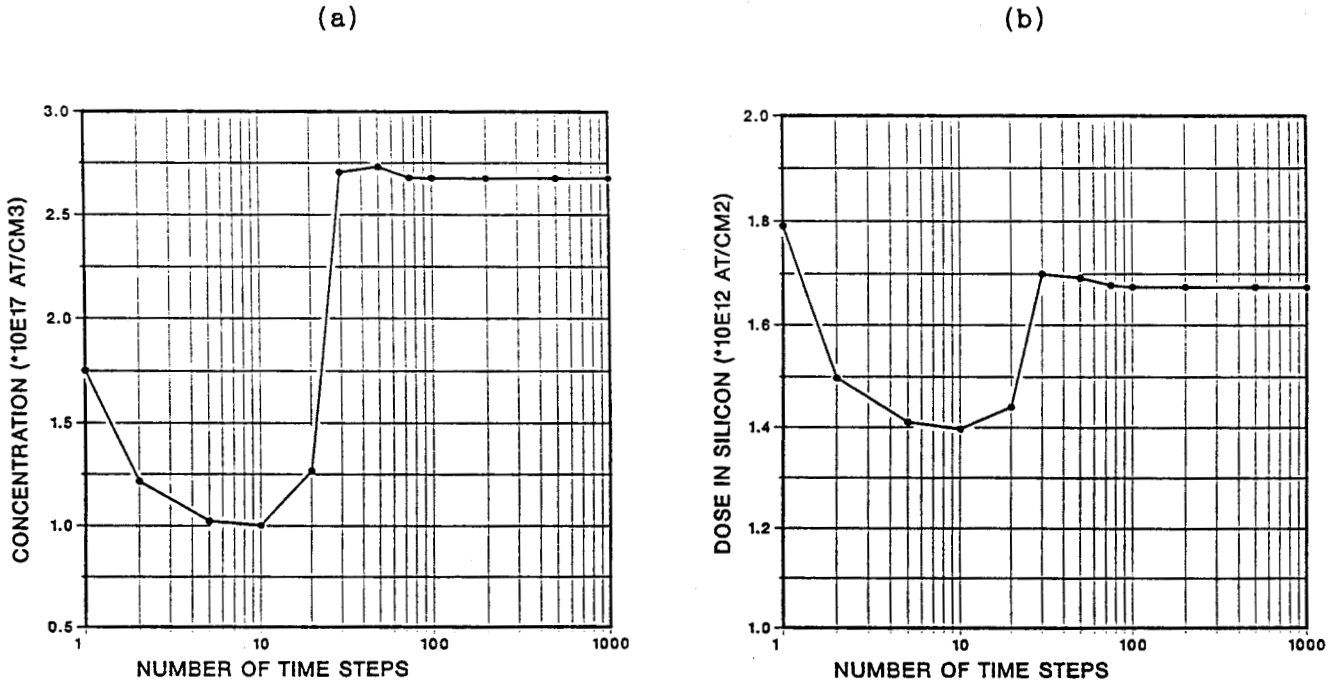


Fig. 47 Phosphorus surface concentration in silicon (a), and resulting dose in silicon (b) after ion implantation and oxidation

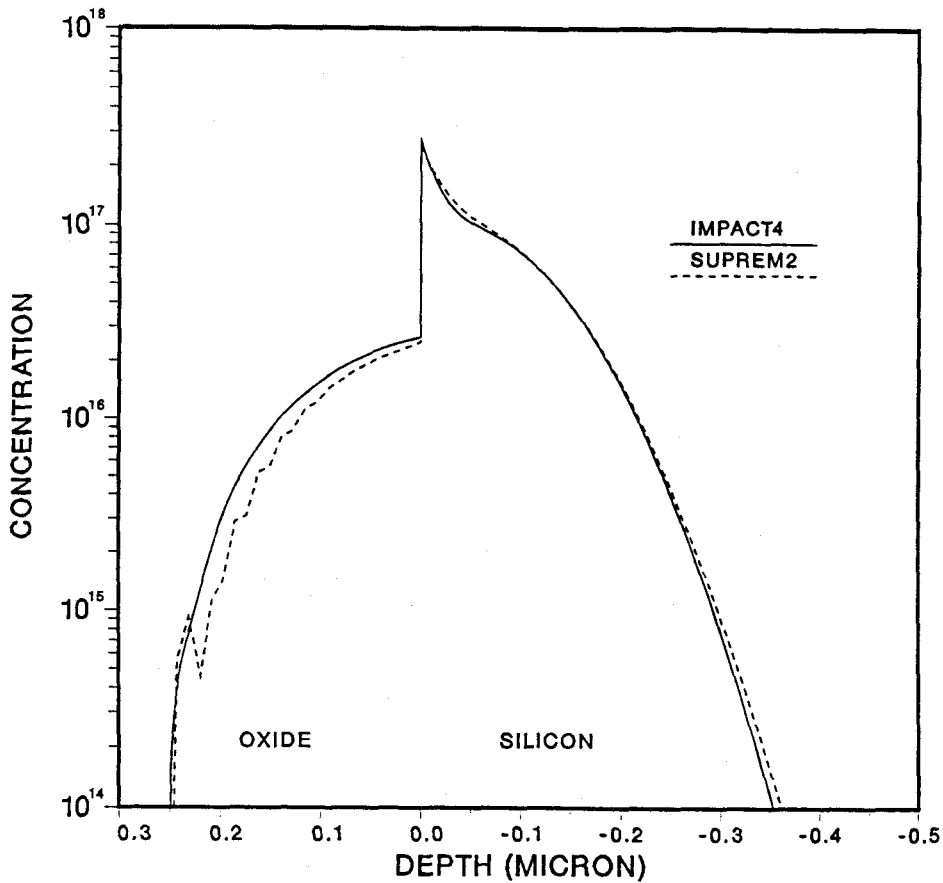


Fig. 48 Comparison between the results of SUPREM II and IMPACT 4 for the 1D phosphorus doped silicon oxidation test example.

When considering the interpolations schemes presented above, it is obvious that, if the value of the time step is sufficiently small, the new incremental oxide thickness will be reduced and consequently there will be no place enough to introduce new nodes between i) the new boundary of the oxide layer with silicon, and ii) the old oxide displaced. This means that above a given value, the evolution of the profile should be nearly "insensible" to the value of the time step since the interpolation errors are minimized: the band of oxide nodes interpolated with  $\alpha C_{s1}$  has not to be considered (the value of each time step is calculated so that an equal amount of oxide is grown during each of this time duration). This phenomenon is observed in Fig. 47.

In fact, it is possible to determine the needed relation between the number of time steps and the mesh spacing, in order to ensure that these values correspond to the "insensibility" level described qualitatively above (30 time steps and 50 Å respectively in the example of Fig. 47-48). Following the meshing procedure proposed in chapter II, we know that regular grids are used for mesh generation in the oxide layers, whose spacing is the mean value of boundary nodes spacing. Assuming that an oxide thickness  $T_{ox}$  is grown in  $NT$  time steps, and that an equal incremental oxide thickness is added during each time step, the insensibility level is approximatively reached if:

$$\frac{T_{ox}}{2.NT} \leq \Delta X \quad (3.133)$$

where  $\Delta X$  is the mean distance between nodes at the silicon-silicon oxide interface. In the examples presented in Fig. 47, about 2500 Å of oxide are grown and the value of  $\Delta X$  is 50 Å. By applying (3.133), we find that  $NT$  should be greater than 25, that is fully coherent with the value found in Fig. 47a and 47b.

It is of practical interest to plot the relation between  $NT$  and  $\Delta X$  for several oxide thicknesses (Fig. 49). It can be seen that even for large oxide thickness and reduced mesh spacing, the required number of time steps is always reasonable. As the usual mesh spacing has a value taken in the range 50-500 Å, the resulting number of time steps is coherent with the requirements imposed by the use of the incomplete implicit scheme. In other words, there is no additional constraints associated with the diffusion under oxidizing ambient.

Using again the previous example of phosphorus diffusion, Fig. 50 shows the evolution of the doping profiles for 1, 10, 30 and 100 time steps. We may remark that the scheme is always stable even when the "insensibility" level is not reached. On the other hand, it becomes difficult to differentiate the profiles corresponding to the numbers of time steps greater or equal than 30.

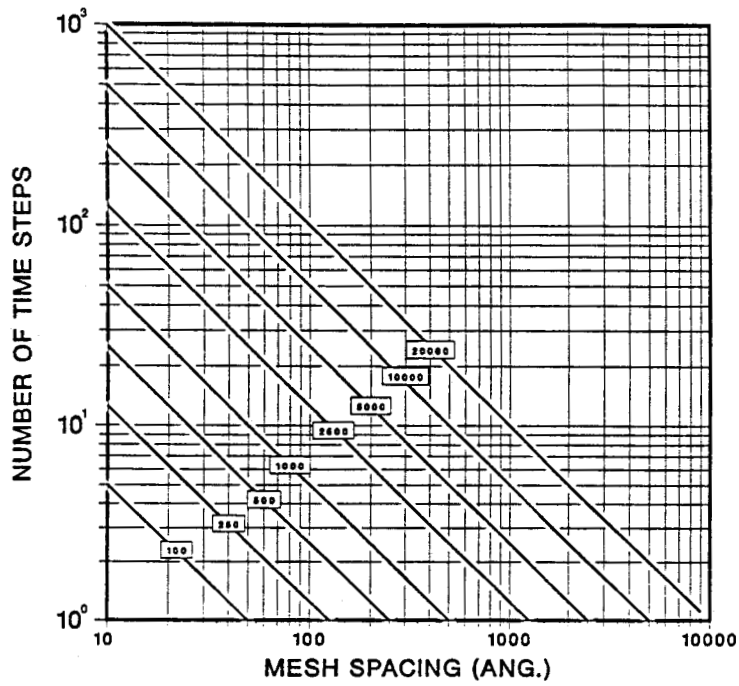


Fig. 49 Relation between the number of time steps and the mesh spacing for several oxide thicknesses ( $\text{\AA}$ ), in order to fullfill condition (3.133).

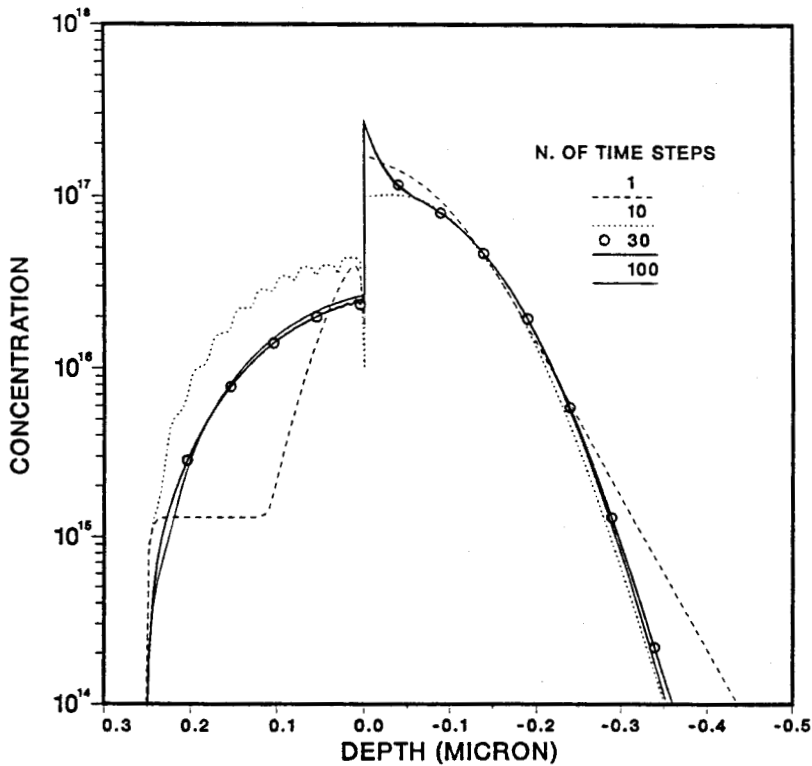


Fig. 50 Evolution of the phosphorus doping profile for different numbers of time steps.

Finally Fig. 51 displays the boron contours in a LOCOS structure at the end of a wet oxidation (at 1000°C for 180 min), in which 36 time steps and an average mesh spacing of 500 Å are used. This set of parameters satisfies the conditions plotted in Fig. 49. Effectively, a good description of the segregation phenomenon is obtained, as also confirmed by a cross-section under the thick oxide (Fig. 52).

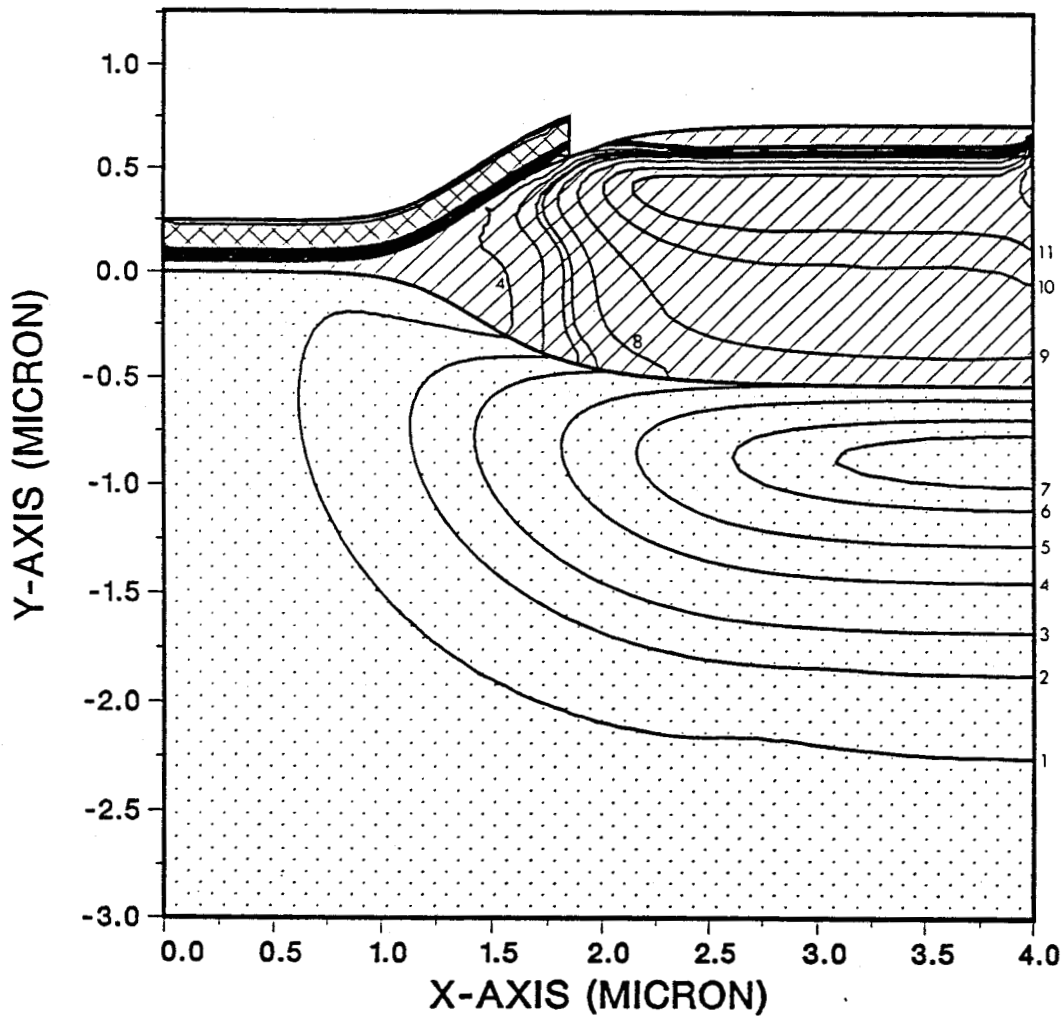


Fig. 51 Boron contours in a LOCOS structure after wet oxidation.  
1:  $1.1 \cdot 10^{15}$ , 2:  $1.5 \cdot 10^{15}$ , 3:  $2 \cdot 10^{15}$ , 4:  $3 \cdot 10^{15}$ , 5:  $4 \cdot 10^{15}$ ,  
6:  $5.0 \cdot 10^{15}$ , 7:  $6.0 \cdot 10^{15}$ , 8:  $8 \cdot 10^{15}$ , 9:  $1 \cdot 10^{16}$ , 10:  $1.5 \cdot 10^{16}$ ,  
and 11:  $2 \cdot 10^{16}$  at/cm<sup>3</sup>.

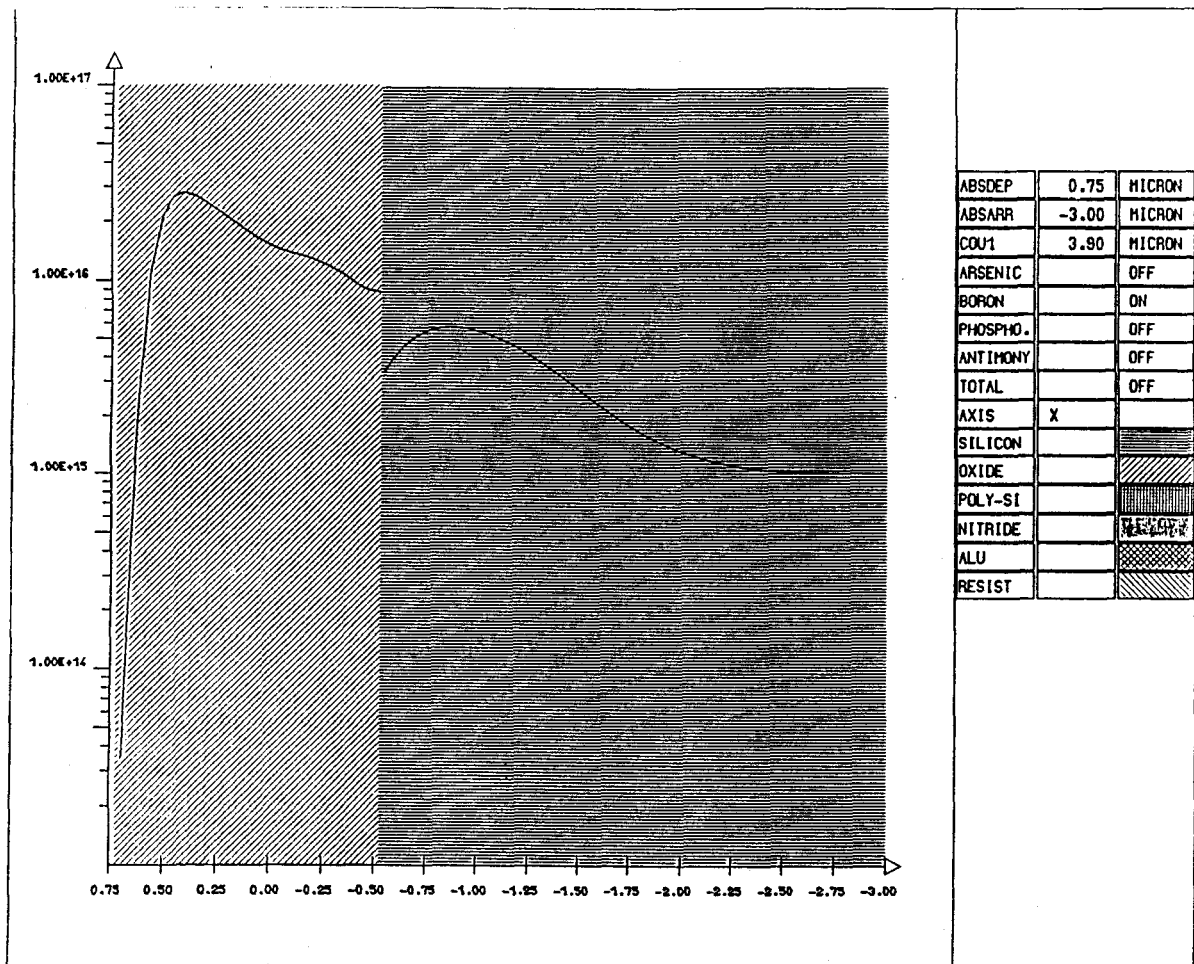


Fig. 52 Boron profile for a cross section in the thick oxide region.

In this latter simulation, we made use of an additional model, related to the coupled aspects of dopant diffusion and oxidation. This is the well-known Oxidation-Enhanced Diffusion effect (OED). During oxidation, interstitials are injected in the silicon bulk. This interstitial supersaturation enhances the diffusion of arsenic, boron and phosphorus. Following the macro-modeling approach for diffusion, we do not calculate the interstitial diffusion [47,68] but use an analytical formula [101-102]:

$$D_{ox} = f_I \cdot D_{int} \cdot \frac{D_{int}}{D_N} \cdot \frac{K(T)}{\varphi} \cdot V_{ox}^n \quad (3.134)$$

where :  $D_{ox}$  is an additive term to the diffusion coefficient.

$f_I$  is a multiplying factor (representative of the contribution of interstitials to the diffusion mechanism).

$D_{int}$  is the intrinsic diffusivity.

$D_N$  is the diffusivity in inert ambient.

$\varphi$  is a factor taking into account the substrate orientation (1 for (100) and 2.7 for (111) ).

$V_{ox}$  is the maximum oxidation velocity.

$K(T)$  is a multiplying factor calculated with an Arrhenius relation.

$n$  is an empirical factor.

The values of these parameters allow to describe with a reasonable accuracy the diffusion of dopants under oxidizing conditions.

REFERENCES :

- [1] W.G. Oldham, S. Nandgaonkar, A. Neureuther and M.O. Toole, "A general simulator for VLSI lithography and etching processes: Part I - Applications to projection lithography", IEEE Trans. Electron Devices, ED-26, 717-722, April 1979.
- [2] A.R. Neureuther, C.H. Ting and C.Y. Liu, "Application of line-edge profile simulation to thin film deposition processes", IEEE Trans. Electron Devices, ED-27, 1449-1455, August 1980.
- [3] W.G. Oldham, A.R. Neureuther, C. Sung, J.L. Reynolds and S.N. Nandgaonkar, "A general simulator for VLSI lithography and etching processes: Part II - Application to deposition and etching", IEEE Trans. Electron Devices, ED-27, 1455-1459, August 1980.
- [4] J. Lorenz, J. Pelka, H. Ryssel, A. Sachs, A. Seidl and M. Svoboda, "COMPOSITE - A complete modeling program of silicon technology", IEEE Trans. Electron Devices, ED-32, No 10, 1977-1986, October 1985.
- [5] M. Ohgo, Y. Takani, A. Moniwa, S. Yamamoto, Y. Sakai, H. Masuda and H. Sunami, "A two-dimensional integrated process simulator: SPIRIT-I", IEEE Trans. Computer Aided Design, CAD-6, 439-445, May 1987.
- [6] G.C. Schwartz, L.B. Rothman and T.J. Schopen, "Competitive mechanisms in reactive ion etching in a  $CF_4$  plasma", J. Electrochem. Soc., 126, 464-469, 1979.
- [7] S. Yamamoto, T. Kure, M. Ohgo, T. Matsuzama, S. Tachi and H. Sunami, "A two-dimensional etching profile simulator: ESPRIT", IEEE Trans. Computer Aided Design, CAD-6, 417-421, May 1987.
- [8] S. Martin, private communication.
- [9] T.J. Cotler, M.S. Barnes and M.E. Elta, "A Monte Carlo microtopography model for investigating plasma/reactive ion etch profile evolution", J. Vac. Sci. Technol., B6(2), 542-550, March/April 1988.
- [10] A. C. Adams, "Dielectric and polysilicon film deposition", in VLSI Technology, p. 111-113, Ed. by S.M. Sze, Mc Graw Hill, 1983.
- [11] I.A. Blech, "Evaporated film profiles over steps in substrates", Thin Solid Films, 6, 113-118, 1970.
- [12] I.A. Blech, D.B. Fraser and S.E. Haszko, "Optimization of Al step coverage through computer simulation and scanning electron microscopy", J. Vac. Sci. technology, 15(1), 13-19, January/February 1978.
- [13] R. Jewett, "A string model etching algorithm", Memorandum No UCB/ERL M79/68, University of California, Berkeley, 1979.



- [14] B. Baccus, D. Collard, E. Dubois and D. Morel, "Two dimensional process simulation of bipolar devices using a multilayer simulator: IMPACT4", BCTM conference, Minneapolis, p. 164-167, September 1988.
- [15] H. Ryssel and K. Hoffmann, "Ion Implantation", in Process and device simulation for MOS-VLSI circuits, Eds. P. Antognetti, D.A. Antoniadis, R.W. Dutton and W.G. Oldham, NATO advanced study institute, Sogesta - Urbino, Italy, July 12-23, 1982.
- [16] J.P. Biersack and L.G. Haggmark, "A Monte Carlo computer program for the transport of energetic ions in amorphous targets", Nucl. Inst. & Methods, 174, 257-267, 1980.
- [17] M.D. Giles, "Ion implantation calculations in two dimensions using the Boltzmann transport equation", IEEE Trans. Computer Aided Design, CAD-5, 679-684, 1986.
- [18] G. Hobler and S. Selberherr, "Monte Carlo simulation of ion implantation into two and three dimensional structures", IEEE Trans. Computer Aided Design, CAD-8, 450-459, May 1989.
- [19] K. Kato, "As ion implantation simulation for trench structures using a Monte Carlo method", IEEE Trans. Electron Devices, ED-35, 1820-1828, November 1988.
- [20] W.P. Petersen, W. Fichtner and E.H. Grosse, "Vectorized Monte Carlo calculation for the transport of ions in amorphous targets", IEEE Trans. Electron Devices, ED-30, 1011-1017, September 1983.
- [21] G. Hobler, E. Langer and S. Selberherr, "Two-dimensional modeling of ion implantation with spatial moments", Solid-State Electronics, Vol. 30, No 4, 445-455, 1987.
- [22] J. Albers, "Monte Carlo simulation of one and two dimensional particle and damage distributions for ion implanted dopants in silicon", IEEE Trans. Electron Devices, ED-32, 1930-1939, October 1985.
- [23] J. Lindhard, M. Scharff and E.H. Schiott, "Range concepts and heavy ion ranges", Mat. Fys. Medd. Dan. Vid. Selsk, 33, No 14, 1-42, 1963
- [24] J.F. Gibbons, W.S. Johnson and S.W. Mylroie, Projected range statistics, Strandsberg, Halstead press, 1975.
- [25] S. Selberherr, Analysis and simulation of semiconductor devices, p. 47-48, Springer Verlag, 1984.
- [26] J.F. Gibbons and S. Mylroie, "Estimation of impurity profiles in ion implanted amorphous targets using joined half Gaussian distributions", Appl. Phys. Lett., 22, No 11, 568-569, 1973.
- [27] W.K. Hofker, "Concentration profiles of boron implantation in amorphous and polycrystalline silicon", Philips Research Reports, 8, 41-57, 1975.

