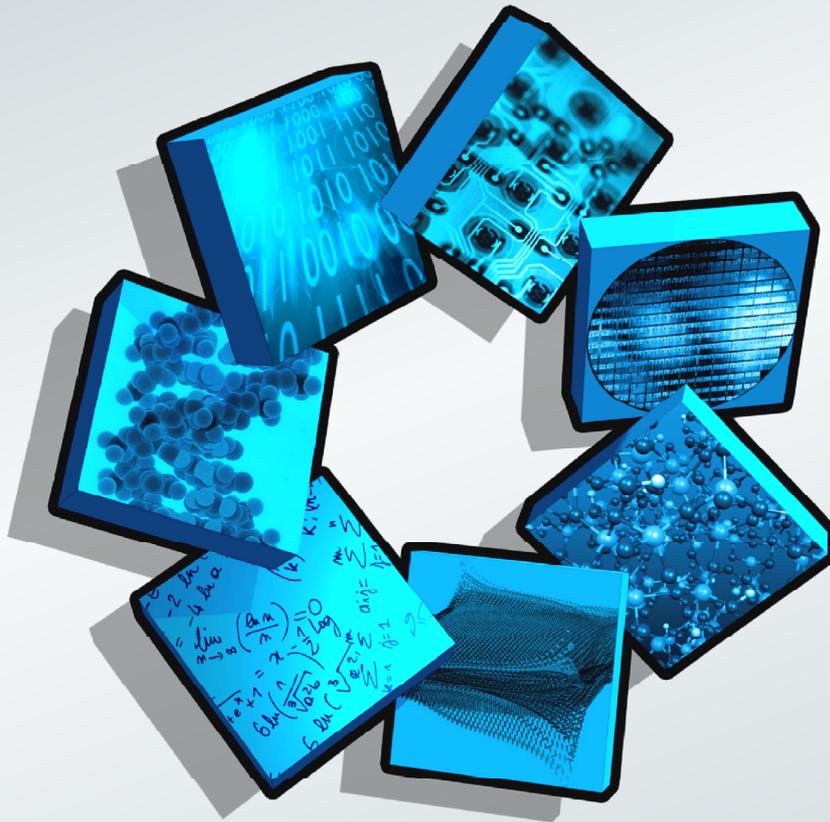


DOCTISS

Conférence des doctorants de l'ED I2S



Recueil des actes



NOM AUTEURS	LABORATOIRE	PAGE
Physique		
Claire DEROT	LCVN	1
Patrick VONLANTHEN	GRAAL	4
Laure BOUGUEN	GES	7
Monika GOLEBIEWSKA	LCVN	11
Teddy ROBERT	GES	15
Mathématiques		
Hilde OUVRARD	I3M	19
Olivier RODRIGUEZ	I3M	23
Guillaume BULTEAU	I3M	26
Selene SANCHEZ - FLORES	I3M	29
Chady ELMIR	I3M	32
Bio-Statistique		
Kevin BLEAKLEY	I3M	35
Florence CHAUBERT PEREIRA	CIRAD INRIA	38
Nadège DOSSAT	IURC	42
Nadine HOUEDE	CRLC	46
Bruno PEREIRA	CRLC	49
Vanessa ROUSSEAU	IURC	53
Électronique		
Giulio SABATINI	IES	56
Vincent CORREAS	IES	59
Olivier CATHABARD	IES	63
Arnaud DUCANCHEZ	IES	67
Alban GASSENQ	IES	71
Simon ROCHEMAN	IES	74
Jean-Michel BEC	Ecole des Mines d'Ales	77
Guillaume TATUR	Ecole des Mines d'Ales	81
Alexandre LAURAIN	IES	85
Ludovic BOYER	IES	89
Thibault LAURENT	IES	92
Thibault DAOUD	IES	95
Mécanique/Génie civil		
Charles VOIVRET	LMGC	99
Vincent TOPIN	LMGC	102
Karl BYTEBIER	LMGC	105
Etienne MALACHANNE	LMGC	109
Michele SERPILLI	LMGC	112
Informatique		
Alexandre LABADIE	LIRMM	116
Nicolas MOREAU	LIRMM	119
Chedy RAÏSSI	LIRMM	122
Nadia EL MRABET	LIRMM	125
Pierre Yves KOENIG	LIRMM	128
Hamidou TEMBINE	LIA	131
David DA SILVA	UMR DAP	134
Florian PINAULT	LIA	138
SYAM		
Julien VIAL	LIRMM	141
Lionel GOUYET	LIRMM	144
Yoann GUILLEMENET	LIRMM	148
Nicolas HOUARCHE	LIRMM	151
Jérémy LAFORET	LIRMM	153

INTERACTIONS ENTRE COLLOÏDES NON UNIFORMEMENT CHARGES

CLAIRE DEROT¹, MARTIN IN²

^{1,2}LCVN – UMR 5587
Université MONTPELLIER II, CC026 place Eugène Bataillon
34095 Montpellier Cedex 05

1 derot@lcvn.univ-montp2.fr

2 in@lcvn.univ-montp2.fr

HTTP://WWW.LCVN.UNIV-MONTP2.FR

Résumé— Nous avons étudié un système modèle constitué de gouttes de décane dispersées dans l'eau grâce à un tensioactif non ionique. Des charges électrostatiques, sous forme monovalente, tétravalente ou octavalente, sont ensuite ajoutées à la surface de ces gouttes. La technique de diffusion de la lumière permet d'analyser les interactions entre ces objets. On apprend ainsi que la répulsion entre les gouttes augmente avec la charge et qu'elle dépend de la répartition superficielle des charges.

Mots clés— colloïdes ; interactions électrostatiques ; patches de charges.

I. INTRODUCTION.

Quel est le point commun entre le brouillard, la peinture à l'eau, le lait et la mayonnaise ?... Ce sont tous des systèmes colloïdaux, c'est-à-dire un ensemble d'objets dispersés dans un milieu, avec une taille comprise entre le nanomètre et le micromètre. Ces particules et le milieu qui les contient sont assimilables à un système colloïdal, dès lors que la taille des objets est telle qu'aucune séparation de phase rapide ne se produit [1]. On considère que c'est Thomas Graham, un physicien britannique, qui introduisit le terme de colloïde, en étudiant les propriétés de diffusion de plusieurs substances en solution. Il souhaitait différencier les espèces capables de traverser une paroi de parchemin de celles qui ne l'étaient pas. Il lui apparut alors que les matériaux qui étaient retenus avaient un aspect gluant (gélatine, blanc d'œuf,...) [1]. Il appela ces derniers « colloïdes », du grec kolla, colle et eidos, forme [2].

Dans ces conditions, la force de gravité qui s'exerce sur chacune des particules est faible devant l'agitation thermique (mouvement incessant dont sont animées les molécules quelque soit leur état et qui devient plus intense quand la température augmente). Une dispersion de telles particules dans un liquide peut alors ne pas décanter spontanément. Cependant, des forces interparticulaires attractives peuvent favoriser l'agrégation de ces petites

particules et donc la séparation de phase. C'est pourquoi, il est nécessaire d'induire des forces répulsives pour garder un système stable au cours du temps.

Pour cela, des charges peuvent être ajoutées à la surface des objets de façon à créer des interactions électrostatiques favorisant la dispersion des gouttes dans le milieu environnant.

Dans cet article, les colloïdes modèles sont étudiés pour analyser l'effet de la répartition des charges sur les interactions électrostatiques. C'est une situation que l'on trouve dans des systèmes colloïdaux plus complexes tels que les protéines et les colloïdes minéraux (argile, ciment). Le système modèle est une microémulsion constituée de gouttelettes d'huile, en suspension dans l'eau, stabilisées par un tensioactif neutre (non ionique). Un tensioactif est une molécule composée d'une partie hydrophile (« qui aime l'eau ») et d'une partie hydrophobe (« qui n'aime pas l'eau »). Il favorise ainsi la formation de gouttes d'huile dispersées dans un milieu aqueux (Figure 1).

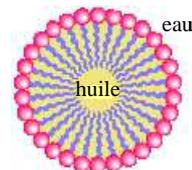


Fig. 1. Représentation schématique d'une goutte de microémulsion.

Des charges sont ensuite ajoutées à la surface des colloïdes grâce à des tensioactifs chargés, qui s'intercalent avec le premier tensioactif ajouté. Lorsque le tensioactif est monovalent, la répartition de charges en surface est statistique et uniforme. Mais, avec un tensioactif multivalent (ou polyélectrolyte), on forme des patches de charges surfaciques, donnant lieu à des colloïdes chargés non uniformément (Figure 2).

Le système ainsi conçu permettrait d'appréhender les interactions entre les protéines, par exemple. En effet, la répartition des charges à la surface de celles-ci n'est pas

uniforme. Il y a des zones fortement chargées positivement, d'autres fortement chargées négativement, et d'autres sans charges. On peut également observer une discrétisation des charges dans le ciment. Ainsi, la présence d'ions multivalents, comme le calcium Ca^{2+} , le magnésium Mg^{2+} ou le phosphate PO_4^{3-} , donne une répartition des charges sous forme de patches pour laquelle la modélisation standard des interactions électrostatiques n'est pas adaptée.

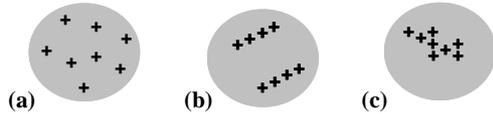


Fig. 2. Schéma de gouttes d'huile avec différentes répartitions de charges : (a) monovalente ; (b) tétravalente ; (c) octavalente.

II. INTERACTIONS ELECTROSTATIQUES.

Il existe deux catégories d'interactions : les forces attractives et les forces répulsives.

Les interactions de Van der Waals, qui ont une très courte portée, ont un effet attractif. En effet, chaque atome comporte un noyau porteur d'une charge positive, et entouré d'un nuage d'électrons chargés négativement. Mais, ces électrons se situent dans des positions excentrées et fluctuantes par rapport au noyau. Cette dissymétrie engendre des forces électromagnétiques interatomiques attractives. Les interactions électrostatiques ion-ion, si les charges sont opposées, sont également attractives [3], [4], [5].

En ce qui concerne les interactions répulsives, la force essentielle qui nous intéresse est la répulsion électrostatique. C'est une interaction longue portée qui existe sur une certaine distance, appelée distance de Debye ($1/\kappa$), et qui est liée à la présence d'espèces ionisées à la surface des colloïdes. Malgré la présence de ces charges superficielles, le milieu reste bien évidemment globalement neutre électriquement. Des contre ions, de charge opposée, entourent les objets chargés et ont tendance à se concentrer au voisinage de la surface. La concentration en contre ions diminue donc lorsque l'on s'éloigne des colloïdes, jusqu'à atteindre la concentration moyenne présente dans le solvant. Quand les surfaces se rapprochent, la concentration en contre ions devient plus grande que la concentration moyenne, ce qui engendre une force répulsive entre les particules colloïdales (Figure 3) [4], [5].

C'est pourquoi, la description théorique classique de l'interaction électrostatique entre des colloïdes s'attache à



Fig. 3. Représentation schématique de la répulsion électrostatique.

décrire la distribution des contre ions autour de la particule colloïdale. Cependant, elle le fait en considérant que la particule porte une charge ponctuelle centrale $Z+$ plongée dans un milieu continu de constante diélectrique déterminée. Les contre ions se répartissent de manière statistique, en suivant la loi de Boltzmann, dans un potentiel exprimé par la loi de Poisson. Ces lois décrivent en général bien le comportement des contre ions, mais des effets subtils apparaissent lorsque la distribution de charges en surface est discrète.

III. PRINCIPE DE LA DIFFUSION DE LUMIERE.

Pour mettre en évidence les interactions électrostatiques entre les colloïdes chargés, nous avons utilisé la diffusion de lumière. Ce dispositif expérimental est constitué d'une source laser qui traverse un petit volume de l'échantillon et d'un détecteur qui récupère l'intensité diffusée à 90° par ce volume (Figure 4).

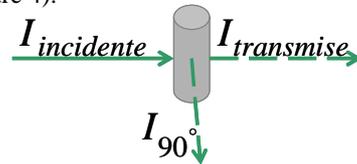


Fig. 4. Principe de l'appareil de diffusion de lumière.

L'intensité ainsi mesurée est la résultante d'interférences d'ondes diffusées par un ensemble de petits sous volumes du volume diffusant. Si tous les sous volumes contenaient strictement la même chose, la résultante des interférences serait nulle. Si un système colloïdal diffuse la lumière c'est, qu'à chaque instant, la composition des sous volumes varie : il y a fluctuation locale de concentration. La valeur moyenne de l'intensité diffusée mesure l'amplitude des fluctuations de concentration (équation (1)), c'est-à-dire la facilité avec laquelle la concentration dans le sous volume peut s'écarter de la concentration moyenne dans toute la solution. C'est ce qu'on appelle la compressibilité osmotique $d\Pi/dc$

$$I_{diffusée} = \frac{2\pi^2 n_0^2}{\lambda^4} \left(\frac{dn}{dc} \right)^2 kTc \left(\frac{d\Pi}{dc} \right)^{-1} \quad (1)$$

avec n_0 l'indice du solvant, λ la longueur d'onde du laser, n l'indice de la solution colloïdale, c la concentration en objets colloïdaux, k la constante de Boltzmann, T la température et Π la pression osmotique [4].

IV. RESULTATS EXPERIMENTAUX.

Nous avons donc mesuré l'intensité diffusée par un système colloïdal lorsque les charges sont réparties de différentes façons à sa surface (Figure 2).

La Figure 5 présente cette intensité en fonction du nombre de charges présent en moyenne à la surface de chaque goutte. On observe que l'intensité diminue lorsque le nombre de charges augmente et ceci dans les trois cas étudiés. Si on se concentre sur le cas de la répartition uniforme, on peut expliquer cette chute d'intensité par la pression osmotique. En effet, en partant de l'équation des gaz parfaits $PV=NRT$ (avec N le nombre de moles et R la constante des gaz parfaits), c'est-à-dire $P=\rho RT$, on sait que dans les cas non idéaux cette pression devient [4]

$$P = \rho RT(1 + B_2 \rho + \dots) \quad (2)$$

L'équation d'état d'une solution colloïdale dans un solvant étant analogue à l'équation d'état d'un gaz, on obtient

$$\Pi = RTc(1 + B_2 c + \dots) \quad (3)$$

On peut alors exprimer la compressibilité osmotique en dérivant l'équation (3) :

$$\frac{d\Pi}{dc} = RT(1 + 2B_2 c + \dots) \quad (4)$$

Le terme B_2 représente ici le second coefficient de Viriel. Ce coefficient caractérise les interactions entre les objets et s'exprime en cm^3 . Il peut, en effet, être positif ou négatif, et plus sa valeur est élevée, plus les interactions répulsives sont importantes. A l'opposé, plus il est négatif, plus l'attraction entre les différents objets est forte.

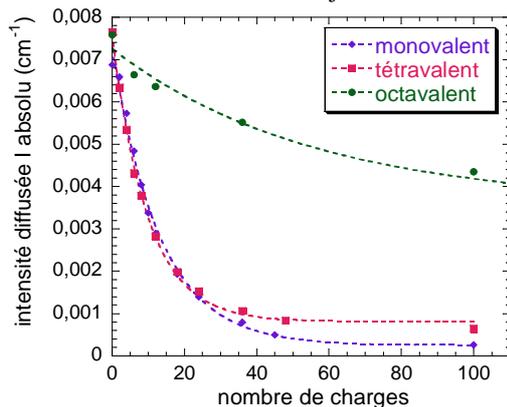


Fig. 5. Variation de l'intensité en fonction du nombre de charges.

En conséquence plus l'intensité diffusée est élevée, plus la compressibilité osmotique est faible (équation (1)), plus le second coefficient de Viriel est faible (équation (4)) et moins les colloïdes se repoussent. Dans le cas des charges

monovalentes, on remarque que l'intensité diminue jusqu'à atteindre un plateau pour un taux de 45 charges par goutte. Les colloïdes se repoussent donc de plus en plus lors de l'ajout de charges sur leur surface. Par contre, à partir de 45 charges par goutte, on atteint une saturation provenant d'un phénomène de condensation. Les contre ions vont, en effet, être attirés à la surface des colloïdes lorsque la charge surfacique devient assez dense. Cette condensation des contre ions donne une nouvelle charge effective plus faible que la charge placée initialement sur la goutte. Entre 45 et 100 charges par goutte, on a donc en réalité toujours le même taux de charges effectif.

On compare ensuite cette distribution uniforme avec les deux autres répartitions. Qualitativement, on observe le même phénomène de diminution d'intensité, et ce pour les mêmes raisons que précédemment, à savoir que la répulsion entre les gouttes augmente avec le taux de charges, ce qui nous donne une intensité plus faible. Mais, quantitativement, cette intensité diminue moins fortement pour le tétramère à partir de 25 charges par goutte, et dès les premières charges ajoutées pour l'octamère. Cette différence s'explique également par la condensation. La répulsion est, en effet, d'autant plus faible que la densité de charges est importante, au niveau du patch. Les contre ions vont donc plus aisément se condenser lorsque la répartition de charges n'est pas uniforme. La charge effective est alors plus faible et ce dès les bas taux de charges. La charge réelle des gouttes étant plus faible, la répulsion diminue donc, ce qui nous donne une pression osmotique plus faible et une intensité diffusée plus importante.

V. CONCLUSION.

La microémulsion choisie comme modèle permet de mettre en évidence les différences de répulsion électrostatique lorsque la répartition de charges n'est pas la même. Ainsi, un patch de charge augmente la densité de charge locale ce qui augmente la condensation des contre ions à la surface des colloïdes.

Par la suite, il sera intéressant d'étudier la distribution des taux de charges surfaciques par électrophorèse capillaire. De plus, nous pourrions nous pencher sur la viscosité de ce système modèle mais à des concentrations en objets chargés beaucoup plus importantes.

REFERENCES.

- [1] J.C. Daniel, R. Audebert, «La juste argile », Les éditions de Physique, 1995, 85-133.
- [2] V. Reus, « Equation d'état de suspensions de latex chargés », thèse Université ParisVI, 1995.
- [3] E. Lécolier, « Suspensions aqueuses de particules colloïdales anisotropes et chargés : structure et dynamique », thèse Université d'Orléans, 1998.
- [4] D. F. Evans, H. Wennerström, « The colloidal domain », VCH Publishers, 1994, 87-185.
- [5] Internet, Cours Ecole Centrale, 2006, http://www.mssmat.ecp.fr/IMG/pdf/CoursEcoleCentral2006_Lecolier1.pdf

DOCTISS 2008

Chimie de l'univers primordial et formation des premières étoiles

Patrick VONLANTHEN¹ (Doctorant 2^{ème} année), Denis PUY² (Directeur de Thèse)

Groupe de Recherches en Astronomie et Astrophysique du Languedoc
UMR 5024 - CC 72
Université de Montpellier II
Place Eugène Bataillon
34095 Montpellier Cedex 05

¹Patrick.Vonlanthen@graal.univ-montp2.fr, ²Denis.Puy@graal.univ-montp2.fr

Résumé—Nous nous intéresserons dans cet article à la chimie de l'univers primordial. Le gaz cosmologique a été le siège de réactions chimiques variées et diverses molécules ont été créées très tôt dans l'histoire de l'univers, les principales étant l'hydrogène moléculaire (H₂) et HD. Ces molécules jouent un rôle probablement déterminant dans la formation des premières étoiles de l'univers, un des sujets les plus actuels en cosmologie. Les propriétés du gaz primordial diffèrent en effet grandement de celles que l'on observe dans les nuages moléculaires actuels, qui sont à notre époque le siège de la formation des étoiles dans les galaxies. En particulier, seules les molécules primordiales sont capables de refroidir les nuages en effondrement gravitationnel, qui sont le lieu de formation des premières étoiles. Après en avoir brièvement expliqué les équations fondamentales, nous donnerons dans cet article les résultats principaux de la chimie primordiale, en incluant pour la première fois une importante correction liée au phénomène de recombinaison cosmologique pour la chimie du deutérium. Nous montrerons ensuite, à l'aide d'un modèle simple en une dimension, l'influence des molécules primordiales sur la formation des premières étoiles de l'univers.

Mots-clés— univers primordial — chimie primordiale — processus moléculaires — premières étoiles

I. INTRODUCTION

La théorie du Big Bang décrit la période extrêmement dense et chaude qu'a connue l'univers il y a un peu moins de 14 milliards d'années. Cette période marque le début de l'expansion de l'univers.

Durant les cent premières secondes, le processus de nucléosynthèse primordiale a donné naissance aux noyaux d'hydrogène, de deutérium (hydrogène lourd), d'hélium, ainsi qu'à quelques traces de lithium, béryllium, bore, etc. La chimie de l'univers primordial sera donc la chimie de ces éléments légers. Après la nucléosynthèse, tous les noyaux sont encore totalement ionisés. Lorsque l'univers s'est suffisamment dilaté et refroidi, on assiste à la recombinaison des noyaux d'hydrogène : environ 380000 ans après le Big Bang, l'hydrogène devient neutre en capturant un électron selon la réaction $H^+ + e^- \rightarrow H + \text{photon}$. Les autres noyaux se recombinent de la même manière. Cet événement capital dans l'histoire de l'univers va ensuite mener, entre autres, à la formation des premières molécules. En effet, tant que

la température est suffisamment élevée, les photons qui baignent l'univers sont capables de détruire immédiatement toute molécule qui pourrait se former (photodestruction). D'un autre côté, la décroissance de la densité, causée par l'expansion, agit à l'encontre des réactions collisionnelles. Toutefois, lorsque la température a suffisamment baissé, les processus de photodestruction deviennent inefficaces. L'apparition des premiers atomes neutres va permettre l'échange de charge avec les autres espèces encore ionisées et ainsi permettre la formation des premières molécules, essentiellement H₂ et HD. La chimie de ces molécules est actuellement l'objet de nombreuses études, à cause du rôle probablement déterminant que ces molécules ont joué dans la formation des premières structures cosmologiques. L'une des questions les plus actuelles en cosmologie est en effet celle de la formation des premières étoiles. Or les propriétés du gaz primordial diffèrent grandement de celles que l'on observe dans les nuages moléculaires actuels, qui sont à notre époque le siège de la formation des étoiles dans les galaxies. En particulier, seules les molécules primordiales induisent un processus de refroidissement du gaz lors des effondrements gravitationnels des premières structures.

II. MOLÉCULES DANS L'UNIVERS PRIMORDIAL

Les espèces chimiques que l'on veut étudier sont couplées par de nombreuses réactions. Il est donc nécessaire de résoudre un ensemble complexe d'équations différentielles. En outre, les taux de ces réactions chimiques dépendent de la densité et de la température, qui elles-mêmes décroissent à cause de l'expansion de l'univers. Nous avons donc l'ensemble d'équations suivant :

$$\frac{dT_r}{dt} = -H(z)T_r \quad (1)$$

$$\frac{dT_m}{dt} = -2H(z)T_m + \frac{8}{3} \frac{\sigma_T a_r}{m_e c} T_r^4 (T_r - T_m) x_e + \frac{2}{3} \frac{\Psi_{\text{mol}}}{n_b k_B} \quad (2)$$

$$\frac{dz}{dt} = -H(z)(1+z) \quad (3)$$

$$\frac{dn_b}{dt} = -3H(z)n_b + \sum_{\xi} \left(\frac{dn_{\xi}}{dt} \right)_{\text{chem}} \quad (4)$$

$$\frac{dn_{\xi}}{dt} = -3H(z)n_{\xi} + \left(\frac{dn_{\xi}}{dt} \right)_{\text{chem}} \quad (5)$$

T_r est la température du rayonnement qui baigne l'univers dans son ensemble (rayonnement de fond cosmologique), T_m la température du gaz, z le redshift¹, n_b la densité numérique totale et n_{ξ} la densité numérique de l'espèce chimique ξ (atome, ion ou molécule). $H(z)$ est le paramètre de Hubble, qui décrit le taux d'expansion de l'univers au redshift z . Le membre de droite de l'équation (1) et le premier terme des membres de droite des équations (2), (4) et (5) représentent la décroissance de la température et de la densité due à l'expansion de l'univers. Le deuxième terme du membre de droite de l'équation (2) est le transfert d'énergie du rayonnement vers le gaz via la diffusion Compton des photons du fond cosmologique sur les électrons. Dans ce terme σ_T est la section efficace Thomson, m_e la masse de l'électron, a_r la constante de radiation, c la vitesse de la lumière et x_e la fraction d'ionisation du gaz. Le troisième terme caractérise le transfert d'énergie via les transitions moléculaires (voir la section III pour la définition de Ψ_{mol}). Notons que k_B est la constante de Boltzmann. Dans les équations (4) et (5), le second terme du membre de droite est la contribution des réactions chimiques à la décroissance de la densité :

$$\left(\frac{dn_{\xi}}{dt} \right)_{\text{chem}} = \sum_{\xi_1 \xi_2} k_{\xi_1 \xi_2} n_{\xi_1} n_{\xi_2} - \sum_{\xi'} k_{\xi \xi'} n_{\xi} n_{\xi'}, \quad (6)$$

où les k sont les taux des réactions. Cette équation indique simplement que la densité numérique de l'espèce ξ change à cause des réactions qui forment ξ à partir des réactants ξ_1 et ξ_2 , mais aussi à cause des réactions qui détruisent ξ par collisions avec ξ' .

C'est à la fin des années 60 que différents auteurs ont mis en avant la possibilité de former de l'hydrogène moléculaire dans le contexte des nuages de gaz prégalactiques [1] [2]. Par la suite, plusieurs groupes de chercheurs ont élaboré un réseau de réactions chimiques basées sur H, D, He et Li principalement [3] [4] [5]. Il faut mentionner également l'article récent de Hirata & Padmanabhan [6], dans lequel il est montré que les photons émis lors de la recombinaison influencent la chimie de l'hydrogène, ce qui a pour effet de diminuer sensiblement la quantité de H_2 (d'un facteur légèrement supérieur à 4). Les molécules issues de la chimie primordiales sont principalement H_2 et HD, mais l'on note également la présence de LiH, H_2^+ , HD^+ , H_3^+ , H_2D^+ , HeH^+ ou encore LiH^+ . La molécule la plus abondante est H_2 avec une abondance numérique relative (nombre de molécules de H_2 par unité de volume divisé par le nombre total d'atomes, ions ou molécules par unité de volume) de l'ordre de 10^{-7} . Nous montrons sur la Fig. 1 l'évolution des abondances des espèces basées sur le deutérium, qui sont calculées ici pour la première fois en tenant compte de la correction de Hirata & Padmanabhan relative à la recombinaison cosmologique.

¹Le redshift, ou décalage vers le rouge, est le phénomène cosmologique de décalage des spectres lumineux des objets lointains vers les plus grandes longueurs d'onde. Il est considéré comme l'une des preuves du modèle du Big Bang.

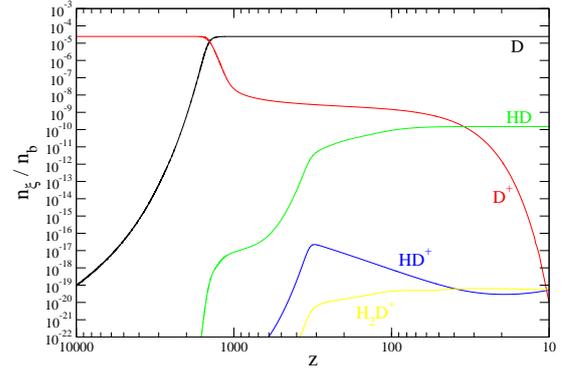


Fig. 1. Evolution des abondances des espèces basées sur le deutérium. En abscisse est donné le décalage spectral z et en ordonnée l'abondance numérique relative (nombre d'atomes, ions ou molécules de l'espèce ξ par unité de volume, n_{ξ} , divisé par le nombre total d'atomes, ions ou molécules par unité de volume, n_b).

III. REFROIDISSEMENT MOLÉCULAIRE ET PREMIÈRES ÉTOILES

L'une des questions les plus actuelles en cosmologie est celle de la formation des premières étoiles dans notre univers. Le problème est très complexe, car il faut comprendre le passage de l'uniformité relative des premiers temps à la structure hiérarchique de l'univers que l'on observe aujourd'hui : les galaxies s'organisent en amas, les amas en super-amas, etc. La structure à très grande échelle de notre univers est ainsi comparable à celle d'une éponge : de grosses condensations de matière sont reliées entre elles par des filaments qui délimitent de grosses régions vides semblables aux trous de l'éponge.

En fait, l'univers n'était pas parfaitement homogène à ses débuts. Certaines régions étaient plus denses que la moyenne alors que d'autres l'étaient moins. A l'intérieur des perturbations surdenses, l'auto-gravitation va agir contre l'expansion de l'univers. Concrètement, il existe un moment appelé décrochage, où l'expansion n'agit plus du tout sur la surdensité. Celle-ci commence alors à s'effondrer sur elle-même par l'effet de la gravitation. La pression interne du gaz ne cesse alors d'augmenter et au bout d'un certain temps elle va devenir du même ordre de grandeur que la force de gravité qui contracte le nuage de gaz. Pour diminuer la pression et permettre à l'effondrement de se poursuivre, il est nécessaire de refroidir le nuage afin de diminuer les collisions entre particules, donc la pression interne. Nous voyons ainsi clairement que la taille finale de l'objet en effondrement est influencée par les processus de refroidissement. C'est ici qu'interviennent les molécules primordiales.

Chaque molécule possède différents niveaux d'énergie rotationnelle. Une collision entre la molécule et un atome, un ion ou une autre molécule peut transformer une partie de l'énergie cinétique de la collision en énergie rotationnelle de la molécule, qui passe ainsi à un niveau d'énergie supérieur. Pour revenir au niveau inférieur (désexcitation), la molécule peut émettre un photon dont l'énergie est égale à la différence entre les deux niveaux. De cette manière, le

gaz perd de l'énergie grâce aux molécules primordiales et se refroidit. Un processus similaire permet au gaz de gagner de l'énergie (chauffage) : dans ce cas, la molécule est excitée en absorbant un photon et se désexcite par collision. On définit la fonction thermique Ψ_{mol} (qui apparaît dans l'équation 2) comme la différence entre le chauffage et le refroidissement dus à une molécule donnée.

Lorsqu'un nuage de gaz passe le point de décrochage et commence à s'effondrer, la température du gaz va progressivement devenir plus importante que celle du rayonnement. Dans ce cas, le refroidissement moléculaire est supérieur au chauffage et, globalement, les molécules refroidissent leur milieu. Sans ce refroidissement, le nuage cesserait de s'effondrer à partir du moment où la pression interne compenserait la force de gravité. Les molécules permettent donc à l'effondrement de se poursuivre et de former des objets plus petits. Mentionnons également que ce refroidissement peut engendrer un phénomène d'instabilité thermique, qui peut fragmenter le nuage en plusieurs morceaux plus petits.

La Fig. 2 montre l'évolution de la température d'un nuage de 10^{10} masses solaires. Nous considérons un modèle très simple d'effondrement d'un nuage sphérique parfaitement homogène [7] [8]. On voit clairement, vers la fin de l'effondrement, que la température du gaz T_m décroît grâce au refroidissement moléculaire. Par contre, lorsqu'on ne tient pas compte du refroidissement, elle ne cesse d'augmenter. Il convient toutefois de mentionner que le résultat de la

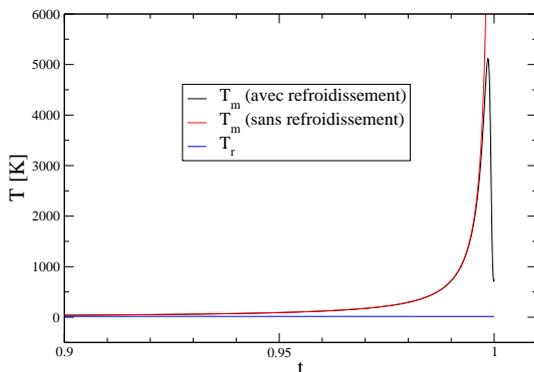


Fig. 2. Evolution de la température au cours de l'effondrement d'un nuage de 10^{10} masses solaires. L'axe des ordonnées indique la température en kelvin, l'axe des abscisses indique le temps écoulé depuis le début de l'effondrement en unités de 8.56×10^8 années.

Fig. 2 est obtenu sans tenir compte de la correction importante introduite par Hirata & Padmanabhan [6]. Nous étudions actuellement les effets de cette correction au cours d'un effondrement gravitationnel.

Le mécanisme de refroidissement qui vient d'être présenté est nettement moins efficace que les processus en cours dans les zones de formation stellaires actuelles. Pour cette raison les astrophysiciens pensent que les étoiles primordiales étaient beaucoup plus massives que les étoiles actuelles (plusieurs dizaines de fois voire plus de cent fois la masse de notre soleil). Une conséquence directe de leur très grande masse est que leur durée de vie était très courte (quelques millions d'années seulement, contre une dizaine

de milliards d'années pour une étoile semblable au soleil).

Notons d'ailleurs que l'expression "étoile primordiale" est un terme très général qui décrit toutes les étoiles de composition primordiale, quelles que soient les circonstances de leur formation. Un consensus semble se dégager quant à une nomenclature plus précise à utiliser :

- **Etoiles de Population III.1** : étoiles de première génération, aux propriétés déterminées entièrement par les paramètres cosmologiques et le processus de formation des premières structures.
- **Etoiles de Population III.2** : étoiles de seconde génération, dont la composition est toujours primordiale, mais dont la formation a été significativement affectée par les générations précédentes d'étoiles (injection d'énergie cinétique, rayonnement de haute énergie, etc.), ce qui peut changer l'échelle de masse des ces étoiles.
- **Etoiles de Population II.5** : étoiles formées à partir de gaz de métallicité² non nulle, mais dont la quantité de métaux est insuffisante pour modifier les propriétés de refroidissement et donc le processus de formation. Cette métallicité joue toutefois un rôle non négligeable dans l'évolution de l'étoile.

IV. CONCLUSION ET TRAVAIL FUTUR

Nous avons exposé dans cet article les principes de base de la chimie de l'univers primordial. Nous avons illustré, par un modèle simple d'effondrement, l'influence du refroidissement moléculaire sur la formation des premières structures. Par la suite nous allons examiner quelle est l'influence de la correction des abondances de HD sur la fonction thermique de cette molécule (l'importance exacte du refroidissement lié à HD est toujours débattue). Nous considérerons également des modèles d'effondrement de gaz primordial en 3D avec refroidissement moléculaire. Ces modèles, plus élaborés que le modèle dont il a été question ici, permettront de déterminer plus précisément l'importance des fonctions thermiques moléculaires, et en particulier de voir si elles peuvent induire un processus de fragmentation par instabilité thermique, ce qui pourrait avoir un impact majeur sur la masse des premières étoiles.

RÉFÉRENCES

- [1] Peebles, P. J. E. "Recombination of primeval plasma" 1968, ApJ, 153, 1
- [2] Saslaw, W. C. & Zipoy, D. "Molecular hydrogen in pre-galactic gas clouds" 1967, Nature, 216, 976
- [3] Lepp, S. & Schull, J. M. "Molecules in the early universe" 1984, ApJ, 280, 465
- [4] Puy, D. & Alecian, G. & Le Bourlot, J. & Léorat, J. & Pineau des Forêts, G. "Formation of primordial molecules and thermal balance in the early universe" 1993, A&A, 267, 337
- [5] Galli, D. & Palla, F. "The chemistry of the early universe" 1998, A&A, 335, 403
- [6] Hirata, C. M. & Padmanabhan, N. "Cosmological production of H₂ before the formation of the first galaxies" 2006, MNRAS, 372, 1175
- [7] Lahav, O. "Cooling of population III objects in a pressure supported collapse" 1986, MNRAS, 220, 259
- [8] Puy, D. & Signore, M. "Primordial LiH : the chemistry in a collapsing protocloud" 1998, New Astronomy, 3, 27

²En astrophysique, on entend par métallicité l'ensemble des éléments plus lourds que l'hélium.

DERIVE THERMIQUE DE CAPTEURS MAGNETIQUES : PEUT-ON L'ANNULER ?

L. Bouguen^{1,2}, B. Jouault^{1,2}, S. Contreras^{1,2}, Alexandre Kerlain³, Vincent Mosser³, et J-L. Robert^{1,2}

¹Groupe d'étude des Semiconducteurs CC 074, Place Eugène Bataillon 34095 Montpellier Cedex 5, France

²CNRS, UMR 5650, CC074, Place Eugène Bataillon, 34095 Montpellier Cedex 5, France

³ITRON SAS, 76 Avenue Pierre Brossolette, 92240 Malakoff, France

BOUGUEN@GES.UNIV-MONTP2.FR

ROBERT@GES.UNIV-MONTP2.FR

[HTTP://WWW.EDI2S.UNIV-MONTP2.FR/](http://WWW.EDI2S.UNIV-MONTP2.FR/)

Résumé— Nous nous intéressons ici au fonctionnement en température de capteurs magnétiques à Effet Hall à base d'hétérostructures AlGaAs/InGaAs/GaAs. L'utilisation d'une grille à la surface d'une croix de Hall permet de changer la densité du gaz d'électrons bidimensionnel et de modifier, jusqu'à l'annuler, la dérive en température du capteur. Dans cette étude, nous mettons en évidence l'effet de polarisation de la grille ainsi que l'effet de sa géométrie. Nous montrons qu'une modélisation linéaire à une dimension n'est pas adaptée à ce type de géométrie et qu'une méthode d'analyse par éléments finis est nécessaire.

Mots clés— hétérostructures AlGaAs/InGaAs/GaAs, capteur magnétique, Effet Hall, dérive en température.

I. INTRODUCTION.

Les capteurs magnétiques basés sur l'effet Hall sont aussi bien utilisés dans des applications industrielles de type capteurs de position, de rotation, que pour des applications en physique fondamentale telle que l'étude des phénomènes de transport dans les solides ou pour la détection de faibles champs magnétiques. Pour qu'un capteur soit intéressant, il faut qu'il soit très sensible au champ magnétique pour pouvoir détecter de faibles variations et de faibles valeurs et peu sensible aux perturbations extérieures telles que les changements de pression et de température. Des études ont été menées sur l'optimisation de capteurs à base de matériau massif tel que le silicium [1]. En opposition à ces structures à trois dimensions, nous nous intéressons à l'optimisation de capteurs à base d'hétérostructures AlGaAs/InGaAs/GaAs, où un gaz d'électrons bidimensionnel constitue la couche active [2]. Ces capteurs présentent des caractéristiques très intéressantes, une faible dérive en température (-140 ppm/°C) et une forte sensibilité au champ magnétique (750 VA⁻¹T⁻¹). Ces capteurs sont actuellement utilisés pour le comptage de l'électricité. Dans le but de développer une nouvelle famille de capteurs pour de nouvelles applications, il est nécessaire de diminuer voire d'annuler la sensibilité à la température. Nous présentons les résultats obtenus sur des croix de Hall avec grille. Nous verrons comment il est possible de diminuer et d'annuler la dérive en température

en modifiant la polarisation de la grille. Les résultats obtenus avec un modèle linéaire à une dimension seront présentés. Nous montrerons que ce modèle donne une idée des résultats mais qu'il n'est pas adapté si nous voulons faire une modélisation plus précise des effets dus à la géométrie de ces structures.

II. GENERALITES

On appelle effet Hall l'apparition d'un champ électrique transversal et d'une différence de potentiel dans un barreau parallélépipédique de semiconducteur (ou métallique) parcouru par un courant électrique lorsqu'on l'introduit dans un champ magnétique \mathbf{B}^a perpendiculaire à la direction du courant. Lors de l'introduction du barreau dans le champ magnétique les électrons animés d'une vitesse \mathbf{v} sont déviés sous l'action de la force de Lorentz \mathbf{F}_L , des charges positives s'accumulent dans le sens opposé. Le champ électrique transversal s'oppose à la déviation des électrons sous l'action du champ magnétique et il apparaît une différence de potentiel V_H appelée tension de Hall [3] (voir figure 1).

La valeur de la tension de Hall V_H est donnée par la relation suivante :

$$V_H = K_H \cdot I \cdot B \quad (1)$$

où I est l'intensité du courant de polarisation, B l'intensité du champ magnétique et K_H le coefficient de Hall qui est donné par la relation :

$$K_H = \frac{1}{e \cdot n_s} \quad (2)$$

e est la charge de l'électron et n_s la densité surfacique de porteurs dans le gaz bidimensionnel.

Pour que le capteur soit sensible à de faibles variations de champ ou de courant, il faut que la valeur du coefficient de Hall, K_H , soit élevée. Il faut aussi qu'il soit peu sensible aux variations extérieures, particulièrement aux variations de température.

^a Les valeurs en gras représentent des vecteurs.

Un paramètre important est la dérive en température S_T , définie par la relation :

$$S_T \approx \frac{1}{n_{s0}} \frac{dn_s}{dT} \approx f(\phi_s) \quad (3)$$

n_{s0} est la valeur de la densité surfacique de porteurs d'un échantillon sans grille [2].

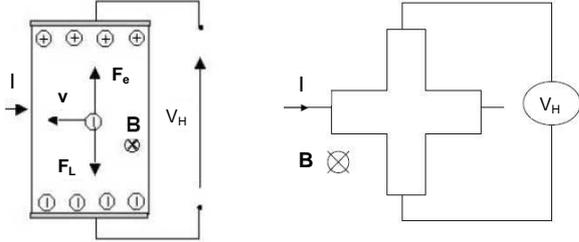


Fig. 1. A droite : Schéma de l'effet Hall dans un barreau de semiconducteur, à gauche : Schéma d'un montage de mesure d'une croix de Hall. F_L est la force de Lorentz, F_e la force électrique, v la vitesse des porteurs, I l'intensité du courant de polarisation, B l'induction magnétique, V_H la tension de Hall.

Le paramètre qui a le plus d'influence sur S_T est l'ancrage du niveau de Fermi E_F en surface. Cet ancrage, qui définit l'énergie ϕ_s , dépend des états de surface (voir figure 2).

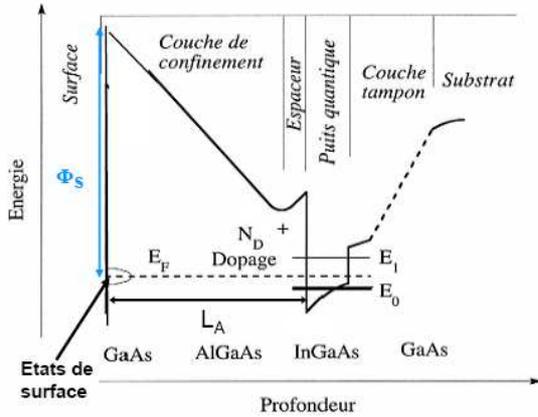


Fig. 2. Schéma du diagramme de bande de l'hétérostructure

Une façon simple de modifier ce paramètre, est d'ajouter une grille en surface et à travers une polarisation, de modifier très facilement le potentiel de surface et donc la dérive en température du capteur.

Pour les structures avec grille, la densité de porteurs dans le gaz d'électrons bidimensionnel est donnée par la relation :

$$n_s = \frac{C_A}{e} (V_g - V_T - V) \quad (4)$$

C_A est la capacité entre la surface et le gaz d'électrons bidimensionnel, e la charge de l'électron, V_g la tension de grille, V_T la tension de seuil du transistor et V la tension au centre de la croix, $V \approx RI/2 \approx V_{in}/2$.

La grille peut être polarisée soit en étant connectée à la masse, dans ce cas, $V_g = 0V$; soit par l'intermédiaire d'un

pont de résistance, et dans ce cas $V_g = \beta V_{in}$, β dépend de la valeur des résistances choisies (voir schéma figure 3).

La dérivée de l'équation 4 par rapport à la température permet d'évaluer la sensibilité à la température. Pour ce faire, considérons que $V \approx RI/2$ et $V_g = \beta V_{in} = \beta RI$. Seule la résistance R dépend de la température. Nous obtenons ainsi la relation :

$$\frac{dn_s}{dT} = \frac{C_A}{e} \frac{dR}{dT} \left(\beta - \frac{1}{2} \right) I + cste \quad (5)$$

Dans cette expression, la variation de la densité de porteurs en fonction de la température (et donc la valeur du coefficient de Hall K_H) dépend du courant de polarisation I , ainsi que du coefficient β et donc de la façon dont est polarisée la grille.

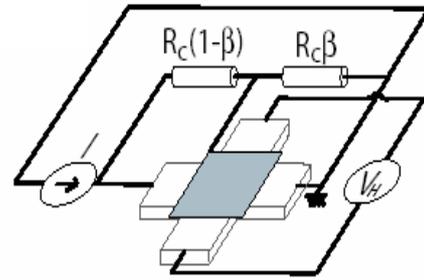


Fig. 3. Schéma de la polarisation de la grille par l'intermédiaire d'un pont diviseur.

D'après les équations 3 et 5, nous en déduisons la dérive en température S_T :

$$S_T = \frac{1}{n_{s0}} \frac{C_A}{e} \left[\left(\beta - \frac{1}{2} \right) \frac{dR}{dT} I \right] + S_{T0} \quad (6)$$

S_{T0} et n_{s0} sont respectivement la dérive en température et la densité de porteurs d'un capteur sans grille. La dérive en température dépend aussi du courant de polarisation et du coefficient β .

D'après l'équation 6, nous pouvons en déduire une valeur de courant de polarisation pour laquelle la dérive en température s'annule. Ce courant dépend de paramètres de la structure tels que C_A et R , paramètres difficiles à modifier, mais aussi de β , paramètre facilement modifiable par l'intermédiaire du pont de résistances.

III. DESCRIPTION DES ECHANTILLONS.

III.1. Structure verticale.

La structure verticale des échantillons étudiés est un empilement successif de différents semiconducteurs avec des paramètres de maille différents, on parle alors d'hétérostructure pseudomorphique. Dans notre cas, l'hétérostructure est à base d'AlGaAs/InGaAs/GaAs, avec un dopage planaire silicium (voir le schéma du diagramme de bande figure 2).

Sur un substrat GaAs, on fait croître la couche active de 20 nm d' $In_{0.15}Ga_{0.85}As$, suivie d'une barrière d' $Al_xGa_{1-x}As$,

et enfin une couche de GaAs. L'épaisseur totale de ces deux dernières couches (L_A) est de 250 nm. La zone active est à l'interface GaAlAs/GaInAs où se situe le gaz d'électrons bidimensionnel. L'alliage $Al_xGa_{1-x}As$ a une concentration d'Al variable qui décroît vers la surface pour éviter la formation d'un autre puit quantique à l'interface AlGaAs/GaAs.

III.2. Géométrie des échantillons.

Les échantillons étudiés sont des croix de Hall de 80 μm de large et 320 μm de long sur lesquelles une grille a été rajoutée. Une photo de ces échantillons est présentée figure 4. L'échantillon nommé S1 a une grille qui recouvre toute la croix. L'échantillon nommé S2 a une grille qui ne recouvre qu'une partie des bras de la croix et l'échantillon S3 a une grille qui ne recouvre que le centre de la croix.

Ces échantillons avec grille ont une structure comparable à celle de transistors à forte mobilité électronique (High Electron Mobility Transistor : HEMT). Il va donc être possible d'utiliser la polarisation de la grille pour modifier la densité de porteurs dans le gaz bidimensionnel.

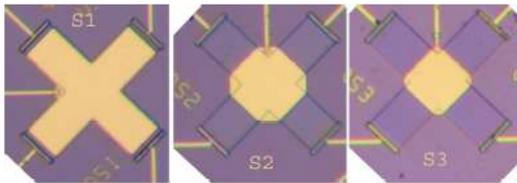


Fig. 4. Photographie des échantillons étudiés (vue de dessus), à droite grille pleine (S1), au centre grille qui recouvre une partie des bras (S2), à gauche grille qui ne recouvre que le centre (S3).

IV. RESULTATS EXPERIMENTAUX.

IV.1. Mesures à température fixée

A une température donnée, la densité de porteurs dans le gaz d'électrons bidimensionnel est donnée par la relation (4).

La figure 5 montre la variation de la densité de porteurs mesurée par effet Hall, en fonction de la tension de grille, les mesures sont faites pour une faible valeur de V et pour les trois types de grille étudiées. Les points sont les résultats expérimentaux et, la ligne pleine représente les résultats de la modélisation linéaire. Contrairement à ce qui était attendu (équation 4), les 3 courbes ne se superposent pas et les résultats de la modélisation ne sont en accord qu'avec les mesures de la croix S1. En effet, lorsque la grille recouvre toute la surface de la croix et lorsque la tension V est faible (régime de fonctionnement linéaire du transistor), la répartition des électrons est homogène et donc $n_s = n_H$ où n_H est la densité de porteurs mesurée par effet Hall. Quand la grille ne recouvre qu'une partie de la surface de la croix, l'effet de la géométrie ne peut être négligé et $n_s \neq n_H$; les mesures d'effet Hall ne nous donnent plus directement la densité de porteurs dans le gaz d'électrons bidimensionnel.

IV.2. Mesures en fonction de la température

Des mesures d'effet Hall ont été réalisées sur les trois géométries pour des températures comprises entre $-60^\circ C$ et $80^\circ C$.

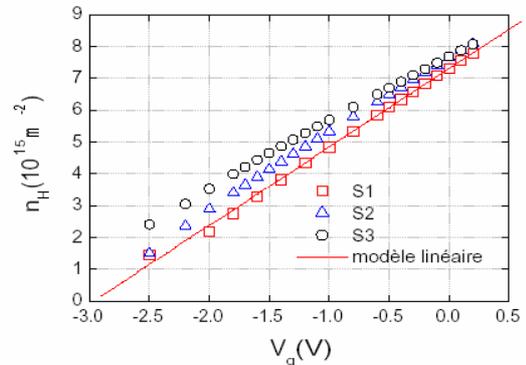


Fig. 5. Variation de la densité de porteurs mesurée par effet Hall n_H en fonction de la tension de grille : les symboles sont les résultats expérimentaux obtenus pour les 3 géométries de grille, la ligne pleine représente le résultat obtenu avec le modèle linéaire.

La figure 6 montre la variation relative $K_H/K_H(T=22^\circ C)$ en fonction de la température pour différents courants de polarisation, pour $V_g = 0V$ dans le cas d'un échantillon de type S2. Cette variation dépend de la valeur du courant de polarisation. Elle peut avoir une pente négative, positive et même nulle. C'est le cas pour un courant de polarisation de $I \approx 80 \mu A$.

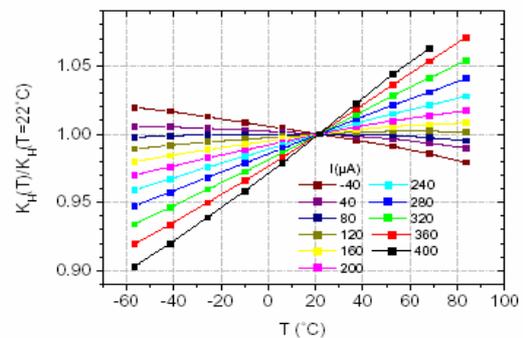


Fig. 6. Echantillon S2 : Variation relative de $K_H/K_H(T=22^\circ C)$ en fonction de la température pour des valeurs de courant comprises entre $-40 \mu A$ et $400 \mu A$, la grille est connectée à la masse.

Pour pouvoir intégrer le capteur dans un montage électronique, il faut que la dérive en température s'annule pour une valeur du courant de polarisation suffisamment importante.

La figure 7 montre la variation de la dérive en température en fonction du courant de polarisation pour les trois géométries de grille lorsque $V_g = 0V$. Les symboles représentent les résultats expérimentaux, la ligne pleine le résultat du modèle linéaire. La valeur du courant pour laquelle $S_T = 0 \text{ ppm}/^\circ C$ dépend de la taille de la grille ; cette

valeur est d'autant plus forte que la grille est petite (respectivement pour les structures S1, S2, S3, $I \approx 60 \mu\text{A}$, $80 \mu\text{A}$, $90 \mu\text{A}$). Cette figure montre que le modèle linéaire n'est pas adapté aux géométries S2 et S3. Ces structures sont intéressantes car elles permettent d'obtenir une dérive en température nulle pour un courant de polarisation plus important que la structure S1. De plus, elles possèdent des courants de fuites de grille (courants qui passent par la grille au lieu de traverser l'échantillon) moins importants. Cependant, ce courant de polarisation est encore trop faible. D'après l'équation 6, la valeur du courant de polarisation dépend de la façon dont est polarisée la grille, c'est-à-dire de la valeur de β .

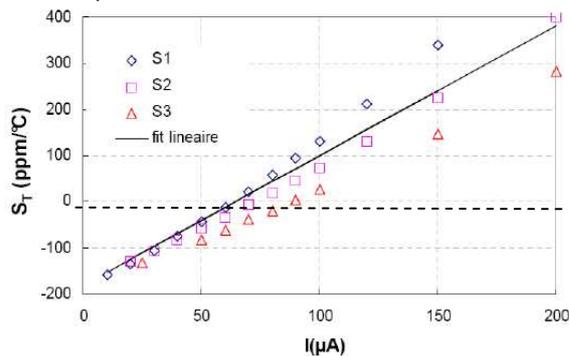


Fig. 7. Variation de la dérive en température en fonction du courant de polarisation pour les trois différentes géométries de grille, $V_g=0\text{V}$. Les points représentent les données expérimentales et les lignes pleines les résultats du modèle linéaire.

La figure 8 montre la variation de la dérive en température pour différentes polarisations de la grille ($\beta = 1/3$ et $1/4$). Les points représentent les résultats expérimentaux et les lignes pleines les résultats de la modélisation linéaire. Il existe un courant de polarisation pour lequel la dérive en température s'annule et ce, quelle que soit la façon dont on polarise la grille. La valeur de ce courant dépend du coefficient β . Le courant est environ égal à $150 \mu\text{A}$ pour $\beta=1/4$ et $300 \mu\text{A}$ pour $\beta=1/3$.

Un autre phénomène qui apparaît est la saturation de cette dérive pour les fortes valeurs de courants de polarisation. Le modèle linéaire ne permet pas d'interpréter ce phénomène. Ce modèle est un modèle à une dimension où l'effet de la géométrie n'est pas pris en compte. Un autre modèle à deux dimensions basé sur la méthode des éléments finis a été développé. Ce modèle prend en compte les géométries à deux dimensions de la croix et de la grille. Les résultats obtenus avec ce modèle sont présentés sur la figure 8 par les lignes pointillées et sont en très bon accord avec les résultats expérimentaux.

V. CONCLUSION.

L'étude de la dérive en température de croix de Hall avec grille a montré qu'il était possible de minimiser la sensibilité à la température ($S_T \rightarrow 0$) en polarisant la grille. Ce résultat

dépend de la géométrie (S1, S2 et S3), de la polarisation de grille (V_g) et du courant de polarisation de l'échantillon (I).

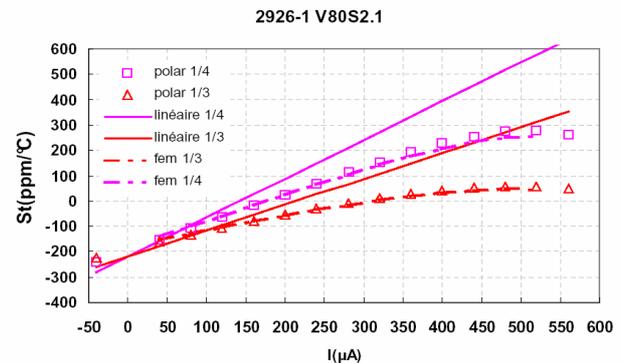


Fig. 8. Echantillon S2 : Variation de la dérive en température en fonction du courant de polarisation pour différentes polarisations de grille. Les points représentent les données expérimentales et les lignes pleines les résultats du modèle linéaire, les lignes pointillées les résultats du modèle par éléments finis.

Par exemple, la dérive en température d'un échantillon de type S2 s'annule pour un courant de polarisation égal à $150 \mu\text{A}$ pour une valeur de coefficient $\beta = 1/4$.

La géométrie de la grille influence les mesures d'effet Hall, la densité de porteurs mesurée par effet Hall n_H diffère de la densité de porteurs dans le gaz d'électrons bidimensionnel n_s . Ceci s'observe dans le cas des géométries S2 et S3.

Cette étude a montré qu'une modélisation linéaire à une dimension n'est pas suffisante pour expliquer les phénomènes dus à la géométrie de la grille ; par contre, nous avons montré qu'une modélisation à deux dimensions par la méthode des éléments finis donne des résultats en parfait accord avec les résultats expérimentaux. Il s'agit là d'une confirmation du rôle de la géométrie de la grille, et on peut donc tirer parti de cet effet pour réaliser des composants qui présentent une dérive en température nulle.

VI. REMERCIEMENT.

Nous remercions Y. Tréguier du Groupe d'Etude des Semiconducteurs et J. Lyonnais de l'ATEMI pour leur support technique.

REFERENCES.

- [1] J.B. Kammerer, « Capteurs intégrés pour la mesure à haute résolution de champs magnétiques », thèse de doctorat, Université Louis-Pasteur-Strasbourg I, 2004
- [2] F.Kobbi, « Etude et optimisation d'hétérostructures pseudomorphiques AlGaAs/InGaAs/GaAs destinées à la réalisation d'un capteur magnétique », thèse de doctorat, Université de Montpellier II, 1996.
- [3] RS Popovic, « Hall Effect Devices », 2nd edition, 2004

Phase Transitions in Nitrogen Monolayer Adsorbed on Graphite: The Monte Carlo Study

Monika R. GOŁĘBIEWSKA¹, Robert FABIAŃSKI², Bogdan KUCHTA³, Lucyna FIRLEJ¹

¹Laboratoire des Colloïdes, Verres et Nanomatériaux, Université Montpellier 2, Place Eugene Bataillon, 34095 Montpellier, Cedex 5, France

²Instytut Chemii Fizycznej i Teoretycznej, Politechnika Wroclawska, Wybrzeże Wyspiańskiego 27, 50-370 Wrocław, Poland

³MADIREL, UMR6121, CNRS-Université de Provence, Centre de St. Jérôme, 13397 Marseille, Cedex 20, France

Monika.Golebiowska@univ-montp2.fr

Summary

We have observed two structural phase transitions in the monolayer of molecular nitrogen adsorbed on graphite. At $\sim 23\text{K}$ the commensurate, herringbone structure undergoes an in-plane rotational transformation without losing its translational order. At 78 K the system melts.

Key words

Canonical Monte Carlo simulations, adsorption mechanism, phase transition

Introduction

Molecular nitrogen (N_2) adsorbed on graphite reveals a very rich phase diagram (Fig.1a). In particular, at low temperatures a variety of structures appears, depending on surface coverage. The sequence of structural phase transitions at densities close to monolayer is already well understood and described in the literature [1-4]. Therefore, our aim is to investigate the phase transitions in multilayer systems. We want to understand whether the presence of subsequent layers modifies the mechanism and temperature of rotational phase transition and melting within a particular layer. We present here the first stage of the work – canonical Monte Carlo simulations of full N_2 monolayer adsorbed on corrugated surface of graphite. The calculation for n -layers systems ($n= 2, 3, 4$) are in progress.

The system

At low temperature and full monolayer coverage the N_2 molecules form so called herringbone structure, commensurate with graphite. It consists of two sublattices of molecules. In solid phase, in both sublattices the molecules are parallel to each other and molecular centers of mass coincide with centers of graphite hexagons. Between the sublattices, the molecules are perpendicular (Fig.1b). This configuration is the starting point of our study.

We have formed a bilayer by adding on the top of the first layer (sites A on Fig.1c) a second one, of the same structure but shifted by the 1.25 \AA in x direction (sites B). The third layer can be added in two ways: by putting it above the first one (AB A hexagonal-like sequence) or above the hexagons that are still “empty” (sites C, ABCA cubic sequence). The four-layer structures are, in consequence, ABAB or ABCA. Both types of molecular stacking are under investigation.

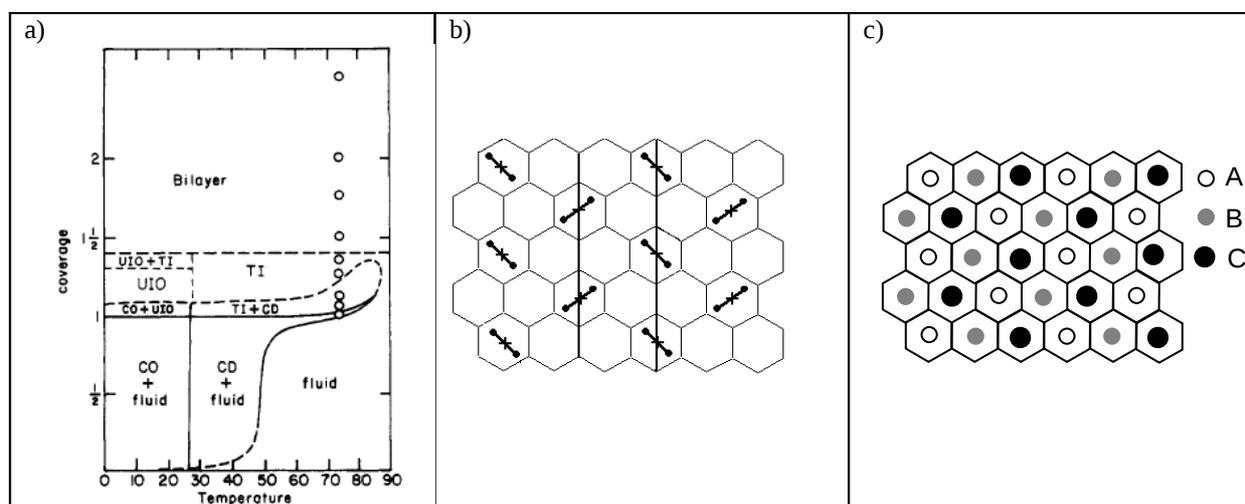


Fig. 1 a) The phase diagram of nitrogen on graphite, [5] b) low temperature herringbone structure of N_2 monolayer, commensurate with underlying graphite, c) possible occupations of sites in multilayer structures (see the text).

Simulation details

Monte Carlo (MC) simulations were performed in canonical ensemble, in the temperature range from 0 K to 100 K. The dimensions of simulation box ($x = 59.028289 \text{ \AA}$, $y = 42.599999 \text{ \AA}$, $z = 20.000000 \text{ \AA}$) were chosen to accommodate a monolayer of 160 N_2 molecules. Periodic boundary conditions were applied to keep the molecules inside the simulation box. Nitrogen molecules were considered as rigid, described by the site-site Eters potential [6]. The electrostatic interaction was modeled by using four point charges: two positive ($+0.373e$) at the distances $\pm 0.847 \text{ \AA}$ from the center of mass and two negative ($-0.373e$), at $\pm 1.044 \text{ \AA}$. Graphite surface was modeled using the Fourier expansion of Steele's potential [7], with $\epsilon_{NC} = 31.92 \text{ K}$ and $\sigma_{NC} = 3.36 \text{ \AA}$. The stabilization runs consist of $8 \cdot 10^6$ MC steps. The final results were taken from 10^6 MC steps from production runs.

Order parameters – definitions

To characterize the structure of the layer we used several order parameters described below.

The herringbone order parameter, OP_{HERR} , is a measure of the in-plane orientational ordering of the molecular axis.

$$OP_{HERR} = \frac{1}{N_M} \left\langle \sum_{i=1}^{N_M} (-1)^j \sin(2\phi_i) \right\rangle \quad (1)$$

N_M is the number of molecules in the layer, ϕ_i is an angle between the axis of the i th molecule and arbitrary chosen direction of the simulation box and j – number of the sublattices. $OP_{HERR} = 1$ for herringbone structure; $OP_{HERR} = 0$ when the molecules are randomly oriented. The commensurate order parameter, OP_{COMM} , monitors whether the adsorbed molecules are in registry with the graphite substrate.

$$OP_{COMM} = \frac{1}{6N_M} \left\langle \sum_{i=1}^{N_M} \sum_{s=1}^6 \exp(-ig_s r_i) \right\rangle \quad (2)$$

The outer sum runs over all N_M molecules at the center of mass positions and inner sum over six graphite reciprocal lattice vectors, g_s . r_i is the position of the center of the i th hexagon in the graphite plane. $OP_{COMM} = 1$ when all center of mass positions of N_2 are directly over the centers of the hexagons. $OP_{COMM} = 0$ for structures incommensurate with graphite.

The Φ_6 order parameter measures the average bond order within a plane layer.

$$\phi_6 = \frac{1}{N_M} \left\langle \sum_{i=1}^{N_M} \exp(in\alpha_i) \right\rangle \quad (4)$$

where $n = 6$, α_i is the angle between two vectors connecting i th molecule with two of the closest neighboring molecules. $\Phi_6 = 1$ for solid triangular

layer, $\Phi_6 = 0$ when in-plane, liquid – like disorder appears.

The tilt order parameter, OP_{TILT} , quantifies the out-of-plane orientation ordering of the system.

$$OP_{TILT} = \frac{1}{2N_M} \left\langle \sum_{i=1}^{N_M} (3\cos^2 \theta_i - 1) \right\rangle \quad (3)$$

where θ_i is an angle between the surface normal and the long axis of the i th molecule. $OP_{TILT} = -0.5$ when all the molecules are parallel to the surface of graphite and increases to $OP_{TILT} = 1$ when the molecules are perpendicular to the substrate.

Results

At low temperatures ($T < 23 \text{ K}$) the nitrogen monolayer forms a 2D solid ($\Phi_6 \sim 1$) commensurate with graphite ($OP_{COMM} \sim 1$). N_2 molecules lay flat over the substrate ($OP_{TILT} \sim -0.5$). The triangular structure is of herringbone type (Fig. 2a). It can be deduced from azimuthal angle distribution (Fig. 3a): two sharp peaks, appearing at 45° and 135° indicate orientations of molecules in herringbone sublattices, perpendicular to each other. As consequence, the value of the herringbone order parameter is close to 1 (Fig. 2a).

At 23 K the system undergoes a rotational phase transition. It means that N_2 molecules loose their orientational order and start to rotate around their centers of mass. The value of OP_{HERR} decreases rapidly to 0 (Fig. 2a). The peaks in azimuthal angle distribution first broaden and then disappear (Fig. 3b). The rotation movements compel the N_2 molecules to move away from the substrate. Subsequently, the out-of-plane movements emerge. At the same time, the translational order is preserved: the commensurate order parameter (Fig. 2b) and the Φ_6 order parameter (Fig. 2c) decrease mildly.

Between 23 K and 78 K the molecules start to rotate beyond the plane parallel to the surface. The rotation of molecules becomes more isotropic. The value of OP_{TILT} becomes close to 0 (Fig. 2d)

At about 78 K, the system loses its translational order. The values of OP_{COMM} and Φ_6 order parameters decrease rapidly to 0 (Fig. 2b and c). The $g(r)$ pair distribution function becomes fluid-like (Fig. 3b). The monolayer melts.

Further increasing of the temperature leads to a progressive desorption of molecules from the surface. The process occurs first by promotion of molecules to the second layer, as shown on Fig. 4c, and when the thermal energy of molecules becomes higher than adsorption energy, a molecular gas is formed inside the box. It should be remembered here that, as the simulations are carried out in canonical ensemble, the total number of molecules in the box remains constant.

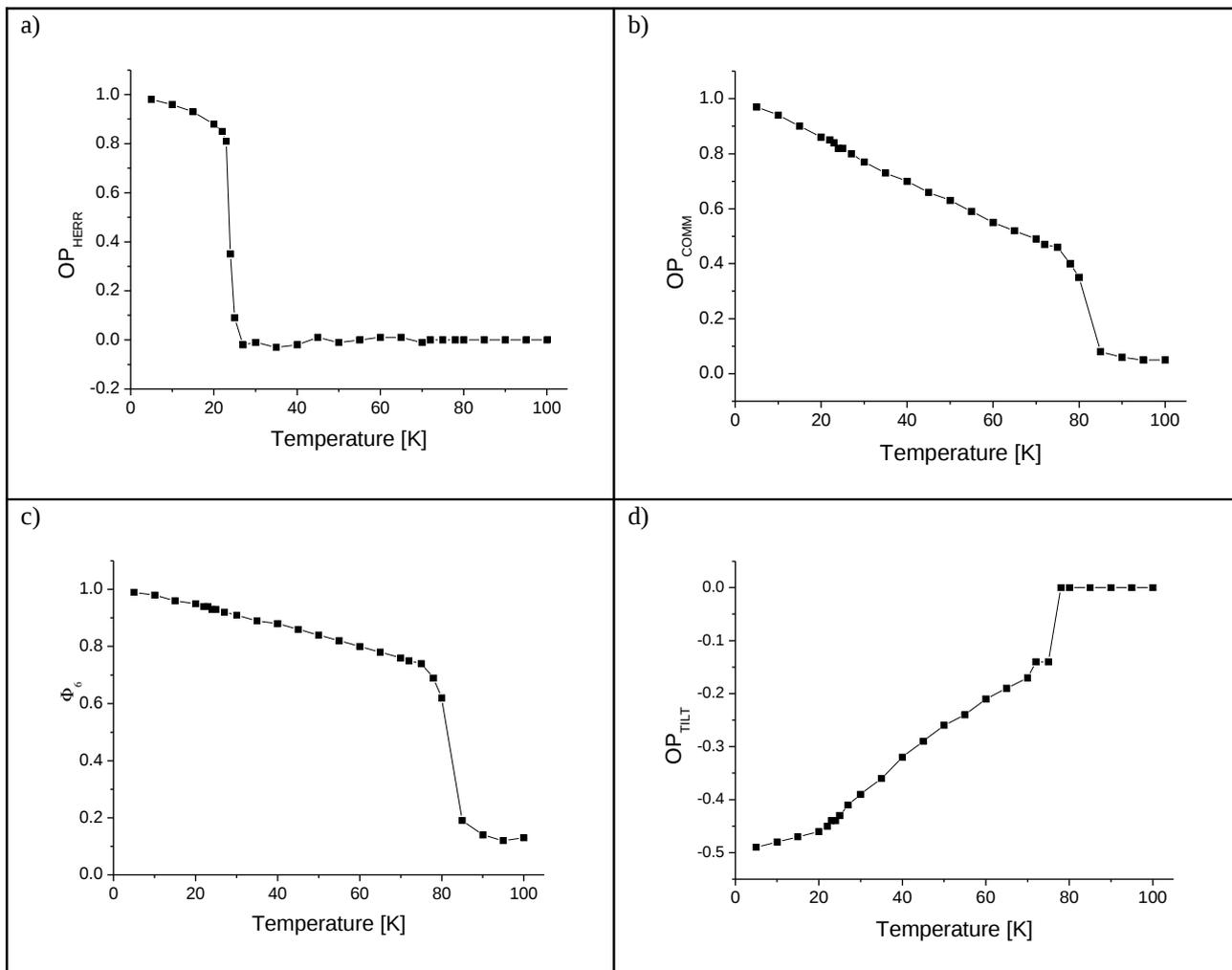


Fig 2 Order parameters plotted as a function of temperature.

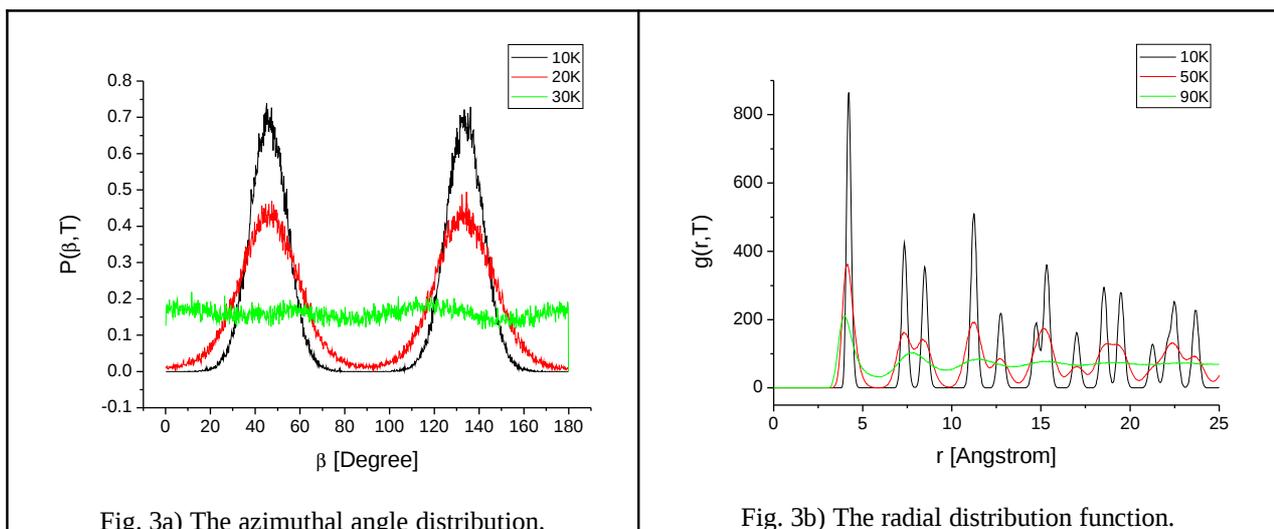


Fig. 3a) The azimuthal angle distribution.

Fig. 3b) The radial distribution function.

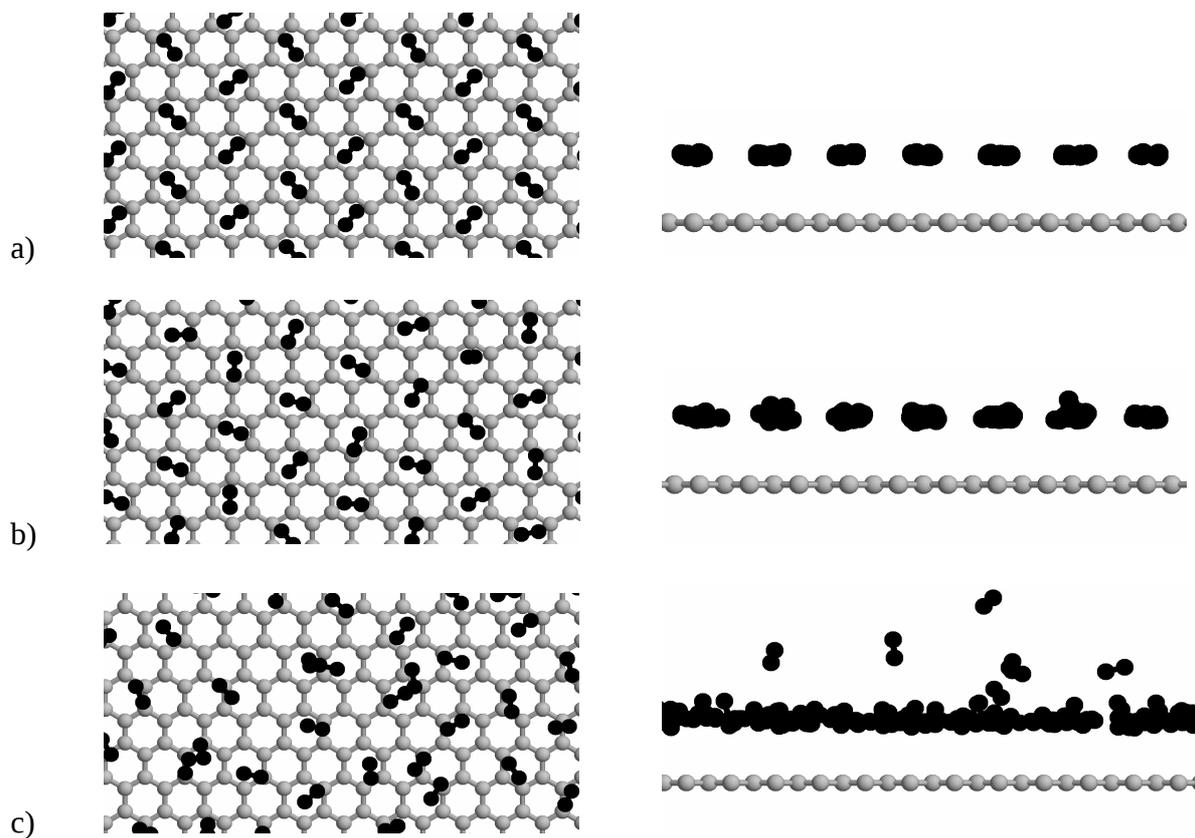


Fig. 4 The nitrogen monolayers over graphite surface at a) 5 K, b) 25 K and c) 85 K.

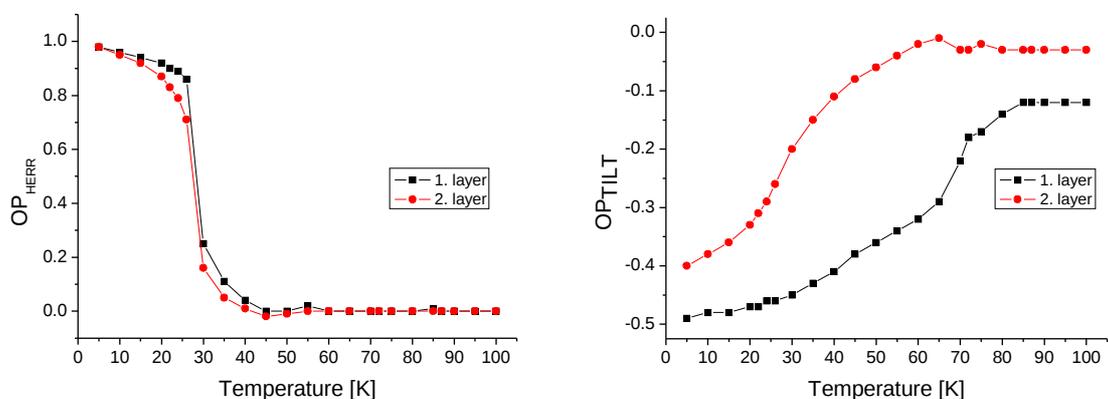


Fig. 5 The herringbone and tilt order parameters for the bilayer plotted as a function of temperature.

Work in progress

The results presented above are in agreement with data referred in previous studies. [5] They will be used as the reference point to investigate the behavior of multilayer systems. Fig. 5 shows the first data obtained for bilayer system. Even these preliminary results show that the temperature and mechanism of both phase transition vary with surface coverage and change from layer to layer.

Further analysis of the data is now in progress.

References

1. Bottani, E.J. and Bakaev, V.A., *Langmuir* **1994**, 10, 1550-1555
2. Kamakashi, J. and Ayappa, K.G., *Langmuir* **2001**, 17, 5245-5255
3. Vernov, A.V. and Steele, W.A., *Langmuir* **1986**, 2, 606-612
4. Kuchta, B. and Etters, R.D., *Phys. Rev. B* **1987**, 36, 3400-3404
5. Vernov, A.V. and Steele, W.A., *Langmuir* **1986**, 2, 219-227
6. Etters, R.D., *Phys. Rev. B* **1986**, 33, 8615-8621
7. Steele, W.A., *Surf. Sci.* **1973**, 36, 317

PUITS QUANTIQUES NATURELS DANS SiC

TEDDY ROBERT¹ (DOCTORANT 2^E ANNÉE), JEAN CAMASSEL², SANDRINE JUILLAGUET³,
POLYCHRONIADIS⁴

Efstathios

^{1,2,3} Laboratoire du Groupe d'étude des Semiconducteurs
Université Montpellier II, CC074 Place Eugène Bataillon, 34095 Montpellier Cedex 5, France

⁴ Département of physics, Aristotle University of Thessaloniki, Thessaloniki 54124 Greece

[1 ROBERT@GES.UNIV-MONTP2.FR](mailto:1.ROBERT@GES.UNIV-MONTP2.FR)

[2 JEAN.CAMASSEL@GES.UNIV-MONTP2.FR](mailto:2.JEAN.CAMASSEL@GES.UNIV-MONTP2.FR)

[3 JUILLAGUET@GES.UNIV-MONTP2.FR](mailto:3.JUILLAGUET@GES.UNIV-MONTP2.FR)

[4 POLYCHR@AUTH.GR](mailto:4.POLYCHR@AUTH.GR)

[HTTP://WWW.GES.UNIV-MONTP2.FR/](http://www.GES.UNIV-MONTP2.FR/)

Résumé— Le carbure de silicium peut se cristalliser sous différentes structures : c'est le polytypisme. Cette faculté rend possible la formation spontanée de puits quantique. Ces puits quantiques apparaissent comme des fautes d'empilement dans le polytype 4H-SiC, qui se développent dans le cristal soit pendant la croissance soit pendant le fonctionnement des composants. Ces puits quantiques sont bidimensionnels de type II. Leur nature et leur épaisseur peuvent être déterminées à partir de la corrélation entre les mesures expérimentales et le calcul des énergies dans le puits quantique. Les techniques expérimentales utilisées sont la photoluminescence et la microscopie électronique en transmission.

Mots clés— Puits quantique; Faute d'empilement; Photoluminescence; Carbure de silicium ; TEM

I. INTRODUCTION.

Le carbure de silicium (SiC) est un semiconducteur IV-IV intéressant pour le développement de composants de puissance fonctionnant dans un large domaine de fréquence et utilisables à des températures élevées. Une particularité du SiC est son polytypisme. Chaque polytype possède des propriétés physiques différentes. Les polytypes les plus étudiés sont le 6H-SiC et le 4H-SiC qui sont deux polytypes hexagonaux et le 3C-SiC qui est un polytype cubique. Leur gap peut varier de 2.3 à 3.4 eV. Le 4H-SiC est préféré au 6H-SiC, en raison d'une meilleure mobilité des électrons. Cependant, dans 4H-SiC, des instabilités dans le fonctionnement des composants ont été associées à la présence de fautes d'empilement notées SF (Stacking Faults). A cause de leur différence de gap, plus petit que celui du polytype, ces fautes se comportent comme des puits quantiques, qui piègent les électrons, ce qui entraîne un échauffement localisé du composant. Son fonctionnement est ainsi limité.

Deux techniques expérimentales sont utilisées pour caractériser les fautes : la photoluminescence (PL) et la microscopie électronique en transmission (TEM). La PL est une méthode optique non destructive. Elle consiste à exciter les électrons de l'échantillon à l'aide d'un laser et à récolter la lumière émise par ce dernier à l'aide d'un spectromètre et d'une caméra CCD. Les spectres obtenus renseignent sur la position des niveaux d'énergie dans le

puits quantique. La TEM est une méthode de microscopie basée sur le principe de la diffraction des électrons avec un grossissement élevé. Elle consiste à placer un échantillon suffisamment mince sous un faisceau d'électrons afin de visualiser, soit la figure de diffraction dans le plan focal de l'objectif, soit la transformée de fourrier de cette figure par l'impact des électrons sur un écran fluorescent. Ces images nous donnent alors des informations sur l'organisation cristalline.

Dans ce travail, nous décrivons la structure des polytypes SiC et celle des SFs. Puis nous vérifions la nature et la signature optique des SFs dans le 4H-SiC. Enfin nous évaluons son gap optique à partir du calcul de puits quantique. Dans ce calcul, nous prenons en compte les effets du champ électrique qui existent dans SiC.

II. ASPECTS STRUCTURAUX.

II.1. SiC polytype.

Certains cristaux ont la faculté de se former dans des structures différentes par un mode d'empilement (translation, rotation, ...) d'un module (couches d'atomes, feuillets, ...) le long d'une direction particulière : c'est le polytypisme. Dans le cas du SiC, le module est une bicouche Si-C. Une bicouche est un empilement compact d'atomes de C et d'atomes de Si. Chaque atome se retrouve au centre d'un tétraèdre (Fig. 1(a)). C'est la séquence d'empilement des bicouches, le long de la direction perpendiculaire au plan de base (c-axis), qui définissent les différents polytypes de SiC. En considérant qu'une bicouche forme une sphère dure (Fig. 1(b)), il n'existe que 3 types de plans compacts pour empiler les sphères : A, B et C. Beaucoup de combinaisons d'empilement sont alors possibles. Chaque combinaison forme un polytype. A ce jour, on en dénombre 200. Les polytypes peuvent avoir une structure cristallographique cubique, hexagonal ou rhomboédrique. Par exemple, le 4H-SiC est hexagonal (H) et possède une séquence périodique de 4 bicouches (4) : c'est la notation de Ramsdell. Elle consiste à associer le nombre de bicouches (1,2,3,4,...) de la période du polytype au symbole correspondant au système cristallin (C pour cubique, H pour hexagonal, R pour rhomboédrique,...).

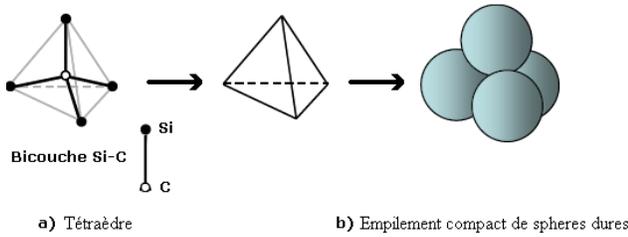


Fig. 1. a) Tétrahédre où tous les atomes de Si et de C sont tétravalents Fig. 1(a); b) Empilement compact de sphères dures Fig. 1(b)

Une autre notation simplifiée, celle de Hägg, est utilisée pour représenter l'enchaînement relatif entre deux bicouches. Les enchaînements AB, BC, CA sont notés + et les enchaînements inverses AC, BA, CB sont notés -. Par exemple, le 4H-SiC a une séquence ABCB et est noté (+ +--). D'autres polytypes sont représentés dans le Tableau 1. On les visualise clairement dans le plan (11-20) (Fig. 2).

Polytype	Séquence	Hägg
2H-SiC	AB	+ -
3C-SiC	ABC	+++
4H-SiC	ABCB	+ +--
8H-SiC	ABCABACB	+ + + - - - -

Tableau 1. Exemple de différents polytypes.

II.2. Faute d'empilement.

Une faute d'empilement est un défaut planaire bidimensionnel. Elle se développe facilement dans le cristal de SiC à cause de sa faible énergie de formation de l'ordre de 15mJ/m² [1]. La conséquence directe de la formation d'une SF est la modification locale de la séquence d'empilement du polytype. La séquence modifiée est provoquée par un glissement de plan entre deux bicouches. La partie du cristal au dessus du plan se déplace par rapport à la partie en dessous (Fig. 3). Ceci est appelé une faute de Shockley notée SSF. Ce glissement entraîne alors la transformation d'une bicouche A, B ou C respectivement en une bicouche B, C ou A ou en bicouche C, A ou B. Par exemple, la séquence parfaite du 4H-SiC, ABCBABC... devient après une SSF (|), ABCBA|CAC... dans lequel B et C se transforme obligatoirement en C et en A. En notation de Hägg, un + devient un -, et vice-versa. Cette séquence modifiée représente alors la séquence d'un autre polytype (séquences encadrées sur la Fig. 3). Théoriquement, plusieurs combinaisons de SSFs, donc plusieurs polytypes, peuvent se former et la plupart des séquences fautées sont possibles. Néanmoins, expérimentalement, seules des séquences de 3C [2] et de 8H [3] ont été observées dans le 4H-SiC.

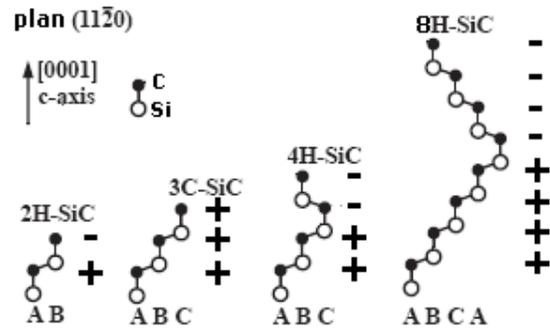


Fig. 2. Séquence d'empilement des différents polytypes dans le plan (11-20)

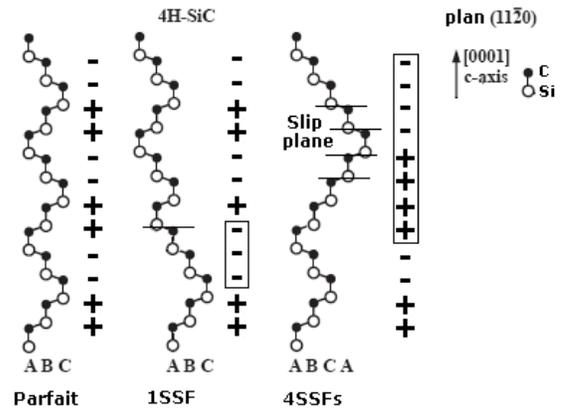


Fig. 3. Séquence de faute d'empilement dans le 4H-SiC : 1SSF et 4SSFs

III. CARACTÉRISATION DES FAUTES.

Dans ce travail, nous observons un échantillon de 4H-SiC épitaxié dans la direction (0001), non intentionnellement dopé. Nous examinons d'abord l'échantillon en TEM pour analyser la séquence des fautes, puis en PL pour déterminer leur gap optique.

III.1. Microscopie électronique en transmission

Les images obtenues par TEM vont nous permettre de déterminer la nature de la SF en examinant sa séquence d'empilement. La Fig. 4(a) montre une image de TEM en cross-section (direction [0001]) en faible résolution sur lequel la densité des SFs est élevée. Les fautes apparaissent en ligne noire à l'intérieur de la région grise qui représente la matrice de 4H-SiC. Pour avoir une structure plus détaillée des SFs, nous utilisons le TEM en haute résolution (HR). Dans ce cas, on augmente le facteur de grossissement et la séquence des bicouches de la faute apparaît. La Fig. 4(b) montre une image de HRTEM en cross-section : chaque spot que l'on voit représente une bicouche de Si-C et chaque rangée de spot représente une rangée de bicouche dans le même plan. On reconnaît la structure d'empilement de la matrice de 4H-SiC. Concernant la faute, l'image en HRTEM (Fig. 5) nous indique une séquence fautée de 8 bicouches : c'est une séquence de 8H. La séquence d'empilement du polytype 4H-SiC (+ + - / + + - / + + -) est transformée par

4SSFs en une cellule unitaire de 8H (++/---+---/---). Actuellement, aucune autre épaisseur que 8 bicouches de 8H n'a été observée. Néanmoins, dans certaines zones de l'échantillon, les fautes sont relativement proches. Le nombre de bicouches de la matrice de 4H-SiC, séparant deux fautes de 8H consécutives, peut être suffisamment petit pour que ces fautes soient couplées ou non.

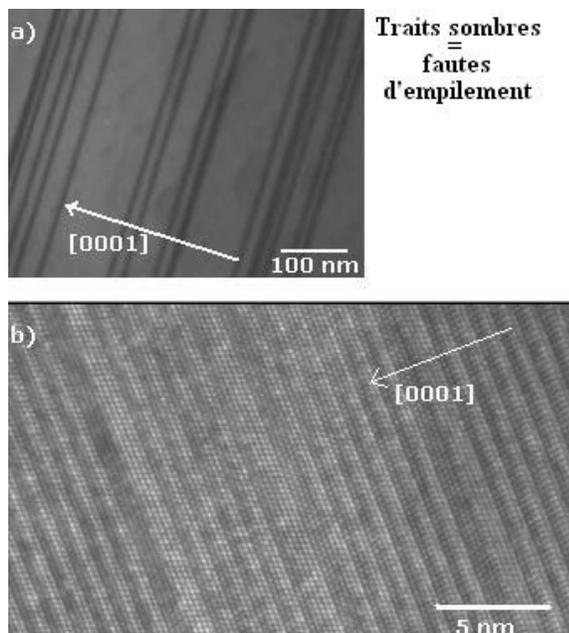


Fig. 4. Image de TEM en cross-section de l'échantillon de 4H-SiC avec des fautes de 8H ; a) faible résolution Fig. 4(a); b) haute résolution Fig. 4(b)

III.2. Photoluminescence.

L'étude de photoluminescence à basse température (5K) va nous permettre de déterminer le polytype de la matrice et d'identifier les fautes. Lorsque l'on éclaire une zone de l'échantillon SiC avec un laser de 244 nm, des excitons (paires électrons-trous) sont créés. Lorsqu'ils se recombinent, ils émettent alors des photons avec une certaine énergie. Comme dans le SiC les gaps sont indirects, une partie de l'énergie émise est dissipée sous forme de phonons (vibrations du réseau cristallin). Sur un spectre de photoluminescence, on s'attend donc à voir des pics d'intensités différentes qui correspondent aux énergies ou aux longueurs d'onde des photons émis. Sur la (Fig. 6), vers 380-400 nm, on observe la signature optique du polytype 4H-SiC et vers 460-510 nm, celle de la faute de 8H. Les pics nommés TA, LA, TO et LO sont les phonons de la faute. A partir de l'énergie de ces phonons, on évalue le gap optique de la faute Egx. Egx est l'énergie à laquelle les photons sont émis pour la faute. Elle est comprise entre le gap du 4H (Eg(4H)) et le gap du 3C (Eg(3C)) qui possède le plus petit gap de tous les polytypes. Elle vaut 2.67 eV.

IV. Puits quantique de 8H ou de 3C ?

Tous les calculs de structure de bande [4] et les observations faites en luminescence [5] ont montré que

dans le 4H-SiC il existe des fautes de 3C et de 8H. Une SF de 3C ou de 8H se comporte dans le 4H-SiC comme un puits quantique bidimensionnel puisque dans ce cas, le gap énergétique de la faute est toujours plus petit que celui du 4H-SiC. Le puits est de type II (Fig. 7(a)).

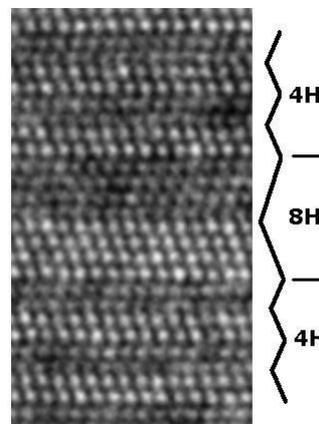


Fig. 5. Image de HRTEM en cross-section de la faute dans le 4H-SiC

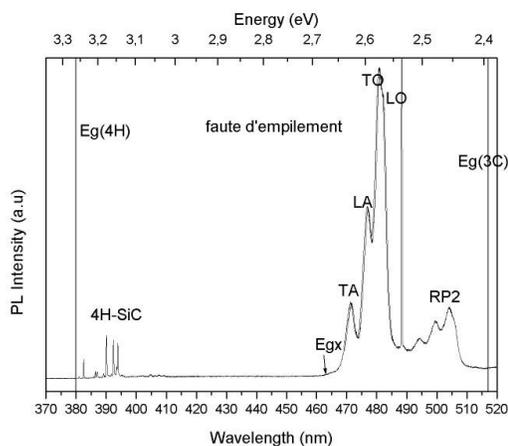


Fig. 6. Spectre de photoluminescence à 5K d'une faute de 8H avec phonons TA, LA, TO, LO et la réplique à 2 phonons (RP2). Le pic fin à 488 nm est le pic dû au laser

Les électrons de la bande de conduction sont confinés dans le puits. Les trous de la bande de valence restent localisés à l'extérieur près des interfaces à cause de l'interaction électron-trou. En effet, nous utilisons un modèle de puits quantique dans lequel 2 paramètres sont importants : la largeur du puits et le champ électrique. La largeur du puits c'est l'épaisseur de la faute en bicouche. Le champ électrique est le champ existant à l'intérieur de la faute. Il est dû à la différence de polarisation spontanée entre le polytype de la matrice et de la faute [6]. Le champ incline donc le potentiel du puits de ΔV et les énergies de transition sont abaissées. On s'attend à avoir deux transitions radiatives possibles : Egx- et Egx+ (Fig. 7(b)). La transition la plus probable est Egx- car elle possède la plus petite énergie. Le champ électrique a été déterminé à partir des résultats expérimentaux de fautes de 3C dans du

4H-SiC dont on connaissait à la fois leurs épaisseurs et leurs gaps optiques [5]. Il a été évalué à 1.2 MV/cm. On sait que le champ électrique est proportionnel à l'hexagonalité du polytype [6]. Nous l'évaluons dans le 8H à 0.6 MV/cm.

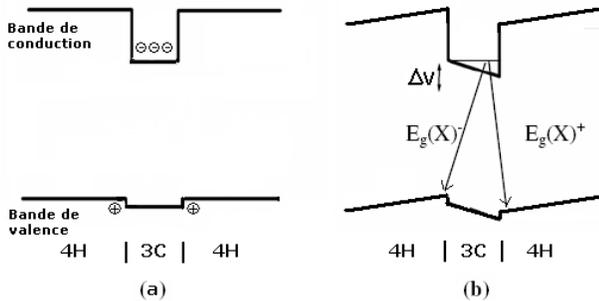


Fig. 7. Puits quantique bidimensionnelle de type II ; a) sans champ électrique ; b) avec champ électrique

Dans notre échantillon de 4H-SiC, le gap optique de la faute de 8H est de 2.67 eV. Cette valeur n'est pas en accord avec notre modèle où on trouve 2,86 eV. Deux hypothèses sont possibles. Soit le champ électrique expérimental n'est pas correct, soit la séquence de 8H peut se lire d'une autre façon. Concernant le champ électrique expérimental, il n'est pas raisonnable de l'augmenter si on veut rester cohérent avec l'hexagonalité du polytype. Par contre on peut montrer que le 8H (----++++) peut se modéliser comme une structure de 3C/2H/3C (---/-+/+++). La structure obtenue est un système de 2 puits quantiques couplés (Fig. 8). Les électrons sont confinés dans les deux puits de 3C. Les trous sont confinés dans la barrière de 2H. 4 transitions sont alors possibles avec 2 transitions identiques. La plus probable est celle où les trous et les électrons sont tous les deux confinés. L'énergie de recombinaison correspondante vaut 2.7 eV ce qui est relativement proche de la valeur expérimental de 8H de 2.67 eV.

V.CONCLUSION.

Les fautes d'empilement dans le SiC sont des puits quantiques naturellement parfaits sans défauts d'interfaces. Dans le 4H-SiC, la nature de la faute de 8H a été déterminée par la microscopie électronique en transmission. Son gap optique a été évaluée à l'aide de la photoluminescence. A partir du modèle physique, nous avons vu qu'une faute de 3C dans du 4H-SiC se comporte comme un simple puit quantique. Tandis qu'une faute de 8H dans du 4H-SiC se comporte comme deux puits quantiques couplés.

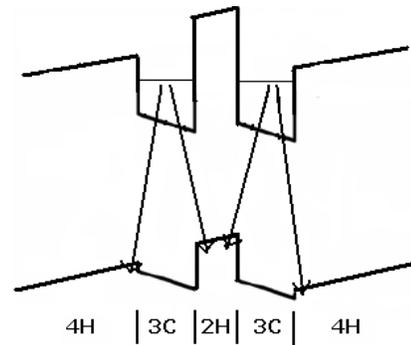


Fig. 8. Structure de 8H : système de 2 puits quantiques couplés de 3C avec champ électrique séparée par une barrière de 2H

REFERENCES.

- [1]M.H. Hong, A.V. Samant, P. Pirouz, dans Philos. Mag. A, 2000, 80, 919.
- [2]Skowronski M. and Ha S., dans J. Appl. Phys., 2006, 99, 011101.
- [3]S. Izumi, H. Isuchida, dans Appl. Phys. Letters, 2005, 86, 202108.
- [4]H. Iwata, U.Lindefelt, dans J. Appl. Phys., 2003, 93, 1577.
- [5]J. Camassel, S. Juillaguet, dans Mater. Sc. Forum, 2005, 331, 483-485
- [6]A. Qteish, V. Veine, dans Phys. Rev. B, 1992, 45, 6534.
- [7]J. Camassel, S. Juillaguet, dans J. Phys. D, 2007, 40, 6264-6277.

DOCTISS 2008

Simulation numérique d'écoulements turbulents

Hilde OUVRARD¹, Bruno KOOBUS²

Institut de Mathématiques et de Modélisation de Montpellier
Université Montpellier 2
Case Courrier 051, Place Eugène Bataillon
34095 MONTPELLIER Cedex , France

¹houvrard@math.univ-montp2.fr, ²koobus@math.univ-montp2.fr

Résumé— Parmi les différentes approches disponibles pour la simulation d'écoulements turbulents, une méthode récente de simulation des grandes échelles, basée sur une formulation variationnelle multi-échelles sera présentée. Pour mesurer les performances de cette approche, nous l'appliquerons au calcul de l'écoulement autour d'un obstacle en 3D dont on connaît bien le comportement. On utilise pour le calcul un schéma mixte élément fini/volume fini du second ordre avec un maillage de type non-structuré.

Mots-clés— turbulence, Navier-Stokes compressible, simulation des grandes échelles, cylindre circulaire.

I. INTRODUCTION

La simulation numérique est un outil de plus en plus utilisé à la fois pour l'analyse des phénomènes physiques, mais aussi pour la conception en ingénierie. Ceci est du tout d'abord à l'augmentation de la puissance de calcul des ordinateurs mais surtout au développement de modèles de turbulence plus performants. La qualité d'un modèle est d'abord testée sur un problème simple et bien connu, afin de pouvoir être appliqué par la suite à des problèmes réels en industrie.

Pour simuler un écoulement turbulent, il est possible de résoudre les équations qui décrivent l'état du fluide sans utiliser de modèle de turbulence. Cette méthode est appelée simulation numérique directe (DNS). Il s'agit d'une méthode très précise mais extrêmement coûteuse en temps de calcul et d'autant plus lorsque le fluide est très turbulent. Elle n'est donc pas utilisable pour les simulations industrielles à géométries complexes.

La méthode RANS (Reynolds Averaged Navier-Stokes) est l'approche la plus utilisée. C'est une méthode statistique dans laquelle les équations de la turbulence sont moyennées, c'est-à-dire que les variables sont séparées en une partie moyenne qui sera calculée et une partie fluctuante qui sera modélisée. Son coût est peu élevé mais elle ne fournit pas de très bons résultats.

La simulation des grosses structures LES (Large Eddy Simulation) est un modèle déterministe qui représente un compromis entre les méthodes DNS et RANS. Cette méthode consiste à différencier les échelles spatiales. Seules les grosses structures sont simulées et un modèle de sous-mailles est ajouté pour les petites échelles. Cette approche fournit des résultats plus précis que la méthode RANS. Elle présente en revanche certaines difficultés lorsque l'on

cherche à simuler l'écoulement autour de corps non-profilés comme les cylindres circulaires. Pour pallier en partie ces difficultés, de nouveaux modèles de sous-mailles ont été développés [7] [6].

Parallèlement, une nouvelle approche de simulation des grosses structures VMS-LES (Variational Multi-Scale LES) a été présentée [3] [4]. L'opération de filtrage des variables pour la LES traditionnelle est remplacée par une projection variationnelle. Cette méthode est conçue pour simuler des écoulements dans lesquels se forment des structures tourbillonnaires de grandes tailles, ce qui est typiquement le cas de l'écoulement autour de cylindres circulaires. Pour évaluer la qualité de l'approche VMS-LES [5] associé au modèle de sous-mailles WALE (Wall-Adapting Local Eddy-viscosity) [6], on considère l'écoulement d'un fluide compressible turbulent autour d'un cylindre circulaire. La simulation est calculée avec un schéma mixte élément fini/volume fini. On utilise un maillage non-régulier dit non-structuré. Ce type de maillage est généralement utilisé pour des problèmes présentant une géométrie plus complexe.

Nous présenterons en partie II le modèle VMS-LES plus en détails. La partie III sera consacrée au schéma numérique utilisée pour cette simulation. Enfin la partie IV traitera du cas test et des résultats de la simulation.

II. MODÉLISATION DE LA TURBULENCE

On considère un fluide parfait Newtonien compressible, c'est-à-dire un gaz ayant une viscosité constante ou dépendant éventuellement de la température, dont la masse volumique n'est pas constante. On s'intéresse à l'écoulement de ce fluide à l'intérieur d'un domaine Ω en trois dimensions sur un intervalle de temps I . Dans l'approximation des milieux continus, les équations de Navier-Stokes décrivent l'état d'un fluide à chaque instant et en tous points du domaine. Elles peuvent être écrites de la manière suivante :

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho u) = 0, \quad (1)$$

$$\frac{\partial (\rho u)}{\partial t} + \nabla \cdot (\rho u \otimes u) = -\nabla P + \nabla \cdot \sigma, \quad (2)$$

$$\frac{\partial E}{\partial t} + \nabla \cdot [(E + P)u] = -\nabla \cdot (\sigma u) + \nabla \cdot (\lambda \nabla T), \quad (3)$$

Les grandeurs qui déterminent l'état d'un fluide sont la vitesse u suivant les trois directions, la densité ρ , la pression P , l'énergie totale E et la température T . La température et l'énergie pouvant être déduites de la densité et de la pression. Les équations (1),(2) et (3) décrivent respectivement la conservation de la masse dans le domaine, la conservation de la quantité de mouvement (loi fondamentale de la dynamique) et la conservation de l'énergie. On désigne par σ le tenseur des contraintes visqueuses et par λ la conductivité thermique.

L'écoulement d'un fluide est dit turbulent lorsqu'il présente des mouvements tourbillonnaires, désordonnés et aléatoires. Le caractère turbulent est déterminé par le nombre de Reynolds $Re = \frac{\rho V D}{\mu}$ où ρ est la masse volumique du fluide, μ la viscosité du fluide, V et D une vitesse et une longueur de référence qui dépendent du problème d'écoulement considéré. Plus ce nombre est grand, plus les instabilités générées par les termes non-linéaires de l'équation de Navier-Stokes seront importantes, autrement dit plus le fluide sera turbulent.

Un écoulement turbulent est constitué de tourbillons de tailles diverses qui possèdent des comportements très différents. Les mécanismes de formation des tourbillons expliquent la construction du modèle LES. Les gros tourbillons conduisent les tourbillons de petites tailles dans l'écoulement et ont une durée de vie longue. Les petits tourbillons naissent par l'étirement des structures de grandes tailles. Ils ont une durée de vie plus courte et possèdent un caractère isotrope localement. Leur comportement universel justifie dans l'approche LES, la modélisation de leurs effets sur le reste de l'écoulement.

Dans l'approche VMS-LES, le mouvement est séparé a priori, c'est-à-dire avant le départ de la simulation. Il est composé, comme dans la méthode LES d'une partie qui sera résolue et d'une partie modélisée. La partie résolue est ensuite séparée en une partie pour les grandes échelles et une partie petites échelles. Les tourbillons génèrent des tourbillons de plus petites tailles et leur transfèrent de l'énergie, jusqu'à ce que les tourbillons créés soient assez petits pour être dissipés par la viscosité. Le principe de la méthode VMS-LES repose donc sur la modélisation des effets des échelles non-résolues sur les petites échelles résolues. On obtient ainsi moins de dissipation dans l'écoulement. On utilise le modèle de sous-mailles de WALE [6] afin d'obtenir un meilleur comportement près des parois.

III. SCHÉMA NUMÉRIQUE

Afin de résoudre numériquement notre système d'équations aux dérivées partielles, chacune des équations continues doit être discrétisée. En d'autres termes, on ne cherche plus à connaître la valeur exacte des fonctions sur tout le domaine de l'espace et du temps, mais simplement sur un nombre fini de points et d'instant. Il faut donc commencer par se donner une discrétisation du domaine spatial et construire ainsi un maillage, puis une discrétisation du domaine temporel. On applique ensuite un schéma numérique à l'équation, c'est-à-dire un algorithme qui nous permet de

calculer une valeur approchée de la solution en un point et à un instant, à partir de solutions en d'autres points.

Pour la simulation d'écoulements tridimensionnels, il existe de nombreuses méthodes de discrétisation spatiale et d'avancement en temps et de plus les termes de chacune des équations ne sont pas traités systématiquement avec la même approche. Dès lors, chaque équipe qui réalise des simulations de turbulences utilise généralement sa propre méthode numérique pour discrétiser les équations.

Afin de valider les modèles VMS-LES sur lesquels on travaille, on utilise pour la discrétisation spatiale une méthode mixte élément-fini/volume-fini dans laquelle les degrés de liberté sont situés sur les noeuds du maillage. Plus précisément, les termes de diffusion sont traités par éléments finis, les termes de convection nécessite quant à eux, un traitement par volumes finis avec plus de précaution. On utilise pour l'avancement en temps un schéma implicite du second ordre.

IV. ÉCOULEMENT AUTOUR D'UN OBSTACLE

A. Description du cas test

On considère l'écoulement autour d'un cylindre circulaire de diamètre D et de hauteur égale à la taille du domaine suivant la direction z . Le nombre de Mach M_∞ qui est égal au rapport de la vitesse locale du fluide u_∞ à l'entrée du domaine sur la vitesse du son, vaut 0.1. L'écoulement est subsonic, il n'y a pas de chocs qui se forment dans le fluide. Le nombre de Reynolds égal à 3900 dépend ici du diamètre D et de la vitesse u_∞ . Le domaine de calcul est représenté Fig. 1, les flèches désignant le sens de l'écoulement du fluide. On impose des conditions limites identiques à l'entrée et à la sortie du domaine (plans $x = -10$ et $x = 25$), ainsi que sur les plans latéraux du domaine $y = \pm 20$, qui correspondent à des conditions de Steger-Warming [8]. Sur la dernière direction, des conditions limites périodiques sont appliquées, à la surface du cylindre on impose des conditions de non-glissement.

Le domaine est discrétisé avec un maillage non-structuré

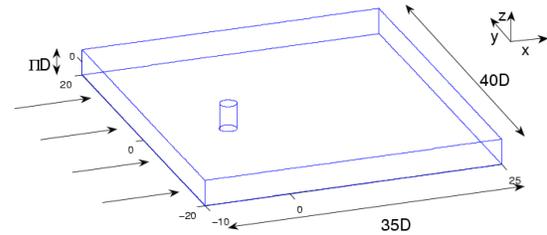


Fig. 1. Domaine de calcul.

constitué de tétraèdres. La Fig. 2 représente une coupe suivant l'axe $z = 0$ du maillage utilisé pour ces calculs. Celui-ci est naturellement plus raffiné autour du cylindre et dans son sillage dans lequel vont se former les tourbillons que l'on cherche à simuler. Il est constitué d'environ 300 000 noeuds, ce qui correspond à un maillage assez grossier. Le modèle de turbulence utilisé pour cette simulation est le modèle VMS-LES avec le modèle de sous-grille WALE. Elle est lancée pendant 24 heures sur 32 processeurs en parallèles au CINES (supercalculateur basé à Montpellier).

Le domaine de calcul est subdivisé en 32 sous-domaines, un processeur calcule l'écoulement dans un seul sous-domaine, s'en suit une phase de communication entre les différents processeurs.

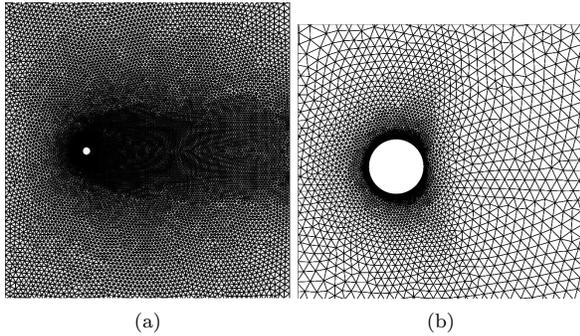


Fig. 2. Coupe du maillage à $z = 0$.

B. Résultats de la simulation

Pour cette valeur du nombre de Reynolds et dans cette configuration, des expériences en laboratoire ont été réalisées et on dispose ainsi d'un certain nombre de grandeurs caractéristiques mesurées [10] [9]. Ce problème a motivé de nombreuses études numériques, entre autres pour la validation de nouveaux modèles de turbulence. Le comportement du fluide est très bien connu. En particulier, l'écoulement présente des instabilités tridimensionnelles dans le sillage du cylindre avec différentes échelles de tourbillons. On visualise Fig. 3 une isosurface de pression instantanée issue de notre simulation qui représente l'ensemble des points soumis à une même pression à un instant donné. On y observe la formation de structures tourbillonnaires, ainsi qu'un phénomène de dissipation des tubes lorsque l'on s'éloigne dans le sillage du cylindre.

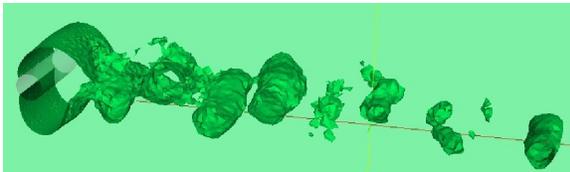


Fig. 3. Isosurface de pression instantanée.

Pour valider un modèle, on souhaite étudier un certain nombre de phénomènes. Citons par exemple le décollement du fluide à la paroi du cylindre, la dissipation des tourbillons et le caractère tridimensionnel de l'écoulement. A l'issue d'une étape de post-traitement des données renvoyées par le programme de calcul, les grandeurs caractéristiques expérimentales sont alors comparées avec celles obtenues lors des simulations. Les grandeurs évaluées dans ce problème sont par exemple, le coefficient de traînée moyen, la longueur de recirculation, le nombre de Strouhal et l'angle de séparation. Le coefficient de traînée moyen est une grandeur relative à la composante suivant la direction x de la force aérodynamique subit par la cylindre. La longueur de recirculation permet de mesurer la taille moyenne de la zone de recirculation située en aval du cylindre. Le nombre de Strouhal est un nombre sans dimen-

sion qui mesure la fréquence de formation des grosses structures. L'angle de séparation est aussi une valeur moyenne qui mesure la position à laquelle le fluide se détache du cylindre avant la formation de tourbillons.

Pour obtenir des résultats exploitables, c'est-à-dire comparables avec l'expérience, on réalise les simulations sur une période pendant laquelle une vingtaine de tourbillons se forment dans le sillage. Les grandeurs caractéristiques calculées pour cette simulation sont très proches des valeurs expérimentales. Fig. 4 représente la composante de la vorticitée instantanée suivant l'axe z . La vorticitée ω est égale au rotationnel de la vitesse du fluide. C'est un vecteur porté par la direction de l'axe de rotation des tourbillons. Il représente la quantité de rotation du fluide. L'axe de rotation est dans ce cas assez proche de la direction z comme on l'observe Fig. 3, ainsi la composante suivant z en 2D nous donne une bonne idée de la formation des tourbillons. On observe très peu de structures de petites tailles, l'emploi d'un maillage aussi grossier ne nous permet probablement pas de capturer tous les détails de turbulence présents dans l'écoulement.

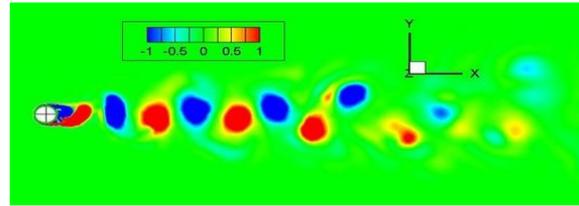


Fig. 4. Vorticitée instantanée ω_z sur la section $z = 0$.

V. CONCLUSION

Nous avons présenté le modèle de turbulence VMS-LES associé au modèle de sous-grille WALE appliqué à la simulation de l'écoulement d'un fluide autour d'un cylindre circulaire. De bons résultats en accord avec l'expérience sont obtenus pour cette simulation sur un maillage assez grossier et avec un schéma numérique d'ordre 2. Il avait été conclu dans d'autres études sur la LES que pour obtenir des résultats précis, il était nécessaire de travailler avec des maillages fins de l'ordre du million de noeuds avec des schémas d'ordre supérieur à 2.

Les résultats présentés sont assez prometteurs pour la simulation d'écoulements à haut nombre de Reynolds, qui sont des écoulements que l'on rencontre en industrie, en utilisant des schémas hybrides dans lesquels ce modèle VMS-LES est combiné avec un modèle statistique RANS.

RÉFÉRENCES

- [1] L. Landau et E. Lifchitz. *Physique Théorique*, Editions MIR, Moscou, 1971.
- [2] S. B. Pope, *Turbulent Flows*, Cambridge University Press, 2000.
- [3] T.J.R. Hughes, L. Mazzei et K.E. Jansen, « Large eddy simulation and the variational multiscale method, » *Comput. Vis. Sci.*, vol. 3, pp. 47–59, 2000.
- [4] T.J.R. Hughes, A.A. Oberai et L. Mazzei, « Large eddy simulation of turbulent channel flows by the variational multiscale method, » *Phys Fluids*, vol. 13, pp. 1784–1799, 2001.
- [5] B. Koobus et C. Farhat, « A variational multiscale method for the large eddy simulation of compressible turbulent flows on unstructured meshes, application to vortex shedding, » *Comput. Methods Appl. Mech. Engrg.*, vol. 193, pp. 1367–1383, 2004.

- [6] F. Nicoud et F. Ducros, « Subgrid-scale stress modeling based on the square of the velocity gradient tensor, » *Flow, Turbulence and Combustion*, vol. 62, pp. 183–200, 1999.
- [7] A.W. Vreman, « An eddy-viscosity subgrid-scale model for turbulent shear flow : algebraic theory and application, » *Physics of Fluids*, vol. 16, pp. 3670–3681, 2004.
- [8] J.L. Steger et R.F. Warming, « Flux vector splitting for the inviscid gas dynamic equations with applications to the finite difference methods, » *J. Comp. Phys.*, vol. 40(2), pp. 263–293, 1981.
- [9] L. Ong et J. Wallace, « The velocity field of the turbulent very near wake of a circular cylinder, » *Exp. in Fluids*, vol. 20, pp. 441–453, 1996.
- [10] L.M. Lourenco et C. Shih, « Characteristics of the plane turbulent near wake of a circular cylinder, a particle image velocimetry study, » (data taken from Kravchenko and Moin 1994), 1993.

DOCTISS 2008

Billards & Beignets, Bretzels, etc. Surfaces plates et uniformisation explicite

Olivier RODRIGUEZ¹ (Doctorant 2ème année),
Emmanuel ROYER², Robert SILHOL³ (Directeurs de Thèse)

Institut de Mathématiques et de Modélisation de Montpellier (UMR 5149)
Université Montpellier 2, place Eugène Bataillon, 34090 Montpellier, France

¹orodrig@math.univ-montp2.fr,
²royer@math.univ-bpclermont.fr, ³rs@math.univ-montp2.fr

Résumé— Ce texte est un essai de vulgarisation destiné à présenter les notions de surface plate et de surface de Riemann et quelques sujets connexes. On expose en premier lieu une construction topologique du tore et la propriété des trajectoires du flot géodésique sur celui-ci (théorème de Weyl). Il existe d'autres surfaces plates vérifiant cette propriété appelée alternative de Veech, dont on donne un exemple non trivial *a priori*. On indique ensuite comment l'étude des surfaces plates permet d'aborder sous un autre angle le problème de l'uniformisation des surfaces de Riemann.

Mots-clés— Surface plate, billards, surface de Riemann, uniformisation.

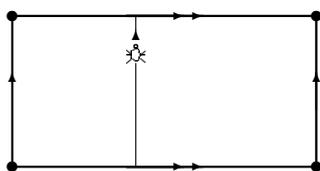
I. INTRODUCTION

La première partie de ce texte est consacrée à une brève présentation de la notion de surface plate et le lien existant entre les trajectoires sur ces dernières et les billards polygonaux. On aborde ensuite les concepts de surface (topologique) et de surface de Riemann, pour ensuite effleurer une facette du problème de l'uniformisation.

Tirillé entre la précision qu'offre la rigueur du formalisme mathématique et le désir de faire découvrir aux non initiés quelques concepts surprenants, l'auteur a choisi de présenter ces derniers dans un style voulu accessible au plus grand nombre. Néanmoins, pour ne pas sacrifier à l'exactitude, certains résultats sont cités en note de bas de page et pourront être omis en première lecture.

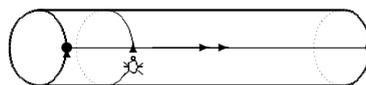
II. TRAJECTOIRES SUR UN TORE

Vous connaissez sans doute ce genre de jeu vidéo, où le personnage (ici, un insecte) se déplace sur un écran statique, soit horizontalement, soit verticalement, de telle sorte qu'une fois arrivé sur un bord de l'écran, il réapparaisse en face sur le bord opposé ?

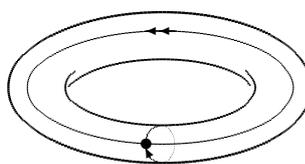


Lorsqu'on se déplace verticalement de bas en haut, le "saut" d'un bord à un autre peut paraître assez angoissant. Pour

pallier à cela, recollons les deux bords horizontaux : on s'aperçoit que le personnage se retrouve en fait à se déplacer sur un cylindre :

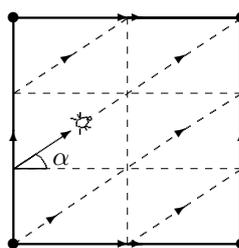


Et pour les trajets horizontaux ? qu'à cela ne tienne, identifions les deux bords circulaires du cylindre, qui étaient les côtés verticaux du rectangle :



Pour effectuer idéalement ce recollement, il faut s'imaginer que le rectangle original est fait d'une matière infiniment élastique, de façon à éliminer les plis qui ne manqueraient pas d'apparaître si vous faisiez la construction avec du papier. La surface qu'on obtient et sur laquelle l'insecte se déplace a la forme d'une chambre à air, ou d'un beignet : c'est ce qu'on appelle un *tore*. Notons au passage que les quatre sommets du rectangle sont identifiés en un seul point du tore. Formellement, la construction qu'on vient de décrire peut en fait s'effectuer à partir d'un carré.

Il est clair que s'il se contente de se déplacer uniquement horizontalement (ou verticalement), l'insecte va sans cesse repasser par son point de départ, en suivant à chaque fois une trajectoire identique à la première. Lui vient alors naturellement la question : « Et si je me déplaçais toujours en suivant une trajectoire rectiligne, mais avec un autre angle de départ ? Serai-je de nouveau condamné à suivre la même trajectoire *ad vitam aeternam* ? ».



S'il était encore vivant, le grand mathématicien Hermann Weyl lui répondrait alors :

« Si l'angle α formé par ta trajectoire et le bord horizontal du carré est un multiple *rationnel* de π , alors tu repasseras, tôt ou tard, par ton point de départ. Après quoi, tu reprendras *exactement* la même route

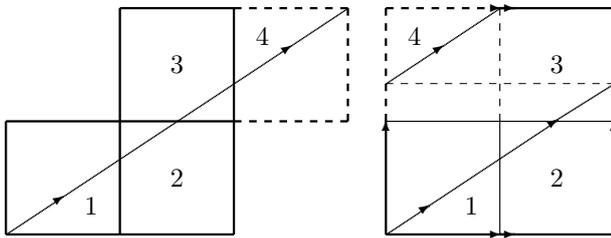
que tu as suivie la première fois, et ainsi de suite jusqu'à ce que mort s'ensuive. Mais dans le cas où cet angle α est

un multiple *irrationnel* de π , alors tu finiras par passer par *presque tous les points* du carré, et tu passeras dans chaque petit morceau du carré un temps proportionnel à l'aire de ce morceau! Il n'y aucune autre alternative...¹ »

« Formidable! », s'écrie alors l'insecte, frappé par cette révélation. « Mais existe-t'il d'autres surfaces que le tore "carré" sur lesquelles cette alternative soit la seule possible? ». Mais là, Weyl serait resté muet...

III. BILLARDS ET SURFACES PLATES

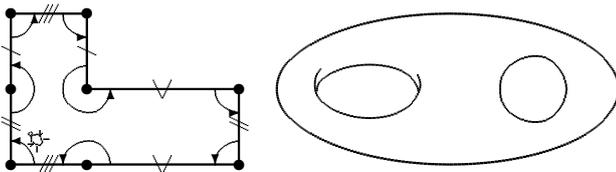
Regardons à présent notre carré comme une table de billard : une bille part du coin inférieur gauche avec une pente ni horizontale, ni verticale et se déplace idéalement selon une trajectoire rectiligne. Elle va finir par heurter le côté droit et rebondir. Mais au lieu de regarder la trajectoire de la bille se réfléchir, imaginons plutôt ce qui se passe si c'est... la *table* qui se réfléchit!



On obtient une image miroir du premier carré, sur laquelle la bille continue sa course, avec la même pente. Lorsqu'elle atteint le côté supérieur de ce nouveau carré, celui-ci est à son tour réfléchi par image miroir, et ainsi de suite. On se retrouve finalement avec quatre copies identiques de notre carré d'origine, qu'on peut réassembler pour former un carré plus gros, dont les côtés opposés sont identifiés : revoilà notre tore.² On comprend ici qu'on peut interpréter les trajectoires "rectilignes" sur le tore en termes de dynamique dans un billard carré.

On peut en fait appliquer la construction précédente à d'autres polygones que le carré : on peut alors obtenir d'autres surfaces, dites *plates*. Et de même, les trajectoires "rectilignes" sur ces surfaces plates peuvent être interprétées en termes de dynamique dans un billard polygonal. On peut obtenir de la sorte d'autres exemples de surfaces vérifiant l'alternative de la section précédente, appelée *alternative de Veech*.

Considérons par exemple les trajectoires dans un billard en forme de L. Ceci revient à considérer les trajectoires rectilignes sur la surface ci-dessous, obtenue en recollant deux-à-deux les côtés marqués d'un même symbole : cette surface *ressemble* au "double tore" de droite.



Un peu comme deux beignets qui auraient cuit ensemble collés l'un à l'autre, ou encore une espèce de bretzel. Par

¹Dans le jargon, on dit que le flot géodésique est soit périodique, soit uniquement ergodique.

²Cet exemple provient de [1].

ailleurs, les huit points noirs se retrouvent identifiés en un seul point... assez spécial. Ce type de points spéciaux est le prix à payer pour obtenir par ce procédé des surfaces plates autres que le tore.

L'insecte, se retrouvant sur cette nouvelle surface et intrigué par le gros point noir, décide d'en faire le tour. Ô stupeur! une fois revenu à son point de départ, il se rend compte (difficilement, certes) qu'il a décrit un angle de 6π en faisant *un seul tour* autour du point noir! Vous pouvez vous en persuader en suivant les identifications des côtés sur la figure et en comptant les angles... Si jamais vous essayez d'effectuer cette construction avec du papier, ce point devrait vous poser pas mal de problèmes.³

Il existe une condition, découverte par William Veech [2] à la fin des années 1980, satisfaite par toutes les surfaces plates qui vérifient l'alternative de Veech.⁴ Cette condition est en général difficile à valider, excepté dans certains cas particuliers, tel celui de la surface de la figure précédente qui vérifie cette fameuse alternative.

IV. SURFACES DE RIEMANN ET UNIFORMISATION

Le tore, comme la sphère, est un exemple de *surface (topologique) compacte orientable*. Nous ne dirons rien sur ces deux derniers termes techniques, si ce n'est qu'ils signifient bel et bien ce qu'ils suggèrent. Sans en donner une définition formelle, on peut dire qu'une surface est un objet mathématique obtenu "en recollant des morceaux de plan déformés". Ainsi, si on se trouve sur une surface, on aura toujours l'impression d'être sur un sol (plus ou moins) plat, et ceci en tout point : c'est précisément pour cette raison que les gens ont longtemps pensé que la Terre était plate...

On peut attribuer à chaque surface un nombre entier appelé le *genre*, qu'on peut voir comme le nombre de "trous", ou plutôt de "poignées", de la surface. Par exemple, le tore est de genre 1, la surface de la figure précédente a pour genre 2, et la sphère 0. Le genre caractérise les surfaces au sens suivant : si deux surfaces ont le même genre, alors on peut déformer de façon élastique l'une des deux pour la transformer en la seconde.⁵ Ainsi, la surface d'un beignet, d'une tasse à café ou d'une bouée sont (topologiquement) équivalentes, car elles sont toutes de genre 1.

Sur une surface, on peut voir les morceaux de plan qu'on a recollés comme des morceaux du plan réel \mathbb{R}^2 . Mais on peut aussi identifier le plan réel au plan des nombres complexes \mathbb{C} : on est alors conduit à la notion de *surface de Riemann*.

La notion de surface de Riemann est un raffinement de celle de surface topologique. En effet, toute surface de Riemann compacte peut être considérée comme un *ensemble de points à coordonnées complexes vérifiant une équation polynomiale à coefficients complexes*⁶, ce qui n'est

³Il s'agit d'une singularité conique de la métrique plate induite par une 1-forme holomorphe dont ce point est un zéro double. En termes exacts, le polygone en forme de L correspond à la donnée de la surface de droite munie de cette 1-forme.

⁴Le stabilisateur d'une telle surface sous l'action de $SL(2, \mathbb{R})$, appelé *groupe de Veech*, est un réseau.

⁵Le genre est un invariant topologique.

⁶Plus précisément, toute surface de Riemann compacte est biholomorphe à une courbe algébrique projective lisse (théorème d'existence de Riemann).

pas le cas d'une surface topologique en général. Il s'agit d'un résultat extrêmement puissant, car il affirme qu'un objet défini de façon *locale* (par des recollements appropriés de morceaux du plan complexe) admet une description *globale* (une "simple" équation polynomiale).

Le tore, par exemple, est une surface de Riemann compacte, et c'est aussi l'ensemble des points de coordonnées (z, w) , avec z et w des nombres complexes vérifiant une équation de la forme $w^2 = z(z-1)(z-\mu)$, où μ est un complexe fixé. Ainsi, bien que toutes les surfaces de genre 1 soient *topologiquement* équivalentes, ce n'est pas le cas *en tant que surfaces de Riemann* (une infinité de μ est possible). L'équation ci-dessus est bien connue des cryptologues, pour qui le tore complexe est une *courbe elliptique*.

L'ensemble des points de coordonnées complexes (z, w) tels que $w^2 = z(z-1)(z-a)(z-b)(z-c)$ où a, b et c sont des nombres complexes fixés, correspond quant à lui à une surface de Riemann compacte de genre 2 (comme celle de la dernière figure), également connue des cryptologues et appelée *courbe hyperelliptique*.

La question qui se pose naturellement est de savoir comment faire *explicitement* la jonction entre l'aspect géométrique (forme et dimensions du polygone) et l'aspect algébrique (calcul des paramètres de l'équation correspondante). Dans le cas du tore, il existe toute une théorie permettant d'y répondre avec une grande précision (voir [3]). Mais dans le cas des surfaces de Riemann de genre supérieur ou égal à 2, il s'agit d'un problème très difficile et jusqu'à présent irrésolu en toute généralité.

Contrairement au tore, les surfaces de Riemann de genre au moins égal à 2 ne sont pas des objets plats par nature : c'est pour cette raison qu'apparaît une singularité sur la surface obtenue à partir du polygone en L de la dernière figure. Cependant, l'étude de la surface plate correspondant à la donnée d'une surface de Riemann⁷ permet dans certains cas de calculer une certaine matrice, appelée *matrice des périodes*, associée à cette surface et qui la caractérise complètement (voir [4]). Il est ensuite possible, à partir de cette matrice, de déterminer avec une grande précision les paramètres de l'équation polynomiale associée, via des fonctions analytiques appelées *fonctions Thêta*, possédant des propriétés de symétrie remarquables.

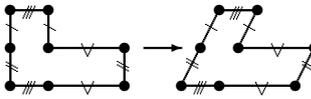
V. FAMILLES DE SURFACES ET APPLICATIONS

L'étude des surfaces plates s'avère précieuse dans l'exploration de la géométrie et la dynamique des *espaces de modules des surfaces de Riemann*, qui "paramètrent" l'ensemble des surfaces de Riemann d'un genre fixé.

On peut voir l'espace des modules comme un "super-espace" dont les points sont des (classes de) surfaces de Riemann d'un même genre. Cette espace admet une géométrie à part entière, il existe en particulier une notion de longueur entre deux points de cet espace : pour se fixer les idées, on pourrait interpréter cette "longueur" comme une quantité minimale d'énergie nécessaire pour déformer une surface en une autre.

Certaines surfaces plates vérifiant l'alternative de Veech possèdent une caractéristique exceptionnelle. Si on représente une telle surface par un polygone du plan, et qu'on considère les déformations de ce polygone conservant les parallèles (figure ci-dessous), alors on obtient toute une

famille de surfaces plates. Il lui correspond une famille de surfaces de Riemann, qui forme en certaines occasions une courbe dans l'espace des modules, appelée *courbe de Teichmüller*. Une telle courbe a la propriété de minimiser localement la distance entre deux de ses points, c'est-à-dire deux surfaces de Riemann de cette famille.⁸



Pour aborder le problème de l'uniformisation décrit à la fin de la section précédente, une approche initiée par Peter Buser et Robert Silhol [5] et

développée par Aline Aigon-Dupuy et Robert Silhol [6], consiste à envisager une "uniformisation en famille". Étant données deux surfaces de Riemann compactes de même genre appartenant à une certaine famille et pour lesquelles on sait explicitement comment déformer l'une en l'autre, si on connaît avec précision les paramètres de l'équation correspondant à l'une des surfaces d'une part, et qu'on sait exprimer, à partir de ceux-ci, les paramètres pour l'autre surface d'autre part, on peut alors obtenir avec exactitude les équations de toutes les surfaces de cette famille.

Les familles considérées dans ces précédents travaux sont définies par le fait que les surfaces les constituant peuvent être pavées par des copies d'un même polygone, comme le carrelage d'un sol. Nous cherchons à présent à obtenir des résultats pour d'autres familles de surfaces de Riemann, en particulier celles décrites ci-dessus et obtenues à partir de familles de surfaces plates en genre 2.

Outre leur importance fondamentale en Mathématiques, les recherches sur les surfaces de Riemann et leurs espaces de modules sont motivées par le fait que ces concepts apparaissent de façon naturelle en Physique théorique, en particulier en théorie conforme des champs, dans les théories des cordes et la théorie M d'Edward Witten. Ces dernières ont pour but d'unifier la relativité générale (qui explique l'infiniment grand) et la mécanique quantique (qui s'intéresse à l'infiniment petit), qui bien qu'étant les théories physiques parmi les plus fiables à ce jour, divergent totalement sur de nombreux points fondamentaux. Nous n'en dirons pas davantage ici, l'auteur se trouvant bien loin d'être un spécialiste de ces questions.

RÉFÉRENCES

- [1] P. Hubert et T. Schmidt. *Introduction to Veech surfaces*, in *Handbook of dynamical systems* vol.1B pp. 501-526. Elsevier B.V., 2006.
- [2] W. Veech. *Teichmüller curves, Eisenstein series and an application to triangular billiards*. *Inventiones Mathematicae* **97** :3 (1989) pp. 553-583.
- [3] R. Godement. *Analyse mathématique IV : Intégration et théorie spectrale, analyse harmonique, le jardin des délices modulaires*. Springer-Verlag, 2003.
- [4] R. Silhol. *Genus 2 translation surfaces with an order 4 automorphism in The geometry of Riemann surfaces and Abelian varieties* pp. 207-213, *Contemporary Mathematics* **397**, AMS, Providence (RI), 2006.
- [5] P. Buser et R. Silhol. *Geodesics, periods, and equations of real hyperelliptic curves*. *Duke Mathematical Journal* **108** (2001), pp.211-250.
- [6] A. Aigon-Dupuy et R. Silhol. *Hyperbolic hexagons and algebraic curves in genus 3*. *Journal of the London Mathematical Society* **66** (2002), pp. 671-690.
- [7] A. Zorich. *Flat surfaces*, in *Frontiers in Number Theory, Physics and Geometry* vol. 1, pp. 439-586. Springer-Verlag, 2006.

⁸L'orbite sous l'action de $SL(2, \mathbb{R})$ d'une 1-forme holomorphe de genre 2 à unique zéro double se projette sur une géodésique complexe dans l'espace des modules si, et seulement si, son stabilisateur est un réseau.

⁷et d'une 1-forme holomorphe

Systoles et triangulations minimales des surfaces

Guillaume BULTEAU¹ (Doctorant 3e année), Ivan BABENKO² (Directeur de Thèse)

Institut de Mathématiques et Modélisation de Montpellier
 Université de Montpellier 2
 Case Courrier 051
 Place Eugène Bataillon
 34095 MONTPELLIER Cedex France

¹guillaume.bulteau@univ-montp2.fr, ²babenko@univ-montp2.fr

Résumé— Ce texte est un essai de vulgarisation et de présentation de la notion de systole sur une surface compacte en résumant certains des résultats obtenus sur le sujet. Il présente ensuite les triangulations minimales de surfaces ainsi qu'un angle d'attaque pour améliorer certaines estimations asymptotiques sur le quotient systolique.

Mots-clés— Surface, systole, quotient systolique, triangulation minimale.

I. INTRODUCTION

La première partie est consacrée à une vulgarisation de la notion de systole sur les surfaces compactes, partie que j'espère à la portée de tous. La seconde, plus technique, donne certains résultats essentiels de la géométrie systolique des surfaces : elle nécessite un peu plus de connaissances. Pour une introduction bien plus poussée à la notion de systole, voir l'article [1] et pour une vision globale des derniers résultats obtenus sur le sujet, voir [2]. La dernière partie présente la notion de triangulation minimale d'une surface ainsi que des applications possibles à la géométrie systolique.

II. GÉNÉRALITÉS ET INÉGALITÉ DE LOEWNER

A. Systole d'une surface compacte

Pour l'instant le mot *surface* désignera une surface compacte plongée dans l'espace à trois dimensions \mathbb{R}^3 . Les exemples classiques de surfaces compactes sont la sphère, le tore.

Un *lacet* désignera ici une courbe simple fermée tracée sur la surface. Un lacet *contractile* est un lacet que l'on peut déformer continûment en un point sur la surface.

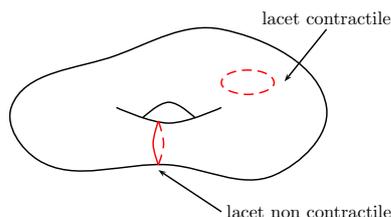


Fig. 1. Un tore.

Pour mesurer des grandeurs telles que l'aire ou la lon-

gueur de courbes sur une surface S plongée dans \mathbb{R}^3 , il n'y a pas de problème : on se sert de la structure euclidienne canonique de \mathbb{R}^3 .

La *systole* d'une surface S est la longueur du plus petit lacet *non contractile* de cette surface. On la désignera dans la suite par $\text{syst}(S)$.

Par extension le mot systole désigne autant tout lacet dont la longueur réalise la systole au sens précédent. Par exemple la sphère de dimension deux n'a pas de systole : tous ses lacets sont contractiles. Un tore de révolution peut avoir deux types de systoles.

B. Systole et aire du tore

En regardant la figure 2, on peut s'attendre au fait que la systole du tore contrôle son aire : si la systole n'est pas trop petite, l'aire du tore ne pourra être arbitrairement petite.

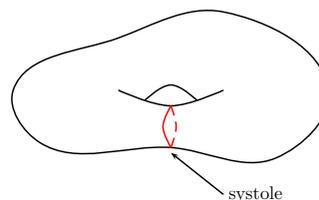


Fig. 2. Systole.

C'est effectivement le cas et on dispose du résultat suivant découvert en 1946 par LOEWNER.

INÉGALITÉ DE LOEWNER. Si $T \subset \mathbb{R}^3$ est un tore plongé alors on a :

$$\frac{\text{aire}(T)}{\text{syst}(T)^2} \geq \frac{\sqrt{3}}{2}.$$

En fait l'inégalité de LOEWNER est bien plus générale et reste valable lorsque T est un tore abstrait de dimension 2 muni d'une métrique riemannienne. De plus le cas d'égalité a lieu si et seulement si T est un tore plat hexagonal c'est à dire obtenu en recollant les côtés opposés du losange de la figure 3.

Les définitions formelles d'une surface abstraite et d'une métrique riemannienne peuvent se trouver dans [3]

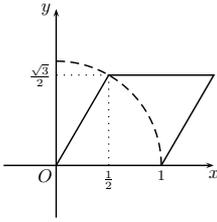


Fig. 3. Le tore hexagonal.

III. LE QUOTIENT SYSTOLIQUE DES SURFACES ABSTRAITES

A partir de maintenant, le mot surface désigne une variété compacte de dimension 2 (surface abstraite). Pour pouvoir mesurer des quantités telles que l'aire et la longueur des courbes sur une surface S on peut la munir d'une métrique riemannienne g . L'aire de S pour cette métrique g se note $\text{aire}(S, g)$ et on peut encore considérer la systole de S pour cette métrique qui sera notée $\text{syst}(S, g)$.

A. Systole et aire des surfaces. Définition du quotient systolique.

Soit S une surface abstraite. Si g est une métrique riemannienne sur S on pose

$$\sigma(S, g) = \frac{\text{aire}(S, g)}{\text{syst}(S, g)^2}.$$

On a alors le résultat suivant. Si S est une surface qui n'est pas homéomorphe à une sphère, munie d'une métrique riemannienne g , alors :

$$\sigma(S, g) \geq \frac{2}{\pi}.$$

Le cas d'égalité est atteint si et seulement si S est le plan projectif (surface obtenue en identifiant les points diamétralement opposés des bords d'un disque) muni d'une métrique de courbure constante -1 .

On peut alors définir le *quotient systolique* de S qui est :

$$\sigma(S) = \inf \sigma(S, g).$$

où l'infimum est pris sur toutes les métriques riemanniennes g sur S .

On ne connaît que trois cas pour lesquels $\sigma(S)$ est atteint : le tore, le plan projectif et la bouteille de Klein (et dans ce dernier cas la métrique est à singularités).

B. Comportement asymptotique en fonction du genre

Brièvement dit, le genre d'une surface mesure le nombre de trous de cette surface. Par exemple la sphère est de genre 0, le tore est de genre 1 et la figure 4 présente une surface de genre 2.



Fig. 4. Surface de genre 2.

Lorsque S_h est une surface de genre h , GROMOV a démontré (voir [2]) que :

$$\sigma(S_h) > \frac{4\sqrt{h} + 27}{64}.$$

On dispose aussi d'un comportement asymptotique du quotient systolique en fonction du genre h . Pour tout λ réel tel que $0 < \lambda < \pi$ et toute surface S_h de genre h on a, pour h suffisamment grand :

$$\sigma(S_h) \geq \frac{\lambda h}{\log^2 h} \quad \text{donc} \quad \sigma(S_h) \underset{+\infty}{\gtrsim} \pi \frac{h}{\log^2 h}.$$

De plus BUSER et SARNAK ont montré (voir [5]) qu'il existe des surfaces S_h de genre h arbitrairement grand telles que :

$$\sigma(S_h) \underset{+\infty}{\lesssim} \frac{9\pi}{4} \frac{h}{\log^2 h} \quad (*)$$

Cette dernière constante $\frac{9\pi}{4}$ est susceptible d'être améliorée et cela fait actuellement l'objet de travaux de ma thèse.

IV. TRIANGULATIONS MINIMALES

A. Introduction

Une idée pour essayer d'améliorer la constante $\frac{9\pi}{4}$ de (*) est de considérer des *triangulations minimales* des surfaces. Une *triangulation* d'une surface compacte S est un recouvrement de S en « triangles », les côtés des triangles étant des courbes sur la surface, et on impose la contrainte suivante : l'intersection de deux triangles est soit vide, soit constituée d'un seul sommet ou d'un seul côté (sommets compris). Lorsque le nombre de triangles est le plus petit possible, on parle alors de *triangulation minimale* de S .

Pour chaque surface S_h de genre h on sait construire une telle triangulation en plongeant des graphes dans la surfaces S_h (voir [4]). Pour une infinité de valeurs de h il est remarquable que c'est le graphe complet K_n (graphe ayant n sommets et dont tous les sommets sont reliés par une arête) qui va engendrer une triangulation minimale de la surface. C'est le cas lorsque n est congru à 0,3,4,7 modulo 12 et le genre de la surface (orientable) est alors donné par :

$$h = \frac{n^2 - 7n + 12}{12}.$$

Par exemple la figure 5 présente la triangulation minimale du tore engendrée par le plongement du graphe complet K_7 dont les sommets sont numérotés de 0 à 6 (je rappelle que le tore s'obtient en identifiant les côtés opposés du rectangle).

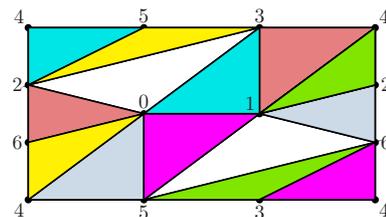


Fig. 5. Triangulation minimale du tore.

B. Des surfaces remarquables.

Dans le cas où $n \equiv 3$ ou 7 (12) le graphe complet K_n engendre une triangulation minimale d'une surface orientable S_h de genre $h = \frac{n^2-7n+12}{12}$. En passant au graphe dual (c'est à dire le graphe dont les sommets sont les centres des triangles et les arêtes sont les segments reliant les centres des triangles adjacents -voir figure 6-) on peut découper la surface de manière à ce qu'elle s'obtienne en recollant de manière idoine n polygones à $n-1$ côtés.

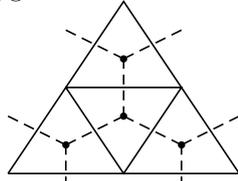


Fig. 6. Graphe dual (en pointillé).

En considérant ces polygones comme des polygones réguliers du plan hyperbolique, le recollement donne alors une surface Σ_h de même genre h munie d'une métrique hyperbolique. La figure 7 montre ces polygones dans le modèle du disque pour $n = 15$ (surface de genre 11). On peut remarquer que les côtés des triangles se prolongent en des géodésiques de la surface.

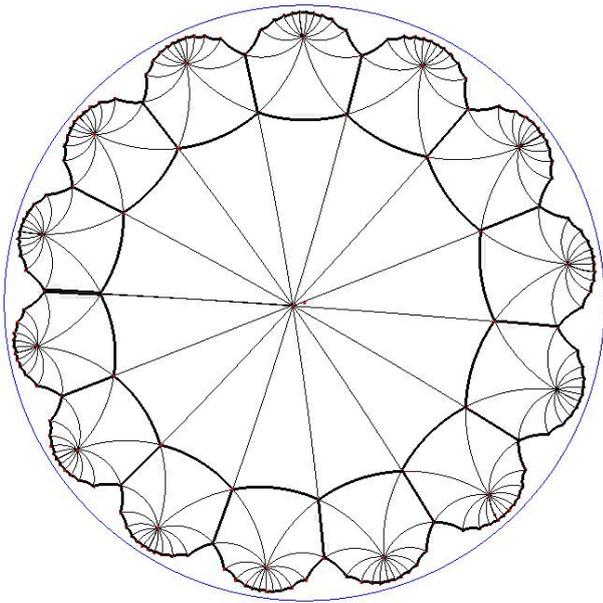


Fig. 7. Polygones hyperboliques.

C. Perspectives

On espère que le type de surfaces Σ_h décrites ci-dessus peuvent permettre de donner une meilleure borne supérieure pour le comportement asymptotique du quotient systolique en fonction du genre : il est facile de trouver l'aire de telle surface qui est donnée par la formule de Gauß-Bonnet (pour $h \geq 2$)

$$\text{aire}(\Sigma_h) = 4\pi(h - 1).$$

Il reste alors à essayer de minorer le mieux possible la systole...

RÉFÉRENCES

- [1] M. Berger, « Systoles et applications selon Gromov » Séminaire N. Bourbaki, exposé 771 *Astérisque*, vol. 216, pp. 279–310, 1993.
- [2] M. G. Katz, *Systolic geometry and topology*, American Mathematical Society, New-York, 2007.
- [3] S. Gallot, D. Hulin, J. Lafontaine, *Riemannian geometry*, Springer-Verlag, Berlin, 1987.
- [4] M. Jungerman, G. Ringel, « Minimal triangulations on orientable surfaces » *Acta Math.*, vol. 145, pp. 121–154, 1980.
- [5] P. Buser, P. Sarnak, « On the period matrix of a Riemann surface of large genus » *Invent. math.*, vol. 117, pp. 27–56, 1994.

DOCTISS 2008

Algebraic Structures

Selene SANCHEZ-FLORES ¹ (Doctorante 3e année), Claude CIBILS ² (Directeur de Thèse)

GTA - Laboratoire de Géométrie, Topologie et Algèbre.

Université Montpellier II - Case Courrier 051 - Place Eugene Bataillon - 34095 Montpellier Cedex

¹sanchez@math.univ-montp2.fr ²cibils@math.univ-montp2.fr

Résumé— I will explain some algebraic structures with the help of combinatorial examples.

Mots-clés— algebra, Lie algebra, Gerstenhaber bracket, graded Lie algebra.

I. INTRODUCTION

Explaining mathematics to a general public is quite a challenge. The problem is caused by the complexity of the language of mathematics. For example, words like "group", "ring" or "field" involve precise technical meanings. Therefore, understanding them becomes difficult and annoying : it is not like we are reading a story ! However, I believe that with the help of examples, explaining and understanding abstract concepts turn out to be not so hard. Besides, using diagrams, graphs, figures or any visual tool helps mathematics to become more pleasant.

The aim of this article is rather modest and consists to give a general idea of some concepts and techniques used in my PhD thesis. I will try to illustrate them by explaining examples that can be seen using diagrams and combinatorics. Briefly, my research work is about a particular algebraic structure defined over the cohomology groups of an algebra. Precisely, the algebraic structure I deal with is the Lie structure defined on the Hochschild cohomology which is given by the so called Gerstenhaber bracket. It is related to deformation theory and noncommutative calculus.

II. ALGEBRA STRUCTURE

Mathematicians refer to "algebraic structure" as a set together with an operation which associates to a couple of elements a third one. "Groups", "rings", "fields" and "algebras" are some examples of algebraic structures. Now, given a set and an operation on this set, in order to carry the name of "group", "ring" "field" or "algebra", both the set and the operation have to satisfy some axioms. In this paragraph, we will discuss the case of algebras which are very powerful tools in the study of geometric objects. A well known example of algebra is the set of polynomials in one variable together with the addition, subtraction and multiplication (but not division). Let us give the precise definition of algebra. An *algebra* A (not necessarily associative) is a vector space (over a field k) together with a bilinear map $\mu : A \times A \rightarrow A$. So an algebra is a set A together with an operation μ and the axioms they must satisfy are : A is a vector space and μ is a bilinear map. An

algebra A is called *associative* if $\mu(\mu(a, b), c) = \mu(a, \mu(b, c))$ for all elements a, b and c in A . An algebra A is called *unital* if there exists a linear application $\iota : k \rightarrow A$ such that $\mu(a, \iota(\lambda)) = \mu(\iota(\lambda), a) = \lambda a$ for all element a in A and λ in k . An algebra is called *commutative* if $\mu(a, b) = \mu(b, a)$ for all element a and b in A .

For example, consider $\mathbb{R}[x]$ the set of all polynomials in one variable. Recall that a polynomial is a linear combination of monomial expressions : $a_0 + a_1x + a_2x^2 + \dots + a_nx^n$ where each a_i is a real number, x is the variable and n is a natural number. The set $\mathbb{R}[x]$ together with the multiplication of polynomials is an algebra. Why ? well, first of all notice that the set of polynomial together with the addition of polynomials and the multiplication by scalars is a vector space (over the real numbers). Now, let us remark that the multiplication of polynomials is a bilinear map because it is distributive. In fact we have that the algebra $\mathbb{R}[x]$ is associative, unital and commutative. Another example of associative unital algebra is the set of 2×2 matrices with coefficients in k together the multiplication of matrices. This algebra is not commutative.

A. Path algebras

There is another way to think of a polynomial algebra. In order to explain it let us introduce the concept of quiver. A *quiver* Q consists of two sets Q_0 and Q_1 together with two applications $s, t : Q_1 \rightarrow Q_0$. The set Q_0 is called the set of *vertices* and Q_1 the set of *arrows*. We call the applications s and t , the *source* and the *target* respectively.

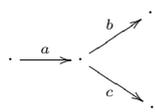
For example, consider the quiver Q whose underlying sets are $Q_0 = \{e\}$ and $Q_1 = \{x\}$. In fact there is no other possibility for the applications s and t than to associate to the arrow x the vertex e . Hence, x is a *loop* because it has the same source and the same target. We can visualize the quiver Q as :



where the loop represents the arrow x and the dot represents the vertex e .

Given a quiver Q , we can construct *paths* by concatenating arrows as in a chain. For example, let a and b two arrows such that the source of b is the same as the target a then we can construct a path that goes from the source of a to the target of b . Such path goes first through a and then through b . We denote such path by the expression ba . For

example, in the quiver



and ca are paths but we cannot construct a path that goes through the arrow b and through the arrow c . The *length* of a path is the number of arrows in its decomposition. By convention, we have that the paths of length zero are the vertices and the paths of length one are the arrows. One more thing, notice that the source and target application are also defined for paths.

We can now define the path algebra. Given a quiver Q and a field k , we denote by kQ the vector space generated by all paths (arrows and vertices included). An element of kQ is a linear combination of paths. Let μ be the bilinear map defined by the concatenation of paths, so μ associates to a couple of paths (p, p') , the path $p'p$ if $t(p) = s(p')$ and zero otherwise. We call kQ together with μ the *path algebra of Q* . If the quiver has a finite number of vertices, what does μ associates to the couple consisting of an arrow and the sum of all vertices?

Let us give an example of a path algebra. Consider the quiver of the one loop, let us find the set of all paths. We have that the only path of length 0 is the vertex e and that the only path of length 1 is the arrow x . What about the paths of other lengths? Notice that the only path of length n is the path given by the expression $\underbrace{x \cdots x \cdots x}_n$ which we will denote by x^n . It is also clear that μ associates to the couple of paths (x^n, x^m) the path x^{n+m} . In fact, it is not difficult to see that $\mathbb{R}[x]$ is the path algebra of the one loop quiver. What is the path algebra of the two loops quiver? Can we describe $\mathbb{R}[x, y]$ as a path algebra?

Actually, in the study of any finite-dimensional algebras, the path algebra of a quiver is the principal tool. The reason is due to a result of Peter Gabriel that states that any finite dimensional, basic and irreducible algebra over an algebraically closed field can be actually seen as a quotient of a path algebra by an admissible ideal.

B. Lie algebra

A *Lie algebra* L is a vector space together with a bilinear map called *bracket* $[-, -] : L \times L \rightarrow L$ such that $[x, y] = -[y, x]$ and that $[x, [y, z]] + [z, [x, y]] + [y, [z, x]] = 0$. For example, the matrices $M_{2 \times 2}(k)$ have a Lie algebra structure with an operation called the *commutator bracket* of matrices, defined as follows $[M, N] = MN - NM$. Such bracket satisfies the above identities, thanks to the associativity of the multiplication of matrices. The example above is a particular case of a more general fact : let A together with μ an associative algebra, the commutator bracket defined as $[a, b] = \mu(a, b) - \mu(b, a)$ provides the set A with a Lie algebra structure.

Now, we are going to use quivers to give an example of Lie algebra. Given a quiver Q , we say that two paths are *parallels* if both have the same source and the same target. We will denote $Q_1 \parallel Q_1$ the set of couples (a, b) where a and b are parallel arrows and $k(Q_1 \parallel Q_1)$ denotes the vector space generated by such set. We define a bracket in $k(Q_1 \parallel Q_1)$ as follows : $[(a, b), (c, d)] = \delta_{a,d}(c, b) - \delta_{b,c}(a, d)$ where $\delta_{x,y}$ means zero if $x \neq y$ and means one if $x = y$.

This bracket endows the set of parallel arrows with a Lie algebra structure. In fact, if you consider the quiver of only two parallel arrows : $\cdot \rightrightarrows \cdot$, we have that $k(Q_1 \parallel Q_1)$ coincides with the Lie algebra $M_{2 \times 2}(k)$.

Another example of Lie algebra is the space of derivations of an algebra. Let A together with μ be an associative algebra, we define a *derivation* of A as a linear map $D : A \rightarrow A$ such that $D(\mu(a, b)) = \mu(a, D(b)) + \mu(D(a), b)$. Notice that if D and D' are derivations, then the composition DD' is not necessarily a derivation. In order to have a well defined operation we consider the commutator bracket as in the case of matrices. Hence, the vector space of all derivations together with the commutator bracket, i.e. $[D, D'] = DD' - D'D$, is a Lie algebra.

Now, there is a result [4] that tell us that if $A = kQ$ where Q is the quiver of two parallel arrows (which are not loops), we have that the Lie algebra of derivations of A can be seen as the Lie algebra $k(Q_1 \parallel Q_1)$ which is the same as the Lie algebra $M_{2 \times 2}(k)$.

III. GRADED ALGEBRA AND LIE GRADED ALGEBRA

In this paragraph we study algebra structures which are endowed with an additional data : graduation. We will define graded algebras and Lie graded algebras. Then we will present the Gerstenhaber bracket which provides an example of Lie graded algebra.

IV. DEFINITIONS

An example of graded algebra is the polynomial algebra. The polynomials can be seen as the direct sum of the vector spaces generated by each monomial expression :

$$\mathbb{R}[x] = \bigoplus_{n=0}^{\infty} \mathbb{R}x^n \quad .$$

The graduation is evident : elements of the vector space $\mathbb{R}x^n$ have degree n . Moreover, the degree of an element obtained by the multiplication of an element of degree n and an element of degree m is the sum of the degrees : $n+m$. Let us give the exact definition. An algebra A is called *graded* if A has a direct sum decomposition (this means that $A = A_0 \oplus A_1 \oplus \cdots \oplus A_n \oplus \cdots$) and the multiplication μ maps $A_i \times A_j$ to A_{i+j} .

A *graded Lie algebra* is a vector space L with a direct sum decomposition, i.e. $L = L_0 \oplus L_1 \oplus \cdots \oplus L_n \oplus \cdots$, together with a *graded bracket*, i.e. $[-, -]$ that maps $L_n \times L_m$ to L_{n+m} , such that we have the following axioms satisfied : if x^n is in L_n , y^m in L_m and z^p in L_p , then we have that $[x^n, y^m] = -(-1)^{nm}[y^m, x^n]$ and that

$$(-1)^{np}[x^n, [y^m, z^p]] + (-1)^{pm}[z^p, [x^n, y^m]] + (-1)^{mn}[y^m, [z^p, x^n]] = 0.$$

We say that the elements of L_n have *degree n* . Notice that L_0 is in fact a Lie algebra.

An example of Lie graded algebra is given by the Gerstenhaber bracket that we will present below. Some formulas have to be introduced first.

A. The formula for \circ_i and \circ

In order to give the definition of the Gerstenhaber bracket, we need to give two formulas : one for \circ_i and for \circ .

Let V be a vector space. For $n = 1, 2, \dots$, consider $C^n(V, V)$ the set of all n -multilinear maps on V , i.e.

$$C^n(V, V) := \{f : \underbrace{V \times \dots \times V}_{n \text{ times}} \rightarrow V \mid f \text{ is multilinear}\}$$

Now let us fix n and let $i = 1, \dots, n$. Let f be in $C^n(V, V)$ and g be in $C^m(V, V)$ we denote by $f \circ_i g$ the $(n+m-1)$ -multilinear function that associates to the tuple $(x_1, x_2, \dots, x_{n+m-1})$ the element

$$f(x_1, x_2, \dots, g(x_i, \dots, x_{i+m-1}), \dots, x_{n+m-1}).$$

Perhaps this formula seems very abstract but do not discourage because in the next section we will describe it using quivers and parallel paths. The next formula is given in terms of the maps \circ_i . We denote by $f \circ g$ the $(n+m-1)$ -multilinear function given by

$$f \circ g := \sum_{i=1}^n (-1)^{(i-1)(m-1)} f \circ_i g \quad (1)$$

Notice that if $n = 1$ and $m = 1$ then $f \circ g$ is just the composition of linear functions.

B. Gerstenhaber bracket

We will denote by $C^*(V, V)$ the following direct sum :

$$C^*(V, V) = \bigoplus_{n=1}^{\infty} C^n(V, V) \quad .$$

The Gerstenhaber bracket is the bilinear map

$$[-, -]_G : C^*(V, V) \times C^*(V, V) \rightarrow C^*(V, V)$$

defined by

$$[f, g]_G := f \circ g - (-1)^{(n-1)(m-1)} g \circ f. \quad (2)$$

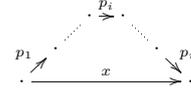
Notice that if $n = 1$ and $m = 1$ then we find that such formula is the commutator bracket. Now, in order to have a graded bracket, we will declare that for $n = 1, 2, \dots$, elements of $C^{n+1}(V, V)$ are of degree n : denote an element of $C^{n+1}(V, V)$ by $f^{(n)}$. In this way, we have that the bracket maps the couple $(f^{(n)}, g^{(m)})$ to an element in $C^{n+m+1}(V, V)$, this means that such element is of degree $n+m$ which is the sum of the degree n and the degree m . The proof that $C^*(V, V)$ together with the bracket $[-, -]_G$ is a Lie graded algebra is given by Gerstenhaber in [2]. It is not difficult but rather technical.

As a matter of fact, Gerstenhaber proved in [2] that the bracket can be defined on the Hochschild cohomology of an associative algebra. So the Lie graded algebra structure can be carried to the Hochschild cohomology. Right now, probably you are asking yourselves, what does Hochschild cohomology means? Well, given an associative algebra, it can be defined a sequence of vector spaces called the Hochschild cohomology groups. Such vector spaces play an important role in the study of the algebra. For example, we have that the zero cohomology group is the center of the algebra. The first cohomology group is a quotient of the derivations of A , so in fact it is a Lie algebra. The second Hochschild cohomology group is related to the deformations of the multiplication of an associative algebra.

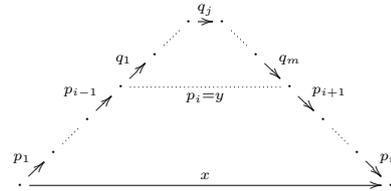
V. COMBINATORIAL GERSTENHABER BRACKET

In this section, the objective is to give another bracket : which is defined as the Gerstenhaber bracket but using quivers. The idea is to clarify the formula of \circ_i which is involved in the definition of the Gerstenhaber bracket. Thus, we will have a more visual example of a Lie graded algebra.