- [28] G.A.J. Amaratunga, K. Sabine and A.G.R. Evans, "The modeling of ion implantation in three layer structure using the method of dose matching", IEEE Trans. Electron Devices, ED-32, 1889-1890, September 1985.
- [29] H. Ryssel, "Implantation and diffusion models for process simulation", in VLSI process and device modeling, summer course 1983, ed. by K.M. De Meyer, Heverlee, Belgium, 1983.
- [30] M. Simard-Normandin and C. Slaby, "Empirical modeling of low energy boron implants in silicon", J. Electrochem. Soc, vol 132, No 9, 2218-2223, 1985.
- [31] H. Ryssel, J. Lorenz and K. Hoffman, "Models for implantation into multilayer targets", Appl. Phys. A 41, 201-207, 1986.
- [32] G. Hobler and S. Selberherr, "Verification of ion implantation models by Monte Carlo simulations", in Proc. ESSDERC Conf., 445-448, September 1987.
- [33] H. Runge, "Distribution of implanted ions under arbitrary shaped mask edges", Phys. Status Solidi (A), Vol. 39, 595-599, 1977
- [34] M. Kump and R. Dutton, "Two dimensional process simulation SUPRA", in Process and device simulation for MOS-VLSI circuits, Eds; P. Antognetti, D.A. Antoniadis, R.W. Dutton and W.G. Oldham, NATO advanced study institute, Sogesta - Urbino, Italy, July 12-23, 1982
- [35] G. Hobler and S. Selberherr, "Efficient two-dimensional Monte Carlo simulation of ion implantation", in Proc. 5th Int. Conf. NASECODE, pp. 225-230, 1987.
- [36] R. Oven, D.G. Ashworth and C. Hill, "Simulation and measurement of the lateral spreading of ions implanted into amorphous targets", in Proc. SISDEP-3, 429-440, Bologna, Italy, 1988.
- [37] J. Lorenz, W. Krüger and A. Barthel, "Simulation of the lateral spread of implanted ions: theory", in Proc. NASECODE VI, 513-520, 1989.
- [38] H. Sunami, "Cell structures for future DRAM's", in IEDM Tech. Dig., 694-697, 1985.
- [39] D.P. Kennedy and R.R. O'Brien, "Analysis of the impurity atom distribution near the diffusion mask for a planar p-n junction", IBM J. Res. Dev., 9, 179-186, 1965.
- [40] D. Chin, M.R. Kump, H.G. Lee, and R.W. Dutton, "Process design using coupled two-dimensional process and device simulators", IEEE Trans. Electron Devices, ED-29, No 2, 336-340, 1982.
- [41] M.E. Law and R.W. Dutton, "Verification of analytic point defect models using SUPREM-IV", IEEE Trans. on computer-aided design, CAD-7, No 2, 181-190, February 1988.

- [42] M. Orlowski, "Unified model for impurity diffusion in silicon", Appl. Phys. Lett., vol. 53, No 14, 1323-1325, 3 October 1988.
- [43] N.E.B. Cowern, "General model for intrinsic dopant diffusion in silicon under non-equilibrium point-defect conditions", J. Appl. Phys., Vol. 54, No 9, 4484-4490, 1 November 1988.
- [44] M.D. Giles, "Defect-coupled diffusion at high concentrations", IEEE Trans. Computer Aided Design, CAD-8, No 5, 460-467, May 1989.
- [45] M.R. Kump and R.W. Dutton, "The efficient simulation of coupled point defect and impurity diffusion", IEEE Trans. on computer-aided design, CAD-7, No 2, 191-204, February 1988.
- [46] D. Mathiot and J.C. Pfister, "Dopant diffusion in silicon: a consistent view involving nonequilibrium defects", J. Appl. Phys., Vol. 55, No 10, 3518-3530, 1984.
- [47] P.M. Fahey, P.B. Griffin and J.D. Plummer, "Point defects and dopant diffusion in silicon", Review of Modern Physics, Vol 61, No 2, 289-384, April 1989.
- [48] W. Jüngling, P. Pichler, S. Selberherr, E. Guerrero and H.W. Pötzl, "Simulation of critical IC fabrication processes using advanced physical and numerical methods", IEEE Trans. Electron Devices, ED-32, No 2, 156-167, February 1985.
- [49] W. Nuyts and R. Van Overstraeten, "Computer calculations of impurity profiles in silicon (I)", Phys. Stat. Sol. (a), 15, 329-341, 1973.
- [50] C.P. Ho, J.D. Plummer, S.E. Hansen and R.W. Dutton, "VLSI process modeling - SUPREM III", IEEE Trans. Electron Devices, ED-30, 1438-1453, 1983.
- [51] R.B. Fair, "Concentration profiles of diffused dopants", in Impurity doping processes in silicon, 315-442, Amsterdam, North-Holland, 1981.
- [52] R.B. Fair and J.C.C. Tsai, "A quantitative model for the diffusion of phosphorus in silicon and the emitter dip effect", J. Electrochem. Soc., 124, No 7, 1107-1117, 1977.
- [53] R.B. Fair, "The effect of strain-induced band-gap narrowing on high concentration phosphorus diffusion in silicon", J. Appl. Phys., Vol. 50, 860-868, 1979.
- [54] E. Guerrero, H. Pötzl, R. Tielert, M. Grasserbauer and G. Stingeder, "Generalized model for the clustering of As dopants in Si", J. Electrochem. Soc., vol. 129, No 8, 1826-1831, August 1982.
- [55] D. Collard, "Application de la méthode des éléments finis à la simulation bidimensionnelle de technologies silicium", Thèse de docteur ingénieur, Université de Lille, 1984.

- [56] H. Ryssel, K. Habeger, K. Hoffmann, G. Prinke, R. Dümcke and A. Sachs, "Simulation of doping processes", IEEE Trans. Electron Devices, ED-27, No 8, 1484-1492, 1980.
- [57] H. Ryssel, K. Müller, K. Habeger, R. Henkelmann and F. Jahnel, "High concentration effects of ion implanted boron in silicon", Appl. Phys., 22, 35-38, 1980.
- [58] A.H. Van Ommen, "Diffusion of group III and V elements in SiO<sub>2</sub>", in Proc. INFOS 87, 244-264, Eds G. Declerck and R. de Keersmaecker, North-Holland, 1987.
- [59] L. Mei and R.W. Dutton, "A process simulation model for multilayer structures involving polycrystalline silicon", IEEE Trans. Electron Devices, ED-29, 1726-1734, November 1982.
- [60] K. Sakamoto, K. Nishi, T. Yamaji, T. Miyoshi and S. Ushio, "Complete process modeling for VLSI multilayer structures", J. Electrochem. Soc., vol. 132, 2457-2463, 1985.
- [61] A.G. O'Neill, C. Hill, J. King and C. Please, "A new model for the diffusion of arsenic in polycrystalline silicon", J. Appl. Phys., 64 (1), 167-174, 1 July 1988.
- [62] S.K. Jones and C. Hill, "Modelling dopant diffusion in polysilicon", in proc. SISDEP III conf., 441-449, Eds G. Baccarani and M. Rudan, Bologna, 1988.
- [63] O.C. Zienkiewicz, "The finite element method", Mc Graw-Hill, New-York, 1977.
- [64] D. Collard and J.N. Decarpigny, "Two-dimensional process simulation using a quadratic finite element discretization", COMPEL, Vol. 3, No 1, 17-33, 1984.
- [65] L. Borucki, H.H. Hansen and K. Varahramyan, "FEDSS - a 2D semiconductor fabrication process simulator", IBM J. Res. Develop., vol. 29, no 3, 263-276, 1985.
- [66] R.R. O'Brien, C.M. Hsieh, J.S. Moore, R.F. Lever, P.C. Murley, K.W. Brannon, G.R. Srinivasan and R.W. Knepper, "Two dimensional process modeling: a description of the SAFEPRO program", IBM J. Res. Develop., vol. 29, no 3, 229-241, 1985.
- [67] E. Caquot and A. Marrocco, "Numerical simulations with finite elements of non-linear impurity diffusion in silicon", in proc. NASECODE II conf., Dublin, Ireland, June 17-19, 1981.
- [68] D. Collard and K. Taniguchi, "IMPACT - a point-defect based two dimensional process simulator: modeling the lateral oxidation enhanced diffusion of dopants in silicon", IEEE Trans. Electron Devices, ED-33, No 10, 1454-1462, 1986.

- [69] B. Baccus, D. Collard, E. Dubois and D. Morel, "IMPACT<sup>4</sup> - a general two-dimensional multilayer process simulator", in proc. SISDEP III conf., 255-266, Eds G. Baccarani and M. Rudan, Bologna, 1988.
- [70] J.A. Meijerink and H.A. van der Vorst, "An iterative solution method for linear systems of which the coefficient matrix is a symmetric M-matrix", Mathematics of Computation, Vol. 31, No 137, 148-162, January 1977.
- [71] T. Wada and R.L.M. Dang, "Modification of ICCG method for application to semiconductor device simulators", Electronic letters, vol. 18, No 6, 265-266, 18th March 1982.
- [72] M. Hane and K. Hane, "Finite and boundary element approach to process simulation with conjugate gradient-based method", in Proc. 5th Int. Conf. NASECODE, pp. 219-224, 1987.
- [73] B. Baccus, E. Dubois, D. Collard and D. Morel, "Efficient two-dimensional multilayer process simulation of advanced bipolar devices", Solid-State Electronics, Vol. 32, No 11, 1013-1023, 1989.
- [74] B.E. Deal and A.S. Grove, "General relationship for the oxidation of silicon", J. Appl. Phys., vol. 36, No 12, 3770-3778, 1965.
- [75] E.A. Irene, "Silicon oxidation studies: a revised model for thermal oxidation", J. Appl. Phys., Vol. 54 (9), 5416-5420, 1983.
- [76] A. Fargheix, G. Ghibaudo and G. Kamarinos, "A revised analysis of dry oxidation of silicon", J. Appl. Phys., Vol 54 (5), 2878-2880, 1983.
- [77] A.G. Revesz, B.J. Mrstik, H.L. Hughes and D. McCarthy, "Structure of SiO<sub>2</sub> films on silicon as revealed by oxygen transport", J. Electrochem. Soc., Vol. 133, No 3, 586-592, March 1986.
- [78] R.H. Doremus, "Oxidation of silicon - stress relaxation in silica", J. Electrochem. Soc., Vol. 134, No 8, 2001-2003, August 1987.
- [79] E.A. Lewis and E.A. Irene, "Models for the oxidation of silicon", J. Vac. Sci Technol. A<sup>4</sup> (3), 916-925, May/June 1986.
- [80] H.Z. Massoud, J.D. Plummer and E.A. Irene, "Thermal oxidation of silicon in dry oxygen growth-rate enhancement in the thin regime - I. Experimental results", J. Electrochem. Soc., Vol. 132, No 11, 2685-2693, November 1985.
- [81] H.Z. Massoud, J.D. Plummer and E.A. Irene, "Thermal oxidation of silicon in dry oxygen growth-rate enhancement in the thin regime - I. Physical mechanisms", J. Electrochem. Soc., Vol. 132, No 11, 2693-2700, November 1985.
- [82] C.-J. Han and C.R. Helms, "Parallel oxidation mechanism for Si oxidation in dry O<sub>2</sub>", J. Electrochem. Soc., Vol. 134, No 5, 1297-1302, May 1987.

- [83] D.-B. Kao, J.P. McVittie, W.D. Nix and K.C. Saraswat, "Two-dimensional thermal oxidation of silicon -I. Experiments", IEEE Trans. Electron Devices, ED-34, No 5, 1008-1017, 1987.
- [84] D. Chin, S.Y. Oh, S.M. Hu, R.W. Dutton and J.L. Moll, "Two-dimensional oxidation", IEEE Trans. Electron Devices, ED-30, No 7, 744-749, 1983.
- [85] A. Poncet, "Finite element simulation of local oxidation of silicon", IEEE Tr. CAD, Vol. CAD-4, No 1, 41-53, 1985.
- [86] H. Matsumoto and M. Fukuma, "Numerical modeling of nonuniform Si thermal oxidation", IEEE Trans. Electron Devices, ED-32, No 2, 132-140, 1985.
- [87] R.B. Marcus and T.T. Sheng, "The oxidation of shaped silicon surfaces", J. Electrochem. Soc, Vol. 129, No 6, 1278-1282.
- [88] B.R. Penumalli, "Lateral oxidation and redistribution of dopants", in proc. NASECODE II conf., Dublin Ireland, June 17-19, 1981.
- [89] N. Guillemot, G. Pananakakis and P. Chenevier, "A new analytical model of the "Bird's Beak"", IEEE Trans. Electron Devices, ED-34, No 5, 1033-1038, 1987.
- [90] D. Chin, "Two-dimensional oxidation - modeling and applications", Integrated circuits laboratory, Stanford University, June 1983.
- [91] E.P. EerNisse, "Viscous flow of thermal SiO<sub>2</sub>", Appl. Phys. Letters, Vol. 30, No 6, 290-293, 15 March 1977.
- [92] H. Umimoto, S. Odanaka, I. Nakao and H. Esaki, "Numerical modeling of non-planar oxidation coupled with stress effects", IEEE Trans. Computer Aided Design, CAD-8, No 6, 599-607, June 1989.
- [93] C.P. Ho and J.D. Plummer, "Si/SiO<sub>2</sub> interface oxidation kinetics: a physical model for the influence of high substrate doping levels I. Theory", J. Electrochem. Soc., Vol. 126, No 9, 1516-1522, 1979.
- [94] D.A. Antoniadis and R.W. Dutton, "Models for computer simulation of complete IC fabrication processes", IEEE J. Solid-State Circuits, Vol. SC-14, No 2, 412-422, April 1979.
- [95] A.J. Chorin, "A numerical method for solving incompressible viscous flow problems", J. Comp. Phys., vol. 2, 12-26, 1967.
- [96] L. Borucki and J. Slinkman, "An efficient finite element algorithm for modeling two-dimensional oxidation and impurity redistribution in silicon", in proc. NASECODE IV conf., Dublin, Ireland, 226-233, June 1985.
- [97] M.E. Law, C.S. Rafferty and R.W. Dutton, "The SUPREM-IV program", Integrated Circuits laboratory, Stanford University, Stanford, 1987.
- [98] M. Av-Ron, M. Shatzkes, P.J. Burkhardt and I. Cadoff, "Distribution of dopant in SiO<sub>2</sub>-Si ", J. Appl. Phys., Vol. 47, No 7, 3159-3166, July 1976.

- [99] D. A. Antoniadis, M. Rodoni and R.W. Dutton, "Impurity redistribution in  $\text{SiO}_2$ -Si during oxidation: a numerical solution including interfacial fluxes", J. Electrochem. Soc., Vol. 126, No 11, 1939-1945, November 1979.
- [100] I. Desoutter, D. Collard and J.N. Decarpigny, "Two-dimensional simulation of diffusion under oxidizing ambient using the finite element method", in proc. NASECODE IV conf., Dublin, June 17-24, 1985.
- [101] K. Taniguchi, K. Kurosawa and M. Kashiwagi, "Oxidation-enhanced diffusion of boron and phosphorus in (100) silicon", J. Electrochem. Soc., Vol. 127, 2243-2248, 1980.
- [102] Y. Ishikawa, Y. Sakina, H. Tanaka, S. Matsumoto and T. Numi, "The oxidation-enhanced diffusion of arsenic and phosphorus in silicon by thermal oxidation", J. Electrochem. Soc., Vol. 129, 644-648, 1982.

## CHAPTER IV

### ADAPTIVE MESHING

#### Abstract

In this chapter, an adaptive mesh refinement technique is proposed for the two-dimensional finite element multilayer process simulation. Mesh refinement is based on dopant concentration ratio inside each element, together with a prediction technique, minimising the interpolation errors. Significant reduction in CPU time is obtained by automatic grid refresh, several methods being tested. The mesh adequacy is evaluated with bipolar test structures, using analytical punchthrough voltage calculations. Applications of these methods are presented in two ways: first, an advanced trench isolated polysilicon bipolar transistor is simulated to show the general possibilities of the techniques, and finally a coupled process and device simulation approach allows the evaluation of the scheme on real structures, compared with experimental measurements.



## I. INTRODUCTION

As already pointed out in chapter II (part III), mesh generation is a key point for two-dimensional process simulation efficiency. The meshing procedures previously proposed (3.2:strategy and 3.3:mesh generation) have proved their capabilities on standard structures. However, it is obvious that there exist some limitations to these approaches:

- i) There is always an interaction between the user and the mesh generation module. In our case, this is limited to the definition of the first grid.
- ii) This first grid is always rectangular, but its shapes can be considerably altered during the different processing steps (oxidation and etching). This means that 1) the initial mesh may not be so efficient at the end of a complete simulation and that 2) it is difficult to mesh with a sufficient node density some critical regions (bottom of a trench for the channel-stopper).
- iii) There is no evaluation of the error made during the diffusion steps and so the accuracy is only estimated by the user's experience.

These limitations have motivated a study on adaptive mesh refinement. It is also believed that IMPACT<sup>4</sup> is a very interesting tool for such purpose. On the other hand, if more sophisticated process models have to be implemented in the future versions of the program, care must be taken to avoid an influence of the spatial discretisation on the results precision (and consequently on models evaluation).

At this point, it is useful to have a rapid overview on other process simulators. Most of the codes use fixed *a priori* grids, for finite difference method (SUPREM IV [1]), as well as for finite element method (FEDSS [2], SAFEPRO [3]). Some attempts have been done to use adaptative grids, but the application of these techniques to complex multilayer structures has not been performed nor seems to be straightforward. Using finite difference method, interesting work has been undertaken by Pichler

et al. [4] for the 1D and 2D cases (rectangular shape). The ASWR method has been also proposed [5]. Concerning the finite element method, refined windows techniques have been used (TITAN [6]).

Other schemes for the finite element method have been studied such as the refinement technique proposed by Amaratunga [7] on simple test structures and finally the moving finite element has been applied to process simulation only for 1D problems [8]. On the other hand, adaptative mesh refinement is now often used in device simulation programs for general structures [9], but the non-stationary problem in process simulation makes the proposed methods not directly usable.

In this chapter, we present the numerical techniques aimed to solve these problems, taking into account the following considerations [10-11]:

- \* there should not be any interaction between the user and the mesh generation, even when defining the initial structure (silicon substrate).
- \* Non planar (silicon trench), multilayer structures (doping from polysilicon) and moving boundary problems (oxidation) must be treated.
- \* The spatial discretisation error associated with the chosen mesh must be quantitatively evaluated.
- \* Significant reduction in the CPU time is expected (compared to usual fixed grid strategy). However, as mesh refinement has little influence on the convergence of the numerical procedure, and only on results precision, this is not the major motivation in our application. Nevertheless, it will be shown that interesting characteristics will be reached.

Section II describes the numerical techniques and the mesh generation principles. In section III, simple test examples are studied, evaluating the spatial discretisation error using bipolar transistor simulation, and determining the best parameters for refinement and refresh procedure. Applications are performed in two ways in section IV: the complete simulation of a polysilicon, trench isolated bipolar device is first presented, and finally a coupled process and device approach allows to determine the refinement procedure efficiency by the calculation of electrical parameters in a VLSI technology.

## II. NUMERICAL TECHNIQUES AND MESH GENERATION

### 2.1 Introduction

As a first step, it is important to show that in the scheme we have proposed for the resolution of the diffusion equation (chapter III, 3.3), the temporal discretisation error can be decoupled from the spatial one. For this purpose, a simple 1D test example is presented. Ion implantation of arsenic, followed by a diffusion step at 1000°C for 20 min is simulated with different number of time steps (constant through the diffusion), and different grid sizes. Fig. 1 indicates the effect of the number of time steps on the diffusion profile, for a small mesh size (50 Å). As the number of timesteps increases, the profile is steeper, with a convergence for small time steps increments (such results have been already mentioned when studying the relation between the non-linear and temporal aspects of diffusion).

In the same manner, Fig. 2 shows the effect of the spatial discretization for a reasonable number of time steps (30). If the initial mesh is not fine enough for the ion implantation step, the integrated dose in the substrate is incorrect (curve for 500Å grid spacing, Fig. 2). Again, for finer mesh size, a saturation in the doping profile is obtained. Two important informations can be deduced: first, the effect of the temporal scheme can be decoupled from the mesh size, allowing to focus on the grid refinement techniques. Secondly, the behaviour depicted in Fig.2 allows to determine the adequacy of the mesh by simple punchthrough calculations as will be described in section III.

An other point is to examine the type of refinement to be studied: h-refinement (increasing the number of elements) or p-refinement (increasing the polynomial order of selected elements). In preliminary works, 6-noded triangles (quadratic shape function) were used for the diffusion problem [12], but this imposed very severe requirements on the mesh. This experience indicates that it can not be extended to a general multilayer problem. For this reason, h-refinement on linear triangular elements is chosen.

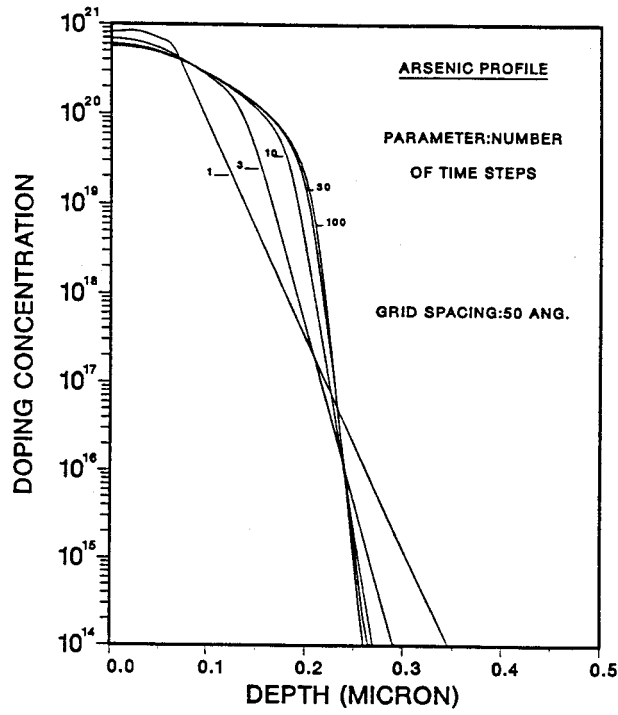


Fig. 1 Effect of the number of time steps on the diffused profiles.

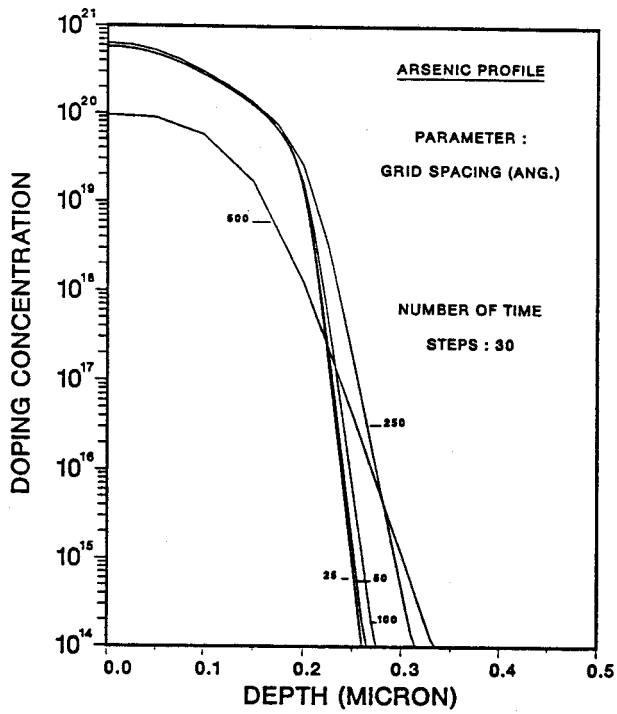


Fig. 2 Effect of the grid spacing on the diffused profiles.

## 2.2 Mesh generation

Due to the special characteristics of process simulation, original procedures are introduced for mesh generation and refinement. They are mainly based on node connectivity limitation, and allow also to solve the boundary consistency between different layers.

In a standard manner, an element to be refined is divided into four sub-triangles by adding a node at each mid-side. In order to obtain a conforming finite element triangulation, if one node is added on a side, the triangle is divided into two sub-triangles, if two nodes are added on the sides, the triangle is always divided into four sub-triangles. This method generally preserves the quality of the initial grid, or in other words, the new elements have not an element aspect ratio too much greater than that of the original element.

However, if the initial grid is very coarse, these operations may not be sufficient to obtain the desired implantation dose in the target (fig. 2). Moreover, a smooth transition between refined and unrefined zones is needed. For these purposes, a new "propagation" procedure, based on connectivity limitation, has been developed, as explained in fig. 3 [13]. As this results in a propagation of the refined zones, the problems posed by the non-stationary equations are facilitated, since the diffusion front movements are anticipated. This proposed method reveals to be more efficient in the case of process simulation than other proposed schemes [14].

In order to obtain the best triangulation, the Delaunay method is used. Here we take again advantage of the MSHPTS mesh generator from INRIA software library [15]. As a first step for this study, we use this mesh generator after each mesh refinement iteration. This is perphas not the most efficient procedure in term of CPU time but it allows to focus on other important aspects of the complete problem. In conjunction, the regularisation or mesh smoothing is also widely used (each node is moved to the the center of gravity of the polygon defined by its adjacent nodes). If this latter technique is not needed for ion implantation, it is absolutely necessary for diffusion processes, otherwise the mesh quality degrades rapidly as the diffusion fronts move. On the other hand, it has the disadvantage to add interpolation calculations.

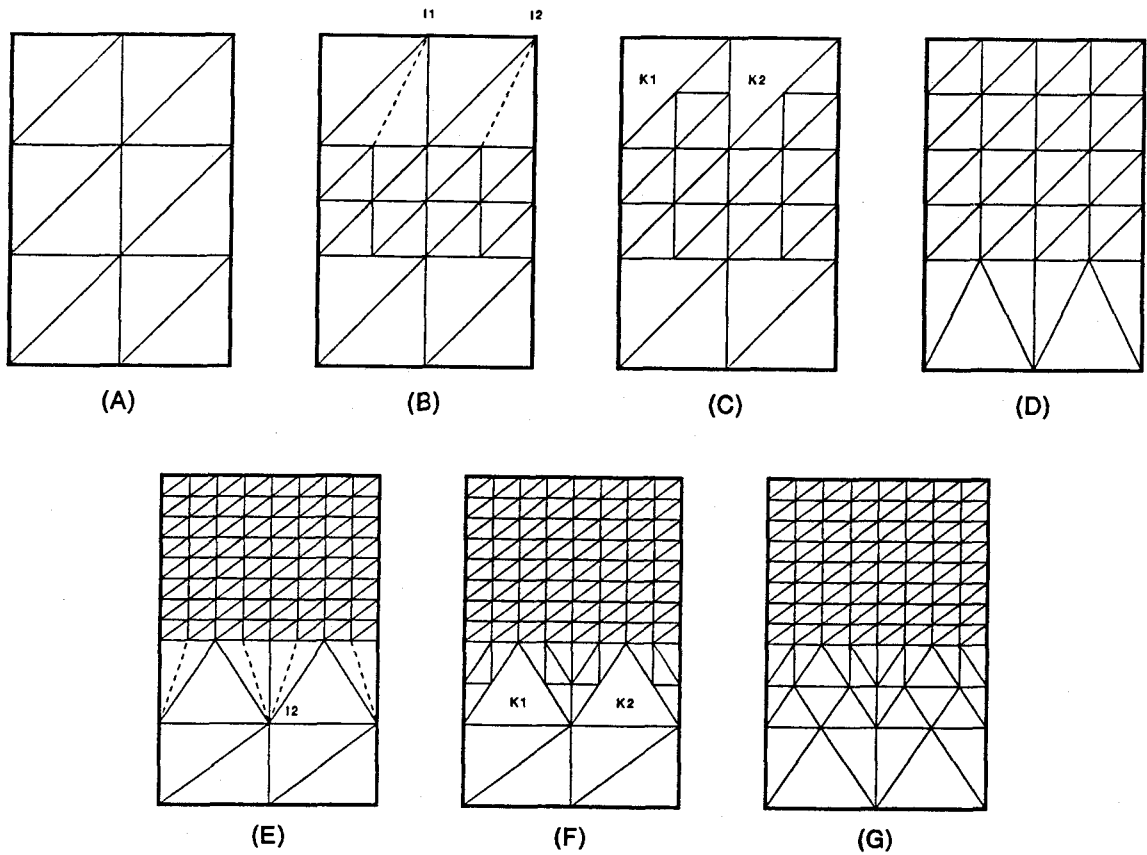


Fig. 3 Principle of mesh refinement propagation. Ion implantation is performed on the mesh (A). If all the distribution is located in the top triangle layer, only the second layer is refined (B), as no sufficient gradient is detected in the first layer. In the algorithm, if I1 and I2 are boundary nodes and originate at least a bisection, the concerned elements are split into four ( (B) and (C) ). Triangles K1 and K2 have two new nodes on their sides: a third node is then added ( (C) and (D) ). The same procedure is used from (E) to (G) except that, as I2 is an interior node originating at least 2 bisections (E), the concerned elements are split into four.