Let Q be a quiver, we denote by Q_n the set of paths of length n . Now the set $Q_n \parallel Q_1$ is the set of all couples consisting of a path in Q_n and an arrow that are parallels. An element (p, x) in $Q_n \parallel Q_1$ is called a *shortcut* :



We denote $k(Q_n \parallel Q_1)$ the vector space generated by the set $Q_n \parallel Q_1$. We will define a bilinear map \diamond_i from the product $k(Q_n \parallel Q_1) \times k(Q_m \parallel Q_1)$ to $k(Q_{n+m-1} \parallel Q_1)$ as follows. Given two shortcuts (p, x) and (q, y) we will define $(p, x) \diamond_i (q, y)$ to be the shortcut obtained by replacing the arrow p_i by the path q if $p_i = y$ where p_i is in the decomposition of p , like the figure shows :



Now, if the arrow p_i is not y then $(p, x) \diamond_i (q, y) = 0$. In the context of shortcuts, it is easy to see that the operation \diamond_i replaces arrows by paths. Once \diamond_i is defined we will define the map \diamond as we defined \circ , (see eq. (1)), but using this time the maps \diamond_i instead of \circ_i . Finally using \diamond , let us define the combinatorial bracket $[-, -]_Q$ as we did for the Gerstenhaber bracket, (see eq. (2)). It is proven in [3] that the vector space of all shortcut $\bigoplus_{n=1}^{\infty} k(Q_n \parallel Q_1)$ together with the combinatorial bracket $[-, -]_Q$ is a Lie graded algebra by declaring that shortcuts in $k(Q_{n+1} \parallel Q_1)$ are of degree n . We will call this Lie graded algebra the *shortcut Lie algebra*.

Let me point out two interesting facts about shortcuts. First of all, Cibils has demonstrated in [1] that for a particular family of algebras (monomial algebras of radical square zero), the Hochschild cohomology can be described using the language of shortcuts. Second, we proved in [3] that the Lie graded algebra structure on the Hochschild cohomology groups of such algebras can be describe using the shortcut Lie algebra.

In conclusion, quivers help visualize abstract notions.

RÉFÉRENCES

- [1] Cibils, Claude *Hochschild cohomology algebra of radical square zero algebras*, Algebras and modules, CMS Conf. Proc. vol. 24, 93–101, 1998.
- [2] Gerstenhaber, Murray *The cohomology structure of an associative ring*, Ann. of Mathematics (2) 78 , 267-288, 1963.
- [3] Sanchez-Flores, Selene , *The Lie module structure on the Hochschild cohomology groups of a monomial algebra of radical square zero*, arXiv/math :0711.2810v1[math.RT], 2007.
- [4] Strametz, Claudia *The Lie algebra structure on the first Hochschild cohomology group of a monomial algebra* J. Algebra Appl.5, 245-270, 2006.

DOCTISS 2008

Systoles et variétés plates

Chady ELMIR¹ (Doctorant 3e année), Jacques LAFONTAINE² (Directeur de Thèse)

Institut de Mathématiques et Modélisation de Montpellier
Université Montpellier 2
Place Eugène Bataillon
34095 Montpellier cedex France

¹elmir@math.univ-montp2.fr, ²jaclaf@math.univ-montp2.fr

Résumé— Ce texte est un essai de vulgarisation de la notion de métrique riemannienne, longueur de courbes et systole sur les surfaces compactes. En même temps je résume les résultats les plus classiques sur l'inégalité systolique en dimension 2. Enfin, j'explique rapidement la notion de variété plate et je présente de façon simple le sujet de ma thèse et un résultat qu'on a obtenu.

Mots-clés— Surface, géodésique, systole, rapport systolique, variété riemannienne plate.

I. INTRODUCTION

On s'est tous probablement demandé : en prenant l'avion de Paris à Los Angeles (par exemple) le chemin que prend l'avion est-il le plus court parmi tous les chemins possibles sur la surface de la Terre entre ces deux villes ? Est-ce que les chemins les plus courts sur la sphère ressemblent à ceux du plan (qui sont évidemment les segments de droites) ? Ou peut être tout simplement ces chemins n'existent pas.

En fait pour les surfaces suffisamment "régulières et compactes", de tels chemins existent entre deux points quelconques de la surface, on les appelle "les géodésiques". La sphère est un bel exemple de telle surface, il y a évidemment plein d'autres jolis exemples ; on cite l'ellipsoïde, la bouée, la 2-bouée (bouée à 2 trous), etc.. Les géodésiques sont les solutions d'un système d'équations différentielles non linéaires un peu compliqué et leur existence est due à un théorème d'existence de solutions pour ce type d'équations différentielles.

Si on veut être très précis et exact il faut tenir compte du fait que la Terre n'est pas toujours ronde, en fait elle ne l'est jamais. Elle passe par des formes plutôt ellipsoïdales à cause d'un nombre de phénomènes qui arrivent au noyau de la Terre. Par conséquent ces plus courts chemins changent mais bien évidemment la surface de la Terre ne change pas. Là plein de questions se posent ; quel est le lien entre les anciens chemins minimisants et les nouveaux ? Est-ce qu'il y a une formule qui relie les longueurs des géodésiques fermés avec la surface de la Terre et qui reste valable même si la Terre se déforme ? Si une telle relation existe, est-ce qu'il y en a d'autres pour les autres surfaces déjà citées, ou peut-être pour des surfaces plus compliquées ?

Dans le présent papier on explicite les plus courts chemins sur les surfaces les plus "élégantes" de la géométrie : la sphère, le tore, le ruban de Möbius et la bouteille de Klein. Pour les deux premières il est clair qu'on les voit comme des surfaces de l'espace à trois dimensions. On donnera pour les deux dernières des dessins qui permettent plus ou moins de voir concrètement leur structure. En plus on évoquera une inégalité qui relie l'aire de ces surfaces avec la plus petite longueur des géodésiques fermées et qui reste vraie si on change la métrique de notre surface, en géométrie riemannienne une telle inégalité s'appelle "inégalité systolique".

II. INÉGALITÉ SYSTOLIQUE SUR LES SURFACES

A. La sphère ronde et déformée

Dans le monde des mathématiques la sphère ronde, notée S^2 , est par définition (en rapportant à l'espace un système de coordonnées orthonormées) l'ensemble des points (x, y, z) de l'espace \mathbb{R}^3 qui vérifient l'équation $x^2 + y^2 + z^2 = R^2$ où R est en fait le rayon de la sphère. Comme on a déjà dit dans l'introduction la sphère est une surface compacte, c'est-à-dire entre deux points quelconques il existe toujours une géodésique dont la longueur est minimale parmi tous les chemins possibles entre ces deux points..

Les géodésiques de la sphère S^2 sont les grands cercles : l'intersection avec S^2 des plans passant par son centre. Ce sont les courbes sur S^2 dont la projection sur le plan tangent d'un point de la courbe est un segment. En déformant la sphère ronde en changeant la métrique canonique on obtient des surfaces homéomorphes à la sphère qui ne sont pas des sphères rondes : il suffit d'imaginer la sphère ronde étirée ou bosselée. On notera la sphère munie d'une métrique g par (S^2, g) . On définit la systole (notée Sys) de (S^2, g) par la plus petite longueur des géodésiques fermées. Cette définition ne peut pas être adoptée pour les autres surfaces, elle caractérise la sphère parce qu'elle est simplement connexe c'est-à-dire toute courbe fermée peut être déformée en un point. Il faut noter que contrairement au n -tore l'existence de géodésiques fermées sur une sphère est un résultat difficile à démontrer, ça se fait en utilisant des "techniques qualitatives".

L'inégalité qu'on a promis dans l'introduction est la suivante : il existe une constante strictement positive ϵ tel que

pour toute métrique g sur la sphère on ait

$$\text{Aire}(S^2, g) / (\text{Sys}(S^2, g))^2 \geq \epsilon$$

Essayons d'expliquer cette inégalité sur l'exemple qu'on a évoqué au début. La surface de la terre étant constante, cette inégalité montre que quelque soit la déformation de la Terre on peut toujours trouver des géodésiques fermées de longueurs pas trop grandes. Ceci est en fait équivalent à dire qu'à systole fixée le volume ne peut pas être trop petit. Ce résultat récemment trouvé par S.Sabourau et F.Balacheff est valable pour toute métrique possible sur la sphère donc les sphères à n -doigts ne sont pas exclues.

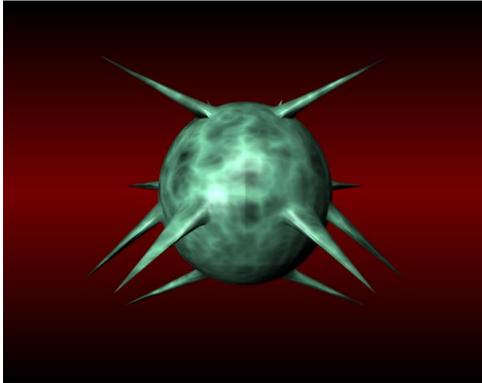


Fig. 1. Sphère à n -doigts.

Dans la suite la systole d'une surface Σ sera la plus petite longueur des courbes fermées non contractiles, c'est-à-dire non déformable en un point.

B. Le Tore

La bouée ou le tore T^2 en langage mathématique, est une surface qui ressemble comme le dit son nom à une chambre à air. Cette surface supporte une inégalité systolique, c'est un résultat dû à Lowner en 1949 : il existe une constante strictement positive ϵ tel que pour toute métrique g sur le tore on ait

$$\text{Aire}(T^2, g) / (\text{Sys}(T^2, g))^2 \geq \epsilon$$

mais dans ce cas on connaît la meilleure constante ϵ , elle est égale à $\sqrt{3}/2$. En plus on sait pour quelle métrique g a lieu l'égalité. Pour l'obtenir il faut voir le tore d'une façon un peu abstraite.

Etant donné un tore plongé dans l'espace \mathbb{R}^3 on commence par le découper suivant un méridien et on le déforme pour obtenir un cylindre. On redécoupe ce dernier suivant un parallèle et on l'aplatit pour obtenir un rectangle.

Autrement dit en partant d'un rectangle du plan si on recolle les côtés opposés de la façon canonique on obtient un tore. Pour un rectangle muni de ce recollement les géodésiques fermées sont les segments parallèles à un côté du rectangle et de même longueur. On appellera une telle métrique du tore une métrique plate, pour cette dernière la systole est donc égale au minimum des longueurs des côtés du rectangle. Enfin, on obtient une égalité dans l'inégalité ci dessus pour un tore plat équilatéral c'est à dire pour le tore obtenu en recollant les côtés d'un losange équilatéral qui est à son tour obtenu en recollant deux triangles équilatéraux suivant un côté.



Fig. 2. Le Tore de dimension 2.

C. Le ruban de Möbius

On passe maintenant à l'étude d'un type différent de surfaces : les non orientables. Une surface compacte Σ est non orientable si pour un point quelconque p de Σ en faisant tourner le vecteur normal en p à Σ d'un tour complet alors le vecteur change de sens en son opposé. Le ruban de Möbius est l'exemple le plus classique de surface non orientable. On l'obtient en partant d'un rectangle du plan et en recollant deux côtés opposés mais en tournant. Un rectangle muni de cette identification est un ruban de Möbius plat, ses géodésiques sont les segments de droites. On peut vérifier qu'une telle surface n'a qu'un seul bord (un cercle) et qu'elle n'est pas orientable.



Fig. 3. Le ruban de Möbius.

En 1952 Pu a démontré qu'il vérifie une inégalité systolique, la métrique qui donne le meilleur rapport systolique étant la déformation sphérique :

On part de la sphère ronde de rayon 1, soit Σ la surface obtenue en otant de la sphère les disques ouverts dont les centres sont deux points diamétralement opposés et de rayon r inférieur à π . La relation d'équivalence qui identifie les points diamétralement opposés de Σ donne le ruban de Möbius sphérique.

D. La bouteille de Klein

La complexité de cette surface se traduit dans le fait qu'on ne peut pas la voir comme une surface de \mathbb{R}^3 , néanmoins on peut sous certaines conventions faire une figure dans l'espace qui montre un peu comment elle est faite.

On part d'un rectangle du plan, on recolle deux côtés opposés de la façon canonique pour obtenir un cylindre. Ce dernier ayant deux cercles de bord, on les recolle tout en passant à l'intérieur du cylindre comme le montre la figure ci dessous.

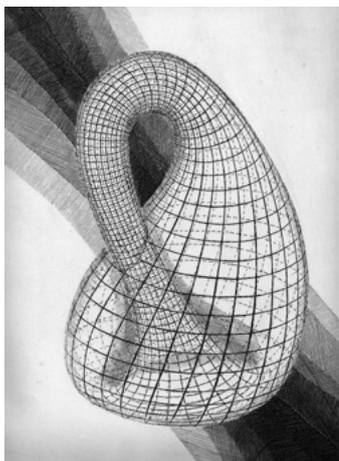


Fig. 4. La bouteille de Klein.

La bouteille de Klein est une surface non orientable qui vérifie une inégalité systolique comme l'a montré C. Bavard en 1986. Il utilise le fait que la bouteille de Klein (notée K) est obtenue en recollant deux rubans de Möbius suivant leur bord et le résultat de Pu cité ci-dessus.

Pour toute métrique g sur la bouteille de Klein on a

$$\text{Aire}(K^2, g) / (\text{Sys}(K^2, g))^2 \geq \frac{2\sqrt{2}}{\pi}$$

l'égalité caractérise une métrique singulière g_0 de la bouteille de Klein, elle est obtenue à partir de la métrique sphérique du ruban de Moebius avec $r = \pi/2$. En fait il suffit de recoller deux tels rubans de Moebius pour obtenir la bouteille de Klein (K, g_0) .

Il faut noter que les surfaces qu'on a décrites ne sont pas les seules qui vérifient une inégalité systolique, en fait J. Hebda a démontré que toute surface riemannienne qui n'est pas une sphère vérifie une telle inégalité. Le point faible des autres surfaces est qu'on ne connaît pas les meilleurs constantes systoliques et donc on sait rien sur la métrique optimale.

III. SYSTOLE EN DIMENSION TROIS

Pour une variété de dimension quelconque l'inégalité systolique n'est pas toujours vraie, le contre exemple le plus connu est $S^2 \times S^1$. Pour cet exemple la systole est toujours égale à $\text{long}(S^1)$ puisque S^2 est simplement connexe et le volume est égal à $\text{Vol}(S^2)\text{long}(S^1)$, d'où en fixant $\text{long}(S^1)$ et en faisant tendre $\text{Vol}(S^2)$ vers zéro le rapport systolique de $S^2 \times S^1$ tend lui aussi vers zéro. En 1982 M. Gromov a donné une condition topologique pour qu'une telle variété vérifie l'inégalité, cette condition élimine les espaces qui sont le produit de deux espaces dont l'un est simplement connexe.

En dimension 3, pour les variétés qui vérifient la condition de Gromov, on ne connaît rien en dehors de l'existence de la constante, même pour les variétés les plus simples comme le tore T^3 et l'espace projectif $\mathbb{R}P^3$. Dans ma thèse je m'intéresse aux variétés de dimension 3 qui portent une métrique plate, ces variétés vérifient la condition de Gromov donc son théorème s'applique. Topologiquement, ces

variétés sont des quotients d'un tore T^3 par un groupe fini d'isométries opérant sans points fixes.

En se servant de la bouteille de Klein singulière construite par Bavard et de la classification des variétés plates compactes et non orientables, on a démontré que pour ces variétés les métriques plates ne sont jamais optimales, c'est-à-dire il existe sur ces variétés des métriques non plates dont le rapport systolique est inférieur à celui des plates. On espère trouver le même résultat pour les variétés plates compactes et orientables autre que le tore, le cas du tore lui même semble beaucoup plus difficile.

RÉFÉRENCES

- [1] C. Bavard. *Inégalité isosystolique pour la bouteille de Klein*. Math. Ann. 274, 439-441, 1986.
- [2] M. Berger, *Systoles et applications selon Gromov*. Séminaire N. Bourbaki, exposé 771 Astérisque, vol. 216, pp 279-310, 1993.
- [3] L.S. Charlap. *Bieberbach Groups and flat manifolds*. Universitext, Berlin, 1986.
- [4] M.G. Katz, *Systolic Geometry and Topology*. Math. Surveys and Monographs 137, Amer. Math. Soc., Providence, R.I. 2007.
- [5] P. Pu, *Some inequalities in certain non-orientable riemannian manifolds*. Pacific J. Math. 2, 55-71, 1952.
- [6] J.A. Wolf, *Spaces of constant curvature*,

DOCTISS 2008

The Mysterious and Magical Immune System !

Kevin BLEAKLEY¹ (Doctorant 3e année), Gérard BIAU² (Directeur de Thèse)

Institut de Mathématiques et de Modélisation de Montpellier,
Place Eugène Bataillon,
34095 Montpellier.

¹bleakley@math.univ-montp2.fr, ²gerard.biau@upmc.fr

Résumé—Immunobiology is the study of the immune system - our protection against illness and death. Immunoglobulins and T cell receptors, which make up part of the immune system, are the objects that patrol our body like a protecting army. We have billions of different types of them ! Many poorly-understood processes are needed to make this huge number, some of which are random. Since more and more data is being collected on these random processes, we can start to use statistics to look closely at the results of these random processes, and try to infer the actual physical process taking place !

Mots-clés—Immune system, immunobiology, statistics, trimming, eating, IMGT.

I. INTRODUCTION

The immune system is an amazing thing. It has one main goal : keep us alive. It may have only one main goal but that doesn't stop it being indescribably complex.

We start the story with *lymphocytes*, which are a type of *white blood cell*. They make up an important part of our immune system. A picture of one is shown in Fig. 1.

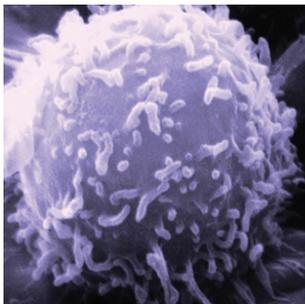


Fig. 1. A lymphocyte seen from a scanning electron microscope.

There are three major types of lymphocyte :

- **natural killer cells.** These are part of the *innate immune system*.
- **B cells.** These are part of the *adaptive immune system*. B cells are produced in the bone marrow of most mammals.
- **T cells.** These are another part of the adaptive immune system. The *T* in *T* cell receptor stands for *thymus*, which is where they are made.

The human body makes millions of B cells every day. Like soldiers, they patrol the *blood* and the *lymph*. Each one of them is built slightly differently, and can therefore recognise only one (or none) of the invading enemy. Each is different because each has sitting on its surface a unique *B cell receptor*, commonly

known as an *immunoglobulin*. This is the part of the B cell that recognises the one invader (antigen).

What happens next is shown graphically in Fig. 2. The B cell

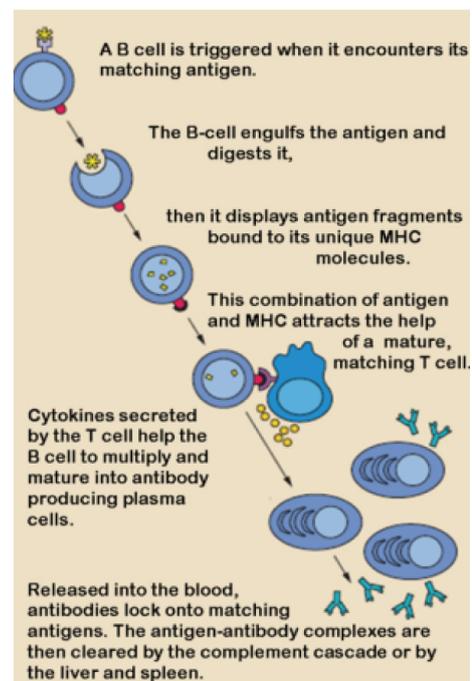


Fig. 2. How B cells and T cells work together.

recognises the antigen, catches it, and breaks it up into small pieces. It then shows off pieces of the antigen in the hope of attracting the attention of one (if any !) helper T cell.

Just like the B cell, each helper T cell is built slightly differently. It too has something on its surface, which, as you may have guessed, is called the *T cell receptor*, and can recognise *one* specific piece of antigen attached to the B cell.

When the T cell signals to the B cell that it has arrived and starts to help, the B cell eventually transforms into either a *memory B cell* or a *plasma B cell*. Plasma B cells transform into antibody producing factories, essentially making many copies of more or less the same antibody. These antibodies are then sent out into the body to try and stick to the microbes, making them visible to other parts of the immune system that will collect them and take them away to be destroyed.

Now we take a step back to ponder the question : “So how

can there be enough different B cells and T cells to recognise the vast possible number of infecting agents?"

To answer this question, we must look at the formation of the B cell receptor or T cell receptor, the parts that make *contact* with the infecting element.

The important genes for B cell receptors and T cell receptors are located in groups on our genome (i.e., on our DNA). As can be seen in Fig. 3, these groups are V, (D sometimes, but not in the type of receptor shown in Fig. 3), and J genes. V means *variable*, D means *diversity* and J means *joining*. The one gene called C is the *constant* gene. **The D gene is only present in certain receptors.** This is why this **we will often put the D gene in parentheses.**

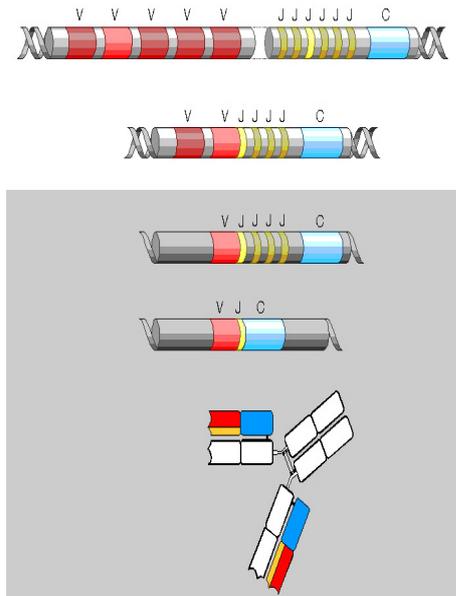


Fig. 3. Simple overview of V(D)J recombination for T and B cells.

Through a random (as far as we know) recombination process (as shown in Fig. 3), **one V, (one D) and one J** gene are chosen and **put together**. We can already see that this will create different resulting DNA sequences, depending on which V, (D) and J are chosen. If this was the only important process, we would end up with thousands of possible DNA sequences, thus thousands of B cell receptors and T cell receptors. This is nowhere near enough.

So where do we get more? Nature comes to the rescue with some beautiful processes, as shown in Fig. 4. Before the 3 (or 2)

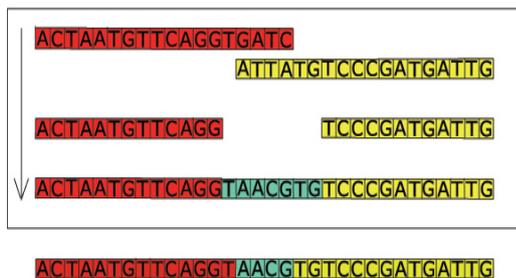


Fig. 4. Beautiful processes.

genes are joined together, some hungry guys called *exonuclease* come along and eat a random (as far as we know) number of the nucleotides on the ends of each of the V, (D) and J genes.

Already, this whole diversifying process is amazing. But it gets better. After the exonuclease has finished, another guy, called *TdT* (for short) comes along and starts adding a random number (as far as we know) of nucleotides to the trimmed ends of the V, (D) and J genes! What's more, for the B cell receptors (not for the T cell receptors), there is even another step where mutations are introduced into the sequence!

The resulting sequence of nucleotides has therefore a significant random element, and means that a huge number of different sequences (and therefore receptors) can be created. If we want to talk in numbers, this set of processes is thought to be able to produce 10^{12} different receptors! This crazy number is part of what helps keep us alive.

II. A CLOSER LOOK AT NUCLEOTIDE EATERS

The nucleotide eating process is very little understood. Usually, it is considered *random*. But *random* can mean a lot of things. When we roll a dice, the number that comes up is *random*, but we know very well the probability of rolling a 1, 2, 3, 4, 5 or 6. Each has a probability $1/6$.

To give another example, the number of cars arriving at a petrol station every hour is a *random variable*. It could be 0, 1, 2, or even 20. It turns out that the probability of 0, 1, 2, ..., 20, etc. cars follows what is called the *Poisson distribution*. Mr Poisson and one of his distributions are shown in Fig. 5.

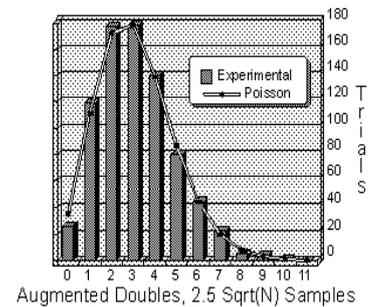


Fig. 5. Siméon-Denis Poisson and his distribution.

In order to understand better the eating process, we assembled a set of several hundred DNA sequences representing different V(D)J junctions. Then we used automatic tools from **IMGT** (the world's foremost immunoinformatics website/database/tools resource, based in Montpellier, directed by Marie-Paule Lefranc) to find the V gene, J gene and to predict the *number of eaten nucleotides* and number of randomly added nucleotides. Here is an example of a DNA sequence:

tgtgccttgtgggagcaactggaagagttgggcaaaaaatcaaggtattt

When we give this to IMGT, it returns the prediction in Fig. 6.

3'V-REGION N-REGION 5'J-REGION
 tgtgccttgtgggag... caactggaagagttgggcaaaaaatcaaggtattt

Fig. 6. IMGT output

The little dots are the predictions of the number of eaten nucleotides from the V and J genes (3 for the V gene and 5 for the J gene). If we count the same things for hundreds of sequences, we can then count the number of times 0, 1, 2, 3, ... nucleotides

were eaten. If we then show that count on a histogram, we get for example Fig. 7.

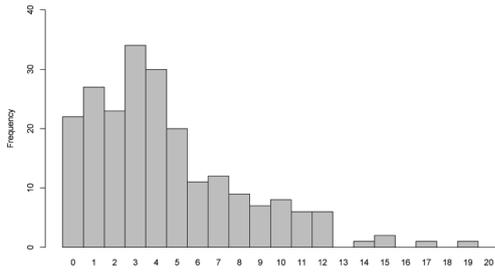


Fig. 7. Histogram count of number of eaten nucleotides

Before testing whether this distribution could be a Poisson distribution, we realised that we should try and remove biases from the automatic tool predictions. To understand why, we look at Fig. 8 (the same as Fig. 4).

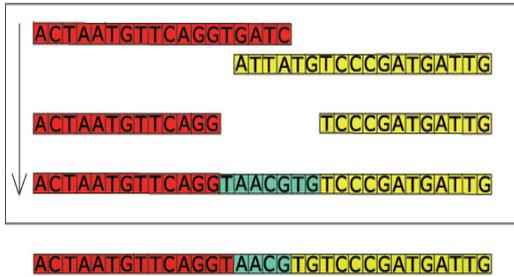


Fig. 8. What really happened ?

The bottom line inside the frame shows the finished sequence (i.e., **what really happened**), with the colours representing the remaining V gene (red), J gene (yellow) and added nucleotides (blue). The last line (outside the frame) shows an automatic tool prediction. You can see **it is not the same as what really happened!** This is because, by bad luck, some eaten nucleotides were replaced by **exactly the same** added nucleotides ! Indeed, on the V end, an eaten **T** nucleotide was replaced by a **T**, and on the J end, the eaten **TG** were replaced by new **TG**.

To remove this bias, we had to use some mathematics. We called B the probability law of **what really happened** and F the probability law of **what we see via the automatic prediction**. Our goal was to transform the data we had which follows the law F (i.e., Fig. 7) to data that follows the law B , the **real law**.

With some simple hypotheses and a bit of work, we were able to show the following result :

$$\mathbb{P}\{B = k\} = \frac{4}{3} \mathbb{P}\{F = k\} - \frac{1}{3} \mathbb{P}\{F = k + 1\}.$$

This means that the true probability of there being k eaten nucleotides is equal to $4/3$ the probability the automatic prediction sees k eaten nucleotides, **minus** $1/3$ the probability the automatic prediction sees $k + 1$ eaten nucleotides.

We applied this transformation to the data in Fig. 7, and this gave us the result shown in Fig. 9.

Looking back at the Poisson distribution next to Mr Poisson (Fig. 5), we can see that it looks quite a lot like the distribution in Fig. 9. Using what are called hypothesis tests, we were able to show that the hypothesis that the distribution B was Poisson could not be rejected.

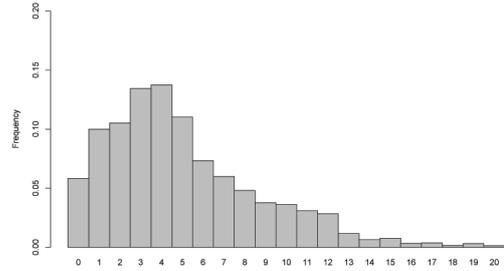


Fig. 9. Unbiased probability distribution B

III. CONCLUSION

These results are consistent with the following scenario : there are many exonuclease (eaters) being thrown around by heat and chemical charges, at random. Occasionally one of them accidentally gets close to the end of one of the V or J genes, and takes a bite ! In a fixed time window, the number of them randomly arriving would follow a Poisson law.

If only it were so simple ! After talking to immunobiologists, this hypothesis is unlikely. They believe that one exonuclease comes along (randomly ?), and eats a random number of nucleotides, on its own. Therefore, we must come up with a better reason for the appearance of Poisson-like distributions.

Time to get back to work !

IV. ACKNOWLEDGMENT

We would like to thank Marie-Paule Lefranc, Gérard Biau, Véronique Giudicelli and Yan Wu, who collaborated on this project.

RÉFÉRENCES

- [1] Bleakley, K., Giudicelli, V., Wu, Y., Lefranc, M.-P. and Biau, G. « IMGT standardization for statistical analyses of T cell receptor junctions : The TRAV-TRAJ example, » *In Silico Biology*, vol. 23, pp. 573–588, 2006.
- [2] Bleakley, K., Biau, G. and Lefranc, M.-P. « Recovering true probabilities for nucleotide trimming processes in T cell receptor TRA and TRG junction formation with IMGT tools, » *Submitted*.

UN MODELE STATISTIQUE POUR IDENTIFIER ET CARACTERISER LES COMPOSANTES DE LA CROISSANCE D'UN ARBRE

FLORENCE CHAUBERT-PEREIRA (DOCTORANTE 3E ANNEE)

CIRAD, UMR Développement et Amélioration des Plantes
INRIA, équipe Virtual Plants
TA A-96 / 02, Avenue Agropolis, 34398 Montpellier Cedex 5, France

CHAUBERT@CIRAD.FR
[HTTP://WWW.EDI2S.UNIV-MONTP2.FR/](http://www.edi2s.univ-montp2.fr/)

Résumé— La croissance d'un arbre, caractérisée par exemple par la longueur de pousses annuelles successives, est principalement le résultat de trois composantes : ontogénique, environnementale et individuelle. La composante ontogénique est supposée structurée comme une succession de phases de croissance bien séparées, asynchrones entre arbres tandis que la composante environnementale (principalement d'origine climatique) est supposée prendre la forme de fluctuations locales, synchrones entre les arbres. Afin d'identifier et de caractériser ces trois composantes, nous proposons comme modèle, une combinaison semi-markovienne de modèles linéaires mixtes; c'est à dire un mélange de modèles linéaires mixtes avec des dépendances semi-markoviennes. La semi chaîne de Markov sous-jacente représente la succession de phases de croissance tandis que les modèles linéaires mixtes rattachés à chaque état de la semi chaîne de Markov sous-jacente représentent à la fois l'influence des variables climatiques comme un effet fixe et l'hétérogénéité inter-individuelle comme un effet aléatoire. Cette modélisation statistique est illustrée par l'analyse de pousses annuelles de pins Laricio en fonction du climat. Une combinaison semi-markovienne de modèles linéaires mixtes, de type « gauche-droite » à 3 états composée de deux états transitoires successifs et d'un état final absorbant, a été estimée; chaque phase de croissance correspondant à une augmentation de la croissance. Le modèle estimé montre que l'influence du climat est faible sur la première phase de croissance, puis augmente fortement tandis que la part d'hétérogénéité inter-individuelle diminue plus légèrement.

Mots clés— Semi-chaînes de Markov cachées; modèles linéaires mixtes; combinaison semi-markovienne; algorithme MCEM; composantes de la croissance d'un arbre.

I. INTRODUCTION.

La croissance d'un arbre est principalement le résultat de trois composantes : une composante ontogénique, une composante environnementale et une composante individuelle. A noter que l'ontogénie décrit le développement progressif d'un organisme depuis sa conception jusqu'à sa mort.

Selon Guédon *et al.* [3], la composante ontogénique est supposée être structurée comme une succession de phases de croissances, asynchrones entre arbres tandis que la composante environnementale (principalement d'origine climatique) est supposée prendre la forme de fluctuations locales, synchrones entre les arbres. Guédon *et al.* [3] ont proposé un ensemble de méthodes pour l'analyse de ces deux composantes. Dans leur approche, la composante environnementale est simplement caractérisée globalement, mais il est intéressant d'étudier en détail la réponse d'un arbre aux conditions environnementales. Nous proposons ici un modèle statistique qui permet de différencier et de caractériser la structure ontogénique sous-jacente, le rôle des variables environnementales et la part d'hétérogénéité inter-individuelle sur chaque phase de croissance.

II. MATERIELS.

Le jeu de données comprend quatre sous-échantillons de pins Laricio plantés dans un peuplement forestier dans la région Centre de la France : 31 arbres âgés de 6 ans (première année non mesurée), 29 arbres âgés de 12 ans (première année non mesurée), 30 arbres âgés de 18 ans (première année non mesurée) et 13 arbres âgés de 23 ans (deux premières années non mesurées). Les arbres dans les deux premiers sous-échantillons (6 et 12 ans) ont été sélectionnés en fonction de la répartition de la hauteur des arbres à l'intérieur du peuplement alors que les arbres dans les deux autres sous-échantillons (18 et 23 ans) ont été sélectionnés en fonction de la répartition du diamètre des arbres à hauteur de poitrine à l'intérieur du peuplement. Les arbres du premier sous-échantillon (6 ans) sont restés en pépinière pendant deux ans avant d'être transplantés alors que les arbres des trois autres sous-échantillons sont restés en pépinière pendant trois ans avant d'être transplantés. La densité de plantation était de 1800 tiges/ha pour le premier sous-échantillon (6 ans) et de 2200 tiges/ha pour les trois autres sous-échantillons.

Les données observées ont été collectées rétrospectivement, c'est-à-dire que le développement de chaque arbre a été

reconstitué à une date d'observation donnée à partir de marqueurs morphologiques (comme par exemple des cicatrices de ramification) correspondant aux événements passés. Les arbres sont décrits de la base au sommet par les pousses annuelles où la longueur en cm est enregistrée pour chaque pousse annuelle. La pousse annuelle est définie comme le segment de tige établi durant une année. Les longueurs de pousses des arbres âgés de 18 ans sont représentées sur la Figure 1 ; chaque courbe correspondant à un arbre.

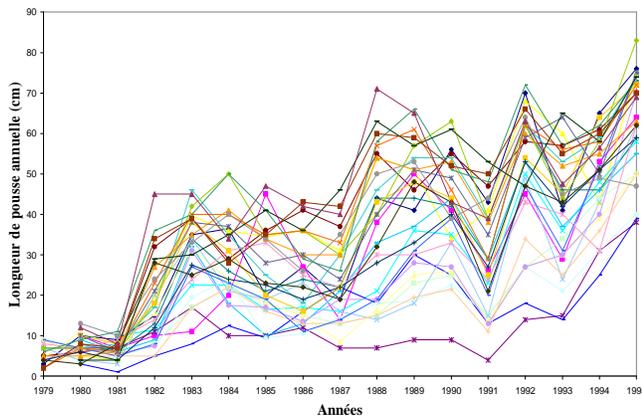


Fig. 1. Pins Laricio de 18 ans : longueurs (en cm) de pousses annuelles successives en fonction de l'année.

III. MÉTHODES.

Avant de proposer un modèle statistique qui permet de différencier et de caractériser les différentes composantes de la croissance d'un arbre, nous allons présenter deux grandes familles statistiques sur lesquelles il repose.

III.1. Semi-chaînes de Markov cachées.

Une chaîne de Markov d'ordre 1 est un processus stochastique à temps discret et à espace d'états discrets caractérisé par le fait que tout le passé du processus (du temps 1 au temps t) est seulement dépendant de l'état précédent (au temps $t-1$), ou encore, que le présent étant connu (temps $t-1$), le futur (temps t) est indépendant du passé. Une chaîne de Markov $\{S_t\}$ d'ordre 1 à J états, homogène dans le temps, est définie par des probabilités initiales d'être dans l'état j au temps 1 et par des probabilités de transition de passer dans un état au temps t sachant l'état dans lequel on se trouvait au temps $t-1$.

A partir d'une chaîne de Markov, il est possible de définir de nouvelles classes de modèles. Une solution classique consiste à faire l'hypothèse qu'une séquence n'est pas directement générée par une chaîne de Markov mais indirectement par des lois de probabilités attachées aux états de la chaîne de Markov. Une chaîne de Markov cachée [1] peut être vue comme un couple de processus stochastiques $\{S_t, Y_t\}$ tel que le processus $\{S_t\}$, appelé processus d'état, soit une chaîne de Markov d'ordre 1 et que le processus $\{Y_t\}$, appelé processus d'observation, soit lié au processus

d'état par une fonction probabiliste f telle que $Y_t = f(S_t)$. A noter que le processus d'état est « caché », c'est-à-dire non observable.

Les semi-chaînes de Markov cachées généralisent les chaînes de Markov cachées avec la propriété distinctive de modéliser explicitement le temps d'occupation d'un état par opposition aux chaînes de Markov cachées où le temps d'occupation d'un état suit implicitement une loi géométrique. Dans notre application biologique, le processus d'observation représente les longueurs de pousses annuelles, un état représente une phase de croissance et le processus d'état (semi-chaîne de Markov sous-jacente) représente la succession des phases de croissances d'un arbre.

III.2. Modèles linéaires mixtes.

Les données échantillonnées présentent une certaine variabilité dont on voudrait déterminer les différentes sources et mesurer leurs importances. Dans le cas de données observées supposées gaussiennes, les modèles linéaires mixtes (LMM) constituent un outil statistique performant pour analyser et étudier cette variabilité [5]. La méthodologie repose sur l'introduction d'effets aléatoires dans la modélisation afin de séparer les différentes sources de variation. A noter que dans les modèles linéaires mixtes, on distingue deux types d'effets :

- les effets fixes : ils ont un nombre fini de niveaux sur lesquels se répartissent les données. On souhaite en retirer une information sur l'influence de chaque niveau sur la variable observée.
- et les effets aléatoires : ils ont un nombre infini de niveaux (en opposition avec les effets fixes). Un échantillon seulement de ces niveaux étant représenté, on ne cherche pas à connaître l'influence de chaque niveau mais on souhaiterait connaître la part de variabilité induite par cet effet.

Ainsi, l'introduction d'effets aléatoires dans la modélisation permet de séparer la variabilité en deux parts : la variabilité liée aux effets aléatoires et la variabilité résiduelle. Dans notre application biologique, la composante environnementale sera modélisée par un effet fixe et la composante individuelle par un effet aléatoire.

III.3. Combinaison semi-markovienne de modèles linéaires mixtes.

Guédon *et al.* [3] ont mis en évidence que les fluctuations environnementales, principalement d'origines climatiques ainsi que la variabilité individuelle variaient d'une phase de croissance à une autre. Comme évoqué précédemment, la variabilité est soit liée à l'hétérogénéité inter-individuelle (composante individuelle), soit résiduelle. Nous proposons ici de modéliser ce phénomène en associant un modèle linéaire mixte à chaque phase de croissance. Le modèle proposé, appelé « combinaison semi-markovienne de modèles linéaires mixtes », appartient à la famille des combinaisons markoviennes [2]. Une semi chaîne de

Markov représente la succession de phases de croissance alors que des modèles linéaires mixtes rattachés à chaque état de la semi chaîne de Markov sous-jacente représentent à la fois l'influence des variables climatiques comme effets fixes et l'hétérogénéité inter-individuelle comme un effet aléatoire ([5] et partie III.2.).

Nous proposons maintenant le modèle de croissance suivant. Soit Y_{at} la longueur de pousse annuelle et S_{at} la phase de croissance non-observable pour l'individu a , $a=1, \dots, N$, au temps t , $t=1, \dots, T_a$. Une **combinaison semi-markovienne de modèles linéaires mixtes** peut être vue comme une paire de processus stochastiques $\{S_{at}, Y_{at}\}$ où le processus d'observation $\{Y_{at}\}$ est lié au processus $\{S_{at}\}$, qui est une semi-chaîne de Markov d'ordre 1, par le modèle linéaire mixte suivant :

$$Y_{at} = \underbrace{X_{at} \beta_{s_{at}}}_{\text{partie effets fixes}} + \underbrace{\tau_{s_{at}} \xi_{a,s_{at}}}_{\text{parties effets aléatoires}} + \underbrace{\varepsilon_{at}}_{\text{résidus}}$$

où X_{at} est le vecteur des variables environnementales (effets fixes).

Sur la phase de croissance $S_{at}=s_{at}$,

- $\beta_{s_{at}}$ est le vecteur des paramètres inconnus des effets fixes,
- $\xi_{a,s_{at}}$ est la réalisation inconnue de l'hétérogénéité inter-individuelle pour l'individu a dont la distribution suit la loi normale centrée réduite $N(0,1)$,
- $\tau_{s_{at}}$ est la contribution inconnue de l'hétérogénéité inter-individuelle,
- ε_{at} est l'erreur résiduelle dont la distribution $\varepsilon_{at} | S_{at} = s_{at}$ suit la loi $N(0, \sigma_{s_{at}}^2)$,
- et $\sigma_{s_{at}}^2$ est la variance résiduelle inconnue.

Il est important de noter que dans ces modèles, nous avons deux structures non-observées (ou « cachées »): les phases de croissance et les réalisations de l'effet aléatoire modélisant l'hétérogénéité inter-individuelle. Les combinaisons semi-markoviennes de modèles linéaires mixtes sont estimées à partir d'un algorithme de type MCEM (Monte Carlo Expectation Maximization, [4, 6]) dont les itérations se décomposent en 3 étapes : simulation des séquences d'états sachant les réalisations de l'effet aléatoire, prédiction des réalisations de l'effet aléatoire sachant les séquences d'états et maximisation.

IV. RESULTATS.

Une combinaison semi-markovienne de modèles linéaires mixtes, de type « gauche-droite » à 3 états composée de

deux états transitoires successifs et d'un état final absorbant, a été estimée sur la base des quatre sous-échantillons de pins Laricio. Un état est dit transitoire si après avoir quitté cet état, il est impossible d'y revenir. Un état est dit absorbant, si après avoir pénétré dans cet état, il est impossible d'en sortir. Dans les régions tempérées, les précipitations peuvent avoir soit un effet retard d'une année (sur le nombre d'éléments), soit un effet immédiat (sur l'allongement des pousses) selon qu'elles surviennent au cours de l'organogenèse¹ ou de l'allongement. Nous avons donc choisi comme effets fixes, une constante et le cumul des précipitations (en mm) durant une période recouvrant l'organogenèse et l'allongement d'une pousse annuelle.

Les distributions de la longueur de pousses annuelles sont bien séparées entre les phases de croissance (peu de chevauchement des distributions de la longueur de pousses annuelles entre deux états successifs); comparaison de la différence moyenne $m_{j+1} - m_j$ entre les états successifs et des écarts types σ_j et σ_{j+1} dans le tableau 1.

Tableau 1. Décomposition de la variabilité sur chaque phase de croissance. Moyenne (m_j) et écart type (σ_j) de la longueur de pousses annuelles sur chaque phase de croissance.

	Phase de croissance		
	1	2	3
Variance aléatoire	6.80	52.05	72.29
Variance résiduelle	5.11	39.84	76.54
Variance totale	11.91	91.89	148.83
Part d'hétérogénéité	57.08%	56.64%	48.57%
Longueur de pousses annuelles (m_j, σ_j)	7.19, 3.45	26.06, 9.59	50.43, 12.20

Le modèle estimé montre que l'influence de la pluie cumulée est faible sur la première phase de croissance (correspondant au début de la vie de l'arbre), puis augmente fortement avec les phases de croissances (un petit peu moins dans la seconde phase de croissance que dans la troisième phase de croissance); voir par exemple les arbres âgés de 18 ans (Fig.2); tandis que bien que la variabilité augmente le long de la vie de l'arbre, la part d'hétérogénéité inter-individuelle décroît plus lentement (Tableau 1).

¹ Origine et développement des organes et des ensembles d'organes chez les plantes

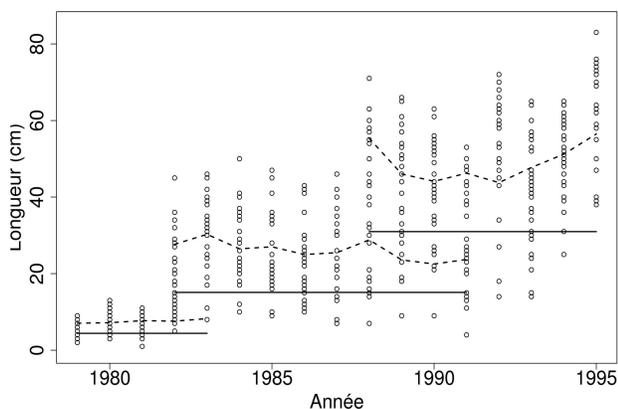


Fig. 2. Pins Laricio âgés de 18 ans: représentation graphique des longueurs (en cm) de pousses annuelles observées (points) comme une succession de 3 phases de croissance. Sur chaque phase de croissance, la constante est représentée par un trait et la constante plus l'influence de la pluie cumulée par des pointillés.

V. CONCLUSION.

Les combinaisons semi-markoviennes de modèles linéaires mixtes permettent d'identifier et de caractériser les différentes composantes (ontogénique, environnementale et individuelle) de la croissance d'un arbre. Le comportement de chaque arbre au sein de la population peut être étudié sur la base des réalisations de l'effet aléatoire hétérogénéité inter-individuelle prédites sur chaque phase de croissance.

Une intéressante direction pour des travaux futurs serait d'étudier les combinaisons semi-markoviennes de modèles linéaires mixtes pour d'autres types d'effets aléatoires. Par exemple, si les données météorologiques ne sont pas en notre possession, il serait intéressant de modéliser la composante environnementale par un effet aléatoire environnement commun à tous les individus pour une année donnée.

Une perspective serait de développer la méthodologie statistique pour des combinaisons semi-markoviennes de modèles linéaires généralisés mixtes où les observations sont supposées appartenir à la famille exponentielle (par exemple présence/absence de bourgeon, nombre de branches, ect...).

RÉFÉRENCES.

- [1] Y. Ephraïm, N. Merhav. « Hidden Markov processes », dans IEEE Transactions on Information Theory, 2002, 48, 6, 1518-1569.
- [2] S. Frühwirth-Schnatter. « Finite Mixture and Markov Switching models », 2006, Springer Series in Statistics, New York, NY: Springer, 492 p.
- [3] Y. Guédon, Y Caraglio, P. Heuret, E. Lebarbier, C. Meredieu. « Analyzing growth components in trees », dans Journal of Theoretical Biology, 2007, 248, 3, 418-447.
- [4] G.J. McLachlan, T. Krishnan. « The EM algorithm and Extensions », 1997, Wiley Series in Probability and

Mathematical Statistics, Applied Probability and Statistics, New York, NY: John Wiley & Sons, 274 p.

- [5] G. Verbeke, G. Molenberghs. « Linear mixed models for longitudinal data », 2000, Springer Series in Statistics, Berlin: Springer, 568 p.
- [6] G. Wei, M. Tanner. « A Monte Carlo implementation of the EM algorithm and the poor's man data augmentation algorithm », dans Journal of American Statistical Association, 1990, 85, 699-704.

Mélanges de lois normales bivariées appliqués à l'identification de marqueurs dans les spectres de masse SELDI-TOF.

Nadège DOSSAT¹ (Doctorant 3e année), Jean-Pierre DAURÈS², Nicolas MOLINARI³
(Directeurs de Thèse)

Institut Universitaire de Recherche Clinique
641, avenue du doyen Gaston GIRAUD 34093 Montpellier cedex 5, France

¹nadege.dossat@iurc.montp.inserm.fr, ²daures@iurc.montp.inserm.fr, ³molinari@iurc.montp.inserm.fr

Résumé—La protéomique est aujourd'hui indispensable dans la compréhension de maladies, telle que le cancer. Les protéines sont les caractères observables de la maladie. La spectrométrie de masse SELDI-TOF est un des outils laboratoires utilisés dans l'identification des protéines. Les spectres produits sont des signaux bruités et non alignés. Les biomarqueurs qui sont des protéines ou des morceaux de protéines sont caractérisés par leur rapport masse sur charge (unité Dalton). Cette mesure est présente en abscisse sur les spectres de masse. L'ordonnée d'un biomarqueur sur le spectre est l'intensité qui représente la quantité de celui-ci chez chaque individu. Le bruit et le non alignement des spectres rendent difficile l'identification des biomarqueurs. Après un débruitage des spectres par ondelettes, la difficulté d'identification des protéines a été résolue par l'utilisation des mélanges de lois normales bivariées hétérogènes.

Mots-clés—Cancer du sein, Protéomique, Spectres de masse SELDI-TOF, Débruitage, Mélange fini de lois normales bivariées.

I. INTRODUCTION

La protéomique est définie comme l'étude des protéines exprimées, incluant l'identification et la résolution du rapport structure/fonction dans des conditions saine et de maladie. La protéomique en fournissant une vue intégrée du processus individuel de la maladie au niveau de la protéine peut permettre de diagnostiquer la maladie et de choisir une pharmacothérapie spécifique. Les profils protéiques de 29 femmes atteintes du cancer du sein et de 29 témoins sains ont été analysés à l'aide de la spectrométrie de masse. Un spectre fournit la mesure d'un ensemble de biomarqueurs, i.e. de protéines ou de morceaux de protéines. Un biomarqueur est caractérisé par son abscisse (masse/charge électrique nette) sur le spectre. L'intensité en ordonnée représente la quantité de celui-ci. Le spectromètre de masse SELDI-TOF fournit des signaux bruités et non alignés. Cette variabilité en abscisse et en ordonnée rend difficile l'identification des biomarqueurs. Nous avons pris en compte cette double variabilité en utilisant des lois normales bivariées pour modéliser la loi de chaque biomarqueur. L'identification de l'ensemble des biomarqueurs revenant à identifier le mélange fini de lois normales bivariées

optimum.

II. DÉBRUITAGE DES SPECTRES PAR ONDELETTES ET DÉTECTION DES BIOMARQUEURS

Dans un premier temps, il a été procédé à un débruitage des spectres par ondelettes [1, 3]. La figure 1 représente l'analyse multirésolution d'un des spectres. Le spectre s est décomposé en une approximation \hat{A}_d et en une somme de détails D_d .

$$s = \hat{A}_d + D_d$$

\hat{A}_d est la projection orthogonale du signal sur l'espace d'approximation de niveau 6 (a_6 sur la partie gauche de la figure 1). Le niveau d'approximation de niveau 6 a été choisi car les approximations de niveau 5 et 6 sont très proches.

$$\hat{A}_d = \sum_k \hat{\beta}_{6,k} \varphi_{6,k}$$

Les fonctions φ sont les fonctions d'échelles. D_d est la somme de détails D_j qui sont les projections orthogonales du signal sur les espaces des détails de niveaux j .

$$D_d = \sum_{1 \leq j \leq 6} D_j$$

avec $D_j = \sum_k \alpha_{j,k} \psi_{j,k}$, où les fonctions $\psi_{j,k}$ et $\alpha_{j,k}$ sont respectivement les atomes de la base d'ondelette et les coefficients d'ondelette. Les indices j et k sont le niveau de dilatation et la translation de l'ondelette mère. Le débruitage d'un spectre consiste en un choix du niveau d'approximation (ici le niveau 6) et d'un seuillage des coefficients d'ondelettes $\alpha_{j,k}$ ($d_j|_{j=1,\dots,6}$ sur la partie droite de la figure 1). Nous avons effectué un seuillage dur, i.e. tous les coefficients compris dans les intervalles en pointillés sur la figure 1 ont été mis à zéro. Le signal débruité s'écrit alors de la manière suivante :

$$\hat{s} = \hat{A}_d + \hat{D}_d$$

avec $\hat{D}_d = \sum_{1 \leq j \leq 6} \sum_k \hat{\alpha}_{j,k} \psi_{j,k}$ qui est la somme des détails seuillés. Le seuillage des coefficients d'ondelettes

permet de récupérer les informations les plus importantes perdues dans l'approximation du signal \hat{A}_d . Nous définissons les biomarqueurs comme étant les maxima locaux des spectres débruités obtenus précédemment. Une fois les biomarqueurs détectés, il reste encore à les identifier.

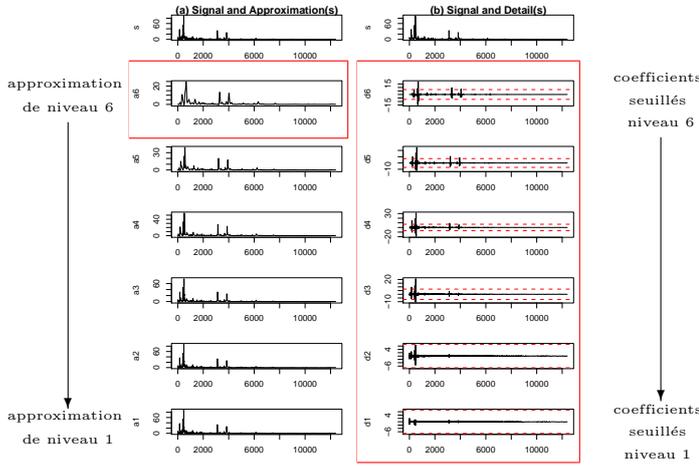


Fig. 1. Analyse multirésolution du spectre s , avec les ondelettes symplettes $la8$. Sur la droite, les approximations a_j du niveau le plus grossier (niveau 6) au niveau le plus fin (niveau 1). Sur la gauche, les coefficients d'ondelettes du niveau le plus grossier (niveau 6) au niveau le plus fin (niveau 1). Les coefficients seuillés sont les coefficients d'ondelettes modifiés de telle sorte que les coefficients compris dans les intervalles rouges sont mis à zéro.

III. IDENTIFICATION DES BIOMARQUEURS

L'identification des biomarqueurs revient à les regrouper en fonction de leur nature. Autrement dit des biomarqueurs qui sont considérés comme ayant la même abscisse sur le spectre sont pris comme étant identiques. Ceci revient à une classification des biomarqueurs. Comme nous l'avons dit précédemment la variabilité en abscisse et en ordonnée rend cette classification difficile. De plus, un biomarqueur peut ne pas être présent chez tous les individus, il y a non reproductibilité à 100% des biomarqueurs. Deux biomarqueurs peuvent être très proches en abscisse et ne différer que par leur intensité, autrement dit par la quantité de ceux-ci dans l'échantillon analysé. Il s'agit donc de considérer cette problématique en deux dimensions. Pour ce faire nous avons émis l'hypothèse que les biomarqueurs sont indépendants et que tous suivent une loi normale bivariable. L'identification des biomarqueurs revient à trouver un mélange fini de loi normale bivariable [2] optimum en terme de BIC . La densité d'un mélange fini de lois s'écrit sous la forme d'une somme finie et pondérée de densités de distribution.

$$f(y) = \sum_{j=1}^K p_j \phi_j(y)$$

avec $p_j|_{j=1,\dots,K}$ les poids qui sont aussi appelés proportions du mélange. Ces proportions respectent les contraintes $0 < p_j < 1$ et $\sum_{j=1}^K p_j = 1$. Les densités

$\phi_j|_{j=1,\dots,K}$ sont choisis comme étant de la même famille paramétrique de lois de probabilité, chaque composante du mélange suit une loi normale bivariable. L'indice j fait référence aux différentes valeurs des paramètres. La densité de la j ème composante (ou de la loi de la classe j) ϕ_j peut aussi s'écrire $\phi(y|\mu_j, \Sigma_j)$, avec $\mu_j \in \mathbf{R}^2$ la moyenne et $\Sigma_j \in \mathbf{R}^{2 \times 2}$ la matrice de variance-covariance. Le nombre de biomarqueurs ainsi que la nature de ceux-ci n'étant pas forcément les mêmes dans les deux groupes, nous modéliserons un mélange de modèles pour chaque groupe (cancer/témoin).

A. L'algorithme EM

Si nous nous plaçons dans un des groupes, nous noterons $y = (y_1, \dots, y_n) \in \mathbf{R}^{2 \times n}$ l'ensemble des n biomarqueurs observés pour tous les individus du groupe. Nous ne connaissons pas la composante pour laquelle le i ème biomarqueur y_i est observé, i.e. la valeur j telle que $Y_i \sim N_2(\mu_j, \Sigma_j)$. Nous noterons $z = (z_1, \dots, z_n) \in \mathbf{R}^{K \times n}$ les données manquantes. Tel que $z_i|_{i=1,\dots,n}$ est un vecteur de données binaires de dimension K avec $z_{ij} = 1$ si et seulement si le biomarqueur y_i appartient à la composante j . L'échantillon complet est $x = (x_1, \dots, x_n)$ avec $x_i = (y_i, z_i)$. L'espace des paramètres est noté Θ , tel que les paramètres $\theta \in \Theta$ s'écrivent sous la forme :

$$\theta = (p_1, \dots, p_K, \mu_1, \dots, \mu_K, \Sigma_1, \dots, \Sigma_K)$$

avec $0 < p_j < K$ et $\sum_{j=1}^K p_j = 1$. Notre problématique correspond au calcul de l'estimateur de maximum de vraisemblance $\hat{\theta}$ avec des données manquantes, ce qui nécessite l'emploi de l'algorithme **EM**. Nous noterons $L(x; \theta)$ la log vraisemblance des données complètes.

$$\begin{aligned} L(x; \theta) &= \log \left(\prod_{i=1}^n \prod_{j=1}^K p_j^{z_{ij}} \phi(y_i | \mu_j, \Sigma_j)^{z_{ij}} \right) \\ &= \sum_{i=1}^n \sum_{j=1}^K z_{ij} [\log p_j + \log \phi(y_i | \mu_j, \Sigma_j)] \end{aligned}$$

Une première étape (0) permet d'initialiser les paramètres du mélange $\theta^{(0)}$. Les étapes suivantes sont constituées d'une phase **Expectation** et **Maximization** d'où l'algorithme **EM** tire son nom. La description de la $(r+1)$ ème étape est comme suit :

1. Phase **Expectation** : Calcul de $Q(\theta, \theta^{(r)})$.

$$\begin{aligned} Q(\theta, \theta^{(r)}) &= E_{\theta} [L(x; \theta)] \\ &= \sum_{i=1}^n \sum_{j=1}^K E_{\theta} [z_{ij}] [\log p_j + \log \phi(y_i | \mu_j, \Sigma_j)] \end{aligned}$$

2. Phase **Maximization** : Calcul de $\theta^{(r+1)} = \arg \max_{\theta \in \Theta} Q(\theta, \theta^{(r)})$

Pour maximiser $Q(\theta, \theta^{(r)})$ par rapport à θ ; il suffit de maximiser par rapport à θ la log vraisemblance complète $L(x; \theta)$ et de remplacer dans le résultat z_{ij} par $E_{\theta^{(r)}} [z_{ij}]$. Ce qui revient à calculer $\theta^{(r+1)} = (p_1^{(r+1)}, \dots, p_K^{(r+1)}, \mu_1^{(r+1)}, \dots, \mu_K^{(r+1)}, \Sigma_1^{(r+1)}, \dots, \Sigma_K^{(r+1)})$:

- $p_j^{(r+1)} = \frac{1}{n} \sum_{i=1}^n E_{\theta^{(r)}} [z_{ij}]$
- $\mu_j^{(r+1)} = \frac{\sum_{i=1}^n E_{\theta^{(r)}} [z_{ij}] y_i}{\sum_{i=1}^n E_{\theta^{(r)}} [z_{ij}]}$

- $\Sigma_j^{(r+1)} = \frac{\sum_{i=1}^n E_{\theta^{(r)}} [z_{ij}] (y_i - \mu_i^{(r+1)})(y_i - \mu_i^{(r+1)})^T}{\sum_{i=1}^n E_{\theta^{(r)}} [z_{ij}]}$
- avec $E_{\theta^{(r)}} [z_{ij}] = P(z_{ij} = 1/\theta^{(r)}, y_i)$

Ces deux phases sont exécutées jusqu'à ce que le nombre d'itérations soit atteint, ou jusqu'à ce que les paramètres $\theta^{(r)}$ aient une évolution entre l'étape (r) et ($r+1$) inférieure à un certain ε fixé. L'algorithme **EM** garantit la convergence vers θ^* un point stationnaire de la vraisemblance $L(\theta|y)$, et non pas vers un maximum global. Il faut donc un grand nombre de départs aléatoires $\theta^{(0)}$.

B. L'algorithme **EM** contraint

Nous n'avons aucune information sur l'identité des biomarqueurs. Aussi aucune hypothèse sur la forme des matrices de variance-covariance n'a été imposée. Ceci pose un problème car la log vraisemblance pour un mélange fini de distributions normales bivariées avec des matrices de variance-covariance inégales dans les différentes composantes est non bornée. Autrement dit un maximum global n'existe pas. Concrètement, il se peut que lors d'une étape de l'algorithme **EM** les paramètres $\theta^{(r)}$ se trouve sur le bord de l'espace des paramètres Θ avec une classe j singulière qui ne contiendra qu'un seul individu i (i.e. $\mu_j = y_i$ et $\sigma_j \rightarrow 0$). La solution est de limiter l'espace des paramètres Θ à un sous-ensemble compact de celui-ci en imposant des contraintes sur la forme des matrices de variance-covariance pour qu'elles ne soient pas trop disparates d'une composante à l'autre. Nous recherchons donc un maximum global contraint. Une façon de poser des contraintes sur la forme des matrices de variance-covariance est de limiter les valeurs propres de celles-ci à un intervalle $[a, b]$. Nous recherchons un maximum local de la vraisemblance pour choisir les constantes a et b . Nous avons fait une classification k-means pour initialiser l'algorithme **EM**, puis nous avons employé l'algorithme **EM** pour trouver des paramètres θ^* tel que $L(\theta^*|y)$ est un maximum local de $L(\theta|y)$. Ensuite, nous avons calculé toutes les valeurs propres des matrices de variance-covariance $\Sigma_1^*, \dots, \Sigma_K^*$. Les constantes a et b ont été prises comme étant la plus petite et respectivement la plus grande des valeurs propres. Cette contrainte sur les valeurs propres des matrices de variance-covariance s'applique dans la phase de **Maximization**. A chaque étape (r), les matrices de variance-covariance sont diagonalisées.

$$\Sigma_j^{(r)} = P_j^{-1} D_j P_j = \text{diag}(\lambda_{1j}, \lambda_{2j}), j = 1, \dots, K$$

La contraintes sur les valeurs propres sont alors appliquées, en modifiant la matrice diagonale D_j en une nouvelle matrice D'_j .

$$\begin{cases} \text{Si } \min(\lambda_{1j}, \lambda_{2j}) < a, \text{ alors on lui affecte la valeur } a. \\ \text{Si } \max(\lambda_{1j}, \lambda_{2j}) > b, \text{ alors on lui affecte la valeur } b. \end{cases}$$

Nous obtenons de nouvelles matrices de variance-covariance $\Sigma_j^{(r)} = P_j^{-1} D'_j P_j$. Cet algorithme **EM** contraint a été exécuté sur 500 départs aléatoires pour un nombre de composantes K fixé. Le mélange à K composantes retenu est celui qui donne une vraisemblance maximale.

C. Choix du nombre de composantes

L'autre inconnue dans notre mélange fini de lois normales bivariées est le nombre K de composantes, puisque nous n'avons aucune information priori sur le nombre de biomarqueurs présents dans l'échantillon. Nous avons choisi le critère BIC pour trouver le modèle le plus parcimonieux en pénalisant la vraisemblance par un terme tenant compte de la complexité du modèle. Le critère de BIC est très simple à calculer. Sa formule est la suivante :

$$BIC(\hat{\theta}, K) = -2 \log f(y|\hat{\theta}, K) + \nu_K \log(n)$$

avec ν_K le nombre de paramètres libres du mélange de modèle. Nous avons choisi le nombre de composantes K de telle sorte que les trois modèles suivants avec $K+1$, $K+2$, $K+3$ composantes aient une valeur de BIC supérieure au modèle à K composantes.

IV. CONCLUSION ET PERSPECTIVES

Cette méthode d'identification des marqueurs a plusieurs avantages. Elle permet de détecter deux biomarqueurs qui ont des abscisses (m/z) proches voir qui se chevauchent et qui ne peuvent être identifiés que par leurs intensités qui diffèrent. Nous pouvons prendre comme exemple les deux marqueurs identifiés vers l'abscisse $3200Da$ (c.f. figure2) qui ont clairement une expression (intensité) très différente dans les deux groupes. Les biomarqueurs n'ont pas besoin d'être présents chez tous les individus pour être détectés. Nous avons modélisé indépendamment les mélanges de lois normales bivariées pour les deux groupes. Les marqueurs identifiés sont de même nature dans les deux groupes. Une comparaison de plusieurs débruitages et d'autres critères que le BIC sont en cours. La réalité sur l'identité des biomarqueurs n'étant pas connue, le seul moyen de comparer les résultats obtenus est d'évaluer le pouvoir discriminant des biomarqueurs identifiés.

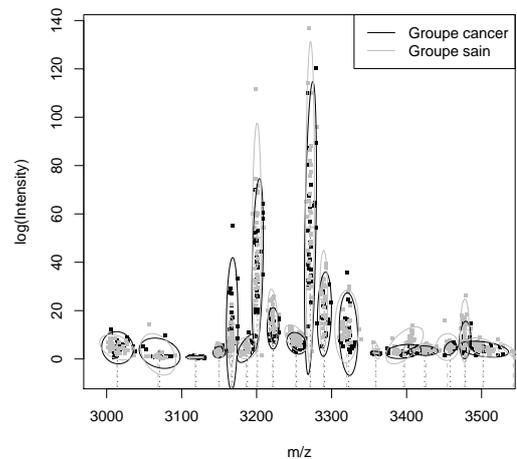


Fig. 2. Extrait des biomarqueurs détectés dans la première partie du spectre entre les abscisses $3000Da$ et $3537Da$. Les mélanges finis de lois normales bivariées obtenus pour les deux groupes (cancer/témoin) ont été représentés sur le même graphique. Les ellipses de confiance à 95% pour chaque biomarqueur permettent de visualiser les biomarqueurs identifiés.

RÉFÉRENCES

- [1] KR. Coombes, S. Tsavachidis, JS. Morris, KA. Baggerly, MC. Hung, HM. Kuerer, « Improved peak detection and quantification of mass spectrometry data acquired from surface-enhanced laser desorption and ionization by denoising spectra with the undecimated discrete wavelet transform. » *Proteomics*, vol. 5(16),pp. 4107–4117, Nov, 2005
- [2] G. McLachlan, D. Peel, *Finite Mixture Models*, Wiley, New York, États-Unis, 2000.
- [3] M. Misiti, Y. Misiti, G. Oppenheim, et al., *Les ondelettes et leurs applications*, Hermès Science, Paris, France, 2003

RECHERCHE DE STRATEGIE BIOSTATISTIQUE DANS LES ESSAIS CLINIQUES DE PHASE I/II D'ASSOCIATION COMPRENANT UN AGENT NON CYTOTOXIQUE

NADINE HOUEDE¹ (DOCTORANTE 4E ANNEE), ANDREW KRAMAR² (DIRECTEUR DE THESE)

Unité de Biostatistiques, CRLC Val d'Aurelle Paul Lamarque
Parc Euromédecine 208 Rue des Apothicaires, 34298 Montpellier Cedex, France

HOUEDE@BERGONIE.ORG

ANDREW.KRAMAR@VALDOREL.FNCLCC.FR

[HTTP://WWW.EDI2S.UNIV-MONTP2.FR/](http://WWW.EDI2S.UNIV-MONTP2.FR/)

Résumé : Nous proposons une nouvelle stratégie pour déterminer la combinaison de dose optimale d'une association cytotoxique et thérapie ciblée, basée sur trois variables d'évaluation : la toxicité associée à chacun des médicaments et l'efficacité de l'association.

Mots clefs : Bayésien, phase I/II, Efficacité, Toxicité, association thérapeutique.

I. INTRODUCTION.

L'émergence de nouvelles thérapies ciblées dans le traitement du cancer a entraîné de nouveaux challenges pour définir le meilleur index thérapeutique. L'objectif principal d'un essai de phase I est de recommander une dose tolérable pour pouvoir ensuite évaluer son efficacité dans un essai de phase II. Les index thérapeutiques des thérapies ciblées sont en général élevés et ils peuvent être efficaces sans que l'on ait besoin de s'approcher de la dose maximale tolérée. Avec le développement d'essais d'association et de thérapies ciblées, l'escalade de dose de type 3+3 n'est plus optimale. La méthode CRM [1], basée sur un modèle Bayésien, a permis d'ouvrir la voie vers d'autres perspectives : évaluation de la relation dose-réponse en association, évaluation de deux paramètres de façon concomitante: toxicité/efficacité. L'objectif est de choisir la meilleure dose à utiliser dans un essai clinique de phase III, pour estimer les probabilités d'efficacité et de toxicité à cette dose, et d'arrêter l'essai de façon précoce s'il devient improbable de trouver une dose qui est à la fois non toxique et efficace.

II. MOTIVATION.

Le cancer de vessie représente environ 10000 cas par an en France en moyenne [2], dans un tiers des cas il se présente sous forme invasive. La durée médiane de survie est de 18 mois quand il existe une atteinte ganglionnaire, de 9 mois

quand il existe des métastases viscérales. Le traitement de référence en chimiothérapie pour les cancers de vessie métastatique, comme gemcitabine/cisplatine (GC) ou methotrexate, vinblastine, doxorubicine, et cisplatine (M-VAC), permette d'obtenir des taux de réponses intéressant allant de 45% à 70% [3-5]. Malgré cela la survie sans rechute à long terme est rare et la plupart des patients meurent de l'évolution de leur maladie. Aucun progrès en survie n'a été réalisé ces dix dernières années pour ces patients. Basé sur des données précliniques et leur synergie, l'association d'un anti-VEGFR, combiné à la chimiothérapie de référence pourrait améliorer la survie prolongée de nos patients. Un facteur qui limite le progrès du développement de ces combinaisons de traitement (chimiothérapie-thérapie ciblée) est l'existence d'un schéma approprié d'essai clinique évaluant effectivement et efficacement ces thérapies ciblées.

III. SCHEMA D'ETUDE.

Il s'agit d'un essai de phase I/II de combinaison de traitement de stratégie d'escalade de dose d'une association anti-VEGFR-gemcitabine-CDDP chez des patients non pré-traités par chimiothérapie, suivis pour un cancer des voies urinaires inopérable ou métastatique (stade IV). La recherche de dose est basée sur trois variables ordinales : l'efficacité, la toxicité liée à la gemcitabine, la toxicité liée à l'anti-VEGFR. Des cohortes de trois patients sont traitées par palier, suivant la grille de dose (figure 1). Les scores d'utilité pour chaque paire de dose sont calculés, basés sur les données obtenues de patients déjà traités, utilisant une formulation de type Bayésien. Pour des raisons de faisabilité, nous avons fait le choix de traiter au maximum 48 patients (3 par palier et 12 au palier de doses optimales). Pour des raisons de sécurité, si toutes les combinaisons de dose ont une toxicité inacceptable, l'essai doit être clos prématurément.

Schéma d'administration dans le cas d'une association entre une thérapie ciblée et un agent cytotoxique

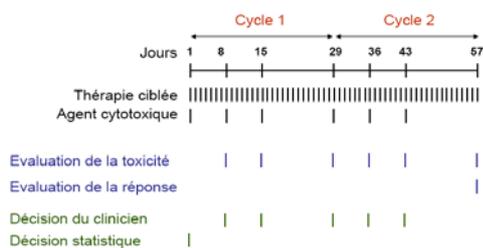


Figure 1 : Schéma de l'étude

IV. GRILLE D'ESCALADE DE DOSE.

Le problème que nous posons est de choisir la paire de dose optimale pour l'association des deux, pour les deux premiers cycles de traitement. Dans le cas de notre essai, il existe douze combinaisons de dose, comme illustré en figure 2 :

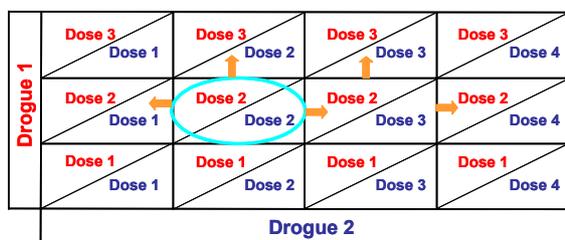


Figure 2 : Grille de combinaison de dose

V. MODELE.

Dans l'essai sur le cancer de la vessie, tous les critères sont des variables ordonnées à trois modalités. On définit les toxicités associées à l'agent biologique par Y_1 et les toxicités associées à l'agent chimique par Y_2 . Puisque la méthode fait la distinction entre les toxicités sévères qui sont ou non résolues, pour chaque type de toxicité $k = 1, 2$, on définit $Y_k = 0$ si la toxicité sévère (grade 3, 4) de type k n'est pas observée, $Y_k = 1$ si la toxicité sévère de type k est observée mais qu'elle est résolue en moins de 2 semaines, et $Y_k = 2$ si la toxicité sévère de type k est observée mais qu'elle n'est pas résolue dans les 2 semaines.

La variable d'efficacité $Y_3 = 0$ si la réponse tumorale (PR or CR) est obtenue, $Y_3 = 1$ si la maladie est stable (SD) à la fin du deuxième cycle de traitement, et $Y_3 = 2$ dès que la maladie progresse (PD). Le traitement est arrêté si on observe une progression précoce, c'est-à-dire dès la fin du premier cycle de 28-jours.

Pour tenir compte des effets conjoints de d_1 et d_2 pour chaque entrée de Y , ainsi que l'association entre ces éléments, la procédure commence par une modélisation de la distribution marginal de chaque critère, en utilisant le modèle très flexible d'Aranda-Ordaz [6] et ensuite ces distributions marginales sont combinées en utilisant une copula gaussienne trivariée [7].

VI. REGLES DE DECISION.

La définition d'utilité entend la qualité de bénéfice rendu au patient pour chaque couple de résultat obtenu en termes d'efficacité et de toxicité. Afin de définir ces scores d'utilités, nous avons utilisé la technique Delphi qui est une technique interactive qui permet d'obtenir un consensus d'experts, en tout anonymat des experts les uns par rapport aux autres.

Algorithme de décision :

- 1- Traiter la première cohorte de patients au couple de dose situé au milieu de la grille.
- 2- La dose d'un seul médicament peut-être modifié à la fois, d'un seul palier, aucune case de la grille ne sera sautée.
- 3- Avant d'inclure une nouvelle cohorte de patient, il faut recueillir les données des deux premiers cycles des patients précédents.
- 4- Le palier de dose de la nouvelle cohorte de patient est calculée afin d'optimiser la fonction d'utilité
- 5- Règles d'arrêt de l'essai : si aucun palier de dose n'a une toxicité et une efficacité acceptable, l'essai est terminé précocement.

VII. CONCLUSION ET PERSPECTIVES.

Notre travail découle du besoin clinique d'avoir des outils appropriés aux progrès apportés en recherche fondamentale, avec des modèles statistiques permettant d'analyser les associations chimiothérapie et thérapie ciblée, qui représente actuellement un nombre important d'essais cliniques. Le problème est alors de choisir la paire de dose optimale de l'association.

C'est une stratégie innovante sur de nombreux points :

- Tout d'abord la différenciation de deux types de toxicité.
- L'efficacité est prise en compte dès les premières étapes de développement.
- La modélisation de chaque relation marginale dose-toxicité 1, dose-toxicité 2 et dose efficacité est donnée par le modèle très flexible d'Aranda-Ordaz, permettant de s'ajuster aux données cliniques.
- L'utilisation de Copula permet d'analyser la nature de la dépendance entre chaque variable d'évaluation.
- L'utilisation des valeurs d'utilités permet de classer les différents résultats obtenus de façon concrète en terme de bénéfice rendu au patient, selon l'avis d'un groupe d'expert, qui utilisent leur connaissance a priori des situations cliniques identifiées.
- L'utilisation d'une méthode Bayésienne, pour une recherche de dose séquentielle et adaptée, basée sur des scores d'utilités à posteriori.

Cette nouvelle stratégie permet de minimiser le nombre de patients traités à une dose inefficace, de minimiser le nombre de patients traités à dose toxique et de donner le maximum de chance pour le patient inclus de répondre au traitement. Bien sur, elle nécessite un partenariat entre clinicien et statisticien plus importants tout au long de l'essai, elle nécessite un monitoring rigoureux de chaque patient inclus, avec recueil des données au fur et à mesure de l'essai, mais cet effort est largement récompensé par les avantages éthiques et scientifiques.

Cette méthode a été développée à partir d'un modèle d'essai clinique particulier, mais peu tout a fait être adaptée à différentes situations d'association cytotoxique/thérapie ciblée, quelque soit le rythme d'administration. Elle nécessite simplement de redéfinir des scores d'utilités lié à chaque situation clinique envisageable dans le cadre de l'essai.

RÉFÉRENCES

[1] O'Quigley J, Pepe M, Fisher L, Continual reassessment method: a practical design for phase 1 clinical trials in cancer, *Biometrics*, 1990, 46, 33-48

[2] Réseau Français des registres du cancer, Francim Hôpitaux de Lyon, INSERM, Institut de Veille Sanitaire, Evolution de l'incidence et de la mortalité par cancer en France de 1978 à 2000, Saint-Maurice, 2002

[3] Beekman KW, Bradley D, Hussain M, New molecular targets and novel agents in the treatment of advanced urothelial cancer. *Semin Oncol*, 2007, 34, 154-64

[4] Bellmunt J, Albiol S, Chemotherapy for metastatic or unresectable bladder cancer, *Semin Oncol*, 2007, 34, 135-44

[5] Roberts JT, von der Maase H., Sengelov L, Conte PF, Dogliotti L, Oliver T, et al, Long-term survival results of a randomized trial comparing gemcitabine cisplatin and methotrexate vinblastine doxorubicin cisplatin in patients with locally advanced and metastatic bladder cancer, *Ann Oncol*, 2006, 17, 118-122

[6] Aranda-Ordaz FJ, An extension of the proportional-hazards model for grouped data, *Biometrics*, 1983, 39, 109-17

[7] Trivedi PK, Zimmer DM, Copula Modeling : an introduction for practitioners, *Foundations and Trends in Econometrics*, 2005, 1, 1-111

GLOSSAIRE

Index thérapeutique: Ecart entre une dose toxique et une dose active

Essai d'association : Essai de combinaisons de traitement

Thérapies ciblées : Modulateurs de la réponse biologique

CRM : Continual Reassessment Method

PD : Maladie progressive

RC=CR : Réponse complète

RP=PR : Réponse partielle

SD : Maladie stable

VEGFR : Vascular Epidermal Growth Factor receptor

Y_j : Toxicité liée à l'agent j, j=1,2

Y_j=0 : pas de toxicité

Y_j=1 : toxicité sévère (CTC grade 3-4), récupérant dans les deux semaines

Y_j=2 : toxicité sévère (CTC grade 3-4), ne récupérant pas dans les deux semaines

Y3: Efficacité

Y3=0 : réponse partielle ou complète

Y3=1 : stabilisation

Y3=2 : progression ou décès

DEVELOPPEMENTS METHODOLOGIQUES DES ESSAIS RANDOMISES EN CLUSTERS APPLICATIONS AUX ESSAIS D'INTERVENTION

BRUNO PEREIRA¹ (DOCTORANT 3E ANNEE), ANDREW KRAMAR² (DIRECTEUR DE THESE)

Unité de Biostatistiques, CRLC Val d'Aurelle Paul Lamarque
Parc Euromédecine 208 Rue des Apothicaires, 34298 Montpellier Cedex, France

1 BRUNO.PEREIRA@VALDOREL.FNCLCC.FR

2 ANDREW.KRAMAR@VALDOREL.FNCLCC.FR

[HTTP://WWW.EDI2S.UNIV-MONTP2.FR/](http://www.edi2s.univ-montp2.fr/)

Résumé— Un essai randomisé en clusters (ERC) est un essai dans lequel des clusters d'individus, plutôt que les individus eux-mêmes, sont assignés aléatoirement à différents groupes d'intervention. Les unités de randomisation dans de tels essais peuvent alors être de taille plus ou moins grande s'étendant des clusters relativement petits tels des ménages à des communautés beaucoup plus importantes tels des services d'hôpitaux, des pratiques médicales ou encore des régions administratives. Ces essais sont devenus un outil essentiel dans l'évaluation des interventions non thérapeutiques telles les changements de mode de vie, les programmes d'éducation et les innovations en termes de prestations de santé. L'objectif de cet exposé repose sur la présentation des notions fondamentales utilisées dans les choix de stratégie expérimentale des ERC. Nous discuterons également des outils méthodologiques proposés pour le développement de ces essais et aborderons les challenges posés par ce sujet.