Finally, in order to avoid very long and narrow elements, a technique based on connectivity limitation is used. Considering a patch of elements as depicted in fig. 4a, if a node connectivity is greater typically than 11, nodes are added on the mid-sides of the concerned elements (fig. 4b), in order to obtain, by a new triangulation, the structure of fig. 4c.

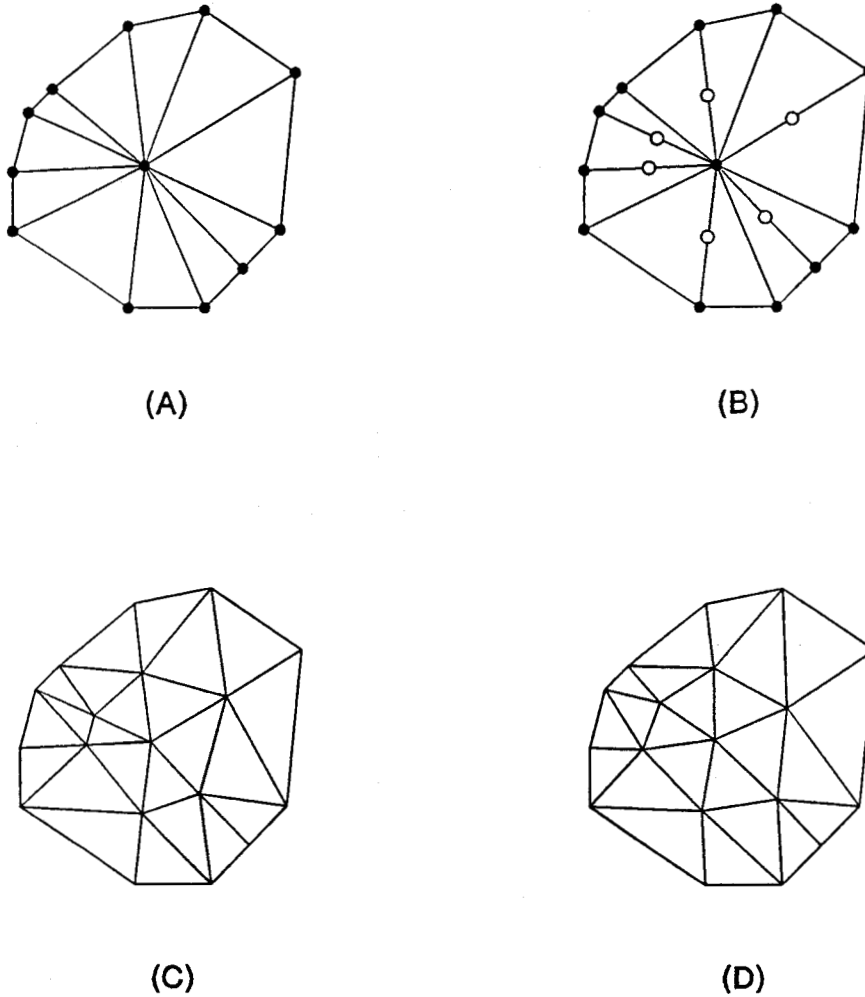


Fig. 4 Principle of node connectivity limitation technique. Considering a patch of elements, if a node connectivity is too great (A), nodes are created at mid-sides. The number of created nodes is half the number of node connectivity (B). A new triangulation is then performed (C) and finally a regularization step (D).

In this study, it has been found sufficient, in a first approach, to apply the above procedure only in the silicon layer. The other layers have then to be updated, as the segregation phenomenon must be taken into account (at the interface between two layers, there are double nodes at the same coordinates, belonging to each layer). After update (new nodes are created, corresponding to the boundary added silicon nodes, followed by a re-triangulation of the concerned layer), the connectivity limitation of fig. 4 is applied iteratively in these oxide, polysilicon or nitride layers. When applying this connectivity limitation, the new nodes can not be created on the interface so that the mesh in the silicon domain must not be modified. From the physical point of view, this permits also to take carefully into account the segregation effect extension.

### 2.3 Definition of criteria

In order to determine the regions where local refinement must be performed during the simulation, some criterion must be defined. This can be achieved in the present case, for example with an *a posteriori* error estimate using local Neumann error criteria as proposed in [16]. However this leads to supplementary calculations, and the question is whether a simpler criterion could be as efficient in a similar way. For this purpose, the basic idea is to use the concentration ratio inside each element to decide if this element must be splitted or not. Such procedure is widely used in device simulation to construct the first grid, but until now, it has not been fully exploited in process simulation. By introducing  $\alpha$ , we estimate the quality of a linear element upon the nodal concentration  $C_{ij}$ , for impurity  $i$  (As, B, P, Sb or other diffusing species) and node  $j$ :

$$\max \left( \frac{C_{1j}}{C_{1k}} \right) \leq \alpha \quad \text{with } j=1,3 \quad k=1,3 \quad (1)$$



However, it has been found that this criterion is not directly applicable in actual 2D simulations because it leads to a too large number of nodes. It is then proposed that  $\alpha$  depends on the concentration level:

$$\alpha = \alpha_0 \cdot 2 \left( \text{LOG}_{10} \left( \frac{C_{\text{max}_i}}{\max(C_{ij})} \right) \right) \quad (2)$$

$C_{\text{max}_i}$  being the maximum of concentration of impurity  $i$  in the structure. This ensures a reasonable number of nodes and moreover the meshes created are nearly regular. From the physical point of view, the diffusion coefficient is related to the doping level, giving consistency to this  $\alpha$  criterion [7].

However, in case of smooth profiles at a low concentration level, conditions (1) and (2) are not sufficient. Moreover, another aspect is not well solved by the presented method: the transition between a steep profile (arsenic emitter) and the uniform background level (arsenic epitaxy at low doping level). In order to solve these two limitations, a 'prediction' technique is introduced as follows: the ratio between concentrations before and after diffusion step is computed for each node:

$$\beta_{ij} = \frac{(C_{ij})_t}{(C_{ij})_{t+\Delta t}} \quad (3)$$

The ratio between the values of  $\beta_{ij}$  for the nodes contained in a patch of elements is computed and compared to a limiting value:

$$\beta = \frac{\max (\beta_{ij})}{\min (\beta_{ik})} \quad (4)$$

This  $\beta$  value allows to determine regions where an important change in the concentrations will arise at the next time step. If  $\beta$  is greater than 2, nodes are added following the algorithm presented in fig. 4.

These methods are illustrated in the case of boron implantation in a silicon trench structure, with oxide masks. A value of 5 is used for  $\alpha_0$ , and 6 iterations on the refinement procedure are needed to satisfy condition (1) in silicon. Fig. 5a shows the initial mesh, fig. 5b the final mesh and fig. 5c the corresponding boron countours. This a typical example where mesh refinement is very useful, since an *a priori* fixed grid would contain numerous nodes to describe precisely the channel stopper region, as illustrated in fig. 5c.

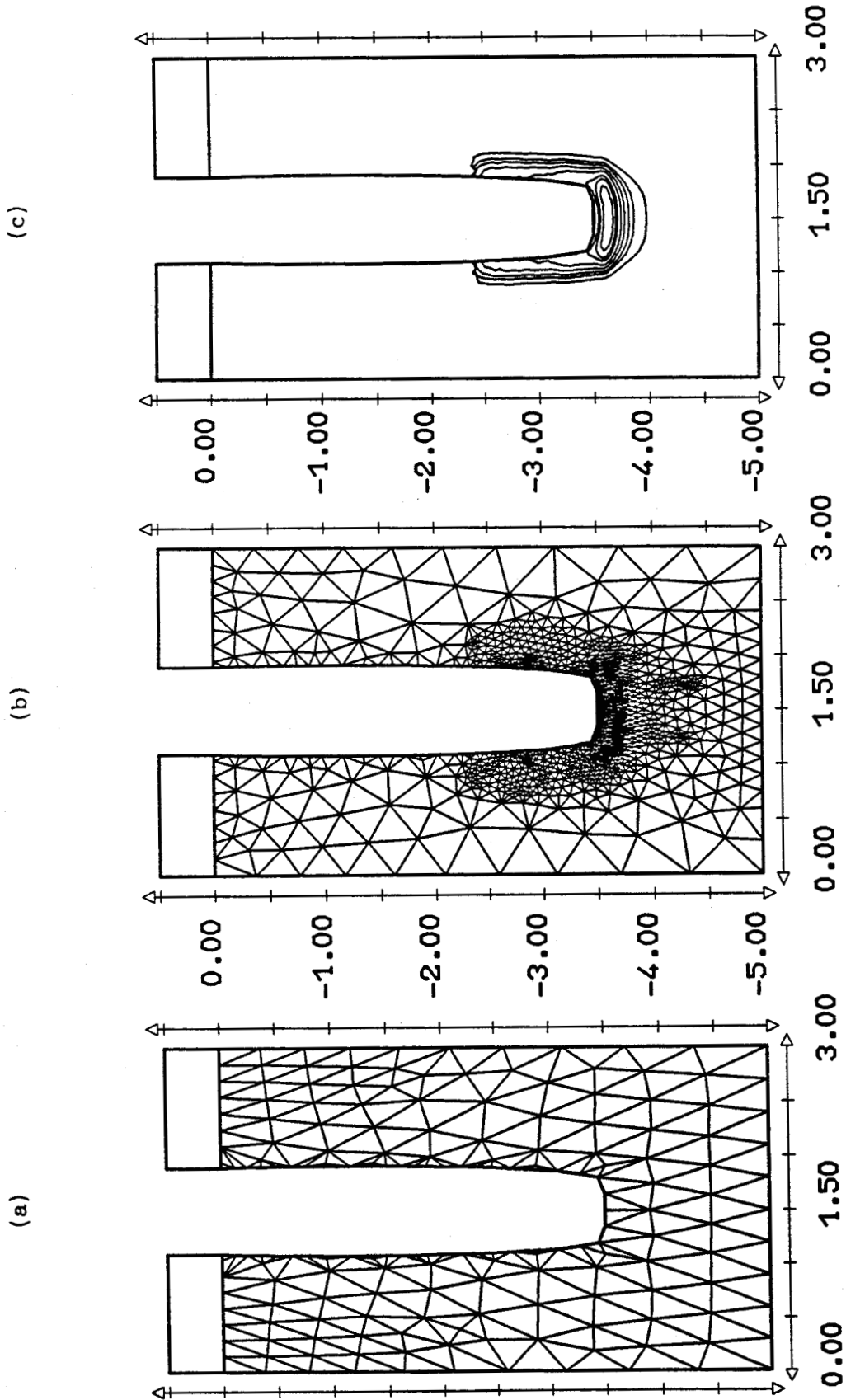


Fig. 5 Boron implantation in a silicon trench structure. From the initial coarse mesh, 6 iterations are used to obtain the final mesh. Boron contours are also displayed.

## 2.4 Refresh procedure

It is expected that a significant amount of CPU time can be saved by applying the algorithms and criteria presented above. However, in this case, precision is reached with an important increase of computation time. This is mainly due to the fact that node elimination is not used, since condition (1) does not seem to be adequate for this aim. On the other hand, a systematic refresh at each time step is not realistic. For this purpose a "refresh" technique has been introduced, automatically performed and determined by the program upon the dopant distribution evolution. The mean value of  $\alpha$  is calculated at the beginning of each time step in order to determine if a refresh procedure is needed before the diffusion equation resolution:

$$\alpha_{\text{mean}} = \frac{1}{NT} \cdot \sum_{1=1,NT} \alpha_1$$

with:

$$\alpha_1 = \max \left( \frac{C_{ij}}{C_{ik}} \right) \quad (5)$$

with NT the number of elements in silicon.

When the ratio between  $\alpha_{\text{mean}}$  at the first time step and at the current time step is greater than a given value (experiments show that 1.5 is a suitable value), this means that the mesh is no more optimal with respect to the doping profile evolution : the condition (1) can be fulfilled with much less nodes. This is typically the case after a little number of time steps following ion implantation, the profile being more flattened near the surface. This method allows to optimize the total number of nodes for a given value of  $\alpha_0$ .

From the topological point of view, a general refresh procedure has been developed for complex geometries and is described in fig. 6 where, as a test, the refresh procedure is applied on the trench structure of fig. 5b.

The first task is to eliminate completely interior nodes and to reconstruct a list of nodes describing the boundary. Triple points (where three boundary lines converge) and points at sharp corners are extracted as proposed in [17]. Between these key positions, new points are created with the initial node spacing as a minimal value. A first triangulation is performed (fig. 6a). The connectivity limitation of fig. 4 is then iteratively applied to obtain the triangulation presented in fig. 6 (again during this procedure, no nodes are added on the boundaries). Six iterations are generally sufficient to reach convergence. This mesh is now the primitive grid where condition (1) is applied and allows to obtain the structure described in fig. 6c. The other layers are then updated taking into account this new topology.

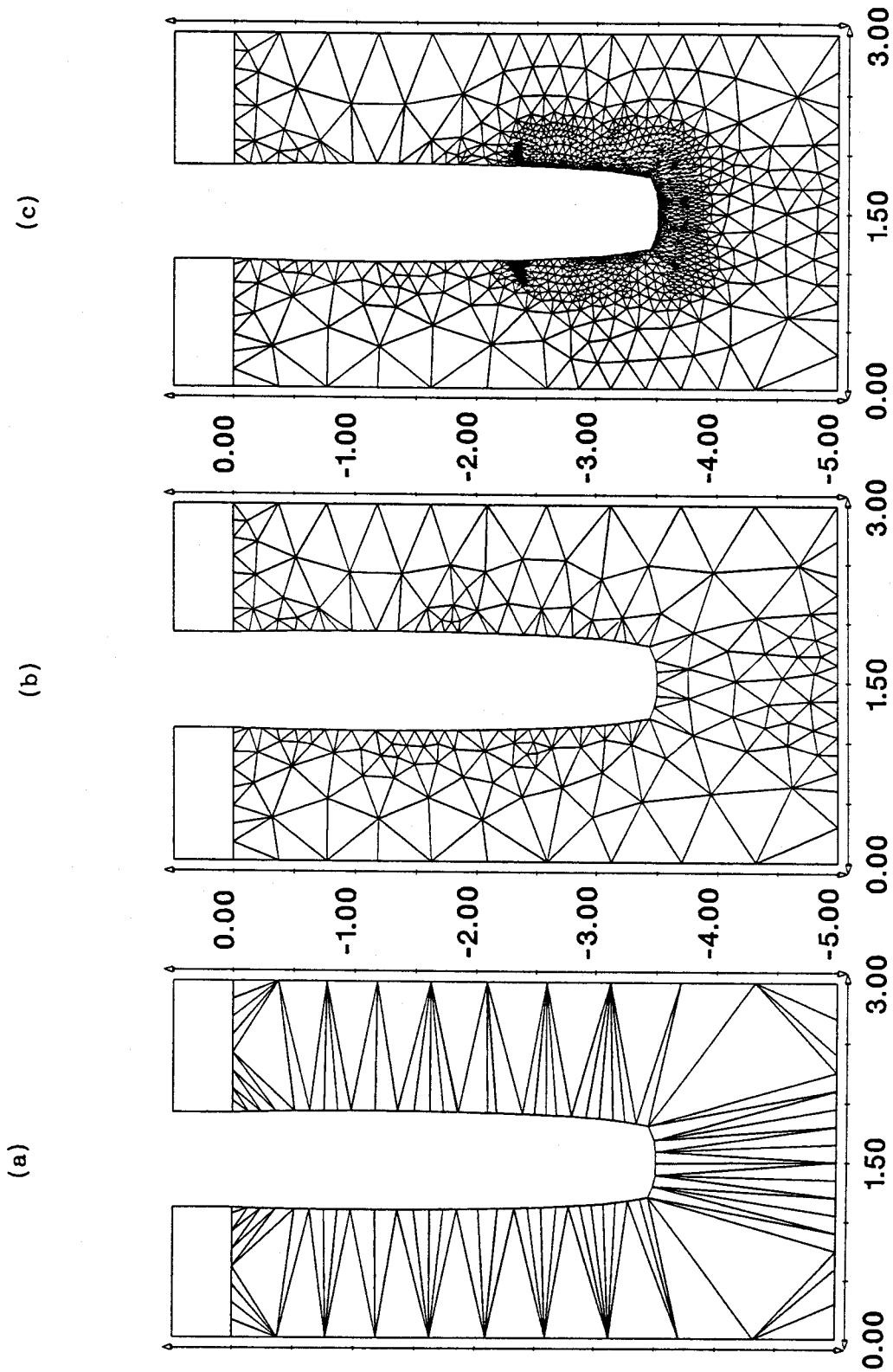


Fig. 6 Example of refresh procedure on the trench structure of Fig. 5. Internal nodes are first removed and boundary nodes recalculated. Iterations on geometrical considerations allow to obtain an intermediate grid ( use of the node connectivity limitation with a value of 11 for internal nodes and a value of 6 for boundary nodes).  $\alpha$  criterion is then applied.

### III. TEST EXAMPLES ON BIPOLAR STRUCTURES

#### 3.1 Quantitative error evaluation

When conditions (1) and (2) are applied, the problem is to define a "reasonable" value for  $\alpha_0$ . For this purpose, we take advantage of the doping profile evolution with respect to the mesh size (fig. 2) and use an evaluation of results precision in a test structure. As the electrical characteristics of bipolar devices are very sensitive to the doping profile, a calculation of the punchthrough voltage is used [18]. If we consider an uniform doping profile for the collector, an analytical expression can be derived for the punchthrough voltage  $V_{PT}$  [19]:

$$V_{PT} = \frac{Q_{B0}^2}{2 \epsilon_0 \epsilon_r q N_{DC}} \quad (6)$$

where  $Q_{B0}$  is the base Gummel number and  $N_{DC}$  the epitaxial doping concentration. This dose evaluation has 3 major advantages:

- \* The calculation is independent on the sophistication level of process models (as could be the problem if one wants to compare simulation results with experiments). It can then be used for many process program evaluation.
- \* An electrical parameter is determined, but interpolation errors that occur during the interface with a device simulator are avoided.
- \* The calculated quantity is meaningful for process engineers.

In the case of 1D simulation, this calculation is directly applied, and in 2D, a cross section in the intrinsic part of the device is used.

### 3.2 Single impurity problem

A first test example consists in the formation of an NPN bipolar transistor, where only the emitter implantation and diffusion are simulated. In this case, the errors due to dopant diffusion coupling are eliminated, in order to study precisely the intrinsic behaviour of mesh refinement. Arsenic is implanted at an energy of 40 keV and at a dose of  $6 \cdot 10^{15}$  at/cm<sup>2</sup>, followed by an annealing at 1000°C for 20 min. The resulting profile is then evaluated with the punchthrough voltage calculation, assuming a constant boron base doping of  $1 \cdot 10^{17}$  at/cm<sup>3</sup>, a constant collector doping of  $2 \cdot 10^{16}$  at/cm<sup>3</sup>, and a constant base/collector junction depth of 0.7 μm.

As a first test, at each time step, nodes are added with respect to condition (1) without the use of refresh procedure or prediction technique. The same value is used for  $\alpha_0$  for ion implantation and diffusion. For the 1D case, fig. 7 displays the evolution of  $V_{PT}$  versus  $\alpha_0$  for different initial mesh spacings in the depth direction. For the 2D case, a 2 μm by 2 μm structure is used, arsenic is implanted in a window between 0.7 and 1.3 μm. In this case a vertical cross section is extracted in the middle of the structure to determine  $V_{PT}$ . Fig. 8 displays also  $V_{PT}$  versus  $\alpha_0$  for different initial mesh spacing and a comparison with the 1D case.

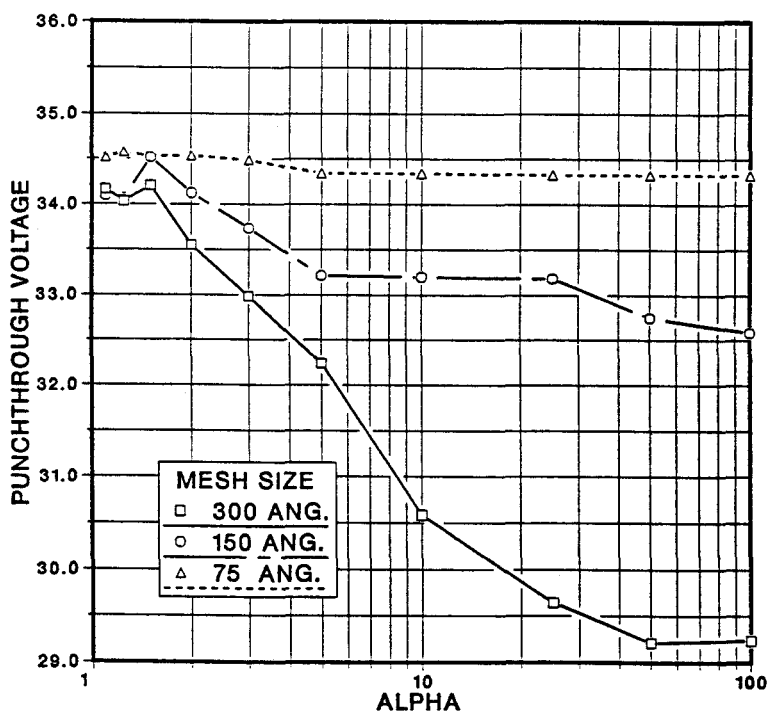


Fig 7 1D test example (single diffusing impurity). Evolution of  $V_{PT}$  with  $\alpha_0$ .



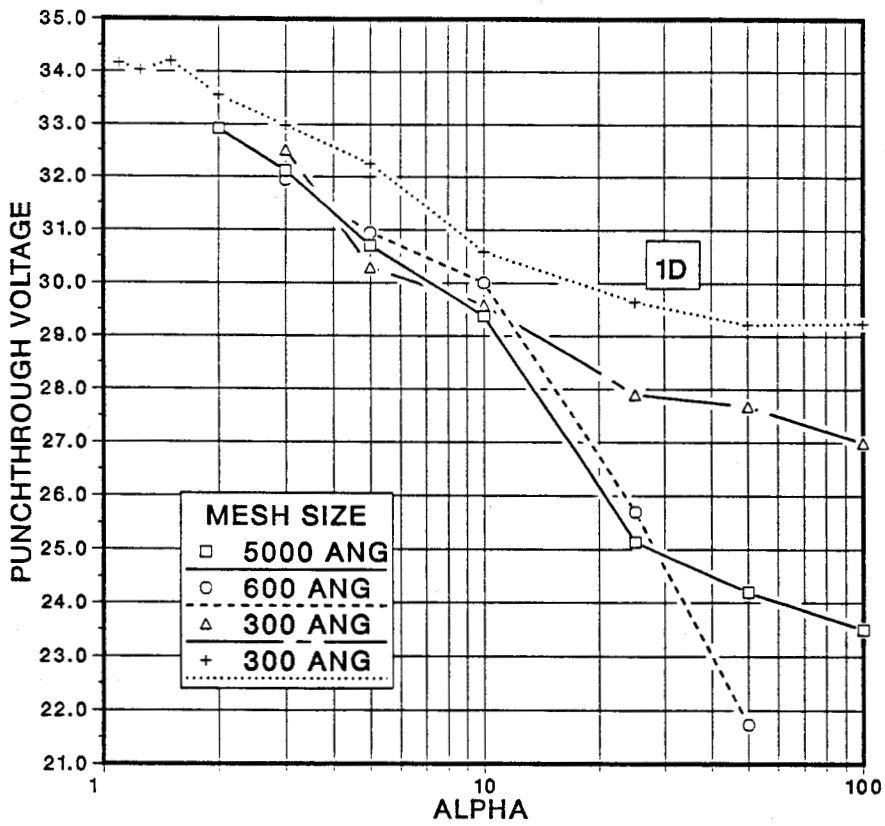


Fig. 8 2D test example. Three initial mesh sizes are tested and a comparison with a 1D calculation is also provided.

From these two figures, three conclusions can be outlined:

- i) 75 Å point spacing is fine enough to obtain accurate results without refinement.
- ii) For  $\alpha_0$  values less than 10 (2D calculations), the initial mesh spacing has no influence on results precision, this is a direct consequence of the "propagation" method.
- iii) It is possible to obtain very accurate results with the refinement procedure in 2D simulations, at the expense of a large number of nodes (about 5,000 for  $\alpha_0=3$  and 10,000 for  $\alpha_0=2$ ).

An other way of representation allows to compare the results in this test case for different refresh procedures. Generally, the number of nodes generated after refresh is significantly lower than that in the initial grid. Two alternatives can be studied: a) this new number of node is not changed, b) a constant number of nodes is considered after refresh. In this latter case, the value of  $\alpha_0$  is automatically decreased during the refresh procedure, the final value of this parameter being typically divided by 5 to 10 compared to the user specified value.

Figure 9 represents the CPU time needed to obtain a given precision (specified by  $V_{PT}$ ), the maximum number of nodes being limited to 5,000. In all cases (except for the fixed grid strategy), the initial grid spacing is  $0.5 \mu\text{m}$ . The six curves represent:

- 1) A classical fixed grid strategy. Even in this simple case (rectangular domain), the number of nodes is rapidly the main limitation. It shows the possibilities of existing schemes.
- 2) The refinement technique without refresh (corresponding to fig. 8). High precision is only reached for excessive CPU time.
- 3) The refinement technique with refresh (alternative a)). CPU times are comparable to the standard fixed grid strategy, since the number of nodes after refresh is decreased.
- 4) The refinement technique with a constant number of nodes after refresh procedure (alternative b):  $\alpha_0$  decreased during the refresh procedure).
- 5) Simple refresh (curve 3), with the prediction technique (equations (3) and (4)). This latter is applied at the end of refresh procedure, when condition (1) is satisfied.
- 6) Refresh with constant number of nodes and prediction technique.

This last curve clearly shows the efficiency of the proposed schemes for this test example, since very high precision is possible even with a low value of  $\alpha_0$ , for low CPU time requirements. The CPU time needed for refinement and refresh is highly compensated by the mesh adequacy with the doping distribution. In this case, the number of nodes is completely optimized with respect to the doping profile evolution. When considering the trends of silicon processing towards reduction of the diffusion cycle by the use of rapid thermal annealing, this method will be considered as very satisfactory.

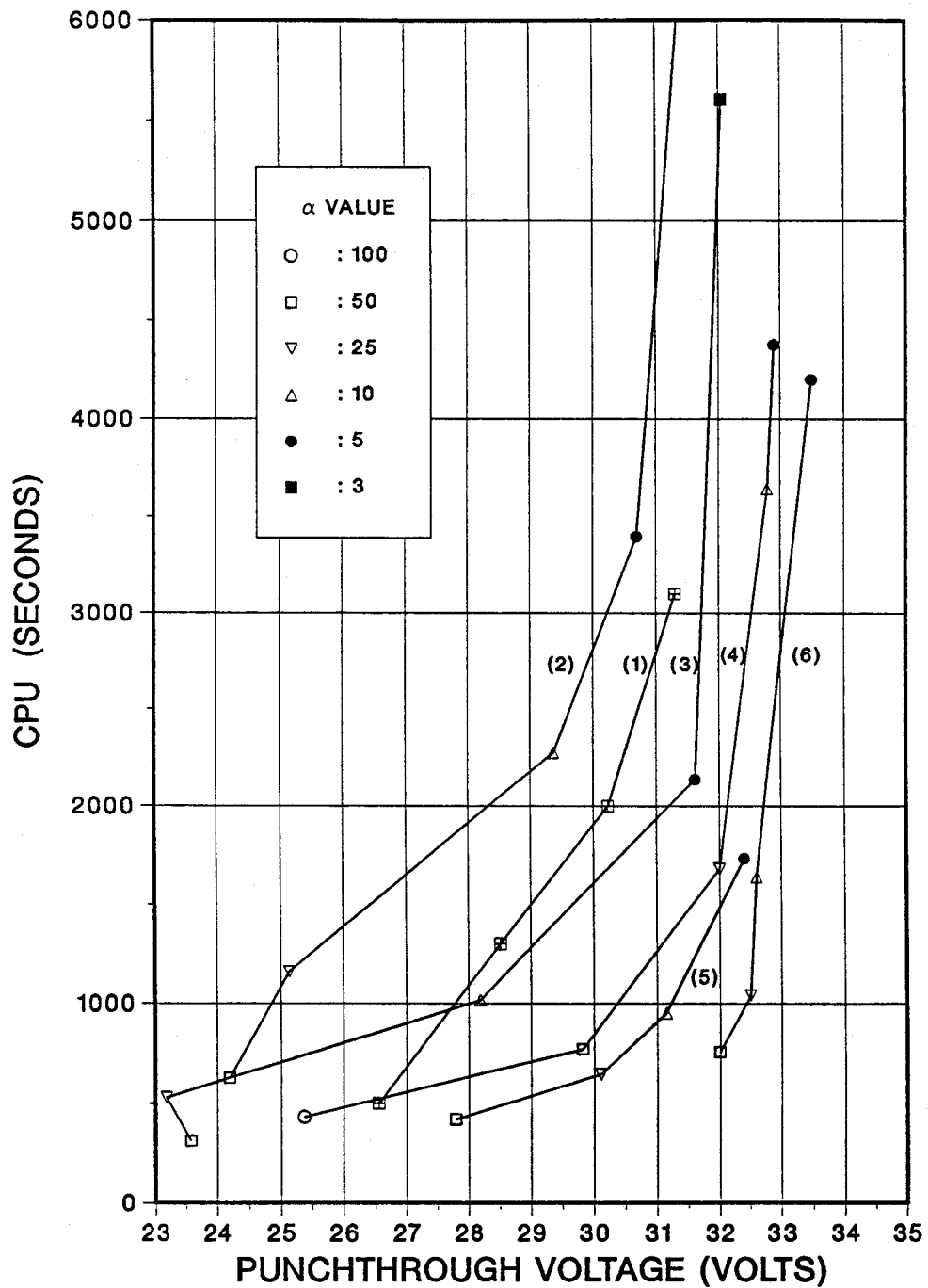


Fig. 9 CPU time needed to obtain a given accuracy (punchthrough voltage) on 2D test example. The value of  $\alpha_0$  for each point is indicated. To obtain a significant comparison, the number of nodes has been limited to 5,000.

### 3.3 Coupled diffusion problem

A more severe test is a coupled diffusion simulation, as in this case the errors for each specy cumulate. Boron is first implanted at an energy of 40 keV and a dose of  $2.5 \cdot 10^{13}$  at/cm<sup>2</sup>, followed by an annealing at 1000°C for 20 min. Arsenic is then implanted at 40 keV with a dose of  $6 \cdot 10^{15}$  at/cm<sup>2</sup>. The final annealing is also performed at 1000°C for 20 min.

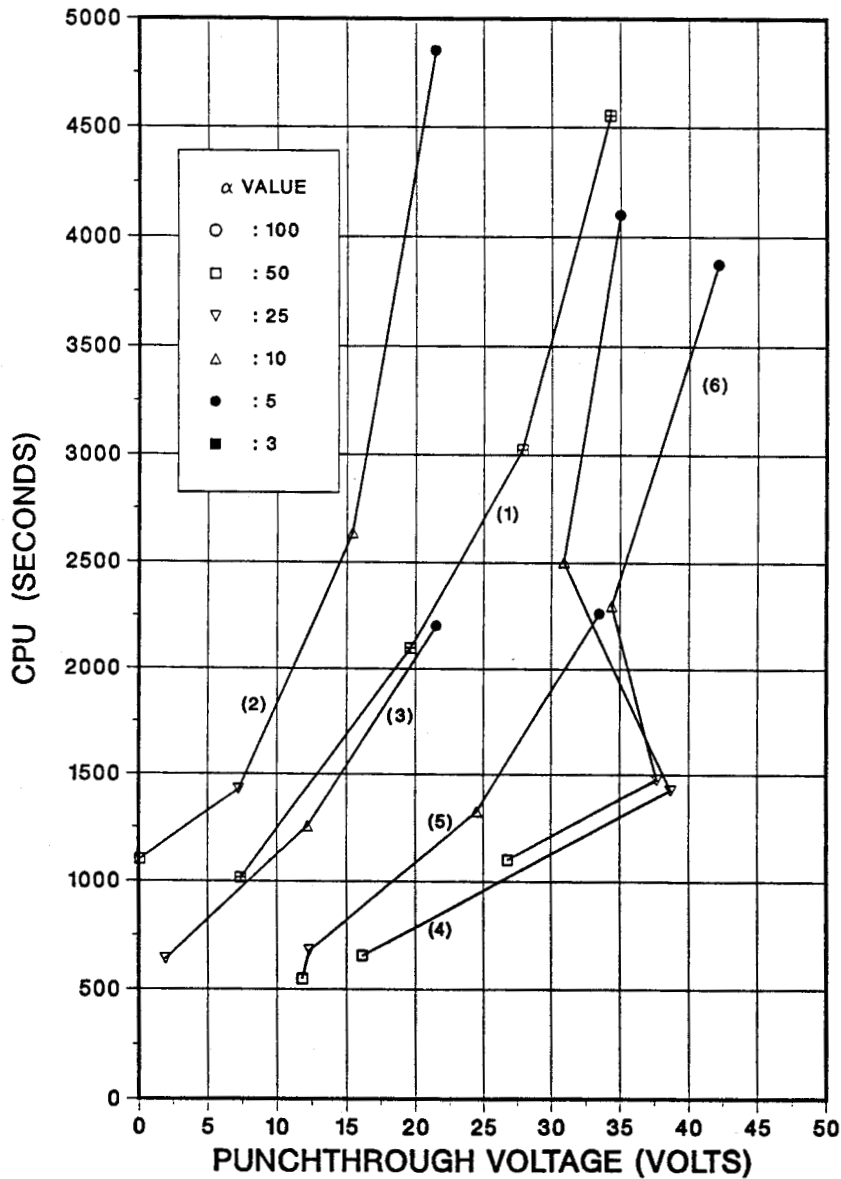


Fig. 10 CPU time versus  $V_{PT}$  for the different schemes (coupled arsenic/boron diffusion).

Figure 10 shows the absolute need for refinement as results evaluate between 1 V and 45 V, the exact value being 50 V (1D calculation). Most of the remarks of the previous test example holds again. However, a close analysis reveals that the good results obtained in curves 4 and 6 for  $\alpha_0=25$  are due to interpolation errors during the refresh procedures. These errors occur mainly in the base/collector junction region. This is confirmed in Fig. 11 where the doping profiles (zoom in the base region) corresponding to curve 6 (for  $\alpha_0=25$  and  $\alpha_0=10$ ) are compared to the most precise available solution (1D simulation). This kind of situation did not appear in the single impurity problem (3.2) since the arsenic gradient prevents such interpolation errors. In any case, results are coherent when  $\alpha_0$  is lower or equal than 10. Again, the "prediction" technique with the automatic evolution of  $\alpha_0$  (curve 6), is efficient in terms of the compromise between precision and CPU requirements.

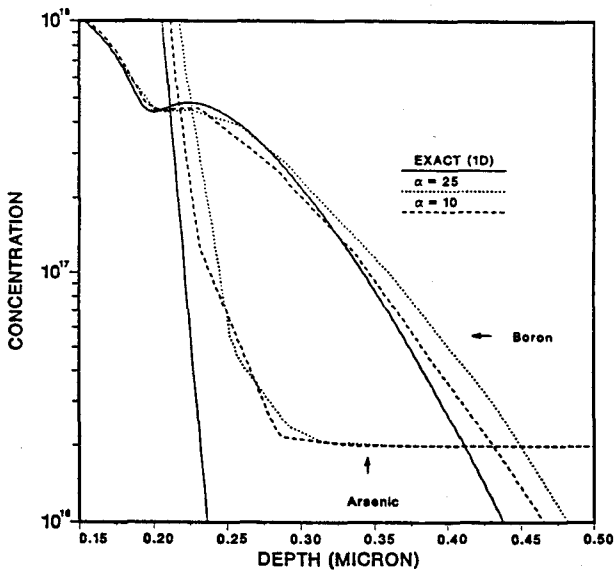


Fig. 11 Comparison between:  
1) best available result (1D)  
2)  $\alpha_0=25$  (curve 6, Fig. 10)  
3)  $\alpha_0=10$  (curve 6, Fig. 10)

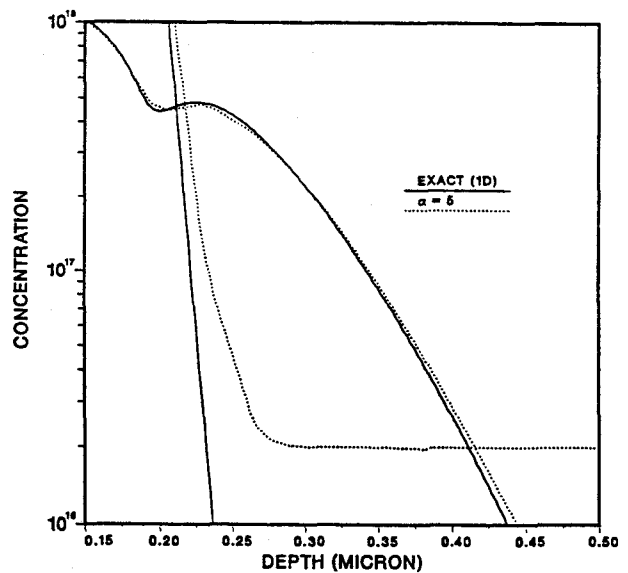


Fig. 12 Comparison between:  
1) best available result (1D)  
2)  $\alpha_0=5$  (curve 6, Fig. 10)

Focusing now on the results obtained with a value of 5 for  $\alpha_0$  (curve 6), it is obvious that a very precise description of the doping profile is reached. This is demonstrated in Fig. 10 and 12. Fig. 13 displays the mesh generated just after a refresh procedure where it can be seen that the high triangle density is located in the emitter/base junction vicinity. The final boron contours are reported on Fig. 14, showing the strong effect of the arsenic gradient on boron diffusion.

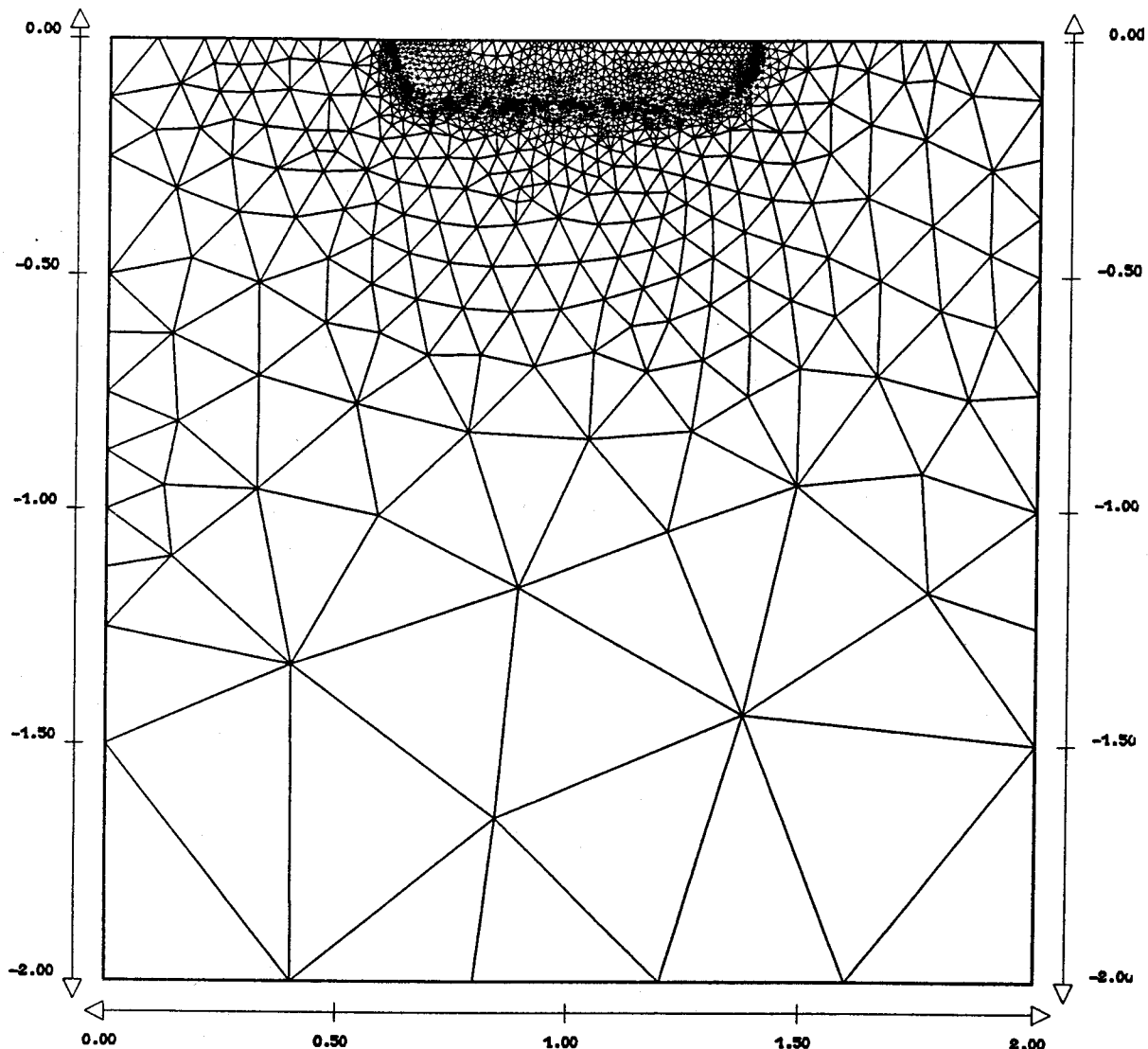


Fig. 13 Mesh generated after a refresh procedure in the case  $\alpha_0=5$  (curve 6,  $V_{PT}=42.15$  V). As the arsenic profile is relatively flat near the surface, strong refinement is only produced in the vicinity of the emitter/base junction.

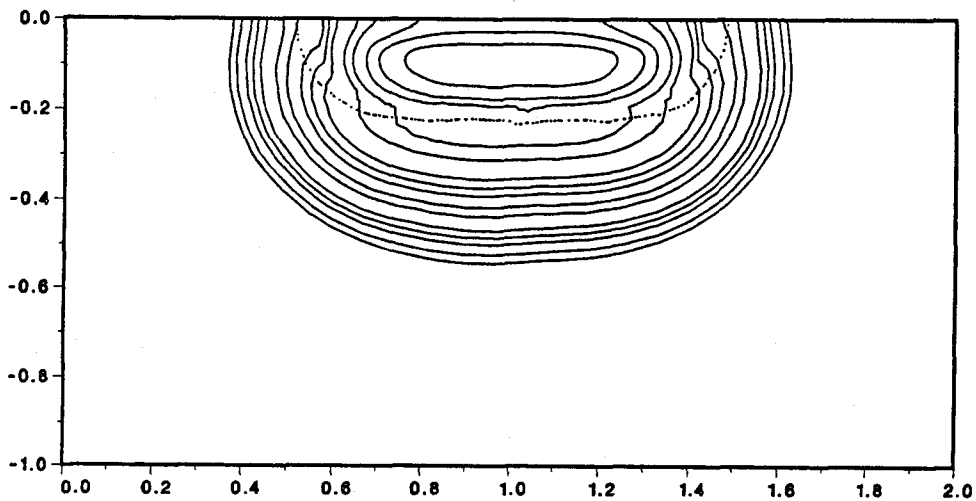


Fig. 14 Boron contours at the end of coupled diffusion.  
The emitter/base junction is also displayed  
(dot line). The final number of nodes is 2794.

In conclusion, the following rules can be proposed:

- \* Very precise results are obtained using the "prediction" method together with a refresh technique where the value of  $\alpha_0$  is adjusted during the refresh.
- \* In this case an initial value for  $\alpha_0$  of 10 allows to simulate with a reasonable accuracy very complex structures where the total number of nodes must be limited: it will be shown in the first application of section IV. On the other hand, the precision of results in 2D seems to be difficultly reached with a standard fixed grid strategy.
- \* In the case of process modeling, a value of 5 allows accurate results, for simple or more complex examples (Fig. 10, 12-14).

## IV. APPLICATIONS

### 4.1 Trench isolated bipolar transistor

Advanced bipolar device is probably one of the most difficult structure to handle in the field of process simulation. On the one hand, segregation effects (and the associated multilayer problem from the numerical point of view) are extremely important [20]. On the other hand, the size of the domains and the precision required on the doping profiles make it very attractive for the test of mesh refinement.

The structure simulated here is taken from [21], using real dimensions concerning topology and the same processing steps. The main characteristics are: deep trench isolation ( $4\ \mu\text{m}$ ) with boron implant at the bottom, a base coupling implant in order to reduce the emitter-to-collector punchthrough current, and finally intrinsic base and emitter are diffused from the same polysilicon layer. Emitter junction depth is of the order of  $500\ \text{\AA}$  and the base width is approximately  $1000\ \text{\AA}$ .

In order to obtain enough precision in this active area, mesh refinement is necessary since the simulation domain needed is at least  $6\ \mu\text{m}$  by  $6\ \mu\text{m}$ . Following the remark of the previous section, a value of 10 is chosen for  $\alpha_0$ . Figure 15 displays the final junctions: the buried layer/channel stopper and P substrate junction, emitter/base and base/collector junctions. Three polysilicon layers (for extrinsic and intrinsic bases formations, trench refill), three oxide layers (thin oxide along the trench, LOCOS and sidewall spacer), are included. Figure 16 shows the boron contours, that exhibit a good stability, for the bases and the channel stopper region.



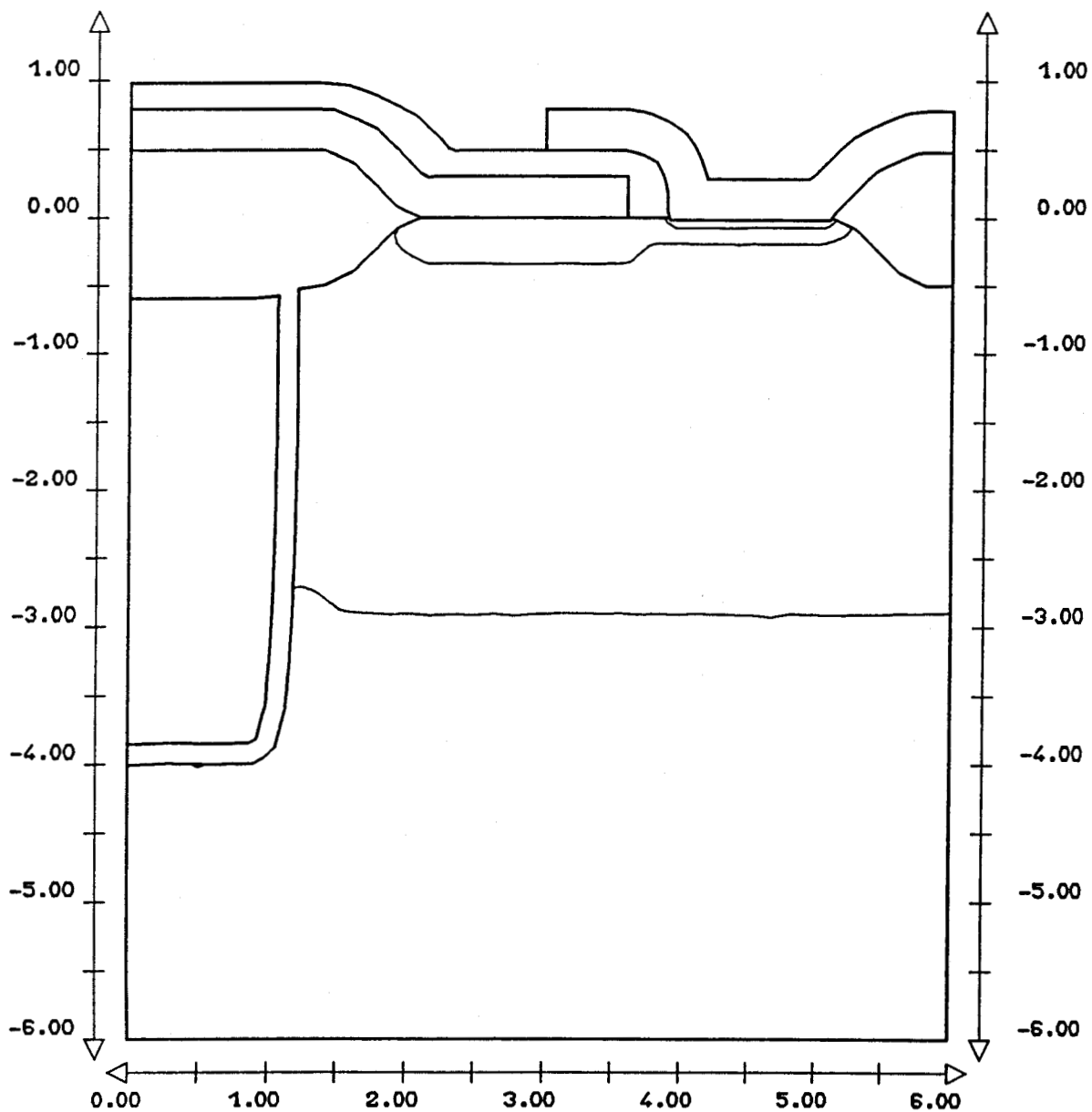


Fig. 15 Structure and junctions of the advanced bipolar simulation.

Despite the important dimensions and the small values of base/emitter and base/collector junctions depths, a good stability is achieved.

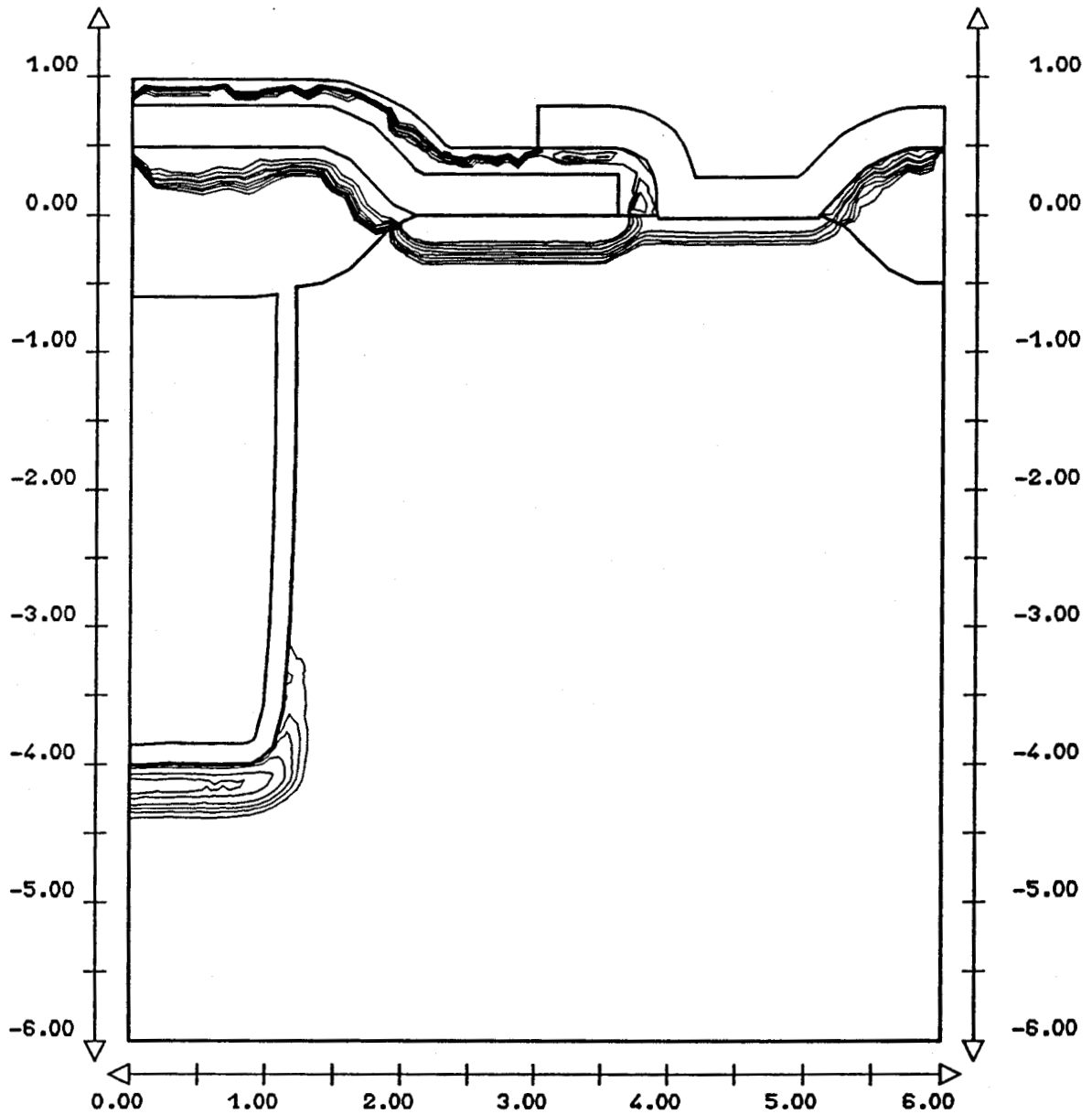


Fig. 16 Boron contours. Exchange of dopants between layers is clearly seen. As mesh refinement is performed in silicon, the mesh is fine in the other layers only in the vicinity of the interfaces with silicon, in order to obtain a good description of segregation effects.

These results are the direct consequence of the refinement method as demonstrated in fig. 17. A medium refinement has been created in silicon at trench bottom, whereas a very important density of triangles is obtained in the extrinsic base region and in the intrinsic part of the device. The buried layer has also an influence on the generated mesh in the middle of the structure. Finally, the conforming triangulation between the different layers is obtained, with the minimal amount of nodes in the other layers than silicon. Obviously, the refinement procedures could be also applied in these layers, but it seems to be unnecessary for the silicon doping profiles precision. The total number of nodes is 4,231 and the total number of elements is 7,384 , for all the layers.

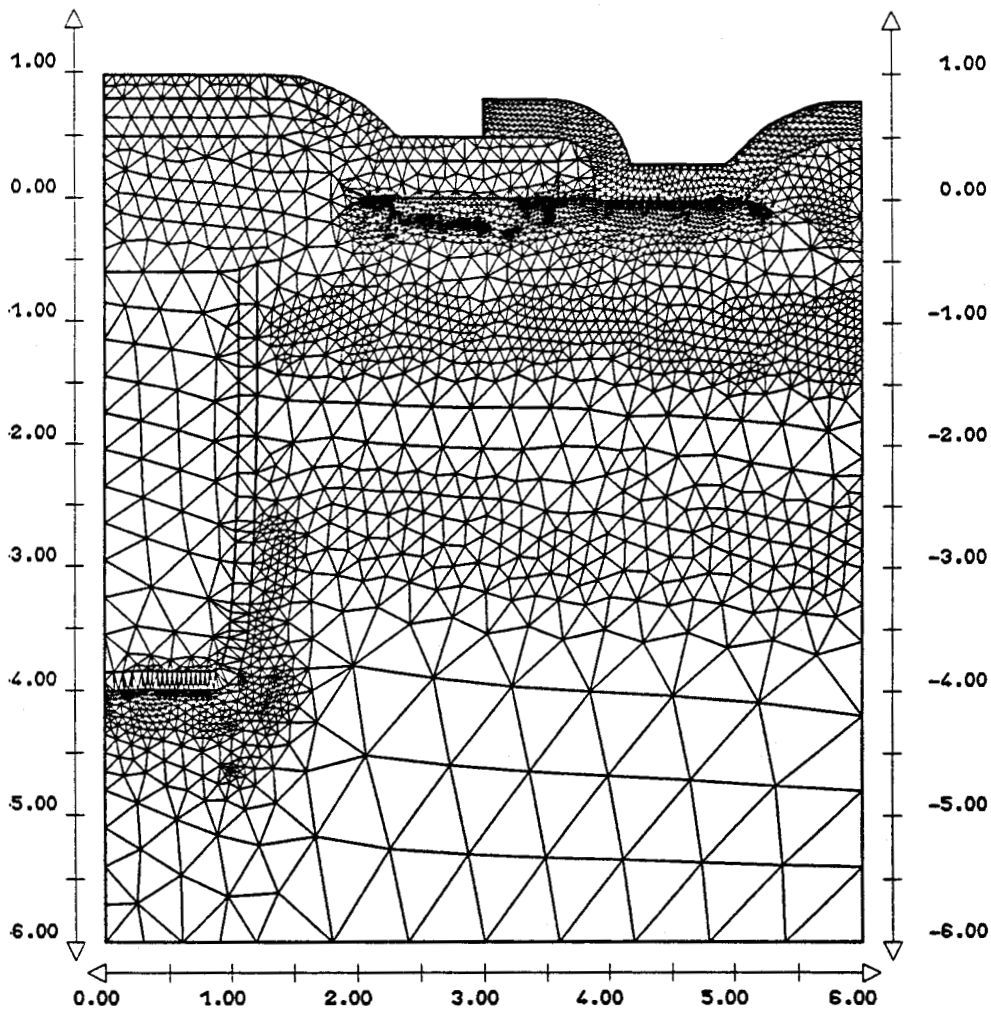


Fig. 17 Final mesh. The refined areas indicate high gradient regions specially in the intrinsic part of the device and for the channel stopper in the bottom trench.