Mots clés — clusters ; randomisation ; design ; analyse statistique.

I. INTRODUCTION.

Les essais randomisés restent à ce jour la méthode la plus rigoureuse pour déterminer une relation de cause à effet entre une intervention (thérapeutique ou non) et les résultats observés chez les sujets. Dans certaines situations, il peut être pertinent de réaliser des essais avec une unité de randomisation collective. Dans ces essais randomisés récents portant sur l'évaluation d'une intervention ou d'une stratégie de prévention, on ne randomise pas individuellement des sujets, mais des groupes de sujets appelés clusters [1]. Ces unités de randomisation peuvent être des médecins, des familles ou des classes. Il s'agit d'unités sociales pour lesquelles les sujets qui composent une unité ne peuvent être considérés comme indépendants les uns des autres : l'intervention testée ne s'appliquant pas à l'échelon individuel mais à un échelon supérieur.

Pourquoi considérer de tels essais ? Il se peut que l'intervention doive être faite naturellement au niveau du cluster et puisse affecter des populations entières et non pas des individus (campagne d'informations sur le tabac par

exemple). L'intervention peut également être destinée à des individus mais peut contaminer d'autres individus. Lors d'une intervention destinée à faire changer le comportement de médecins, on peut s'attendre à ce qu'un médecin applique une recommandation pour un malade et l'oublie pour le malade suivant. Ce type de randomisation est donc justifié lorsqu'on craint une contamination entre les groupes à comparer, c'est-à-dire que les patients d'un des bras de l'essai bénéficient également de l'intervention mise en place dans l'autre bras de l'essai. L'existence d'une telle contamination inter-groupes aurait alors une double conséquence : d'une part l'effet estimé s'en trouverait biaisé, et d'autre part, les hypothèses statistiques sous-jacentes au test réalisé ne seraient plus vérifiées puisque les deux échantillons ne seraient plus indépendants. L'évolution rapide des méthodes statistiques posant de nombreux défis; l'objectif principal de mes travaux réside dans le développement de méthodologies statistiques permettant de mieux planifier et analyser les données issues d'essais randomisés en clusters (ERC). Ces avancées peuvent être de nature purement théorique ou bien informatiques. Par ailleurs, la recherche en épidémiologie représente également une part importante de ce travail. Ainsi, nous nous intéressons à l'amélioration de la prévention des comportements à risque et des connaissances épidémiologiques ainsi qu'à l'analyse d'actions de prévention et de dépistage au niveau collectif.

Dans ce qui suit, nous préférons différencier les deux étapes essentielles de la recherche en biostatistique. Nous considérerons d'une part les différents travaux réalisés dans le choix d'une stratégie expérimentale optimale ; en particulier quel est le meilleur design associé aux données issues d'ERC ? Puis, dans une seconde partie, nous nous intéresserons à l'analyse statistique en tentant de montrer comment les travaux réalisés au cours de cette thèse répondent aux questions suivantes : quelle unité d'analyse faut-il considérer et quelles méthodes statistiques doivent être mises en œuvre ?

II. STRATEGIE EXPERIMENTALE.

Contrairement aux essais où l'on randomise chaque sujet individuellement, la randomisation par clusters conduit de façon quasi systématique à une situation où la randomisation est antérieure à l'inclusion des sujets. En pratique, on commence par randomiser les clusters, charge ensuite à chaque responsable de cluster d'inclure les sujets ; cette inclusion se faisant alors en ayant connaissance du bras d'intervention dans lequel ils seront inclus.

II.1. Ethique, consentement et biais.

Les ERC peuvent être de type cluster-cluster (randomisation et intervention dirigées vers l'ensemble d'un groupe) ou cluster-individus (randomisation de groupes et intervention dirigée vers les individus). Ils peuvent poser des problèmes éthiques principalement dans ce dernier cas où l'on considère qu'il est nécessaire d'obtenir, en plus du consentement des responsables de clusters, un consentement de tous les individus inclus (double consentement). Dans l'essai Vivre avec le Soleil (VAS) impliquant des classes de CM1/CM2, un consentement a été demandé aux enseignants mais non aux enfants ou à leurs parents. Dans ce protocole, nous demandons aux enseignants d'informer les parents de cette évaluation mais ne faisons pas signer de consentement. Conscients que la grande majorité des ERC ne présentera pas autant de facilités que l'essai VAS, nous avons réfléchi à de nouvelles solutions de randomisation. Il suffit pour cela de noter que nous sommes dans une situation semblable à celle des schémas de Zelen [2] où la randomisation précède le consentement.

Au-delà de ce problème de consentement, se pose bien souvent le problème d'inclusion des sujets notamment dans les designs de type cluster-individu. Dans l'essai VAS par exemple, l'objectif premier est d'améliorer les connaissances et les comportements des enfants en matière de prévention solaire mais la randomisation se fait au niveau de leurs enseignants. Or, même si ces derniers sont conscients de l'importance du problème de santé publique, nombreux sont ceux qui ont désiré ne pas participer à l'essai. Il nous a donc fallu réfléchir à différents types d'appel à participation et à un plan expérimental permettant de prendre en compte cette réalité.

Enfin, le fait que la randomisation soit antérieure à l'inclusion des sujets peut devenir potentiellement source de biais. La comparaison *a posteriori* des groupes issus de la randomisation est alors d'autant plus délicate. En effet, même si nous prenons nos précautions dans le choix des facteurs de stratification associés à la randomisation des clusters (par exemple région géographique par rapport aux classes), rien ne nous indique que la composition des clusters reste homogène. Le problème est dû au fait que l'on randomise sur des qualités spécifiques aux clusters mais que l'on analyse les données sur des caractéristiques associées aux sujets appartenant aux clusters. Pour illustrer cela, prenons l'exemple du programme MG Tabac LR,

impliquant 61 médecins généralistes (MG) et 1075 sujets fumeurs, qui est un ERC qui a pour objectif de montrer qu'un programme de santé publique peut modifier les pratiques d'aide des médecins à l'arrêt et augmenter les taux de sevrage tabagique de leurs sujets (à un an). A l'inclusion, les MG ayant reçu une formation particulière en tabacologie (MG dit formés dans le groupe intervention) recrutent des patients plus jeunes, plus dépendants, plus dépressifs et ayant reçu plus de prescriptions nicotiques. Que pourra-t-on conclure au moment de l'analyse des données à 1 an puisque nous sommes en fait dans le cas d'un biais de sélection majeur ? Même s'il semble difficile de se lancer dans l'interprétation de résultats dans cette étude, le fait que ces MG aient recruté des patients différents et *a priori* plus enclins à arrêter de fumer semble être un résultat en soi. L'analyse de l'effet de l'intervention, dans un tel essai, pourrait se faire donc dès la fin de l'inclusion des sujets ; reste que ce n'était pas le critère de jugement pris en compte dans cet exemple.

II.2. Coefficient de corrélation intra-cluster (ICC).

La prise en compte de corrélation entre mesures d'une même caractéristique se pose lorsque plusieurs observations proviennent d'un même cluster ; une structure de groupe apparaissant à l'intérieur de l'échantillon et donc une corrélation entre les mesures. La plupart des techniques classiques font l'hypothèse d'indépendance entre les observations ; cependant, cette hypothèse n'est pas appropriée lorsqu'il s'agit d'observations répétées chez un même cluster, dans la mesure où ces observations tendent à être corrélées entre elles. Si cette corrélation existe, celle-ci ne peut être ignorée. Le coefficient de corrélation intra-cluster permet de mesurer la part de la variance due à la variabilité inter-clusters c'est à dire le degré de similarité (homogénéité) intra-cluster :

$$\rho = \frac{\sigma_{\lambda}^2}{\sigma_{\lambda}^2 + \sigma_{\epsilon}^2} \quad (1)$$

avec σ_{λ}^2 variance inter-cluster et σ_{ϵ}^2 variance intra-cluster. Cette valeur ρ , qui fait l'objet de nombreuses publications, a retenu toute notre attention au cours des travaux. ρ étant quasi systématiquement fixé *a priori* selon la taille des clusters, nous recommandons d'effectuer une estimation de ρ (par bras) *a posteriori* lors de chaque ERC dans le but de planifier les futures études et d'en faire une meilleure interprétation. Différents résultats permettant de mesurer la perte de puissance statistique engendrée par un mauvais choix de ρ (*a priori*) selon le nombre de clusters impliqués dans l'essai ont ainsi été obtenus et seront présentés.

II.3. Calcul du nombre de sujets nécessaires (NSN).

Les résultats observés chez deux sujets d'un même cluster auront tendance à être plus similaires que si les résultats étaient associés à deux sujets provenant de deux clusters distincts. Le calcul du NSN devant prendre en compte cette

corrélation intra-cluster, il en résulte une augmentation de la taille des échantillons. En pratique, on calcule la taille d'échantillon nécessaire en utilisant les formules classiques pour des données indépendantes puis on applique un facteur de correction appelé facteur d'inflation, défini par $IF = 1 + (m-1)\rho$ avec m la taille (fixe) des clusters. Ainsi, si on suppose que chaque bras expérimental contient le même nombre de clusters N et si on veut détecter une différence $\Delta = \mu_1 - \mu_2$ entre deux moyennes μ_1 et μ_2 de même écart type σ , il faut inclure M sujets dans chaque bras d'intervention pour un risque alpha bilatéral et une puissance de $1-\beta$:

$$M = mN = \frac{2\sigma^2(z_{\alpha/2} - z_{1-\beta})^2}{\Delta^2} [1 + (m-1)\rho] \quad (2)$$

On peut alors montrer que plus le nombre de clusters par bras est faible, plus il est nécessaire d'avoir un ICC *a priori* élevé pour assurer une puissance théorique élevée. La formule classique présentée précédemment ne permet pas de considérer la variation des tailles des clusters impliqués dans un essai mais se réduit au fait que tous les clusters sont de taille unique m . Le cas le plus fréquent restant tout de même celui où les clusters n'ont pas tous le même nombre d'observations, certains auteurs [3] ont proposé récemment des ajustements du facteur d'inflation :

$$IF = 1 + ((1 + cv^2 \frac{N-1}{N})m - 1)\rho \quad (6)$$

où N correspond au nombre de clusters impliqués dans chaque bras de l'ERC et cv est un coefficient associé à la variation de taille des clusters défini par le rapport de l'écart-type de la taille des clusters par la moyenne de la taille des clusters. Connaissant cette taille moyenne \bar{m} des clusters (écoles, hôpitaux, régions administratives), on peut reprendre la valeur du coefficient de variation cv utilisée dans des ERC similaires comme on peut le faire lorsqu'on fixe *a priori* la valeur de l'ICC sur les bases d'études antérieures. Une seconde issue est possible à partir du moment où l'on connaît la taille de tous les clusters avant leur inclusion. Or ce n'est pas le cas dans la majorité des ERC d'où l'intérêt de refaire le calcul du NSN *a posteriori* et d'analyser par exemple l'évolution de la puissance. Dans le cadre de cette thèse, nous proposons donc de mener ce type de travail en comparant, de plus, les méthodes les plus récentes d'ajustement du facteur d'inflation (dont une que nous proposons) ; l'intérêt final étant de mesurer l'incidence par le biais de simulations que peut avoir le choix de telle ou telle méthode sur les résultats de l'analyse statistique. Comme, à ce jour, aucune des techniques présentées dans cet article ne sont implémentées dans des logiciels statistiques, nous les avons programmées sous forme de macros sous le logiciel R.

III. ANALYSE STATISTIQUE

L'analyse statistique de données des ERC doit tenir compte de la corrélation induite par le dispositif expérimental au risque d'obtenir des degrés de signification sous-estimés et des interprétations erronées. Ceci explique que stratégie expérimentale et analyse statistique soient intimement liées. Une des questions récurrente de cette thèse consiste à se demander si les différentes approches d'estimation statistique sont sensibles à la variation de taille d'échantillon intra et inter cluster.

III.1. Unité d'analyse : le cluster.

Dans un premier temps, nous considérons que l'unité de randomisation (le cluster) correspond à l'unité d'analyse. Dans ce cas, la perte d'information peut s'avérer importante du fait que pour chaque variable, le cluster est représenté par la moyenne (voire la médiane) des valeurs associées aux individus appartenant à ce cluster. Par conséquent, le problème de la corrélation intra-cluster est annihilé puisque l'individu statistique considéré n'est plus le sujet mais le cluster; nous pouvons dès lors utiliser les méthodes statistiques classiques. De part nos travaux, nous sommes arrivés à la conclusion que ce type d'analyse entraîne une perte de puissance statistique relativement importante mais recommandons, néanmoins, de maintenir cette première analyse, à nos yeux primordiale, essentiellement pour deux raisons. Tout d'abord, l'unité de randomisation étant le cluster, de nombreuses covariables sont associées au cluster (entre autres le traitement) et non à l'individu d'où l'intérêt d'une analyse même simplement descriptive au niveau cluster. Par ailleurs, comme nous l'avons déjà noté, la cible de l'intervention peut naturellement être le cluster.

III.2. Unité d'analyse : l'individu dans son cluster.

L'autre voie que l'on peut emprunter lors de l'analyse des ERC revient à considérer l'individu comme unité d'analyse. A partir de là, les questions méthodologiques sont nombreuses selon que l'on s'intéresse à l'analyse statistique descriptive ou multivariée des données.

Allan Donner [4] a montré que les résultats obtenus avec les tests statistiques utilisés généralement (Chi2, Student, etc.) ne correspondent pas à la réalité puisqu'à aucun moment la corrélation intra-cluster n'est prise en compte dans ces méthodes. Toutes n'ont pas été adaptées à l'analyse des ERC comme le test du Chi2 et aucune d'entre elles n'est implémentée dans les logiciels statistiques classiques. Comme pour le calcul du NSN présenté précédemment, nous avons mis en place des outils informatiques (macros R) permettant de réaliser de manière systématique les tests du Chi2 et de Student ajustés, quels que soit la taille et le nombre de clusters considérées.

Lors de l'analyse multivariée, il faut bien veiller à prendre en compte l'éventuelle corrélation intra-cluster. L'unité statistique à considérer ici étant le sujet dans son

cluster, l'analyse statistique se fait alors par le biais de modèles singuliers. Nous proposons de comparer ici les deux familles de modèles statistiques et leurs méthodes d'estimation utilisées dans la comparaison de proportions au niveau individuel : modèles marginaux et les modèles mixtes. Les objectifs, les hypothèses sur les données, les méthodes d'estimation et l'interprétation des résultats sont propres à chacun de ces deux modèles. Pour le modèle marginal appliqué à des données randomisées en clusters, l'objectif est de prédire la prévalence moyenne selon les variables explicatives du modèle. On parle alors de "Population Averaged". L'estimation des paramètres du modèle se fait alors par la méthode d'équations d'estimation généralisées (GEE) [5]. A contrario, l'utilisation de modèles mixtes est liée à la notion "Cluster Specific" et permet d'étudier les facteurs de risque individuels dans une perspective étiologique. Ces modèles consistent en une extension des modèles de régression appliqués à des données structurées de manière hiérarchique par l'introduction d'un effet dit aléatoire (l'effet cluster); de façon à modéliser les différences, non seulement, entre les micro-unités (les sujets), mais aussi entre les macro-unités (les clusters). Les algorithmes d'estimation des paramètres sont assez lourds voire peu performants. Un de nos objectifs de recherche est de comparer la robustesse et la qualité de convergence de ces différentes méthodes d'estimation face à des ERC atypiques (essais avec peu de clusters ou essais avec de nombreux clusters de petite taille). On peut penser que ces différentes approches, basées sur des approximations à différents niveaux, ne réagiront pas de la même façon à la variabilité inter et intra cluster ainsi qu'à la taille et au nombre de clusters. Parmi les approches étudiées : techniques d'intégration numérique telles les quadratures de Gauss-Hermite ou adaptative consistant à maximiser la vraisemblance marginale, algorithme de type MC-EM où les quantités calculées usuellement à l'étape E sont approximées stochastiquement par l'algorithme de Métropolis, méthodes dites de linéarisation basées sur la linéarisation de la fonction de lien telles les méthodes de Schall, GAR (Gilmour, Anderson et Rae) et approche fondée sur la quasi-vraisemblance

Dans l'essai MG Tabac LR, les patients ont été suivis pendant une année au cours de visites régulières sous la forme de 9 consultations et de dossiers à renvoyer à leurs MG. Ainsi, nous possédons pour chaque sujet des mesures répétées. Dans un premier temps, nous avons mesuré, avec les modèles décrits précédemment, l'effet de l'intervention (à savoir la formation de certains MG) ainsi que le taux de sevrage dans chaque groupe à 1 an. Nous n'avons donc utilisé que les questionnaires au temps 0 et à un an. Néanmoins, la figure 1 montre très nettement que le pourcentage de fumeurs du groupe 1 (MG formés) est inférieur à celui du groupe 2 (MG témoins) tout au long de l'intervalle considéré. Pour mener à bien une analyse plus poussée et tenter d'expliquer au mieux la dynamique de l'arrêt du tabac, nous considérons un modèle mixte

relativement plus complexe que ceux étudiés jusqu'à présent puisque nous prenons en compte deux effets aléatoires : un effet aléatoire λ associé au cluster et un autre ψ lié à l'individu.

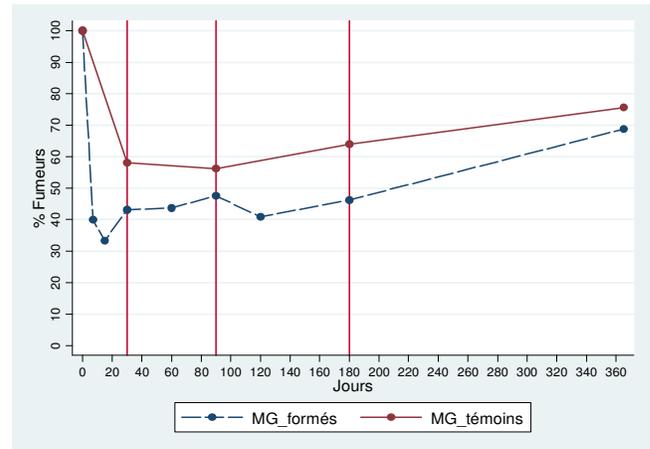


Fig. 1. Evolution de la dynamique du sevrage

IV. CONCLUSION.

Essais utilisés de plus en plus régulièrement pour leur côté pratique lors d'évaluation d'interventions ou de pratiques médicales, les ERC nécessitent une attention toute particulière dans leur mise en place et leur analyse statistique. Une structure de groupe apparaissant à l'intérieur de l'échantillon, il est primordial de prendre en compte la corrélation entre les mesures. Les travaux présentés dans ce papier permettent à notre sens de répondre aux carences bioinformatiques posées par les ERC. De plus, les deux travaux phares de cette thèse que sont l'analyse d'un essai randomisé en clusters avec mesures répétées et la mise en relation de la qualité de l'analyse statistique selon la méthode du calcul de NSN employée, permettent de répondre à des questions épidémiologiques précises telles l'explication de la dynamique du sevrage tabagique ou l'amélioration de la prévention des connaissances et des comportements à risque.

REFERENCES.

- [1] A. Donner, N. Klar, «Design and Analysis of Cluster Randomization Trials in Health Research», Arnold Publishers, 2000
- [2] P. Hu, M. Zelen, «Planning of randomized early detection trials», dans Stat Methods Med Res, 2004, 13, 491-506
- [3] S. Eldridge, D. Ashby, S. Kerry, «Sample size for cluster randomized trials: effect of coefficient of variation of cluster size and analysis method», dans Int J Epidemiol, 2006, 35(5), 1292-300
- [4] A. Donner, «Statistical methods in ophthalmology: an adjusted chi-square approach», dans Biometrics, 1989, 45(2), 605-11
- [5] S. Zeger, K. Liang, «Longitudinal data analysis for discrete and continuous outcomes», dans Biometrics, 1986, 42, 121-130

Un modèle de survie relative appliqué aux données censurées dans le cas de la transplantation rénale.

Vanessa ROUSSEAU¹ (Doctorant 3e année), Jean-Pierre DAURÈS² (Directeur de Thèse)

Laboratoire de Biostatistique et Epidémiologie - Institut Universitaire de Recherche clinique
641, av du Doyen Gaston Giraud, 34093 Montpellier cedex 05, France

¹vanessa.rousseau@inserm.fr, ²daures@inserm.fr

Résumé— Dans de nombreuses études, la survie est le critère principal d'évaluation thérapeutique. Analyser les données de survie c'est s'intéresser à l'apparition d'un évènement au cours du temps, le décès ou la récurrence d'une maladie par exemple. Ceci implique de tenir compte des particularités des données de survie, la censure entre autre. Comme dans toute analyse statistique, nous voulons décrire les observations, donc fournir les estimations de paramètres pertinents, comparer la survie de plusieurs groupes de sujets, expliquer et prédire la durée de survie en fonction de certains facteurs. Dans cet article, nous nous intéressons particulièrement aux notions de survie relative et aux problèmes engendrés par les données censurées. La survie relative permet d'ajuster la survie d'une population spécifique par rapport à la survie d'une population de référence. Les modèles de survie relative sont soit de type additif soit de type multiplicatif. Dans le cas additif, le taux de mortalité relative est la différence entre le taux de mortalité spécifique d'une pathologie et le taux de mortalité de la population de référence. Dans le cas multiplicatif, c'est le ratio. Cependant, la plupart des analyses de survie font généralement l'hypothèse d'indépendance entre le délai de survenue de l'évènement et le délai de survenue de la censure. Dans certaines situations telle que les censures non planifiées ou non administratives, cette hypothèse n'est pas valable. Etant en épidémiologie étiologique, nous nous intéressons plus particulièrement aux modèles de survie relative multiplicatifs et proposons de considérer la censure comme informative dans un de ces modèles particulier. La théorie proposée est appliquée à l'étude de la survie des patients transplantés rénaux, les patients censurés étant ceux qui retournent en dialyse. Dans ce contexte, la censure semble donc informative.

Mots-clés— Survie relative, Censure informative, Modèle de Cox, Risques compétitifs, Transplantation rénale.

I. INTRODUCTION

La survie relative se définit comme le ratio entre la survie d'une population spécifique et la survie d'une population de référence. La survie d'une population spécifique correspond à la survie observée d'un groupe de patients étudiés. La survie d'une population de référence peut être, par exemple, la survie de la population nationale française. Chaque individu de la population observée peut être comparé à un individu ayant les mêmes caractéristiques que lui (sexe, âge, etc...) dans la population de référence. Ainsi, la survie relative permet de prendre en compte la survie de la population générale dans l'étude de la survie d'une cohorte spécifique. L'intérêt de cette méthode est qu'elle donne une mesure de survie nette, c'est-à-dire l'excès de mortalité due à la

maladie étudiée. Pour pouvoir estimer la survie nette, la cause spécifique du décès étudié est supposée indépendante des autres causes de décès, mais elle n'est pas connue (registre pas assez précis). Classiquement, lorsque les causes de décès sont connues, il est possible d'étudier la mortalité spécifique en considérant soit un modèle de survie où les décès indépendants de la maladie sont censurés [3], soit un modèle à risques compétitifs où les décès peuvent être distingués selon leur cause [4]. Un problème majeur est de pouvoir mesurer la part des décès imputables à la maladie étudiée. Dans le cadre de la transplantation rénale par exemple, les décès liés à un cancer peuvent être d'origine indépendante de la transplantation ou liés aux traitements immunosuppresseurs. Ces deux causes peuvent même être imbriquées. Le principal avantage de la survie relative est qu'elle permet d'estimer la survie spécifique à la maladie étudiée sans nécessité de connaître l'information sur la cause de décès.

Il existe essentiellement deux sortes de modèle de survie relative : les modèles additifs et les multiplicatifs. Breslow et al. [2] ont étudié un modèle multiplicatif qui suppose que le risque de décès d'un individu est le produit du risque de décès pour un individu du même sexe et même âge de la population de référence et une fonction des covariables. Pocock et al [6] ont analysé un modèle additif ayant une signification similaire sauf qu'il s'agit de la différence entre les deux risques. Dans les deux approches, les mortalités de la population sont considérées comme des quantités connues et ces deux modèles sont de type paramétrique. Notre objectif est de rechercher les facteurs explicatifs d'un décès après une greffe de rein, nous avons choisi de nous concentrer sur les modèles multiplicatifs. Nous nous sommes donc particulièrement intéressés aux travaux d'Andersen et al. [1] qui ont introduit une généralisation du modèle de régression multiplicatif suivant des spécifications de type Cox semi-paramétrique [3]. Cependant, ces modèles, comme la plupart des modèles de survie, font l'hypothèse d'indépendance entre le délai de survenue de l'évènement et le délai de survenue de la censure. La censure est alors non-informative. Cette hypothèse est respectée dans le cas de censure planifiée. En revanche, elle peut être discutable dans le cas de censure non-planifiée. On peut supposer dans ces situations que les

censures sont informatives, c'est-à-dire le fait que la censure se soit produite apporte une information sur le délai de survenue de l'évènement considéré.

Dans cet article, nous proposons donc de prendre en compte la censure informative dans les modèles de survie relative multiplicatifs. Nous considérons seulement la censure à droite. Nous supposons qu'il y a deux types de censure, appelées perdus de vue et administrative. Les sujets perdus de vue sont considérés comme informatifs. En effet, les perdus de vue sont les patients dont on ne connaît pas l'état à la date de point. Ils représentent donc une perte d'information et sont source de biais. Dans le cas des patients atteints du VIH par exemple, le traitement est souvent lourd avec des effets secondaires multiples et provoque ainsi des arrêts de suivi. De plus, il a été observé que ce sont les patients les plus malades qui arrêtent leur traitement. Dans ce genre de cas, la censure devient dépendante de l'évènement puisqu'elle apporte une information sur l'évènement étudié. Par contre, la censure administrative est seulement engendrée par la fin de l'étude. Elle est donc considérée comme non-informative.

Le reste du document est organisé de la manière suivante. Dans la deuxième section, la théorie du modèle proposé est introduite. Dans la troisième section, le modèle est appliqué à une cohorte française de patients transplantés rénaux (DIVAT) depuis 1990. La section 4 comporte certains points de conclusion et de discussion.

II. MODÈLE

Soient n individus observés et indexés par i , $i = 1, \dots, n$. Soit E , l'ensemble des états pouvant être observés pour chaque individu. Ces états sont :

- L'état initial, ie l'état dans lequel les individus entrent dans l'étude. Tous les individus entrent dans ce même état noté "Etat 0". Il peut correspondre, par exemple, au diagnostic d'une maladie. Par ailleurs, un individu peut rester dans cet état jusqu'à la date de point (date de l'extraction de la base ou date de la fin de l'étude par exemple).
- Un état absorbant ie une fois que l'individu entre dans cet état, il ne peut en sortir. Il s'agit souvent du décès. C'est pourquoi nous allons le noter "Etat D".
- Un état inconnu que l'on notera "Etat I". Nous l'appellons ainsi car une fois qu'un individu entre dans cet état, nous ne savons pas ce qu'il advient. De plus, à partir de cet état, nous n'observons jamais de passage vers d'autres états. C'est pour cela que nous incluons cet état dans notre schéma, il s'agit d'une censure originale, celle des perdus de vue. En effet, cette censure semblerait apporter de l'information au décès.

Les états sont récapitulés dans la figure 1. Pour chaque individu, nous n'observons qu'une seule transition (ou aucune). L'état à l'inclusion dans l'étude est l'état 0. Chaque individu transite soit de l'état 0 à l'état D soit de l'état 0 à l'état I. L'individu peut aussi ne jamais passer vers un des deux états et donc rester dans l'état initial. Pour tous les individus, la transition de l'état I à l'état D ne sera jamais observée. Par ailleurs, nous émettons l'hypothèse assez forte que les probabilités de transition d'un état vers un autre ne dépendent pas du temps.

Les deux fonctions de risque (décès et censure informa-

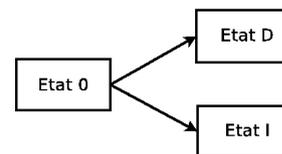


Fig. 1. Modèle à trois états transitoires.

tive) pour un individu i , $i = 1, \dots, n$, s'écrivent donc sous la forme de modèle de régression :

$$\lambda^I(t_i|Z_i^I) = \lambda_0^I(t_i) \exp(\beta^I Z_i^I) \quad (1)$$

$$\lambda^D(t_i|Z_i^D) = \lambda_0^D(t_i) \lambda_*^D(t_i + a_{i0}) \exp(\beta^D Z_i^D) \quad (2)$$

avec a_{i0} qui représente l'âge de l'individu i à l'inclusion dans l'étude. $\lambda_*^D(t_i + a_{i0})$ est la fonction de risque de décès de la population générale de même sexe, même âge $t_i + a_{i0}$ et né la même année que l'individu i observé. Dans l'analyse, les $\lambda_*^D(t_i + a_{i0})$ seront considérées comme des quantités connues (obtenues grâce aux tables vitales). λ_0^D est la fonction de risque relatif de base spécifique à la maladie étudiée. λ_0^I est la fonction de risque de base spécifique à la censure informative étudiée. Z_i^I et Z_i^D sont les vecteurs de covariables associés à la censure et au décès, respectivement. β^I et β^D sont les vecteurs de coefficients de régression inconnus associés à Z_i^I et Z_i^D , respectivement. Le lien entre la censure et le décès est modélisé au travers des covariables. Nous pourrions conclure à une forte dépendance entre la censure et le décès si, après estimation des paramètres du modèle, nous avons exactement les mêmes covariables significatives dans les deux risques. Si quelques covariables significatives sont communes, nous pourrions alors déduire que les perdus de vue apporte de l'information au risque de décéder. Par contre, si aucune covariable significative est identique pour les deux risques, nous pourrions conclure qu'il n'y certainement pas de lien entre la censure et le décès.

Pour estimer les paramètres de ce modèle, nous utilisons la méthode du maximum de vraisemblance. La logvraisemblance est maximisée en utilisant la fonction *optim* du logiciel R. Cette fonction utilise l'algorithme de *quasi-Newton* et permet d'obtenir la matrice hessienne.

III. APPLICATION

A. Présentation de la base de données

La cohorte française DIVAT (Données Informatiques VALidées en Transplantations) concerne 2864 patients transplantés de rein parmi les centres de Nancy, Nantes, Paris-Necker et Toulouse. Les transplantations sont considérées depuis l'année 1990. Les principales caractéristiques renseignées sur les patients sont fournies dans le tableau I. Le codage de chaque covariable nécessaire pour les futures analyses est défini par les médecins.

La date d'inclusion d'un patient dans la cohorte correspond à sa date de première transplantation. La cohorte DIVAT contient des patients transplantés à plusieurs reprises mais notre intérêt ne porte que sur les premières transplantations. Les transplantés présentant une dysfonction chronique du greffon sont censurés à leur date de retour en dialyse, ces patients n'étant plus suivis par la suite. Les

Codage	Nom
1 si ≥ 55 ans, 0 sinon	Date de naissance
1 si femme, 0 sinon	Sexe
1 si ≥ 1996 , 0 sinon	Date de transplantation
jj/mm/aaaa	Date de retour en dialyse
jj/mm/aaaa	Date de décès
1 si ≥ 6 jours, 0 sinon	Reprise de fonction du greffon en jours (dgf)
1 si ≥ 4 , 0 sinon	Incompatibilité entre les antigènes d'histocompatibilité du donneur et ceux du receveur
1,2 ou 3	Traitement immunosuppresseur d'induction au moment de la transplantation
1 si maladie à récurrence, 0 sinon	Type de maladie initiale ayant conduit à l'insuffisance rénale terminale
1 si oui, 0 sinon	Antécédents cardiologiques du receveur
1 si cadavre, 0 sinon	Relation entre le donneur et le receveur et le type de greffe (donneur cadavre ou vivant)
1 si ≥ 24 heures, 0 sinon	Ischémie : temps de conservation du greffon entre le clampage et le déclampage
1 si oui, 0 sinon	Néoplasie : antécédent de cancer du patient avant la greffe
1 si ≥ 55 ans, 0 sinon	Age du donneur
1 si femme, 0 sinon	Sexe du donneur
1, 2, 3 ou 4	Centre de transplantation

TABLE I

CARACTÉRISTIQUES DES PATIENTS TRANSPLANTÉS ET LEUR CODAGE.

patients qui sont encore en vie à la date de l'étude sont censurés à cette date. L'intérêt de l'étude est de mesurer l'effet des différentes covariables sur la mortalité spécifique des transplantés avec un rein fonctionnel. Le temps d'intérêt est le délai entre la greffe et le décès.

B. Modélisation

L'état initial de chaque individu est "greffé en bonne santé". A un certain temps, soit les individus retournent en dialyse et passent donc dans l'état que l'on a appelé l'état I, c'est-à-dire que l'on ne connaît pas leur devenir après cette date de retour en dialyse, soit l'individu décède et passe donc dans l'état D (figure 2).

Nous adoptons une stratégie de sélection des covariables

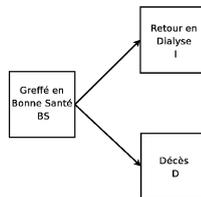


Fig. 2. Schéma des trois états possibles pour chaque patient.

descendante. Chaque covariable est insérée dans les deux types de risque (donc deux fois). Ensuite, les covariables sont supprimées une à une du modèle selon une stratégie de sélection décroissante avec un seuil de signification à 5%. Le test de Wald nous permettra d'estimer l'erreur de première espèce.

C. Résultats

Pour le risque de décès, seules les covariables sexe du receveur, âge du receveur, dgf et année de la greffe sont significatives. Pour le retour en dialyse, ce sont les covariables âge du donneur, sexe du donneur, incompatibilité, anti-hla et année de la greffe qui sont significatives. Il semblerait donc que la censure soit informative car la covariable année de la greffe est significative pour les deux risques. Mais, on ne peut pas conclure à un lien très fort entre le retour en dialyse et le décès car il y a qu'une seule covariable commune à ces deux risques. En conclusion, d'après ce modèle, les femmes, les transplantés de moins de 55 ans, les transplantés dont le greffon reprend fonction plus de 6 jours après la greffe et les patients qui ont été transplantés

avant 1996 ont plus de chance de mourir après une greffe que les hommes, les moins de 55 ans, les transplantés dont le greffon reprend fonction avant 6 jours et ceux qui ont été transplantés après 1996, respectivement.

IV. CONCLUSION

Le modèle de survie relative présenté dans cet article diffère de ceux déjà connus dans la littérature. En effet, c'est un modèle à trois états transitoires qui permet de prendre en compte la censure informative et les taux de mortalité de la population générale. Ce modèle paramétrique qui, de plus, est multiplicatif est facilement interprétable. Il permet, entre autre, de connaître les facteurs explicatifs d'un décès après une greffe en considérant le fait que les greffés peuvent retourner en dialyse et qu'alors ils ne sont plus suivis. Ce modèle suppose que les perdus de vue apporte de l'information sur le décès.

Par ailleurs, les taux de mortalité de la population générale sont inclus dans le risque de décès de la population observée. Pour chaque individu observé, on prend en compte son taux de mortalité de la population générale à travers un individu qui a les mêmes caractéristiques que lui (âge, sexe et autres facteurs potentiels). Ceci signifie que l'on prend en compte tous les autres risques de décéder que celui de décéder après une greffe. Le modèle introduit à la fois les caractéristiques individuelles de la cohorte étudiée et les changements démographiques de la population de référence. Ceci peut-être très intéressant dans le cas de cohorte suivie sur de longues périodes.

Par ailleurs, nous supposons qu'il y a un lien entre le retour en dialyse et le décès, ceci par l'intermédiaire des covariables. Dans notre exemple, il semble que les perdus de vue apporte de l'information au décès. Un point négatif de ce modèle est que nous ne pouvons pas quantifier ce lien. Il serait donc intéressant de modéliser le lien entre la censure et le décès.

RÉFÉRENCES

- [1] Andersen, P.K., & Borch-Johnsen, K., & Deckert, T., & Green, A., & Hougaard, P., & Keiding, N., & Kreiner, S., A Cox regression model for the relative mortality and its application to diabetes mellitus survival data 1985, *Biometrics*, 41, 921-932.
- [2] Breslow, N.E., Lubin, J.H., Marek, P., & Langholz, B., Multiplicative models and cohort analysis, 1983, *Journal of the American Statistical Association*, 78, 1-12.
- [3] Cox, D.R., Regression models and life tables (with discussion), 1972, *Journal of the Royal Statistical Society : Series B*, 34, 187-220.
- [4] Gjertson, D.W., Dabrowska, D.M., Cui, X., & Cecka, J.M., Four Causes of Cadaveric Kidney Transplant Failure : A Competing Risk Analysis, 2002, *American Journal of Transplantation*, 2, 84-93.
- [5] Huang, X., & Wolfe, R.A., frailty model for informative censoring,
- [6] Pocock, S.J., & Gore, S.M., & Kerr, G.R., Long-term survival analysis : The curability of breast cancer, 1982, *Statistics in Medicine*, 1, 93-104.

Etude microscopique d'un matériau pour l'électronique rapide : l'InGaAs

Giulio SABATINI¹ (Doctorant 2e année), Luca VARANI² (Directeur de Thèse)

Institut d'Electronique du Sud UMR 5214
 Université Montpellier II, Place Eugène Bataillon, 34095 Montpellier, France

¹sabatini@ies.univ-montp2.fr, ²varani@ies.univ-montp2.fr

Résumé— La réalisation de composants microélectroniques nécessite une connaissance approfondie des matériaux utilisés. Nous étudions donc les propriétés de transport électronique dans l' $\text{In}_{0.53}\text{Ga}_{0.47}\text{As}$, aujourd'hui l'un des matériaux III-V les plus utilisés dans le domaine de la microélectronique à hautes fréquences. Les valeurs de la vitesse de dérive, de l'énergie, de la masse effective moyenne et de la mobilité en fonction du champ électrique ont été calculées en utilisant un simulateur Monte Carlo.

Mots-clés— InGaAs, matériaux à haute mobilité, transport électronique, Monte Carlo.

I. INTRODUCTION

La réalisation des nouveaux dispositifs pouvant fonctionner à très grande fréquence exige des matériaux semi-conducteurs avec d'excellentes propriétés de mobilité et de transport. Dans ce domaine, l'InGaAs montre un grand potentiel dû aux vitesses de seuil élevées des électrons. Le but principal de cet article est donc d'approfondir la connaissance de l'un des matériaux III-V les plus utilisés : l' $\text{In}_{0.53}\text{Ga}_{0.47}\text{As}$, pour pouvoir guider la conception de nouveaux dispositifs microélectroniques et aider l'optimisation de leurs performances.

Aujourd'hui la nécessité d'augmenter la fréquence de fonctionnement des dispositifs électroniques et leur capacité d'intégration mène à une diminution progressive de leurs dimensions. La réduction des dimensions des dispositifs jusqu'à l'échelle nanométrique rend possible la diminution du temps de transit des porteurs à l'intérieur des structures et, par conséquent, une augmentation de la fréquence de fonctionnement par rapport aux composants traditionnels. Quand on arrive à ces dimensions, les longueurs typiques des dispositifs sont très proches de la distance moyenne entre les collisions des électrons dans les matériaux qui les constituent. La réduction des dimensions et l'augmentation de la vitesse de transport, grâce à un matériau à haute mobilité comme l' $\text{In}_{0.53}\text{Ga}_{0.47}\text{As}$, permettent au transport électronique d'avoir un caractère balistique. Des études récentes ont montré que les transistors tels que des HEMTs à canal $\text{In}_{0.53}\text{Ga}_{0.47}\text{As}$ peuvent fonctionner comme des cavités résonantes pour des ondes de plasma électronique. En effet, si les dimensions du dispositif atteignent le domaine nanométrique et si, de ce fait, le régime balistique est atteint, les ondes de

plasma oscillent à des fréquences pouvant atteindre plusieurs TeraHertz (THz), comme prévu théoriquement [1] et comme récemment confirmé expérimentalement [2]. Le délicat choix des matériaux joue donc un rôle important dans la montée en fréquence.

Pour approfondir la connaissance de l' $\text{In}_{0.53}\text{Ga}_{0.47}\text{As}$, nous calculons, grâce à un simulateur Monte Carlo [3] un certain nombre de paramètres de transport microscopique importants dans la modélisation physique des dispositifs et nécessaires au bon fonctionnement des simulateurs macroscopiques. Pour cela, nous commençons en présentant le modèle physique employé dans nos simulations. Ensuite, avant de tirer nos conclusions, nous montrons les résultats obtenus dans le cas de l' $\text{In}_{0.53}\text{Ga}_{0.47}\text{As}$ à une température $T=300\text{K}$ avec une densité d'impuretés ionisées N_D comprise entre 10^{16} et 10^{18} cm^{-3} , et nous comparons les résultats des simulations avec les données expérimentales présentes dans la littérature.

II. MODÈLE PHYSIQUE

Pour étudier les propriétés de transport de l' $\text{In}_{0.53}\text{Ga}_{0.47}\text{As}$ à température ambiante nous employons un simulateur Monte Carlo standard [3] où les propriétés électroniques du matériau massif sont directement liées à la structure de bande et aux mécanismes de collision.

Le modèle ici employé [4] a été précédemment validé par comparaison avec les données expérimentales [5]. Les données d'entrée du simulateur ont été obtenues en interpolant les valeurs des matériaux binaires correspondants [6]. Pour la bande de conduction, nous avons employé un modèle avec trois vallées sphériques non-paraboliques. Les mécanismes de collisions inclus dans notre modèle physique sont : la collision avec les impuretés ionisées, les transitions dues à l'absorption et à l'émission des phonons optiques polaires et non polaires, les collisions avec des phonons acoustiques élastiques, les collisions intervallées et d'alliage et l'ionisation par impact [7].

Cette méthode de modélisation microscopique nous permet de simuler le comportement de chaque électron dans le matériau et donc de calculer avec précision les paramètres nécessaires à la modélisation physique du matériau en question.

III. RÉSULTATS

Nous étudions dans un premier temps les propriétés de l' $\text{In}_{0.53}\text{Ga}_{0.47}\text{As}$ massif à température ambiante par le calcul des paramètres statiques les plus importants.

Nous commençons en calculant la fraction d'électrons présents dans les différentes vallées (Fig.1). À l'équilibre thermique, les électrons se trouvent dans la vallée Γ , c'est à dire au niveau d'énergie minimal. Quand on applique un faible champ électrique (régime ohmique) les porteurs restent majoritairement dans la même vallée et un petit nombre de porteurs gagne assez d'énergie, grâce au champ électrique, pour passer dans les vallées supérieures (L et X). Donc, au fur et à mesure que le champ électrique augmente, le nombre de porteurs dans la vallée Γ diminue et celui dans les vallées à plus haute énergie, L et X , augmente. Nous avons reporté dans la Fig. 1 les résultats des simulations pour la fraction de population électronique dans les vallées Γ , L et X en fonction du champ électrique, avec une densité d'impuretés ionisées $N_D = 10^{16} \text{ cm}^{-3}$, comparés avec les résultats expérimentaux obtenus par Choo [8]. Nous remarquons que le transfert d'électrons dans la vallée L commence approximativement entre 1 et 2 kV/cm et il y a plus d'électrons dans la vallée L que dans la vallée Γ pour des champs électriques supérieurs à 10 kV/cm. Les électrons sont transférés dans la vallée X à partir de 5 kV/cm et la fraction de population dans cette vallée devient non négligeable à partir de champs électriques proches de 30 kV/cm. La fraction de population s'avère pratiquement indépendante de N_D dans la gamme $10^{15} - 10^{18} \text{ cm}^{-3}$.

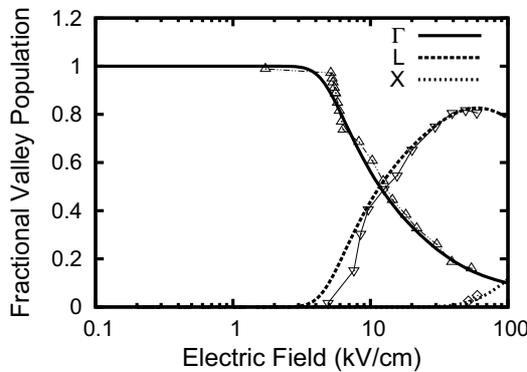


Fig. 1. Fraction de population dans les vallées Γ , L et X en fonction du champ électrique pour $N_D = 10^{16} \text{ cm}^{-3}$. Les lignes représentent notre simulation et les symboles les résultats expérimentaux présents dans la littérature. Les lignes relient simplement les points reportés.

Dans les composés III-V, comme c'est le cas de l' InGaAs , la masse effective moyenne des porteurs \bar{m}_x varie selon la vallée dans laquelle l'électron se trouve. Par la suite, nous avons reporté dans la Fig. 2 la masse moyenne en fonction du champ électrique. Pour de faibles valeurs du champ électrique, tous les électrons sont dans la vallée de Γ et \bar{m}_x correspond aux valeurs de la masse effective au fond de cette vallée. Ensuite \bar{m}_x augmente avec le champ électrique, atteignant la valeur de la masse effective au fond de la vallée L à environ 70 kV/cm. Enfin, pour des champs électriques plus élevés, \bar{m}_x augmente jusqu'à la valeur associée à la vallée X .

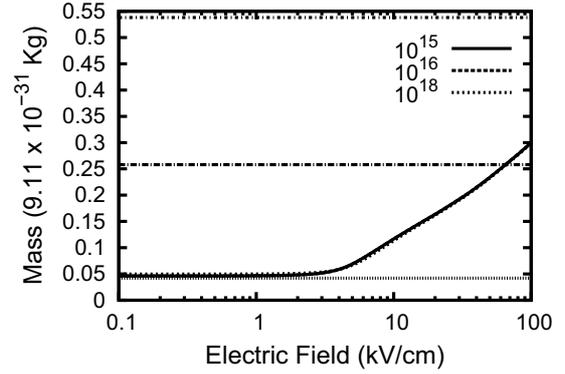


Fig. 2. Masse effective moyenne en fonction du champ électrique, pour les densités d'impuretés ionisées reportées et exprimées en cm^{-3} . Les lignes horizontales représentent les valeurs de la masse effective dans les vallées a) Γ , b) L et c) X .

Un paramètre important pour évaluer la prédisposition d'un matériau III-V à être utilisé comme zone active dans des dispositifs à haute fréquence est la vitesse électronique. Nous représentons dans la Fig. 3 les résultats des simulations pour la vitesse de dérive moyenne en fonction du champ électrique, confrontée avec plusieurs données expérimentales présentes dans la littérature. Quand nous augmentons le champ électrique un certain nombre de porteurs gagnent donc assez d'énergie pour passer dans les vallées supérieures (Fig.1). Quand les porteurs passent de la vallée Γ aux vallées supérieures, leur masse effective moyenne augmente (Fig.2) et donc la vitesse commence à diminuer. Nous remarquons aussi que le début du transfert dans la vallée X est à l'origine d'une petite bosse dans la caractéristique de la vitesse pour un champ électrique autour de 30 kV/cm. Pour des valeurs élevées du champ électrique, tous les porteurs sont transférés dans la vallée X . La vitesse se stabilise alors sur une valeur donnée. Dans la Fig. 3 nous constatons un bon accord de nos simulations avec les résultats expérimentaux. Nous reportons aussi

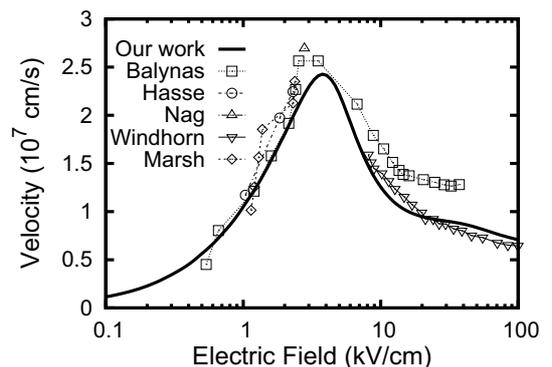


Fig. 3. Vitesse de dérive des électrons en fonction du champ électrique. La ligne continue représente notre simulation et les symboles les résultats expérimentaux présents dans la littérature. Les lignes en traits fins relient simplement les points reportés.

dans la Fig. 4 la vitesse de dérive moyenne en fonction du champ électrique pour différentes valeurs de la densité d'impuretés ionisées. Nous remarquons que la vitesse de pic est située autour de $2.3 \times 10^7 \text{ cm/s}$ pour un champ de seuil de 3.75 kV/cm dans le cas où $N_D = 10^{15}$ et 10^{16} cm^{-3} , par contre il est égal à $1.85 \times 10^7 \text{ cm/s}$ à 4.75 kV/cm

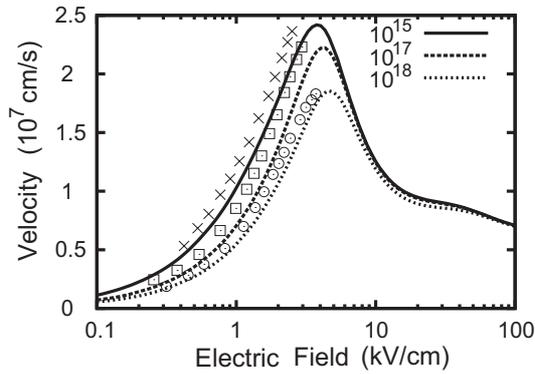


Fig. 4. Vitesse de dérive des électrons en fonction du champ électrique pour les densités d'impuretés ionisées reportées et exprimées en cm^{-3} . Les lignes représentent nos simulations et les symboles les résultats expérimentaux présents dans la littérature.

pour $N_D = 10^{18} \text{ cm}^{-3}$. La vitesse diminue avec l'augmentation du dopage et nous notons aussi une légère augmentation du champ de seuil. Nous remarquons enfin le bon accord avec les résultats expérimentaux obtenus par Haase [9]. En utilisant les résultats de la Fig. 3 nous avons calculé la mobilité différentielle $\mu' = dv/dE$ à fréquence nulle et nous reportons les résultats dans la Fig. 5 en fonction du champ électrique. Nous remarquons que la mobilité oh-

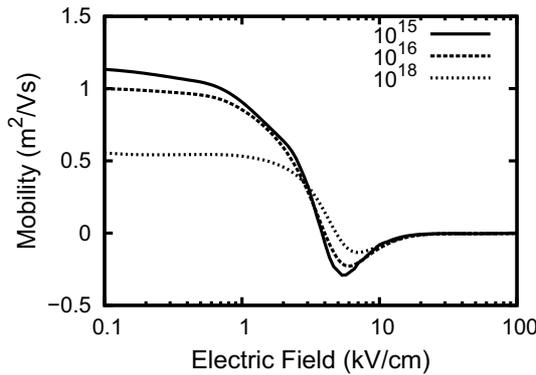


Fig. 5. Mobilité différentielle à fréquence nulle en fonction du champ électrique pour les densités d'impuretés ionisées reportées et exprimées en cm^{-3} .

mique dépend fortement de la densité de donneurs et qu'elle varie de 5 500 à 10 000 et 11 500 cm^2/Vs pour les trois valeurs de N_D considérées dans cet article. Comme nous pouvons le voir, elle montre le comportement typique des semi-conducteurs III-V avec une mobilité différentielle négative après le champ de seuil.

IV. CONCLUSION

Dans le but d'approfondir la connaissance de l' $\text{In}_{0.53}\text{Ga}_{0.47}\text{As}$, le point important de cette étude a consisté en l'extraction des paramètres de transport en régime stationnaire à l'aide de notre simulateur Monte Carlo.

Les valeurs stationnaires de la population, de la masse effective moyenne, de l'énergie, de la vitesse de dérive, ont été de fait calculées en fonction du champ électrique et confrontés avec les données expérimentales présentes dans la littérature. En même temps, la mobilité, paramètre

très important pour évaluer les performances des dispositifs électroniques, a été décrite. Le calcul des paramètres cinétiques nous a permis de mettre en évidence la grande mobilité ohmique de ce matériau, de l'ordre de $1.1 \times 10^4 \text{ cm}^2/\text{Vs}$, qui confirme sa prédisposition à être utilisé dans la réalisation de dispositifs électroniques à vitesses élevées de fonctionnement, comme par exemple dans le cas du canal d'un nanotransistor HEMT.

Ce travail constitue seulement le point de départ d'une étude plus vaste qui vise une compréhension plus approfondie du nanotransistor HEMT, des parties qui le constituent et donc des matériaux utilisés, aussi bien que dans la recherche d'oscillations de plasma et d'instabilités qui devraient faire osciller le transistor aux fréquences THz.

RÉFÉRENCES

- [1] M. Dyakonov and M. Shur. *Plasma wave electronics for terahertz applications*. Terahertz Systems and Sources, Kluwer Academic Publishers, Netherland, 2001.
- [2] W. Knap, J. Lusakowski, T. Parenty, S. Bollaeret, A. Cappy, and M. Shur. Terahertz emission by plasma waves in 60nm gate high electron mobility transistors. *Appl. Phys. Lett.*, 84 :2331-2333, 2004.
- [3] C. Jacoboni and L. Reggiani. The Monte Carlo method for the simulation of charge transport in semiconductors with applications to covalent materials. *Reviews of modern Physics*, 55 :645, 1983.
- [4] B. Garcia Vassallo. *Transporte Electronico en Estructuras de Heterounion InAlAs/InGaAs : HEMTs Dispositivos Balísticos de Dimensiones Nanométricas*. PhD thesis, Universidad de Salamanca, 2005.
- [5] J. Mateos, T. Gonzalez, D. Pardo, V. Hoel, and A. Cappy. Effect of the T-gate on the performance of recessed HEMTs. A Monte Carlo analysis. *Semicond. Sci. Technol.*, 14 :864, 1999.
- [6] S. Adachi. *Physical Properties of III-V Semiconductors InGaAsP*. Wiley, New York, 1992.
- [7] C. Jacoboni and P. Lugli. *The Monte Carlo method for Semiconductor Device Simulation*. Springer-Verlag Wien, Computational Microelectronics, New York, 1990.
- [8] K. Y. Choo and D. S. Ong. Analytical band Monte Carlo simulation of electron impact ionization in $\text{In}_{0.53}\text{Ga}_{0.47}\text{As}$. *J. Appl. Phys.*, 96 :5649, 2004.
- [9] M. A. Haase, V. M. Robins, N. Tabatabaie, and G.E. Stillman. Impact ionization rates for electrons and holes in $\text{Ga}_{0.47}\text{In}_{0.53}\text{As}$. *J. Appl. Phys.*, 56 :2295-2298, 1985.

DEVELOPPEMENT D'UN OUTIL DE PREDICTION D'ALEAS LOGIQUES DUS AUX RADIATIONS DE TYPE IONS LOURDS

VINCENT CORREAS ¹ (DOCTORANT 3E ANNEE), FREDERIC SAIGNE ² (DIRECTEUR DE THESE)

IES – Institut d'Electronique du Sud – Groupe RADIAC
Place Eugène Bataillon, 34095 Montpellier Cedex 05, France

[1 VINCENT.CORREAS@IES.UNIV-MONTP2.FR](mailto:1.VINCENT.CORREAS@IES.UNIV-MONTP2.FR)

[2 FREDERIC.SAIGNE@IES.UNIV-MONTP2.FR](mailto:2.FREDERIC.SAIGNE@IES.UNIV-MONTP2.FR)

Résumé — Avec l'intégration technologique, les circuits intégrés deviennent sensibles à de nouvelles contraintes, la fiabilité devient un enjeu majeur pour les fabricants. Depuis quelques années, la tenue aux radiations des circuits intégrés est une composante indispensable de la fiabilité, et se caractérise notamment par une évaluation sur des mémoires de type SRAM du taux de défaillances fonctionnelles induites suite au passage d'un seul ion lourd. Les fabricants sont donc particulièrement intéressés par le développement d'outils susceptibles de prédire le comportement de leurs technologies sous faisceau d'ions lourds avant même que le composant ne soit fabriqué. L'outil de prédiction PHISco (Prediction of the Heavy Ion Sensitivity code) permet de reproduire les tests sous faisceaux en vue d'établir la section efficace aux ions lourds. Les paramètres d'entrée de cet outil nécessitent deux types de données technologiques : la structure 3D et les modèles électriques.

Mots clés— Single Event Upset (SEU) ; Ion lourd ; SRAM ; Simulateur PHISco ; Modèle de diffusion collection.

I. INTRODUCTION.

Lorsque l'homme s'est lancé dans la conquête de l'espace, il a dû faire face à de nouvelles contraintes. En 1961, une première publication a porté sur l'existence de particules chargées autour de l'atmosphère terrestre [1]. Depuis, de nombreuses recherches sont orientées dans la quantification des spectres de particules ionisantes. En 1962, Wallmark et Marcus [2] annoncent la théorie du basculement logique d'un bit induit par une charge ionisante. En 1975, à bord de Intelsat IV, le premier aléa logique, appelé communément Single Event Upset (SEU) est observé [3]. Cette nouvelle contrainte radiative a pour conséquence l'altération du fonctionnement du composant.

Les fabricants de satellite, ne pouvant pas se permettre de perdre les satellites durant la mission, ont demandé aux fabricants de circuits intégrés de mettre au point des composants susceptibles de tenir aux radiations.

Pour cela, les concepteurs de systèmes électroniques ont recours à des solutions telles que l'implantation de blocs redondants, pour assurer la continuité du fonctionnement, le développement de code correcteur d'erreur pour localiser et rétablir en cas de SEU un changement d'état logique. D'un point de vue plus technologique, les fondeurs ont ajouté des

profils de dopage en profondeur dans les couches épitaxiées permettant de réduire la diffusion des porteurs vers les électrodes. Dans le même objectif, ils ont développé la technologie SOI (Silicon On Insulator) permettant l'isolation de la partie active des transistors vis-à-vis du substrat par l'intermédiaire d'une couche d'isolant.

L'équipe RADIAC du laboratoire IES de l'Université Montpellier 2, a initié en 1996 les premiers travaux sur la problématique SEU [4] afin d'établir une compréhension des phénomènes physiques majeurs conduisant à un changement d'état logique. Une succession d'études sur le sujet a permis d'élaborer un outil de prédiction de la sensibilité des mémoires SRAM aux neutrons et aux alphas (DASIE : Detailed Analysis of the Secondary Ion Effect) [5]. Cet outil permet de prédire avec une très bonne précision les résultats des tests expérimentaux. L'outil présenté dans ce travail est l'évolution dédiée à la prédiction de la section efficace SEU due aux ions lourds.

Dans une première partie, seront décrits les tests expérimentaux réalisés sur une technologie 130nm STMicroelectronics. Dans une seconde partie, les critères d'apparition de SEU et l'outil de prédiction ion lourd PHISco seront développés. La troisième partie présentera la comparaison des simulations avec l'expérience. Un exemple d'utilisation de l'outil sera développé dans la quatrième partie. Pour conclure, les avantages et les possibilités de l'outil seront énumérés.

II. RESULTATS EXPERIMENTAUX.

L'expérience s'est déroulée au Laboratoire National de Brookhaven (BNL) sur un accélérateur de type Tandem Van de Graaff. Les mémoires SRAMs ont été conçues et implantées dans un seul véhicule de test en technologie CMOS 130nm par STMicroelectronics.

Lors des mesures expérimentales, il est compté le nombre de changement d'état logique induit par les particules. La section efficace (σ) représente la probabilité d'apparition d'un SEU ramenée à une surface de sensibilité par bit logique. La relation fait intervenir la fluence et le nombre de bit (Eq. 1).

$$\sigma = \frac{NbreSEU}{Fluence \times NbreBit} \quad (1)$$

Afin de pouvoir estimer le taux d'erreur d'un composant dans un environnement radiatif donné (lié à l'orbite spatiale

du satellite), les résultats sont présentés sous forme de probabilité d'apparition d'un SEU par cm² et par bit, appelée section efficace SEU. La figure 1 représente cette section efficace SEU normalisée en fonction du LET effectif des ions. Le LET représente le transfert d'énergie linéique de l'ion dans la matière (LET = dE/dx).

La notation de LET effectif est utilisée à partir du moment où l'on réalise des tests avec un angle d'incidence θ par rapport à la normale (\perp à la surface) et est déterminée par l'Eq. 2 [6].

$$LET_{eff} = \frac{LET}{\cos\theta} \quad (2)$$

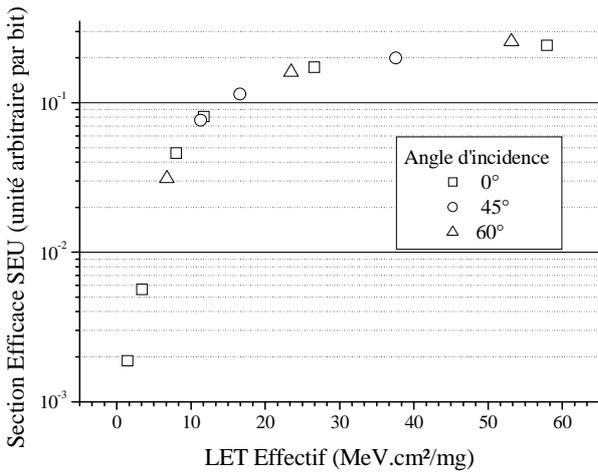


Figure 1. Section efficace SEU en unité arbitraire en fonction du LET effectif en MeV.cm²/mg

III. DESCRIPTION DE L'OUTIL DE SIMULATION.

III.1. Principes fonctionnels du simulateur PHISco (Prediction of the Heavy Ion Sensitivity code)

Une structure SRAM est constituée de quatre transistors. Sa principale fonction est de garder en mémoire un niveau logique ('1' ou '0'). Dans le mode de mémorisation, il y a toujours deux transistors bloqués et les deux autres sont en mode résistif. L'étude porte sur les drains des transistors bloqués démontrés comme étant les zones de plus grande sensibilité pouvant induire un basculement [7].

Pour évaluer la sensibilité des SRAMs, PHISco considère deux cas :

- lorsque le parcours de l'ion ne traverse pas de drain,
- lorsque l'ion impacte un des drains sensibles.

III.1.1. L'ion ne traverse pas de drain

Pour évaluer l'apparition d'un SEU, il faut considérer le processus de création de paires électron/trou, qui par diffusion vont générer un courant parasite entraînant le basculement. A partir des tables de Ziegler [8], on détermine l'énergie déposée par l'ion le long de la trace pour en déduire la densité de paires électron/trou. Dans [9], il a été montré, dans ce cas, que la collection des charges arrivant par diffusion aux électrodes est à l'origine des basculements. L'outil de simulation détermine le courant

parasite arrivant sur chacun des drains étudiés à l'aide d'un modèle de diffusion collection [10]. Le fonctionnement du modèle est schématisé dans la figure 2. Pour chaque ion, le parcours va être segmenté en unité de longueur dl pour prendre en compte la variation du LET dans la matière.

Pour chaque segment, le modèle va être appliqué afin de déterminer le nombre de porteurs collectés par unité de surface du drain (dx x dy) en fonction du temps (Éq. 3). Ce modèle dépend de la technologie à travers deux paramètres d'entrée : la constante de diffusion D et la vitesse de collection des porteurs v à la jonction obtenues à l'aide de simulation composant 3D TCAD [11].

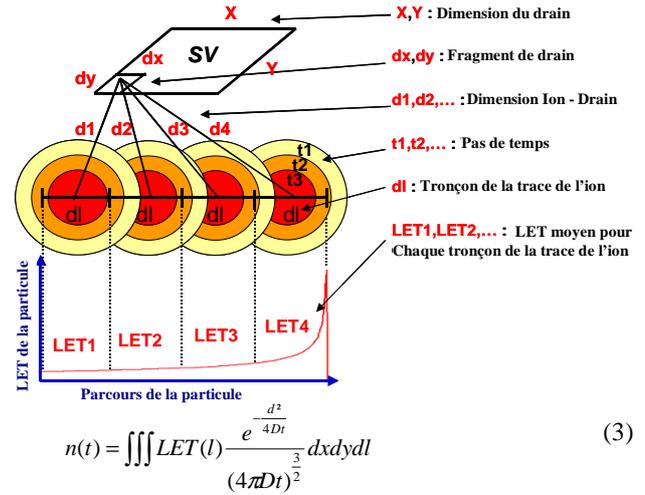


Figure 2. Modèle de diffusion ambipolaire

Le calcul du courant est caractérisé par les valeurs I_{max} et T_{max}, où T_{max} correspond au temps pour lequel est observé le maximum du courant (I_{max}) (Fig.3 A). Une fois obtenu par simulation électrique SPICE [11], le couple de valeurs induisant un basculement, il est reporté sur un abaque I_{max}=f(T_{max}) (Fig.3 B). Pour déterminer l'apparition d'un SEU, on compare les coordonnées max du courant parasite généré par l'impact simulé avec l'abaque. Un SEU est comptabilisé lorsque le couple est au dessus de l'abaque. Cet abaque caractérise l'apparition de SEU pour une technologie spécifique.

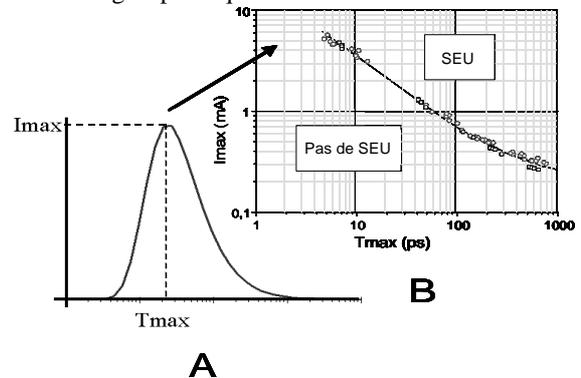


Figure 3 (A). Courant de diffusion collection. (B) Abaque I_{max}=f(T_{max})

III.1.2. L'ion traverse le drain

Lorsque l'ion traverse le drain, le phénomène de « funneling » est observé (Fig. 4). Les charges générées par la particule vont créer une déformation des lignes de champs, qui induiront une collection rapide des porteurs. Le critère de sensibilité SEU va être déterminé par simulation composant 3D TCAD. Des simulations, faisant varier le LET de la trace modélisée, permettent de déterminer un LET seuil de basculement, c'est-à-dire le LET minimal induisant un basculement [12].

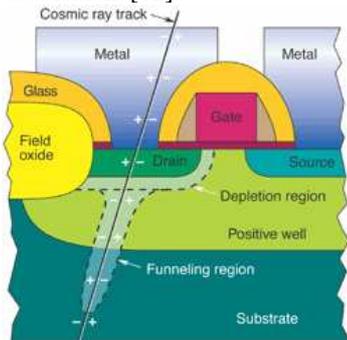


Figure 4. Phénomène de Funneling.

III.1.3. Paramètres d'entrée

Les paramètres d'entrées sont donc relatifs à :

- ◊ le processus de diffusion collection à travers la constante de diffusion D, la vitesse de collection du porteur v à la jonction et les dimensions des drains.
- ◊ la valeur du LET seuil.
- ◊ l'abaque $I_{max}=f(T_{max})$.

III.2. Architecture du simulateur

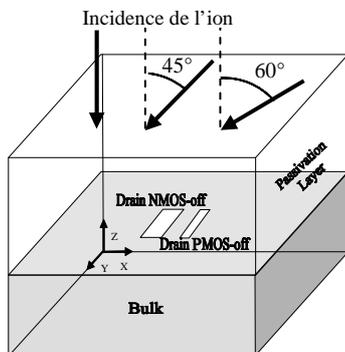


Figure 5. Structure du simulateur.

Dans la figure 5, schématise l'architecture de la structure simulée, dans laquelle est représenté :

- ◊ Les surfaces des drains des transistors NMOS et PMOS bloqués ainsi que la couche passivation (métal et oxydes d'isolation) et le substrat.
- ◊ Le faisceau d'ions se situe au dessus et possède trois directions.

Chaque test expérimental est simulé séparément. Les coordonnées des points d'impacts sont obtenues aléatoirement sur la surface du composant. L'outil calcule la perte d'énergie à travers la couche de passivation pour obtenir le LET à la surface de la zone active, afin d'y être

comparé au LET seuil lorsqu'il l'ion passe par un des drains. Dans le cas contraire, on va appliquer le modèle de diffusion collection dans le substrat.

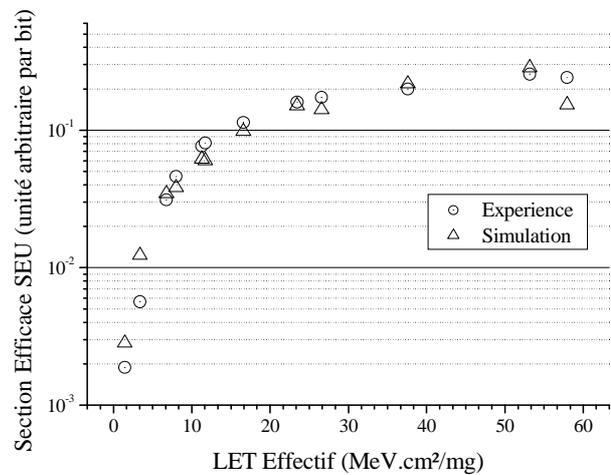


Figure 6. Section efficace SEU expérimentale et simulé en unité arbitraire en fonction du LET effectif en MeV.cm²/mg

Nous avons volontairement réalisé une seule configuration NMOS-PMOS, car l'architecture du point mémoire est symétrique. La distance entre les drains des transistors bloqués est la même pour la mémorisation d'un '1' ou d'un '0' logique.

Dans la figure 6, les sections efficaces expérimentales et simulées sont normalisées et représentées en fonction du LET effectif de l'ion. Les simulations sont en bon accord avec les résultats expérimentaux. L'erreur moyenne est de seulement 20%. Pour chaque ion, la section efficace est obtenue après 10 minutes de simulation sur un processeur Pentium 4 de 3.4GHz.

IV. POSSIBILITES DE L'OUTIL

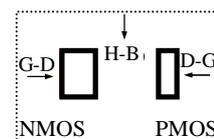


Figure 7. Direction du faisceau (G : Gauche, D : Droite, H : Haut et B : Bas)

L'outil permet en outre de faire une étude sur la direction du faisceau par rapport à la cellule mémoire (Fig. 7). Trois configurations ont été simulées et la variation sur les tests avec angle d'incidence a été reportée dans le Table 1. Ces résultats montrent des possibilités de variation de l'ordre de 10% de la valeur de section efficace suivant la direction du faisceau. On observe également un effet de LET : plus le LET augmente, moins on va observer d'écart suivant la configuration.

L'effet du LET s'explique par le recouvrement ou non de la zone de sensibilité du PMOS par celle du NMOS. Les figures 8 à 10 représentent les surfaces sensibles de chaque transistor, et mettent en évidence l'effet de recouvrement des zones de sensibilités pour une simulation sous faisceau

d'ion Chlore à 60° d'incidence. Dans les figures 8 et 10, le recouvrement n'est pas complet, tandis que pour la figure 9, pour un faisceau d'ion dirigé du PMOS vers le NMOS, le recouvrement est complet et la sensibilité du point mémoire est réduite à la sensibilité du NMOS.

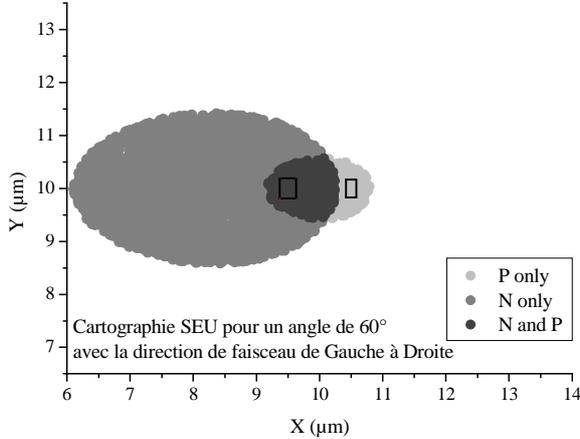


Figure 8. Cartographie des SEU pour un angle de 60° de gauche à droite

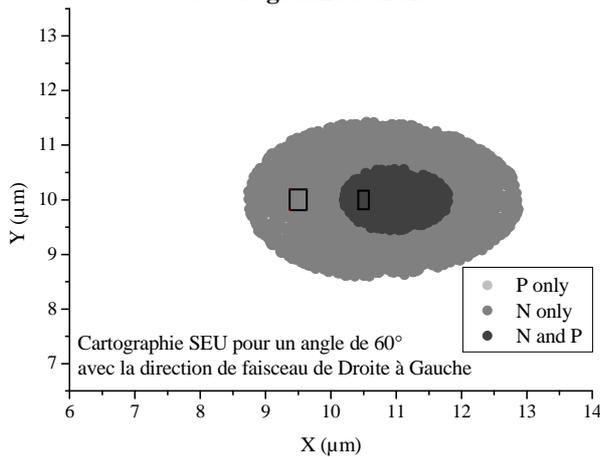


Figure 9 Cartographie des SEU pour un angle de 60° de droite à gauche

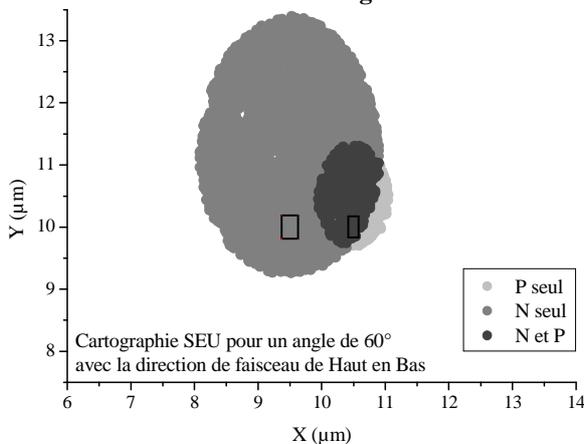


Figure 10. Cartographie des SEU pour un angle de 60° de haut en bas

Ion / LET (MeV.cm ² /mg)	H-B	D-G	G-D	Variation max (%)
F 60°/6.76	3.44E-02	3.09E-02	3.45E-02	10.55
Si 45°/11.3	6.01E-02	5.44E-02	6.17E-02	11.87
Cl 45°/16.59	9.58E-02	8.77E-02	9.80E-02	10.50
Cl 60°/23.46	1.50E-01	1.44E-01	1.51E-01	4.45
Ni 45°/37.59	2.16E-01	2.01E-01	2.18E-01	7.70

Table 1. Section efficace SEU normalisé pour tous les tests avec angle suivant les trois directions.

V. CONCLUSION

Le simulateur PHISco permet :

- ◇ d'étudier la sensibilité de chaque drain,
- ◇ d'évaluer l'impact de la passivation [13],
- ◇ d'obtenir la forme de la zone sensible, information importante dans l'étude des basculements multiples induits par une seule particule,
- ◇ de faire varier la direction du faisceau,
- ◇ de déterminer le paramètre physique qui conditionne la sensibilité, pour pouvoir proposer des pistes de durcissement aux contraintes radiatives.

REFERENCES.

- [1] K.P. Chopra, "Interactions of Rapidly Moving Bodies in Terrestrial Atmosphere," Rev. Mod. Phys. (1960) p. 153.
- [2] J.T. Wallmark et al., "Minimum Size and Maximum Packing Density of Nonredundant Semiconductor Devices," Proc. IRE. (1962) pp. 286-298.
- [3] D. Binder, et al., "Satellite Anomalies from Galactic Cosmic Rays," IEEE Trans. Nucl. Sci., NS-22 (1975) pp. 2675-2680.
- [4] P. Roche, "Etude des Aléas Logiques (SEU) Induit par une Particule Ionisante dans des mémoires SRAM Développées en Technologies CMOS Submicroniques," Thèse. (1999).
- [5] T. Mérelle, "Conception et validation d'un simulateur monte-carlo d'aleas logiques induits par l'environnement radiatif terrestre sur les technologies très largement submicroniques," Thèse. (2005).
- [6] D. Binder, "Analytic SEU rate calculation compared to space data," IEEE. Trans. Nucl. Sci. vol 35 (1988) pp 1570-1572.
- [7] P.E. Dodd, et al., "Impact of technology trends on SEU in CMOS SRAMs," IEEE. Trans. Nucl. Sci. vol 43 (1996) pp 2797-2804.
- [8] J.F. Ziegler, "The transport of ions in matter," IBM-Research.
- [9] J.M. Palau et al., "Device simulation study of the SEU sensitivity of SRAMs to internal ion tracks generated by nuclear reactions," IEEE. Trans. Nucl. Sci. vol 48 (2001) pp 225-231.
- [10] G. Hubert et al., "Detailed analysis of secondary ions effect for the calculation of neutron-induced SER in SRAMs," IEEE. Trans. Nucl. Sci. vol 48 (2001) pp 1953-1959.
- [11] T. Mérelle et al., "Criterion for SEU occurrence in SRAM deduced from circuit and device simulations in case of neutron-induced SER," IEEE. Trans. Nucl. Sci. vol 48 (2001) pp 1148-1155.
- [12] P. Roche et al., "Determination of Key parameters for SEU occurrence using 3D full cell SRAM simulations," IEEE. Trans. Nucl. Sci. vol 46 (1999) pp 1354-1362.
- [13] V. CORREAS et al., "Innovative simulations of heavy ion cross sections in 130nm CMOS SRAM," IEEE. Trans. Nucl. Sci. vol 54 (2007) pp 2413-2418.

LE LASER A CASCADE QUANTIQUE InAs/AlSb EMETTANT ENTRE 3 ET 5 μm

OLIVIER CATHABARD (DOCTORANT 2^o ANNEE), ROLAND TEISSIER (DIRECTEUR DE THESE), ALEXEÏ BARANOV (CO-DIRECTEUR DE THESE).

Institut d'Electronique du Sud
Place Eugène Bataillon, 34095 Montpellier Cedex 05, France

CATHABARD@IES.UNIV-MONTP2.FR
[HTTP://WWW.EDI2S.UNVI-MOTP2.FR](http://WWW.EDI2S.UNVI-MOTP2.FR)

Résumé— L'objectif de ce travail concerne la conception, la réalisation et la caractérisation de diodes lasers à cascade quantique pour les applications du moyen infrarouge dans la plage de longueurs d'onde 3 – 5 μm . Cet article rapporte le principe de fonctionnement et les résultats obtenus sur des structures laser à cascade quantique dans le système de matériau InAs/AlSb. Ceci afin d'obtenir des composants présentant de faibles courants de seuil, des puissances de sortie de l'ordre du Watt et un fonctionnement à température ambiante.

Mots clés— Infrarouge ; Hétérostructure ; Laser à semiconducteur ; Antimoniures.

I. INTRODUCTION.

Les composants optoélectroniques trouvent des applications dans de nombreux domaines très variés comme les télécommunications, l'affichage ou encore le stockage de données.

Les longueurs d'onde utilisées par ces composants diffèrent selon leurs applications. Il n'existe cependant pas de composants efficaces dans la zone infrarouge comprise entre 3 et 5 μm malgré de nombreuses applications (télécommunications, spectroscopie infrarouge...). Une solution envisagée pour combler ce manque est le laser à cascade quantique (LCQ).

Le principe de fonctionnement de ce type de laser est basé sur l'utilisation de transitions optiques entre niveaux confinés de la bande de conduction [1]. Ceci permet d'atteindre des longueurs d'onde plus élevées que les diodes laser utilisant des recombinaisons entre les bandes de valence et de conduction (recombinaisons interbandes).

II. PRINCIPES FONDAMENTAUX DES LASERS A CASCADE QUANTIQUE.

Les diodes lasers à semiconducteurs les plus communes utilisent des recombinaisons électrons-trous à travers le gap. Ce dernier détermine entièrement la longueur d'onde d'émission du laser. Le fonctionnement des LCQ diffère des

diodes lasers sur deux points fondamentaux : l'unipolarité et le schéma en cascade.

Dans les lasers unipolaires, un seul porteur de charge (ici les électrons) participe à l'émission. L'unipolarité est la conséquence de l'utilisation de transitions optiques entre niveaux confinés de la bande de conduction, appelées transitions intersousbandes. La longueur d'onde d'émission sera déterminée grâce au choix des épaisseurs des matériaux constituant les puits quantiques.

Le schéma en cascade permettra le « recyclage » des électrons d'une période à la suivante. Ceci va permettre à tous les électrons de contribuer au gain et à l'émission du laser à chaque période.

III. LA ZONE ACTIVE.

La zone de gain ou zone active d'un LCQ est composée de deux parties, la région active et l'injecteur, formant une période répétée entre 20 et 50 fois.

III.1. La région active.

La région active est composée généralement de trois puits actifs épais (50Å) séparés par des barrières fines (10 Å). Les transitions s'effectuant dans ces puits, ils joueront un rôle primordial dans les caractéristiques optiques du laser.

Leurs tailles et donc la position des niveaux d'énergie vont déterminer la longueur d'onde d'émission. De plus, le gain intersousbande dépend fortement de la durée de vie des électrons sur le niveau haut, noté e3, ainsi que du couplage entre les niveaux haut et bas (e2) de la transition (Fig.1). Ces deux paramètres ayant des évolutions antagonistes, nous pourrions favoriser l'un ou l'autre selon le dessin de ces puits actifs. Un dessin confinant les deux niveaux dans le même puits (transition verticale) permettra d'augmenter le couplage alors qu'un dessin les séparant spatialement (transition diagonale) rallongera la durée de vie des porteurs.

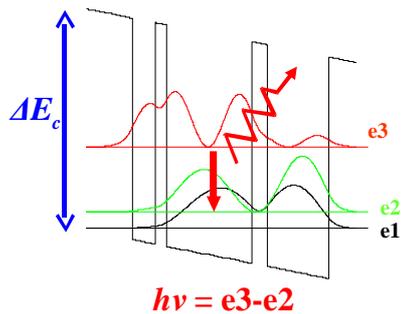


Fig.1 : Schéma des puits actifs d'un LCQ.

III.2. L'injecteur.

L'injecteur a pour fonction de permettre le transport des électrons du niveau e2 vers le niveau e3 de la période suivante (Fig.2).

L'effet laser ne peut être obtenu que lorsque l'inversion de population entre les deux niveaux de la transition optique est obtenue. L'injecteur devra donc injecter efficacement les électrons sur le niveau e3 et, d'autre part, extraire rapidement ceux-ci vers la période suivante.

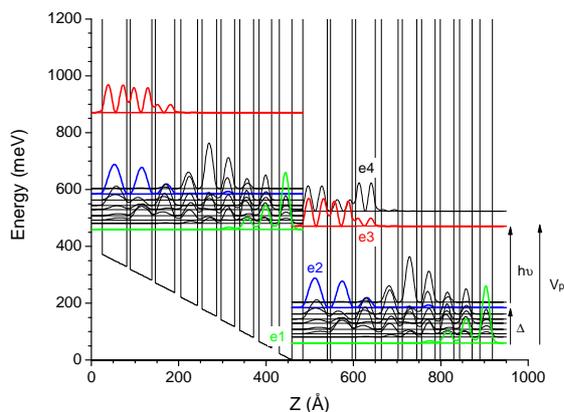


Fig.2 : Diagramme de bande d'une période de zone active de LCQ.

Un superréseau graduel est utilisé pour réaliser cette fonction. La taille de chaque puits est choisie afin d'aligner les niveaux pour un champ électrique de 50-100 kV/cm. Il y a alors apparition de minibandes d'énergie dont la première est mise en résonance avec le niveau e3. Les électrons pourront alors être injectés sur ce niveau à travers une barrière épaisse. Cette barrière, la barrière d'injection, devra être assez épaisse pour éviter les courants de fuite vers les niveaux e1 ou e2 mais assez fine pour permettre l'injection de porteurs vers le niveau e3 par effet tunnel résonnant. Il faudra donc trouver un compromis sur l'épaisseur de cette barrière qui se situe entre 25 et 35Å.

IV. LE SYSTEME DE MATERIAU INAS/ALSB.

Plusieurs systèmes de matériau sont aujourd'hui utilisés pour le développement de LCQ.

Le système InGaAs/InAlAs sur InP est celui qui permet la réalisation du premier LCQ en 1994 [2]. Plus tard, l'effet laser fût obtenu avec des LCQ à base de GaAs [3]. Ces deux systèmes sont aujourd'hui bien maîtrisés mais ne permettent pas d'atteindre des longueurs d'ondes inférieures à 4 μm du fait de leur faible discontinuité de bande conduction. Le choix des matériaux pour ces travaux s'est porté sur un troisième système basé sur les antimoniures.

Ce dernier utilise des barrières en AlSb, des puits en InAs et présente plusieurs avantages sur ces concurrents. Tout d'abord la possibilité de réaliser les structures sur des wafers d'InAs ou de GaSb. Mais sa caractéristique la plus intéressante pour la réalisation de LCQ est la grande discontinuité de bande de conduction entre InAs et AlSb (2.1eV) ce qui permet d'atteindre de très courtes longueurs d'ondes. De plus, les électrons ont une faible masse effective dans l'InAs ce qui a pour conséquence d'accroître le gain intersousbande. Le premier LCQ à antimonium fut réalisé en 2003 [4] et émettait à 10μm en pulsé.

V. LE GUIDE D'ONDE.

Comme dans les diodes lasers, le confinement des photons dans la direction de croissance est obtenu par un guide d'onde diélectrique plan qui repose sur un contraste d'indice entre le cœur du guide et les couches de confinement. Chaque couche du guide est décrite par son indice complexe qui est fonction de la longueur d'onde. La résolution des équations de propagation de l'onde électromagnétique nous donne l'indice complexe effectif des modes confinés dans le guide. $N_{eff} = n + i \cdot k$ est l'indice effectif du mode où la partie réelle représente l'indice optique et la partie imaginaire les pertes du mode. Les modes guidés sont les états propres du champ électromagnétique dans le guide.

L'onde se propage dans le plan perpendiculaire à la direction de croissance z et son champ électrique peut être parallèle ou perpendiculaire au plan des couches. D'après la règle d'or de Fermi, le couplage ne se fera qu'avec les modes transverses magnétiques (TM).

Le guide d'onde utilisé est réalisé par deux couches d'InAs fortement dopé prenant en « sandwich » la zone active. Les pertes optiques étant trop importantes dans l'InAs fortement dopé, il faut insérer une couche d'InAs faiblement dopé (les spacers) afin de confiner le mode dans la zone active [5] (Fig.3).

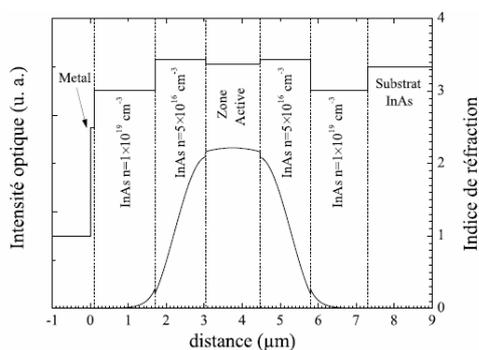


Fig.3 : Indice optique des différentes couches et mode optique TM0 (à droite).

Ce type de guide d'onde est efficace pour des composants émettant au-delà de 4 μm mais pose différents problèmes pour des longueurs d'onde inférieures [6]. En effet, cette limite correspond à l'énergie d'absorption interbande dans les spacers. Pour palier à ce problème, la solution retenue a été de remplacer les spacers d'InAs par un superréseau InAs/AlSb (typiquement 20 Å d'InAs + 20 Å d'AlSb). L'utilisation de ce superréseau va augmenter le gap optique et permettre une émission en dessous de 4 μm. L'épaisseur du spacer ainsi que le niveau de dopage devront être correctement choisis afin d'optimiser le recouvrement entre le mode optique et la zone active tout en réduisant les pertes.

VI. EPITAXIE ET TECHNOLOGIE DES STRUCTURES A CASCADE QUANTIQUE.

Les structures à cascade quantique sont réalisées par épitaxie par jets moléculaires (EJM) dans un bâti Riber Compact 21E. L'EJM est la condensation d'une phase vapeur en phase cristalline dans un environnement ultravide. Le bâti est constitué de cinq cellules d'effusion pour les éléments III (Al, Ga et In) et trois cellules pour les dopants (Si, Te et Be). Les éléments V (Sb et As) sont contenus dans des cellules à vannes. La qualité du vide dans ce type de bâti (10⁻¹⁰ Torr) permet la croissance de couches d'une qualité extrême ainsi que l'accès à différents types de contrôles in situ tel que le RHEED (Reflection High Energy Electron Diffraction). La croissance des structures à cascade quantique s'effectue grâce à des séquences de cache des cellules. L'épaisseur des couches d'une structure à cascade quantique est de l'ordre du nanomètre et rend l'épitaxie délicate au niveau des interfaces entre les matériaux. L'optimisation des conditions de croissance et des séquences de caches ont permis d'obtenir des structures de haute qualité cristalline ainsi que des interfaces nettes. L'épitaxie se fait sur des wafer de deux pouces en InAs ou GaSb.

Il va ensuite falloir faire la technologie des lasers, c'est-à-dire réaliser des rubans sur le wafer. Cette étape va permettre de confiner le mode optique dans le plan perpendiculaire au plan des couches. Classiquement, les rubans font 12 μm de large et 3 mm de long (Fig.4).

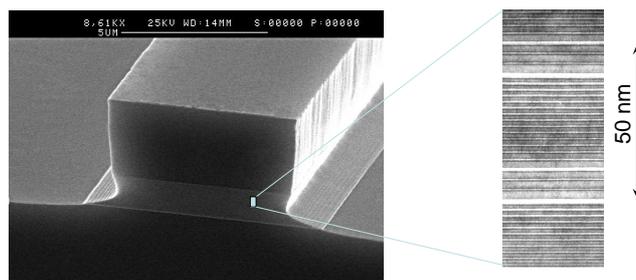


Fig.4 : Image MEB d'un ruban laser après gravure (à gauche). Image TEM d'une période de zone active (à droite).

Ils sont réalisés grâce à plusieurs étapes de photolithographie, de gravure et un dépôt de métal qui va permettre la prise de contact électrique. La gravure doit traverser la zone active afin de rendre homogène la densité de courant et le champ électrique appliqué dans tout le volume. Nous utilisons généralement une gravure humide dont la vitesse et le profil dépend de l'orientation et de la nature du matériau attaqué. Une solution H₃PO₄/H₂O₂/H₂O en proportion (2 : 1 : 2) est utilisée et permet d'obtenir des flancs de gravure sans rugosité trop importante. De plus, la vitesse d'attaque est d'environ 1 μm/min à 20°C ce qui permet un contrôle assez précis de la profondeur gravée. Les contacts électriques sont réalisés par évaporation de 20 nm de chrome et 200 nm de d'or. Les échantillons sont ensuite amincis à la poudre d'alumine et montés sur des embases de cuivre pour la caractérisation. Le cristal clivé est soudé à l'indium sur l'embase qui fait office de dissipateur thermique. Notons que c'est le côté épitaxié de l'échantillon qui est soudé afin d'approcher au maximum la zone active du cuivre. Plusieurs études ont été effectuées sur les étapes de la technologie des rubans laser afin de diminuer les effets de la température ou encore de réduire le courant de seuil. Des simulations ont permis de calculer la résistance thermique de nos structures et d'envisager des solutions pour diminuer cette dernière par le dépôt d'un diélectrique (SiO₂ ou Si₃N₄) ou de contact d'or épais (3 à 6 μm) permettant une meilleure dissipation thermique. Le courant de seuil peut être diminué d'environ 20% par le dépôt d'un miroir de haute réflectivité sur une facette du ruban.

VII. RESULTATS.

Les caractérisations électriques sont effectuées en utilisant des impulsions de 100 nsec à une fréquence de 1 kHz et les spectres laser sont récoltés par un spectromètre à transformée de Fourier avec un détecteur DTGS. Nous avons tout d'abord étudié les structures laser émettant à 4.5 μm. Nous avons opté pour un dessin vertical qui favorise le couplage entre les niveaux e3 et e2 et un spacer en superréseau. Après optimisation des conditions de croissance et du dessin de la zone active, nous avons obtenu un laser avec un courant de seuil de 1.7 kA/cm² à température ambiante et fonctionnant jusqu'à 400 K avec une puissance de sortie de 100 mW (Fig .5).

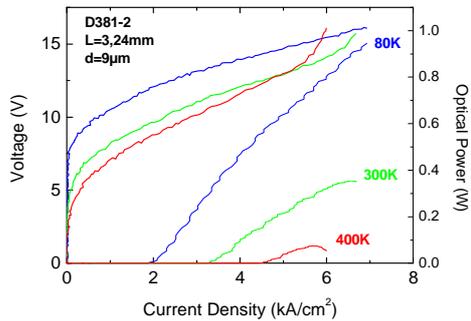


Fig.5 : Caractéristiques tension-courant et puissance électrique-courant.

Fort des résultats obtenus à $4.5\mu\text{m}$, nous avons développé plusieurs structures émettant autour de $3\mu\text{m}$. Pendant longtemps, la plus courte longueur d'onde obtenue avec un LCQ était de $3.3\mu\text{m}$ [7]. Record obtenu avec une structure réalisée sur InP. Le contrôle de l'alignement des niveaux devient de plus en plus délicat lorsque l'énergie de transition augmente. Pour cette raison, la réalisation de ces structures requiert un grand contrôle de la croissance des matériaux. Ceci afin de conserver une injection efficace pour un plus grand champ électrique. Après plusieurs essais nous avons réalisé une structure émettant à $3.3\mu\text{m}$ et qui présente un courant de seuil de 3kA/cm^2 et une puissance de sortie supérieure au Watt à température ambiante [8] (Fig.6).

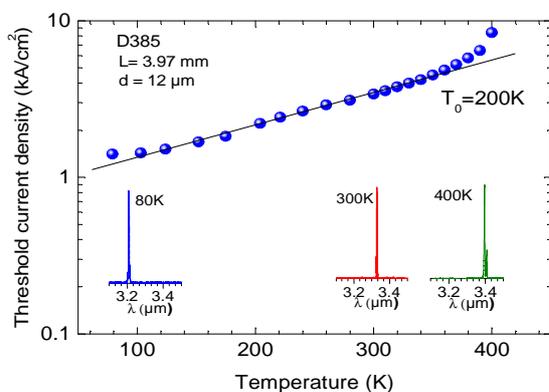


Fig.6 : Densité de courant de seuil en fonction de la température. Spectres laser mesurés pour 80K, 300K et 400K (en encart).

Nous avons ensuite réalisé le premier laser à cascade quantique émettant en dessous de $3\mu\text{m}$ tout système de matériaux confondus et fonctionnant à température ambiante [9].

VIII. CONCLUSION ET PERSPECTIVES

Nous avons donc démontré la possibilité de couvrir complètement la gamme de longueur d'onde entre 3 et $5\mu\text{m}$ par des LCQ dans la filière antimonure. Les travaux actuels visent à améliorer les performances de ces composants autour de $3\mu\text{m}$. Ceci passe par l'optimisation du dessin des structures, l'amélioration de la dissipation thermique ainsi que la diminution des pertes optiques dans la structure.

REFERENCES.

- [1] R.F. Kazarinov and R.A. Suris, «*Possibility of amplification of electromagnetic waves in a semiconductor with a superlattice*», Sov. Phys. Semicond., **5**, 707 (1971).
- [2] J. Faist, F. Capasso, D.L. Sivco, C. Sirtori, A.L. Hutchinson, A.Y. Cho, «*Quantum cascade laser*», Science, **264**, 553 (1994).
- [3] C. Sirtori, P. Kruck, S. Barbieri, P. Collot, J. Nagle, M. Beck, J. Faist, U. Oesterle, «*GaAs/Al_xGa_{1-x}As quantum cascade lasers*», Appl. Phys. Lett., **73**, 3486 (1998a).
- [4] K. Othani, H. Ohno, «*InAs/AlSb quantum cascade lasers operating 10µm*», Appl. Phys. Lett., **82**, 1003 (2003).
- [5] K. Ohtani, H. Ohno, «*InAs-Based quantum cascade light emitting structures containing a double plasmon waveguide*», Journal of Crystal Growth, **251**, 718 (2003).
- [6] R. Teissier, D. Barate, A. Vicet, C. Alibert, A.N. Baranov, X. Mercadet, C. Renard, M. Garcia, C. Sirtori, D. Revin, J. Cockburn, «*Room temperature operation of InAs/AlSb quantum cascade structures*», Appl. Phys. Lett., **85**, 167 (2004).
- [7] M.P. Semtiv, M. Wienold, S. Dressler, W.T. Masselink, «*Short wavelength ($\lambda \approx 3.3\mu\text{m}$) InP-based strain compensated quantum-cascade laser*», Appl. Phys. Lett., **89**, 211124 (2006).
- [8] J. Devenson, O. Cathabard, R. Teissier, A.N. Baranov, «*High temperature operation of $\lambda = 3.3\mu\text{m}$ quantum cascade lasers*», Appl. Phys. Lett., **91**, 141106 (2007).
- [9] J. Devenson, R. Teissier, O. Cathabard, A.N. Baranov, «*InAs/AlSb quantum cascade lasers emitting below 3µm*», Appl. Phys. Lett., **90**, 111118 (2007).

Réalisation d'un laser VCSEL monolithique à base d'antimoine pompé électriquement émettant à 2,3 μm

ARNAUD DUCANCHEZ¹, LAURENT CERUTTI², FREDERIC GENTY³

IES - CNRS UMR5214
Université Montpellier 2
Place Eugène bataillon
Bât. 13, CC067
34095 Montpellier FRANCE

¹ ducanchez@ies.univ-montp2.fr

² cerutti@ies.univ-montp2.fr

³ genty@ies.univ-montp2.fr
<http://www.ies.univ-montp2.fr/>

Résumé—Nous reportons dans ce travail, la fabrication et la caractérisation d'un laser à cavité verticale à base d'antimoine pompé électriquement, émettant à 2,3 μm , opérant en régime continu et à température ambiante jusqu'à 287 K. Il se compose de deux miroirs de Bragg type n, d'une jonction tunnel et d'une zone active à multi puits quantique. L'objectif final étant d'intégrer et d'utiliser ces sources lasers dans des dispositifs destinés à la détection de gaz.

Mots clés—VCSEL, antimoine, monolithique, jonction tunnel, pompage électrique.

I. INTRODUCTION.

De nos jours, les sources lasers à semi-conducteurs opérant dans le domaine du moyen infrarouge (2 – 5 μm) peuvent être utilisées dans de nombreuses applications, et notamment, pour la détection et l'analyse de traces de gaz par spectroscopie d'absorption.

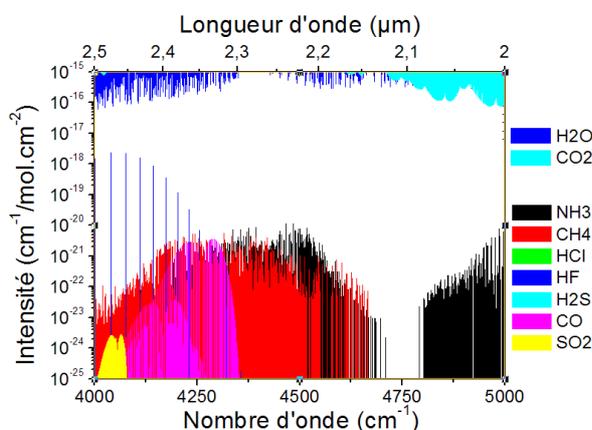


Figure 1. Spectre d'intensité des raies d'absorption des principaux gaz polluants entre 2-2.5 μm .

En effet, dans cet intervalle de longueur d'onde, et plus particulièrement entre 2,0 μm et 2,5 μm , les raies d'absorption de molécules polluantes telles que le méthane (CH₄) ou l'ammoniac (NH₃) sont très intenses comme on peut le voir sur la figure 1. En revanche, les raies d'absorption de l'eau sont moins intenses et surtout beaucoup moins denses. Par conséquent, le domaine spectral 2,0-2,5 μm correspond à une fenêtre de transmission de l'atmosphère.

Comparé aux diodes lasers classiques à guidage par l'indice, les lasers à cavité verticale à émission par la surface (VCSEL) sont particulièrement bien adaptés pour ce type d'application. En effet, ils possèdent de nombreux avantages comme un faisceau émergent circulaire et peu divergent, de faible courant de seuil, une émission monomode (une seule longueur d'onde) et la possibilité de fabriquer aisément des matrices de lasers. Bien que des VCSELs à base de GaSb pompés optiquement et émettant à 2,3 μm avec des propriétés tout à fait remarquables ont été réalisés [1]-[3], il est nécessaire de disposer de lasers pompés électriquement (EP-VCSEL) afin de développer et de réaliser des dispositifs expérimentaux tout intégrés et compacts.

De très bons résultats pour des EP-VCSELs à base d'InP émettant au-dessus de 2 μm et utilisant une jonction tunnel enterrée ont été reportés [4], mais semble représenter la limite que l'on puisse atteindre en terme de longueur d'onde avec ce type de matériau. A l'inverse, les hétéro-structures à base de GaSb permettent de couvrir l'intégralité du spectre du moyen-infrarouge [5], et récemment, le premier EP-VCSEL à base de GaSb émettant à 2,3 μm et opérant en régime continu à température ambiante a été présenté [6]. Néanmoins, il s'agit d'un procédé technologique très élaboré nécessitant une reprise d'épitaxie, alors que l'intérêt de notre approche réside dans le fait qu'il s'agit d'un procédé technologique moins contraignant et qui nous a permis d'obtenir des résultats tout à fait encourageants pour

la suite [6]. Le but final étant d'aboutir à un fonctionnement monomode en continu à température ambiante et avec une température de fonctionnement élevée.

II. EPITAXIE ET PROCEDE TECHNOLOGIQUE

La structure VCSEL a été réalisée par épitaxie par jets moléculaires (EJM) sur un substrat GaSb (n) dopé Te, d'orientation (100). Elle se compose d'un miroir de Bragg inférieur de 20 paires en $\text{AlAs}_{0,07}\text{Sb}_{0,93}/\text{GaSb}$ dopé n et centré à $2,3 \mu\text{m}$, en accord de maille sur le substrat. D'une cavité de longueur $1,5\lambda$ qui contient une zone active avec 5 puits quantiques de type-I en $\text{Ga}_{0,65}\text{In}_{0,35}\text{As}_{0,1}\text{Sb}_{0,9}$ séparés par des barrières en $\text{Al}_{0,65}\text{Ga}_{0,35}\text{As}_{0,04}\text{Sb}_{0,96}$ de 20 nm d'épaisseur et placée au maximum du champ. Cette demi-structure est ensuite surmontée d'une jonction tunnel en GaSb (p^{++})/InAsSb (n^{++}) et terminée par un miroir de Bragg supérieur de 18 paires en $\text{AlAs}_{0,07}\text{Sb}_{0,93}/\text{GaSb}$ dopé n à $1,2 \times 10^{18} \text{ cm}^{-3}$. Une coupe transversale de la structure complète est donnée figure 2.

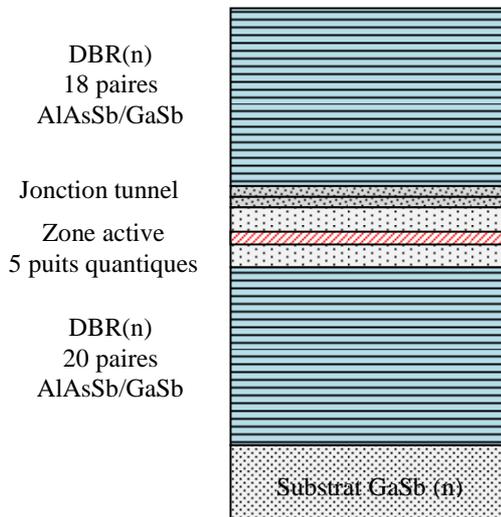


Figure 2. Coupe transversale de la structure VCSEL.

A partir de cette hétéro-structure, un procédé technologique est utilisé afin d'arriver au composant final. Les principales étapes sont illustrées figure 3. Tout d'abord, une première couche de résine négative est déposée et insolée afin de définir les contacts supérieurs en anneaux, où une couche d'épaisseur de 300 nm d' AuGeNi est déposée par pulvérisation cathodique. Au préalable, et avant de rentrer l'échantillon dans le bâti de dépôt, celui-ci subit une désoxydation avec une solution d'acide chlorhydrique diluée pendant deux minutes. Ensuite, la résine superflue est enlevée par un procédé classique de « lift-off ». La deuxième étape est consacrée à la gravure et à l'isolation. Pour ce faire, une couche de résine positive est déposée afin de protéger les contacts supérieurs de la gravure humide. Concernant la gravure, deux solutions sont utilisées. L'une à

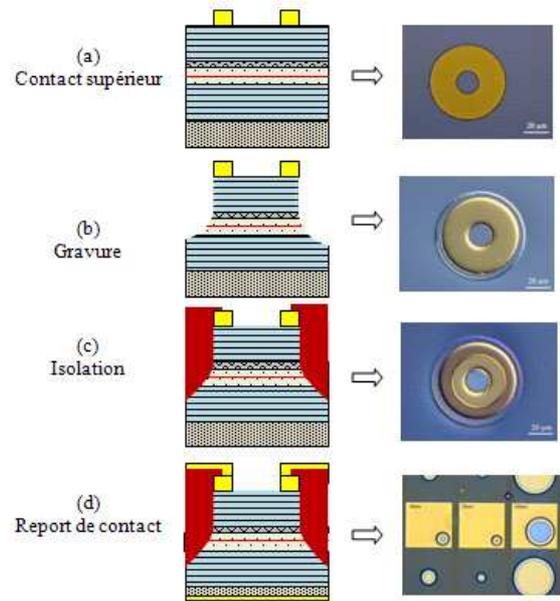


Figure 3. Principales étapes du procédé technologique pour la fabrication du composant : vue en coupe et au microscope optique.

base de $\text{CrO}_3:\text{HF}:\text{H}_2\text{O}$ pour graver les miroirs de Bragg et la zone active, et une autre à base d'acide citrique [7] pour la jonction tunnel, qui sert aussi de couche d'arrêt et qui permet d'obtenir des profils de gravure droit comme on peut le voir sur la figure 4.

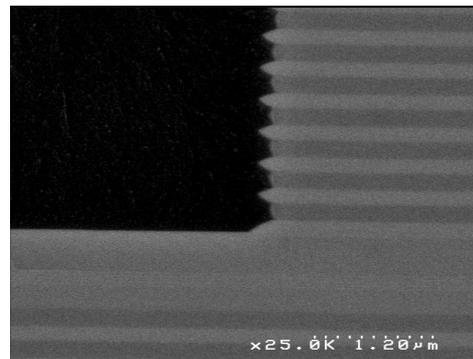


Figure 4. Image au MEB du profil de la gravure jusqu'à la couche d'InAsSb de la jonction tunnel.

Après cette étape de gravure, nous venons déposer de la résine positive qui est chauffée pendant deux heures jusqu'à une température maximum de 200°C pour assurer l'isolation électrique latérale. Enfin, le report de contact est obtenu par le dépôt d'une couche d'or d'épaisseur de 50 nm. La dernière étape consiste à amincir le substrat jusqu'à $150 \mu\text{m}$ par un polissage mécanique puis chimique et de métalliser la

face arrière avec une couche de 300 nm d' AuGeNi. A la fin du procédé technologique, le composant est soudé à l'indium sur une embase en cuivre pour être testé et caractérisé (Figure. 5).

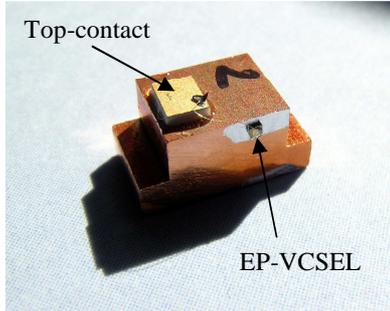


Figure 5. EP-VCSEL soudé sur une embase en cuivre pour test.

III. CARACTERISATION EN CONTINU

Les mesures et les caractéristiques ont été réalisées sur des composants qui possédaient un diamètre interne de 20 μm avec un contact en anneau de 20 μm de large, soit un diamètre total de 60 μm . La figure 6 montre la réflectivité et le spectre laser à 30 mA en régime continu à 284 K. Ils ont été mesurés grâce à un spectromètre pour l'infrarouge à Transformée de Fourier (FTIR) et montre une « stop band » assez large centrée à 2,308 μm avec un mode de la microcavité centré à 2,311 μm , qui démontre ainsi le bon accord entre la les miroirs de Braggs et le mode de la microcavité.

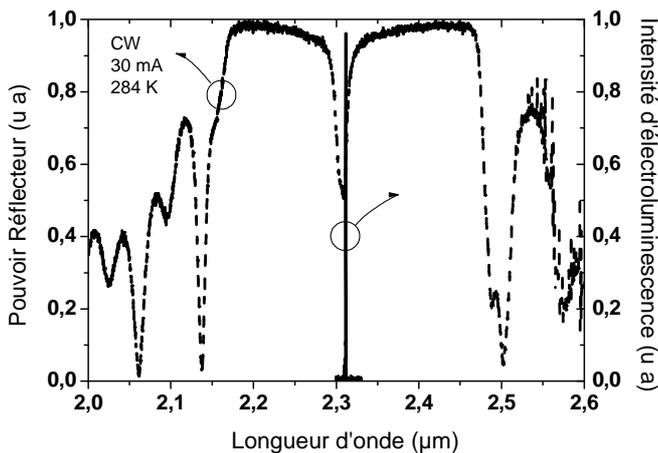


Figure 6. Réflectivité et spectre laser pour un composant de 20 μm de diamètre interne.

Les caractéristiques P(I) et V(I) en régime continu à différentes températures sont données figure 7 et permettent de mettre en évidence un fonctionnement jusqu'à 287 K

avec un courant de seuil de 33 mA, correspondant à une densité de courant de 1,16 kA/cm^2 .

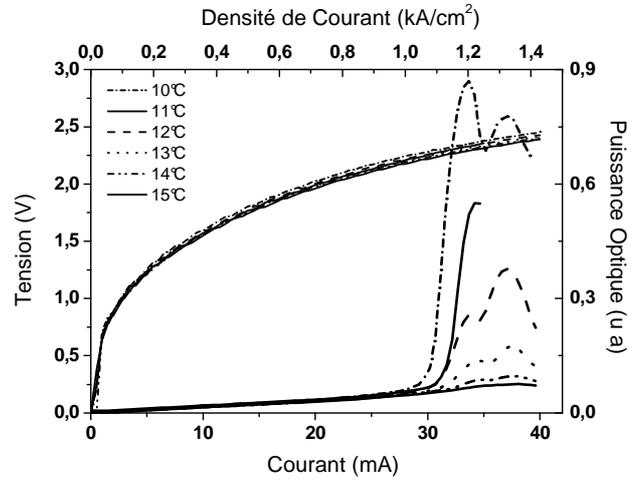


Figure 7. Caractéristiques P(I) et V(I) en régime continu pour différentes températures.

Enfin, l'évolution du courant de seuil en fonction de la température en régime quasi-continu (1 μs , 10 %) a été étudiée. La courbe obtenue est représentée sur la figure 8. On observe une variation parabolique du seuil qui est typique pour les lasers VCSEL. Une densité de courant de seuil minimum de 0,26 kA/cm^2 à 246K a été mesurée. Ce seuil minimum étant atteint lorsque le maximum du gain des puits quantiques est parfaitement accordé avec le mode de la microcavité.

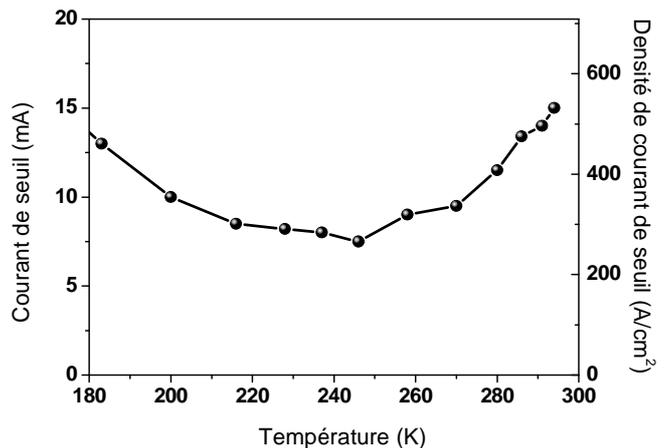


Figure 7. Courant de seuil en fonction de la T° en régime quasi-continu (1 μs , 10 %).

IV. CONCLUSION

Nous avons donc réalisé une structure VCSEL monolithique pompée électriquement à base de GaSb et émettant à 2,3 μm . Les facteurs clés responsables de ces

résultats résident dans l'utilisation des deux miroirs de Braggs (n) et de la jonction tunnel comparée à une structure VCSEL p.i.n classique. Les principales caractéristiques sont une température maximale de fonctionnement de 287 K avec une densité de courant de seuil de $1,16 \text{ kA/cm}^2$ pour des composants possédant un diamètre total de $60 \mu\text{m}$ et un procédé technologique assez simple pour la fabrication du composant lui-même. Cependant, afin d'améliorer et d'étendre la plage de fonctionnement en température de ces composants jusqu'à 300 K, il est nécessaire d'améliorer l'aspect thermique en limitant l'échauffement responsable en partie de la dégradation des caractéristiques et des performances du composant. Enfin, cela passe ensuite par la réalisation d'EP-VCSELS de plus petits diamètres ($\approx 10 \mu\text{m}$).

V. REMERCIEMENT.

Ce travail est financé en grande partie par le projet Européen NEMIS n°IST-2005-31845.

Enfin, les auteurs voudraient remercier A. Vicet pour son aide concernant les caractérisations électriques et optiques des composants.

REFERENCES.

- [1] A. Ouvrard, A. Garnache, L. Cerutti, F. Genty and D. Romanini, "Single frequency tunable Sb-based VCSEL emitting at $2.3 \mu\text{m}$," *IEEE Photon. Technol. Lett.*, vol. 17, pp. 2020-2022, October 2005.
- [2] N. Schulz, M. Rattunde, C. Manz, K. Köhler, C. Wild, J. Wagner, S. -S. Beyertt, U. Brauch, T. Kübler and A. Giesen, "Optically pumped GaSb-based VCSEL Emitting 0.6W at $2.3 \mu\text{m}$ " *IEEE Photon. Technol. Lett.*, vol. 18, pp. 1070-1072, 2006.
- [3] J. -M. Hopkins, A. J. Maclean, D. Burns, E. Riis, N. Schulz, M. Rattunde, C. Manz, K. Köhler, and J. Wagner, "Tunable, Single-frequency, diode pumped $2.3 \mu\text{m}$ VCSEL," *Optics Express*, vol. 15, issue 13, pp. 8212-8217, June 2007.
- [4] M. Ortsiefer, G. Böhm, M. Grau, K. Windhorn, E. Rönneberg, J. Roskopf, R. Shau, O. Dier, and M. -C. Amann, "Electrically pumped room temperature CW VCSELS with $2.3 \mu\text{m}$ emission wavelength," *Electron. Lett.*, vol. 42, no. 11, pp. 640-641, May 2006.
- [5] A. Krier, "Mid infrared Semiconductor Optoelectronics", *Springer-Verlag (Optical Sciences)*, 2006.
- [6] A. Bachmann, T. Lim, K. Kashani-Shirazi, O. Dier, C. Lauer and M. -C. Amann, "Continuous-wave operation of electrically pumped GaSb-based vertical cavity surface emitting laser at $2.3 \mu\text{m}$," *Electron. Lett.*, vol. 44, no. 3, pp. 202-203, January 2008.
- [7] O. Dier, C. Lin, M. Grau and M.C. Amann, "Selective and non-selective wet-chemical etchants for GaSb-based materials," *Semicond. Sci Technol.*, vol. 19, no. 11, pp. 1250-1253, November 2004.

CYCLE DE CONCEPTION D'UN LASER A SEMI-CONDUCTEUR POUR LA DETECTION DE GAZ

ALBAN GASSENQ (DOCTORANT 1ERE ANNEE), LAURENT CERUTTI, ALEXEÏ BARANOV,
ERIC TOURNIE (DIRECTEUR DE THESE)

Institut Electronique du Sud – UMR5214 du CNRS
Université Montpellier 2, Case courrier 067, 34095 Montpellier Cedex 5. France

GASSENQ@IES.UNIV-MONTP2.FR
ERIC.TOURNIE@UNIV-MONTP2.FR
[HTTP://WWW.EDI2S.UNIV-MONTP2.FR/](http://WWW.EDI2S.UNIV-MONTP2.FR/)

Résumé—Le groupe « Nanomir » de l'IES travaille sur la conception et la fabrication de nouveaux systèmes photoniques moyens infrarouges (M.I.R). Le M.I.R comprend la gamme de longueurs d'onde allant de 2 à 5 μm , elle permet de réaliser des émetteurs lasers très bien adaptés à la détection de polluants gazeux par spectroscopie. La conception d'un laser comprend quatre étapes. Premièrement, la simulation du composant à fabriquer permettra de se faire une idée des résultats espérés. Ensuite la structure est réalisée par épitaxie par jet moléculaire. Cette technologie permet de contrôler avec précision des croissances allant de la monocouche atomique ($\sim 0.1\text{nm}$) jusqu'aux micromètres. Troisièmement, plusieurs étapes technologiques permettent, à partir de la plaquette épitaxiée, de mettre en forme ce laser. Enfin des études du composant indiquent les paramètres à modifier qui sont réinjectés dans de nouvelles modélisations afin de repartir pour un nouveau cycle.

Mots clés— Laser ; semi-conducteur ; infrarouge; détection de gaz.

I. INTRODUCTION.

La protection de l'environnement et la lutte contre la pollution sont des thèmes majeurs du XXI^{ème} siècle qui nécessitent des efforts particulièrement appuyés. La spectroscopie optique est le moyen le plus précis et le plus sensible pour détecter et analyser les gaz. Le MIR présente un grand intérêt pour ce type d'application, car il existe dans cette gamme de longueur d'onde de nombreuses raies d'absorption de gaz polluants (CH_4 , HCl , CO_2 , HF , NH_3 ...). Les systèmes d'analyses photoniques nécessitent l'emploi de composants lasers opérant à température ambiante en courant continu. Pour le moment il n'existe ce type de composant qu'entre 2 et 3 μm [1]. La recherche de nouveaux lasers émettant à plus grande longueur d'onde permettra d'augmenter la sensibilité de détection. En effet, les molécules polluantes sont d'autant plus absorbantes que la longueur d'onde est grande.

II. LES SEMI-CONDUCTEUR

Les semi-conducteurs sont des matériaux présentant une conductivité électrique intermédiaire entre les métaux et les isolants. La plupart des alliages semi-conducteurs des colonnes III et V du tableau de Mendeleïev cristallisent sous la forme d'une structure « Blende de Zinc ». Comme montré sur la figure1.(b) l'arrête de ce cube « 2a » définira le paramètre de maille.

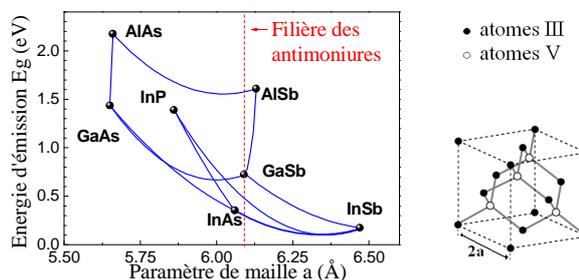


figure1.(a)

figure1.(b)

Figure 1. Energie d'émission en fonction du paramètre de maille ; (a). $E_g=f(a)$; (b) structure cristalline.

Ces matériaux sont utilisés pour l'optoélectronique car ils peuvent émettre de la lumière. Leur énergie d'émission ou « gap » (E_g), dépend de leur paramètre de maille, comme le montre le graphique de la figure 1 (a). C'est pour cela que la filière du semi-conducteur définira la longueur d'onde d'émission, reliée à E_g par la relation (1).

$$\lambda(\mu\text{m}) \approx \frac{1.24}{E_g(\text{eV})} \quad (1)$$

Un laser à semi-conducteur [2] sera réalisé à partir d'un empilement de couche de différents alliages de matériaux III- V ayant des paramètres de mailles proches.

III. CYCLE DE CONCEPTION D'UN LASER

La réalisation d'un nouveau laser se présente sous la forme d'une boucle comprenant quatre étapes. Chaque nouveau composant est sensé être plus performant que le précédent grâce aux modifications qui lui ont été apportées.

III.1. Modélisation.

La modélisation est un outil très utile permettant de tirer des tendances expérimentales. Les simulations permettent de s'intéresser aux caractéristiques optiques, électriques et thermiques. Par exemple les diagrammes de bandes et les fonctions d'onde des électrons et des trous présentent l'énergie ou la longueur d'onde d'émission du futur laser. Ce type de graphique est représenté sur la figure 2.(b). Dans ce cas, le laser possède deux puits quantiques, ce sont les zones à faibles gaps où les photons sont émis. Lorsqu'un électron passe du niveau haut au niveau bas, l'énergie de ce photon (E_{photons}) est égale à la différence entre ces deux niveaux. La lumière est ensuite confinée par les couches de « cladding » (en blanc sur la figure 2.(a)) dans le guide d'onde. Ces couches de confinement optique possèdent toujours un grand gap (E_{g1}) et le guide d'onde un gap intermédiaire (E_{g2}). Cette différence donne un contraste d'indice optique permettant de guider la lumière.

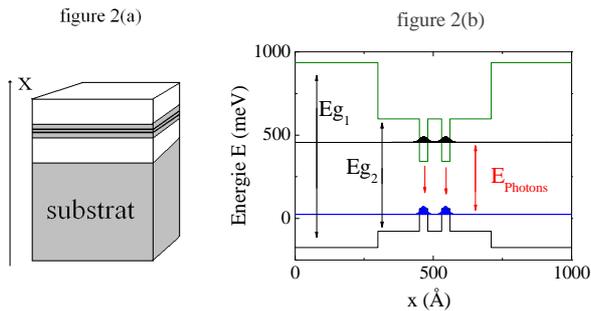


Figure 2. Exemple de diagramme de bande d'un laser
(a). Structure de l'échantillon ; (b) Diagramme de l'échantillon.

Ce type de modélisation indique que la famille des antimoniures (Sb) est la meilleure candidate pour réaliser des lasers émettant dans le moyen infra rouge.

III.2. Epitaxie par jets moléculaires.

Après la modélisation, on passe à l'étape de fabrication. L'épitaxie par jets moléculaires consiste à déposer une couche cristalline sur un monocristal de même composition, ou de composition différente mais de paramètre de maille identique ou très proche, en envoyant des flux atomiques ou moléculaires [3]. L'épitaxie par jets moléculaires peut être résumée à un appareil d'évaporation sous ultra vide. Le chauffage des matériaux contenus dans des cellules indépendantes permet un dépôt par évaporation des atomes qui cristallisent au contact du substrat (figure 2(a)).

III.3. Etapes technologiques.

Une fois le cristal réalisé, il doit subir plusieurs étapes technologiques pour passer du matériau au composant final. Premièrement, des étapes de lithographie et de gravure permettront de dessiner un ruban. Sur l'exemple de la figure 3 ce « mesa » est de 100µm. Ensuite des dépôts d'or permettront d'amener le courant pour alimenter le laser. Pour finir, on clive (coupe) le laser à la longueur souhaitée (2 mm sur la figure 3), et on soude la puce laser sur un support en cuivre (figure 4).

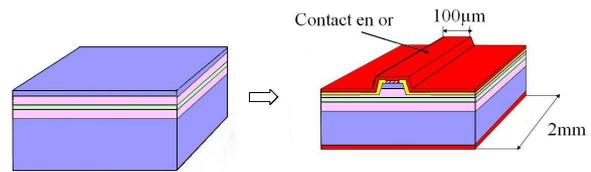


Figure 3. Modifications technologiques apportées à une plaquette épitaxiée pour faire un laser.

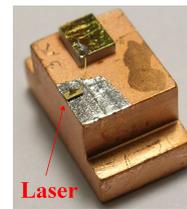


Figure 4. Laser monté sur son embase en cuivre.

Ces étapes utilisent des méthodes de la microélectronique permettant d'atteindre des précisions de masquage inférieures au micromètre.

III.4. Caractérisation.

La caractérisation du composant livre le verdict. Bien que l'application définira toujours le cahier des charges, on cherchera, la plupart du temps, à avoir un laser fonctionnant à température ambiante, alimenté en courant continu et ayant une longueur d'onde précise. Les figures 4 et 5 montrent les deux principaux types de caractérisations électro-optiques réalisées sur un laser à semi-conducteur. La caractéristique « P(J) » (figure 5) est la variation de la puissance optique en fonction du courant d'alimentation. Sur cet exemple on peut observer que la puissance émise augmente rapidement au delà de 0.4kA/cm² à 93K, c'est le « seuil » de l'effet laser. Il augmente avec la température de fonctionnement.

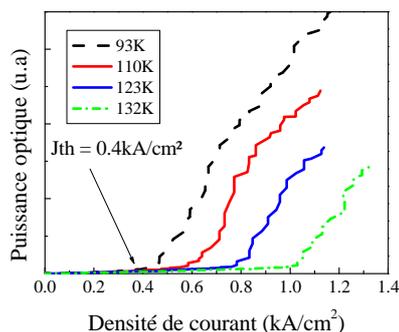


Figure 5. Variation de la puissance d'émission d'un laser en fonction du courant injecté

La figure 6 montre un exemple de caractéristique « $P(\lambda)$ » à 90K, c'est la variation de la puissance de sortie en fonction de la longueur d'onde. Pour cet exemple, l'effet laser apparaît au dessus de 1A.

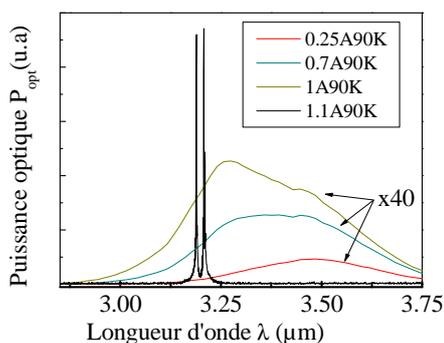


Figure 6. Variation de la puissance d'émission d'un laser en fonction de la longueur d'onde.

Ces deux caractérisations fournissent des paramètres primordiaux dans l'étude d'un laser. Leurs variations peuvent être étudiées en fonction de la température et du mode d'alimentation.

IV. APPLICATION.

Toutes les molécules possèdent des bandes d'absorption spécifiques liées aux énergies de vibrations et de rotations moléculaires. Comme il est rappelé dans l'introduction, les raies d'absorption de certains polluants sont intenses dans le moyen infra rouge entre 2 et 5 μm . La filière des antimoniures permet une émission dans la gamme de longueurs d'onde comprise entre $\lambda = 1.8$ et 5 μm (fig.1(a) et équation 1), c'est pourquoi leur utilisation est avantageuse pour ce genre d'application [4]. La figure 7 montre l'effet de l'absorption d'un gaz sur la caractéristique $P(I)$, l'analyse de la puissance optique reçue à travers une cellule permettra d'analyser et de quantifier tout les gaz contenus dans celle ci.

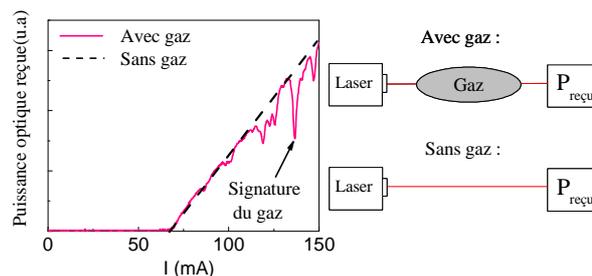


Figure 7. Variation de la puissance reçue par un détecteur avec ou sans cellule gazeuse

Ce type de système peut par exemple permettre de mesurer l'émission de polluants dans l'atmosphère par les incinérateurs industriels en temps réel, ce qui est un enjeu de taille pour bien des industriels.

V. CONCLUSION.

Ainsi la réalisation de ces nouveaux types de sources permettrait de détecter des gaz encore difficilement détectable avec précision. Il va de soit que les quatre étapes présentées sont toutes aussi importantes les unes que les autres et doivent se réaliser en synergie pour arriver à l'objectif d'une source laser émettant entre 3 et 5 μm en continu à température ambiante.

VI. REMERCIEMENTS.

Les travaux que j'effectue actuellement se déroulent dans le cadre du projet (FP6-017383) DOMINO financé par la Communauté Européenne regroupant les compétences de six laboratoires : le Paul-Drude-Institute à Berlin (Allemagne), le CNR National Nanotechnology Laboratory de Lecce (Italie), l'Université de Hull (Angleterre), le Ioffe Physico-Technical Institute de St. Petersburg (Russie) et Alcatel-Thales III-V Lab de Orsay (France).

REFERENCES.

- [1] D. Barat, «Animonide-based DFB laser emitting above 2.6 μm », 2007, Electronics Letters 43:23
- [2] E. Rosencher «Optoélectronique » 1998.
- [3] M. A. Herman «Molecular Beam Epitaxy.» 1996, 2nd edition
- [4] A. Vicet «Etude et réalisation d'un analyseur multigaz à diodes lasers accordables», 2001, Thèse Université Montpellier II, IES – Nanomir

VALIDATION DE BASES DE DONNEES DE REACTIONS NUCLEAIRES $n + \text{Si}$ ET $p + \text{Si}$ PAR MESURE DE L'ENERGIE DEPOSEE DANS UN CAPTEUR CCD

SIMON ROCHEMAN¹ (DOCTORANT 3E ANNEE), FREDERIC SAIGNE² ET FREDERIC WROBEL³ (DIRECTEURS DE THESE)

IES, Université Montpellier II
Place Eugène Bataillon, 34000 Montpellier Cedex, France

[1 SIMONROCHEMAN@GMAIL.COM](mailto:1.SIMONROCHEMAN@GMAIL.COM)

[2 SAIGNE@IES.UNIV-MONTP2.FR](mailto:2.SAIGNE@IES.UNIV-MONTP2.FR)

[3 WROBEL@IES.UNIV-MONTP2.FR](mailto:3.WROBEL@IES.UNIV-MONTP2.FR)

[HTTP://WWW.EDI2S.UNIV-MONTP2.FR/](http://WWW.EDI2S.UNIV-MONTP2.FR/)

Résumé— L'irradiation de proton ou de neutron provoque des réactions nucléaires dans les noyaux de silicium des composants électronique. Les produits de ces réactions déposent leur énergie en ionisant le semi-conducteur. L'irradiation d'un capteur CCD permet de mesurer l'énergie déposée par ces réactions nucléaires sous forme d'images. On compare les résultats de ces expériences avec la simulation Monte Carlo du dépôt de charge par réaction nucléaire dans une structure CCD pour valider notre connaissance des réactions nucléaires.

Mots clés— réaction nucléaire, CCD, Monte Carlo, irradiation, silicium.

I. INTRODUCTION.

Lorsque des rayons cosmiques se rapprochent de la Terre, les particules chargées sont happées par le champ magnétique terrestre et restent piégées dans la magnétosphère. Les particules non chargées, elles, atteignent l'atmosphère, interagissent avec les noyaux atomiques de l'air et créent ainsi des produits secondaires. Ces particules secondaires produisent eux-mêmes d'autres particules par réaction nucléaire avec l'air, créant ainsi des cascades où les neutrons sont prépondérants [1]. Les satellites sont donc soumis à des radiations de protons, électrons et ions tandis que les avions sont soumis à des radiations de neutrons. Les radiations de neutrons ou protons peuvent causer des effets singuliers dans les composants en déposant des charges parasites.

Les neutrons n'étant pas chargés, ils ne ionisent pas la matière directement. Les protons, eux, ionisent peu la matière. Par contre, ces deux nucléons peuvent interagir avec les noyaux d'atome de silicium du composant. Les ions produits, eux, ionisent le silicium [2] le long de leur trace. On appelle ce transfert d'énergie le LET (Linear Energy

Transert) [3]. Ce sont ces charges qui peuvent modifier le comportement du composant.

Dans le but de prévoir la fréquence des effets singuliers d'un composant, on simule ce processus. Les réactions nucléaires neutron + silicium ou proton + silicium peuvent donner des produits secondaires très variés. De l'hydrogène au phosphore, ces produits peuvent avoir différentes énergies et directions suivant des lois statistiques qui dépendent de l'énergie du nucléon incident.

Dans des travaux précédent, un code de physique nucléaire nommé MC-RED [4] (Monte Carlo Recoil Energy Determination) a été développé. Ce code nous fournit les détails des réactions pour des énergies incidentes comprises entre 1 et 200 MeV.

Les résultats de ce code ont été partiellement validés par des expériences dans de précédents travaux : un faisceau de nucléon mono-énergétique est dirigé sur une cible en silicium, on mesure le type, l'énergie et la direction des ions produits par réaction. Mais ces informations ne peuvent pas être obtenues pour les ions lourds ($Z > 2$) car ceux-ci perdent leur énergie très rapidement et ne sortent donc pas de l'échantillon-cible.

Le but de ce travail est de valider les résultats de MC-RED par méthode indirecte : on irradie de neutrons ou de protons un détecteur lui-même composé de silicium. Le détecteur présenté ici est un capteur de caméra CCD. La charge déposée dans le silicium par chaque réaction nucléaire est collectée dans les pixels et traduite en image par la caméra.

Parallèlement, on simule cette expérience par méthode Monte Carlo : on génère la réaction nucléaire à une position aléatoire et on calcule l'énergie collectée par les pixels un grand nombre de fois. On compare ensuite la simulation avec l'expérience.

II. EXPERIENCE.

II.1. Système expérimental.

Le boîtier et l'optique d'une caméra CCD sont retirés pour laisser le capteur accessible. Le capteur est constitué de 384 x 288 pixels, chaque pixel mesure 23 μm x 23 μm. La surface totale du CCD est de 0.6 cm². La caméra est reliée à une carte d'acquisition sur ordinateur. Les fichiers sont enregistrés sous format numérique. La vidéo fonctionne en mode entrelacé. Cela signifie que chaque image du fichier numérique est l'addition de deux trames successives enregistrées dans le capteur CCD. Les lignes paires correspondent à une trame et les lignes impaires à la suivante. Chaque image correspond à 40 ms. Le temps de collection de la charge par le capteur CCD pour chaque trame est donc de 20 ms.

Le système est étalonné pour fournir une loi de conversion de niveau de gris d'un pixel en charge. Cet étalonnage est réalisé par des tirs de laser sur la CCD. La fréquence des tirs laser est fixée à 20Hz. (un tir ou moins par image) Connaissant l'énergie déposée par un tir laser, on en déduit la quantité de charge libérée par son passage. La figure 1 présente le détail d'une image obtenue par tir laser. On fait varier l'intensité du laser du seuil de détection minimum jusqu'à saturation des pixels centraux de l'amas de pixel.

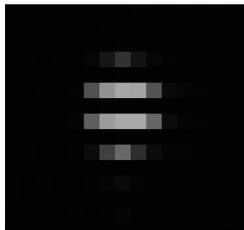


Fig. 1. Détail d'une image obtenue lors d'un tir laser sur le capteur CCD.

L'énergie calculée pour l'ensemble des pixels est ajoutée et comparée à l'énergie déposée par le laser. On cherche une fonction sous la forme correspondant à la correction gamma effectuée par l'électronique de la caméra.

$$E = A \cdot \frac{e^{\text{greyscale}/B} - 1}{e^{255/B} - 1} \quad (1)$$

II.2. Irradiations

Le capteur est placé dans l'obscurité de façon à ce que seules les réactions nucléaires y déposent des charges. Il est irradié par un faisceau de proton ou de neutron orienté suivant la normale au capteur. Plusieurs expériences ont été réalisées : Le capteur a été irradié par des faisceau mono-énergétiques de protons de 20, 50 et 150 MeV et de neutrons de 30 et 63 MeV. Les vidéos obtenues présentent des tâches qui correspondent à la trace d'ions secondaires dans les pixels. Il arrive qu'on obtienne de longues traces correspondant à des ions légers qui déposent leur énergie sur une grande distance figure 2 (a) ou bien des tâches très

lumineuses correspondant au dépôt de charge par un ion lourd figure 2 (b).

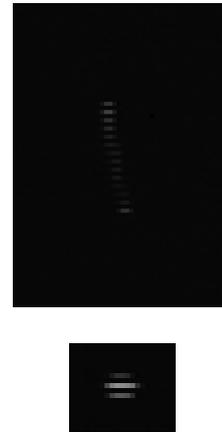


Fig. 2. Détail d'images obtenues lors d'irradiation du capteur CCD par des neutrons de 30 MeV ; (a) trace d'un ion léger; (b) trace d'un ion lourd.

III. SIMULATION.

L'expérience est simulée par méthode Monte Carlo. Il est nécessaire de simuler un grand nombre de réactions nucléaires dans le composant. La simulation de réaction nucléaire demande beaucoup de temps de calcul. Pour rendre les simulations plus rapides, 100 000 réactions nucléaires n+Si et p+Si ont été calculées pour 12 énergies distinctes. Pour chaque réaction nucléaire, les détails suivants sont enregistrés dans une base de données : le type des ions secondaires, leurs énergies et leurs directions par rapport à celle du nucléon incident.

Ces bases de données sont utilisées pour la simulation d'irradiation de composants. Dans la simulation de nos expériences, la position de la réaction est déterminée aléatoirement. Chaque trace d'ion est discrétisée en 30 points. La quantité de charge déposée en chaque point est donnée par le programme SRIM en fonction du type d'ion et de son énergie [5]. Les charges déposées en profondeur sont collectées dans les pixels par diffusion. On calcule la quantité de charge collectée par chaque pixel.

IV. ANALYSE.

Il existe deux méthodes pour analyser les vidéos obtenues et ainsi comparer les résultats expérimentaux aux simulations.

IV.1. Charge par pixel

La plus simple consiste à créer un histogramme de tous les niveaux de gris obtenus dans une vidéo. Cet histogramme de niveau de gris peut se traduire ensuite en terme de charge par pixel sans aucune information géométrique. La figure 3 présente les résultats obtenus au Centre de Protonthérapie d'Orsay pour des protons d'énergie 50 et 150 MeV.

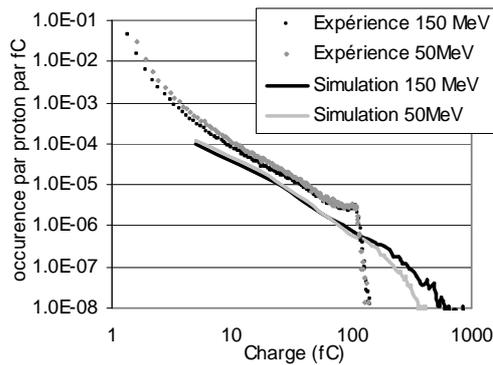


Fig. 3. Comparaison expérience simulation de la charge déposée par pixel par irradiation d'une CCD par protons de 50 et 150 MeV au CPO.

On constate que les simulations sont d'une bonne forme dans la gamme 5 fC– 100 fC mais elles sous-estiment le nombre de pixel touchés d'un facteur 1.7. Au-delà de 100 fC, le pixel sature et fournit son niveau de gris maximum.

IV.2. Energie par amas de pixels

La seconde méthode d'analyse consiste à additionner l'énergie contenue dans des pixels adjacents correspondant à un même évènement. On effectue le même calcul dans la simulation pour comparer les deux histogrammes obtenus. La figure 4 (a) présente les résultats pour une irradiation de neutrons de 30 MeV. La figure 5 (b) présente les résultats pour une irradiation de protons de 50 MeV. On constate un très bon accord dans le cas des neutrons pour des énergies supérieures à 0.5 MeV. Dans la gamme inférieure, le grand nombre d'évènements observés est du au bruit dans les mesures. Dans le cas des protons, on observe un très bon accord pour des énergies supérieures à 2 MeV. En dessous de 2 MeV, le grand nombre d'évènement observé est causé par le dépôt d'énergie de protons à travers la CCD.

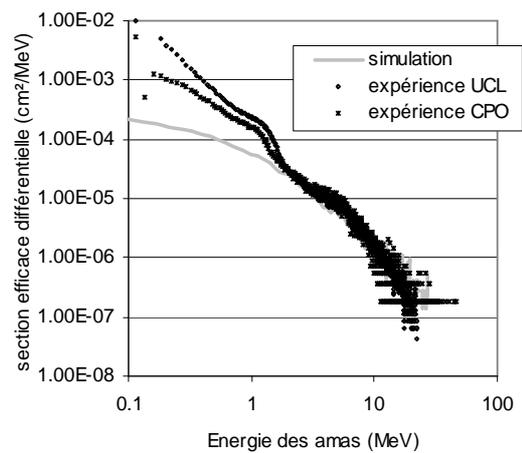
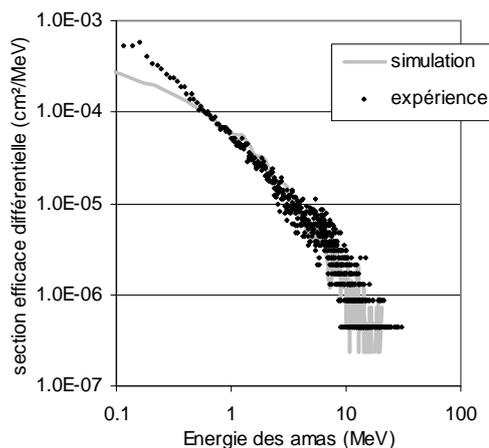


Fig. 3. Comparaison simulation expériences de l'énergie par amas de pixels CCD irradiée ; (a) par neutrons de 30 MeV à l'UCL; (b) et protons de 50 MeV à l'UCL et au CPO.

En effet, les protons incidents peuvent ioniser la matière sur leur passage. Un proton peu énergétique (d'énergie inférieure à 2 MeV) dépose toute son énergie dans le composant. Malgré le fait que le faisceau de proton soit mono-énergétique, le capteur CCD a reçu des protons de basse énergie. Ils sont très rares en comparaison des protons mono-énergétiques (environ 1%) mais chacun d'eux est observable alors que dans le cas des protons de 50 MeV, seuls ceux qui provoquent une réaction nucléaire avec le silicium sont observés.

V. CONCLUSION.

Simulations et expériences sont en bon accord dans le cas de l'irradiation d'une CCD. En particulier pour l'énergie collectée par amas de pixels. D'autres capteurs sont utilisés pour mesurer l'énergie déposée par réaction nucléaire dans des volumes de silicium différents.

REFERENCES.

- [1] M.S. Gordon, P. Goldhagen, K.P. Rodbell, T.H. Zabel, H.H.K. Tang, J.M. Clem, and P. Bailey, "Measurement of the Flux and Energy Spectrum of Cosmic-Ray Induced Neutrons on the Ground", IEEE Trans. Nucl. Sci., 2004, vol. 51, no. 6, pp. 3427-3434.
- [2] E. L. Petersen, "Nuclear reactions in semiconductors," IEEE Trans. Nucl. Sci., 1980, vol 27, no. 6, pp. 1494-1499.
- [3] C. Vial, J. M. Palau, J. Gasiot, M.C. Calvet, and S. Fournier, "A new approach for the prediction of the neutron-induced SEU rate", IEEE Trans. Nucl. Sci., 1998, vol. 45, no. 6, p. 2915-2920
- [4] F. Wrobel, J.M. Palau, M.C. Calvet and P. Iaconi, "Contribution of SiO₂ in Neutron-Induced SEU in SRAMs", IEEE Trans. Nucl. Sci., 2003, vol. 50, no. 6, pp. 2055-2059
- [5] J.F. Ziegler, J.P. Biersack, U.Littmark, "The stopping and Range of ions", Pergamon Press, New-York, 1985

STIMULATION LASER DE LA RETINE

JEAN-MICHEL BEC¹ (DOCTORANT 2E ANNEE), MICHEL DUMAS² (DIRECTEUR DE THESE)

Laboratoire IES

Equipe STINIM, Ecole des Mines d'Alès, Site EERIE, Parc Scientifique George Besse, 30035 Nîmes Cedex 1, France

1 BEC@IES.UNIV-MONTP2.FR

2 DUMAS@UNIV-MONTP2.FR

[HTTP://WWW.EDI2S.UNIV-MONTP2.FR/](http://WWW.EDI2S.UNIV-MONTP2.FR/)

Résumé— Les prothèses visuelles rétiniennes semblent être la solution la plus intéressante pour les patients atteints d'une maladie des photorécepteurs telles que la Dégénérescence Maculaire Liée à l'Age « DMLA » ou la rétinite pigmentaire « RP ». Jusqu'à présent, toutes les équipes travaillant à la conception d'une prothèse visuelle, utilisent des réseaux d'électrodes pour injecter un courant et faire varier localement le potentiel de membrane des neurones. Ce papier présente une nouvelle méthode sans contact pour stimuler avec une grande précision la rétine. Cette technique est basée sur la stimulation des neurones de la rétine par la lumière d'un laser pulsé infrarouge. Les résultats obtenus montrent la faisabilité d'une telle stimulation qui pourrait fournir une solution alternative aux techniques plus classiques de stimulation électrique.

Mots clés— Prothèse visuelle ; Laser Infrarouge ; Rétine, Potentiel d'action.

I. INTRODUCTION.

Une prothèse visuelle vise à remplacer le maillon défaillant de la chaîne psychosensorielle. Dans le cas de maladies qui atteignent les photorécepteurs, les solutions proposées visent à stimuler directement le dernier étage du traitement de l'information de la rétine assuré par des neurones, appelés cellules ganglionnaires. Ce sont ces dernières qui assurent le transfert de l'information le long de leur axone jusqu'au cerveau. La plupart des équipes travaillant à la conception d'une prothèse visuelle ont choisi d'utiliser des réseaux d'électrodes pour déposer des charges et faire varier localement le potentiel de membrane des neurones. Ces techniques d'implantation et de stimulation électrique sont variées (stimulation corticale [1], épi rétinienne [2] [3] [4] [5], sub rétinienne [6] ou du nerf optique [7] et ont permis d'obtenir des sensations visuelles chez des patients, nommées « phosphènes » [8] [9] rendant possible, dans certains cas, la reconnaissance de formes simples. Néanmoins, la stimulation par voie électrique présente des inconvénients directement liés à la nature invasive du système (accès chirurgical difficile) pour certains ou pour d'autres raisons d'ordre plus technologiques. Dans cet article nous présentons une méthode novatrice de stimulation optique par laser infrarouge. Cette méthode optique permet une stimulation sans contact possédant une très bonne

résolution spatiale (plus le diamètre du spot du laser sera petit, plus la définition sera précise). De plus, les dimensions actuelles des lasers à semi conducteurs permettent d'envisager des dispositifs de stimulation très compacts pouvant être implantés. Les résultats obtenus résultent d'une collaboration entre l'équipe Système de Traitement de l'Information Neuronale et d'Ingénierie Médicale STINIM de l'Institut d'électronique du Sud (IES) de Montpellier et l'institut de Neurosciences de Montpellier (INM). Les propriétés de fonctionnements des lasers font d'eux, des instruments couramment utilisés dans le milieu médical. Différents phénomènes d'interactions laser/tissu [10-11] tels que l'hyperthermie, la coagulation et la volatilisation sont utilisés dans les applications médicales et dépendent directement des paramètres de stimulation du laser (longueur d'onde, puissance, temps et mode d'émission, forme du faisceau et surface du spot) et du tissu à traiter (coefficients optiques, paramètres thermiques et coefficients de la réaction de dénaturation thermique) Cette méthode de stimulation de la rétine n'a pas fait l'objet d'étude à notre connaissance. Néanmoins, d'autres équipes ont réalisé des études qui ont pour thème l'interaction d'un laser infrarouge avec le tissu vivant et plus spécialement, de cellules nerveuses du nerf sciatique [12] et auditif [13]. Ils ont ainsi montré qu'une excitation infra rouge de 1,8 ou 2,1µm permettait d'avoir une énergie suffisamment absorbée pour exciter le neurone sans risquer de l'endommager.

Nous présentons dans ce papier deux séries d'expériences mises en place pour vérifier qu'en l'absence de photorécepteurs, il est possible de stimuler les cellules ganglionnaires de la rétine avec une excitation laser infrarouge. Dans une première série d'analyse in vitro, nous enregistrons par la technique du patch clamp la réponse de cellules ganglionnaires d'une tranche de rétine de souris, excitées directement par le laser focalisé. La deuxième série d'analyse in vivo en préparation permettra l'étude de la rétine d'un œil de souris endormie. Nous détaillons chaque résultat en indiquant les paramètres de stimulation utilisés. En effet, c'est en faisant varier les paramètres géométriques, fréquentiels, temporels ou énergétiques du signal de commande du laser que nous optimisons l'interaction du faisceau avec le neurone.

Nous présentons enfin les travaux futurs qui doivent permettre d'étudier le transfert de l'information jusqu'au cerveau.

II. TECHNIQUES DE PRISE DE MESURE.

Deux techniques de mesures ont été développées en collaboration avec l'INM. Nous avons, d'une part, utilisé la technique du patch Clamp qui permet de vérifier la génération d'un potentiel d'action. D'autre part, la technique de détection de potentiel évoqué permet de contrôler que les potentiels générés sont correctement interprétés par le cerveau.

II.1. Le patch clamp (*in vitro*).

Il s'agit de la technique permettant l'étude électrophysiologique *in vitro* d'une cellule. On mesure l'activité d'une cellule par le nombre d'ions spécifiques qui traversent sa membrane en créant un courant ionique pouvant être mesuré. Le passage de ces ions s'effectue par des canaux spécialisés présents sur la totalité de la surface de la cellule. Cette méthode consiste à sceller sur la membrane plasmique, la pointe micrométrique d'une électrode en verre creuse contenant une solution identique au milieu extracellulaire.

II.2. Le potentiel évoqué (*in vivo*).

Le Potentiel Evoqué Visuel par flash (PEV flash) est la réponse électrique du cortex, provoquée par une stimulation visuelle. Il est possible par cette technique d'observer la conduction des voies visuelles. Il est enregistré au moyen d'électrodes placées en regard du cortex occipital. L'étude du PEV par flash est réalisée sur un œil avec une pupille dilatée pour permettre l'éclairage uniforme de la rétine. La forme du PEV et la latence de ses différents composants renseignent sur l'état des voies de conduction visuelle (Fig. 1).

II.3. Modèle animal.

L'INM possède un appareil permettant ce type de mesure sur des animaux. La souris a été finalement préférée pour réaliser les tests malgré la taille réduite de son globe oculaire qui rend l'opération plus délicate en termes d'accessibilité et de visibilité. La pertinence de notre choix réside dans la possibilité de réaliser des tests sur un modèle animal possédant une pathologie de dégénérescence rétinienne collant parfaitement à notre problématique. En effet, il est possible (exclusivement avec des souris) de disposer de sujets génétiquement modifiés (Retinal degeneration 10 (rd10) à mice). Ces animaux, naissent avec des yeux normaux qui leur permettent de passer par les étapes d'apprentissage de la vision pendant les dix premiers jours de leur vie. Une fois passé ce délai, ils développent une dégénérescence progressive des photorécepteurs de leur rétine qui entrainera la cécité. Travailler sur de tels animaux permettra d'observer des réponses qui prendront en compte la destruction des photorécepteurs.

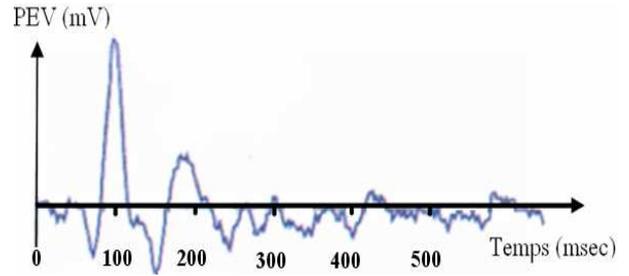


Fig. 1. Le Potentiel Evoqué Visuel. < 75 ms : réponse sous corticale du CGL, 100 ms : réponse corticale (aire V1), > 135 ms : cognition visuelle [O. Payer- INM]

III. BANCS EXPERIMENTAUX.

Chaque série de mesures est effectuée avec un laser moyen infra rouge de longueur d'onde λ égale à $2,12\mu\text{m}$ d'une puissance optique maximum de 16 mW fabriqué par l'équipe Composants à NANOstructures pour Moyen Infrarouge (NANOMIR) du laboratoire IES de l'Université Montpellier 2 (UM2).

III.1. Tests *in vitro*

Ce système permet la visualisation et la prise de mesure directe en temps réel des variations du potentiel membranaire d'une cellule ganglionnaire (CGR) située dans une tranche de rétine. Il peut être scindé en deux parties distinctes (Fig. 2).

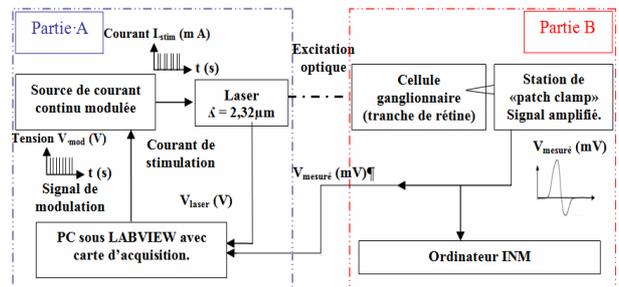


Fig. 2. Banc de mesure *in vitro*

La partie « A » permet de piloter le laser de stimulation et d'enregistrer la réponse du neurone. Elle se compose d'un ordinateur équipé du logiciel de programmation LABVIEW et d'une carte d'acquisition. Cette dernière permet de commander la source de courant continu modulée en fréquence qui alimente le laser et de sauvegarder le signal mesuré par la station patch clamp. La partie « B » permet de mesurer l'activité électrique de la cellule et d'amplifier ce signal.

III.2. Test *in vivo*.

Le test *in vivo* en cours de réalisation permettra de stimuler la rétine d'un animal vivant avec le laser et de détecter les potentiels évoqués au niveau du cortex visuel.

Pour pallier au problème résultant de la grande absorption des rayons infrarouges par les couches successives de l'œil (corné, cristallin, corps vitré ...), le laser est couplé à une fibre optique que l'on viendra insérer dans l'œil de la souris. Pour la protéger et faciliter son insertion dans l'œil, la fibre optique est introduite au préalable dans une seringue (Fig.3).

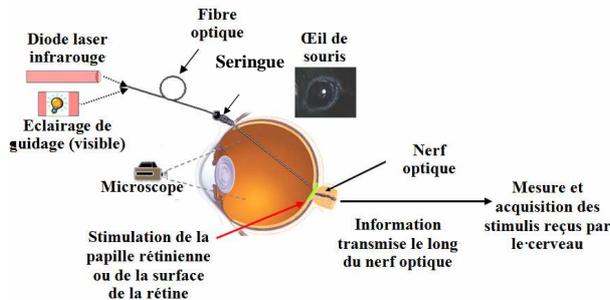


Fig. 3. Banc de mesure in vivo

L'absorption du corps vitré à notre longueur d'onde n'est pas négligeable. Pour diminuer cet effet d'absorption et réaliser une excitation suffisante, nous allons placer la sortie de notre fibre au plus près de la rétine. De plus, un éclairage de guidage (spectre visible) nous permettra de placer notre fibre optique de manière optimale pour nos mesures. Pour réaliser le couplage du laser et de la fibre sélectionnée, un banc optique est en cours de réalisation. La procédure de commande du laser et de l'acquisition des données reste identique à celle utilisée pour la stimulation in vitro.

IV. RESULTATS ET DISCUSSION

Les résultats obtenus portent sur les tests in vitro réalisés à l'INM. Les tranches de rétine, la pose des électrodes et le suivi des différentes étapes du protocole spécifique à ce genre de mesures ont été réalisés par Rima Baroum Tannous. Le protocole de test permet de vérifier que l'électrode de mesure est bien placée et que la « CGR » n'est pas endommagée. La stimulation laser est définie par l'amplitude la forme et la fréquence du signal de commande employé. Pour le test présenté, nous avons réalisé dix stimulations d'une puissance de 6 mW toutes les deux secondes. Le terme « stimulation » correspond à un train d'impulsion (ici 25) à la période (ici 2ms) et au rapport cyclique (ici 1/2) fixés avant le début du test. Appliquer ces paramètres pour la stimulation laser nous a permis d'observer une réponse de la cellule cible sous la forme de potentiels d'actions. (Fig.4). La période du signal d'excitation ne doit pas dépasser 200ms, ce qui permet à la cellule de revenir à sa température initiale en limitant l'accumulation de chaleur pouvant causer une dégradation du tissu ciblé. Les tests réalisés avec des temps de stimulation compris entre 49 et 200ms ont tous entraînés une réponse de la CGR.

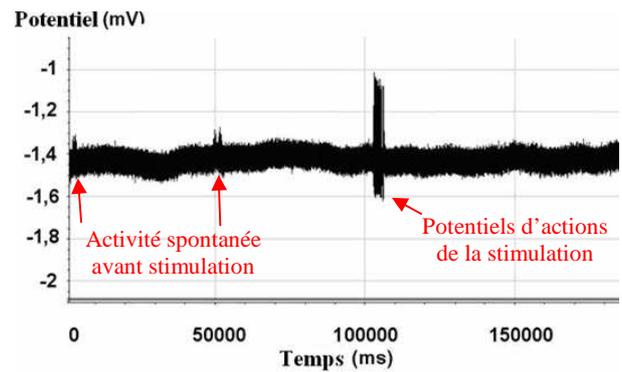


Fig. 4. Potentiel d'action stimulation = f (Temps) pour une stimulation de 49ms.

Les temps inférieurs à 49ms n'ont eu aucun effet. Un calcul de la densité d'énergie transférée montre que ce temps de stimulation correspond à une valeur seuil approximative de $0,3J.cm^{-2}$. Ces valeurs coïncident aux valeurs seuils de déclenchement relevées aux cours des études réalisées sur les cellules nerveuses auditives et du nerf sciatique [12-14]. Un signe de « bonne santé » de la cellule est de pouvoir visualiser les potentiels d'actions spontanés que les CGR envoient périodiquement au cerveau. Ces relevés sont effectués en début et en fin de test. En début, ils servent à vérifier le bon positionnement de l'électrode de mesure et le bon fonctionnement de la cellule avant de commencer les tests. Ceux réalisés une fois les tests terminés permettent une comparaison avec ceux obtenus avant la stimulation. En étudiant les relevés des potentiels d'actions spontanés (Fig.5), nous avons observé un phénomène qui n'apparaît pas lors d'une stimulation électrique classique réalisée avec l'électrode de la station patch clamp. L'effet constaté est une amplification de tous les potentiels spontanés observés après la stimulation. Le rapport entre les deux relevés est compris entre 1,6 et 2,5 suivant les temps de stimulation utilisés. Il semble que l'amplitude atteinte de manière spontanée corresponde à celle atteinte lors des potentiels d'actions de stimulation. Si l'on ne peut avancer de réponse définitive à ce phénomène, la première possibilité serait que la cellule a subi des dommages et que ces réponses spontanées en sont les signes. La deuxième possibilité serait un « effet mémoire » de la cellule. Celle-ci, après avoir subi une excitation laser, générerait des potentiels d'actions spontanés d'amplitude équivalente à ceux observés pendant la stimulation. Si la faisabilité d'une telle stimulation est maintenant prouvée, il reste à expliquer quels mécanismes d'interactions laser/tissu sont en cause. Plusieurs hypothèses sont avancées : contribution électromagnétique, photo chimique, photomécanique ou thermique [14]. Des tests futurs sont prévus pour étudier ces mécanismes d'interactions.

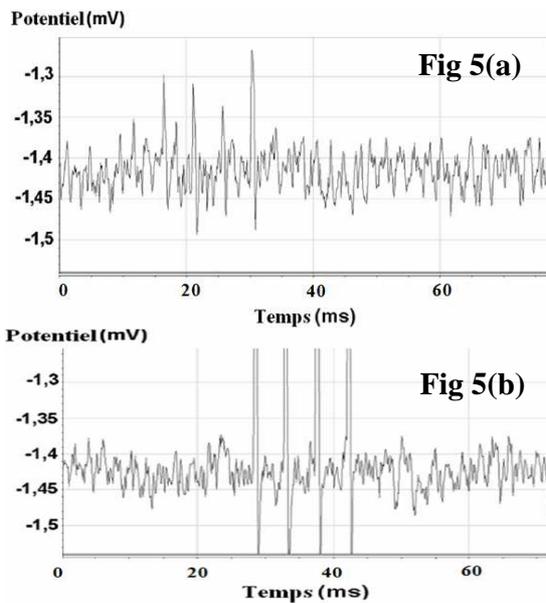


Fig. 5. Potentiel d'action spontané; (a) avant stimulation laser Fig. 5(a); (b) après la stimulation Fig. 5(b).

V. CONCLUSION.

Cet article présente une nouvelle méthode de stimulation d'une rétine dépourvue de photorécepteurs. Les deux séries d'expériences réalisées pour étudier la faisabilité d'une stimulation laser pulsée infrarouge des neurones de la rétine autres que les photorécepteurs y sont décrites. Les résultats portant sur les tests *in vitro* démontrent que la cellule excitée par laser réagit en générant un potentiel d'action. La série d'analyse *in vivo* en cours sur l'excitation laser de la rétine d'un œil de souris endormie devrait nous permettre de confirmer que les potentiels électriques générés au niveau des cellules ganglionnaires sont bien transférés jusqu'au cerveau. Des études supplémentaires vont également être réalisées pour comprendre les phénomènes observés et pour optimiser le signal de stimulation

VI. REMERCIEMENT.

Ces recherches sont financées par la Fondation « EADS » (European Aeronautic Defence and Space company foundation). Nous remercions les autres membres de l'équipe « STINIM » : Isabelle Marc, Fabrice Bardin, Jean-Claude Borie et Guillaume Tatur. Nous remercions également nos collaborateurs sur le projet LASDOP : Christian Amel, Christian Chabbert, Olivier Payet, Agnès Muller, Rima Baroum Tannous et Guy Lenaers de l'INM. Yves Rouillard et Aurore Vicet de l'UM2.