#### 4.2 Coupled process and device approach

The best justification for adaptive mesh refinement arises from the coupled process and device simulation approach. Device optimisation using simulation can only be investigated if a sufficient confidence in process simulation results has been obtained. This is not always achieved, due to the state of the art in process modeling and due to the numerical errors. In order to appreciate the effects of mesh refinement in real cases, the complete simulation of a standard bipolar technology is realised. The IMPACT package is used for this purpose. Other works [22,23] have already described the numerical techniques concerning the interface between process and device domains, in order to minimize the interpolation errors for layer shapes and doping distributions. The process used in this study is the SUBILO-N technology from RTC PHILIPS and has the following characteristics: NPN transistor with a cut-off frequency of 7.5 GHz ( $V_{cb} = 2$  V), a full implanted process and a buried LOCOS isolation technique. Successful comparisons with experiments have been reported, where a fixed grid strategy with a great number of nodes for process simulation was used [24].

In our case, process models and device simulation parameters have been adjusted in 1D calculations and are used in the 2D structures. Figure 18 displays the device structure with the different junctions and the main dimensions.

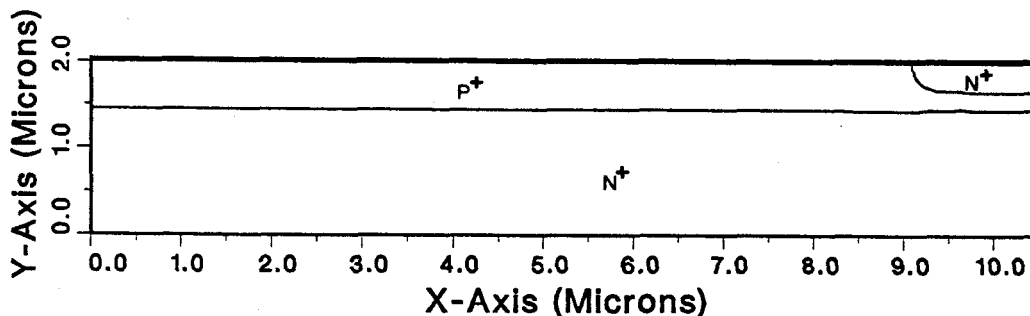


Fig. 18 Bipolar structure used in the coupled process and device approach: N<sup>+</sup> emitter, P<sup>+</sup> base and N<sup>+</sup> buried collector.

IMPACT 3 device simulator is then used to simulate the electrical behaviour of the transistor. One of the most important parameter in bipolar technology is the gain. Figure 19 shows the gain evolution with the collector current for different values of  $\alpha_0$  (the most efficient refresh method determined from section III is used). A good agreement with experiment is obtained when  $\alpha_0 = 5$ , that is coherent with the analytical calculations of punchthrough voltages. For values greater or equal to 25, the simulated characteristics are not satisfactory and hence are not displayed in this figure.

This example shows the efficiency of the methods proposed in section II, since the precision of results obtained in 2D can be compared to the 1D calculations. This allows to simulate complex structures or analyse the effect of process models with a great deal of confidence.

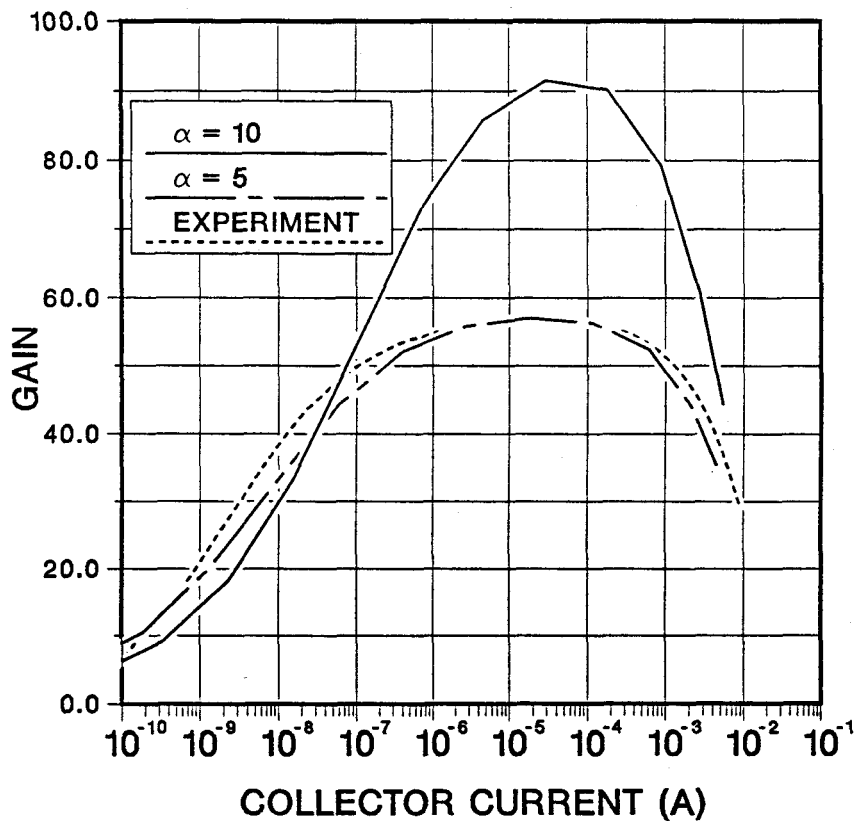


Fig. 19 Gain versus collector current for different values of  $\alpha_0$ .

A comparison with experimental measurements is also provided.

REFERENCES :

- [1] M.E. Law and R.W. Dutton, "Verification of analytical point defect models using SUPREM IV", IEEE Trans. CAD, CAD-7, 181-190, 1988.
- [2] L. Borucki, H.H. Hansen and K. Varahramyan, "FEDSS - A 2D semiconductor fabrication process simulator", IBM J. Res. Develop., vol. 29, n°3, 263-276, 1985.
- [3] R.R O'Brien, C.M. Hsieh, J.S. Moore, R.F. Lever, P.C. Murley, K.W. Brannon, G.R. Srinivasan and R.W. Knepper, "Two-dimensional process modeling: a description of the SAFEPRO program", IBM J. Res. Develop., vol. 29, n°3, 229-241, 1985.
- [4] P. Pichler, W. Jüngling, S. Selberherr, E. Guerrero and H.W. Pötzl, "Simulation of critical IC fabrication steps", IEEE Trans. Electr. Dev., vol ED-32, n°10, 1940-1953, October 1985.
- [5] J. Lorenz and M. Svoboda, "ASWR - method for the simulation of dopant redistribution in silicon", in Proc. Third Int. Conf. on Simulation of Semiconductor Devices and Processes, (G. Baccarani and M. Rudan Eds.), 243-254, September 1988.
- [6] S. Martin, A. Gerodolle and V. Pimont, "Performance evaluation of the TITAN 5 process simulator", in Lecture notes of the software forum, NASECODE VI conf., 168-172, July 1989.
- [7] G.A.J. Amaratunga and R. Imail, "Quasilinear formulation with a simple remeshing scheme for the finite element based simulation of dopant diffusion in silicon", IEE Proceedings, vol. 133, n°6, 221-228, 1986.
- [8] M.J. Baines, C.P. Please and P.K. Sweby, "Numerical solution of dopant diffusion equations", in Proc. Second Int. Conf. on Simulation of Semiconductor Devices and Processes (K. Board and D.R.J. Owen, Eds.), 271-286, 1986.
- [9] P. Ciampolini, A. Forghieri, A. Pierantoni, A. Gnudi, M. Rudan and G. Baccarani, "Adaptative mesh generation preserving the quality of the initial grid", IEEE Trans. CAD, vol. 8, n°5, 490-500, 1989.
- [10] B. Baccus, D. Collard and E. Dubois, "Adaptive mesh refinement for multilayer process simulation", accepted NUPAD III conf., June 3-4 1990, Hawaii.
- [11] B. Baccus, D. Collard and E. Dubois, "Adaptive mesh refinement for multilayer process simulation using the finite element method", submitted IEEE Trans. CAD.

- [12] D. Collard and K. Taniguchi, "IMPACT - a point-defect-based two-dimensional process simulator: modeling the lateral oxidation enhanced diffusion of dopants in silicon", IEEE Trans. Electron Devices, vol. ED-33, 1454-1462, 1986.
- [13] B. Baccus, D. Collard and E. Dubois, "Dynamic mesh refinement for multilayer process simulation", in lecture notes of the software forum, NASECODE VI conf., 108-113, Dublin, Ireland, July 1989.
- [14] M.C. Rivara, "Algorithms for refining triangular grids suitable for adaptative and multigrid techniques", Int. J. Numer. Methods Eng., vol. 20, 745-756, 1984.
- [15] F. Hecht and A. Marrocco, "Internal technical report on MSHPTS", INRIA Rocquencourt, France, Feb. 1986.
- [16] R. Ismail and G. Amaratunga, "Application of local neumann error criteria for remeshing in dopant diffusion problems", in Proc. Third Int. Conf. on Simulation of Semiconductor Devices and Processes, (G. Baccarani and M. Rudan Eds.), 549-560, September 1988.
- [17] N. Shigyo, K. Sato, K. Kato and T. Wada, "TRIMEDES : a triangular mesh device simulator linked with topography/process simulation", The transactions of the IEICE, vol. E 71, n° 10, 992-999, October 1988.
- [18] D. Collard, B. Baccus and E. Dubois, "Two-dimensional multilayer process simulation", VLSI process/device modeling workshop, 15-16, Osaka, May 26-27, 1989.
- [19] D.K. Ferry, L.A. Akers and E.W. Greeneich, "Ultra large scale integrated microelectronics", p. 140, Prentice Hall, 1988.
- [20] B. Baccus, E. Dubois, D. Collard and D. Morel, "Efficient two dimensional multilayer process simulation of advanced bipolar devices", Solid-State Electronics, Vol. 32, No 11, 1013-1023, November 1989.
- [21] T. Yamaguchi, Y.-C.S. Yu, E.E. Lane, J.S. Lee, E.E. Patton, R.D. Herman, D.R. Ahrendt, V.F. Drobny, T.H. Yuzuriha and V.E. Garuts , "Process and device performance of a high-speed double poly-si bipolar technology using borosenic-poly process with coupling-base implant", IEEE Trans. Electron Devices, vol ED-35, n° 8, 1247-1256, August 1988.
- [22] E. Dubois, J.L. Coppée, B. Baccus and D. Collard, "A study of the electrical performances of isolation structures", accepted IEEE Trans. Electron Devices.

- [23] E. Dubois, J.L. Coppée, B. Baccus and D. Collard, "Electrical performances evaluation of isolation structures by coupled process and device simulation", in Proc. Third Int. Conf. on simulation of semiconductor Devices and Processes, (G. Baccarani and M. Rudan Eds.), 151-161, September 1988.
- [24] E. Dubois, B. Baccus and D. Collard, "Analysis of the punchthrough effect in walled emitter bipolar transistors", in Proc. ESSDERC conf., 789-792, September 1989.



## CHAPTER V

### APPLICATIONS

#### Abstract

In this chapter, three applications are described. The first one concerns the simulation of an LDDMOS transistor, including oxide sidewall spacer formation and gate re-oxidation. This first result shows the complete integration of the models and numerical procedures reported in chapter III.

A VLSI bipolar technology is simulated in the second application. We focus our attention on the doping profiles in the vicinity of the fully recessed oxide, in a walled emitter configuration. Conditions for a base thinning formation are presented.

The last example describes an advanced bipolar technology, including doping from polysilicon and trench isolation. This allows to study the influence of the sidewall spacer width on the link between intrinsic and extrinsic bases.

## I. LDD-MOS TRANSISTOR

### 1.1 Introduction

In VLSI circuits, the dimensions of the devices are continually decreased to achieve higher density and speed. In order to maintain acceptable characteristics, scaling rules are used for the down-scaling of devices. For MOS technology, the channel length, junction depth and gate oxide thickness are scaled down, while the channel doping is scaled up. In this scheme, the power supply voltage is not changed, resulting in a higher channel electric field. Apart from the classical short channel effect (decrease of the threshold voltage with decreasing channel length), it has dramatic effects on device performances. High energy electrons ("hot electrons") may have enough energy to cross the silicon/silicon oxide interface and to create damage at this interface. The exact mechanisms are not well understood, but it results in threshold voltage shift and reduction of electron mobility.

One of the main alternative to this problem is to change the source/drain doping profiles by introducing  $n^-$  regions [1]. This lightly doped region (hence the modified device is called a LDD MOS transistor) drops off some drain voltage and reduces the drain electric field. The optimisation of such doping profiles (or other kinds of advanced MOS structures [2]) is greatly facilitated by the systematic use of coupled process/device simulations [3-4].

### 1.2 Simulation results

This first complete simulation shows the application of IMPACT4 to a standard LDD MOS device. Starting from a P-Well substrate, the channel profile is formed by a phosphorus implant ( $15 \text{ keV}$ ,  $7 \cdot 10^{11} \text{ at/cm}^2$ ) followed by a diffusion in neutral ambient ( $950^\circ\text{C}$ , 40 Min). The gate oxide has a thickness of  $185 \text{ \AA}$  ( $900^\circ\text{C}$ , 60 min in dry ambient). The polysilicon gate and a thin oxide layer are then deposited and an RIE etching determines the

gate structure. Phosphorus is implanted to form the LDD region ( $4 \cdot 10^{13}$  at/cm<sup>2</sup> at 20 keV) . Fig. 1 shows the phosphorus contours after a small annealing step (900°C, 30 min). The interesting points here are the polysilicon gate edge profile, showing the result of dry etching. It can also be seen that, as polysilicon was *in situ* doped with phosphorus (constant level at  $10^{19}$  at/cm<sup>3</sup>), this impurity is slightly incorporated in the gate oxide on the left part of the structure by static segregation.

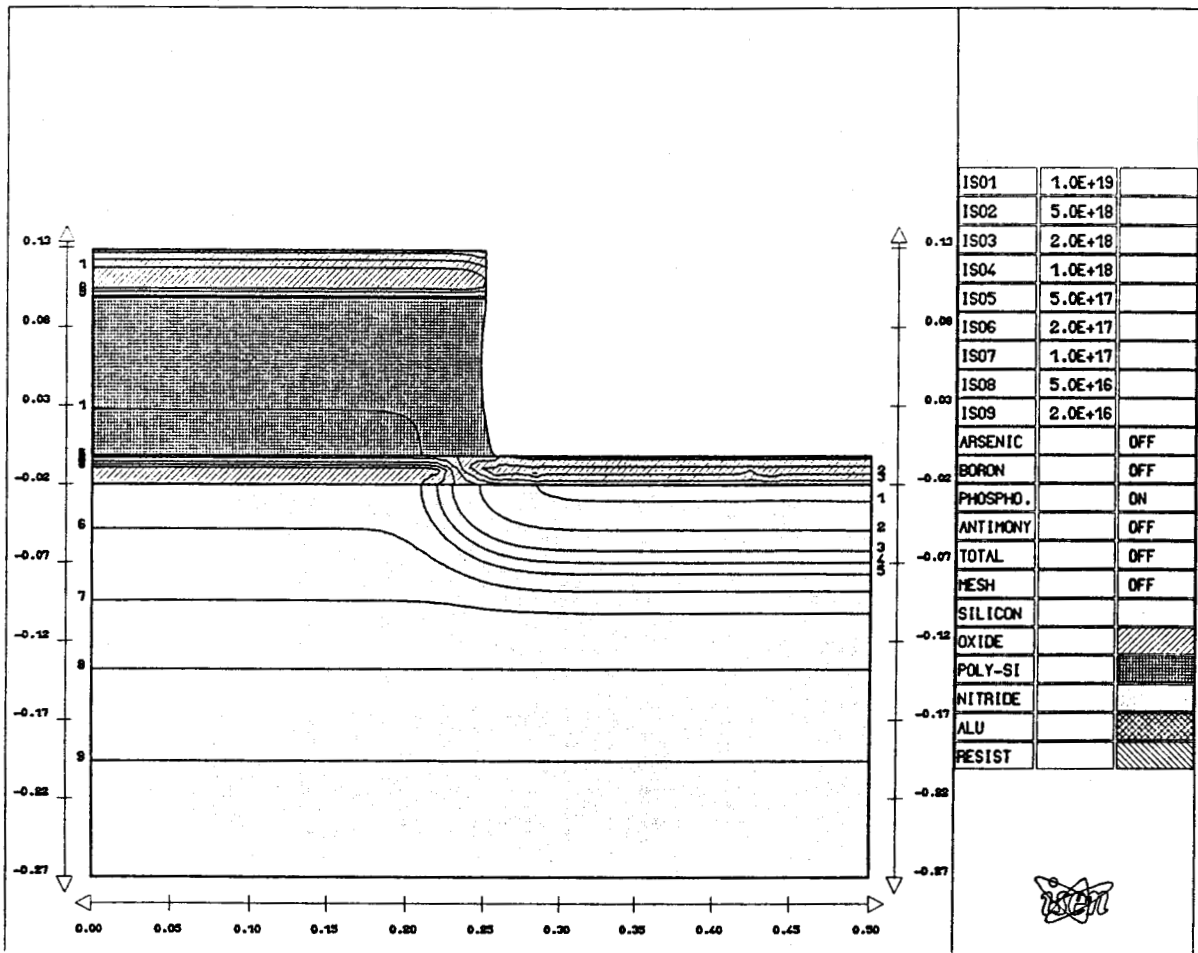


Fig. 1 Phosphorus contours after LDD ion implantation and annealing.  
 Note that the upper oxide layer acted as mask for the RIE polysilicon gate etching.

In a standard manner, the oxide sidewall spacer is formed by a conformal oxide deposition (layer thickness:  $0.05 \mu\text{m}$ ), followed by an anisotropic etching ( $0.07 \mu\text{m}$ ). Arsenic is then implanted to form the source/drain regions ( $40 \text{ keV}$ ,  $5 \cdot 10^{15} \text{ at/cm}^2$ ). Fig. 2 displays the arsenic contours, where it can be seen that we have made use of the depth dependent lateral standard deviation modeling. The final diffusion is performed at  $900^\circ\text{C}$ , for 84 min in dry ambient. In this case we take advantage of the numerical formulation of oxidation, since it is nearly impossible to use analytical formula of simplified oxide displacements. Fig. 3 and 4 display the arsenic and phosphorus countours in the final structure, respectively.

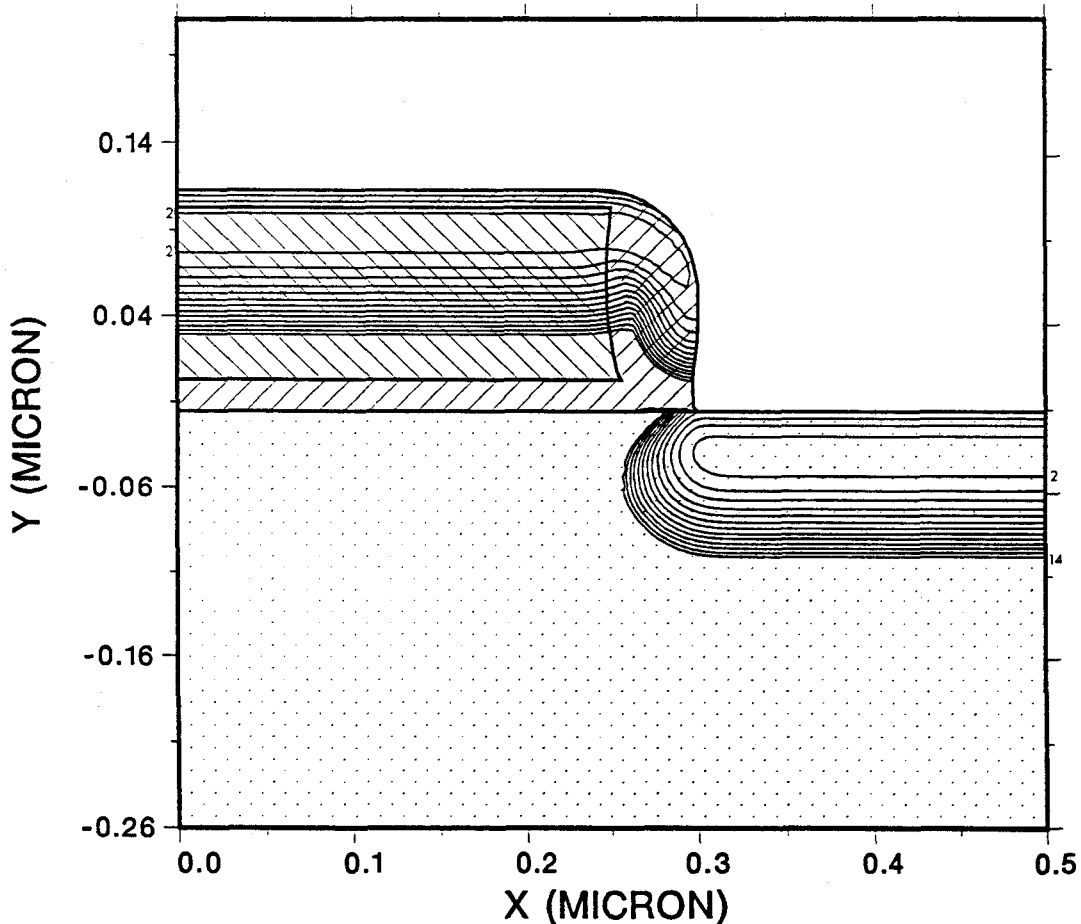


Fig. 2 Arsenic contours after source/drain ion implantation.

1:  $3 \cdot 10^{21}$ , 2:  $1 \cdot 10^{21}$ , 3:  $3 \cdot 10^{20}$ , 4:  $1 \cdot 10^{20}$ .....  
14:  $1 \cdot 10^{15} \text{ at/cm}^3$ .

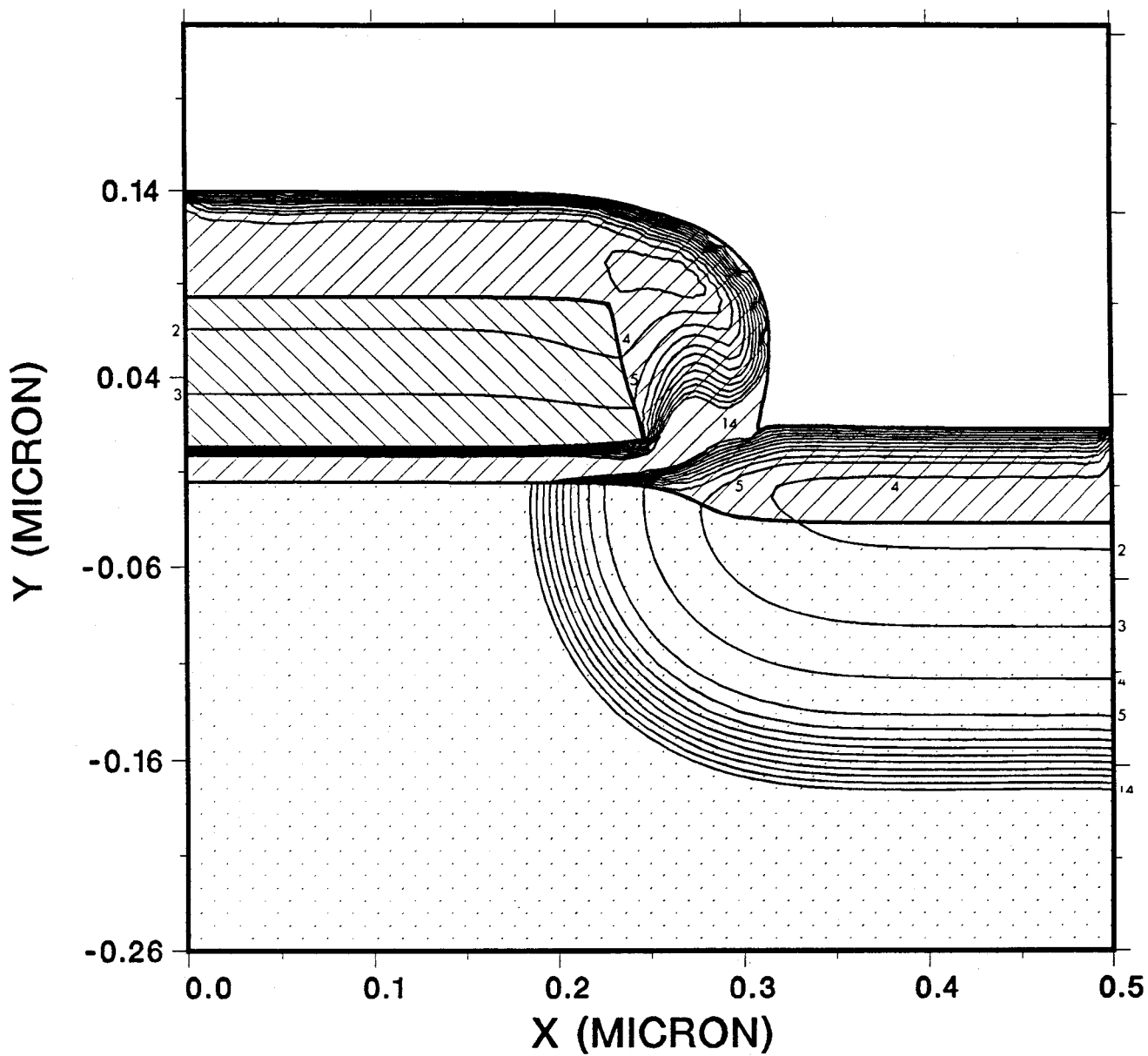


Fig. 3 Arsenic contours at the end of the simulation.

1:  $3 \cdot 10^{21}$ , 2:  $1 \cdot 10^{21}$ , 3:  $3 \cdot 10^{20}$ , 4:  $1 \cdot 10^{20}$ .....  
14:  $1 \cdot 10^{15}$  at/cm<sup>3</sup>.