REFERENCES.

[1] R. A. Normann, E. M. Maynard, P. J. Rousche, D. J. Warren, "A neural interface for a cortical vision

prosthesis", dans *Vision research* 39, 1999, 2577-2587.

- [2] M. S. Humayum, E. de Juan, Jr, J. D. Weiland, G. Dagnelie, S. Katona, R. Greenberg, S. Suzuki, "Intraocular retinal prosthesis", dans *Tr.Am.Ophth.Soc*, 2001, Vol 99.
- [3] J.F.Rizzo, J.Wyatt, J. Loewenstein, S. Kelly, D. Shire, "Perceptual efficacy of electrical stimulation of human retina with a microelectrode array during short term surgical trials", dans *Investigative Ophthalmology & Visual Science*, 2003, 44, 12.
- [4] N. H. Lovell, J. Hallum, E. Chen, et al, "Advances in retinal neuroprosthetics", dans *Handbook of Neural Engineering*, 2007, Akay (ed), Wiley IEEE press.
- [5] R. Eckmiller, D. Neumann, O. Baruth, "Tunable retina encoders for retina implants: why and how", *J Neural Engineering*, 2005, 2 S91.
- [6] E. Zrenner, K. D. Millicez, V. P. Gabel et al "The development of subretinal microrphotodiodes for replacement of degenerated photoreceptors", dans *Ophthalmic*, 1997, Res. 29: 269.
- [7] J. Delbeke, M. C. Wanet-Defalque, B. Gerard, "The microsystems based visual prosthesis for optic nerve stimulation", dans *Artificials organs*, 2002, 26(3) 232.
- [8] W. H. Dobelle, "Artificial vision for the blind by connecting a television camera to the brain", dans *ASAIO Journal*, 2000, 46:3-9.
- [9] G. J. Suaning, L. E. Hallum, S. Chen, P. J. Preston, N. H. Novell, "Phosphene Vision: development of a portable visual prosthesis system for the blind", dans *IEEE EMBS*, 2003, Cancun, Mexico, 17-21.
- [10] M. Niemz, "Laser-tissue interactions: fundamentals and applications", dans Ed: Springer, 2002.
- [11] J. M. Brunetaud, S. Mordon, T. Desmettre et C. Beacco, "Les applications thérapeutiques des lasers", EDP Sciences, 2002.
- [12] J. Wells, Kao, Mariappan, Albea, D. Jansen, Konrad, Mahadevan-Jansen, "Optical stimulation of neural tissue *in vivo*", dans *Optics letters*, 2005, 30, 5, 504.
- [13] A. D. Izzo, E. Duco. Jansen, J. Y. Walsh, "Laser Stimulation of Auditory Nerve", dans *Lasers in Surgery and Medicine*, 2006, 38:745-753.
- [14] J. Wells, C. Kao, P. Konrad, A. Mahadevan-Jansen, and E.D. Jansen, "Biophysical mechanisms responsible for pulsed low-level laser excitation of neural tissue", dans *Biophys J BioFAST*, 2007 doi:10.1529/biophysj.107.104786.

CONCEPTION D'UN SYSTEME DE VISION PAR PHOSPHENES

GUILLAUME TATUR¹ (DOCTORANT 1^{IERE} ANNEE), MICHEL DUMAS² (DIRECTEUR DE THESE)

12S LGI2P, Ecole des Mines d'Ales, équipe STINIM (Systèmes de Traitement de l'Information Neuronales et Ingénierie Médicale)
Site Eerie Parc Scientifique et Technique Georges Besse, 69 rue Georges Besse 30035 NÎMES

¹GUILLAUME.TATUR@EMA.FR

²DUMAS@UNIV-MONTP2.FR

[HTTP://WWW.EDI2S.UNIV-MOTNP2.FR](http://WWW.EDI2S.UNIV-MOTNP2.FR)

Résumé— Cette thèse s'inscrit dans le cadre plus général de la réalisation d'une prothèse visuelle destinée aux personnes atteintes d'une maladie dégénérative des cellules photoréceptrices de la rétine (« Retinitis Pigmentosa », DMLA). Les travaux consisteront à l'extraction et à la mise en forme adéquate de données issues de capteurs optiques pour l'établissement d'une vision cohérente basée sur la génération de phosphènes. Les équipes travaillant sur différents types de prothèses ainsi que les recherches sur le contenu informationnel de la vision prothétique, proposent à quelques exceptions près, une simple diminution de la résolution des images issues d'une camera. L'objectif va donc être de fournir une vision axée sur les problèmes que pose la navigation, tout en exploitant au mieux l'utilisation d'un moyen de stimulation original pour l'excitation des voies optiques. Nous travaillons actuellement sur un mode de représentation symbolique basé sur le développement d'un algorithme d'analyse invariant à l'illuminant. Cet algorithme servira notamment à la détection de marqueurs couleurs.

Mots clés— prothèses visuelles ; navigation ; constance chromatique; analyse d'images couleurs.

I. INTRODUCTION.

Le projet LASer Driven Optical Prosthesis (LASDOP) a pour objectif la création d'une prothèse visuelle destinée aux personnes atteintes d'une maladie dégénérative des cellules photoréceptrices de la rétine. Ce projet tire son originalité du fait que la stimulation des voies optiques se fait, non par injection de courants par le biais d'électrodes mais, par exposition de la fibre nerveuse à un signal laser. Une autre thèse en cours étudie la stimulation par le laser des cellules nerveuses de la couche interne de la rétine. L'étude qui nous intéresse ici devra fournir aux patients une vision pertinente de l'environnement basée sur la perception de phosphènes. Elle devra permettre de retrouver une mobilité, une autonomie dans les tâches courantes de navigation en milieux connus ou non. Cet article propose tout d'abord un rappel en physiologie de la perception humaine puis nous verrons succinctement les différents types de prothèses et le choix de vision prothétique par

différentes équipes. Enfin, nous expliciterons l'orientation initialement choisie pour ce début de thèse.

II. RAPPEL EN PHYSIOLOGIE DE LA PERCEPTION.

Dans cette partie nous rappelons dans les grandes lignes et dans la perspective qui nous intéresse, les étapes de la transduction rétinienne. Ensuite, nous décrirons succinctement l'une des maladies dégénératives de la rétine : la rétinite pigmentaire. Enfin, nous fournirons quelques éléments d'informations concernant les phosphènes.

II.1. La transduction rétinienne.

La rétine (*Fig.1*) est une structure du fond de l'œil organisée en couche de cellules dont la partie externe est constituée de cellules photoréceptrices qui transforment l'énergie lumineuse en potentiel de membrane. Les cônes, grâce à leurs photo pigments spécifiques, sont sensibles à trois longueurs d'ondes : rouge, bleu ou vert. Les bâtonnets sont constitués d'un seul type de pigment et possèdent une haute sensibilité en faible éclaircissement. Ainsi, lors de l'absorption de photons par les pigments des photorécepteurs ces derniers subissent un changement de conformation qui va induire une variation du potentiel membranaire proportionnellement à l'intensité reçue. Si nous remontons le long de l'axe optique nous aboutissons à la couche la plus interne constituée des cellules ganglionnaires, premiers neurones de la chaîne visuelle. La convergence de leurs axones forme le nerf optique qui s'insère dans une zone du fond de l'œil appelé papille optique. Ce sont les cellules ganglionnaires qui transmettent l'information issue des photorécepteurs vers le cortex visuel. Enfin, c'est le passage des résidus de la phototransduction dans l'épithélium pigmentaire qui permet de renouveler les éléments photosensibles et donc leurs retours à un état excitable.

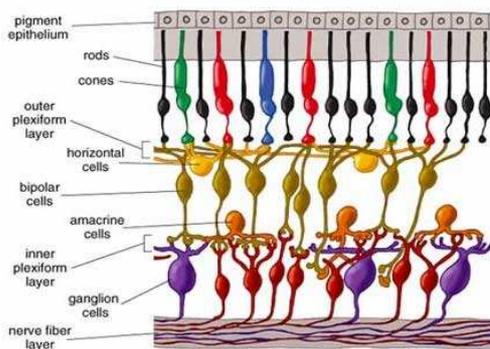


Fig. 1. Schéma d'un coupe de rétine.

Credit: Helga Kolb, New Scientist, vol. 91, janvier 2003.

II.2. Rétinite pigmentaire.

Il s'agit d'une maladie héréditaire de l'œil impliquant une dégénérescence rétinienne progressive périphérique puis centrale dans 85 % des cas et se caractérisant principalement par une héméralopie¹ ainsi qu'une atrophie de la papille. 35000 personnes en sont atteintes en France et environ 400000 en Europe². En effet, l'épithélium ne jouant plus son rôle de traitement des déchets de la transduction, ceux-ci s'accumulent et progressivement les cellules deviennent inopérantes. L'atrophie de la papille optique à pour cause la dégénérescence des cellules ganglionnaires qui ne sont plus sollicitées. Pendant une durée relativement longue le taux de survie des cellules ganglionnaires reste élevé. Ainsi cela combiné à leurs relatives accessibilités fait de ces dernières des cibles intéressantes pour la méthode de stimulation choisie.

II.3. Les phosphènes.

L'apparition de phosphènes est un phénomène entoptique³ caractérisé par la sensation de percevoir de la lumière sous forme de points ou zones lumineuses. Elle est causée par une stimulation mécanique, électrique ou magnétique de la rétine ou du cortex visuel. Volta (1745-1827) avait constaté que sa pile pouvait, au contact de l'œil, provoquer des sensations lumineuses. Cette idée fut reprise notamment dans [1] pour vérifier l'état du nerf optique et estimer si une implantation épi rétinienne est possible. Ainsi la stimulation du nerf optique et *a fortiori* des cellules ganglionnaires, peut permettre de rétablir une forme de perception dont l'élément de base est le phosphène.

III. LA VISION PROTHETIQUE.

Comme nous l'avons vu au II.3, il existe différentes manières de stimuler les voies visuelles mais l'essentiel des travaux sur les prothèses visuelles portent sur la stimulation électrique par le biais de microélectrodes.

III.1. Les différents types d'implants.

Nous distinguons plusieurs types d'implants.

- L'implant cortical et l'implant intra cortical se présentent sous la forme d'une matrice disposée au niveau de l'aire visuelle primaire. Son avantage est qu'il passe outre les anomalies liées aux diverses maladies de la vision en amont du cortex visuel. Malheureusement l'équivalence entre la position de la zone stimulée est le lieu d'apparition du phosphène dans le champ visuel est mal connue.
- Les implants sous rétinien sont constitués d'un réseau d'électrodes se substituant aux photorécepteurs en stimulant la couche de cellules en aval de ceux-ci.
- La stimulation du nerf optique au même titre qu'un implant épi rétinien, stimule directement les cellules ganglionnaires et permet d'éviter la couche de photorécepteurs défectueux ce qui est le cas dans les maladies qui nous intéressent. De plus la topographie des fibres nerveuses est telle qu'il est possible de lier positions de stimulations et positions induites dans le champ visuel. C'est cette dernière solution qui, au regard de nos connaissances et hypothèses actuelles, semble la plus adéquate.

III.2. Les représentations de l'environnement.

Les études de CHA *et al* [2] sont souvent citées en référence que ce soit sur les performances en acuités ou encore en navigation suivant les caractéristiques de la vision fournie. Il est à noter que ces études sont basées sur des tests sur personnes voyantes dont la vue est artificiellement altérée. Les paramètres testés sont principalement le nombre de points lumineux ou phosphènes ainsi que la taille du champ visuel. L'équipe conclut que des résultats optimaux en navigation sont obtenus pour un champ visuel de 30° et une image, de 25x25 pixels uniformes régulièrement espacés, capturée par une simple camera. Ces résultats se retrouvent également dans [3] qui propose 500 points de stimulations pour réaliser la plupart des tâches de mobilité. Augmenter ce nombre jusqu'à 1000 semble améliorer la sensation de sécurité. Dans [4] (Fig.2), il est question de tests réalisés sur des patients possédant des implants corticaux. Le capteur est une camera miniature dont les images sont transformées puis transmises au stimulateur. Leur espace de tests permet de mesurer les performances en acuité et dans des tâches de suivi de lignes, de lumières au plafond ou encore de personnes dans une salle. Les développements en cours sont l'utilisation de filtres de Sobel pour la détection de contours significatifs et l'affichage de l'information de distance par modulation de l'intensité du phosphène.

¹ Forte difficulté à voir la nuit.

² Source: La Rétinite pigmentaire (faf.asso.fr).

³ Excitation du système visuel avec un stimulus autre que la lumière

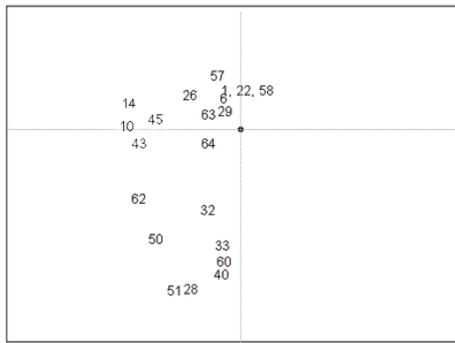


Fig. 2. Cartographie des phosphènes, dans le champ visuel d'un patient, obtenue expérimentalement dans [4].

Dans [5] les auteurs utilisent la puissance de calcul d'un simple PDA⁴ équipé d'une camera pour analyser l'environnement ce qui implique quelques concessions et prétraitements : conversion de l'image camera (capture 160x120 pixels) en image faible résolution (32x16 pixels) avec seulement 8 niveaux de gris. L'analyse consiste en la segmentation en zones de niveaux de gris homogènes et suivi temporel de leurs expansions. Leurs résultats sont très variables suivant les conditions d'éclairage et ce manque de robustesse les rend difficilement exploitables. Dans [6] les auteurs se basent sur un implant prenant la forme d'un anneau composé de quatre électrodes autour du nerf optique. Les paramètres de stimulation déterminés expérimentalement, leur permettent de générer des phosphènes de manière reproductible dans certaines parties du champ visuel. Ainsi, ils utilisent la connaissance de cette carte de stimulation en la superposant à un image prétraitée issue d'une camera pour sélectionner les points de stimulations. Chaque point est activé individuellement et aléatoirement toutes les 80 ms jouant ainsi sur l'intégration temporelle des stimulations. Enfin, une autre équipe [7] importante dans ce domaine a choisi, à l'instar de la plupart des équipes, de représenter l'environnement sous la forme d'une matrice de points. La différence est qu'ils considèrent un bruit sur les caractéristiques des phosphènes. Ce bruit a été modélisé selon les observations subjectives de patients implantés. De plus, une défaillance aléatoire de certaines électrodes est prise en compte. Il faut souligner l'importance de l'apprentissage dans toutes ces expériences car en effet, l'amélioration des performances est liée à l'adaptation de notre perception à l'information, extrêmement appauvrie, qu'offre un implant.

IV. ORIENTATIONS CHOISIES

Nous avons choisi de travailler sur deux axes en parallèle : premièrement l'étude d'une représentation pertinente axée sur la stimulation idoine des cellules ganglionnaires. En effet, nous pensons dans un premier temps nous baser sur un modèle biomimétique de rétine. Ce modèle pourrait permettre de déterminer les paramètres de stimulation

adéquats c'est-à-dire porteurs de sens pour le système visuel humain (SVH). Il faut également prendre en compte l'imprécision du positionnement du faisceau laser ainsi qu'un diamètre de faisceau pouvant subir des altérations. L'information que doit contenir l'image en entrée sera discutée avec des spécialistes de la réadaptation fonctionnelle⁵.

Deuxièmement, nous pouvons imaginer dans un premier temps, un mode de représentation symbolique basé sur la détection de marqueur couleur. L'exploitation de l'information couleur dans une scène serait le vecteur d'une quantité d'information plus grande que ne le permet la simple luminance (valeur achromatique synonyme d'intensité). Cependant l'analyse de la couleur nécessite un prétraitement pour ne pas tenir compte des modifications de couleurs lorsque la ou les sources de lumières (*Fig.3*) varient induisant une altération des couleurs perçues par le capteur. Deux approches sont possibles : l'application de techniques de constance chromatique ou bien la recherche d'un espace indépendant de l'illuminant.



Fig. 3. Même objet sous différents illuminants. Crédit : base publique de l'université de Simon Fraser

IV.1. La constance chromatique.

C'est la combinaison de l'émission spectrale de la source et de la réponse spectrale de l'objet qui définit le rayonnement qui parvient à notre œil où il faut alors prendre en compte la réponse du capteur. L'invariance de la couleur perçue par le SVH, en dépit du changement de la couleur de l'illuminant est appelée constance chromatique ou constance de couleur. La constance chromatique implique que le SVH essaie de décrire la réflectance indépendamment de la couleur de l'illuminant. Il semble alors que cette constance soit due à : une adaptation des cônes, une analyse spatiale des couleurs ainsi qu'à une connaissance à priori de la couleur de certaines surfaces (comme la teinte de la peau par exemple). Nous présenterons ici l'essentiel de l'approche basique qu'est la règle d'adaptation de Von Kries, mais il existe d'autres méthodes tel que l'algorithme RETINEX de Land [8] et l'algorithme GRAY WORLD ou plus récemment les approches probabilistes et les réseaux de neurones. Pour Von Kries, l'adaptation du système visuel à différents illuminants est réalisée en ajustant trois coefficients de gain (A, B et C) associés aux trois canaux de couleur (L, M, S)⁶ de manière à ce que les stimuli ($L_w, M_w,$

⁵ Institut ARAMAV : Clinique de réadaptation et rééducations fonctionnelles pour déficients visuels, Nîmes, France.

⁶ Décrit les domaines de sensibilités de canaux : grandes (rouge), moyennes (vert) et courtes longueurs d'onde (bleu).

⁴ Personal Digital assistant

S_w) d'une surface blanche de référence donnent ($L'=1$, $M'=1$, $S'=1$) après correction.

$$\begin{bmatrix} L' \\ M' \\ S' \end{bmatrix} = \begin{bmatrix} A & 0 & 0 \\ 0 & B & 0 \\ 0 & 0 & C \end{bmatrix} * \begin{bmatrix} L \\ M \\ S \end{bmatrix}$$

En réalité la matrice est non diagonale du fait des recouvrements entre les domaines de sensibilités des composantes. Ces méthodes ont le désavantage de ne pas prendre en compte les variations locales d'illuminant et nécessitent de fortes hypothèses sur ses caractéristiques. Nous préférons alors l'autre approche qui est l'invariance à l'illuminant.

IV.2. Invariance à l'illuminant.

Devant le problème posé par ces variations de couleurs, certains auteurs [9][10], notamment dans le domaine de l'indexation d'images couleurs, proposent des méthodes qui exploitent des paramètres tels que la relation d'ordre des niveaux RGB d'un pixel (modèle de Hordley [10]) et l'intersection d'histogrammes adaptés. Sans détailler ces méthodes, nous retiendrons le principe qui est de changer de « point de vue » pour l'interprétation invariante de la scène alors que les conditions d'éclairage peuvent changer localement. Suivant cette idée nous avons choisi d'étudier le contenu fréquentiel et spatial chromatique dans l'image et recherchons à optimiser les paramètres invariants tout autant que les matériaux dans le contexte applicatif du marquage.

Un programme MATLAB est en développement ainsi que la réalisation d'un espace de test pour la réalisation de mesures colorimétriques et l'essai des algorithmes.

V. CONCLUSION.

Pour la question de l'information à transmettre et de sa forme nous avons présenté les approches en vision prothétique des principales équipes. Nous avons alors fait le choix d'étudier une stimulation plus adaptée au SVH par le biais, entre autres, des modèles de rétine. Parallèlement, l'extraction invariante à l'illuminant de l'information couleur de l'image est en cours d'étude. Elle permettrait par exemple, la détection robuste de marqueurs autorisant une représentation simple et rapide de l'environnement sous forme symbolique.

VI. REMERCIEMENTS.

Nous remercions les membres de l'équipe STINIM : Isabelle Marc, Dominique Lafon, Jean-claude Borie, Fabrice Bardin et Jean Michel Bec.

REFERENCES.

- [1] J. Delbeke, D. Pins, G. Michaux, MC. Wanet-Defalque, S. Parrini. C. Veraart, «Electrical Stimulation of Anterior Visual Pathways in Retinitis Pigmentosa », dans *investigative ophthalmology & visual science* , janvier 2001, vol 42, 291-297
- [2] K. Cha, K.W. Horch, R.A. Normann, «Mobility performance with a pixelized vision system», dans *Vision research (Vis. res.)*, 1992, vol. 32, 1367-1372.
- [3] J.R. Sommerhalder, A. Perez Fornos, K. Chanderli, L. Colin, X. Schaer, F. Mauler, A.B. Safran and M. Pelizzone, «Minimum Requirements for Mobility in Unpredictable Environments», dans *Invest Ophthalmol Vis. Sci.* 2006, vol 47
- [4] W. H. Dobelle, «Artificial Vision for the Blind by Connecting a Television Camera to the Visual Cortex», dans *ASAIO Journal*, 2000, vol. 46, 3-9.
- [5] J. Dowling, A. Maeder, W. Boles, «A PDA based artificial human vision simulator», dans *Proceedings APRS Workshop on Digital Image Computing*, Brisbane, 2005, 109-114.
- [6] M E Brelén, F Duret, B Gérard, J Delbeke, C Veraart, «Creating a meaningful visual perception in blind volunteers by optic nerve stimulation», dans *J. Neural Eng.*, 2005, 22-28.
- [7] G. Dagnelie, P. Keane, V. Narla, L. Yang, J. Weiland , M. Humayun «Real and virtual mobility performance in simulated prosthetic vision», dans *J. Neural Eng.*, 2007, 92-101.
- [8] E. H. Land, «Recent advances in retinex theory and some implications for cortical applications: Color vision and the natural image», dans *Proc. Nat. Acad. Sci., USA*, 1983, 5163-5169.
- [9] M. Exbrayat, L. Martin, «Calcul et visualisation d'une distance inter enluminures à partir des couleurs», dans *Le Médiéviste et l'ordinateur*, 2006, 45.
- [10] D. Muselet, L. Macaire, P. Bonnet, J. Postaire, «Reconnaissance d'objets grâce à l'analyse des composantes couleur », dans *GRETSI*, 2003.

Développement de nouvelle source laser à émission verticale de haute cohérence.

Alexandre LAURAIN¹ (Doctorant 1e année), Arnaud GARNACHE² (Directeur de Thèse), Mikhaël MYARA³ (co-directeur)

Institut d'Electronique du Sud UMR 5214
Place Eugène Bataillon, Université Montpellier II
34095 Montpellier Cedex 05, France

¹alexandre.laurain@ies.univ-montp2.fr,
²agarnache@univ-montp2.fr, ³myara@opto.univ-montp2.fr

Résumé— Nous présentons le principe de fonctionnement d'un laser à base de semi-conducteur. Nous verrons ensuite les propriétés d'un laser à émission verticale à cavité externe (VECSEL). Ces sources lasers mono-fréquences émettant entre 2 et 3 μm trouvent des applications dans les domaines de l'analyse de gaz et du biomédical. Ensuite, nous définissons les concepts nécessaires à la compréhension des relaxations de contraintes dans les couches épitaxiées. L'étude de ces relaxations nous permet de mieux comprendre les différentes valeurs de gain obtenue. Enfin nous présentons le système de caractérisation par micro photoluminescence réalisé au laboratoire ainsi que les résultats obtenus.

Mots-clés— Laser ; VECSEL ; dislocation ; photoluminescence.

I. INTRODUCTION

Ces trente dernières années, l'utilisation du laser s'est largement répandue dans de multiples domaines d'applications. On peut sommairement classer ces applications en trois catégories [1]. La première est la recherche en laboratoire (spectroscopie, optique non linéaire), la seconde est celle des applications industrielles lourdes (médical, usinage). Le troisième domaine d'utilisation concerne le grand public (CD, lecteur de code barres, télécom). Cet engouement pour le laser s'explique par ses caractéristiques exceptionnelles : monochromaticité, faible divergence, cohérence temporelle et spatiale, très grandes énergies possibles, impulsion ultra courte, possibilité d'accord en longueur d'onde,...

On classe les lasers en fonction de la nature du milieu excité. Parmi les plus utilisés on trouve les lasers à gaz (CO₂), à colorant (rhodamine 6G), à solide (YAG) et à semi-conducteur. C'est ce dernier type de laser qui représente l'immense majorité des lasers utilisés dans l'industrie. En effet, ses avantages sont nombreux : tout d'abord, il permet un couplage direct entre l'énergie électrique et la lumière, d'où les applications en télécommunications (à l'entrée des réseaux de fibres optiques). En plus de la pureté spectrale (une seule longueur d'onde émise), les lasers à semi-conducteur présentent une consommation d'énergie très basse (rendement de l'ordre de 30 à 40%), une grande durée de vie (>10000 h), une grande stabilité de la longueur d'onde (λ_e) et la possibilité de balayer continûment λ_e sur de large plage [2]. Le transport d'information par

fibre optique (1.55 μm) et la lecture/écriture de données sur cédérom (0.78 μm et 0.6 μm pour les DVD) sont les deux principaux secteurs qui ont permis le développement de ce type de composant. D'autres applications (militaires, médicales, spectroscopie) ont néanmoins vu le jour grâce au développement de laser émettant dans le proche et le moyen-infrarouge (0.7 μm -10 μm). Les applications militaires concernent à la fois le guidage laser de missile et la réalisation de dispositifs de contre-mesure avec des lasers de fortes puissances. Dans le domaine médical, l'utilisation de laser infrarouge permet de réaliser des opérations de l'oeil (myopie) avec une grande précision tout en cautérisant la plaie. Enfin l'application à la détection de gaz de haute sensibilité a vu le jour avec l'avènement de ce type d'émetteur de lumière dans cette gamme de longueur d'onde. Effectivement l'absorption des molécules gazeuses y est particulièrement forte en comparaison à la gamme visible du spectre de la lumière. C'est principalement, cette application que nous visons à l'IES. Dans la première partie, nous définirons le principe de fonctionnement d'un laser. Puis nous présenterons l'intérêt des lasers à émission verticale à cavité externe pour la détection de gaz. Enfin, nous détaillerons les processus de relaxation dans les structures réalisées ainsi que l'outil de caractérisation qui a été développé.

II. PRINCIPE DE FONCTIONNEMENT DU LASER

A. Interaction rayonnement matière

Le mot **laser** est l'abréviation de Light Amplification by Stimulated Emission of Radiation. Un laser est donc un système qui génère et amplifie la lumière tout comme un transistor amplifie un signal électronique. Pour réaliser cette amplification, on exploite les propriétés d'interaction de la matière avec un rayonnement. Les trois types d'interactions fondamentales sont l'absorption, l'émission spontanée et l'émission stimulée. (Fig.1). Les électrons gravitant autour d'un atome peuvent occuper plusieurs états d'énergie différent, s'il occupe un état qui correspond à l'énergie minimale (bande de valence pour un semi-conducteur), on dit que l'atome est à l'état fondamental. Si au contraire, il occupe un état d'énergie supérieure (bande de conduction), on dit qu'il est dans un état excité. Si le

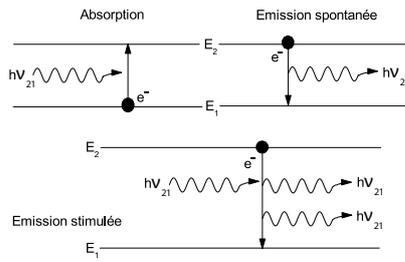


Fig. 1. présentation des différents processus d'interaction

passage de l'état fondamental (E_1) à l'état excité (E_2) se traduit par l'absorption d'une énergie électromagnétique, on dit qu'il y a absorption du photon d'énergie $h\nu = E_2 - E_1$.

Considérons maintenant un atome dans un état excité, l'énergie qu'il a acquise pourra être perdue spontanément en retombant sur un niveau inférieur. L'énergie perdue peut soit être transformée en chaleur (thermalisation), soit donner lieu à l'émission d'un photon d'énergie $h\nu = E_2 - E_1$. Mais cette émission spontanée n'a pas lieu simultanément pour tous les centres excités. L'émission spontanée est incohérente [2] : chaque atome émettant indépendamment des autres, à un instant quelconque, avec un état de phase initial aléatoire et pendant une durée très brève. Il n'y a pas de relation de phase, de direction ou de polarisation entre toutes ces émissions, le rayonnement est omnidirectionnel. Seule la fréquence est fixée par le gap.

Enfin le phénomène d'émission stimulée, découvert au niveau théorique par Einstein en 1917, correspond à la libération d'un photon par retombée d'un atome excité à un niveau d'énergie inférieur, mais cette fois provoqué par un autre photon. Ce nouveau photon aura la même fréquence que celui qui lui a donné naissance. Cette émission stimulée possède des propriétés remarquables : les deux rayonnements considérés ont la même phase, la même direction et la même polarisation, c'est un phénomène de duplication de photon. La lumière émise est cohérente, c'est donc une amplification de lumière par émission stimulée de rayonnement.

B. Conditions laser

La première condition laser est l'inversion de population. Pour qu'il y ait une véritable amplification du rayonnement, il faut que la probabilité d'émission stimulée soit plus grande que la probabilité d'absorption. Autrement dit, il faut qu'il y ait plus d'électron dans l'état excité que dans l'état fondamental. Or, à l'équilibre thermodynamique, cette condition est impossible, il est donc nécessaire d'exciter ces électrons par un pompage extérieur. Ce pompage peut être électrique ou optique.

La deuxième condition laser est l'oscillation. Jusqu'à présent, nous avons vu les conditions d'amplification, mais nous voulons utiliser le laser comme source de rayonnement. Nous devons donc transformer cet amplificateur en oscillateur. Pour cela, on utilise une cavité résonnante (cavité Fabry-Pérot) constituée de deux miroirs plan (ou concave). Le laser oscille sur les modes de cette cavité, elle détermine en grande partie les propriétés du faisceau émis. Donc en résumé, pour obtenir un laser il faut un mi-

lieu amplificateur, un dispositif de pompage et une cavité résonnante (Fig.2).

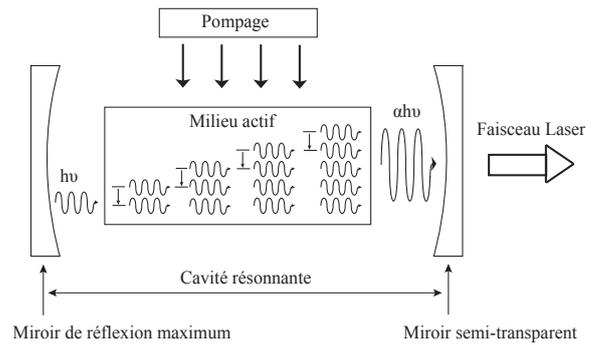


Fig. 2. Présentation des éléments d'un laser

III. LASER À CAVITÉ EXTERNE ÉMETTANT PAR LA SURFACE

A. Fabrication du composant

On va s'intéresser ici à un type particulier de laser à semi-conducteur. Il s'agit d'un laser dont la cavité est verticale, c'est à dire que la propagation de la lumière se fait perpendiculairement au plan des couches épitaxiées (Fig.3). Nous avons vu que pour fabriquer un laser il faut une zone active (milieu amplificateur), et une cavité résonnante (miroirs). Dans ce type de structure, l'épaisseur de la zone active est de l'ordre de quelques microns, donc le rayonnement qui se propage à travers ces couches n'est pas beaucoup amplifié ($< 5\%$) [3]. Contrairement aux lasers à émission par la tranche qui ont un gain beaucoup plus fort (zone active de quelques centaines de μm), les VECSEL doivent compenser cette faible amplification par l'utilisation de miroirs très réfléchissants. Le rayonnement traversera ainsi plus de fois le milieu amplificateur avant d'être extrait de la cavité. Pour réaliser ces miroirs de grande réflectivité ($> 99\%$), on dépose une succession de couches à fort contraste d'indice et d'épaisseur $\lambda/2$, on nomme cette structure miroir de Bragg. Ainsi, le rayonnement de longueur d'onde λ interfère constructivement après réflexion sur les couches. le nombre de couches que l'on dépose dépend de la réflectivité que l'on veut obtenir.

La structure d'un VECSEL est composée d'un miroir de Bragg inférieur, une zone active à base de puits quantiques et d'un miroir concave situé dans l'air à quelques millimètres ou centimètres de la surface (Fig.3).

B. Propriétés du VECSEL

Un avantage du laser à émission par la surface vient du très faible volume du milieu amplificateur ($< 20 \mu\text{m}^3$), qui permet de diminuer la densité de puissance nécessaire pour atteindre le seuil laser (inversion de population et compensation des pertes interne). Grâce à une surface émettrice de quelques micromètres de diamètre, le faisceau de sortie est circulaire TEM_{00} et de quelques degrés de divergence. Mais les propriétés les plus intéressantes du VECSEL sont liées à la nature de sa cavité. En effet, pour une cavité externe, le résonateur est de type plan-concave qui est naturellement stable (contrairement aux cavités plan-plan). La condition pour que le faisceau dans la cavité soit sur un mode stable,

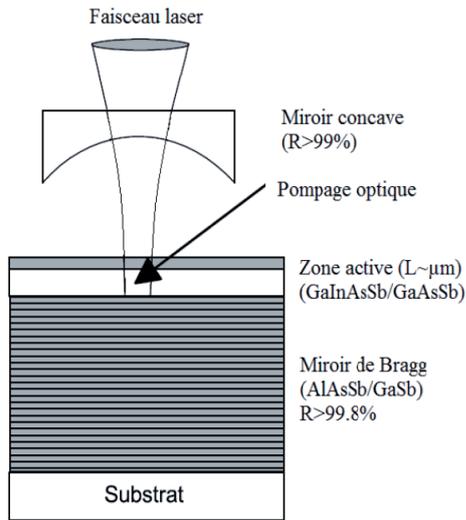


Fig. 3. Illustration d'un laser à émission verticale à cavité externe

est que le front d'onde se superpose à lui-même après un aller retour. Ainsi, il est nécessaire que le rayon de courbure du faisceau (gaussien) coïncident avec le rayon de courbure des miroirs de la cavité. C'est pour cette raison que dans le cas des lasers à émission par la tranche, où la cavité est de type plan-plan, le laser n'est pas sur un mode stable car les rayons de courbure du faisceau gaussien ne sont pas plans. Les modes longitudinaux et transverses du VECSEL sont fixés par la cavité externe. Il existe un grand nombre de modes longitudinaux coïncidant avec la courbe de gain (Fig.4).

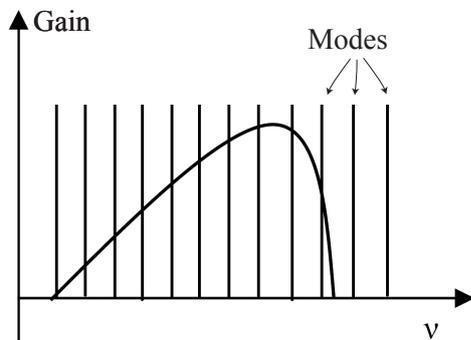


Fig. 4. Courbe de gain et modes longitudinaux dans un VECSEL

De plus, dans ce type de cavité la durée de vie du photon est très importante (10 ns), la dynamique de ce type de laser est donc beaucoup plus stable (pas de saut de mode). La cavité est plus longue (quelques mm) et n'amplifie pas les émissions spontanées, l'onde lumineuse est par conséquent de haute cohérence. Cela engendre une très faible largeur de raie d'émission et une faible sensibilité au retour inverse de la lumière. De plus, la cavité est découplée du matériau à gain, le VECSEL est alors très peu sensible aux fluctuations de la puissance de pompage ou de température de la zone active. Ceci permet d'obtenir de larges plages d'accordabilité continue (> 250GHz) et de fortes puissances de sortie avec un faisceau monomode transverse [3]. Par ailleurs, ce type de cavité permet d'insérer des éléments intra-cavité (filtres, gaz,...), faisant des VECSELS de bons candidats pour la spectroscopie intra-cavité ou la réalisation de la-

sers monofréquences largement accordables.

IV. RELAXATION DE CONTRAINTE DANS LES COUCHES ÉPITAXIÉES

A. Epaisseur critique

Pour optimiser les performances du VECSEL et ainsi satisfaire aux exigences liées à l'application de détection de gaz, il est nécessaire d'augmenter le gain de nos structures. On montre que pour augmenter le gain, il faut mettre plus de puits dans la zone active. Cela pose un problème car le matériau utilisé pour fabriquer les puits n'est pas en accord de maille sur le substrat. Si on augmente trop l'épaisseur des couches que l'on dépose sur le substrat, on atteint l'épaisseur critique. Au delà de cette épaisseur, la contrainte liée au désaccord de maille (1.6%), ne peut plus être relaxée élastiquement, il apparaît alors des dislocations qui dégradent les propriétés optiques des couches épitaxiées (Fig.5).

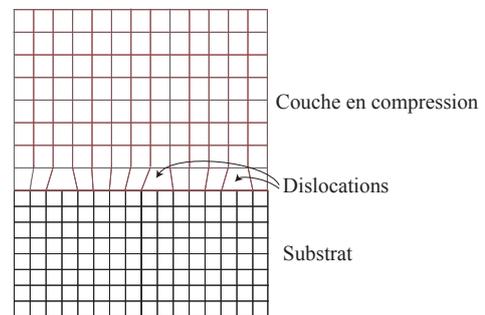


Fig. 5. désaccord de maille entre le substrat et la couche épitaxiée

Il n'existe pas de modèle théorique clairement défini permettant de prédire la valeur de cette épaisseur critique de manière générale. En effet, les modèles diffèrent suivant la nature des dislocations générées et des matériaux utilisés. Toutefois nous avons adapté un de ces modèles [4] à nos conditions de croissance. La confrontation du modèle avec les résultats expérimentaux confirme le bon accord entre les deux (Fig.6).

Epaisseur critique en fonction de la contrainte

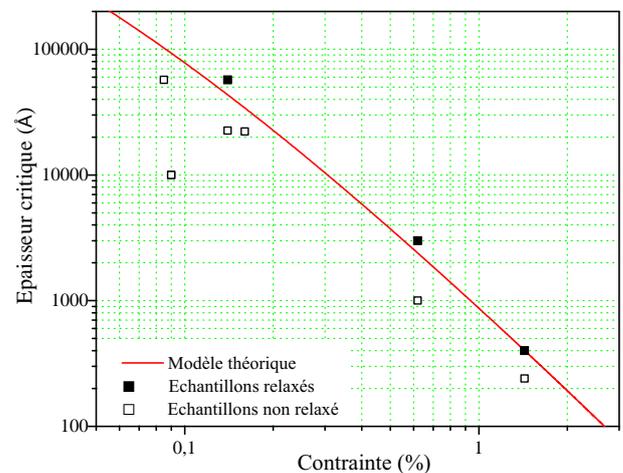


Fig. 6. Modèle théorique et points expérimentaux de l'épaisseur critique en fonction de la contrainte.

La relaxation d'une couche entraîne la formation de lignes de dislocation dans le plan des couches. Le fonctionnement du laser peut être perturbé par ces dislocations notamment si leur densité devient importante. Or, il n'existe aucun moyen de caractérisation simple pour évaluer cette densité. C'est la raison pour laquelle nous avons développé un outil de caractérisation par micro photoluminescence.

B. Caractérisation par micro-photoluminescence

Lorsqu'un échantillon est relaxé, il présente généralement un réseau quadrillé de lignes de dislocations suivant les axes cristallographiques [110] et [1-10]. Lorsque la densité de ces lignes est très élevée (>1 tous les $10\ \mu\text{m}$), on peut généralement les observer avec un microscope optique de haute résolution ou par diffraction des rayons X (DRX). Cependant, si la densité est faible ou si les dislocations sont enterrées, il devient très difficile de les observer. C'est la raison pour laquelle nous avons développé un système d'imagerie par micro-photoluminescence.

Une dislocation perturbe le réseau cristallin et son énergie se situe généralement au centre de la bande interdite du semiconducteur, c'est un centre recombinant. La durée de vie d'un électron sur ce défaut est très faible, aussi les électrons sont piégés sur une distance voisine de la longueur de diffusion. Ces recombinaisons sont non radiatives, c'est à dire que la disparition de l'électron n'est pas accompagnée par l'émission d'un photon. Donc si l'on observe la photoluminescence de l'échantillon, nous devons observer des zones noires au voisinage des lignes de dislocation appelé darklines. Le montage utilisé pour visualiser la photoluminescence est illustré (Fig.7).

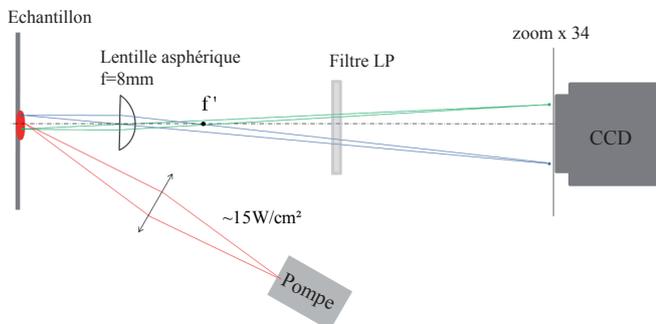


Fig. 7. Système d'imagerie par micro-PL

L'échantillon est excité optiquement par une diode laser commerciale. L'absorption des photons par le milieu à gain crée des porteurs dans la bande de conduction du semiconducteur. En se recombinant, ces porteurs génèrent un rayonnement à la longueur d'onde fixé par le gap du matériau utilisé. La luminescence ainsi obtenue est appelée photoluminescence. Ensuite on focalise le spot de luminescence sur la caméra avec une lentille asphérique. Puis on sélectionne la luminescence en insérant un filtre passe haut qui coupe la longueur d'onde de la pompe (diffusée par la surface de l'échantillon). La résolution optique de ce système est limitée par la diffraction (critère de Rayleigh). Pour une longueur d'onde de $1\ \mu\text{m}$ elle est d'environ $1\ \mu\text{m}$. La longueur de diffusion des électrons dans le matériau utilisé (InGaAs) étant supérieure à cette résolution, nous devons visualiser des darklines si toutefois l'échantillon

est relaxé (Fig.8). La caractérisation par micro-PL des

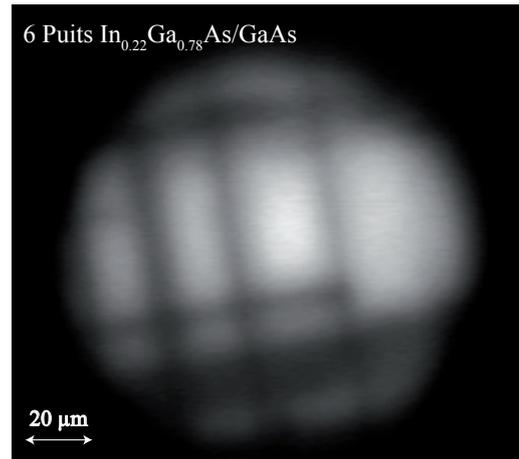


Fig. 8. Image par μ -PL de lignes de dislocation

échantillons émettant à $1\ \mu\text{m}$ est maintenant bien maîtrisée, par contre la caractérisation des échantillons à $2.3\ \mu\text{m}$ est plus délicate. En effet, à cette longueur d'onde, la résolution optique est moins bonne ($2-3\ \mu\text{m}$) et il n'existe pas de caméra 2D suffisamment sensible pour pouvoir effectuer une image de la photoluminescence. Cependant, nous avons développé un système équivalent avec une caméra linéaire sensible à $2.3\ \mu\text{m}$, on reconstitue l'image en balayant l'échantillon. Ce travail de caractérisation nous permet de mieux contrôler la compensation de contrainte (le désaccord de maille est compensé par une couche tampon de contrainte opposée) sur nos échantillons. Cela permet également de mieux comprendre la différence de gain entre les structures.

V. CONCLUSION

Après avoir présenté les éléments de compréhension des lasers à semiconducteurs, nous nous sommes intéressés à une configuration particulière du laser. Nous avons vu que les VECSELS présentent des propriétés de cohérence et de stabilité remarquable. Son fonctionnement intrinsèquement monomode et son accordabilité en fréquence plus importante (découplage du gain et du mode) lui donnent de sérieux avantages pour des applications d'analyse de gaz. Ensuite, nous avons vu que l'optimisation du gain pose des problèmes de relaxation de couches. Enfin, nous avons présenté le système de caractérisation par micro-PL qui nous a permis de révéler la présence de ligne de dislocation dans nos structures émettant à 1 et à $2.3\ \mu\text{m}$.

RÉFÉRENCES

- [1] R. Farcy. *Application des lasers*. Paris, 1993.
- [2] H. Maillat. *Le laser*. Paris, 1990.
- [3] A. Ouvrard. *Nouvelles sources lasers à base d'antimoniures pour l'application à l'analyse de traces de gaz dans le domaine spectral 2-2.5 μm* . PhD thesis, Université Montpellier II, 2005.
- [4] P.M.J. Marée, J.C. Barbour, and J.F. van der Veen. Generation of misfit dislocations in semiconductors. *J.A.P.*, 62(11) :4413, 1 december 1987.

ÉTUDE DE L'ÉVOLUTION DE L'OXYDE DE GRILLE DES COMPOSANTS DE PUISSANCE

LUDOVIC BOYER¹ (DOCTORANT 2^E ANNÉE), SERGE AGNEL² ET P. NOTINGHER (CO-DIRECTEURS DE THÈSE)

Institut d'Électronique du Sud (IES) – Groupe Énergie et Matériaux

¹ludovic.boyer@univ-montp2.fr

²serge.agnel@ies.univ-montp2.fr

Résumé— Les convertisseurs statiques embarqués dans des applications aéronautiques et terrestres doivent répondre à des critères de fiabilité très stricts pour des raisons à la fois liées à la sécurité des personnes et à des critères économiques. Cet article porte sur un travail visant à quantifier la fiabilité des composants semi-conducteurs de puissance, qui sont les parties stratégiques des convertisseurs. À travers des techniques de caractérisation traditionnelles et novatrices, nous nous intéressons à l'évolution de l'oxyde de grille de ces composants, susceptible d'être un indicateur de leur âge opérationnel.

Mots clés— fiabilité; IGBT; oxyde; C(V); Méthode de l'Onde Thermique.

I. INTRODUCTION

Les composants de puissance à base de semi-conducteurs sont aujourd'hui au coeur des systèmes de conversion d'énergie. Durant les 20 dernières années, leur intégration de plus en plus poussée a permis d'améliorer les performances des systèmes de conversion, tout en réduisant leur coût et leur encombrement. Parallèlement, les contraintes (thermiques, électriques, mécaniques) auxquelles sont soumis ces composants sont devenues de plus en plus sévères, alors que leur utilisation est de plus en plus fréquente dans des applications à forte criticité. Ceci a fait augmenter de manière significative les exigences concernant la fiabilité de ces dispositifs.

A la base de ces composants, se trouve une structure appelée Métal Oxyde Semi-conducteur (MOS) qui comporte une électrode métallique (la grille) connectée à un substrat en silicium dont la partie supérieure a été oxydée (oxyde de grille). La majorité de l'énergie véhiculée par une structure MOS transite par un canal étroit situé sous l'oxyde de grille. L'oxyde de grille et le canal subissent pendant le fonctionnement de forts champs électriques associés à des contraintes thermiques significatives, qui en font des maillons faibles dans la chaîne de fiabilité du système de conversion d'énergie.

Il y a alors aujourd'hui une très forte demande de l'industrie concernant la quantification de la fiabilité des composants de puissance. Afin de répondre à ces demandes, une compréhension plus approfondie des mécanismes de défaillance des structures MOS des composants de puissance est nécessaire. Pour ce faire, l'adaptation de méthodes de caractérisation des structures MOS aux composants de puissance est indispensable. De plus, l'étude de

l'évolution de l'oxyde de grille est susceptible d'apporter des informations précieuses sur l'âge opérationnel du composant.

Le travail décrit dans cet article porte sur l'analyse de l'évolution des propriétés de l'oxyde de grille d'un transistor de puissance à grille isolée (IGBT). Cette analyse est faite en utilisant deux techniques; l'une traditionnelle (la méthode Capacité-Tension); l'autre originale (la Méthode de l'Onde Thermique). Une partie du travail est dédiée à l'application de la première technique aux composants de puissance. La dernière technique, qui a été mise au point à l'IES, est susceptible d'apporter des informations complémentaires grâce à sa grande sensibilité.

II. DESCRIPTION DES STRUCTURES ET DES MÉTHODES EMPLOYÉES

II.1. La structure MOS et l'IGBT à grille en tranchée.

Une illustration de structure MOS est donnée Fig. 1(a). Le substrat semi-conducteur d'une structure MOS est plus ou moins dopé en électrons. Lorsqu'une tension est appliquée sur l'électrode de grille, un champ électrique est induit dans l'oxyde. Ce champ électrique a pour effet d'attirer des électrons du substrat sous la grille (ou de les repousser de la grille).

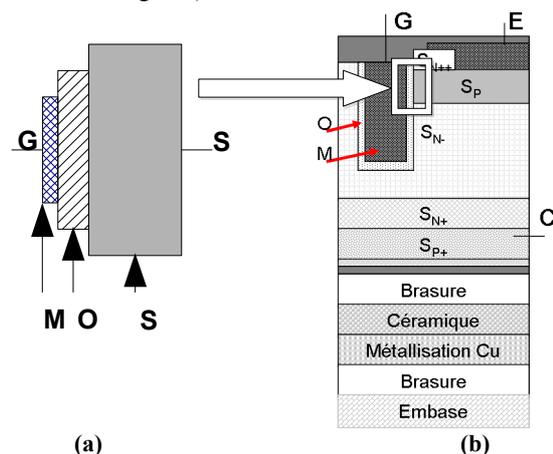


Fig. 1. Structures MOS (a) et IGBT à grille en tranchée (b)

L'IGBT (*Insulated Gate Bipolar Transistor*) à grille en tranchée est un composant de l'électronique de puissance utilisé en interrupteur commandé en tension. Il est notamment composé de 3 électrodes (Grille G, Émetteur E et Collecteur C), de couches de silicium différemment dopées S, et d'une couche isolante (O) en forme de tranchée (Fig. 1(b)). Les différentes couches de l'embase de

l'IGBT servent à renforcer sa rigidité et à évacuer la chaleur produite par la puce lors de son fonctionnement. La partie de l'IGBT correspondant à une structure MOS élémentaire est encadrée dans la Fig. 1(b). Cette zone du composant est la plus vulnérable aux contraintes électriques.

Les méthodes de caractérisation décrites ci-dessous visent à analyser les différents paramètres de la zone encadrée de la Fig. 1(b).

II.2. La méthode Capacité-Tension

La méthode Capacité-Tension (C(V)) est très utilisée pour analyser les composants de l'électronique à base de structure MOS. Elle permet de caractériser les défauts (pièges) – situés dans l'oxyde et le silicium sous l'oxyde – qui apparaissent et évoluent lors du vieillissement du composant.

Une structure MOS donne naissance à une capacité dont la valeur dépend de la géométrie de l'oxyde de grille et de la charge électrique (électrons) présente dans le semi-conducteur (sous l'oxyde). Sous l'effet des différentes tensions de polarisation appliquées à la structure, la valeur de la charge dans le semi-conducteur (donc de la capacité mesurée) change. L'apparition de pièges lors du vieillissement de la structure MOS a pour effet de capturer certaines charges. Il apparaît alors une variation de la capacité mesurée lors de l'application de différentes polarisations à la structure. Cette variation permet, par le biais de différents traitements [1], d'obtenir des informations sur la localisation des pièges et leur densité. Les informations ainsi obtenues sur les pièges présents dans la structure permettent de quantifier le vieillissement de cette dernière.

II.3. La Méthode de l'Onde Thermique

La Méthode de l'Onde Thermique (MOT) est une méthode de caractérisation non destructive qui a été développée au sein du Groupe Énergie et Matériaux (GEM) de l'IES. Elle permet d'obtenir la distribution volumique des charges piégées dans un isolant.

La MOT a initialement été développée pour l'analyse des isolants des systèmes de transport d'énergie (câbles, ...) dont les épaisseurs peuvent atteindre plusieurs centaines de micromètres [2]. Différentes études menées au GEM ont montré qu'il était possible de caractériser par la MOT des structures MOS soumises à des tensions de polarisation continues avec des épaisseurs d'oxyde de quelques dizaines de nanomètres. Ceci a pu être fait avec une sensibilité supérieure à celle de la méthode C(V) lors de la détermination des tensions caractéristiques des structures [3].

L'illustration du principe de la MOT est donnée Fig. 2 pour la caractérisation d'un isolant "simple" contenant une charge piégée. L'application d'un échelon de température sur une face de l'isolant produit une onde thermique qui, en traversant l'échantillon, entraîne un léger déplacement relatif de la charge piégée. Ce déséquilibre provoque une redistribution des charges d'influence présentes aux électrodes. On obtient ainsi dans le circuit extérieur un courant dit "d'onde thermique" $I_{MOT}(t)$, qui dépend de la capacité C de l'échantillon, de la variation

de celle-ci avec la température α , de son épaisseur d , du champ électrique dans l'échantillon $E(x)$ et de la variation temporelle de la température dans l'échantillon $\Delta T(x,t)$:

$$I_{MOT}(t) = -\alpha C \int_0^d E(x) \frac{\delta \Delta T(x,t)}{\delta t} dx \quad (1)$$

À partir de l'expression (1) et par le biais de différents traitements [2], il est possible de retrouver $E(x)$ et de localiser les charges piégées dans l'isolant.

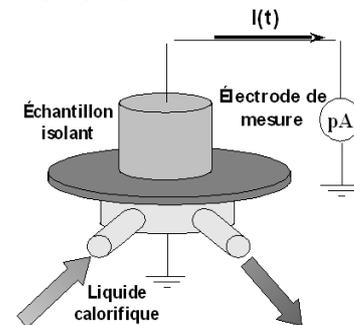


Fig. 2. Principe d'application de la MOT

III. RÉSULTATS EXPÉRIMENTAUX

III.1. Suivi de l'évolution de l'oxyde de grille d'une structure MOS

La structure de l'IGBT à grille en tranchée étant complexe, la première étape de l'étude de quantification de vieillissement est faite sur une structure MOS de base (voir Fig. 1(a)). L'évolution des propriétés de l'oxyde de cette structure élémentaire est suivie lorsqu'elle est soumise à des contraintes électriques similaires à celles subies par les IGBT en fonctionnement (cycles rectangulaires de tension, de fréquence variable dans la gamme des dizaines de kHz). Pour cela, nous avons conçu et réalisé un module électronique et une cellule de vieillissement accéléré adaptés à ce type d'échantillon (voir Fig. 3). On peut ainsi soumettre simultanément une dizaine d'échantillons à une contrainte électrique d'amplitude et de fréquence variables, à une température donnée.

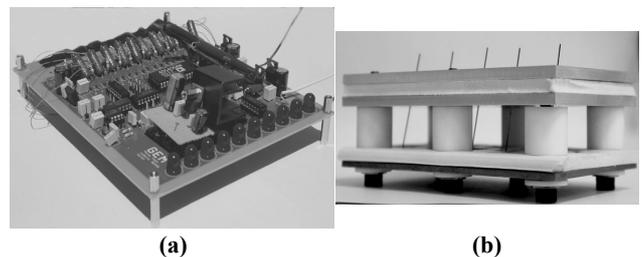


Fig. 3. Module électronique (a) et cellule de vieillissement (b)

Cette campagne de vieillissement est actuellement en cours. Des caractérisations par les méthodes C(V) et MOT sont faites périodiquement afin de suivre l'évolution des échantillons. Elle se poursuivra en faisant varier l'amplitude des contraintes électriques et thermiques.

III.2. Adaptation de la méthode Capacité-Tension aux IGBT à grille en tranchée

Des mesures d'applicabilité de la C(V) ont été effectuées sur des IGBT à grille en tranchée. En effet, l'analyse d'oxydes de structures plus simples, ceux des transistors

MOS à grille en tranchée, a déjà été effectuée [4] mais nous n'avons pas trouvé d'études traitant de l'applicabilité de la $C(V)$ sur les IGBT à grille en tranchée. Les résultats de ces mesures sont présentés en Fig. 4. Elles ont été faites entre chaque électrode du composant : Grille-Émetteur C_{GE} , Grille-Collecteur C_{GC} et Grille-Émetteur/Collecteur en court-circuit C_{tot} . Cette décomposition des mesures $C(V)$ est appelé split- $C(V)$. D'après la littérature [5] [4], l'utilisation de la split- $C(V)$, qui consiste à faire des mesures $C(V)$ de façon sélective, permet une caractérisation plus aisée de la structure étudiée. Ainsi, en effectuant les mesures uniquement sur une partie du composant (ici C_{GE}) et par le biais de traitements minimes, on pourrait analyser la partie de l'IGBT qui nous intéresse (indiquée dans la Fig. 2(b)) comme une structure MOS plane élémentaire (Fig. 1(a)) [1].

Les formes des courbes $C_{tot}(V)$ et $C_{GE}(V)$ sont semblables à des formes de courbes $C(V)$ pour des structures MOS simples [1]. Ces mesures correspondent donc à la caractérisation de la structure MOS de l'IGBT.

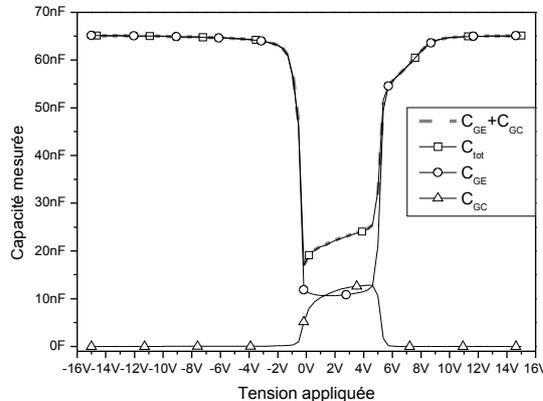


Fig. 4. Mesures $C(V)$ sur IGBT à grille en tranchée effectuées entre les différentes électrodes

L'expression de C_{tot} est :

$$C_{tot}(V) = C_{GC}(V) + C_{GE}(V) \quad (2)$$

Les courbes $C_{tot}(V)$ et $C_{GC}(V) + C_{GE}(V)$ ont été tracées dans la Fig. 4. Comme elles se superposent, la relation (2) est vérifiée, ce qui valide l'applicabilité des mesures split- $C(V)$ sur l'IGBT à grille en tranchée sur toute la gamme de polarisation appliquée ($\pm 15V$).

L'étude des possibilités de caractérisation du composant étudié est en cours. Nous espérons ainsi pouvoir analyser plus facilement l'évolution des charges électriques piégées dans l'oxyde de l'IGBT.

III.3. Adaptation de la Méthode de l'Onde Thermique aux IGBT à grille en tranchée

Des mesures MOT de faisabilité ont été effectuées sur des IGBT à grille en tranchée pour différentes polarisations. Les résultats de ces mesures sont présentés dans la Fig. 5. La forme des courbes obtenues et le changement de l'amplitude du courant en fonction de la polarisation appliquée montrent la faisabilité des mesures [3].

Le banc de caractérisation actuel doit cependant être adapté à la structure étudiée (voir Fig. 1(b)) afin d'optimiser le transfert thermique, l'amplitude de la tension de polarisation appliquée et d'atténuer le bruit venant perturber la mesure du courant d'onde thermique.

Ceci se fera en plusieurs étapes dont la principale est la conception d'un nouveau système d'application d'échellons de température à l'IGBT. Plusieurs approches sont actuellement à l'étude.

Nous espérons, en développant cette technique, pouvoir détecter des faibles évolutions de l'oxyde corrélables à l'âge opérationnel du composant.

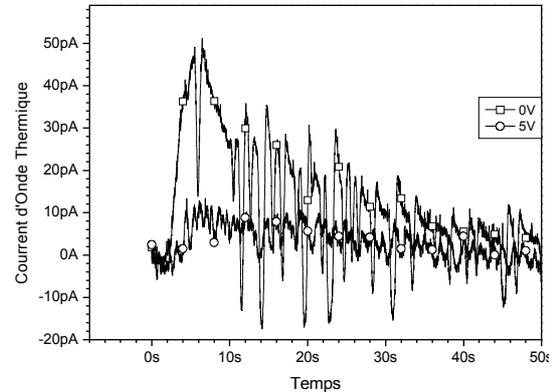


Fig. 5. Mesures MOT sur IGBT à grille en tranchée pour différentes polarisations

IV. CONCLUSION ET PERSPECTIVES

Le travail présenté dans cet article se situe dans le cadre de l'étude de l'évolution des oxydes de grille des composants de puissance. Les résultats présentés ont confirmé l'applicabilité de la méthode split- $C(V)$ à des IGBT à grille en tranchée pour toute la gamme de polarisation.

De même, les mesures effectuées par la Méthode de l'Onde Thermique semblent confirmer l'applicabilité de cette méthode originale sur les composants finis.

Nous envisageons de poursuivre l'étude des oxydes des structures MOS élémentaires et d'optimiser l'application de la Méthode de l'Onde Thermique aux composants finis. Dans un deuxième temps, l'étude du vieillissement sera poursuivie sur les composants finis en conditions similaires à celles subies en service sur des bancs de test spécifiques déjà mis au point à l'IES [6].

V. RÉFÉRENCES

- [1] E. H. Nicollian, J. R. Brews, "MOS Physics and Technology", 2003.
- [2] A. Toureille, "Mesures de charges d'espace par la méthode de l'onde thermique: différentes techniques de validation numérique et expérimentale", Journal of Electrostatics, 1994, Vol. 32, p. 277-286.
- [3] Olivier Fruchier, "Étude du comportement de la charge d'espace dans les structures MOS ...", Thèse Doctorale, UM2, 2006.
- [4] A. Yahata, et al., "Exact evaluation of channel mobility for trench MOSFET using split C-V method", Elsevier Science, Applied Surface Science, 1997, Vol. 117/118, p.181-186.
- [5] S. Mileusnic, et al., "MOS transistors characterization by split C-V method", IEEE Semiconductor Conference, 2001, Vol. 2, p.503 – 506.
- [6] Mohamed EL GHAZOUANI, "Conception d'une plate-forme flexible de vieillissement haute température pour modules IGBT sous contraintes", Thèse Doctorale, UM2, 2006.

Exploitation des phonons optiques polaires dans le GaN en vue de la réalisation d'une source térahertz

Thibault LAURENT¹ (Doctorant 1^{ère} année), Luca VARANI²

Institut d'Électronique du Sud, GEHF
CC 084, place Eugène Bataillon, 34090 Montpellier

¹thibault.laurent@ies.univ-montp2.fr, ²luca.varani@ies.univ-montp2.fr

Résumé— Le projet AITHER développé au sein de l'équipe GEHF permettra d'émettre des ondes électromagnétiques dans le domaine térahertz. Le phénomène physique à la base de ce nouveau composant est étudié dans cet article. Il consiste en l'exploitation d'une propriété particulière de l'interaction des électrons avec les phonons optiques polaires tout en limitant les autres interactions que pourrait subir l'électron lors de sa traversée du composant.

Mots-clés— TéraHertz, GaN, phonons, interactions électroniques.

I. INTRODUCTION

Le domaine térahertz (300 GHz – 30 THz) suscite de plus en plus d'intérêt au vu du champ d'applications industrielles qu'il offre, en particulier dans les secteurs militaire, médical ou encore des technologies de l'information [1–3]. Réaliser des sources et des détecteurs susceptibles de fonctionner dans cette gamme de fréquences est l'un des principaux enjeux de la recherche en hyperfréquences.

Afin d'aboutir à une source de rayonnement térahertz, il a été développé au sein de l'équipe GEHF de l'IES un nouveau dispositif baptisé AITHER (Amplifying Integrable TeraHertz Emitter of Radiations), qui, d'un point de vue microscopique exploite l'interaction entre électrons et phonons optiques polaires (qui seront définis dans la partie II). Pour que ce dispositif puisse fonctionner, il faut que cette interaction devienne prépondérante par rapport aux autres, c'est pourquoi nous évoquerons dans la partie II les différentes interactions que peut rencontrer un électron lors de son passage dans un matériau, puis, dans la partie III, les moyens dont nous disposons pour favoriser dans le GaN (Nitrure de Gallium) ce type d'interaction par rapport aux autres.

II. INTERACTIONS ÉLECTRONIQUES

Nous ne dresserons pas ici une liste exhaustive des différents mécanismes d'interaction que peut subir un électron lors de sa traversée d'un matériau, mais en rappellerons les principales conséquences.

Lors de leur passage au sein d'un matériau, les porteurs interagissent avec une multitude de particules différentes. Citons par exemple les interactions [4] :

- entre porteurs (indifféremment électron ou trou) ;
- avec les impuretés ionisées ;

– avec les atomes du réseau cristallin.

Elles induisent des échanges d'énergie entre le porteur considéré et la particule en jeu. Ainsi la trajectoire du porteur au cœur du matériau sera déterminée par ces interactions et sa quantité de mouvement initiale s'en trouvera modifiée comme l'illustre la figure 1.

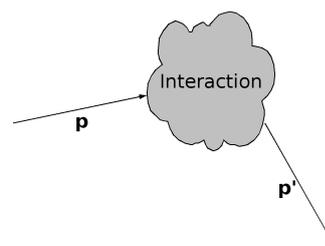


Fig. 1. Interaction d'un porteur possédant une quantité de mouvement initiale \mathbf{p} et finale \mathbf{p}' .

L'interaction qui nous intéresse plus précisément est celle impliquant le porteur et le réseau cristallin [5]. La collision entre le porteur considéré et un atome du réseau provoque la vibration de ce dernier qui se propage de proche en proche tout au long de la maille du réseau. Selon l'énergie dont dispose initialement le porteur, elle peut donner lieu à l'émission ou à l'absorption d'un phonon, quantum de vibration du réseau cristallin, comme l'illustre la figure 2. Ces phonons existent naturellement dans le réseau cristallin à une température différente du zéro absolu puisque l'énergie thermique des atomes les font vibrer autour de leur position d'équilibre. Les liaisons entre chaque atome (qualifiables d'élastiques) permettent la propagation de ces vibrations élémentaires tout au long du réseau.

La longueur d'onde du phonon émis ou absorbé lors de la collision entre le porteur et le réseau cristallin dépend directement de la quantité d'énergie échangée entre les deux protagonistes. Toutes les longueurs d'onde ne sont pas envisageables, et il a été montré [5] qu'en fonction du vecteur d'onde du phonon \mathbf{q} , le nombre de ces valeurs est limité. Ce nombre dépend des différents modes de propagation possibles dans le réseau. La figure 3 illustre l'allure des courbes de dispersion dans un cristal pour une direction de propagation donnée. Nous pouvons y constater que deux domaines de fréquences en ressortent. Suivant la gamme d'énergies

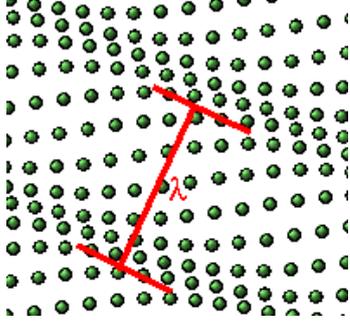


Fig. 2. Illustration de la propagation d'un phonon de longueur d'onde λ (exagérée pour rendre compréhensible le phénomène) : chaque point représente un atome du réseau se déplaçant autour de sa position d'équilibre [6].

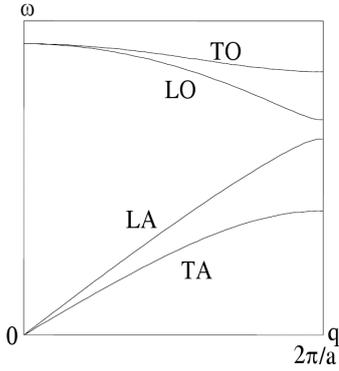


Fig. 3. Allure des courbes de dispersion dans un cristal pour des vibrations se propageant le long de l'axe [100] : vibration longitudinale optique (LO), transverse optique (TO), longitudinale acoustique (LA) ou transverse acoustique (TA).

(ou de fréquences) mise en jeu, nous parlerons alors de phonons *acoustiques* ou *optiques*.

Enfin, l'appellation *polaire* est juxtaposée aux phonons optiques lorsque le cristal considéré est le résultat d'un alliage de matériaux électriquement différents (par exemple un alliage III/V de GaN). En effet, un écartement des atomes élémentaires d'un réseau de ce type crée un moment dipolaire, et par conséquent, la propagation d'une onde polarisée. Pour les phonons acoustiques, nous parlerons d'effets piézoélectriques, observables en basse fréquence. Il est donc impossible d'obtenir de tels effets dans des matériaux uniquement constitués d'atomes de la colonne IV comme le Si ou le SiC car l'écartement de deux atomes ne crée pas de moment dipolaire.

Comme nous l'avons vu, un électron est susceptible d'interagir avec une multitude de particules différentes, mais la réussite finale de notre dispositif dépend fondamentalement de la capacité que nous aurons à sélectionner l'interaction intéressante pour notre application : celle des électrons avec les phonons optiques polaires.

Dans la suite, nous ne parlerons plus de *porteur* mais d'*électron*, puisqu'afin de s'affranchir des interactions électrons - trous très gênantes et d'obtenir une meilleure mobilité, le canal considéré sera unipolaire.

III. SÉLECTIVITÉ DE L'INTERACTION AVEC LES PHONONS OPTIQUES POLAIRES DANS LE GAN

Considérons un électron pénétrant dans un matériau semi-conducteur à base de GaN avec une quantité de mouvement \mathbf{p}_0 . Nous supposons que ce porteur subit des interactions successives, approche qui découle de la règle d'or de Fermi [7], et nous appellerons $\tau_{\mathbf{p}_0}$ le temps moyen entre deux interactions, ou temps de relaxation. Il nous est alors possible de définir le taux d'interaction $\frac{1}{\tau_{\mathbf{p}_0}}$ qui quantifie la probabilité par unité de temps qu'une interaction entre l'électron ainsi défini et la particule considérée ait effectivement lieu.

Il est à noter que, selon la nature de l'interaction, ce taux dépendra d'une série de paramètres de transport : température, champ électrique, canal 2D-3D, dopage, ... Ainsi, l'action sur ces paramètres nous permet théoriquement de rendre un phénomène prépondérant [7]. Illustrons ceci à l'aide d'un exemple concret : celui de l'interaction entre l'électron et le réseau cristallin induisant l'émission des deux types de phonons précédemment évoqués, que nous dénommerons dans la suite « interaction avec les phonons ».

Le taux d'interaction avec les phonons acoustiques est donné par la relation :

$$\frac{1}{\tau_{\mathbf{p}_0}} = \frac{\pi D_A^2 k_B T_L}{\hbar c_1} \frac{1}{W_{fi}} g_{2Df}(E)$$

où D_A , T_L , W_{fi} et g_{2Df} correspondent respectivement au potentiel de déformation (dépendant du matériau), à la température du réseau, à l'effet du confinement dans une direction et à la densité d'états.

Le taux d'interaction avec les phonons optiques polaires vérifie quant à lui :

$$\frac{1}{\tau_{\mathbf{p}_0}} = \sum_{\mathbf{p}'_{\parallel}} \frac{2\pi}{\hbar} |D_{2D}(\beta)|^2 \delta_{\mathbf{p}'_{\parallel}, \mathbf{p}_{\parallel} \pm \hbar\beta} \delta(E' - E \mp \hbar\omega)$$

où D_{2D} , $\delta_{\mathbf{p}'_{\parallel}, \mathbf{p}_{\parallel} \pm \hbar\beta}$ et $\delta(E' - E \mp \hbar\omega)$ correspondent respectivement à un terme relatif au confinement des électrons dans une couche restreinte (puits quantique), à une condition sur la quantité de mouvement de l'électron et au fait que l'énergie ne peut être cédée que par paquets $\hbar\omega$ d'énergie.

Nous constatons que si l'abaissement de la température permet de diminuer le taux d'interaction entre porteur et phonon acoustique, elle n'a en revanche aucune influence sur l'interaction entre porteur et phonon optique polaire. Ainsi, jouer sur la température est l'un des moyens d'action possible pour favoriser l'émission de ces phonons optiques polaires.

De plus, il a été montré [8] que dans le GaN les principales interactions que subit un électron lorsqu'il traverse un barreau massif de GaN, sont celles avec les impuretés, les phonons acoustiques et les phonons optiques polaires, comme l'illustre la figure 4.

D'après ce qui précède, nous pouvons donc d'ores et déjà nous affranchir de l'interaction avec les phonons acoustiques en abaissant la température du matériau. La deuxième interaction dont nous souhaiterions nous acquitter est celle avec les impuretés : pour cela, nous pouvons faire en sorte que les impuretés nécessaires à l'apport d'électrons dans le canal soient elles-mêmes déportées

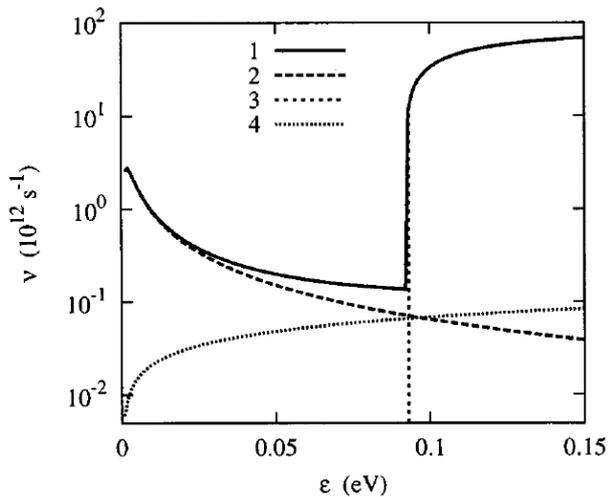


Fig. 4. Taux d'interaction des principaux mécanismes de collision électronique dans le GaN en fonction de leur énergie à $T_0 = 10 \text{ K}$ et $N_D = 10^{16} \text{ cm}^{-3}$: (1) total, (2) avec les impuretés, (3) avec les phonons optiques et (4) avec les phonons acoustiques [8].

du canal afin qu'elle n'apparaissent pas à l'endroit où les électrons se meuvent. Cette technique est largement répandue puisque d'une manière générale la présence d'impuretés dans le canal conducteur d'un composant est gênante.

Ainsi, nous avons montré qu'il est possible, dans le GaN, de rendre prépondérante l'interaction entre électrons et phonons optiques polaires.

IV. CONCLUSION

Il est donc possible d'exploiter pleinement l'interaction électrons - phonons optiques polaires dans le GaN en abaissant la température du matériau et en séparant la couche de canal et de donneurs (couche d'impuretés fournissant les électrons pour le canal). Ces deux conditions réunies permettront le transport des électrons dans le canal sans autre interaction que celles avec les phonons optiques polaires. Dans ces conditions, il a été montré [8,9] que le GaN est capable d'amplifier une onde électromagnétique incidente dont la fréquence appartient au domaine térahertz. L'ajout d'une cavité extérieure permet alors au phénomène de s'auto-amplifier et ainsi de générer une onde sans source extérieure.

Le développement et la campagne d'expérimentation de ce dispositif innovant sont en cours au sein de notre équipe, de la réalisation des échantillons à leur caractérisation, en partenariat avec le CHREA et Lumilog.

RÉFÉRENCES

- [1] P.H. Siegel. Terahertz technology in biology and medicine. *IEEE Transactions on Microwave Theory and Techniques*, 52(10) :2438–2447, 2004.
- [2] P.H. Siegel. Terahertz technology. *IEEE Transactions on Microwave Theory and Techniques*, 50(3) :910–928, mar 2002.
- [3] M.C. Kemp, P.F. Taday, B.E. Cole, J.A. Cluff, A.J. Fitzgerald, and W.R. Tribe. Security applications of terahertz technology. volume 5070, pages 44–52. SPIE, 2003.
- [4] L. Reggiani et al. *Hot-Electron Transport in Semiconductors*, volume 58. Springer-Verlag, 1985.
- [5] Neil W. Ashcroft and David N. Mermin. *Solid State Physics*. Saunders College Publishing, 1976.
- [6] <http://fr.wikipedia.org/wiki/Phonon>.

- [7] M. Lundstrom. *Fundamentals of carrier transport*. Press syndicate of the University of Cambridge, 2nd edition, 2000.
- [8] E. Starikov, P. Shiktorov, V. Gruzinskis, L. Reggiani, L. Varani, J. C. Vaissière, and Jian H. Zhao. Monte carlo simulation of the generation of terahertz radiation in gan. *Journal of Applied Physics*, 89(2) :1161–1171, 2001.
- [9] E. Starikov, P. Shiktorov, V. Gruzinskis, L. Reggiani, and L. Varani. Coherent terahertz radiation generation due to carrier interaction with low-temperature optical phonons in semiconductors : Achievements and perspectives. *Journal of Nanoelectronics and Optoelectronics*, 2 :1–25, 2007.

CONCEPTION ET REALISATION D'UN TRANSISTOR QUANTIQUE A ELECTRONS CHAUDS InAs

THIBAUT DAUD (DOCTORANT 3E ANNEE), ROLAND TEISSIER (DIRECTEUR DE THESE)

Laboratoire IES, Groupe NANOMIR
Adresse, 34000 Montpellier Cedex, France

DAUD@IES.UNIV-MONTP2.FR

RESUME— Le présent travail porte sur l'élaboration d'une structure originale de transistor quantique à électrons chauds à base d'InAs. La finalité est de développer un composant dont les limites intrinsèques en fréquence sont notablement plus grandes que celles des composants III-V habituels. Cet article présente les grandes lignes sur la conception d'une telle structure, sa fabrication et ses premières caractéristiques statiques.

Mots clés— antimoniure, hétérojonction, super réseau, transistor, hyperfréquences.

I. INTRODUCTION.

Les réseaux de télécommunication imposent des flux de données de plus en plus importants et nécessitent donc des systèmes électroniques de plus en plus performants. Les matériaux semi-conducteurs alternatifs, comme par exemple la filière InP qui présente d'excellents atouts en termes de transport électroniques, trouvent leur place dans ce domaine. Deux types de composants sont sollicités en très haute fréquence : le High Electron Mobility Transistor (HEMT) et le transistor bipolaire à hétérojonction (TBH). Ces transistors peuvent fonctionner à des fréquences supérieures à 500GHz [1].

Cependant, les limites fréquentielles de ce type de structure sont atteintes au niveau du design du composant et des propriétés intrinsèques des matériaux utilisés.

Mon sujet d'étude porte sur la conception et la fabrication d'un nouveau composant, basé sur le principe unipolaire et réalisé à partir de matériaux innovants avec des propriétés électroniques bien plus performantes. Ce transistor, susceptible d'atteindre un fonctionnement au-delà du THz, peut offrir de nouvelles perspectives pour les futures générations de transistor.

II. PRINCIPE DE FONCTIONNEMENT DU QHET

Ce transistor est basé sur le principe du transistor à électrons chauds (HET) [2]. Sa structure est comparable à celui d'un transistor bipolaire à hétérojonction. Le transport des électrons est perpendiculaire aux couches et l'émission thermoionique régit le passage des électrons de l'émetteur

vers la base. Ce composant a la particularité d'être unipolaire c'est-à-dire d'avoir une base de type N, au même titre que les deux autres régions. Les électrons majoritaires de la base sont bloqués par les discontinuités de bandes des hétérojonctions émetteur-base et base-collecteur, tandis que les électrons chauds peuvent passer librement de l'émetteur au collecteur.

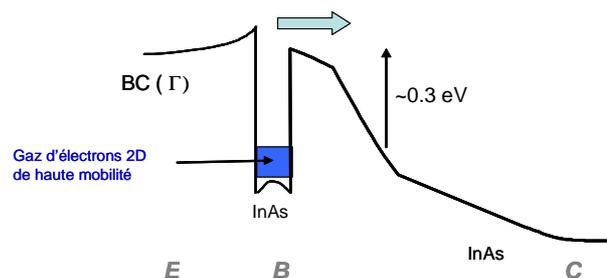


Fig.1. Diagramme de bande schématisé d'un transistor à électrons chauds (bande de conduction)

Ce transistor est développé dans la filière 6.1Å (antimoniure) afin d'exploiter les propriétés de transport électronique remarquables du matériau InAs. Son concept unipolaire et les propriétés intrinsèques du matériau InAs en terme de mobilité électronique (de l'ordre de 20000cm²/Vs) et de vitesse moyenne des électrons (de l'ordre de 1.5x10⁸ cm/s) permettent théoriquement de repousser considérablement les limites fréquentielles actuelles.