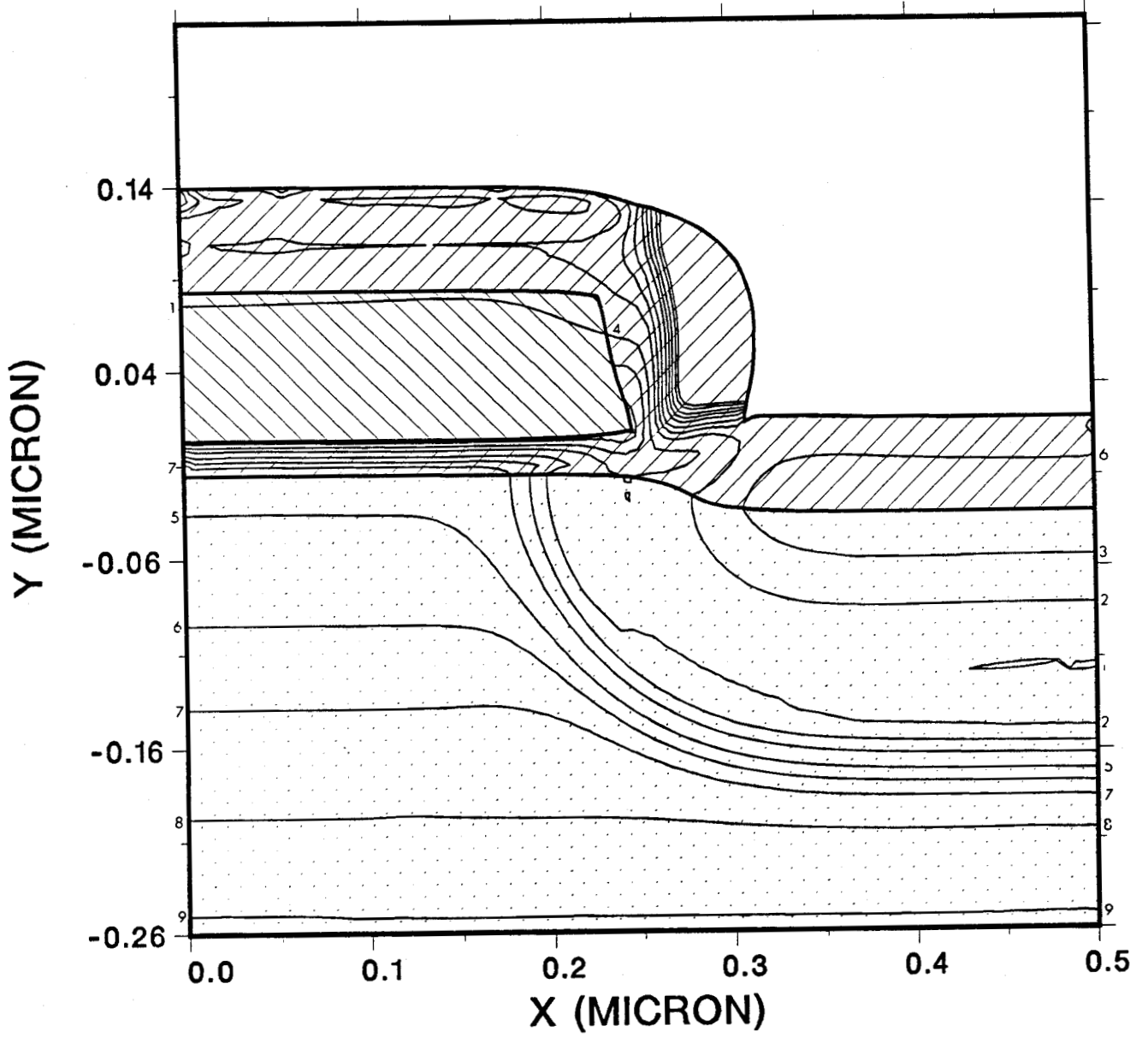


Fig. 4 Phosphorus contours at the end of the simulation.  
1:  $5 \cdot 10^{18}$ , 2:  $2 \cdot 10^{18}$ , 3:  $1 \cdot 10^{18}$ , 4:  $5 \cdot 10^{17}$ , 5:  $2 \cdot 10^{17}$ ,  
6:  $1 \cdot 10^{17}$ , 7:  $5 \cdot 10^{16}$ , 8:  $2 \cdot 10^{16}$ , 9:  $1 \cdot 10^{16}$ , 10:  $5 \cdot 10^{15}$ ,  
11:  $2 \cdot 10^{15}$ , 12:  $1 \cdot 10^{15}$  at/cm<sup>3</sup>.

By examining these figures, several remarkable features are outlined below:

- 1) The viscous flow of oxide results in a slight overhanging structure at the polysilicon gate edge. Moreover this later layer has been more rapidly oxidized at the upper corner than at the lower one, because there was only a very thin layer of oxide above polysilicon (Fig. 2).
- 2) The overall boundary conditions are efficiently solved in the complete structure. In particular, arsenic has diffused in polysilicon and has reached the polysilicon/oxide interface (Fig. 3). This is also confirmed by two cross section in the left and right parts of the simulation domain (Fig. 5 and 6 respectively).
- 3) The isovalues of arsenic and phosphorus are very stable in the oxide, showing the efficiency of the coupled oxidation/diffusion calculation. Especially, the arsenic contours are the reflect of the oxide deformation during growth and of out-diffusion from the gate.

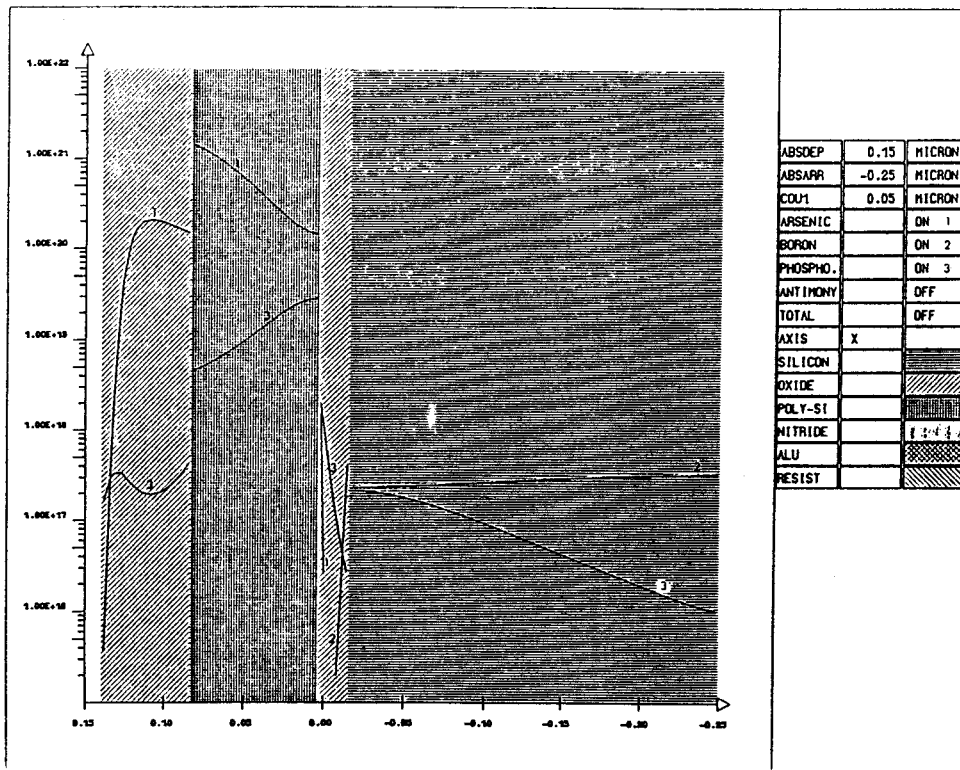


Fig. 5 Cross-section in the left part of the simulation domain (channel region).

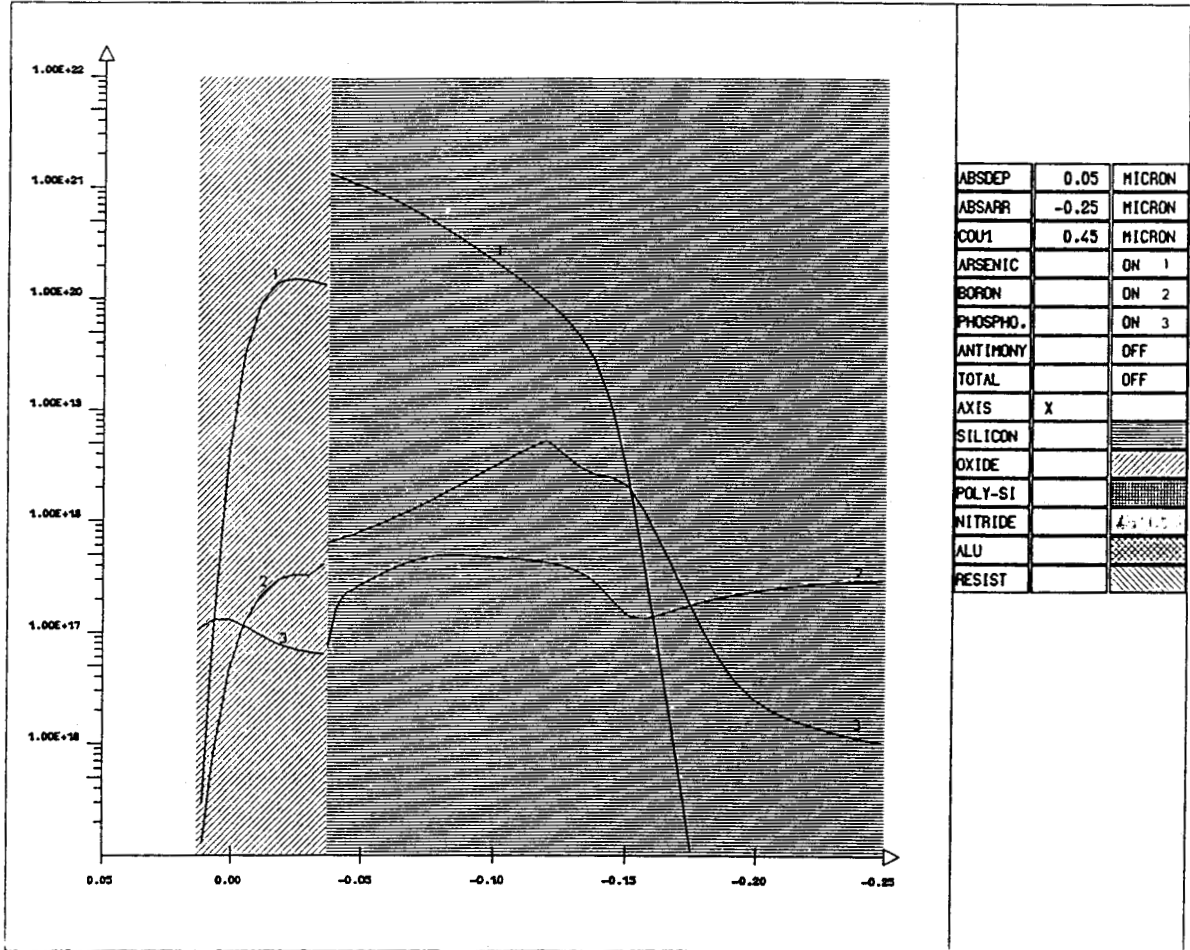


Fig. 6 Cross-section in the right part of the simulation domain. The segregation ratio at the silicon/silicon oxide interface are always obtained. Strong coupling between the impurities can also be seen in silicon.



## II. PARASITIC EFFECTS IN WALLED EMITTER BIPOLAR TRANSISTORS

### 2.1 Introduction

#### 2.1.1 Definition of the problem

It is well-known that high packing density in bipolar integrated circuits is improved by the systematic use of walled emitter transistors [5]. However, both vertical and lateral scalings require a severe control of the process steps, especially when isolation is assumed by a fully-recessed local oxidation of silicon. Indeed, the major drawback related to this technique is its extreme sensitivity to emitter/collector leakage currents caused by a local punchthrough effect.

The goal of this study is to determine:

- i) the key processing steps critical for the punchthrough control.
- ii) the exact location of possible base thinning (and hence of punchthrough current path). Two regions are concerned *a priori*: the silicon bulk near the bird's beak and in the vicinity of the silicon/silicon oxide interface (segregation effects).

For this aim, we have used a coupled process/device approach and comparisons with experimental characteristics [6-8]. In this section we will focus on the process aspects and the simulation results of IMPACT 4.

### 2.1.2 Process description

The processing steps flowchart is given in table I and is typical of a VLSI bipolar technology. In our case, the processing parameters are deduced from the Subilo-N technology (RTC PHILIPS Components) [6].

1 - Buried layer ion implantation and diffusion (Sb)
2 - Epitaxial growth (As)
3 - Channel-stop ion implantation (B)
4 - Buried LOCOS oxidation
5 - Deep N ion implantation (P) for collector contact
6 - Dry oxidation and base ion implantation (B)
7 - Inert base drive-in
8 - Oxide etching
9 - Emitter ion implantation (As)
10 - Final inert annealing

Table I Process flowchart.

### 2.2 Simulations results

The modeling and algorithms used during these simulations have been already described in chapters II and III. The only special point concerns the 2D oxide modeling. In the case of fully recessed oxide, large stress develop during the growth, affecting significantly the bird's beak shape. In order to reproduce as accurately as possible the experimental shapes, fitted analytical formula have been used.

2.2.1 Base thinning formation

A typical 2D result is displayed in Fig. 7 for the boron contours at the end of the process. The channel stopper whose diffusion is significantly reduced by the high antimony concentrations (buried layer) is shown on the right part under the thick oxide. Concerning the base region, segregation effects occur at the Si/SiO<sub>2</sub> interface and results in a slight depletion. As arsenic has an opposite behaviour concerning segregation, a base thinning may occur under the bird's beak, just at the the Si/SiO<sub>2</sub> interface. It will be shown that this situation does not occur.

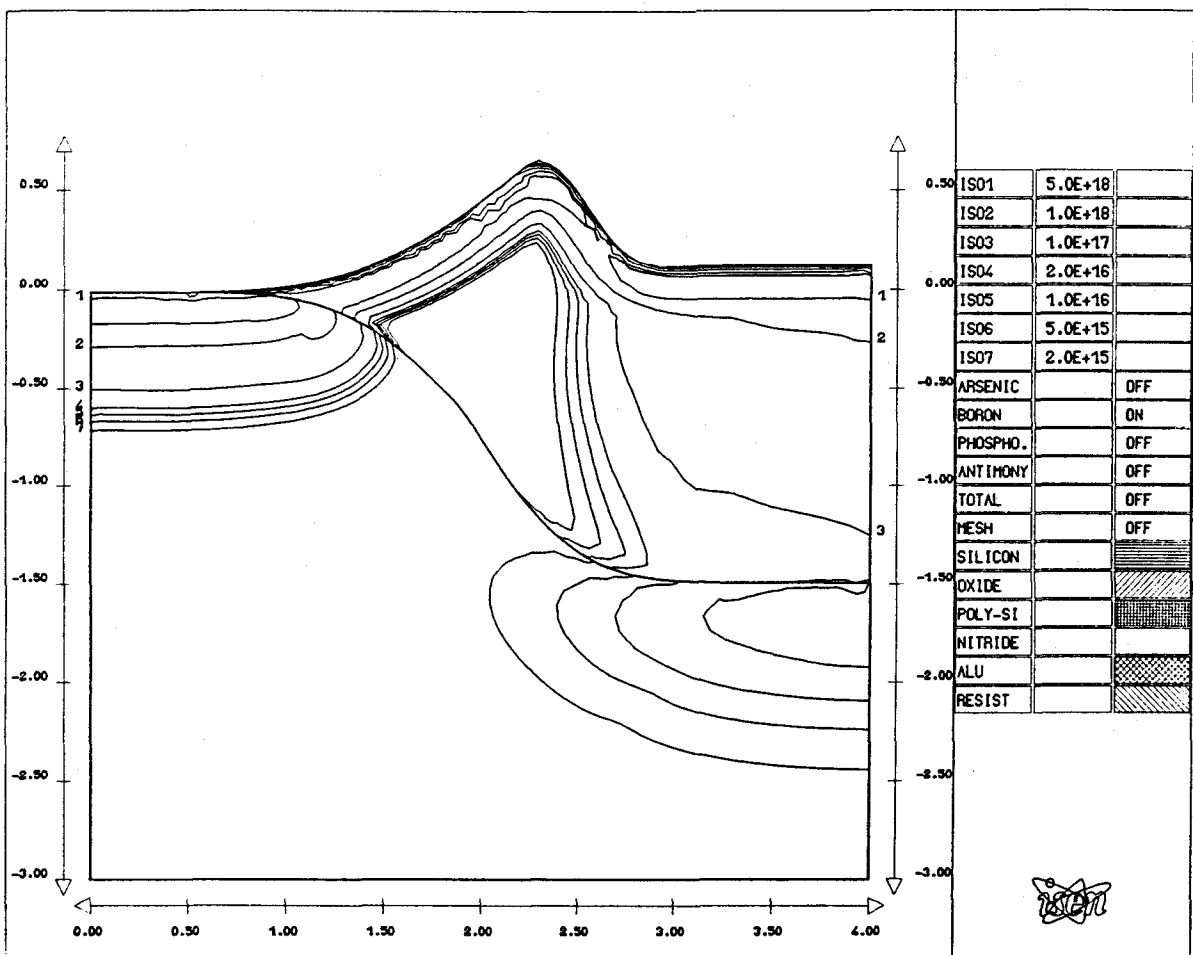


Fig. 7 Boron contours of the walled emitter bipolar transistor.

The masking effect (for boron ion implantation) of the thick oxide is obvious. The shape of the isovalue at a level of concentration of  $1. \cdot 10^{18}$  at/cm<sup>3</sup> is the reflect of coupling with the arsenic emitter during the final diffusion step. Boron incorporated in the oxide during the fully recessed oxide growth is also displayed.

In order to explain the base thinning formation, Fig. 8 shows the structure under study with the different junctions. Numerous calculations have shown that two regions can be distinguished: i) an intrinsic region whose doping profiles are not altered by the bird's beak and ii) a base thinning region. In this latter area, the base width is mainly determined by the field oxide shapes. This geometrical effect can also be seen in Fig. 9 where two cross-sections are performed in these two regions.

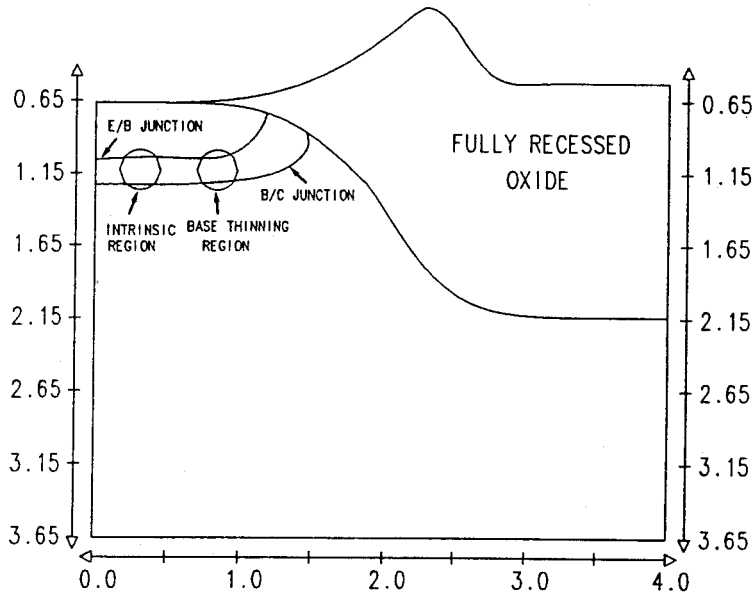


Fig. 8 Base thinning effect and junction locations.

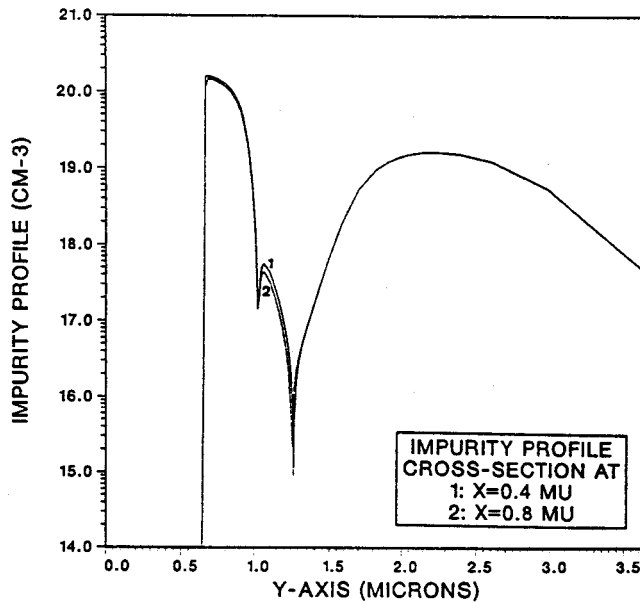


Fig. 9 Doping profiles cross-sections in the intrinsic region and in the base thinning region.

2.2.2 Sensitivity to oxide over-etch

In order to understand some experimental dispersions involving punchthrough effects (Fig. 10), a systematic set of simulations has been undertaken and it has been found that the etching step (8 in the flowchart) induces dramatic changes in the doping profiles. Fig. 11 displays the junctions locations for several values of over-etching: 0, 200 and 400 Å. This latter value is probably over-estimated but allows to understand the influence of an insufficient control of this etching step. As the over-etching increases, the base is narrower, since arsenic penetrates deeper in the substrate during the ion implantation step 9.

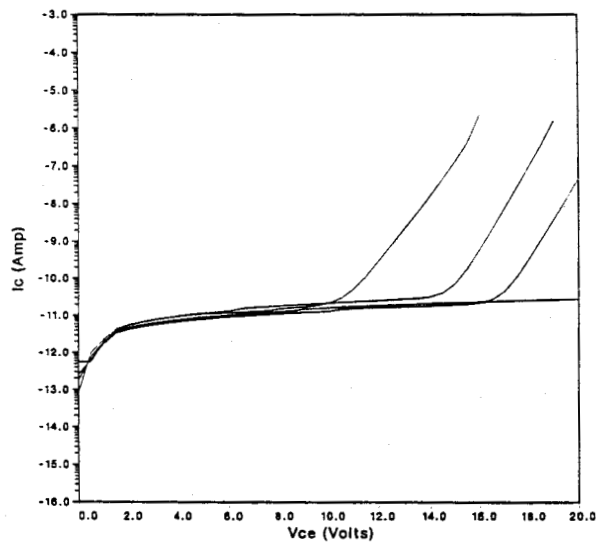


Fig. 10 Measured  $I_c=f(V_{c_e})$  characteristics revealing drastic reduction of the punchthrough voltage, for different wafers.

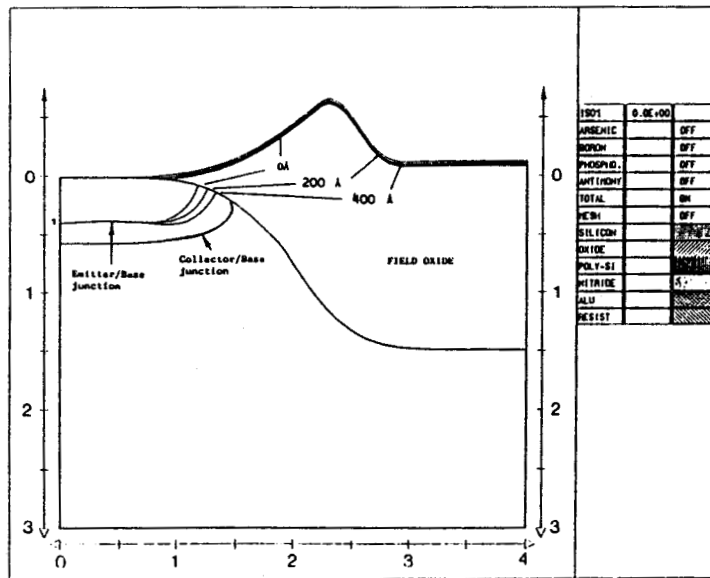


Fig. 11 Junction locations for several over-etching (step 8).

Interesting informations are deduced from Fig. 12 and 13, representing the doping concentrations for each specy along the silicon surface and the Si/SiO<sub>2</sub> interface for 0 and 200 Å of over-etch respectively. Fig. 13 confirms the decrease in base Gummel number (here at the silicon/silicon oxide interface) for the 200 Å case as suggested in Fig. 11. On the other hand, the slight increase of arsenic concentration under the bird's beak (epitaxial layer) occurs during the formation of the thick oxide, but this phenomenon does not affect the active part of the device. Finally, the antimony concentration of the buried layer is also plotted.

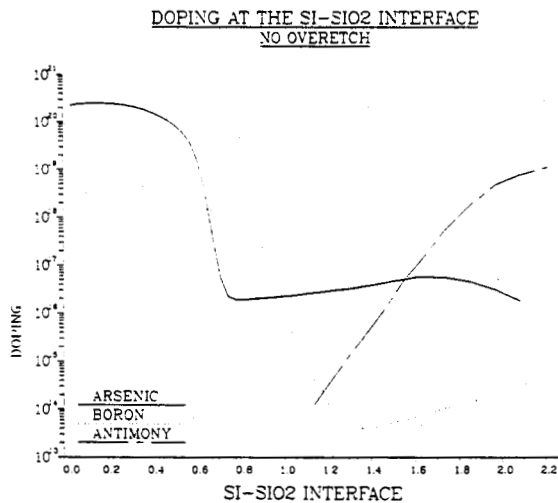


Fig. 12 Doping at the silicon/silicon oxide interface. No overetching has been simulated.

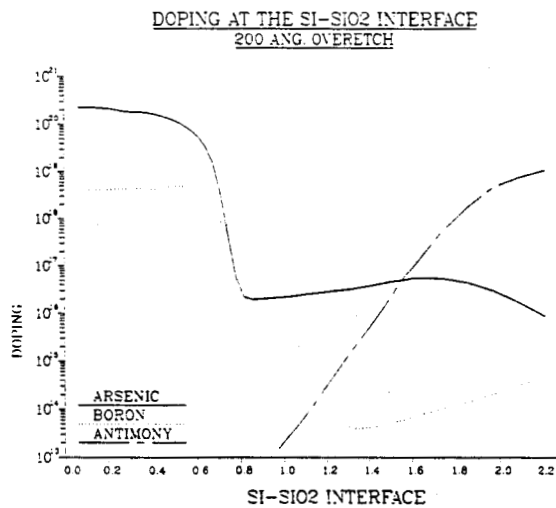


Fig. 13 Same figure in the case of 200 Å of over-etch.

Moreover, Fig. 14 shows the evolution of boron concentration at the emitter/base junction for these three different over-etching values. This concentration decreases with the amount of overetch and can be lower in the base thinning region (Fig.11) than in the intrinsic region (400 Å case), enhancing the probability that the punchthrough current is initiated in the silicon bulk and not at the silicon/silicon oxide interface.

The segregation effect can also be seen on the right part of these curves as the concentration decreases within a short distance, but this effect is not detrimental since the resulting concentration at the interface is always greater than in the intrinsic or "base thinning" region.

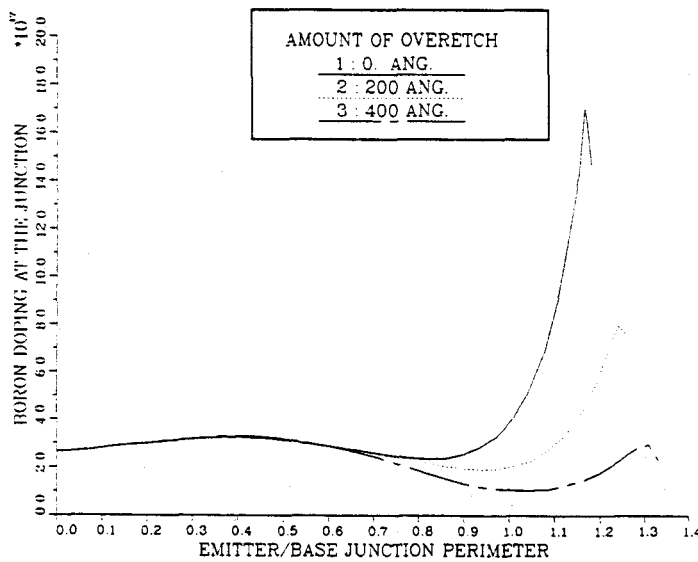


Fig. 14 Boron doping concentration along the emitter/base junction, for three over-etching values.

However, it should be noted that the location of this base thinning region can be displaced when modifying the bird's beak shape. In the case of steep oxide profile, the most severe base thinning occurs along the thick oxide/silicon interface. Such behaviour will be shown for a trench isolated transistor (section III).

This first approach, involving numerous process simulations, shows clearly the location of the base thinning region. Moreover, it seems that a good control of the etching step, just before arsenic ion implantation, is of prime importance. These points are confirmed by using device simulations.

### 2.2.3 Device simulations

This kind of study is a typical case where a coupled process and device simulation approach may be very useful and provide numerous informations. One of the prerequisite for a sufficient confidence in the final results, is an efficient and accurate interface between the two simulators. The grid generation technique and the numerical approach have been discussed in [9]. The electrical behaviour has been computed with an extended version of the 2D device simulator IMPACT 3 [10].

$I_c$ - $V_{ce}$  characteristics have been computed in the case of 0 and 200 Å of over-etch. In this configuration, emitter and base are grounded. The large discrepancies of the punchthrough characteristics of Fig. 10 are fully coherent with the computed currents (Fig. 15). Moreover, the electrons and holes contours for a 200 Å over-etched device at  $V_{ce}=17$  Volts are displayed in Fig. 16. These two figures demonstrate clearly that the punchthrough current path corresponds to the base thinning region evidenced in 2.2.2. A careful control of the critical etching step has also experimentally shown the validity of our simulations [11].

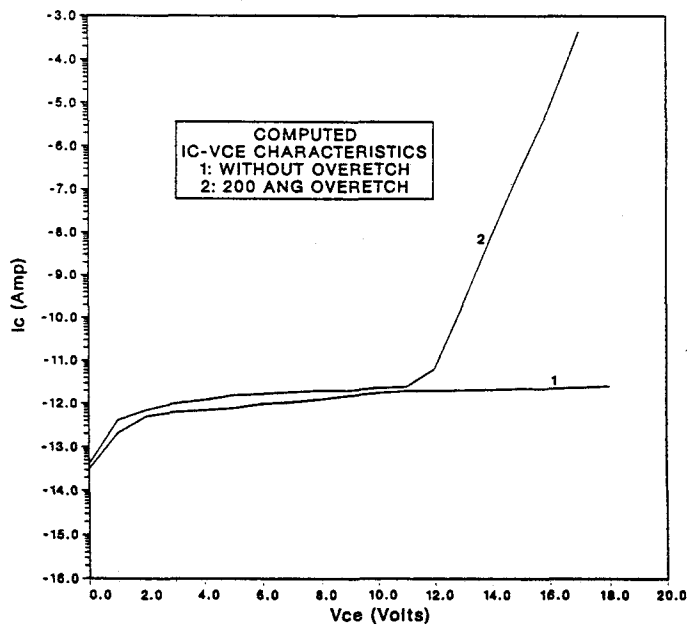


Fig. 15  $I_c=f(V_{ce})$  characteristics for 2 devices with and without oxide over-etch.  
 $V_e = V_b = 0$ .





HOLE DENSITY  
DEVICE WITH  
200 ANG. OVERETCH  
VCE=17 VOLTS

1:	1.0E+18
2:	1.0E+17
3:	1.0E+16
4:	1.0E+15
5:	1.0E+14
6:	1.0E+13
7:	1.0E+12

ELECTRON DENSITY  
DEVICE WITH  
200 ANG. OVERETCH  
VCE=17 VOLTS

1:	1.0E+18
2:	1.0E+17
3:	1.0E+16
4:	1.0E+15
5:	1.0E+14
6:	1.0E+13
7:	1.0E+12

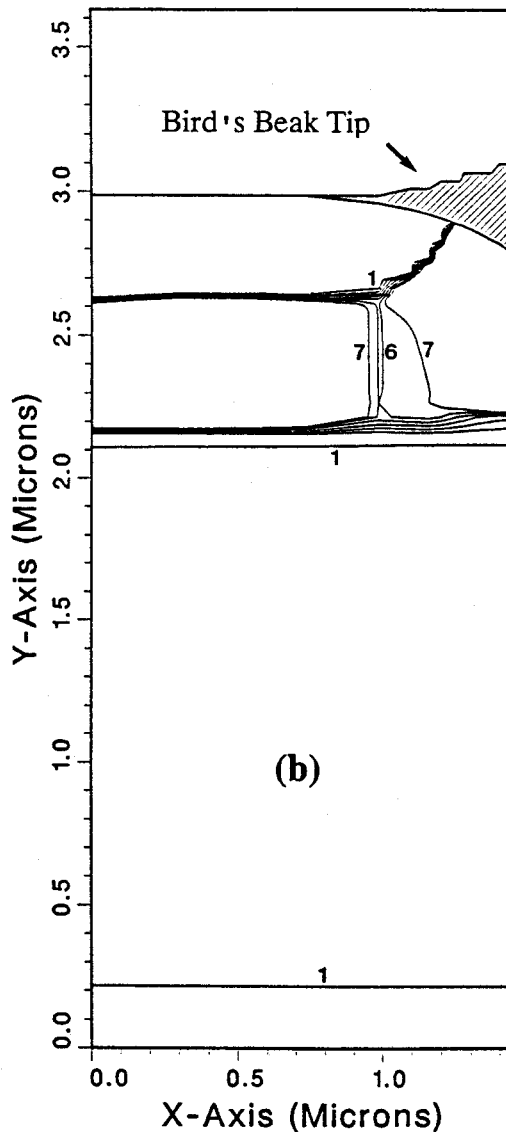
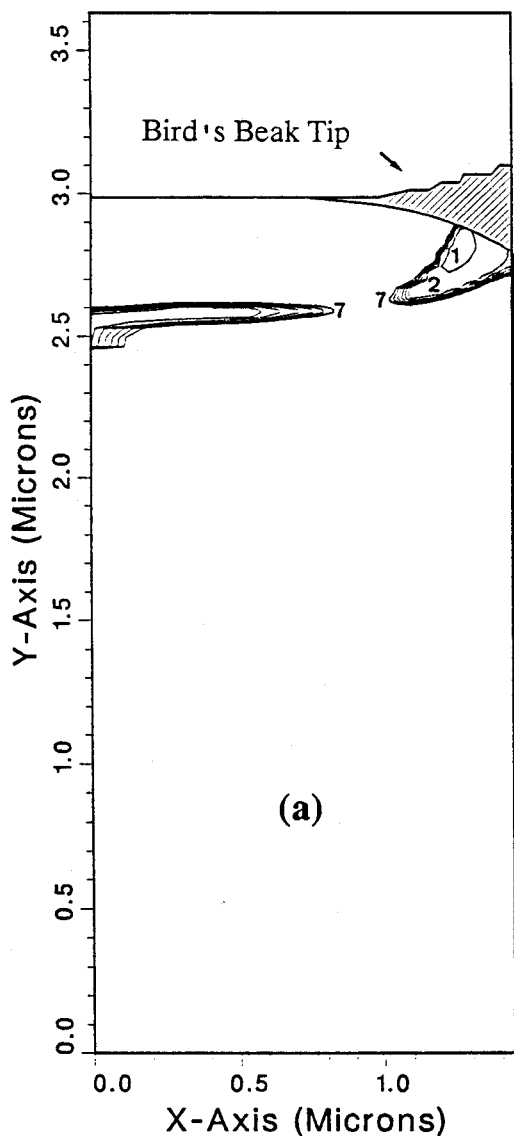


Fig. 16 (a) Hole and (b) electron contours for a device with 200 Å over-etch.  $V_{ce} = 17$  Volts.

### III. APPLICATION TO A POLYSILICON, TRENCH ISOLATED BIPOLAR TRANSISTOR [12]

#### 3.1 Modern bipolar technologies

Bipolar technology has experienced large improvements during the last few years, due partly to the introduction of processing steps coming from CMOS technologies. High performances have been reached when including the following features [13]:

- i) Use of thin polysilicon films for emitter contact. These layers act also as doping source for the diffusion of dopants in silicon, resulting in very shallow and steep  $n'$  emitter doping profiles.
- ii) Use of self-aligned emitter/base configurations. In these schemes, external base resistance and base/collector junction capacitance are significantly reduced.
- iii) Parasitic capacitances are further reduced by the use of deep trench (or U-groove) isolation structures.

Fig. 17 shows one integration of these characteristics [14]. Transit frequencies of more than 25 GHz have been already obtained using different versions of these features.

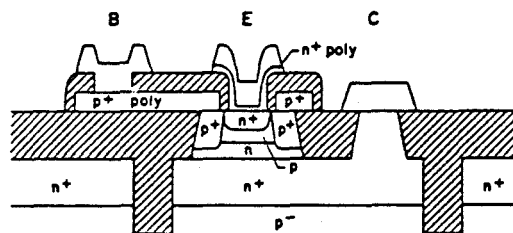


Fig. 17 Example of self-aligned polysilicon emitter bipolar transistor with a trench isolation structure [14].

Process simulation of such devices is obviously an excellent application of the work presented in the preceding chapters. In the following, we simulate the formation of a typical bipolar transistor, and focus on two important aspects: i) the connection between the intrinsic and extrinsic bases, and ii) the walled emitter configuration with trench isolation scheme. The doping profiles in these two regions are mainly governed by the multilayer characteristic and inter-layer dopant exchange.

### 3.2 Device description

Starting with a p-type substrate, an arsenic implantation is simulated to form the buried layer followed by an 1.5 micron thick n-type epitaxy. An RIE U-groove is then performed, and the channel stopper is formed in the trench bottom by an untilted boron implantation. A thin oxide layer is deposited and the U-groove is filled with undoped polysilicon. The extrinsic base contact is defined by a boron doped polysilicon patterning. The oxide sidewall spacer is simulated by an isotropic deposition followed by a purely anisotropic etching. Boron is then diffused from polysilicon by an annealing at 950°C for 40 mn. The intrinsic base is then implanted, self-aligned with the spacer (BF<sub>2</sub> ion is used to obtain a very shallow junction). The resulting structure is displayed in Figure 18 with the boron iso-concentration contours, that clearly show the three P<sup>+</sup> regions: the extrinsic and intrinsic bases and the channel stopper. Arsenic doped polysilicon layer is then deposited to form the emitter contact and the emitter doping by diffusion.

The diffusion barrier effect created by the native oxide lying between the doped polysilicon and the silicon substrate is modeled by a segregation condition for arsenic given by [15]:

$$m_{\text{Si-Poly}} = \frac{C_{\text{Si}}}{C_{\text{Poly}}} = 0.5 \quad (1)$$

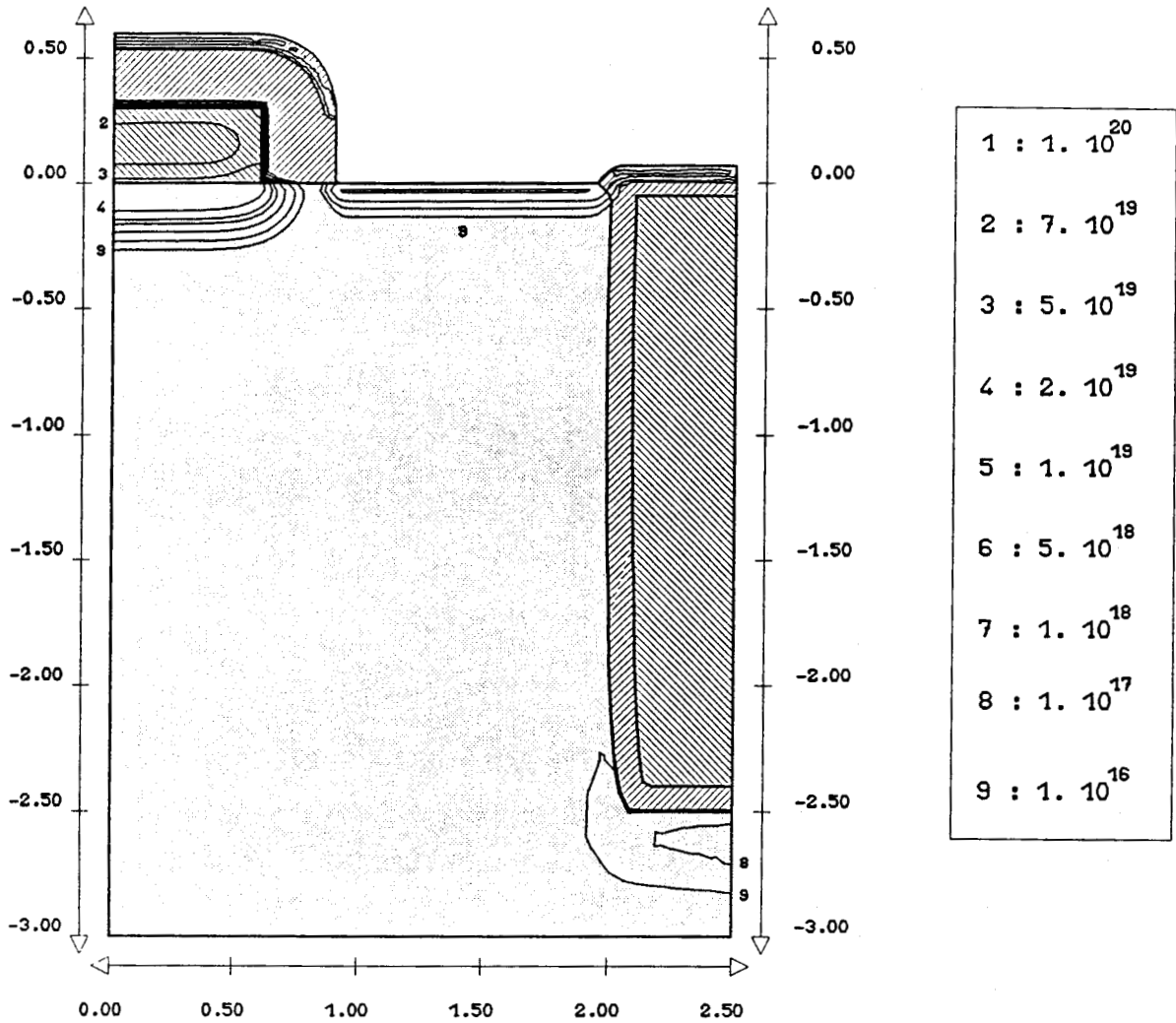


Fig. 18 Boron contours in the polysilicon bipolar transistor after extrinsic base annealing and BF<sub>2</sub> ion implantation.

Finally, the emitter is diffused at 950°C for 40 mn in neutral ambient. The resulting profiles are shown for boron (fig. 19a), arsenic (fig. 19b) and total concentrations (fig. 19c). From these figures, it is obvious that the overall boundary conditions are efficiently solved in the multilayer structure. A cross-section in the active area of the device is displayed in fig. 20, illustrating the arsenic transport from polysilicon to the intrinsic emitter, governed by the segregation condition defined by equation (1). The final mesh is plotted in fig. 21. Typical CPU time for this example is 120 min on a 1-Mips machine.

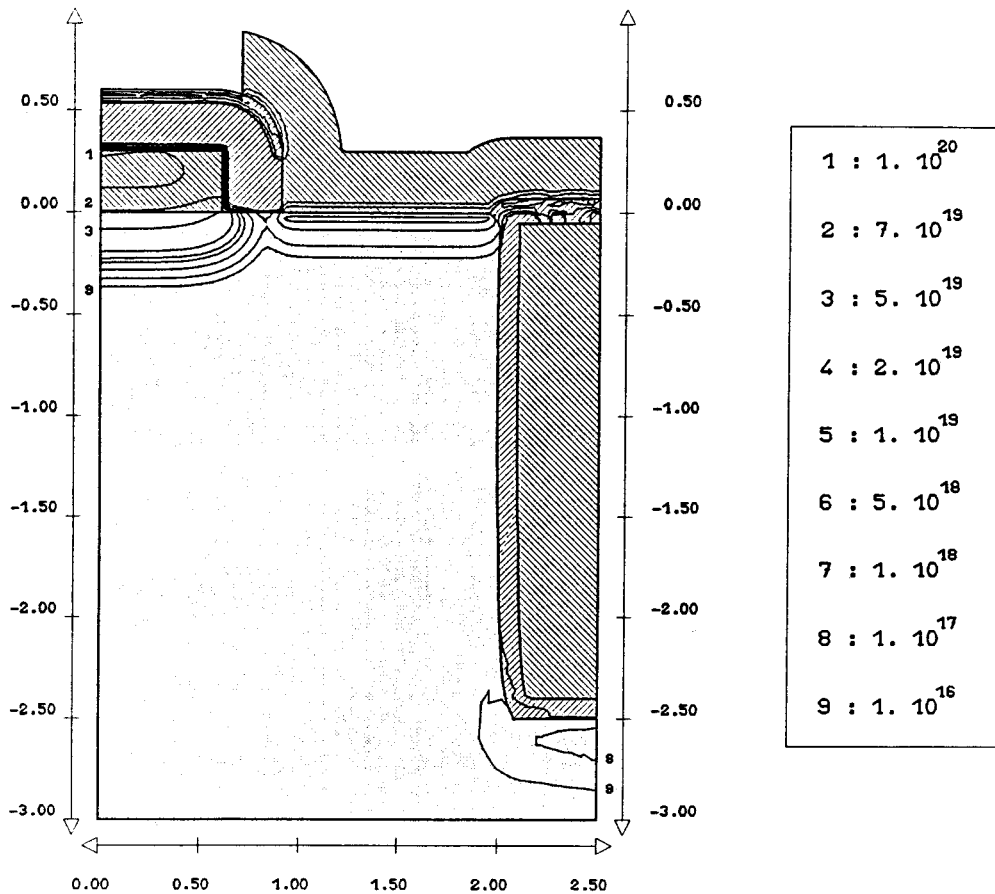


Fig. 19a Boron contours (spacer width: 0.2  $\mu\text{m}$ ).

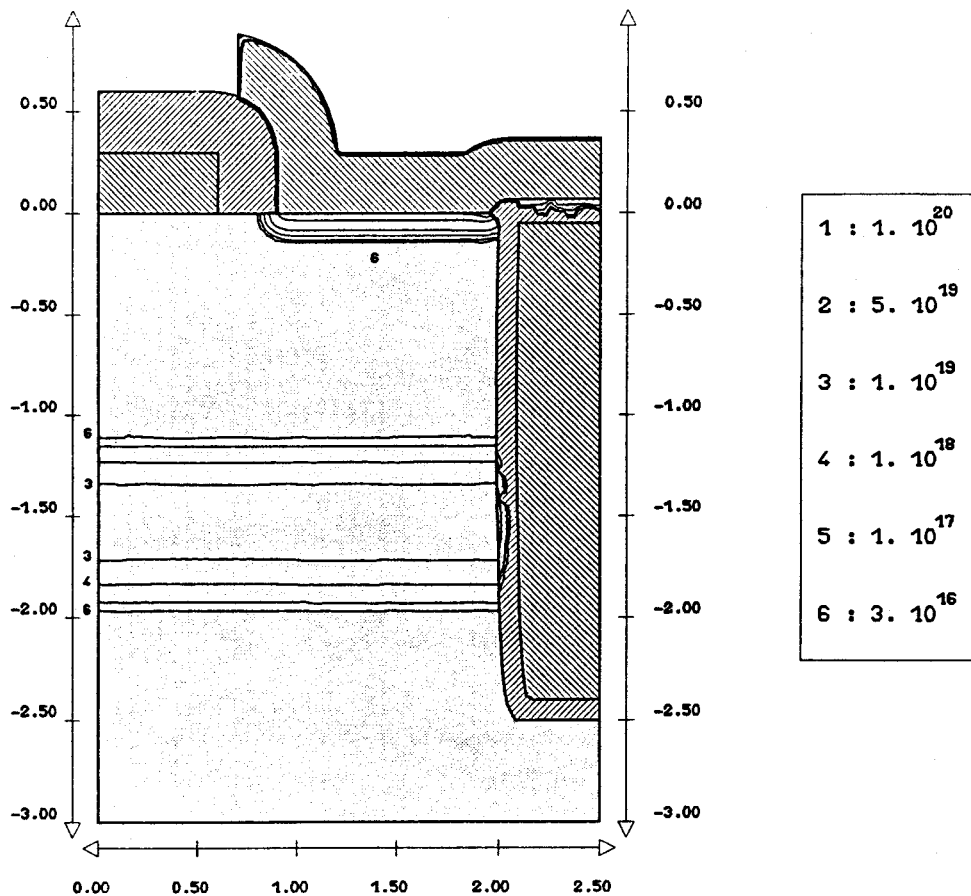


Fig. 19b Arsenic contours (spacer width: 0.2  $\mu\text{m}$ ).

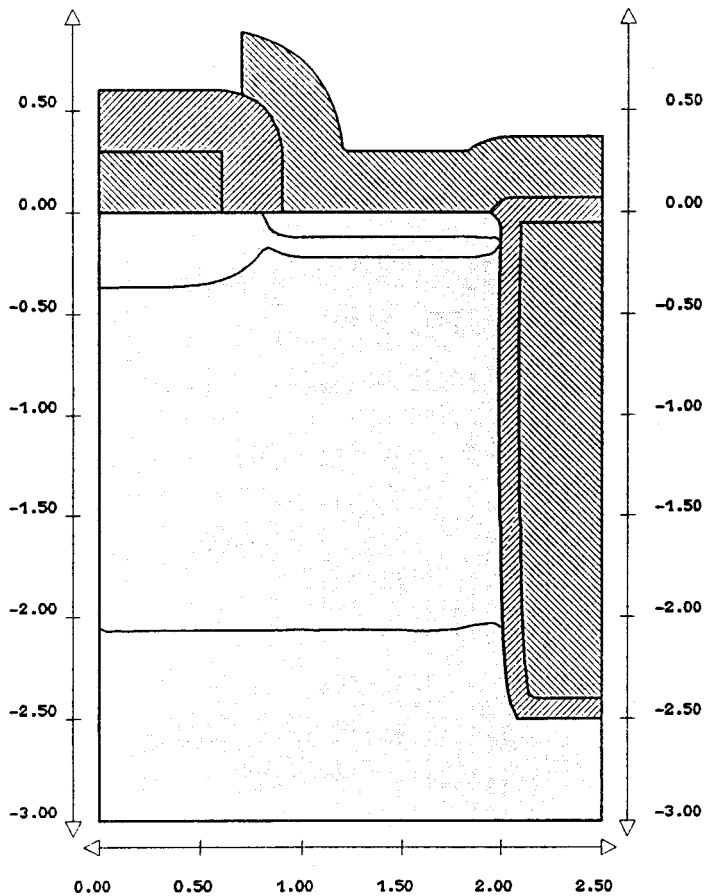


Fig. 19c Junction locations (spacer width: 0.2  $\mu\text{m}$ ).

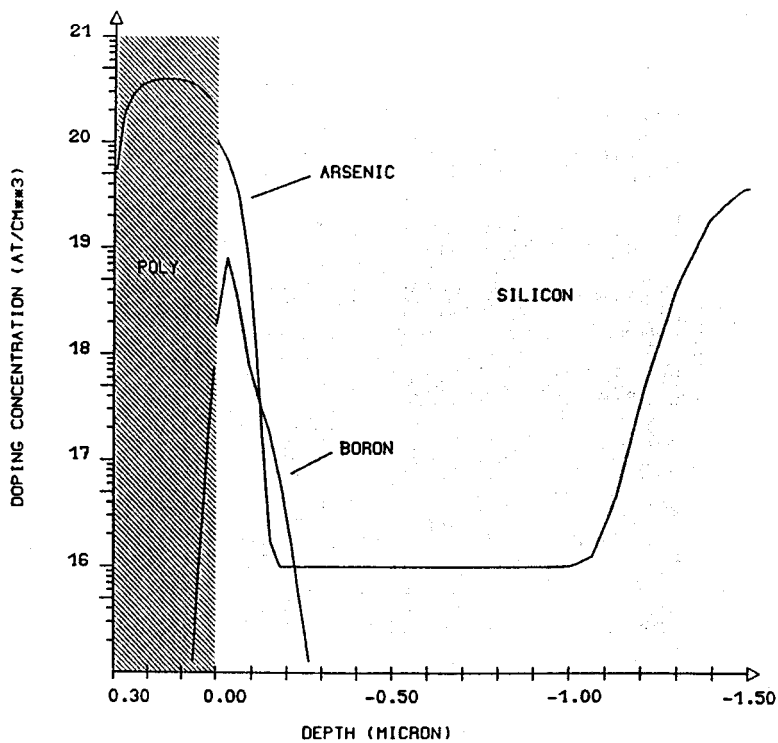


Fig. 20 Cross section in the intrinsic part of the device.

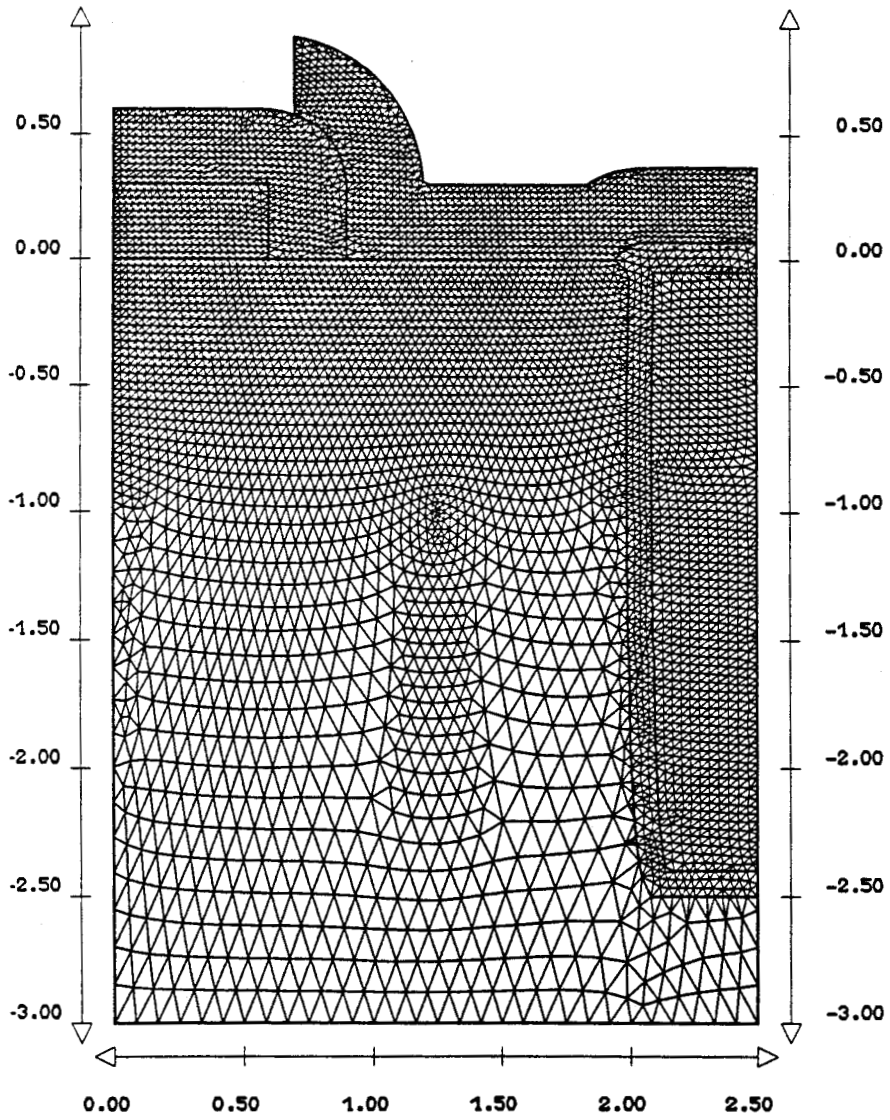


Fig. 21 Final mesh in the polysilicon bipolar transistor.

### 3.3 Extrinsic-intrinsic bases connection

In this typical self-aligned process, the sidewall spacer width has a dramatic influence on the access resistance between the extrinsic and intrinsic bases and on the emitter/collector punchthrough voltage. The boron countour plot displayed on fig. 19a indicates that the doping level at the bases connection region is limited by two factors: the first one, purely geometrical, is due to the weak lateral extension of each P<sup>+</sup> zones, the second one is induced by the static segregation effect that reduces the boron concentration in silicon near the sidewall-oxide/silicon interface.

The resulted structure of the boron concentration profile is more clearly displayed in fig. 22, that shows a zoom of the bases connection region for 0.4 micron spacer width. The segregation effect is obviously demonstrated by the presence of high boron concentration in the oxide spacer near the interface, forcing a boron depletion in silicon. The boron exodiffusion from the doped polysilicon to the spacer as well as the residual implanted profile in the upper part of the structure are also present.