III. CONCEPTION DE LA STRUCTURE

L'équipe NANOMIR du laboratoire IES, spécialisée dans la filière 6.1Å, maîtrise la croissance par jets moléculaires de monocouches cristallines. Bénéficiant du savoir faire de l'équipe d'épitaxie, l'alliage digital InAs/AlSb a été choisi pour réaliser les jonctions émetteur-base et base-collecteur. En effet, cet alliage binaire, équivalent à un alliage quaternaire InAsAlSb qui imposerait une étude plus lourde des paramètres de croissance, offre facilement de la souplesse pour adapter la composition à la valeur voulue. Pour définir les super réseaux InAs/AlSb adaptés, j'ai utilisé

un programme permettant d'étudier les niveaux d'énergies et les probabilités de présence des électrons dans un puits quantique isolé ou couplé (super réseau dans notre cas).

III.1. Jonction émetteur-base.

Dans un puits quantique, la hauteur du potentiel du premier niveau d'énergie varie en fonction de son épaisseur. Plus le puits est fin et plus le niveau fondamental est haut. Il suffit donc de jouer sur la géométrie des puits pour créer la courbure de bande croissante souhaitée dans la jonction émetteur-base. Le phénomène d'effet tunnel résonant à travers le super réseau est prépondérant. Le principe est similaire à celui d'un transistor bipolaire à hétérojonction.

III.2. Base.

La base est un puits quantique InAs de dimension nanométrique dopé de type N. Pour une épaisseur de couche $a=125\text{\AA}$ et un dopage modéré $N = 2 \cdot 10^{18} \text{cm}^{-3}$, la résistance carrée de base est estimée à seulement $R_{\square B} = 75\Omega/\square$. Cette valeur est 10 fois plus faible que celles des transistors bipolaires à hétérojonction classiques. C'est un atout considérable qui joue directement sur l'augmentation de la fréquence maximale d'oscillations.

III.3. Jonction base- collecteur et collecteur.

La jonction base-collecteur est composée d'un super réseau InAs/AlSb et d'une couche non dopée d'InAs massif. Le super réseau joue le rôle d'une barrière qui doit permettre à la fois le blocage des électrons excités thermiquement de la base vers le collecteur et le passage des électrons chauds à travers la base de l'émetteur vers le collecteur. La couche d'InAs massif offre une très grande vitesse électronique et donc un temps de transit faible.

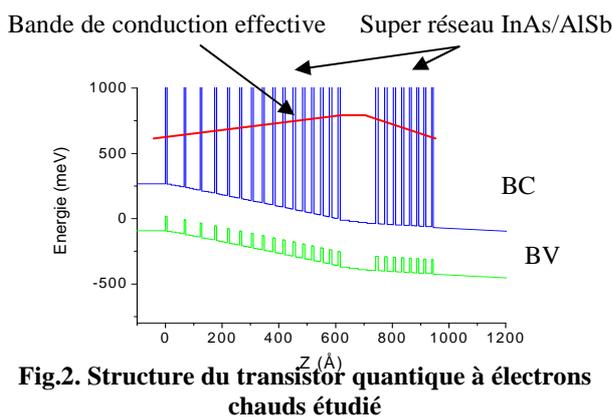


Fig.2. Structure du transistor quantique à électrons chauds étudié

IV. FABRICATION DU COMPOSANT

IV.1. Etapes essentielles de microfabrication.

Les étapes technologiques sont diverses :

- Photolithographie :

Cette étape consiste à transférer les motifs désirés sur l'échantillon. Une couche de résine photosensible aux

rayons UV est déposée sur l'échantillon de façon uniforme. Ensuite une lampe UV expose la résine au travers d'un masque de verre sur lequel sont dessinées les géométries du composant. L'échantillon est finalement immergé dans une solution fortement basique afin de développer les motifs à transférer (principe similaire à la photographie).

- Dépôt ou métallisation :

Ce procédé permet de déposer des couches minces de métaux sur l'échantillon par la méthode de l'évaporation. Un filament est porté à très haute température par le biais d'un courant électrique. Ce filament chauffe le creuset contenant le métal qui s'évapore dans la chambre sous vide et se dépose sur l'échantillon. Une balance à quartz permet en temps réel de contrôler la couche de métal déposée.

- Gravure :

Cette étape consiste à enlever le matériau aux endroits voulus de l'échantillon. Elle peut s'effectuer en milieu plasma, en milieu sec ou humide. Dans mon cas, je procède à des gravures humides.

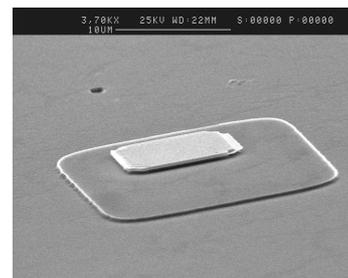


Fig.3. Photo MEB d'un transistor de test

IV.2. Optimisations de la technologie du composant.

Le point critique réside dans le dégagement de la fine couche de base pour y déposer un contact. Nous avons mis au point un protocole de gravure sélective couche à couche. Cette méthode permet d'enlever le nombre exact de paires InAs/AlSb pour dégager au mieux la couche de base InAs se trouvant au milieu d'un super réseau. Les temps de gravure de chacune des couches et l'épaisseur des barrières AlSb jouent un rôle très important dans l'optimisation de ce procédé. Ci-dessous, l'état de surface de la couche de base d'une épaisseur de 110\AA après gravure sélective couche à couche de la jonction émetteur-base :

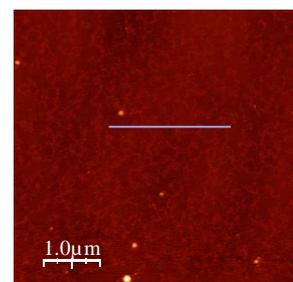


Fig 4(a)

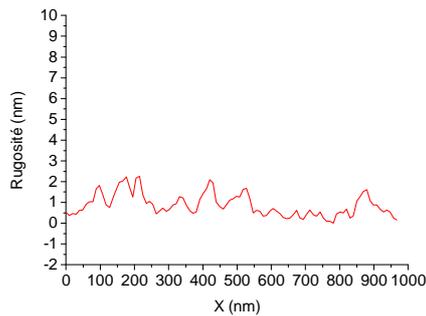


Fig 4(b)

Fig.4. (a) Photo AFM de la surface de la couche de base ; (b) Rugosité de la couche de base

La rugosité en surface n'excède pas les 20\AA soit un cinquième de l'épaisseur de la couche de base, prouvant ainsi la validité du protocole de gravure sélective du super réseau InAs/AlSb.

V. CARACTERISTIQUES

V.1. Caractérisations des jonctions.

Pour comprendre au mieux le comportement du composant, nous avons choisi d'étudier séparément les jonctions.

En ce qui concerne la jonction émetteur-base, nous avons joué sur les épaisseurs de couches du super réseau InAs/AlSb et sur les dopages pour optimiser le courant d'émetteur. En effet plus celui-ci est important et plus le temps de charge de la jonction est court.

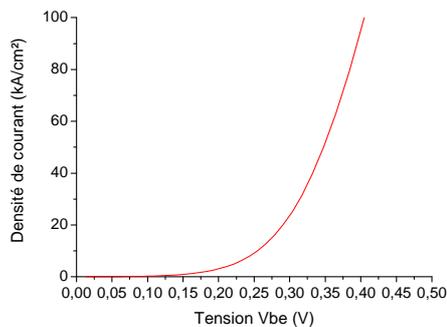


Fig.5. Densité de courant d'émetteur pour une jonction émetteur-base de surface : $S_e = (10 \times 10)\mu\text{m}^2$

Pour fonctionner à de très hautes fréquences, le transistor doit avoir une densité de courant d'émetteur d'au moins 100kA/cm^2 . La condition est vérifiée et permet aussi de valider la théorie de l'effet tunnel à travers un super réseau. Pour une tension de 0.4V la jonction émetteur-base atteint une densité de 100kA/cm^2 .

Le super réseau InAs/AlSb de la jonction base-collecteur a pour fonction de bloquer les électrons venant de la base par effet thermoionique. Il joue le rôle de barrière. Sa

hauteur, c'est-à-dire la différence d'énergie entre le niveau de Fermi de la base et la bande de conduction effective du super réseau, influe directement sur le courant de fuite. Ci-dessous sont représentées les densités de courant de fuite pour différentes hauteurs de barrières.

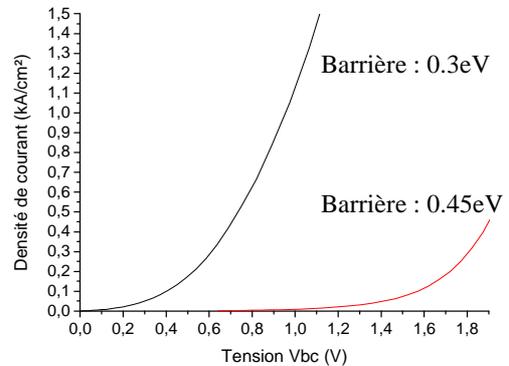


Fig.6. Densité de courant de fuite d'une jonction base-collecteur pour différentes hauteurs de barrière

Les résultats expérimentaux correspondent à la théorie : plus la barrière est haute et plus le courant de fuite est faible. Une épaisseur de 150\AA suffit pour empêcher les électrons de passer à travers le super réseau par effet tunnel.

V.2. Caractéristiques statiques $I_c(V_{ce})$.

Pour des géométries de grandes dimensions, les premières caractéristiques statiques $I_c(V_{ce})$ figurent ci-dessous :

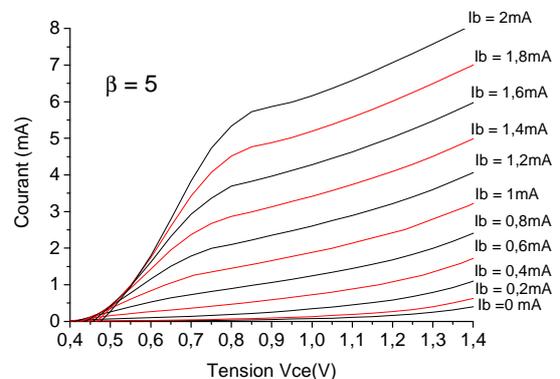


Fig.5. Caractéristiques $I_c(V_{ce})$ pour une surface d'émetteur : $S_e = (10 \times 10)\mu\text{m}^2$

La structure caractérisée présente un gain de 5 sur une plage de tension $V_{ce} > 1\text{V}$. La croissance et la fabrication de ce type de composant sont reproductibles.

V.3. Gain statique en courant.

Le point majeur de l'étude porte maintenant sur l'augmentation du gain statique en courant β qui est un paramètre très important d'un transistor.

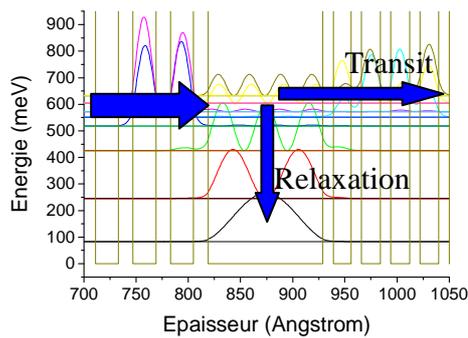


Fig.6. Transfert des électrons autour de la base

Le schéma ci-dessus résume les transferts possibles des électrons autour de la couche de base. En effet, le gain dépend beaucoup du couplage entre le niveau de résonance de la base et les deux autres régions (émetteur et collecteur). Il est directement proportionnel au rapport entre le temps de relaxation et le temps de transit dans la couche de base. Nous pouvons par exemple jouer sur les épaisseurs de couches autour de la base pour optimiser le couplage et donc le gain.

VII. CONCLUSION

Ce sujet est tout nouveau et demande l'étude de différents aspects. Les résultats obtenus prouvent que cette structure innovante fonctionne en tant que transistor. Les perspectives sont multiples car beaucoup de domaines restent encore à exploiter pour acquérir une maturité suffisante sur ce type de composants. L'optimisation du dessin de la structure est un point clé dans l'augmentation du gain en courant. La structure présente est épitaxiée sur un substrat conducteur InAs. Pour évaluer son potentiel fréquentiel, nous travaillons sur la mise au point de la croissance de la structure transistor sur substrat semi-isolant GaAs et une technologie petites dimensions est en cours d'étude pour permettre une caractérisation en régime dynamique (hautes fréquences).

REFERENCES.

- [1] InGaAs/InP DHBTs with 120-nm collector having simultaneously high f_t , $f_{max} \geq 450$ GHz, Griffith Z, Rodwell MJW, Fang XM, Loubychev D, Wu Y, Fastenau JM, Liu AWK, IEEE ELECTRON DEVICE LETTERS 26 (8): 530-532 AUG 2005
- [2] A.F.J. Levi, T.H. Chiu, "Room temperature operation of hot electron transistors", *Appl. Phys. Lett.* 51, 984, (1987).

DOCTISS 2008

Etude géométrique et dynamique de milieux granulaires polydispersés

Charles VOIVRET¹, Jean-Yves DELENNE, Moulay Saïd EL YOUSOUFI et Farhang RADJAI

LMGC

Place Eugène Bataillon 34095 Montpellier cedex 5

¹voivret@lmgc.univ-montp2.fr

Résumé— La granularité (forme et taille des particules) d'un milieu granulaire conditionne fortement son comportement mécanique. Nous nous intéressons ici à son effet sur la texture (arrangement spatial des particules) et sur la transmission des forces au sein d'un échantillon granulaire. Cette étude est menée grâce à des simulations numériques en deux dimensions.

Mots-clés— Milieux granulaires, polydispersité, courbe granulométrique, simulation numérique, texture granulaire.

I. INTRODUCTION

La matière granulaire est omniprésente dans notre environnement. Les sables, poudres, sols, comprimés, céréales, fruits, constituent quelques exemples de granulats que l'on peut rencontrer dans la vie de tous les jours. Malgré une simplicité apparente, cette matière recèle bien des surprises. Tassée elle se comporte comme un solide capable de supporter d'importantes charges (sols) ; au contraire, lâchée sur une pente, elle peut s'écouler comme un liquide (avalanche de neige ou de sable). D'autres phénomènes moins intuitifs comme la ségrégation de taille (regroupement de particules de même taille au sein d'un mélange) ou la dilatance (dilatation d'un milieu initialement dense sous sollicitation) sont propres à la matière en grains. Sans rentrer plus dans les détails, on comprend que la forme et la taille des particules, appelée aussi granularité, vont jouer un rôle important dans le comportement d'un tel milieu. Par exemple, des oranges (rondes) et du ballast (granulat anguleux placé sous les voies de chemin de fer) s'écouleront différemment sur un plan incliné ; on parlera alors de l'effet de forme des particules. De même, si on ajoute des raisins et des pastèques à nos oranges on obtiendra un troisième granulat aux propriétés différentes, on parlera alors d'effet de taille des particules. Dans ce papier on s'intéressera à ce dernier effet en considérant un milieu granulaire modèle constitué de particules de même forme mais de tailles très variées ; le milieu sera alors polydisperse.

Afin de modéliser de tels milieux, on peut utiliser deux méthodes bien distinctes : l'homogénéisation ou l'approche locale (micromécanique). L'homogénéisation consiste à trouver les caractéristiques d'un milieu équivalent à celui étudié au sens de la mécanique des milieux continus (échelle macroscopique). Malheureusement cette approche classique en mécanique ne permet pas d'expliquer des phénomènes comme la dilatance. L'approche locale

consiste à considérer les particules comme indépendantes et interagissant entre elles par des actions de contact dont on cherche à modéliser le comportement (échelle micro ou mésoscopique) pour en déduire le comportement de l'ensemble des particules (échelle macroscopique). Toute la difficulté de cette approche réside dans la compréhension puis la modélisation de la physique du contact entre les particules. Cette modélisation locale se prête particulièrement bien à l'implémentation de lois de comportement locales dans un code de calcul par éléments distincts : en ne supposant que les lois d'interactions on observe numériquement des comportements propres aux milieux granulaires comme la ségrégation ou la dilatance.

II. MODÈLE GRANULOMÉTRIQUE

Un premier pas vers la compréhension de l'influence de la polydispersité est d'étudier uniquement son impact sur la texture (propriétés géométriques) d'un assemblage granulaire. Cette étude, réalisée en deux dimensions, nécessite, d'une part le contrôle de la granularité ici limitée à la distribution des tailles de particules et, d'autre part, de créer des empilements denses contenant un grand nombre de particules afin d'assurer la représentativité de l'empilement.

Il existe deux représentations courantes d'une distribution de tailles de particules : la densité de probabilité de tailles ou la courbe granulométrique cumulée. Cette dernière représente la fraction volumique (ou massique) cumulée des particules dont la taille est inférieure à une taille d , en fonction de d variant entre d_{min} et d_{max} . On l'utilise plus particulièrement en mécanique des sols pour caractériser les propriétés mécaniques des sols, mais cela reste, le plus souvent, basé sur des considérations empiriques [1]. Un des objectifs à long terme de ce travail est d'expliquer cette classification par des considérations à l'échelle locale. Pour cela, nous avons développé une méthode de génération de distributions de tailles de particules circulaires à partir d'une courbe granulométrique modèle. Cette méthode, dont on trouvera des détails dans [2], assure la représentativité statistique de chaque classe de tailles de grains dans la distribution. Nous avons choisi la distribution Bêta comme modèle de courbe granulométrique. Elle permet, à partir de seulement deux paramètres a et b , de faire varier la forme de la distribution. On a ainsi accès avec le même modèle à des distributions

classiques comme les distributions bidisperses (deux tailles de particules), la distribution uniforme par taille des particules et la distribution uniforme par fraction volumique des particules. De plus, des courbes granulométriques à double courbure que l'on retrouve souvent dans les sols sont facilement accessibles. Pour faire varier l'étalement de la distribution des tailles, on introduit le paramètre $s = (d_{max} - d_{min}) / (d_{max} + d_{min})$ variant entre 0 (monodisperse) et 1 (infiniment polydisperse).

III. ETUDE GÉOMÉTRIQUE

Pour réaliser l'assemblage des particules générées, on utilise un protocole purement géométrique mimant un dépôt sous gravité. Cela permet d'effectuer le dépôt de 30 000 particules en quelques minutes plutôt qu'en quelques jours pour des simulations dynamiques. On dépose les particules une à une et le plus bas possible sur la surface constituée des particules immobiles déjà déposées. Afin d'éviter les effets de bords, on introduit une condition de périodicité dans la direction horizontale (voir Fig :1).

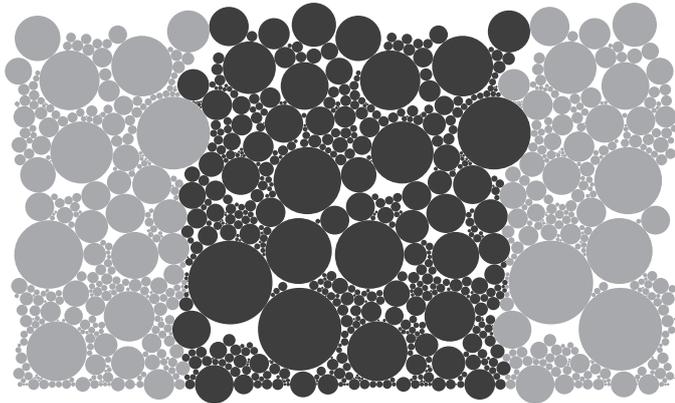


Fig. 1. Echantillon de petite taille déposé géométriquement. Les particules grises sont les images périodiques des particules composant l'échantillon.

Cette étude a permis de mettre en évidence deux phénomènes liés à la dispersion de tailles. Quelle que soit la forme de la distribution granulométrique, il existe un étalement s pour lequel les propriétés de l'assemblage changent radicalement. Par exemple, la figure 2 montre l'évolution de la compacité ρ en fonction de l'étalement s pour différents paramètres de forme a et b . On voit un changement dans l'évolution de la compacité pour $s \simeq 0.4$. A cette valeur d'étalement, les plus petites particules sont suffisamment petites pour s'insérer entre les plus grosses particules (1ère génération d'insertion). En augmentant suffisamment l'étalement de la distribution, ce phénomène d'insertion se reproduit lorsque les petites particules de la génération précédente servent de grosses pour les petites de la suivante. Un autre résultat intéressant est que la distribution par fraction volumique mène aux échantillons les plus compacts et isotropes (propriétés identiques dans toutes les directions de l'espace). Autrement dit, la distribution des tailles des particules et celle des pores générés au cours du dépôt sont proches et mènent ainsi à un remplissage optimal de l'espace.

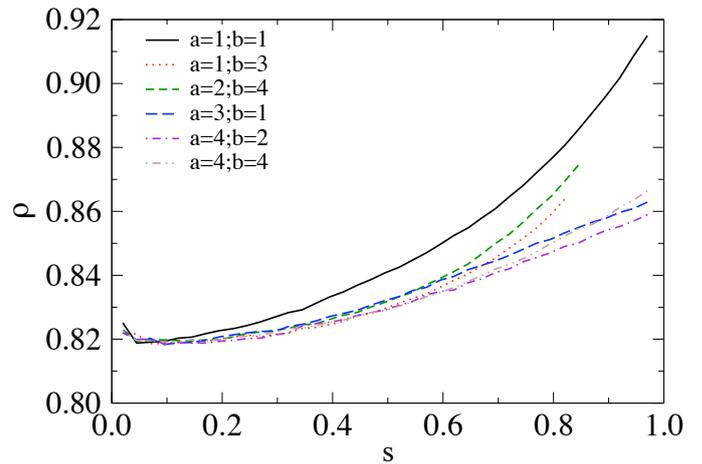


Fig. 2. Compacité des échantillons pour différents paramètres de forme en fonction de l'étalement de la distribution granulométrique.

IV. ETUDE DYNAMIQUE

Au-delà de la géométrie, il est intéressant d'étudier comment se comportent de tels systèmes polydisperses sous sollicitations mécaniques. D'un point de vue macroscopique, il s'agit de cerner l'influence de l'étalement sur les propriétés dynamiques des échantillons. A l'échelle des particules, il s'agit d'identifier le rôle des petites et des grosses particules dans les mécanismes de transmissions des forces.

Pour mener à bien cette étude numérique, nous avons utilisé une stratégie de calcul "Montpellieraine", initiée par J.J. Moreau [3], nommée Dynamique des Contacts (DC). Cette stratégie suppose que les corps simulés sont parfaitement rigides (indéformables) et interagissent par des lois de contact non régulières. Pour les résultats présentés dans ce papier, nous n'avons utilisé que la loi, dite de Signorini, traduisant l'unilatéralité du contact. Elle implique que la vitesse relative normale de deux corps en contact est positive ou nulle, les corps ne peuvent s'interpénétrer. La force normale au contact exercée entre deux particules est non-nulle uniquement si ces particules sont en contact, i.e. les interactions à distance sont exclues. De plus la force normale au contact est positive ou nulle, elle tend à éloigner les corps. A chaque pas de temps de la discrétisation temporelle, on doit trouver les valeurs du couple solution (vitesse relative, force normale) de chaque contact. La force de cette méthode est de gérer une collection de contacts "connectés" tout en levant l'indétermination des forces normales (infinité de forces solutions à vitesse relative nulle). Malgré la puissance de calcul des ordinateurs récents, les simulations dynamiques requièrent souvent plusieurs jours de calculs pour *seulement* une dizaine de milliers de corps.

Les simulations numériques réalisées ont consisté à soumettre les échantillons précédents à une pression verticale constante jusqu'à atteindre un état d'équilibre statique (Fig 3). L'influence de l'étalement est alors identifiable sur plusieurs indicateurs. Par exemple, la distribution de l'intensité des forces de contacts nous renseigne sur l'homogénéité des forces dans le système. En effet, dans un milieu granulaire, même faiblement polydisperse, toutes les particules ne sont pas soumises aux mêmes intensités de force. On décompose généralement les forces en un réseau

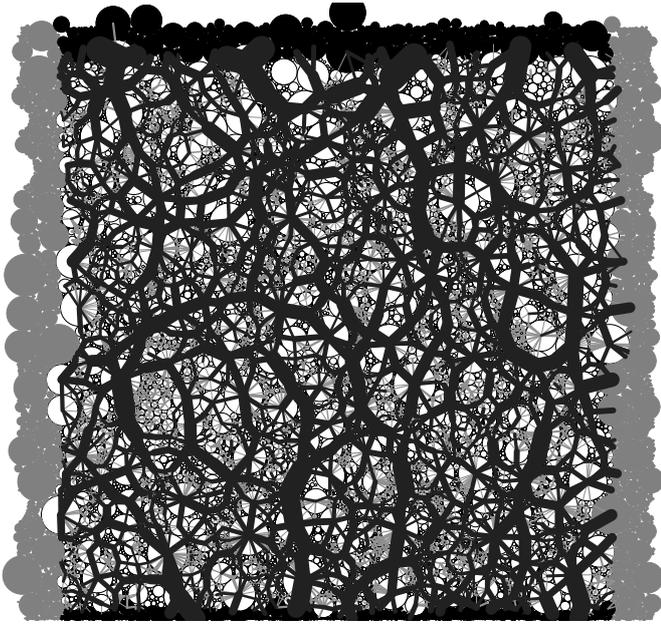


Fig. 3. Echantillon fortement polydisperse ($s = 0.98$) à l'état d'équilibre. La pression est appliquée par le plateau de grains supérieurs (noir) tandis que le plateau inférieur est fixe. La largeur des traits représentant les forces de contacts est proportionnelle à leur intensité. Les forces faibles sont représentées en gris clair tandis que les forces fortes sont représentées en noir.

fort (forces supérieures à la force moyenne) et un réseau faible (forces inférieures à la force moyenne). Le réseau fort est plutôt orienté dans le sens de la sollicitation principale et crée un squelette capable de s'opposer à la sollicitation. Au contraire, le réseau faible a plutôt un rôle de soutien du squelette fort. La figure 4 présente les distributions de forces à l'équilibre pour différents étalements de la distribution uniforme par fraction volumique. L'augmentation de l'étalement de la distribution granulométrique perturbe aussi bien le réseau fort que le réseau faible. La plage de variation des forces fortes augmente avec l'étalement tandis que la probabilité des forces faibles augmente fortement avec l'étalement. En d'autres termes, on dé-homogénéise le système en ce sens que le réseau fort transmet des forces de plus en plus fortes et que le réseau faible mobilise une majorité des forces dont l'intensité faiblit avec l'étalement. Cette dé-homogénéisation des forces est particulièrement importante pour les étalements permettant les effets d'insertions cités plus haut. La figure 5 présente les corrélations entre force moyenne appliquée et taille des particules présentée. On voit qu'au sein d'une même distribution, les grosses particules sont soumises à des forces plus forte que la force moyenne. Un fait intéressant est que la forme de la corrélation change en fonction de l'étalement. Cette transition, correspondant à la transition observée lors de l'étude géométrique, montre un lien fort entre texture et transmission des forces dans les milieux granulaires.

V. CONCLUSION ET PERSPECTIVES

Lors de ces études géométriques et dynamiques, nous avons montré une influence forte de la polydispersité sur la texture et sur les transmissions des forces. Nous avons mis en évidence l'importance de l'étalement d'une distribution granulométrique qui conditionne fortement la structure et

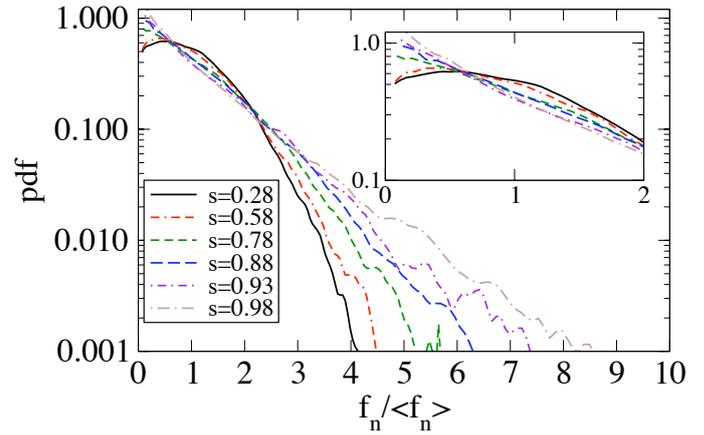


Fig. 4. Densité de probabilité des forces normales pour différents étalements.

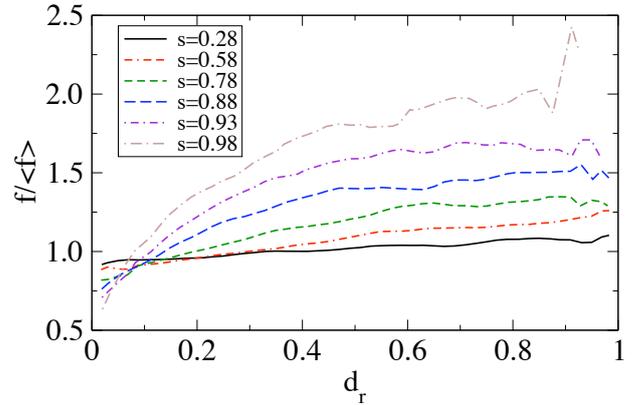


Fig. 5. Force moyenne appliquée aux particules en fonction de leurs tailles pour différents étalements.

les propriétés dynamiques des échantillons. On a montré l'existence d'une valeur critique d'étalement ($s \simeq 0.4$) pour laquelle les mécanismes de transmissions de forces changent.

Les perspectives de ce travail sont multiples. La prochaine étape est de simuler un essai de cisaillement afin d'étudier l'influence de la granulométrie sur le comportement hors équilibre statique. Pour se rapprocher toujours plus de la réalité, une étape importante serait d'effectuer cette étude en trois dimensions. Cela nous permettrait de faire un lien direct avec les sols et d'étudier des phénomènes propres à la troisième dimension comme la ségrégation de taille.

RÉFÉRENCES

- [1] J.K. Mitchell, K. Soga, *Fundamentals of Soil Behavior, third edition* (Wiley, 2005)
- [2] C. Voivret, F. Radjai, J.Y. Delenne, M.S. El Youssoufi, *Phys. Rev. E* **76**(2), 021301 (2007),
- [3] J.J. Moreau, *European Journal of Mechanics A/Solids* **supp.**(4), 93 (1994),

Approche discrete de la fragmentation de l'albumen de blé.

Vincent TOPIN, Jean-Yves DELENNE, Farhang RADJAI
LMGC, UMR CNRS - Université Montpellier 2
34095 Montpellier cedex 5

topin@lmgc.univ-montp2.fr, delenne@lmgc.univ-montp2.fr, radjai@lmgc.univ-montp2.fr

Résumé— La dureté de l'albumen de blé est une propriété importante pour la détermination de la qualité des farines et pour la sélection des variétés dans l'industrie agro-alimentaire. En pratique, la dureté correspond plus à la distribution granulométrique des agrégats après mouture qu'à une propriété mécanique intrinsèque de l'albumen. L'objectif principal de ce travail est d'isoler les propriétés mécaniques sous-jacentes à cette notion de dureté. Dans cet article, un modèle granulaire cohésif prenant en compte à la fois les granules d'amidons et une phase continue (la matrice protéique) est introduit. Ce modèle, fondé sur une discrétisation sur réseau, permet d'intégrer de manière simple la fissuration en incorporant la possibilité de rupture à l'échelle des éléments du réseau. Nous présentons une étude paramétrique détaillée réalisée en fonction de deux paramètres : l'adhérence à l'interface entre les granules d'amidon et la protéine et la quantité de protéine. Plusieurs modes de rupture distincts de l'albumen sont mis en évidence en fonction de ces deux paramètres. On observe que les données relatives au taux d'amidons endommagés, en tant qu'indicateur de la dureté du blé, peuvent être mises à l'échelle en utilisant un paramètre homogène à une ténacité qui combine l'énergie d'interface entre amidon et protéine et la rigidité liée à la quantité de protéine. Cette étude suggère que ce paramètre gouverne la transition entre les comportements de type "soft" et "hard".

Mots-clés— Matériau granulaire cohésif, albumen de blé, cimentation, friabilité, modèle sur réseau, fissuration.

I. INTRODUCTION

L'albumen est le principal constituant du grain de blé. Il est composé essentiellement de granules d'amidon, reliés entre eux par l'intermédiaire d'une matrice protéique laissant place à des pores (Fig. 1 (a)) [1]. La variabilité de la dureté du blé est un aspect important dans la détermination de la qualité des farines dans l'industrie alimentaire. Les classifications commerciales habituelles différencient le blé tendre (soft et hard) du blé dur (durum). Alors que les blés soft nécessitent moins d'énergie à la mouture et produisent des farines fines constituées de petits agrégats, les blés de type hard et durum sont réduits à des agrégats plus volumineux et occasionnent un plus grand nombre de granules d'amidon endommagés [1]. En pratique, la dureté du grain de blé correspond plus à une description empirique de la distribution de la taille des agrégats après mouture qu'à une propriété mécanique du grain. Cependant, il est généralement admis que la dureté est une conséquence de l'adhérence entre les granules d'amidon et la matrice protéique [2; 3] ce qui se traduit d'un point de vue génétique par la présence et la nature de protéines appelées puroindolines. D'autre part, la dureté est aussi fonction de la teneur en protéine dans l'albumen. Cette dernière est une caractéristique liée principalement aux conditions de culture

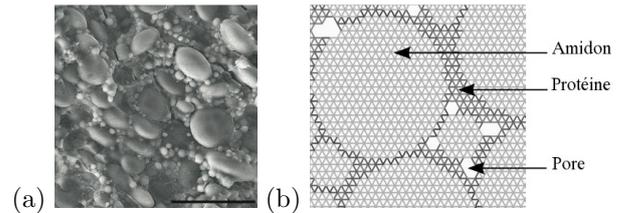


Fig. 1. (a) Vue au microscope électronique de l'albumen de blé, la barre d'échelle correspond à $50\mu\text{m}$. (b) Représentation des particules d'amidon, de la matrice protéique et des pores sur un réseau triangulaire.

et se reflète dans une propriété optique du grain : la vitrosité.

Dans cet article, nous introduisons un modèle de l'albumen de blé prenant en compte les particules d'amidon, la matrice continue de protéine et les pores. La discrétisation de ces phases en éléments linéaires portés par un réseau permet d'initier la rupture d'une manière simple. Le comportement et la rupture de ce milieu dépendent des propriétés des phases et de leurs interfaces. Nous nous intéressons notamment à l'influence de deux paramètres majeurs pour la cohésion de l'albumen : 1) l'adhérence entre les phases ; 2) la fraction volumique de la matrice protéique. Nous montrons la dépendance du module élastique et de la résistance à la rupture par rapport à ces deux paramètres. Une étude paramétrique sera également présentée sur l'endommagement des particules d'amidon en fonction de ces paramètres. Un résultat essentiel de cette étude est la mise en évidence d'un seuil d'endommagement des particules d'amidon qui se met à l'échelle avec une grandeur homogène à une ténacité combinant la quantité de protéine à l'adhérence à l'interface entre la matrice et les particules.

II. DESCRIPTION DU MODÈLE NUMÉRIQUE

L'albumen de blé peut être considéré comme un matériau granulaire composé de particules d'amidon reliés les unes aux autres par la matrice protéique. Dans ce cadre, le passage des interactions microscopiques aux propriétés macroscopiques du milieu est intimement lié à la nature granulaire de l'albumen. L'objectif est de prendre en compte simultanément une phase continue et des particules discrètes, en permettant la rupture dans chacune de ces deux phases et de leurs interfaces. A cette fin, nous avons utilisé un modèle de discrétisation sur réseau appelé LEM (Lattice Element Method). Ce type de modèle a été souvent utilisé en mécanique statistique de la rupture [4; 5], et appliqué

pour étudier la fissuration du béton, des céramiques ou encore des sols [6–8]. L'espace est discrétisé par une grille de nœuds, régulière ou non, connectés par des éléments 1D. Dans notre étude, nous avons utilisé des éléments élastiques fragiles afin de définir un réseau triangulaire régulier avec une longueur d'élément à l'équilibre a . Chaque élément appartenant à une phase ϕ est caractérisé par une raideur k^ϕ ainsi que par un seuil de rupture f^ϕ en force. On définit les paramètres $\sigma^\phi \equiv f^\phi/a$ et $E^\phi \equiv k^\phi/a$, homogènes respectivement à une contrainte et un module d'élasticité.

Dans le cas des matériaux polyphasiques, chaque phase ainsi que ses frontières sont matérialisées par des éléments partageant les mêmes propriétés mécaniques. Dans le cas de l'albumen de blé, ces phases correspondent aux particules d'amidon, à la matrice protéique ainsi qu'aux pores, voir fig. 1 (b) [9]. Les échantillons sont déformés sous des tests en traction ou compression simple en imposant des déplacements sur les bords supérieurs et inférieurs et en laissant libres les bords latéraux. L'état initial correspond à la configuration de référence et l'équilibre à chaque pas est déterminé en cherchant le minimum de l'énergie potentielle totale du système. À chaque pas, tous les éléments excédant un certain seuil sont supprimés, cette opération ayant pour conséquence la fissuration du réseau. Des cycles de relaxation rendent possible la propagation des fissures à l'intérieur du même pas.

Nous avons généré des échantillons bidisperses (voir fig. 1 (a)) avec un rapport de taille de particule égal à quatre. L'ensemble de ces particules représente une fraction totale de 80 % de l'échantillon. Un assemblage de disques rigides est d'abord généré à l'aide d'un code de dynamique des contacts, afin d'en isoler une portion rectangulaire et de la superposer au réseau triangulaire. La matrice protéique est ensuite ajoutée sous la forme de ponts trapézoïdaux connectant les particules voisines suivant un critère de distance. En faisant varier l'épaisseur de ces ponts, on obtient des échantillons dont la fraction volumique de matrice ρ^m évolue de 4% (porosité élevée) à 20% (porosité réduite à zéro).

III. ÉTUDE PARAMÉTRIQUE

Dans cette section, l'influence de la fraction volumique de matrice protéique et de l'adhésion de l'interface entre la matrice et les particules sur le comportement à la rupture de l'albumen est étudiée. Une des spécificités de l'albumen de blé est que ses constituants présentent des propriétés mécaniques similaires [3]. Nous avons donc ici attribué les mêmes valeurs à la raideur et au seuil de rupture pour la phase protéique et amidon. La fraction volumique d'amidon a été fixée à 0,8 en 2D. Une étude paramétrique en traction est effectuée en faisant varier la quantité de protéine ρ^m de 0,04 à 0,2 et le paramètre d'adhésion à l'interface, correspondant au seuil de rupture σ^{pm} entre les particules et la matrice, de 0,3 à 1,05 fois celui des particules ($\sigma^{[p]}$). Tous les autres paramètres géométriques et mécaniques ont été fixés dans ce cadre, et la cohésion entre les particules a été maintenue nulle. Chaque échantillon contient près de 120000 éléments avec environ 500 granules, et chaque simulation a été répétée avec 3 configurations indépendantes.

Les valeurs des modules effectifs E_{eff} des différents échantillons sont reportées à la figure 2 en fonction de ρ^m .

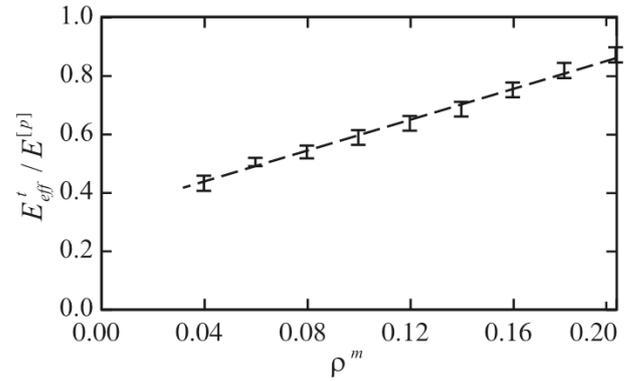


Fig. 2. Module d'élasticité en fonction de la fraction volumique de matrice.

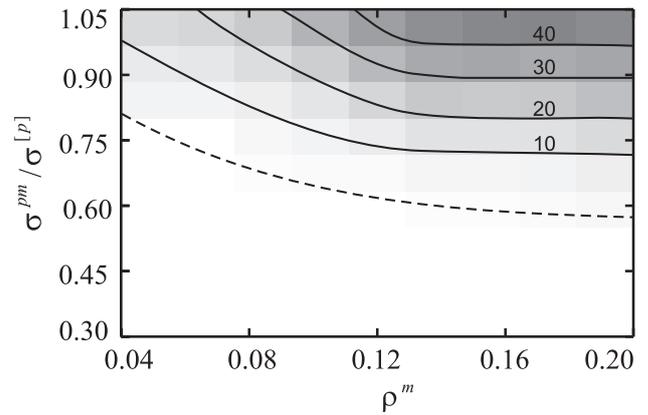


Fig. 3. Carte en niveau de gris de la fraction d'élément rompu dans la phase particulaire dans l'espace paramétrique.

On observe que E_{eff} augmente linéairement avec ρ^m . Le module effectif dépend donc uniquement de la porosité du milieu ainsi que des zones non-cohésives (clivages) entre particules d'amidon.

Dans les modèles sur réseau, il est classique de considérer le nombre d'éléments rompus comme une mesure de l'endommagement [5]. Étant donné que dans les processus liés à la transformation du blé, le taux de granules d'amidon endommagés est un indicateur de la dureté du grain [1], nous nous intéressons plus particulièrement ici la proportion n_b d'éléments rompus dans la phase particulaire par rapport au nombre total d'éléments rompus. La figure 3 montre une carte en niveau de gris représentant n_b dans l'espace paramétrique. On remarque qu'au-dessous d'une frontière clairement définie (ligne en pointillés), il n'y a pas de granules d'amidon endommagés. Cette carte nous permet de distinguer trois principaux régimes de fissuration (fig. 4) : 1) En dessous de la limite d'endommagement des granules d'amidon, la fissure contourne les particules et se propage à travers les pores et l'interface entre l'amidon et la protéine. 2) Au-delà de cette limite et pour $\rho^m < 0,12$ la fissure pénètre partiellement dans les granules depuis les ponts protéiques qui concentrent fortement les contraintes, ce qui mène à l'abrasion de la surface des granules. 3) Enfin, au-delà de cette limite et pour $\rho^m > 0,12$ la fissure se propage aussi bien dans la protéine qu'à travers les granules.

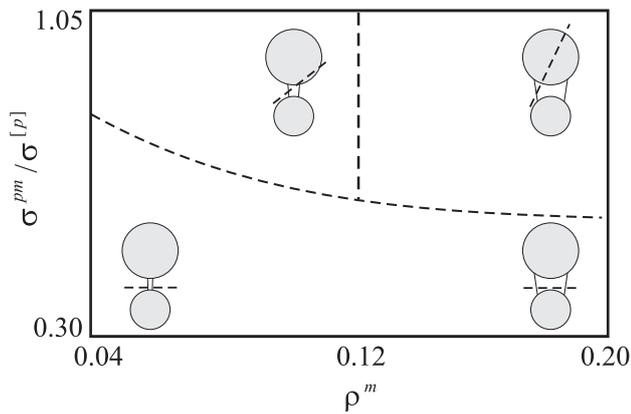


Fig. 4. Représentation schématique des régimes de rupture.

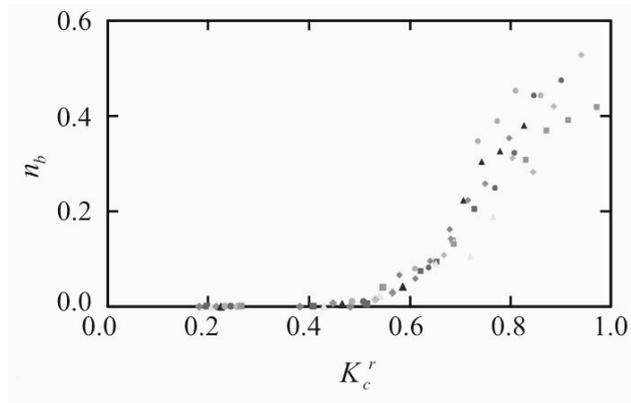


Fig. 5. Fraction des éléments rompus dans la phase des particules en fonction de la ténacité relative.

L'endommagement des granules est contrôlé par une combinaison de la quantité de matrice et de l'adhérence entre la protéine et l'amidon. En mécanique de la rupture, la ténacité $K_c = (EG_c)^{1/2}$ combine la rigidité E et l'énergie par unité de surface G_c nécessaire pour initier une fissure. Nous définissons donc une ténacité relative $K_c^r \equiv K_c^{pm}/K_c^p$ en rapportant la ténacité K_c^{pm} liée à l'interface entre les particules et la matrice à la ténacité K_c^p liée à la phase particulaire. La rigidité E de nos échantillons étant corrélée linéairement avec la quantité de matrice (voir fig. 2), K_c^r combine donc explicitement les deux paramètres de notre étude, la fraction volumique de matrice ρ^m et l'adhérence à l'interface σ^{pm} .

Dans la figure 5, nous avons tracé la fraction n_b des éléments rompus dans la phase particulaire en fonction de la ténacité relative K_c^r . On remarque que tous les points se regroupent sur une seule courbe avec une variabilité liée aux fluctuations entre les échantillons. En dessous d'une ténacité caractéristique $K_c^r = 0,5$ aucun granule n'est endommagé. Au-delà de ce point, la fraction d'éléments rompus augmente linéairement avec K_c^r . Cette ténacité caractéristique correspond à la transition d'un régime où les fissures sont défléchies dans l'interface particule-matrice, à un régime où elles peuvent pénétrer à l'intérieur des particules.

IV. CONCLUSION

Dans ce travail, une méthodologie est proposée pour étudier les propriétés de résistance et la friabilité de l'albumen de blé en fonction de deux paramètres : la quantité de matrice protéique et la nature de l'interface entre les particules d'amidon et la protéine. Le modèle numérique sur réseau introduit permet de discrétiser simultanément les particules et la matrice et intègre de manière simple la possibilité de fissuration dans chaque phase par la rupture des éléments du réseau. Une étude paramétrique a permis de quantifier l'influence de la fraction volumique de protéine et de l'adhérence entre l'amidon et la protéine. L'observation de la fraction d'éléments rompus dans la phase amidon, en tant qu'indicateur de la dureté du grain, suggère qu'il existe une transition entre les comportements soft et hard contrôlée par une grandeur unique. Cette grandeur, homogène à une ténacité, combine la quantité de matrice et la nature de l'interface. D'un point de vue plus général, cette étude met en évidence l'influence forte de la percolation des particules sur la réponse mécanique des matériaux granulaires cimentés comme le béton ou certains conglomerats. Elle permet de distinguer ce type de matériau des composites renforcés par des particules dilués, dont les propriétés mécaniques avant rupture peuvent être décrites en fonction des fractions volumiques des différentes phases.

RÉFÉRENCES

- [1] W.A. ATWELL : *Wheat Flour*. AACCC, St Paul USA, 2001.
- [2] K.K. BARLOW, M.S. BUTTROSE, D.H. SIMMONDS et M. VESK : The nature of the starch-protein interface in wheat endosperm. *Cereal Chemistry*, 50:443-454, 1973.
- [3] G.M. GLENN et R.K. JOHNSTON : Mechanical properties of starch, protein and endosperm and their relationship to hardness in wheat. *Food Structure*, 11:187-199, 1992.
- [4] S. ROUX : *Statistical Models for Fracture in Disordered Media*, chapitre Continuum and discrete description of elasticity and other rheological behavior, pages 87-114. North Holland, Amsterdam, 1990.
- [5] H. J. HERRMANN et S. ROUX, éditeurs. *Statistical Models for Fracture in Disordered Media*. North Holland, Amsterdam, 1990.
- [6] E. SCHLANGEN et E. J. GARBOCZI : Fracture simulations of concrete using lattice models : Computational aspects. *Engineering Fracture Mechanics*, 57(2-3):319-332, 1997.
- [7] J. G. M. VAN MIER, B. M. CHIAIA et Adri VERVUURT : Numerical simulation of chaotic and self-organizing damage in brittle disordered materials. *Computer Methods in Applied Mechanics and Engineering*, 142(1-2):189-201, mars 1997.
- [8] V. TOPIN, J.-Y. DELENNE, F. RADJAÏ, L. BRENDÉL et F. MABILLE : Strength and fracture of cemented granular matter. *The European Physical Journal E*, 23:413-429, 2007.
- [9] Vincent TOPIN, Farhang RADJAÏ, Jean-Yves DELENNE, Abdelkrim SADOUDI et Frederic MABILLE : Wheat endosperm as a cohesive granular material. *Journal of Cereal Science*, In Press, Corrected Proof, 2007.

MESURE DU COMPORTEMENT MECANIQUE DES COUCHES DE LA PAROI CELLULAIRE DU BOIS PAR MICROSCOPIE A FORCE ATOMIQUE

KARL BYTEBIER ¹ (2^{EME} ANNEE), JOSEPH GRIL ²

Laboratoire de Mécanique et Génie Civil
UMR 5508

Université MONTPELLIER II
CC 048 Place Eugène Bataillon
34095 Montpellier cedex 5

BYTEBIER@LMGC.UNIV-MONTP2.FR

GRIL@LMGC.UNIV-MONTP2.FR

[HTTP://WWW.EDI2S.UNIV-MONTP2.FR/](http://WWW.EDI2S.UNIV-MONTP2.FR/)

Résumé— Le bois est un matériau complexe qui n'a pas encore livré tous ses secrets. La détermination des caractéristiques mécaniques du bois se heurte à plusieurs obstacles, notamment le fait que le bois présente une structure « imparfaite », très variable d'un arbre à un autre et extrêmement sensible aux données climatiques (température et hygrométrie). Le but de la présente étude est de déterminer les propriétés mécaniques du bois à l'échelle des cellules, où l'on attend le moins de variabilité. La technique utilisée repose sur l'utilisation de la microscopie à force atomique et d'un mode d'utilisation à contact vibrant. Les résultats permettent la comparaison avec des matériaux étalons.

Mots clés— Microscope à force atomique, parois cellulaires, bois, propriétés mécaniques

Biographie— Après l'obtention du baccalauréat, j'ai fait un DUT Mesures Physiques, puis un IUP Génie Mécanique. Durant cette période est passée la loi LMD qui m'a permis de faire un Master Recherche Mécanique et Ingénierie. J'ai alors effectué mon stage de deuxième année au sein de l'équipe Mécanique de l'Arbre et du Bois au LMGC dont le sujet était l'étude des propriétés mécaniques du bois à l'échelle microscopique par Microscopie à Force Atomique (AFM). Ma thèse est la suite de ce stage.

I. INTRODUCTION

Le bois, tout le monde connaît : matériau vivant dans l'arbre, matériau de construction dans nos maisons, source d'énergie pour nos cheminées. Malgré cela, il recèle bien des surprises, que ce soit au niveau de sa structure ou de son comportement. Bien sûr nous l'avons étudié, observé, nous connaissons sa structure et commençons à comprendre son comportement, mais il reste beaucoup de secrets à découvrir, notamment au niveau de ses caractéristiques mécaniques.

Le bois est l'objet de nombreuses études visant à caractériser son comportement mécanique. Certaines d'entre elles reposent sur des tests de micro-traction [1], ou encore l'utilisation d'ondes ultrasonores [2]. Ces techniques se montrent très intéressantes, mais nécessitent des traitements mécaniques ou chimiques pour les tests de micro-traction, et ne permettent d'obtenir dès résultats qu'au niveau d'une ou quelques cellules. C'est la raison pour laquelle certaines études se sont intéressées à la nanoindentation [3], ainsi qu'à un appareil tout particulier, le microscope à force atomique ou AFM [4], ces techniques permettant d'atteindre des résolutions de l'ordre de quelques dizaines de nanomètres. Ma thèse porte sur l'utilisation de cet appareil dans la détermination des caractéristiques mécaniques du bois à l'échelle des parois cellulaires.

II. LE BOIS

Le bois est un matériau complexe présentant plusieurs niveaux de structure (Fig.1).

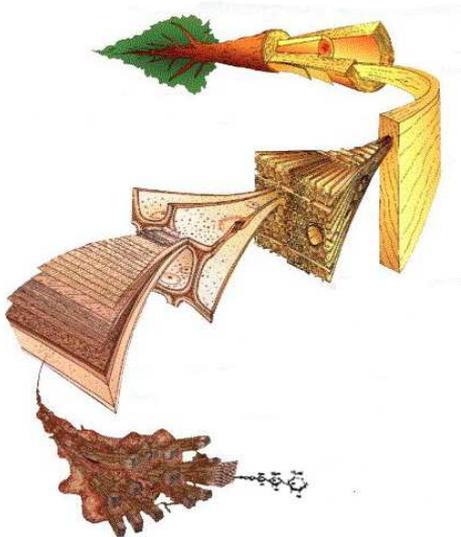


Fig.1. Structure multiéchelle du bois [5]

Le premier niveau est le niveau de l'arbre tel qu'on peut le voir en forêts (1 cm à 10 m). Puis on atteint le niveau du cerne (1 à 10 mm), qui est composé de plusieurs sous ensembles regroupés sous forme de tissus (0,1 à 1 mm) qui sont constitués de cellules de même nature. Ces cellules (20 à 50 μm) présentent une structure multicouche composite (0,1 à 10 μm) qui leur garantit une tenue mécanique quelles que soient les sollicitations mécaniques que peut subir l'arbre. Enfin on atteint le niveau moléculaire des microfibrilles (quelques nm).

III. L'AFM

Inventé en 1982 [6], l'AFM est un microscope qui possède la particularité de ne pas reposer sur un système optique, mais sur l'interaction d'une pointe avec la surface d'un échantillon, la réalisation des images étant effectuée à partir des mouvements de la pointe (Fig. 2).

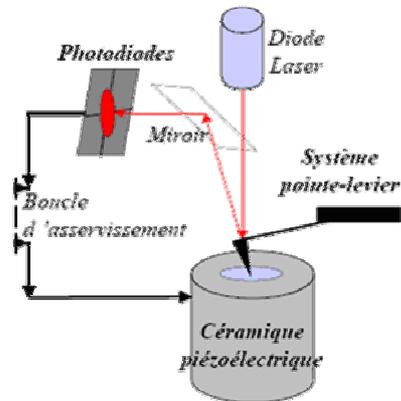


Fig.2. Montage classique d'un AFM.

La pointe, qui est accrochée à un levier, parcourt la surface de l'échantillon. Son déplacement vertical, fonction de la topographie ainsi que de certaines propriétés mécaniques de la surface, est détecté par un système diode/laser. Il est possible d'asservir ce mouvement par l'intermédiaire d'une boucle, permettant l'utilisation de différents modes.

IV. CARTOGRAPHIE DES PROPRIETES MECANIQUES

L'AFM propose différents modes d'utilisation. Le mode de base est dit "mode contact", il permet de visualiser la topographie de l'échantillon. Le mode que nous utilisons dans le cadre de ma thèse est dit « mode contact vibrant » [7], le levier au bout duquel est accroché la pointe est mis en vibration par l'intermédiaire d'un générateur électrique produisant une excitation électrostatique. Cette technique permet non seulement l'imagerie de la topographie, mais aussi, grâce à un traitement numérique adéquat [8], la cartographie des fréquences de résonance et

de la mi-largeur à mi-hauteur, fonction de l'élasticité et de la viscosité respectivement.

Les premiers résultats mettent en avant des différences entre les couches cellulaires (Fig. 3).

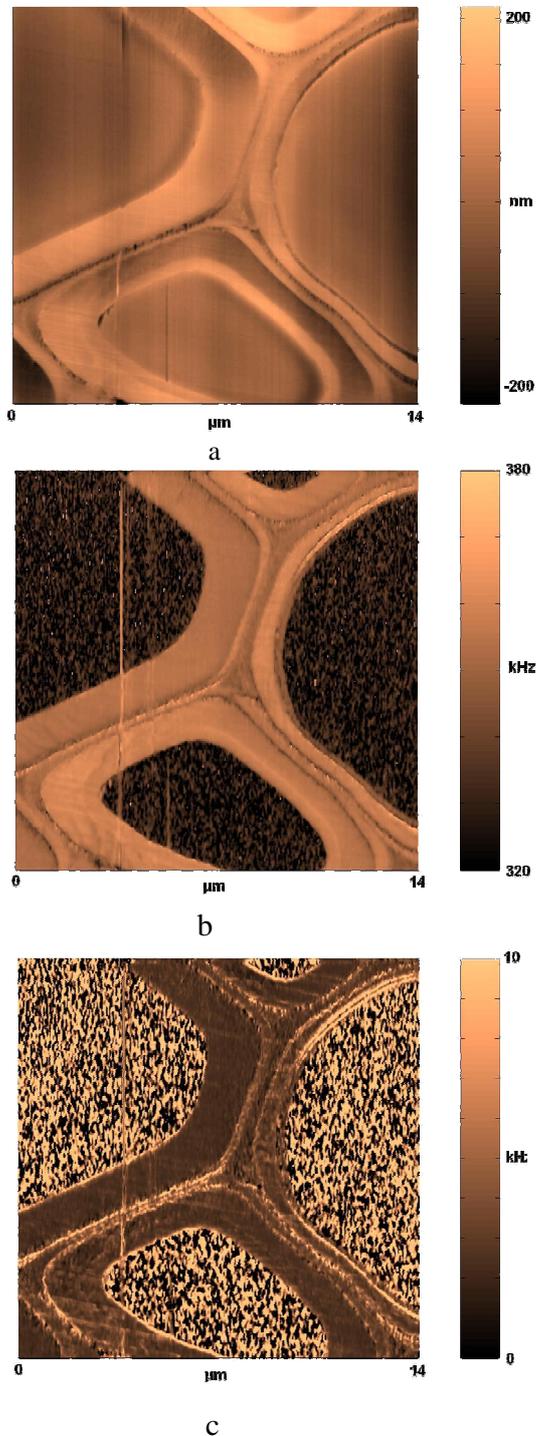


Fig. 3. (a) Topographie de l'échantillon; (b) Image de fréquence de résonance; (c) Image de mi-largeur à mi-hauteur.

L'image de fréquence de résonance permet la comparaison entre le bois et d'autres matériaux aux caractéristiques mécaniques connus dits « étalons » si l'on trace les spectres de fréquence (Fig 4).

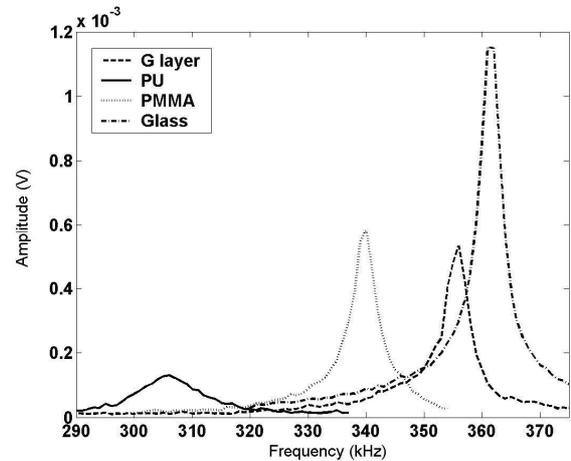


Fig. 4. Comparaison des spectres de fréquences entre plusieurs matériaux, notamment la couche G qui est une couche des parois cellulaires du bois.

On peut ainsi constater que la rigidité « apparente » (donc ne prenant pas en compte l'anisotropie) de la couche G par exemple, est située entre celle du PMMA (1,8 à 3,1 GPa) et celle du verre (70 à 72 GPa).

V. CONCLUSION

A travers la présentation du matériau bois, nous avons pu nous rendre compte de sa complexité et de son aspect multiéchelles. Certaines équipes de chercheurs ont été capables de déterminer ses caractéristiques mécaniques à l'échelle du cerne et de certains de ses constituants chimiques. Mais les données concernant les échelles inférieures manquent, particulièrement au niveau de la paroi cellulaire. Une solution à ce problème était l'utilisation de l'AFM.

La technique de contact vibrant que nous avons utilisée, qui se base sur la mise en vibration du système levier/matériau pour remonter aux caractéristiques de ce dernier, permet d'obtenir des images qualitatives de fréquences de résonance et de mi-largeur à mi-hauteur. Mais la détermination des caractéristiques mécaniques par cette AFM se heurte encore à certains problèmes. En effet, le principe de l'AFM reposant sur l'utilisation d'une pointe, l'aire de contact (et donc la raideur de contact) entre celle-ci et la surface de l'échantillon dépend des données géométriques de la pointe. A l'heure actuelle, l'incertitude sur des grandeurs telles que le rayon de la pointe ou encore la raideur du levier, ne permettent pas de quantifier précisément les caractéristiques mécaniques des échantillons.

VI. REFERENCES

- [1] Burgert et al., A Comparison of Two Techniques for Wood Fibre Isolation - Evaluation by Tensile Tests on Single Fibres with Different Microfibril Angle, *Plant biol.*, 2002, 4, 9-12
- [2] Keunecke et al., Determination of Young's and shear moduli of common yew and Norway spruce by means of ultrasonic waves, *Wood. Sci. Technol.*, 2007, 41, 309–327
- [3] Wimmer et al., Longitudinal hardness and Young's modulus of spruce tracheid secondary walls using nanoindentation technique, *Wood. Sci. Technol.* , 1997, 31, 131–4
- [4] Clair et al., Imaging the mechanical properties of wood cell wall layers by atomic force modulation microscopy, *IAWA Journal*, 2003, 24 (3), 223–230
- [5] Butterfield et al., Microfibril angle in wood, University of Canterbury, 1998, 409
- [6] Binnig et al, *Phys. Rev. Letters*, 1986, 56, 930
- [7] Arinéro and Lévêque, Vibration of the cantilever in Force Modulation analysis by finite element model *Rev. Sci. Instrum.*, 2003, 74, 104
- [8] Arinéro et al., Image processing for resonance frequency mapping in atomic force modulation microscopy, *Rev. Sci. Instrum.*, 2007, 74, 023703

Un modèle poroélastique du remodelage osseux sensible à la fréquence de chargement

Etienne MALACHANNE¹ (Doctorant 3e année), Franck JOURDAN² (Directeur de Thèse)

L M G C

UMR 5508 Université Montpellier 2/CNRS, cc048 Place Eugène Bataillon, 34095 Montpellier cedex 5, France

¹malachanne@lmgc.univ-montp2.fr, ²jourdan@lmgc.univ-montp2.fr

Résumé— L'os est un matériau complexe, qui peut-être assimilé à un milieu poreux. Par ailleurs c'est un matériau vivant qui évolue en permanence et se renouvelle, c'est le phénomène de remodelage osseux. L'objectif de ce travail est d'établir un modèle numérique de remodelage osseux qui prend en compte le comportement poroélastique de l'os et les différents chargements qui lui sont appliqués. Le modèle numérique proposé, validé sur des travaux expérimentaux, rend compte de l'influence de la pression interne dans l'os, ainsi que l'effet de la fréquence de chargement sur le remodelage osseux.

Mots-clés— Poroélasticité, Remodelage Osseux, Éléments finis.

I. INTRODUCTION

L'étude du remodelage osseux débute au XIX^e siècle, avec les observations cliniques faites par un chirurgien allemand [1]. D'après la théorie de Wolff l'architecture interne de l'os change en fonction des différents types de chargement qui lui sont appliqués. D'une manière générale, un os sollicité en contrainte, par exemple par l'intermédiaire d'une activité physique, va avoir une masse volumique qui augmente. Inversement un os peu sollicité aura une densité qui va diminuer. La loi de Wolff puis la loi de Stanford [2], [3] lie la contrainte ou le taux de contrainte et la variation de densité osseuse. Par ailleurs l'os est d'un point de vue mécanique un matériau très complexe. C'est un milieu poreux, saturé, avec des interactions entre les parties fluides et les parties solides. Ainsi une loi de comportement poroélastique issue de la mécanique des sols [4], [5] permet de tenir compte des différents paramètres de cette structure complexe. L'objectif de ce travail va être d'adapter une loi de remodelage du type de celle de Stanford, afin de prendre en compte l'historique de chargement d'un os, et d'implanter une loi de comportement poroélastique dans la loi de remodelage.

II. MÉTHODE ET PROCÉDURE

La théorie de la poroélasticité permet de prendre en compte la structure complexe de l'os et les interactions entre les phases solides et les phases fluides. L'ouvrage [6] détaille les équations mises en jeu du modèle poroélastique : les lois de conservations de la partie fluide et de la partie solide, et les équations de couplage. On rappellera ici seule-

ment la loi de comportement liant la contrainte mécanique et la déformation :

$$\underline{\underline{\sigma}} = \mathbf{D}\underline{\underline{\epsilon}} - bp\underline{\underline{I}}$$

où $\underline{\underline{\sigma}}$ est le tenseur des contraintes, $\underline{\underline{\epsilon}}$ est le tenseur des déformations, \mathbf{D} est le tenseur de Hooke qui est le tenseur de rigidité, p est la pression interne et b le coefficient de Biot. D'autre part, une personne adulte en bonne santé et ayant une activité physique normale, gagne autant d'os que ce qu'il en perd. En revanche un sportif de haut niveau va gagner jusqu'à 18% d'os, quand une personne sédentaire au contraire va perdre de la masse osseuse. Le processus de remodelage est directement lié à la contrainte qui est appliquée à l'os. On introduit donc un paramètre qui correspond à un taux de contrainte quotidien, et qui est défini de la manière suivante :

$$\psi = \frac{1}{T} \sum_i T_i \bar{\sigma}_i g(\|W\|, \|\Gamma\|) \quad (1)$$

où T est la durée d'une journée, T_i la durée du chargement i , $\bar{\sigma}_i$ la contrainte moyenne et $g(\|W\|, \|\Gamma\|)$ une fonction de la vitesse du fluide interstitiel en mouvement, appelée vitesse de Darcy, qui se définit comme suit :

$$W = H \underline{\underline{\text{grad}}} p$$

avec $H = 1.1 \times 10^{-13} \text{ m}^4 \text{ N}^{-1} \text{ s}^{-1}$ la perméabilité macroscopique de l'os cortical [7], et p la pression interstitielle. $\|\Gamma\|$ est la norme de l'accélération de Darcy. ψ est exprimé en Pa s^{-1} . Le paramètre ψ dépend ainsi de l'historique de chargement quotidien. En effet pour chaque type de chargement il tient compte l'amplitude et la durée et aussi de la fréquence à travers la fonction de l'accélération et de la vitesse de Darcy. Le taux de contrainte de référence ψ_{ref} , sera défini à l'aide de (1), en ne prenant que des chargements liés à une activité physique normale. La loi de remodelage, qui s'inspire de la loi de Stanford [2], [3], s'écrit alors :

$$\dot{r} = \begin{cases} c(\psi - \psi_{ref}) + cw & \text{si } \psi - \psi_{ref} < -w \\ c(\psi - \psi_{ref}) - cw & \text{si } \psi - \psi_{ref} > w \\ 0 & \text{si } -w < \psi - \psi_{ref} < w \end{cases}$$

où \dot{r} est un taux de remodelage, c est pris comme étant l'inverse de ψ_{ref} , et w est égal à 10% de ψ_{ref} . La variation de la densité massique osseuse est liée à la variation du taux de remodelage par la relation suivante :

$$\dot{\rho} = \dot{r} S_v \rho_t$$

où S_v est une surface spécifique [8] et ρ_t la densité de tissu osseux totalement minéralisé. Par ailleurs il convient d'appliquer une limite à l'augmentation de la masse osseuse, ainsi au delà d'un gain de 30% on considère que $\psi_{ref} = \psi$.

III. RÉSULTATS NUMÉRIQUES

Le modèle de remodelage présenté précédemment est validé sur les travaux expérimentaux de [9]. La méthode des éléments finis est utilisée pour réaliser une modélisation numérique. Un maillage axisymétrique d'un cubitus de dindon est réalisé et chargé dans un code élément fini. Une pression cyclique de faible amplitude et de haute fréquence (8000 Pa, 20 Hz) est imposée sur la surface interne de l'os, 10 minutes par jour pendant 28 jours. La valeur initiale de la densité massique osseuse est prise égale à 0.5 g cm^{-3} . La Figure 1 montre la répartition de la densité massique osseuse sur le cubitus de dindon après 28 jours.

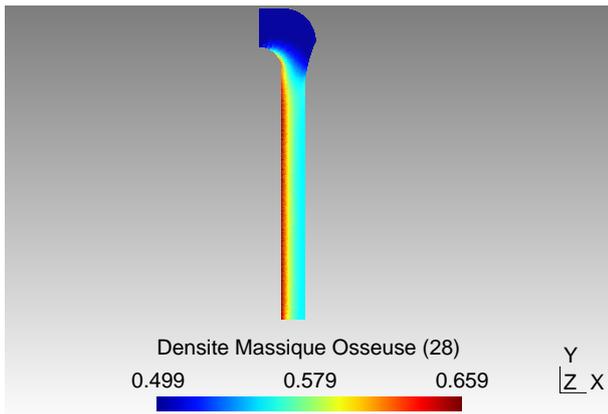


Fig. 1. Modélisation axisymétrique d'un cubitus. On constate que le remodelage le plus important est retrouvé là où la pression est la plus élevée

De plus la Figure 2 montre une dépendance entre la fréquence et le taux de remodelage. En effet celui-ci augmente avec la fréquence, jusqu'à un certain seuil correspondant à un gain maximum de masse osseuse.

IV. DISCUSSION

L'étude expérimentale de [9], compare une population de dindons sédentaires à une population recevant une pression cyclique de faible amplitude et de haute fréquence (8000 Pa, 20 Hz), 10 minutes par jour pendant 28 jours. Les résultats montrent un gain significatif de masse osseuse à la fin de l'expérience, l'augmentation est en moyenne de 18% avec des pics allant jusqu'à 35% dans les zones où le gradient de pression est le plus élevé. L'amplitude de la pression imposée étant faible, cette étude montre l'importance de l'effet de la fréquence sur le processus de remodelage osseux. Une seconde étude expérimentale [10] réalisée sur des cubitus de rat, montre une dépendance entre le

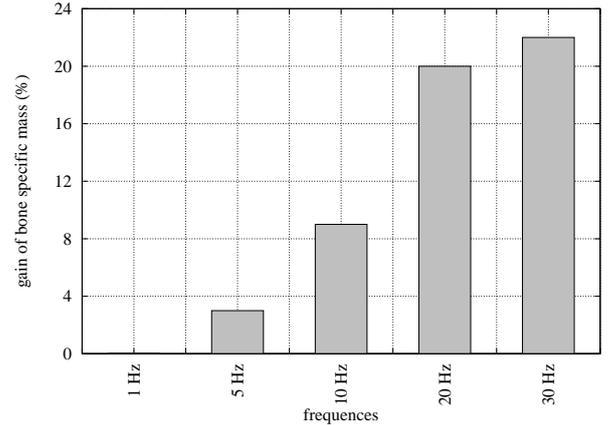


Fig. 2. Evolution moyenne de la densité massique osseuse en fonction de la fréquence de sollicitation

taux de remodelage et la fréquence. En effet sur un domaine de fréquence compris entre 5 Hz et 30 Hz, l'étude montre un remodelage plus significatif à partir de 10 Hz. Le modèle numérique de remodelage osseux proposé dans notre étude, permet de prendre en compte les différentes caractéristiques de chaque chargement : la proportion dans le temps, l'intensité mais aussi la pression interne de l'os et la fréquence à travers la vitesse et l'accélération de Darcy. Ces dernières caractérisent les mouvements de fluide interstitiel et sont souvent évoqués dans la littérature [11], [12], [13] comme ayant un effet sur le remodelage osseux. Les résultats numériques obtenus pour un chargement identique à celui de l'étude expérimentale [9], montrent un maximum de remodelage dans les zones où la pression est la plus élevée (environ 30% d'augmentation, voir Figure 1) et un gain moyen d'environ 20%. Ces résultats numériques sont donc en accord avec les résultats expérimentaux. Par ailleurs les résultats présentés sur la Figure 2 permettent aussi de retrouver une dépendance entre la fréquence et le remodelage osseux, le gain de densité massique osseuse augmentant avec la fréquence. Ceci rejoint les travaux présentés dans [10]. Le modèle proposé prend donc en compte les effets des mouvements de fluide sur le remodelage osseux.

V. CONCLUSION

Cette étude propose un modèle numérique de remodelage osseux prenant en compte la fréquence de chargement à travers les mouvements de fluide. Les résultats numériques rejoignent ceux de plusieurs études expérimentales, rendant compte de l'importance de la fréquence de chargement dans le processus de remodelage osseux.

RÉFÉRENCES

- [1] J. WOLFF : *Das Gesetz der Transformation der Knochen*. Verlag von August Hirshwald, 1892. Translated by P. Maquet et R. Furlong, *The Law of Bone Remodelling*, Springer-Verlag, 1986.
- [2] G.S. BEAUPRÉ, T.E. ORR et D.R. CARTER : An approach for time-dependant bone modeling and remodeling – theoretical development. *Journal of Orthopaedic Research*, 8:651–661, 1990.
- [3] G.S. BEAUPRÉ, T.E. ORR et D.R. CARTER : An approach for

time-dependant bone modeling and remodeling – application :
A preliminary remodeling simulation. *Journal of Orthopaedic Research*, 8:662–670, 1990.

- [4] M.A. BIOT : Problème de la consolidation des matières argileuses sous une charge. *Annales de la Société Scientifique de Bruxelles*, B55:110–113, 1935.
- [5] M.A. BIOT : General theory of three-dimensional consolidation. *Journal of Applied Physics*, 12:155–164, 1941.
- [6] O. COUSSY : *Mechanics of porous continua*. John Wiley & Sons, 1995.
- [7] E. MALACHANNE, D. DUREISSEIX, P. CANADAS et F. JOURDAN : Experimental and numerical identification of cortical bone permeability. *Journal of Biomechanics*, 41:721–725, 2008.
- [8] R.B. MARTIN : Porosity and specific surface of bone. *CRC Critical Reviews in Biomedical Engineering*, 10:179–222, 1984.
- [9] Y-X QIN, T. KAPLAN, A. SALDANHA et C. RUBIN : Fluid pressure gradient, arising from oscillations in intramedullary pressure, is correlated with the formation of bone and inhibition of intracortical porosity. *Journal of Biomechanics*, 36:1427–1437, 2003.
- [10] S.J. WARDEN et C.H TURNER : Mechanotransduction in cortical bone is most efficient at loading frequencies of 5–10 hz. *Bone*, 34:261–270, 2004.
- [11] T. LEMAIRE, S. NAÏLI et A. RÉMOND : Multiscale analysis of the coupled effects governing the movement of interstitial fluid in cortical bone. *Biomechanics and Modeling in Mechanobiology*, 5:39–52, 2006.
- [12] V. KAFKA : On hydraulic strengthening of bones. *Journal of Biomechanics*, 26:761–762, 1993.
- [13] G. LI, J.T. BRONK, K-N. AN et P.J. KELLY : Permeability of cortical bone of canine tibiae. *Microvascular Research*, 34:302–310, 1987.

Une déduction géométrique de la cinématique des modèles classiques de plaque

Michele SERPILLI¹ (Doctorant 3e année),
 Françoise KRASUCKI², Giuseppe GEYMONAT³ (Directeurs de Thèse)

Laboratoire de Mécanique et Génie Civil
 UMR 5508 Université Montpellier II,
 Place Eugène Bataillon, 34695 Montpellier Cedex 5, France

¹serpilli@lmgc.univ-montp2.fr, ²krasucki@math.univ-montp2.fr,
³geymonat@lmgc.univ-montp2.fr

Résumé— Nous donnons une déduction purement géométrique des équations cinématiques de Kirchhoff-Love et de celles de Reissner-Mindlin dans le cas d'une plaque simplement connexe. Cette déduction est obtenue uniquement à partir des relations de compatibilité de Saint Venant et de la représentation intégrale de Cesàro-Volterra. Aucune information concernant la loi constitutive du matériau de la plaque ou le chargement n'est utilisée.

Mots-clés— Conditions de compatibilité, formule de Cesàro-Volterra, modèles de plaque.

I. INTRODUCTION

La justification des modèles classiques de plaque mince a été le sujet de nombreuses recherches dans les vingt dernières années, (P. G. Ciarlet [2] et sa bibliographie, et S. Müller [7]). Les hypothèses cinématiques de Kirchhoff-Love ont été obtenues par développement asymptotique des équations de l'élasticité tridimensionnelle en considérant les forces appliquées et la géométrie propre à la plaque. C'est seulement dernièrement que B. Miara and P. Podio-Guidugli [6], et R. Paroni et al. [8] ont donné une justification du modèle de plaque de Reissner-Mindlin à l'aide respectivement de la méthode de scaling formel et de la Γ -convergence. Dans ce travail, conçu dans le cadre de ma thèse de doctorat, on donne une déduction *purement géométrique* des cinématiques de Kirchhoff-Love et Reissner-Mindlin, à l'aide d'un développement formel appliqué aux équations de compatibilité de Saint-Venant et à la formule de Cesàro-Volterra.

II. LES ÉQUATIONS DE COMPATIBILITÉ ET LA FORMULE DE CESÀRO-VOLTERRA

En mécanique des milieux continus, une question fondamentale est d'établir si la répartition des déformations dans un continuum est quelconque ou si elle est soumise à certaines conditions d'admissibilité ou, mieux, de *compatibilité*. Autrement dit, si on se donne un champ de tenseurs symétriques $\mathbf{e} = (e_{ij})$, existe-t-il un champ de déplacement $\mathbf{v} = (v_i)$ dont il dérive? Ce tenseur donné $\mathbf{e} = (e_{ij})$ est-il le champ de déformation associé au déplacement inconnu $\mathbf{v} = (v_i)$? En général, la réponse est négative.

Au niveau géométrique, ceci s'interprète comme suit : si un objet est coupé en morceaux et si ces morceaux sont déformés de manière quelconque, ils ne peuvent pas être juxtaposés sans vides. Pour pouvoir le faire, et donc afin que les déformations soient *compatibles* avec la continuité du milieu, elles doivent satisfaire des conditions de compatibilité.

La première caractérisation d'un champ de matrice symétrique $\mathbf{e} = (e_{ij})$ s'écrivant sous la forme $\mathbf{e} = \nabla_s \mathbf{v}$, pour un champ de vecteur quelconque \mathbf{v} , a été annoncée par A.J.C.B. de Saint-Venant en 1864, et démontrée rigoureusement par E. Beltrami en 1881. Ce théorème est connu sous le nom de *Théorème de Saint-Venant* et il affirme :

Soit Ω un sous-ensemble ouvert, connexe et simplement connexe. Soit $\mathbf{e} = (e_{ij}) \in \mathcal{C}^2(\Omega; \mathbb{S}^3)$ un champ de matrice donné qui satisfait les *conditions de compatibilité de Saint-Venant* suivantes :

$$\mathcal{R}_{ijkl}(\mathbf{e}) := \partial_{\ell j} e_{ik} + \partial_{ki} e_{j\ell} - \partial_{\ell i} e_{jk} - \partial_{kj} e_{i\ell} = 0 \text{ dans } \Omega. \quad (1)$$

Alors, il existe un champ de déplacement $\mathbf{v} \in \mathcal{C}^3(\Omega; \mathbf{E}^3)$ tel que

$$e_{ij} = (\nabla_s \mathbf{v})_{ij} = \frac{1}{2}(\partial_j v_i + \partial_i v_j).$$

Par ailleurs, on considère le problème lié à un champ de déplacement $\mathbf{v} = (v_i)$ tel que, étant donné un champ de matrice symétrique $\mathbf{e} = (e_{ij})$, \mathbf{v} est la solution du système différentiel aux dérivées partielles $\nabla_s \mathbf{v} = \mathbf{e}$.

En 1906, E. Cesàro a proposé une formule intégrale explicite dans [1], en revisitant et adaptant une autre formule trouvée par V. Volterra [10], qui résout le système différentiel précédent. Étant donné un point fixe $\mathbf{x}_0 \in \Omega$ et un point $\mathbf{x} \neq \mathbf{x}_0 \in \Omega$, et un chemin γ de \mathbf{x}_0 à \mathbf{x} , la *formule intégrale de Cesàro-Volterra* est définie comme suit :

$$v_i(\mathbf{x}_0, \mathbf{x}) = \int_{\gamma} [e_{ij}(\mathbf{y}) + (x_k - y_k)(\partial_k e_{ij}(\mathbf{y}) - \partial_i e_{kj}(\mathbf{y}))] dy_j. \quad (2)$$

L'intégrale est indépendante du chemin si le domaine est simplement connexe.

III. DÉDUCTION GÉOMÉTRIQUE DE LA CINÉMATIQUE DE KIRCHHOFF-LOVE

Soit ε un petit paramètre qui tend vers zéro. Soit ω un domaine simplement connexe de \mathbb{R}^2 . On considère une plaque d'épaisseur $2\varepsilon > 0$ occupant la configuration de référence $\Omega^\varepsilon = \omega \times]-\varepsilon, \varepsilon[$. Soit le tenseur de déformation linéarisée $\mathbf{e}^\varepsilon(\mathbf{x}^\varepsilon) = (e_{ij}^\varepsilon)(\mathbf{x}^\varepsilon)$, défini sur le domaine variable Ω^ε et satisfaisant les conditions de compatibilité de Saint-Venant (1).

Si Ω^ε est simplement connexe, le système d'équations différentielles aux dérivées partielles, $\nabla_s^\varepsilon \mathbf{v}^\varepsilon(\mathbf{x}^\varepsilon) = \mathbf{e}^\varepsilon(\mathbf{x}^\varepsilon)$ admet comme solution explicite $\mathbf{v}^\varepsilon(\mathbf{x}^\varepsilon) = (v_i^\varepsilon)(\mathbf{x}^\varepsilon)$ donnée par la représentation de Cesàro-Volterra (2).