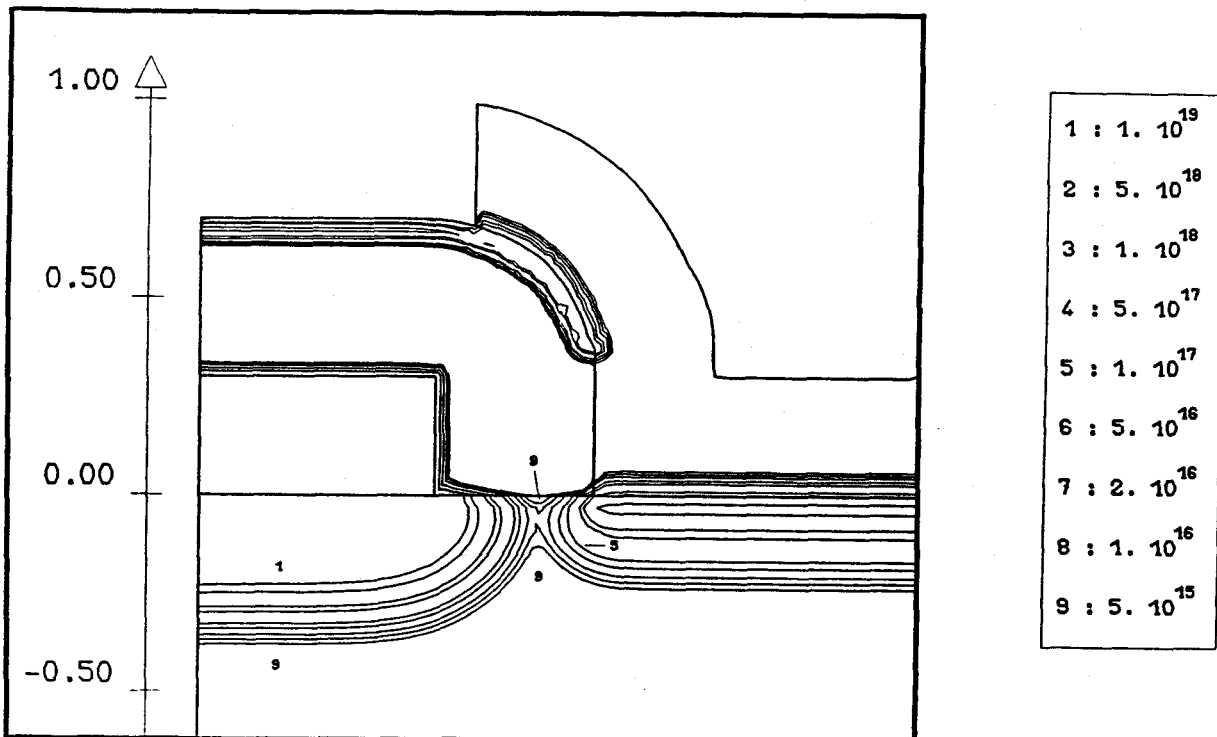


Fig. 22 Boron contours in the sidewall oxide region (0.4 μm width). Boron depletion in silicon and accumulation in the oxide are obvious.



In order to quantify the bases connection degradation with the increase of the sidewall spacer size, the process has been simulated with three different spacer widths : 0.2, 0.3 and 0.4 microns. The arsenic and boron concentration profiles along the silicon surface are plotted in fig. 23 for these three simulations. This figure shows the steep reduction of the minimum boron concentration as the spacer widens (approximately 1 decade/0.1 micron). In the 0.4 micron width case, the minimum boron concentration is lower than the arsenic epitaxial layer concentration, this induces an artificial lateral extension of the emitter towards the extrinsic base, resulting in a local base thinning illustrated by the 2D junction plot of fig. 24. This phenomenon predicts a very low punchthrough voltage.

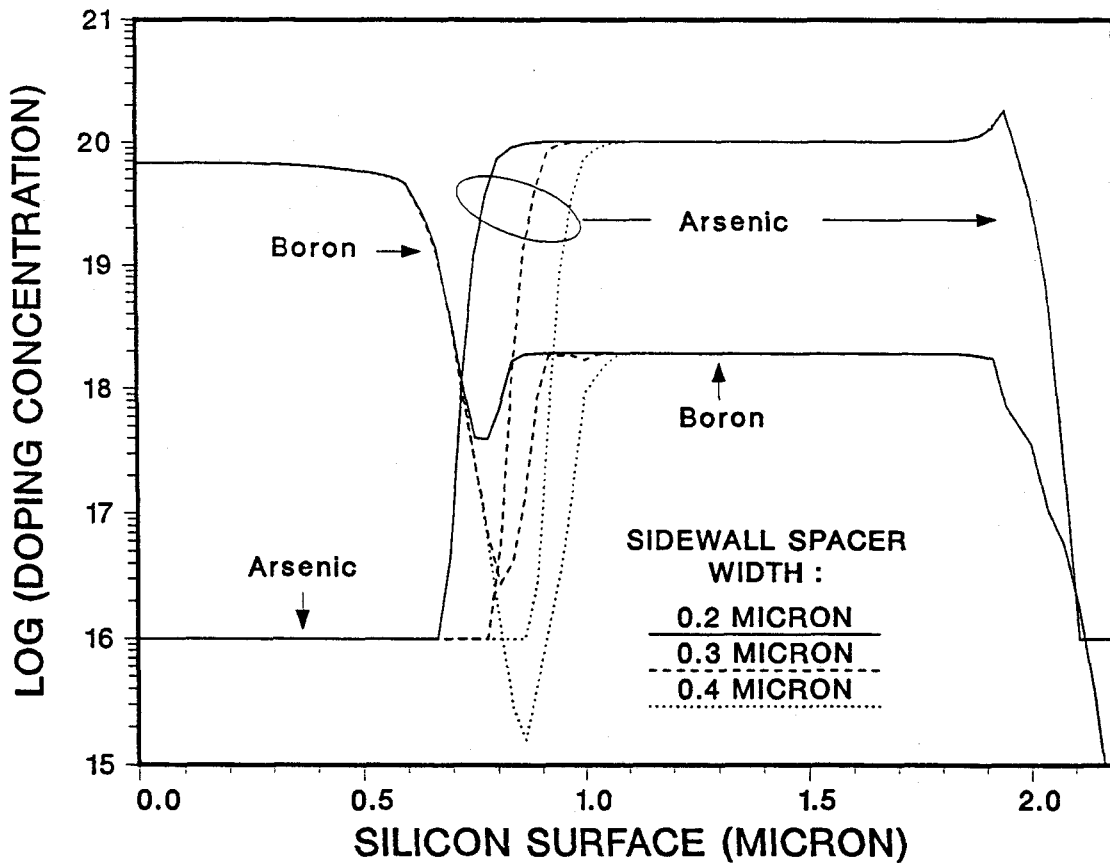


Fig. 23 Arsenic and boron doping along the silicon interface for the three sidewall spacer widths.

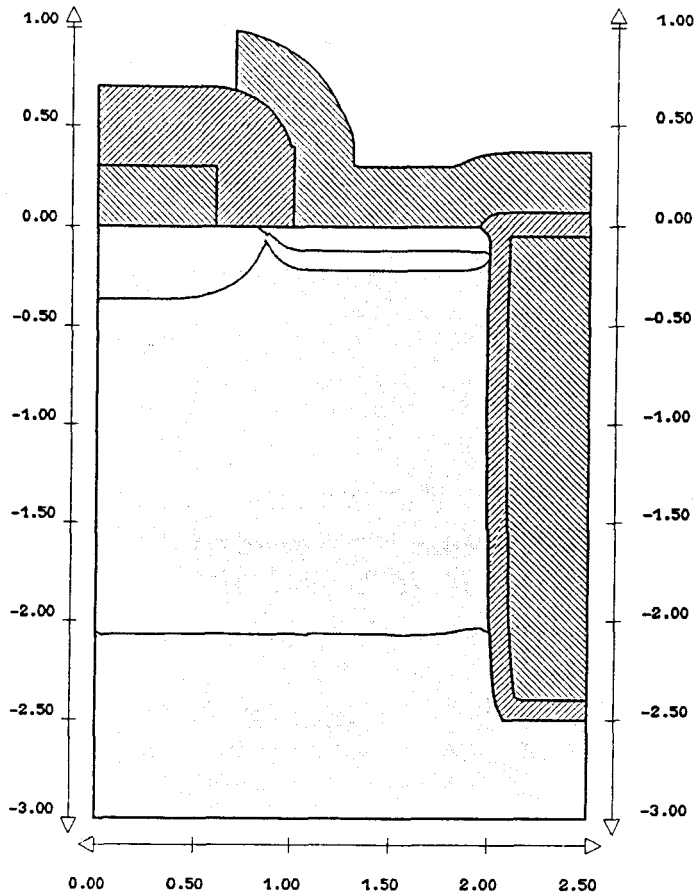


Fig. 24 Junction location in the complete structure in the case of 0.4  $\mu\text{m}$  spacer width. Degradation of the bases connection is clearly plotted under the oxide spacer.

An other way to analyse these effects is to plot the boron concentration along the emitter/base junction. Fig. 25 displays the boron profile along a perimeter defined from the left to the right on the simulation area by: i) P<sup>+</sup> polysilicon/silicon interface, ii) sidewall spacer/silicon interface, until the emitter/base junction is reached, iii) along the emitter-base junction till the trench location. Only the 0.2 micron spacer width gives boron compensation concentrations in the bases connection region higher than those obtained in the intrinsic transistor. In that case, the self-aligned configuration will not suffer from punchthrough voltage limitation. However, too low breakdown voltage could be obtained due to this high compensation concentration. This tradeoff can be optimized by systematic 2D simulations as presented here, experimental evidences have been already published about these parasitic problems [16], coherent with the above simulations.

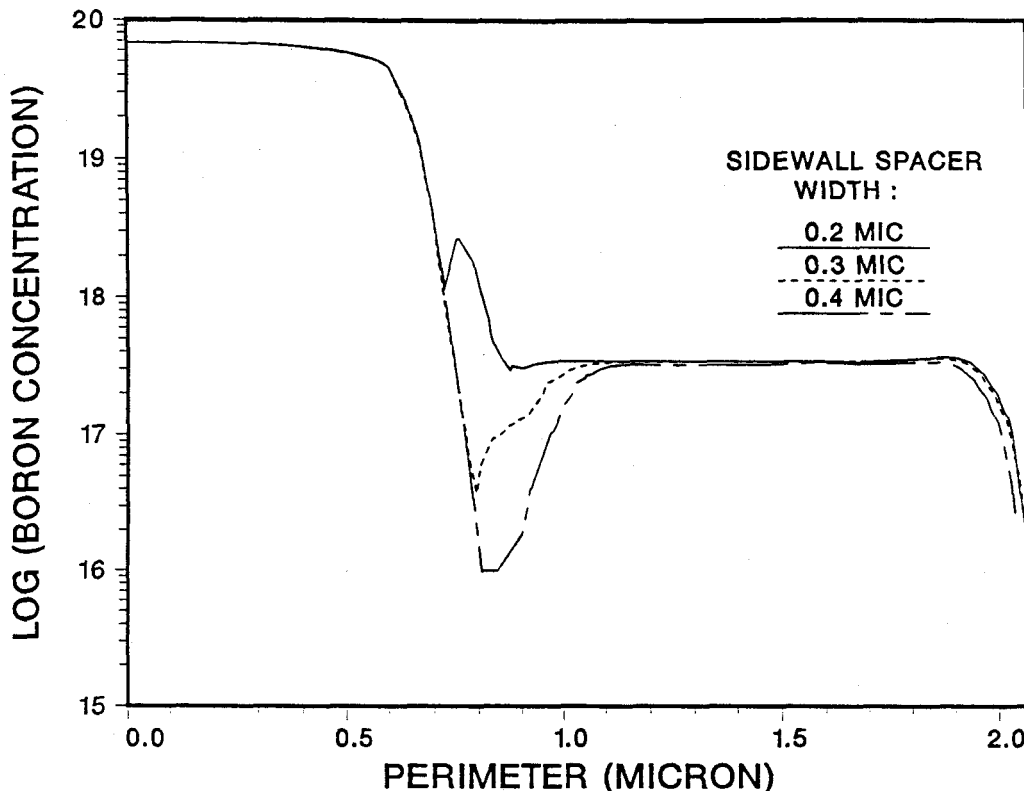


Fig. 25 Boron doping along the silicon/polysilicon interface (until about 0.75  $\mu\text{m}$ ) and along the emitter/base junction. The peak for 0.2  $\mu\text{m}$  spacer width is explained by the boron contours (extrinsic and intrinsic bases connection at this place).

### 3.4 Walled emitter configuration

Walled emitter configuration is widely used in bipolar technologies in order to improve the integration density. However, optimization of such a process is difficult and local base thinning or emitter collector short can easily appear, as clearly demonstrated in section II.

In the simulations presented above, the junction plots of fig. 19c and 24 indicate a severe base thinning formation along the trench oxide. The arsenic and boron concentrations along the silicon surface confirm, in the right part of fig. 23, that the emitter and the collector are almost shorted. The low boron doping is due to the shape of the trench oxide cap

masking the  $\text{BF}_2$  ion implantation (see fig. 18). The doping lowering is reinforced by the boron segregation during the emitter diffusion as shown on fig. 19a. In that region, the arsenic profile diffused from the  $\text{N}^+$  doped polysilicon is not altered by the oxide shape, moreover its segregation coefficient ( $m_{\text{Si-SiO}_2}=10$ ) limits the exodiffusion in the oxide (fig. 19b).

An alternative to reduce the base thinning in the walled region is to smooth the isolation oxide shape. Fig. 26 gives the junction plots for the same process in which the trench is replaced by a planarized recessed oxide [17]. The smooth thin to thick oxide transition allows the implanted boron to penetrate laterally deeper and the base thinning is avoided.

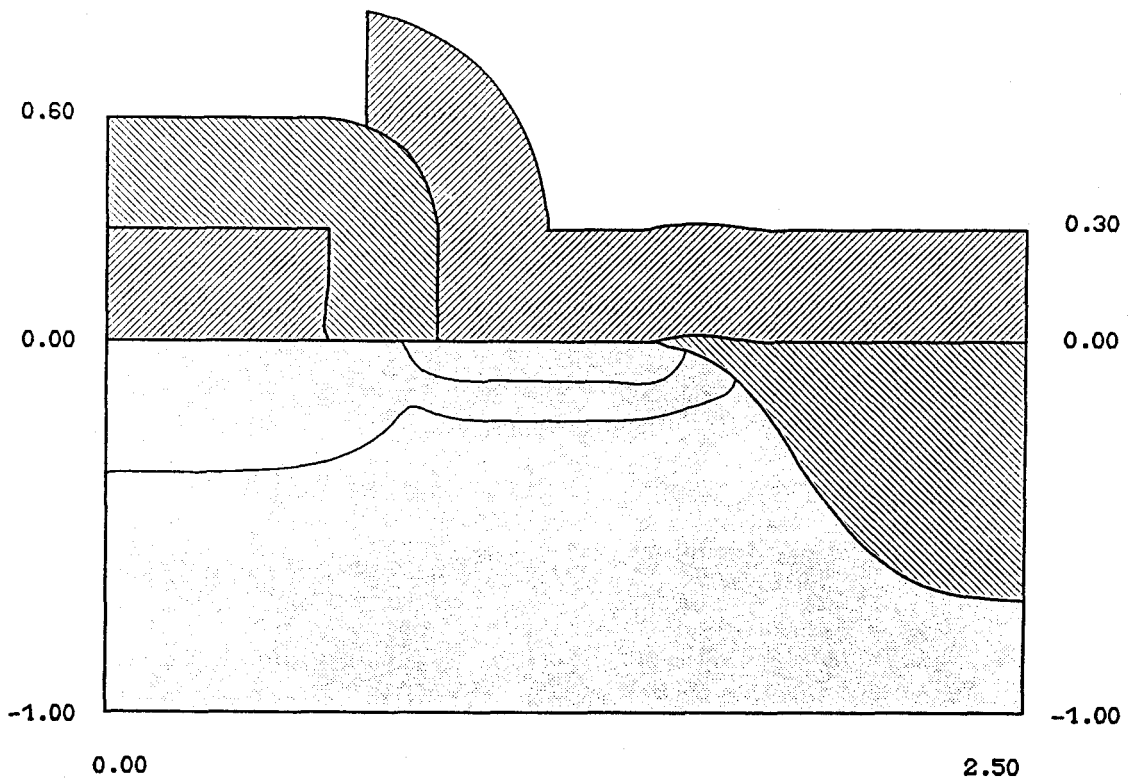


Fig. 26 Junction locations with a smooth oxide shape.  
Base shortening along this oxide is avoided.

REFERENCES :

- [1] P.J. Tsang, S. Ogura, W.W. Walker, J.F. Shepard and D.L. Critchlow, "Fabrication of high-performance LDD-FET's with oxide sidewall-spacer technology", IEEE Trans. Electron Devices, ED-29, No 4, 590-596, 1982.
- [2] J.J. Sanchez, K.K. Hsueh and T.A. DeMassa, "Drain-engineered hot electron resistant device structures : a review", IEEE Trans. Electron Devices, ED-36, No 6, 1125-1132, June 1989.
- [3] K.M. Cham, S.-Y. Oh, D. Chin and J.L. Moll, "A study of LDD device structure using 2D simulations", in Computer-Aided design and VLSI device development, 239-259, Kluwer Academic publishers, 1986.
- [4] D. Collard, B. Baccus, E. Dubois and D. Morel, "IMPACT 1-2-3: an integrated 2D process/device simulator for MOS technology", in Lecture notes of the software forum, NASECODE VI conf., 16-31, July 1989.
- [5] W.C. Ko, T.C. Gwo, P.H. Yeung, S.J. Radigan, "A simplified fully implanted bipolar VLSI technology", IEEE Trans. Electron Devices, ED-, 236-239, March 1983.
- [6] B. Baccus and D. Collard, "Two-dimensional study of Subilo-N technology", research contract for RTC PHILIPS Components, ISEN, November 1988.
- [7] E. Dubois, B. Baccus and D. Collard, "Analysis of the punchthrough effect in walled emitter bipolar transistors", in proc. ESSDERC conf., 789-792, September 1989.
- [8] B. Baccus, D. Collard and E. Dubois, "Simulation bidimensionnelle de technologie bipolaire", Journées GCIS, oral contribution, May 1987.
- [9] E. Dubois, J.L. Coppée, B. Baccus and D. Collard, "Electrical performances evaluation of isolation structures by coupled process and device simulation", in proc. SISDEP III conf., G. Baccarani and M. Rudan Eds., 151-161, September 1989.
- [10] E. Dubois, Thèse de Doctorat en Physique, Université de Lille, Avril 1990.
- [11] J. Lebailly, private communication, RTC PHILIPS Components, Caen, France.
- [12] B. Baccus, E. Dubois, D. Collard and D. Morel, "Efficient two dimensional multilayer process simulation of advanced bipolar devices", Solid-State Electronics, Vol. 32, No 11, 1013-1023, 1989.

- [13] L. Treitinger and M. Miura-Mattausch, "History, present trends, and scaling of silicon bipolar transistors", in Ultra-fast silicon bipolar technology, L. Treitinger and M. Miura-Mattausch Eds., Springer series in electronics and photonics 27, Springer Verlag, 1988.
- [14] D.D Tang, P.M. Solomon, T.H. Ning, D.D. Isaac and R.E Burger, "1.25  $\mu\text{m}$  deep-groove-isolated self aligned bipolar circuits", IEEE Journal of Solid-State Circuits, SC-17, 925-931, October 1982.
- [15] K. Sagara, T. Nakamura, Y. Tamaki and T. Shiba, "The effect of thin interfacial oxides on the electrical characteristics of silicon bipolar devices", IEEE Trans. Electron Devices, ED-34, No 11, 2286-2290, November 1987.
- [16] S.F. Chu, G.R. Srinivasan, H. Bhatia, B.M. Kamlage, F. Barson, J. Mauer and J. Riseman, "A self-aligned bipolar transistor", in VLSI Science and Technology, C.J. dell'Occa and W.M. Bullis Eds., 306-314, 1982.
- [17] B. Baccus, D. Collard, E. Dubois and D. Morel, "Two-dimensional process simulation of bipolar devices using a multilayer simulator", in proc. BCTM conf., 164-167, September 1988.

## CHAPITRE VI

### CONCLUSIONS

En considérant le travail présenté, il est évident que la caractéristique majeure du programme est la possibilité de traiter des structures multicouches, en ce sens que:

- i) 6 matériaux peuvent être pris en compte durant la même simulation. D'autres matériaux sont facilement incorporables dans la structure de données avec seulement des modifications mineures du code.
- ii) La redistribution des dopants durant les étapes de diffusion et d'oxydation est calculée dans chaque couche.
- iii) Le dépôt et la gravure sont inclus dans le programme, sans aucune interface avec les autres modules en ce qui concerne la description des formes.

Afin d'obtenir ces résultats, les points originaux suivants ont été développés:

- i) Une structure de données spécialisée a été incorporée dans le code, facilitant le développement du programme lui-même et le "maniement" des différentes couches.
- ii) Une stratégie unifiée de génération de maillage pour la simulation des étapes de dépôt/gravure et de diffusion/oxydation. Ceci s'est révélé être un bon compromis entre l'aspect numérique, la gestion de la structure de données et la précision obtenue de manière générale. Il convient de noter que cette procédure est cohérente par rapport au niveau de modélisation choisi (pour la diffusion et l'oxydation, par exemple).

- iii) Des procédures numériques efficaces ont été déterminées afin de résoudre l'aspect multicouche et les problèmes associés de ségrégation. L'algorithme de mise en forme de la matrice finale et l'utilisation d'une méthode de gradient conjugué avec préconditionnement se sont montrés particulièrement robustes.
- iv) Une solution au problème de maillage adaptatif a été étudiée et des résultats intéressants ont été obtenus sur des structures complexes et des technologies réelles.

Trois applications de ce programme ont été présentées, montrant les possibilités d'utilisation pour des technologies complexes (transistor bipolaire auto-aligné à émetteur polysilicium), ou actuellement industrialisées de manière plus courante (effets parasites intervenant dans une technologie bipolaire implantée).

Suivant en cela l'approche décrite dans l'introduction (chapitre I), il est opportun d'analyser brièvement les progrès réalisés par ailleurs en simulation de procédé, pendant le temps de développement d'IMPACT 4. En ce qui concerne les simulateurs bidimensionnels, l'accent a surtout été mis sur les aspects de modélisation et principalement sur la diffusion assistée par les défauts ponctuels, en ambiance neutre ou oxydante (par exemple et de manière non exhaustive, SUPREM IV [1] ou TITAN V [2]). Ceci représente bien sûr une approche complémentaire, par rapport aux problèmes topologiques décrits en détail ici. Ces deux simulateurs incluent dans une certaine mesure les possibilités géométriques d'IMPACT 4. D'autres modélisations physiques concernant la croissance bidimensionnelle d'oxyde ont également été développées, mais comme dans le cas de la diffusion, aucune réponse définitive n'a été apportée à ces phénomènes extrêmement complexes.

Dans le même temps, des progrès significatifs ont été effectués dans le domaine de la simulation tridimensionnelle (comme exemple non-unique citons SMART-P [3]). Des structures réalistes ont pu être prises en compte, et généralement l'utilisation de tels programmes est facilitée par l'emploi de super calculateurs. Cependant, il faut noter que la précision absolue de tels codes est souvent difficile à évaluer si l'on examine la génération de maillage, par exemple. Le raffinement et le maillage automatiques sont sans aucun doute des points clés dans le domaine 3D.



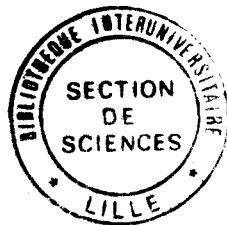
Prenant en compte toutes ces remarques, deux directions principales peuvent être définies en ce qui concerne les travaux futurs.

Les aspects topologiques et géométriques étant résolus, il est maintenant possible de se focaliser sur la modélisation de l'oxydation et de la diffusion. Bien que de nombreuses études aient déjà été entreprises, le microdimensionnement continu des composants élémentaires requiert une compréhension plus fine de ces problèmes. En particulier, les implications sur des structures réelles de telles formulations n'ont pas été complètement abordées. Dans ce cadre, nous pensons que l'utilisation couplée d'outils de simulation de procédé et de dispositif peut être extrêmement efficace.

L'autre direction serait le développement d'un nouveau simulateur 3D (et non pas l'extension d'IMPACT 4 vers des structures 3D). Dans ce cas, et tirant parti de l'expérience acquise durant la présente étude, nous sommes conscients de la tâche formidable que représente la conception et l'écriture d'un tel code [4]. Les aspects de génération de maillage ont déjà été mentionnés ci-dessus, mais les problèmes de génie logiciel et de manipulation d'objets 3D sont également extrêmement importants. Une alternative serait le développement d'un programme pseudo-3D, qui pourrait être considéré comme l'extension d'un simulateur 2D. Les techniques nécessaires à cette approche ont déjà été décrites [5]. Si l'on considère la possibilité de fournir des profils de dopage 3D à des simulateurs de dispositif 3D, c'est là un point très intéressant. Cependant, eu égard à l'évolution des simulateurs de procédé, une question importante est l'opportunité d'un tel travail sur le plan de la recherche, pour les prochaines années. Nous pensons donc que le développement d'un véritable programme 3D est encore un travail innovatif, mais cela demande en contrepartie des choix extrêmement réfléchis afin d'aboutir à un code utilisable.

REFERENCES BIBLIOGRAPHIQUES :

- [1] M.E. Law and R.W. Dutton, "Verification of analytical point defect models using Suprem IV", IEEE Trans. CAD, CAD-7, 181-190, 1988.
- [2] S. Martin, A. Gerodolle and D. Mathiot, "2D diffusion models and non-equilibrium point defects", in proc. NASECODE V conf., 282-288, Boole Press, June 1987.
- [3] S. Odanaka, H. Umimoto, M. Wakabayashi and H. Esaki, "Smart-P: rigorous three-dimensional process simulator on a supercomputer", IEEE Trans. CAD, CAD-7, No 6, 675-683, June 1988.
- [4] S.J. Polak, R.J.G. Goosens and N.E.B. Covern, "Requirements for device and process modeling of submicron devices", in proc. ESSDERC conf., 119-130, September 1989.
- [5] K. Nishi, S. Kuroda, K. Kai and J. Ueda, "Efficient three-dimensional process simulation of MOS devices", in proc. NASECODE VI conf., 297-302, Boole Press, July 1989.



## RESUME

Un code de simulation bidimensionnelle des procédés technologiques silicium a été développé: IMPACT 4. L'approche multicouche permet la redistribution des impuretés dans tous les matériaux courants, pour des technologies MOS ou bipolaires.

La méthode des éléments finis a été retenue pour la discrétisation des équations de diffusion, utilisant des éléments triangulaires linéaires, l'intégration temporelle étant effectuée par un schéma implicite incomplet. Un traitement original de la matrice finale autorise une solution efficace aux problèmes numériques engendrés par la ségrégation des impuretés aux interfaces. L'oxydation bidimensionnelle est modélisée sous la forme de l'écoulement d'un fluide visqueux incompressible. Par ailleurs, une approche unifiée de maillage est utilisée pour toutes ces étapes numériques. Le couplage avec la redistribution des impuretés a également été étudié, se révélant compatible avec les discrétisations spatiales et temporelles des équations de diffusion. Enfin, des procédures originales de maillage adaptatif sont proposées, déterminant des résultats intéressants au niveau du gain en temps de calcul et de la précision des profils de dopage dans le cas de structures complexes de grandes dimensions.

Trois applications valident ce programme. En particulier, l'influence de paramètres technologiques sur les performances de deux procédés bipolaires est analysée.

## MOTS-CLES

Semiconducteur

Diffusion

Simulation

Oxydation

Eléments finis

Maillage adaptatif