Afin d'étudier le comportement asymptotique des champs de déformation $\mathbf{e}^\varepsilon = (e_{ij}^\varepsilon)$ et de déplacement $\mathbf{v}^\varepsilon = (v_i^\varepsilon)$ lorsque l'épaisseur tend vers zéro, on introduit le changement de variable standard qui permet de transformer le problème défini sur un domaine ε -dépendant en un problème défini sur un domaine fixe. Soit $\Omega := \omega \times]-1, 1[$ et soit $\mathbf{x} = (x_i)$ un point générique de Ω ; à chaque point $\mathbf{x} \in \Omega$, on associe le point $\mathbf{x}^\varepsilon \in \Omega^\varepsilon$ à l'aide de la bijection (Fig. 1) :

$$\Pi^\varepsilon : \mathbf{x} = (x_i) \in \Omega \rightarrow \mathbf{x}^\varepsilon = (x_i^\varepsilon) = (x_1, x_2, \varepsilon x_3) \in \overline{\Omega^\varepsilon}. \quad (3)$$

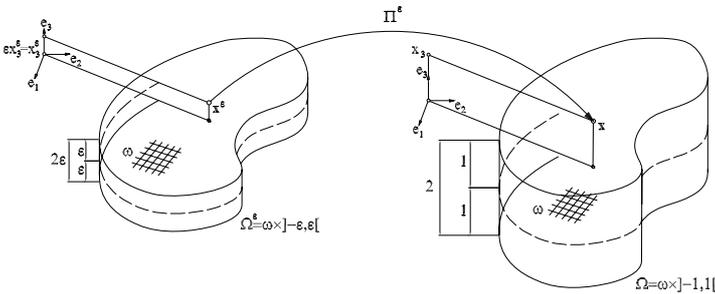


Fig. 1. Changement de variable

Pour cela,

$$\partial_\alpha^\varepsilon = \partial_\alpha \text{ et } \partial_3^\varepsilon = \frac{1}{\varepsilon} \partial_3$$

et $(e_{ij})(\varepsilon)(\mathbf{x}) = (e_{ij}^\varepsilon)(\mathbf{x}^\varepsilon)$ and $(v_i)(\varepsilon)(\mathbf{x}) = (v_i^\varepsilon)(\mathbf{x}^\varepsilon)$ sont respectivement les déformations et les déplacements transformés par Π^ε . Dans la suite, on omettra la dépendance explicite de ces champs par rapport à \mathbf{x} .

En appliquant le changement de variable à (1) et (2), on obtient les équations de compatibilité mises à l'échelle et la formule de Cesàro-Volterra mise à l'échelle définies dans le domaine fixe Ω .

Puisque dans les équations de compatibilité et dans la formule de Cesàro-Volterra mises à l'échelle le paramètre ε apparaît sous forme polynomiale, on suppose l'existence d'un développement asymptotique formel des déformations $e_{ij}(\varepsilon)$:

$$e_{ij}(\varepsilon) = e_{ij}^0 + \varepsilon e_{ij}^1 + \varepsilon^2 e_{ij}^2 + \varepsilon^3 e_{ij}^3 + \dots \quad (4)$$

Les termes d'ordre zéro e_{ij}^0 satisfont :

$$\begin{aligned} \partial_{33} e_{\alpha\beta}^0 &= 0, \\ \partial_{\sigma\alpha} e_{\beta\tau}^0 + \partial_{\beta\tau} e_{\sigma\alpha}^0 - \partial_{\tau\alpha} e_{\beta\sigma}^0 - \partial_{\beta\sigma} e_{\tau\alpha}^0 &= 0, \\ \partial_{3\tau} e_{\sigma\alpha}^0 - \partial_{3\sigma} e_{\tau\alpha}^0 &= 0. \end{aligned} \quad (5)$$

De l'équation (5)₁, on déduit la caractérisation des termes principaux $e_{\alpha\beta}^{KL} := e_{\alpha\beta}^0$:

$$e_{\alpha\beta}^{KL} = c_{\alpha\beta}(x_1, x_2) + x_3 r_{\alpha\beta}(x_1, x_2). \quad (6)$$

On remarque que le champ de déformation est décomposé en deux parties : $c_{\alpha\beta}$, le tenseur de déformation plane, qui caractérise les déformations membranaires ; $r_{\alpha\beta}$, le tenseur de courbure, qui caractérise les déformations flexionnelles.

Les conditions de compatibilité pour les tenseurs $c_{\alpha\beta}$ et $r_{\alpha\beta}$ prennent la forme suivante :

$$\begin{aligned} \partial_{\sigma\alpha} c_{\beta\tau} + \partial_{\beta\tau} c_{\sigma\alpha} - \partial_{\tau\alpha} c_{\beta\sigma} - \partial_{\beta\sigma} c_{\tau\alpha} &= 0, \\ \partial_\tau r_{\sigma\alpha} - \partial_\sigma r_{\tau\alpha} &= 0. \end{aligned}$$

Aucune condition sur les composantes restantes de déformation e_{i3}^{KL} n'a été trouvée. Elles seront obtenues à l'aide de la formule de Cesàro-Volterra mise à l'échelle.

L'hypothèse (4) sur le développement de $e_{ij}(\varepsilon)$ induit les développements asymptotiques formels de $v_i(\varepsilon)$:

$$\begin{aligned} v_\alpha(\varepsilon) &= v_\alpha^0 + \varepsilon v_\alpha^1 + \varepsilon^2 v_\alpha^2 + \varepsilon^3 v_\alpha^3 + \dots, \\ v_3(\varepsilon) &= \frac{1}{\varepsilon} v_3^{-1} + v_3^0 + \varepsilon v_3^1 + \varepsilon^2 v_3^2 + \varepsilon^3 v_3^3 + \dots, \end{aligned} \quad (7)$$

où

$$\begin{aligned} v_\alpha^0(\mathbf{x}) &= \int_{\tilde{\gamma}} [e_{\alpha\beta}^0 + (x_\sigma - y_\sigma)(\partial_\sigma e_{\alpha\beta}^0 - \partial_\alpha e_{\sigma\beta}^0) + x_3 \partial_3 e_{\alpha\beta}^0] dy_\beta, \\ v_3^{-1}(\mathbf{x}) &= - \int_{\tilde{\gamma}} (x_\sigma - y_\sigma) \partial_3 e_{\sigma\beta}^0 dy_\beta. \end{aligned} \quad (8)$$

Si on remplace (6) dans les formules (8), on obtient les formules de Cesàro-Volterra pour $v_\alpha^{KL} := v_\alpha^0$ et $v_3^{KL} := v_3^{-1}$

$$\begin{aligned} v_\alpha^{KL}(x_1, x_2, x_3) &= \int_{\tilde{\gamma}} [c_{\alpha\beta} + (x_\sigma - y_\sigma)(\partial_\sigma c_{\alpha\beta} - \partial_\alpha c_{\sigma\beta}) + \\ &+ x_3 \int_{\tilde{\gamma}} r_{\alpha\beta} dy_\beta, \\ v_3^{KL}(x_1, x_2) &= - \int_{\tilde{\gamma}} (x_\sigma - y_\sigma) r_{\sigma\beta} dy_\beta. \end{aligned} \quad (9)$$

Si on dérive les formules (8) par rapport à la coordonnée transversale x_3 , on peut montrer que certaines composantes de la déformation sont nulles :

$$e_{\alpha 3}^{KL} := \frac{1}{2}(\partial_\alpha u_3^{-1} + \partial_3 u_\alpha^0) = 0 \text{ and } e_{33}^{KL} := \partial_3 u_3^{-1} = 0.$$

Les différences principales entre e_{i3}^{KL} et e_{i3}^0 sont :

- e_{i3}^{KL} sont calculés à partir de la formule de Cesàro-Volterra et ils représentent les déformations associées aux déplacements v_α^{KL} et v_3^{KL} ;
- e_{i3}^0 sont les termes d'ordre zéro du développement asymptotique de $\mathbf{e}(\varepsilon)$.

Remarque : pour obtenir (9), on n'a fait aucune hypothèse sur la loi constitutive du matériau ou sur l'équilibre du système.

IV. DÉDUCTION GÉOMÉTRIQUE DE LA CINÉMATIQUE DE REISSNER-MINDLIN

Afin de retrouver la cinématique de Reissner-Mindlin, on a besoin d'imposer des liaisons internes au champ de matrice $e^\varepsilon = (e_{ij}^\varepsilon)$. L'utilisation des liaisons internes en élasticité est bien connu dans la littérature : B. Miara and P. Podio-Guidugli [6], P. Podio-Guidugli [9].

Notamment, on dit qu'un champ de matrice symétrique, noté par $e^{\varepsilon, RM} = (e_{ij}^{\varepsilon, RM})$, est un champ de matrice de Reissner-Mindlin si, pour tout $\mathbf{x}^\varepsilon = (x_i^\varepsilon) \in \Omega^\varepsilon$, il satisfait :

$$\begin{aligned} e_{\alpha\beta}^{\varepsilon, RM}(\mathbf{x}) &:= e_{\alpha\beta}^\varepsilon(\mathbf{x}), \\ e_{\alpha 3}^{\varepsilon, RM}(\mathbf{x}) &:= e_{\alpha 3}^\varepsilon(x_1, x_2), \\ e_{33}^{\varepsilon, RM}(\mathbf{x}) &= 0. \end{aligned} \quad (10)$$

Il faut remarquer que les deux dernières prescriptions sont formellement équivalentes à $\partial_{33}^\varepsilon v_\alpha^\varepsilon = 0$ et $\partial_3^\varepsilon v_3^\varepsilon = 0$, où $\mathbf{v}^\varepsilon = (v_i^\varepsilon)$ est un champ de déplacement.

Dans ce cas, les conditions de compatibilité deviennent :

$$\begin{aligned} \partial_{33}^\varepsilon e_{\alpha\beta}^\varepsilon &= 0, \\ \partial_{\sigma\alpha}^\varepsilon e_{\beta\tau}^\varepsilon + \partial_{\beta\tau}^\varepsilon e_{\sigma\alpha}^\varepsilon - \partial_{\tau\alpha}^\varepsilon e_{\beta\sigma}^\varepsilon - \partial_{\beta\sigma}^\varepsilon e_{\tau\alpha}^\varepsilon &= 0, \\ \partial_{3\tau}^\varepsilon e_{\sigma\alpha}^\varepsilon - \partial_{3\sigma}^\varepsilon e_{\tau\alpha}^\varepsilon &= \partial_{\tau\alpha}^\varepsilon e_{\sigma 3}^\varepsilon - \partial_{\sigma\alpha}^\varepsilon e_{\tau 3}^\varepsilon. \end{aligned} \quad (11)$$

La formule (11)₁ implique que $e_{\alpha\beta}^\varepsilon$ est une fonction linéaire par rapport à x_3 :

$$e_{\alpha\beta}^\varepsilon(x_1, x_2, x_3^\varepsilon) = c_{\alpha\beta}^\varepsilon(x_1, x_2) + x_3^\varepsilon r_{\alpha\beta}^\varepsilon(x_1, x_2). \quad (12)$$

Donc, en remplaçant (12) dans (11), on a :

$$\begin{aligned} \partial_{\sigma\alpha}^\varepsilon c_{\beta\tau}^\varepsilon + \partial_{\beta\tau}^\varepsilon c_{\sigma\alpha}^\varepsilon - \partial_{\tau\alpha}^\varepsilon c_{\beta\sigma}^\varepsilon - \partial_{\beta\sigma}^\varepsilon c_{\tau\alpha}^\varepsilon &= 0, \\ \partial_{\sigma\alpha}^\varepsilon r_{\beta\tau}^\varepsilon + \partial_{\beta\tau}^\varepsilon r_{\sigma\alpha}^\varepsilon - \partial_{\tau\alpha}^\varepsilon r_{\beta\sigma}^\varepsilon - \partial_{\beta\sigma}^\varepsilon r_{\tau\alpha}^\varepsilon &= 0, \\ \partial_{\tau}^\varepsilon r_{\sigma\alpha}^\varepsilon - \partial_{\sigma}^\varepsilon r_{\tau\alpha}^\varepsilon &= \partial_{\tau\alpha}^\varepsilon e_{\sigma 3}^\varepsilon - \partial_{\sigma\alpha}^\varepsilon e_{\tau 3}^\varepsilon. \end{aligned} \quad (13)$$

Maintenant, si on applique le changement de coordonnées standard Π^ε , $c_{\alpha\beta}^\varepsilon$ devient $c_{\alpha\beta}(\varepsilon)$, $r_{\alpha\beta}^\varepsilon$ devient $r_{\alpha\beta}(\varepsilon)$ et $e_{\alpha 3}^\varepsilon$ devient $e_{\alpha 3}(\varepsilon)$. On suppose ainsi que $c_{\alpha\beta}(\varepsilon)$, $r_{\alpha\beta}(\varepsilon)$ et $e_{\alpha 3}(\varepsilon)$ satisfont les développements suivants :

$$\begin{aligned} c_{\alpha\beta}(\varepsilon) &= c_{\alpha\beta}^0 + \varepsilon c_{\alpha\beta}^1 + \varepsilon^2 c_{\alpha\beta}^2 + \varepsilon^3 c_{\alpha\beta}^3 + \dots, \\ r_{\alpha\beta}(\varepsilon) &= r_{\alpha\beta}^0 + \varepsilon r_{\alpha\beta}^1 + \varepsilon^2 r_{\alpha\beta}^2 + \varepsilon^3 r_{\alpha\beta}^3 + \dots, \\ e_{\alpha 3}(\varepsilon) &= e_{\alpha 3}^0 + \varepsilon e_{\alpha 3}^1 + \varepsilon^2 e_{\alpha 3}^2 + \varepsilon^3 e_{\alpha 3}^3 + \dots, \end{aligned} \quad (14)$$

En remplaçant (14) dans (13), on obtient, pour $i = 0, 1, \dots$, les conditions de compatibilité :

$$\begin{aligned} \partial_{\sigma\alpha} c_{\beta\tau}^i + \partial_{\beta\tau} c_{\sigma\alpha}^i - \partial_{\tau\alpha} c_{\beta\sigma}^i - \partial_{\beta\sigma} c_{\tau\alpha}^i &= 0, \\ \partial_{\sigma\alpha} r_{\beta\tau}^i + \partial_{\beta\tau} r_{\sigma\alpha}^i - \partial_{\tau\alpha} r_{\beta\sigma}^i - \partial_{\beta\sigma} r_{\tau\alpha}^i &= 0, \\ \partial_{\tau} r_{\sigma\alpha}^i - \partial_{\sigma} r_{\tau\alpha}^i &= \partial_{\tau\alpha} e_{\sigma 3}^i - \partial_{\sigma\alpha} e_{\tau 3}^i. \end{aligned}$$

Comme pour la cinématique de Kirchhoff-Love, les renseignements manquants sont obtenus à l'aide de la formule de Cesàro-Volterra.

On considère la représentation de Cesàro-Volterra mise à l'échelle du champ de déplacement $\mathbf{v}^\varepsilon = (v_i^\varepsilon)$ associé au champ de déformation restreint par (10)-(12).

Les hypothèses (14) sur les développements asymptotiques induisent le développement asymptotique suivant pour les déplacements $v_i(\varepsilon)$:

$$v_i(\varepsilon) = v_i^0 + \varepsilon v_i^1 + \varepsilon^2 v_i^2 + \varepsilon^3 v_i^3 + \dots, \quad (15)$$

où

$$\begin{aligned} v_\alpha^0(\mathbf{x}) &= \int_{\widehat{\gamma}} [c_{\alpha\beta}^0 + (x_\sigma - y_\sigma)(\partial_\sigma c_{\alpha\beta}^0 - \partial_\alpha c_{\sigma\beta}^0)] dy_\beta, \\ v_\alpha^1(\mathbf{x}) &= \int_{\widehat{\gamma}} [c_{\alpha\beta}^1 + (x_\sigma - y_\sigma)(\partial_\sigma c_{\alpha\beta}^1 - \partial_\alpha c_{\sigma\beta}^1) + \\ &\quad + x_3(r_{\alpha\beta}^0 - \partial_\alpha e_{3\beta}^0)] dy_\beta + \int_{\gamma_*} e_{\alpha 3}^0 dy_3, \\ v_3^0(\mathbf{x}) &= \int_{\widehat{\gamma}} [e_{3\beta}^0 + (x_\sigma - y_\sigma)(\partial_\sigma e_{3\beta}^0 - r_{\sigma\beta}^0)] dy_\beta. \end{aligned} \quad (16)$$

Si on dérive v_α^0 par rapport à la coordonnée x_3 , on peut remarquer qu'il est indépendant de la coordonnée transversale, et, donc, il gouverne seulement les déformations membranaires. Donc, on définit :

$$v_\alpha^0(\mathbf{x}) = \int_{\widehat{\gamma}} [c_{\alpha\beta}^0 + (x_\sigma - y_\sigma)(\partial_\sigma c_{\alpha\beta}^0 - \partial_\alpha c_{\sigma\beta}^0)] dy_\beta := \bar{v}_\alpha^0(x_1, x_2).$$

On dérive (16)₃ par rapport à x_3 et on a que $\partial_3 v_3^0 = 0$. Donc :

$$v_3^0 := w(x_1, x_2).$$

Puisque la cinématique de Reissner-Mindlin tient compte du comportement membranaire, à flexion et à cisaillement, on est obligé de caractériser le deuxième terme du développement asymptotique v_α^1 :

$$\begin{aligned} v_\alpha^1(\mathbf{x}) &= \int_{\widehat{\gamma}} [c_{\alpha\beta}^1 + (x_\sigma - y_\sigma)(\partial_\sigma c_{\alpha\beta}^1 - \partial_\alpha c_{\sigma\beta}^1) + \\ &\quad + x_3(r_{\alpha\beta}^0 - \partial_\alpha e_{3\beta}^0)] dy_\beta + \int_{\gamma_*} e_{\alpha 3}^0 dy_3 = \\ &= \bar{v}_\alpha^1(x_1, x_2) + x_3(2e_{\alpha 3}^0 - \partial_\alpha w). \end{aligned}$$

On peut maintenant définir la cinématique complète de Reissner-Mindlin :

$$\begin{cases} u_{\alpha\beta}^{RM}(x_1, x_2, x_3) := \bar{u}_\alpha^0(x_1, x_2) + \bar{u}_\alpha^1(x_1, x_2) + x_3(2e_{\alpha 3}^0 - \partial_\alpha w), \\ u_3^{RM}(x_1, x_2) := w(x_1, x_2), \end{cases} \quad (17)$$

et par conséquent, le champ de déformation :

$$\begin{aligned} e_{\alpha\beta}^{RM} &:= c_{\alpha\beta}^0(x_1, x_2) + c_{\alpha\beta}^1(x_1, x_2) + x_3 r_{\alpha\beta}^0(x_1, x_2), \\ e_{\alpha 3}^{RM} &:= e_{\alpha 3}^0(x_1, x_2), \\ e_{33}^{RM} &= 0, \end{aligned}$$

où $c_{\alpha\beta}^i = \frac{1}{2}(\partial_\beta \bar{v}_\alpha^i + \partial_\alpha \bar{v}_\beta^i)$, pour $i = 0, 1$, et $r_{\alpha\beta}^0 = \frac{1}{2}(2\partial_\beta e_{\alpha 3}^0 + 2\partial_\alpha e_{\beta 3}^0 - \partial_{\alpha\beta} w)$.

La quantité $2e_{\alpha 3}^0 - \partial_\alpha w$ dans la fonction (17)₁ représente la rotation de la fibre transversale de la plaque. Cette fibre ne reste pas perpendiculaire au plan moyen lors de la déformation, comme dans le modèle de Kirchhoff-Love : c'est un modèle de plaque capable de se déformer à cisaillement. La fonction $e_{\alpha 3}^0$, qui apparaît dans la rotation du segment transversal, mesure la différence avec l'état de déformation de Kirchhoff-Love.

V. CONCLUSION

La dérivation de la cinématique des modèles classiques de plaque n'est basée que sur des considérations géométriques, puisque les conditions de compatibilité et la formule de Cesàro-Volterra sont profondément liées à la nature métrique de la déformation et de la variété sur laquelle

cette déformation est définie (article [5]). La représentation de Cesàro-Volterra semble importante pour :

- une formulation récente du problème d'élasticité tridimensionnelle, où les composantes de déformation sont considérées comme inconnues primaires (P. G. Ciarlet et P. Jr. Ciarlet [3])
- triangulation FEM.

La généralisation de la formule de Cesàro-Volterra dans le cas des coordonnées curvilignes et de la cinématique des coques et surfaces, [4] constitue deux autres parties de ma thèse.

RÉFÉRENCES

- [1] Cesàro, E. : Sulle formole del Volterra, fondamentali nella teoria delle distorsioni elastiche, Rend. Napoli Serie 3^a Vol. XII, Anno XLV, 311–321 (1906).
- [2] Ciarlet, P.G. : Mathematical Elasticity, vol. II : Theory of Plates, Stud. Math. Appl., North-Holland Amsterdam (1997).
- [3] Ciarlet, P.G., Ciarlet, P.Jr. : Another approach to linearized elasticity and a new proof of Korn's inequality, Math. Mod. Meth. Appl. Sci. **15**, 259–271 (2005).
- [4] Ciarlet, P.G., Serpilli, M. : Forthcoming (2007).
- [5] Geymonat, G., Krasucki, F., Serpilli, M. : The kinematics of plate models : a geometrical deduction, J. Elasticity, **88**, 299–309 (2007).
- [6] Miara B., Podio-Guidugli, P. : Deduction by scaling : a unified approach to classical plate and rod theories, Asymptotic Analysis **51**, 113–131 (2007).
- [7] Müller, S. : Paper presented at the course, Classical and Advanced Theories of Thin Structures : Mechanical and Mathematical Aspects, CISM, Udine, 5–9 June 2006.
- [8] Paroni, R., Podio-Guidugli, P., Tomassetti, G. : A justification of the Reissner-Mindlin plate theory through variational convergence, Analysis and Applications **5**, 165–182 (2007).
- [9] Podio-Guidugli, P. : An exact derivation of the thin plate equation, J. Elasticity **22**, 121–133 (1989).
- [10] Volterra, V. : Sur l'équilibre des corps élastiques multiplement connexes, Annales Scientifiques de l' E.N.S., 3^{ème} Série **24**, 401–517 (1907).

Segmentation thématique de texte : le sac de mots contre la phrase construite.

Alexandre LABADIÉ¹ (Doctorant 3e année), Violaine PRINCE² (Directrice de Thèse)

LIRMM

161 rue Ada,

34392 Montpellier Cedex 5 France

¹labadie@lirmm.fr, ²prince@lirmm.fr

Résumé— Cet article décrit succinctement le domaine de la segmentation thématique de texte en s'appuyant sur la présentation des deux principales visions qui s'affrontent dans le domaine du traitement automatique de la langue naturelle : Les statisticiens, adeptes des méthodes "sac de mots" et les linguistes qui s'appuient sur leur connaissance de la langue.

Mots-clés— Segmentation thématique de texte, TALN, sémantique, syntaxe, "sac de mots".

I. INTRODUCTION

Dans un monde où toujours plus d'information est disponible, disposer facilement de l'information précise que l'on recherche est de plus en plus difficile. En effet la masse d'information noie l'information pertinente dans un océan de données inutiles. Pour retrouver ce dont il a besoin, un utilisateur doit littéralement plonger dans cette mer hostile, et remonte souvent bredouille et à bout de souffle à la surface. C'est particulièrement le cas lorsque l'on effectue des recherches sur de grandes quantités de textes, comme on peut en trouver sur internet. Pour faire face à ces difficultés, des méthodes automatiques sont développées dans le cadre du traitement automatique de la langue naturelle (TALN), parmi elles les méthodes de "segmentation de texte".

Il existe beaucoup de tâches dites de "segmentation de texte". Par exemple, la recherche et l'extraction de textes dans des documents multimédia, où le texte est mélangé à de la vidéo et de l'image, sont des tâches assimilées à la "segmentation de texte" [10]. Regrouper des mots en morphèmes, ou en unités linguistiques plus importantes, est aussi nommé segmentation de texte (e.g. dans le traitement des langues asiatiques, qui utilisent des idéogrammes, les frontières des mots sont difficiles à déterminer [18], [19]). Dans cet article, nous nous intéressons à la "**segmentation thématique de texte**". C'est à dire à l'opération qui a pour but de trouver la structure thématique [6] d'un texte et d'en proposer une décomposition par thème [15]. Si la plupart des textes traitent d'un sujet unique, ils abordent en général plusieurs thèmes en leur sein. Plus le texte est volumineux, plus il contiendra de thèmes. Ainsi, fondamentalement, la segmentation thématique de texte recherche, au sein d'un texte, le début et la fin des thèmes. Pour des raisons pratiques, nous utiliserons pour le reste de cet article le terme "segmentation thématique" plutôt que segmentation thématique de texte.

Si l'on considère que la segmentation thématique doit diviser en plusieurs segments cohérents et distinct sur le plan thématique, alors chaque segment ne doit, à l'idéal, traiter que d'un seul thème. Mais un thème est une unité complexe sur le plan rhétorique, qui nécessite souvent des digressions, des exemples et des argumentations.

Ce qui nous amène à nous poser la question de la définition de la notion de thème. Dans la littérature, nous trouvons plusieurs définitions de ce qu'est un thème. En général, un thème est : *le sujet d'une conversation ou d'une discussion*. En linguistique, on le définit comme : *l'élément d'un énoncé qui est réputé connu par les participants à la communication* (on l'oppose souvent au rhème qui est l'information nouvelle apportée par l'énoncé). Nous admettons ici que le thème d'un segment de texte est : *ce dont il parle*. La segmentation thématique doit donc diviser le texte en segments de textes dont chacune des phrases "parlent" de la même chose. Ce découpage peut ensuite être utilisé pour faciliter la recherche d'informations précises au sein de texte volumineux.

Deux visions majeures s'affrontent dans le domaine de la segmentation thématique (et plus généralement en TALN) : ceux qui considèrent le texte comme une entité statistique et donc l'étudient de ce point de vue, et ceux qui voient le texte comme une structure complexe issue d'un processus humain : la langue. Dans une première partie nous présenterons brièvement les principales approches statistiques en segmentation thématique. Dans une deuxième partie nous présenterons notre approche du domaine, basée sur une étude de la syntaxe, de la sémantique et de la structure du texte.

II. LES APPROCHES "SAC DE MOTS"

Les méthodes de segmentation thématique de type "sac de mots" se basent principalement sur la notion de cohésion lexicale¹, observée au travers de la répétition de termes. On peut regrouper ces méthodes en trois grandes familles que nous allons présenter ici.

A. Segmentation à partir de mesure de similarité entre segments de texte

Les méthodes de segmentation à base de similarité considèrent les différentes portions de texte du document à

¹tel que la décrivent [14]

traiter comme autant de vecteurs. Les composantes des vecteurs étant, dans la plupart des cas, les fréquences d'apparition des mots au sein de la portion de texte, après que celle-ci a été débarrassée des mots inutiles (mots jugés comme peu porteurs de sens). Parfois, cette fréquence des mots est pondérée par un IDF (Inverse Document Frequency), pour renforcer l'importance des mots supposés thématiquement saillants.

L'objectif de ces méthodes est donc de mesurer la proximité ou l'éloignement des portions de texte étudiées grâce à l'angle que forment leurs vecteurs représentatifs. Elles s'appuient donc en général sur le cosinus de cet angle, qu'elles considèrent comme la similarité. La similarité est ensuite exploitée de diverses manières. Choi [4], par exemple, utilise la similarité pour effectuer un classement local des phrases les unes par rapport aux autres.

B. Segmentation à partir de représentation graphique de répétition de termes

En passant par une représentation graphique des termes, il est plus facile de visualiser leur répartition le long du document étudié. Ainsi la méthode du nuage de points, présentée par Helfman [7] emploie cette représentation pour la recherche d'information. Le principe est de positionner sur un graphique chaque occurrence des termes du document. Ainsi, un terme apparaissant à une position i et une position j du texte, sera représenté par les 4 couples (i, i) , (i, j) , (j, i) et (j, j) . Les portions du document où les répétitions de termes sont nombreuses apparaîtront alors sur le graphique comme les zones de forte concentration de points.

Cette approche visuelle de la représentation d'un texte a été reprise et adaptée à la segmentation thématique par Reynar [16] dans son algorithme DotPlotting. L'idée est d'identifier les segments thématiquement cohérents sur le graphique en cherchant les limites des zones les plus denses. La densité d'une région du graphique est calculée en divisant le nombre de points présents dans la région par l'aire de cette dernière. L'objectif de DotPlotting est d'isoler les segments thématiques soit en maximisant leur densité, soit en minimisant la taille des zones "vides" entre les segments. On notera que, dans son principe, cette méthode est très proche de l'algorithme c99 de Choi [4].

Cette approche a même inspiré des méthodes originales, comme celle proposée par Ji et Zha [8], qui consiste à remplacer le problème de segmentation thématique par un problème de segmentation d'image. Cette méthode utilise une technique de diffusion anisotropique sur la représentation graphique de la matrice de distance afin de renforcer les contrastes entre les zones denses et les frontières.

C. Segmentation à partir de chaînes lexicales

La segmentation à base de chaînes lexicales relie les occurrences multiples des mots dans un document et estime qu'une chaîne est rompue si la distance entre deux occurrences du même mot est trop importante. Cette distance est généralement exprimée en nombre de phrases. Ainsi, la méthode *Segmenter* présentée par Kan [9],

procède selon ce principe pour effectuer une segmentation thématique du document étudié. On notera tout de même une subtilité. La distance à partir de laquelle l'algorithme considère qu'il y a rupture dépend de la catégorie syntaxique du mot impliqué dans la chaîne.

Une autre approche basée sur les chaînes lexicales est proposée par Hearst [5] avec son algorithme *Text Tilling*. Un score de cohésion est attribué à chacun des blocs de texte en fonction du bloc qui le suit. Il est quant à lui calculé sur la base d'un premier score dit "lexical" attribué à chaque paire de phrases en fonction de la paire de phrases qui la suit. Ce score lexical est lui même calculé à partir des paramètres que sont le nombre de mots en commun, de mots nouveaux et de chaînes lexicales actives dans les phrases considérées. Le score de chaque segment de texte est alors le produit scalaire normalisé des scores de chacune des paires de phrases qu'il contient. Si un segment présente un score très différent des segments précédents et suivants, alors la rupture thématique se situe au sein de ce segment.

D. Limites du "sac de mots"

Toutes ces approches ont ceci en commun qu'elles se basent sur la cohésion lexicale supposée des segments thématiques. Or il est tout à fait possible que deux portions d'un texte aient peu de mots en commun (et donc une faible cohésion lexicale) tout en véhiculant le même contenu informationnel. Même s'il y a eu des tentatives d'intégrer une information de type sémantique, grâce notamment à l'adjonction d'une analyse sémantique latente à certaines méthodes suscitées ([4]), la base de l'approche reste très "sac de mot".

Or un texte est composé d'unités syntaxiques, qui sont également sémantiques, et dont la granularité est supérieure au mot : les phrases.

III. SYNTAXE, SÉMANTIQUE ET STRUCTURE

Nous avons développé une méthode de segmentation thématique, appelée Transeg, qui se base sur une représentation vectoriel du texte et une vision structurée de ce dernier.

.1 Représentation du texte

La première étape de notre approche est de convertir chaque phrase du texte en un vecteur sémantique. Ce vecteur est obtenu grâce à l'analyseur morpho-syntaxique de la langue française SYGFRAN [1]. Ces vecteurs sont des vecteurs sémantiques à la Roget [17], mais se basant sur le thésaurus Larousse [12] comme référence. Le vecteur de chaque phrase est calculé de manière récursive en combinant linéairement les vecteurs des constituants de la phrase, eux même obtenus par combinaison linéaire des vecteurs de mots. Le poids de chaque constituant dépend du résultat d'une analyse morpho-syntaxique en constituant et en dépendance².

.2 Segmentation du texte

En nous appuyant sur cette représentation de la phrase, nous avons cherché à identifier ce que nous nommons les

²La formule est donnée par [2]

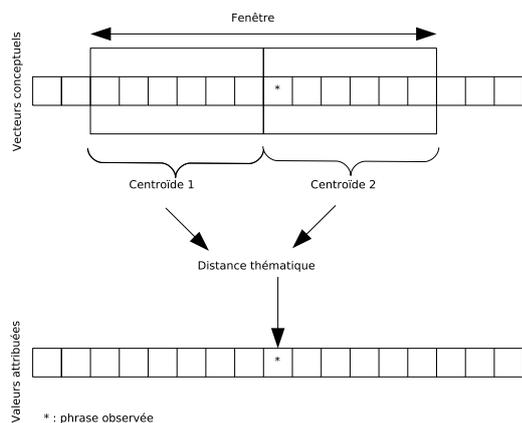


Fig. 1. Attribution d'un score de transition

“zones de transition” à l’intérieur du texte. La notion de zone de transition vient de l’observation que nous avons faite que la frontière entre deux thèmes au sein d’un texte est rarement une phrase unique, mais probablement une courte succession de phrases (2 ou 3). Pour retrouver ces zones de transition nous faisons glisser une fenêtre le long du texte. Chaque moitié de la fenêtre est considérée comme un segment thématique potentiel. On calcule alors un centroïde pour chacun de ces deux segments potentiels. Ce centroïde est un barycentre, ce qui nous permet d’incorporer un peu d’information stylistique lors de la pondération. En effet, les premières phrases et les introductions comportent très souvent plus d’informations pertinentes que les autres phrases ([11],[13]). Les poids de chaque phrase sont donc calculés selon une régression linéaire donnant plus d’importance aux premières phrases d’un segment comparativement aux dernières. Finalement, nous calculons une distance (que nous nommons distance thématique) entre ces deux barycentres. Cette distance est attribuée comme score de transition à la phrase du milieu de la fenêtre (la frontière potentielle donc, voir figure 1).

Les zones de transition sont donc des phrases successives avec un score de transition supérieur à un seuil déterminé. Dans notre première implémentation de la méthode, nous utilisons la distance angulaire pour le calcul du score de transition. Maintenant, nous utilisons une version modifiée de la distance de concordance proposée par Chauché [3]. Nous nous abstenons de la décrire ici pour d’évidentes raisons de place et un souci de vulgarisation. Toutefois, on peut préciser que cette distance est plus discriminante que la distance angulaire lorsqu’il s’agit de traiter des vecteurs n’ayant que peu de composantes activées (ce qui est le cas dans notre approche).

IV. CONCLUSION

La querelle opposant les statisticiens et les linguistes dans le domaine de la segmentation thématique est assez symptomatique du TALN en général. L’opposition entre les partisans du “tout chiffre” et les partisans du “tout langue” se retrouve dans bien d’autres domaines du TALN (avec un avantage pour les statisticiens à l’heure actuelle). Il est alors assez ironique de constater que ces deux approches,

bien qu’en apparence antagonistes, sont complémentaires. L’avenir de la segmentation thématique passera certainement par des approches hybrides.

RÉFÉRENCES

- [1] J. Chauché. Un outil multidimensionnel de l’analyse du discours. *Proceedings of Coling’84*, 1 :11–15, 1984.
- [2] J. Chauché and V. Prince. Classifying texts through natural language parsing and semantic filtering. In *Proceedings of LTC’03*, 2007.
- [3] J. Chauché, V. Prince, S. Jaillet, and M. Teisseire. Classification automatique de textes à partir de leur analyse syntaxico-sémantique. *Proceedings of TALN’03*, pages 55–65, 2003.
- [4] F. Y. Y. Choi. Advances in domain independent linear text segmentation. *Proceedings of NAACL-00*, pages 26–33, 2000.
- [5] M. A. Hearst. Text-tilling : segmenting text into multi-paragraph subtopic passages. *Computational Linguistics*, pages 59–66, 1997.
- [6] M. A. Hearst and C. Plaunt. Subtopic structuring for full-length document access. *Proceedings of the ACM SIGIR-93 International Conference On Research and Development in Information Retrieval*, pages 59–68, 1993.
- [7] J. Helfman. Similarity patterns in language. *Visual Languages*, pages 173–175, 1994.
- [8] X. Ji and H. Zha. Domain-independent segmentation using anisotropic diffusion and dynamic programming. *Proceedings of ACM/SIGIR Conference of Research and Development in Information Retrieval*, 2003.
- [9] M. Kan, J. L. Klavans, and K. R. McKeown. Linear segmentation and segment significance. *Proceedings of WVLC-6*, pages 197–205, 1998.
- [10] D. Karatzas. *Text Segmentation in Web Images Using Color Perception and Topological Features*. ECS Publications, UK, 2003.
- [11] A. Labadié and Chauché. Segmentation thématique par calcul de distance sémantique. *Proceedings of DEFT’06*, 1 :45–59, 2006.
- [12] Larousse. *Thésaurus Larousse - des idées aux mots, des mots aux idées*. Larousse, Paris, 1992.
- [13] A. Lelu, C. M., and S. Aubain. Coopération multiniveau d’approches non-supervisées et supervisées pour la détection des ruptures thématiques dans les discours présidentiels français. In *Proceedings of DEFT’06*, 2006.
- [14] J. Morris and G. Hirst. Lexical cohesion computed by thesaural relations as an indicator of the structure of text. *Computational Linguistics*, 17 :20–48, 1991.
- [15] J. M. Ponte and W. B. Croft. Text segmentation by topic. *European Conference on Digital Libraries*, pages 113–125, 1997.
- [16] J. C. Reynar. *Topic Segmentation : Algorithms and Applications*. Phd thesis, University of Pennsylvania, 1998.
- [17] P. Roget. *Thesaurus of English Words and Phrases*. Longman, London, 1852.
- [18] Z. Wu and G. Tseng. Chinese text segmentation for text retrieval : Achievements and problems. *Journal of the American Society for Information Science*, 44 :532–542, 1993.
- [19] C. C. Yang and K. W. Li. A heuristic method based on a statistical approach for chinese text segmentation. *Journal of the American Society for Information Science and Technology*, 56 :1438–1447, 2005.

Mécanisme de suggestion de ressources touristiques en graphes conceptuels

Nicolas Moreau, Michel Leclère, Marie-Laure Mugnier

LIRMM, Univ. Montpellier 2, CNRS
161, rue Ada
34392 Montpellier, France

{moreau,leclere,mugnier}@lirmm.fr

Résumé— Nous proposons un système de suggestion de ressources touristiques basé sur le formalisme des graphes conceptuels. Ce système doit permettre de proposer des types de ressources à suggérer en fonction de ressources sélectionnées par l'utilisateur, tout en respectant une politique marketing et commerciale d'une région. Pour ce faire, nous introduisons une notion de « règle marketing » permettant de concilier ces deux exigences, et définissons un mécanisme de suggestion basé sur des opérations de graphes.

Mots-clés— Graphes conceptuels, suggestion, e-tourisme, système d'interrogation.

I. INTRODUCTION

Le système de suggestion est réalisé dans le cadre du projet Eiffel [1]. Ce projet a pour objectif principal la conception d'une plateforme pour la mise en œuvre de moteurs de recherche spécialisés dans le tourisme, pour offrir des informations riches et précises aux utilisateurs tout en permettant aux territoires de valoriser leur offre [2].

La valorisation de l'offre touristique passe par la proposition de nouvelles ressources (restaurant, hôtel, organisme, etc.) à l'utilisateur, basées sur ses critères de recherche. Ces ressources permettent de construire une offre touristique complète, alliant par exemple un hébergement, des lieux de restauration, des activités ou événements à proximité, etc. Les suggestions sont calculées à partir de règles marketing, qui permettent de traduire la politique commerciale et marketing de la région et sont construites manuellement par du personnel de la région.

Le mécanisme de suggestion doit prendre en compte des connaissances ontologiques (*i.e.* basées sur un ensemble de concepts hiérarchisés par une relation "est-un"). Ces connaissances peuvent par exemple être d'ordre géographique, comme l'ensemble des lieux-dits, villes et départements composant la région. Il doit également permettre de générer de nouvelles connaissances, par un mécanisme de raisonnement, pour enrichir les connaissances existantes. Enfin, il doit autoriser une construction expressive et intuitive des règles marketing, qui évolueront en fonction de la politique de la région, et qui seront construites par des personnes non spécialistes de la représentation de connaissance. Le formalisme des graphes conceptuels répond à ces problématiques. Les graphes conceptuels offrent une représentation de connaissances visuelle intuitive sous forme de graphes, pour la construction des règles marketing. De plus, les graphes conceptuels disposent de mécanismes de raisonnement logiquement fondés, ce qui permet de garantir les

résultats de notre système de suggestion. Enfin, nous disposons d'une librairie de raisonnement permettant de faire des opérations de graphes, CoGITaNT[3], et d'un logiciel de création graphique de graphes conceptuels CoGUI[4] que nous avons développé.

L'article est organisé comme suit : Dans la section II nous introduisons le formalisme des graphes conceptuels utilisés dans le système de suggestion. Dans la section III nous introduisons les différents éléments et mécanismes du système de suggestion dans le cadre d'un exemple d'utilisation.

II. FORMALISATION PROPOSÉE

Notre proposition s'appuie sur le formalisme de graphes conceptuels que nous développons depuis 15 ans au LIRMM [5] et qui se distingue de la proposition initiale de [6] par le parti pris de conserver un formalisme à base de graphes que ce soit au niveau des représentations ou au niveau du raisonnement. Nous allons tout d'abord introduire les notions fondamentales de notre formalisme, que sont les graphes conceptuels simples, construits sur un support, et comparables par une opération de projection.

A. Les graphes conceptuels simples

Un *graphe conceptuel simple* (SG) est un graphe bipartite étiqueté : l'une des classes de sommets, dite de sommets concepts, représente des entités, et l'autre, dite de sommets relations, représente les relations entre ces entités, ou des propriétés de ces entités. Un sommet concept d'un SG est étiqueté par un couple $t : m$ où t est un type de concept et m est un marqueur. Si le sommet représente une entité précise alors m est un identifiant représentant cette entité (le sommet est dit *individuel*), sinon m est noté $*$ (le sommet est dit *générique*). Un sommet relation est étiqueté par une

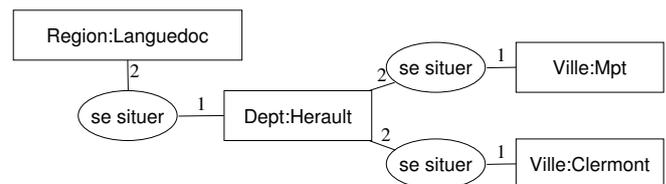


Fig. 1. Un graphe conceptuel.

relation r et, si n est l'arité de r (son nombre d'arguments), n arêtes sont incidentes à ce sommet ; ces arêtes sont totalement ordonnées. De façon classique, dans les dessins, les

sommets concepts sont représentés par des rectangles et les sommets relations par des ovales, et l'ordre sur les arêtes incidentes à un sommet relation n -aire par une numérotation de ces arêtes de 1 à n . On note $G = (C_G, R_G, E_G, l_G)$ un SG où C_G est l'ensemble des sommets concepts, R_G l'ensemble des sommets relations, E_G l'ensemble des arêtes et l_G la fonction d'étiquetage des sommets et des arêtes.

Les étiquettes des sommets sont prises dans un vocabulaire appelé *support* (ou *ontologie*) qui peut être plus ou moins riche. Nous considérerons ici un support (i.e. une ontologie simple) comme une structure $O = (T_C, <_C, T_R, <_R, \sigma)$ où :

- $(T_C, <_C)$ et $(T_R, <_R)$ sont des ensembles ordonnés par spécialisation de types de concept et de relations d'arité quelconque respectivement ($t' \leq t$ s'interprète par " t' est une spécialisation de t ");
- σ associe à chaque relation n -aire, sa signature, un n -uplet de types de concept qui définit le type maximal de chacun de ses arguments (avec respect de la covariance).

Dans [7] des types conjonctifs sont introduits, ils permettent de spécifier comme type d'un concept une conjonction de types de concepts appartenant à T_C . Dans la suite de l'article, nous utiliserons des conjonctions de types, mais pour des raisons de clarté, ces conjonctions de types seront représentées par une bicoloration (les concepts grisés ont un type conjonctif composé du type représenté et d'un type réservé dans l'ontologie, voir fig. 2).

La notion fondamentale pour comparer des SG est une application d'un SG dans un autre appelée *projection* (un homomorphisme de graphes étiquetés). Intuitivement, l'existence d'une projection de G dans H montre que la connaissance représentée dans G est contenue dans H . Une projection π de G dans H est plus précisément une application de C_G dans C_H et de R_G dans R_H qui conserve les arêtes (si on a une arête entre r et c étiquetée i dans G alors on a une arête entre $\pi(r)$ et $\pi(c)$ étiquetée i dans H) et peut "spécialiser" les étiquettes des sommets. Le résultat fondamental sur lequel le formalisme SG s'appuie est l'équivalence entre l'existence d'une projection et la déduction logique sur les formules associées par une transformation ϕ en logique des prédicats aux SG (cf. [6] et [5]).

Des connaissances implicites peuvent être ajoutées par l'intermédiaire de règles d'inférence[8]. Une règle exprime une connaissance de la forme "si hypothèse alors conclusion", où hypothèse et conclusion sont deux SG. La figure 3 présente un exemple de règle. Les pointillés lient certains sommets de l'hypothèse et de la conclusion; ces sommets sont appelés sommets *frontières*. Une règle R s'applique à un SG G s'il existe une projection de l'hypothèse de R dans G . L'application de R à G selon une telle projection π consiste à "attacher" à G la conclusion de R , en fusionnant chaque sommet frontière de la conclusion avec l'image par π du sommet frontière lui correspondant dans l'hypothèse.

III. SCÉNARIO D'UTILISATION

Nous présentons le mécanisme de suggestion par un exemple simple qui fait intervenir tous les éléments du système.

A. Sélection d'une ressource par l'utilisateur

La ressource touristique sélectionnée par l'utilisateur est fournie au système de suggestion sous la forme d'un SG bicoloré. Dans ce SG, un seul concept représente la ressource (qui est celle sélectionnée). Ce concept est rendu générique : on a transformé une ressource en "type de ressource". La bicoloration distingue les caractéristiques choisies par l'utilisateur (informations entrées dans le formulaire de recherche – en gris) des autres caractéristiques connues sur la ressource. Le SG de la figure 2 signifie : « *L'utilisateur a choisi un hôtel car il cible les jeunes célibataires et qu'il est situé dans la ville de Clermont, de plus on sait que cet hôtel interdit les animaux et est proche du lac Salagou* ».

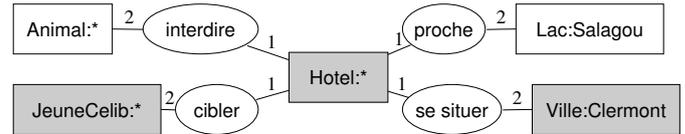


Fig. 2. Un type de ressource sélectionné.

B. Enrichissement de la ressource

La ressource sélectionnée par l'utilisateur va être enrichie des connaissances contenues dans la *base de connaissance*, qui est un SG. Ces connaissances concernent la région : patrimoine culturel, cépages, AOC, villes, proximités etc. (voir figure 1). Cet enrichissement est réalisé par une union des deux graphes : les concepts ayant le même marqueurs individuel (ici le sommet représentant la ville de Clermont) seront fusionnés.

De la connaissance supplémentaire est inférée sur le graphe obtenu, grâce aux *règles d'inférence*. La règle d'inférence de la figure 3 porte sur des informations géographiques de la base de connaissance, et peut se lire comme « *si une ville est dans un département, lui-même dans une région, alors cette ville est dans cette région* ».

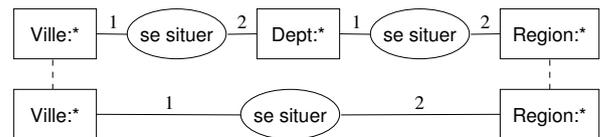


Fig. 3. Une règle d'inférence.

Le résultat de l'enrichissement de la ressource de la figure 2 par la base de connaissance et la règle d'inférence est représenté sur la figure 4 (sur la figure nous n'avons gardé que les sommets qui nous intéressent).

C. Sélection des règles marketing pertinentes

Les *règles marketing* expriment le type de ressource à suggérer en fonction du type de ressource envoyé au système. Une règle marketing exprime une connaissance qui peut être traduite en langue naturelle par une phrase de la forme : « *si un type de ressource a est choisi par l'utilisateur, et que l'on sait également b sur ce type de ressource, alors proposer le type de ressource c* ». Une règle marketing est représentée de la même manière qu'une règle d'inférence (à l'exception près que son hypothèse est bicolorée).

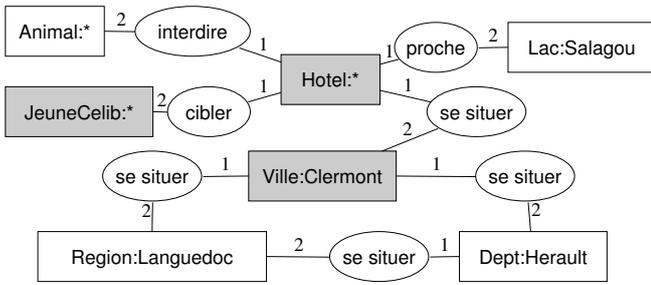


Fig. 4. La ressource sélectionnée enrichie.

La règle marketing de la figure 5, signifie « Si l'utilisateur a choisi une ressource parce qu'elle cible les célibataires et parce qu'elle est dans une certaine ville, et que l'on sait que cette ressource est proche d'un lieu aquatique, alors suggérer les ressources qui se situent dans la même région et qui offrent des activités de pêche ».

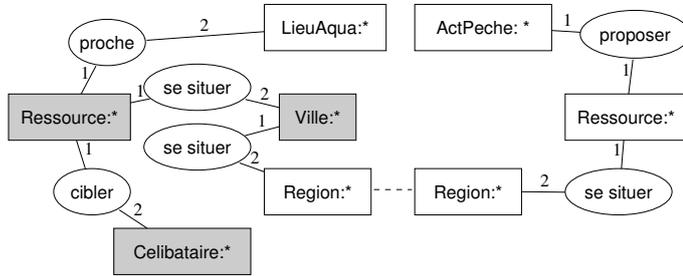


Fig. 5. Une règle marketing.

On sélectionne l'ensemble des règles marketing pertinentes pour le type de ressource sélectionné par l'utilisateur. La pertinence d'une règle marketing est définie par l'existence d'une projection¹ de son hypothèse dans la ressource enrichie (la ressource enrichie est alors considérée comme un cas particulier de l'hypothèse). La règle marketing de la fig. 5 est pertinente pour le type de ressource de la figure 4².

D. Génération des suggestions

Les types de ressources à suggérer sont calculés à partir des règles marketing pertinentes. Ces types de ressources à suggérer correspondent aux deuxièmes parties des règles marketing (dans notre exemple, « Une ressource qui propose une activité de pêche et qui se situe dans [une] région »). Ce « patron de suggestion » est adapté à chaque cas particulier que représente la ressource sélectionnée enrichie. Ces connaissances particulières sont reliées au patron de suggestion par les liens entre l'hypothèse de la règle marketing et son patron (dans notre exemple, sur le concept représentant la région). Comme la règle est pertinente, on sait que la région de la ressource enrichie est un cas particulier (une spécialisation en GC) de la région présente dans l'hypothèse de la règle (et dans le patron). Ainsi, la région du patron de suggestion sera celle de la ressource enrichie, comme le montre la figure 6. Ces liens permettent d'adapter la suggestion à la ressource qui intéresse l'utili-

¹respectant la bicoloration

²Étant donné que *JeuneCelib* < *Celibataire* et *Lac* < *LieuAqua*.

sateur (« Suggérer une ressource ... dans le même ... que la ressource choisie »).

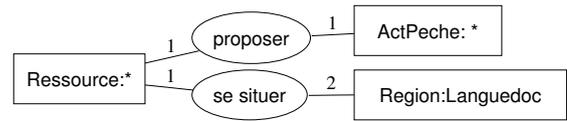


Fig. 6. Un type de ressource à suggérer.

Ces types de suggestions sont ensuite utilisés comme requête dans une autre base (en RDF) pour retourner des ressources à suggérer.

IV. CONCLUSION

Dans le cadre du projet Eiffel, nous avons défini un système de suggestion de ressources touristiques lors du choix d'une ressource par un utilisateur, dont les données et les mécanismes sont basés sur les graphes conceptuels. Ce système doit permettre de valoriser l'offre touristique d'une région, en cohérence avec sa politique marketing de la région. Le système de suggestion se base sur des règles marketing de la forme « si un type de ressource a est choisi par l'utilisateur, et que l'on sait également b sur ce type de ressource, alors proposer le type de ressource c ». Ces règles marketing suivent la politique commerciale de la région puisqu'écrites par le personnel de cette région. De plus les ressources suggérer seront pertinentes pour l'utilisateur puisque les règles marketing permettent de lier les types de ressources suggérées aux ressources sélectionnées (par exemple, le type de ressource suggéré doit être proche, ou près du même lac que la ressource sélectionnée).

Les règles marketing et le mécanisme de construction des types de ressources à suggérer sont à rapprocher des requêtes de type "construction" de SPARQL[9]. En effet une règle marketing et la construction d'un type de ressource à suggérer en fonction de cette règle et de la ressource sélectionnée par l'utilisateur peut être vue comme une règle de construction. La réalisation technique du système de suggestion dans le cadre du projet Eiffel permettra donc l'ajout d'une fonctionnalité de haut niveau dans un serveur de requêtes en graphes conceptuels.

RÉFÉRENCES

- [1] EIFFEL : Site du projet eiffel. <http://www.projet-eiffel.org/>.
- [2] L. NOËL, O. CARLONI, N. MOREAU et S. WEISER : Designing a knowledge-based tourism information system. *Int. J. of Digital Culture and Electronic Tourism, Special Issue on National Tourism Organisations and Exploitation of Information Technologies (à paraître)*, 2008.
- [3] COGITANT : Conceptual graphs integrated tools allowing nested typed graphs. <http://cogitant.sourceforge.net>.
- [4] COGUI : Cogui, outils de création graphique de graphes conceptuels. <http://www.lirmm.fr/cogui>.
- [5] M. CHEIN et M.-L. MUGNIER : Conceptual Graphs : Fundamental Notions. *Revue d'Intelligence Artificielle*, 6(4):365-406, 1992.
- [6] J. F. SOWA : *Conceptual Structures : Information Processing in Mind and Machine*. Addison-Wesley, 1984.
- [7] M. CHEIN et M.-L. MUGNIER : Concept Types and Coreference in Simple Conceptual Graphs. *In Proc. ICCS'04*, LNAI. Springer, 2004.
- [8] Jean-François BAGET et Marie-Laure MUGNIER : Extensions of simple conceptual graphs : the complexity of rules and constraints. *J. Artif. Intell. Res. (JAIR)*, 16:425-465, 2002.
- [9] Eric PRUD'HOMMEAUX et Andy SEABORNE : SPARQL Query Language for RDF. Rapport technique, W3C, 2006.

Echantillonnage pour l'extraction de motifs séquentiels

Chedy RAÏSSI

LIRMM, Université Montpellier 2
161, Rue Ada, 34392 Montpellier, France

chedy.raïssi@lirmm.fr

Résumé— Depuis quelques années, la communauté fouille de données s'est intéressée à la problématique de l'extraction de motifs séquentiels à partir de grandes bases de données en considérant comme hypothèse que les données pouvaient être chargées en mémoire centrale. Cependant, cette hypothèse est mise en défaut lorsque les bases manipulées sont trop volumineuses. Dans cet article, nous étudions une technique d'échantillonnage basée sur des réservoirs et montrons comment cette dernière est particulièrement bien adaptée pour résumer de gros volumes de données. Nous nous intéressons ensuite à la problématique plus récente de la fouille sur des données disponibles sous la forme d'un flot continu et éventuellement infini ("data stream"). Nous étendons l'approche d'échantillonnage à ce nouveau contexte et montrons que nous sommes à même d'extraire des motifs séquentiels de flots tout en garantissant les taux d'erreurs sur les résultats.

Mots-clés— Base de données, extraction de connaissances, motifs séquentiels, échantillonnage.

I. INTRODUCTION

La problématique de l'extraction de motifs séquentiels dans de grandes bases de données intéresse la communauté fouille de données depuis une dizaine d'années et différentes méthodes ont été développées pour extraire des séquences fréquentes. Le développement des nouvelles technologies permet aujourd'hui de générer de très grands volumes de données issues de différentes sources : trafic TCP/IP, transactions financières, enregistrements médicaux, capteurs. Les données apparaissent alors sous la forme d'un flot (*data stream*) de manière continue, à un rythme rapide et éventuellement de manière infinie. L'extraction de connaissances à partir de tels flots a récemment donné lieu à de nombreux travaux de recherche qui se sont focalisés sur la découverte d'itemsets fréquents (e.g., [4], [6], [8]) en utilisant des méthodes telles que le *landmark*, la *fenêtre glissante* ou les modèles de *pondérations temporelles* et peu de travaux se sont intéressés à l'extraction de motifs séquentiels dans les flots de données [9]. Il est vrai qu'outre l'espace de recherche, les approches traditionnelles nécessitent de faire plusieurs passes sur la base ou de stocker cette dernière en mémoire.

Le reste de l'article est organisé de la manière suivante. Dans la section II, nous introduisons plus formellement le problème et les concepts préliminaires de l'extraction de motifs. La section III présente l'échantillonnage dans le cadre d'une base de données statique avec les résultats théoriques sur la précision de l'échantillon et du seuil d'erreur. La section IV étend ces résultats aux flots de données. Une conclusion est proposée dans la section V.

Client	Date	Items
C_1	T_1	a, b, c, d
C_1	T_2	a, c
C_2	T_1	a, b
C_3	T_1	a, d
C_3	T_2	c

Fig. 1. Une base de données exemple \mathcal{D}

II. CONCEPTS PRÉLIMINAIRES

Dans cette section nous présentons la problématique de l'extraction des motifs séquentiels [10], [11] ainsi que le principe d'échantillonnage par réservoir biaisé.

Une transaction constitue, pour un client C , l'ensemble des items achetés par C à une même date. Une transaction s'écrit sous la forme d'un ensemble d'un triplet (id-client, id-date, itemset) où itemset est un ensemble d'items. Une séquence est une liste ordonnée, non vide, d'itemsets notée $\langle s_1 s_2 \dots s_n \rangle$ où s_j est un itemset. Une séquence de données est une séquence représentant les achats d'un client. Soit T_1, T_2, \dots, T_n les transactions d'un client, ordonnées par dates d'achat croissantes et soit $itemset(T_i)$ l'ensemble des items correspondant à T_i , alors la séquence de données de ce client est $\langle itemset(T_1) itemset(T_2) \dots itemset(T_n) \rangle$.

Un client supporte une séquence S si S est incluse dans la séquence de données de ce client. Le support d'une séquence S dans \mathcal{D} , noté $Support(S, \mathcal{D})$, représente le pourcentage des clients qui supportent S . Soit σ un seuil de support minimal défini par l'utilisateur, le problème de l'extraction des motifs séquentiels consiste à extraire toutes les séquences S dans \mathcal{D} telles que $Support(S, \mathcal{D}) \geq \sigma$.

Exemple 1: Considérons la base de données \mathcal{D} de la figure 1. Avec un support minimum de $\sigma = \frac{2}{3}$, les séquences fréquentes dans \mathcal{D} sont les suivantes : $\langle (a) \rangle$, $\langle (b) \rangle$, $\langle (c) \rangle$, $\langle (d) \rangle$, $\langle (a b) \rangle$, $\langle (a d) \rangle$, $\langle (a) (c) \rangle$ et $\langle (d) (c) \rangle$.

L'échantillonnage par réservoir biaisé fut introduit dans [2]. L'idée principale est de réguler l'introduction des points dans le réservoir et ce afin de maîtriser la *fraîcheur* de l'échantillon produit. En d'autres termes, la fonction de biais permet de moduler l'échantillon de manière à se focaliser sur des comportements récents ou anciens en fonction des contraintes de l'application. La fonction de biais est définie de la manière suivante : $f(r, t) = e^{\lambda(t-r)}$ où le paramètre λ correspond au taux de biais. Cette fonction est proportionnelle à $p(r, t)$ avec $r < t$ qui est la probabilité qu'un point introduit dans le réservoir à l'instant r soit encore présent à l'instant t . En outre, l'inclusion d'une fonction de biais exponentielle rend possible l'utilisation d'algorithmes de remplacement simples et surtout cette classe spéciale de fonctions de biais implique aussi une borne

supérieure sur la taille du réservoir qui est indépendante de la longueur du flot. Pour un flot de longueur t , soit $R(t)$ la taille maximale du réservoir qui satisfait la fonction de biais exponentielle, nous avons $R(t) \leq \frac{1}{\lambda}$.

III. ÉCHANTILLONNAGE ET BASE DE DONNÉES STATIQUE

Étant donné que l'un des facteurs clés pour l'extraction est la taille de la base considérée, l'intuition sous-jacente est que l'algorithme d'extraction pourrait être lancé sur un échantillon de la base de données originale afin d'avoir des résultats de manière plus rapide et plus facile.

La première question à laquelle nous devons répondre si nous voulons extraire des motifs à partir d'un échantillon est : *à quel point l'échantillon est-il pertinent par rapport au jeu de données original ?* Nous répondons à cette question en exhibant une garantie sur le taux d'erreur du support d'une séquence. Notons qu'une approche similaire a été proposée pour l'extraction d'itemsets fréquents dans [12].

Définition 1 (Taux d'erreur) Soit \mathcal{D} une base de données transactionnelle de clients et posons $\mathcal{S}_{\mathcal{D}}$ l'échantillon aléatoire généré à partir de \mathcal{D} . Soit s une séquence présente dans \mathcal{D} . Le taux d'erreur absolu en terme d'estimation du support, noté $e(s, \mathcal{S}_{\mathcal{D}})$, est défini tel que : $e(s, \mathcal{S}_{\mathcal{D}}) = |\text{Support}(s, \mathcal{S}_{\mathcal{D}}) - \text{Support}(s, \mathcal{D})|$

Posons $X_{i,s}$ une variable aléatoire indépendante définie telle que :

$$\begin{cases} Pr[X_{i,s} = 1] = p_i & \text{si le } i^{\text{ème}} \text{ client supporte } s, \\ Pr[X_{i,s} = 0] = 1 - p_i & \text{sinon.} \end{cases}$$

$$\text{et } X(s, \mathcal{S}_{\mathcal{D}}) = \sum_i^{|\mathcal{S}_{\mathcal{D}}|} X_{i,s}.$$

$X(s, \mathcal{S}_{\mathcal{D}})$ représente le nombre de clients supportant la séquence s présents dans l'échantillon $\mathcal{S}_{\mathcal{D}}$. Ces clients peuvent être réécrits de la manière suivante : $X(s, \mathcal{S}_{\mathcal{D}}) = \text{Support}(s, \mathcal{S}_{\mathcal{D}}) \times |\mathcal{S}_{\mathcal{D}}|$. De même, l'espérance de la variable $X(s, \mathcal{S}_{\mathcal{D}})$ est $E[X(s, \mathcal{S}_{\mathcal{D}})] = \text{Support}(s, \mathcal{D}) \times |\mathcal{S}_{\mathcal{D}}|$.

Nous voulons estimer la probabilité que le taux d'erreur $e(s, \mathcal{S}_{\mathcal{D}})$ dépasse le seuil ε défini par l'utilisateur, $Pr[e(s, \mathcal{S}_{\mathcal{D}}) > \varepsilon]$.

Pour répondre à cette estimation, nous utilisons une méthode connue en statistiques : les inégalités de concentration (et plus précisément les inégalités de Hoeffding ; [7]) qui permettent de borner la valeur réelle d'une variable aléatoire par rapport à son espérance et un terme d'erreur. Le théorème suivant exhibe une borne inférieure sur la taille de l'échantillon (ou réservoir) :

Théorème 1: Soit s une séquence et $|\mathcal{S}_{\mathcal{D}}|$ la taille du réservoir, alors $Pr[e(s, \mathcal{S}_{\mathcal{D}}) > \varepsilon] \leq \delta$ si $|\mathcal{S}_{\mathcal{D}}| \geq \ln(\frac{2}{\delta}) \frac{1}{2\varepsilon^2}$.

La preuve de ce théorème est présentée dans [5]. Il est à noter que les inégalités de Hoeffding sont souvent considérées comme plus générales et moins précises que les inégalités de concentration de Chernoff, mais le choix de variables aléatoires indépendantes dans l'intervalle $[0, 1]$ donne des résultats similaires pour ces deux inégalités. De plus, il est à remarquer que la taille de l'échantillon $\mathcal{S}_{\mathcal{D}}$ est indépendante de la taille de la base de données initiale et reste conditionnée uniquement par ε et δ .

La table I illustre quelques exemples de tailles d'échantillons en termes de clients pour différentes valeurs de ε et δ . Il est à noter que lorsque les valeurs de ε et

δ sont trop strictes, l'échantillon peut atteindre une taille assez importante.

ε	δ	$ \mathcal{S}_{\mathcal{D}} $
0.01	0.01	26492
0.01	0.001	38005
0.01	0.0001	49518
0.001	0.0025	3333333

TABLE I

DIFFÉRENTES TAILLES D'ÉCHANTILLONS POUR ε ET δ DONNÉS

IV. MOTIFS SÉQUENTIELS, ÉCHANTILLONNAGE ET FLOTS DE DONNÉES

Un des problèmes majeurs qui rend la pratique de l'échantillonnage difficile sur les flots est que l'on ne sait pas à l'avance la taille du flot. Il faut donc développer des algorithmes d'échantillonnages dynamiques qui prennent en compte l'évolution et les changements dans la distribution des données transitant sur le flot. Dans cette section, nous étendons les résultats précédents et présentons un algorithme de maintien d'échantillon (biaisé ou non) de manière dynamique et qui prend en compte les différentes évolutions du flot de données. L'algorithme présenté peut être vu comme une étape de pré-traitement nécessaire afin de permettre l'extraction de séquences fréquentes. Afin que cette étape de pré-traitement soit pertinente, elle doit respecter les conditions suivantes : (i) L'échantillon doit avoir une borne inférieure sur sa taille afin de minimiser le taux d'erreur absolu en terme d'estimation du support. (ii) A cause de la nature même des séquences, les opérations d'insertions et d'enlèvements, nécessaires pour la mise à jour de l'échantillon, doivent se faire au niveau des clients, mais aussi de leurs itemsets. Pour s'en convaincre, il suffit de considérer un ensemble fini de clients avec pour chacun d'eux un grand nombre d'itemsets qui se rajoutent à chaque instant t . Cet ensemble ne peut bien sûr pas être considéré comme un échantillon ou un réservoir car il n'est pas borné et ne fait qu'augmenter avec le flot.

Dans notre modèle de flot de données, un point de données apparaissant à chaque instant t est défini comme un couple constitué d'un identifiant de client et d'une transaction.

Partant de ces contraintes et des résultats théoriques obtenus dans la section III, nous proposons un algorithme simple de remplacement issu de l'approche de réservoir biaisé proposé dans [2] qui permet de réguler l'échantillonnage des transactions des clients sur le flot grâce à une fonction de biaisage temporelle exponentielle. L'idée est la suivante : nous commençons avec un réservoir vide, de capacité maximale $\frac{1}{\lambda}$ et chaque itemset d'un client apparaissant sur le flot est inséré de manière probabiliste dans le réservoir après une opération de *lancer de pièce* : soit par un remplacement des itemsets d'un client déjà présent dans le réservoir, soit par un ajout direct dans une des places encore vacantes. Comme discuté précédemment, nous devons aussi bien contrôler la taille du réservoir en terme de nombre de clients qu'en nombre d'itemsets. Cette opération de contrôle est appliquée grâce à une approche de *fenêtre glissante* (e.g. [1], [3]) qui permet de garder uni-

quement les itemsets les plus récents pour un client donné dans le réservoir. Une fenêtre glissante peut être définie soit comme une fenêtre basée sur les séquences de taille k , contenant les k points les plus récents apparus sur le flot, soit comme une fenêtre basée sur un intervalle de temps de taille t contenant tous les points apparus sur le flot sur une durée de temps t . Dans notre approche nous utilisons des fenêtres glissantes basées sur des séquences afin de garder uniquement les transactions les plus récentes pour les clients présents dans l'échantillon. Ce type de fenêtre glissante permet l'extraction de séquences sur un horizon récent du flot. De plus, la fonction exponentielle de biais permet à l'utilisateur de choisir la taille de son réservoir (avec des contraintes sur l' (ε, δ) -approximation) et ainsi, un échantillon représentatif du flot peut être construit et mis à jour en mémoire selon les besoins de l'application et de l'utilisateur. Le corollaire suivant, issu du théorème 1 exhibe le lien qui existe entre le taux de biais λ et les seuils d'erreurs ε et δ :

Corollaire 1: Soient λ le taux de biais, ε le seuil d'erreur et δ la probabilité maximale telle que $e(s, \mathcal{S}_D) > \varepsilon$, alors :

$$\lambda \leq \frac{2\varepsilon^2}{\ln(2/\delta)}$$

La table II montre quelques valeurs du taux de biaisage λ et la taille minimale du réservoir nécessaire pour la bonne approximation du support des séquences.

ε	δ	λ	$R(t)$
0.01	0.01	0.0000377	26492
0.01	0.001	0.00002631	38005
0.01	0.0001	0.00002019	49518

TABLE II

TAUX DE BIAISAGE λ POUR DIFFÉRENTES VALEURS DE ε ET δ

Le fonctionnement général de l'algorithme est le suivant :

1. A l'arrivée d'un itemset d'un client C_i , voir si le client est dans la liste noire, auquel cas ignorer cet itemset, sinon passer à l'étape 2.
2. Voir si le client C_i est déjà dans le réservoir, si oui, rajouter l'itemset dans la fenêtre et voir s'il y a un décalage nécessaire à faire sinon passer à l'étape 3.
3. Faire un lancer de pièce (tir aléatoire) en cas de succès remplacer un des clients déjà présent dans le réservoir de manière aléatoire (le client remplacé est alors mis dans la liste noire). En cas d'échec, rajouter le client et son itemset dans le réservoir sans rien remplacer.

La partie la plus importante dans l'algorithme est la gestion des décalages des fenêtres glissantes et la mise à jour de la liste noire. Un problème peut apparaître lors de l'étape de remplacement des clients présents dans le réservoir : nous devons détecter si un client était déjà présent dans le réservoir, car le réintroduire sans aucun test préalable rendrait les résultats de l'extraction des motifs séquentiels inconsistants avec la réalité du flot. Le problème d'inconsistance apparaît lorsqu'un client est remplacé par un autre et qu'il revient dans le réservoir à l'instant suivant. Au moment de l'extraction, ce client n'aura pas tous les itemsets qu'il aurait dû avoir dans la fenêtre glissante actuelle. Les points de certains clients doivent donc être ignorés.

Mais d'un autre côté, ignorer des points sur le flot peut introduire un nouveau biais, puisque seuls les clients ayant remplacé d'autres clients seront présents dans l'échantillon. Afin de résoudre ce problème d'inconsistance entre les itemsets des différents clients nous introduisons un système de *liste noire* qui permet d'interdire l'échantillonnage à un certain nombre de clients indésirables. Cette liste noire n'est pas irréversible et est réactualisée à chaque glissement de la fenêtre.

Comme dans [2], nous démontrons dans [5] que la politique de remplacement appliquée dans notre algorithme permet de construire un réservoir biaisé de taille $|\mathcal{S}_D| = \frac{1}{\lambda}$ avec $\lambda = \frac{1}{n}$.

Proposition 1: Notre approche construit un réservoir biaisé respectant la fonction de biaisage temporel $f(r, t) = e^{-\lambda(t-r)}$ avec $\lambda = \frac{1}{S_D}$.

V. CONCLUSION

Dans cet article, nous nous sommes intéressés à de nouvelles techniques de résumés pour représenter des bases de données de motifs séquentiels. Nous avons montré qu'une approche basée sur des échantillons était tout à fait adaptée pour des bases de données statiques et que nous étions capables de maîtriser les taux d'erreur dans les résultats d'extraction de motifs séquentiels. A notre connaissance, ce travail est le premier à utiliser des techniques d'échantillonnage pour extraire des motifs séquentiels dans des bases de données. Nous avons également montré qu'une approche d'échantillonnage basée sur des réservoirs pouvait être adaptée au contexte des flots de données et avons proposé un algorithme de remplacement des éléments du réservoir qui permet de réguler l'échantillonnage des transactions des clients sur le flot via une fonction de biaisage temporelle exponentielle.

RÉFÉRENCES

- [1] C. Aggarwal, editor. *Data Streams : Models and Algorithms*. Springer, 2007.
- [2] C. C. Aggarwal. On biased reservoir sampling in the presence of stream evolution. In *Proc. of VLDB 06*, pages 607–618, 2006.
- [3] Shivnath Babu and Jennifer Widom. Continuous queries over data streams. *SIGMOD Record*, 30(3) :109–120, 2001.
- [4] Y. Chi, H. Wang, P.S. Yu, and R.R. Muntz. Moment : Maintaining closed frequent itemsets over a stream sliding window. In *Proc. of ICDM 04*, pages 59–66, 2004.
- [5] C. Raissi et P. Poncelet. Sampling for sequential pattern mining : From static databases to data streams. In *Proc. of ICDM 2007*, 2007.
- [6] G. Giannella, J. Han, J. Pei, X. Yan, and P. Yu. Mining frequent patterns in data streams at multiple time granularities. In *Next Generation Data Mining*, MIT Press, 2003.
- [7] W. Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301) :13–30, 1963.
- [8] G. Manku and R. Motwani. Approximate frequency counts over data streams. In *Proc. of VLDB 02*, pages 346–357, 2002.
- [9] Chedy Raissi, Pascal Poncelet, and Maguelonne Teisseire. Speed : Mining maximal sequential patterns over data streams. In *Proc. of IS 06*, 2006.
- [10] R. Srikant and R. Agrawal. Mining sequential patterns : Generalizations and performance improvements. In *Proc. of EDBT 96*, pages 3–17, 1996.
- [11] R. Agrawal R. Srikant. Mining sequential patterns. In *Proc. of ICDE 95*, 1995.
- [12] H. Toivonen. Sampling large databases for association rules. In *Proc. of VLDB 96*, 1996.

DOCTISS 2008

Dis moi ce que tu consommes, je te dirai qui tu es.

Nadia EL MRABET¹, Jean Claude BAJARD² & Sylvain DUQUESNE³

LIRMM - I3M - CNRS
Université Montpellier 2

161 rue ada
34092 MONTPELLIER Cedex , France

¹elmrabet@lirmm.fr, ²bajard@lirmm.fr & ³duquesne@math.univ-montp2.fr

Résumé— La cryptographie moderne se fonde sur la difficulté des ordinateurs actuels à faire des opérations sur les grands nombres. Pour pouvoir malgré tout récupérer les secrets qu'ils convoitent les espions utilisent des fuites lors des calculs de déchiffrement. Dans le cas présent, on s'intéresse aux traces de consommation de courant d'un appareil électrique.

Mots-clés— Cryptographie, RSA, clé publique, clé privée, SPA.

I. INTRODUCTION

Dis moi ce que tu manges, et je te dirai qui tu es s'applique habituellement à des personnes ou être vivant. Cet adage a été appliqué à des appareils électroniques utilisant un secret lors de leur calcul. Nous commencerons par une courte introduction à la cryptographie et son utilisation au quotidien. Ensuite, nous nous placerons à la place d'un espion souhaitant récupérer des informations qui ne lui sont pas destinées. Nous verrons de quels moyens il dispose pour parvenir à ses fins ; en particulier la lecture de la consommation de courant. Pour conclure, nous analyserons quelles contre mesures sont applicables à l'attaque que nous aurons décrite, et comment en tant qu'espion nous pouvons l'outre passer.

II. CRYPTOGRAPHIE

La cryptographie est d'après un dictionnaire usuel : "l'ensemble des principes, méthodes et techniques dont l'application assure le chiffrement et le déchiffrement des données afin d'en préserver la confidentialité et l'authenticité". Autrement dit, tout moyen permettant de communiquer avec quelqu'un sans que des personnes étrangères (i.e. des espions) ne puissent récupérer et surtout comprendre les messages. Historiquement, la cryptographie était surtout utilisée à des fins militaires.

Le plus ancien protocole connu est celui des esclaves de Nabuchodonosor, roi de Babylone. Sa technique lui permettait de communiquer avec ses généraux. Il rasait un esclave, et lui faisait tatouer sur le crâne le message à transmettre. Puis il attendait que les cheveux repoussent avant d'envoyer l'esclave à ses généraux. Cette méthode, était non seulement très lente, mais aussi très facile à

décrypter. En effet, il suffisait d'attraper l'esclave et de le raser. De plus, l'esclave ne pouvait "servir" qu'une fois.

Le chiffrement le plus célèbre est le code de César. Il consistait à remplacer chaque lettre du message par la lettre suivante de l'alphabet, ou par la lettre suivante en décalant de trois cran. Exemple, CAESAR deviendrait FDHVDU où $C \rightarrow F$, on décale chaque lettre de trois crans vers la droite. Ce protocole a traversé les époques tout en se sophistiquant. Enigma le système de chiffrement utilisé par les allemands durant la deuxième guerre mondiale est un héritier du code de César.

Pour chiffrer et déchiffrer il fallait posséder une machine Enigma. La méthode de chiffrement est la suivante, à chaque lettre du message la machine associe une lettre différente, cette lettre était tirée au hasard. Les allemands ont pensé que leur protocole cryptographique était sûr car impossible à "casser" à la main. Casser un protocole cryptographique signifie pouvoir retrouver le message clair à partir du message chiffré.

Les protocoles décrits au dessus ont un point commun qu'il est important de souligner : la méthode de chiffrement est elle même secrète. Ce qui n'est pas exempt d'inconvénients. Par exemple, César ne pouvait communiquer avec un général en particulier sans que tous les autres ne puissent décrypter le message.

Ou encore, la découverte d'une machine Enigma dans un sous marin a permis aux alliés de comprendre la méthode de chiffrement. Le déchiffrement nécessitaient un grand nombre de combinaison et les allemands comptaient sur ce point pour rendre leur protocole cryptographique sûr. Ils avaient compté sans un certain Turing, qui avec la machine de Turing a réussi à venir à bout des 10^{16} possibilités associées à un message donné. La machine de Turing nous est plus familière sous le nom d'ordinateur...

Ces méthodes de chiffrement ne vérifient pas le principe de Kerckhoffs. Selon ce principe :

La sécurité d'un protocole cryptographique doit être garantie par les clés de chiffrement et déchiffrement utilisées, et non par la méthode de chiffrement.

Ce qui a donné naissance à la cryptographie moderne, cette cryptographie se caractérise par le fait que la méthode de chiffrement est connue de tous, la méthode de déchiffrement aussi, mais elle nécessite une clé secrète qui elle doit être bien cachée. On parle de cryptographie asymétrique. La clé publique permet de chiffrer un message, cette clé est connue de tous. Pour déchiffrer un message, il faut la clé secrète associée à la clé publique. Le couple (clé publique, clé secrète) est construit de manière à ce que pour un protocole donné il soit difficile de retrouver la clé secrète à partir de la clé publique. Nous allons illustrer le propos en développant l'exemple concret du protocole RSA.

III. PROTOCOLE RSA

A. Principe du protocole RSA

Le protocole RSA du nom de ses inventeurs Rivest Shamir Adleman est un protocole assymétrique à clé publique. Il est utilisé à tous les niveaux de la société. Au gouvernement et bien sur par les armées, le système bancaire, carte à puce, sur internet pour le commerce et l'échange de données, pour le téléphone mobile, la télévision payante...

La sureté du protocole RSA réside dans la difficulté à factoriser des grands nombres. La clé publique est un couple (N, e) où N est le produit de deux grands nombres premiers p, q , et e un entier compris entre 0 et $\phi(N) = (p-1) \times (q-1)$. On considère qu'un message M est un chiffre, la méthode de chiffrement consiste alors à élever le message à la puissance e , modulo N . Le message chiffré $C = M^e \text{ modulo } (N)$. La méthode de déchiffrement utilise des propriétés sur les entiers, et en particulier, le petit théorème de Fermat. Les rappels mathématiques sont dans la partie III-A.1.

L'entier e choisit, n'est pas vraiment pris au hasard, il doit être possible d'associer à e , un entier d compris entre 1 et $\phi(N)$ qui soit l'inverse de $e \text{ modulo } (\phi(N))$. Ce qui se traduit par le fait que $e \times d \equiv 1 \text{ modulo } (\phi(N))$. Pour un tel couple d'entiers (e, d) , la propriété suivante est vérifiée :

Théorème Soit N le produit de nombre premier p et q . Pour tout entier M compris entre 0 et $N - 1 = pq - 1$, pour tout couple (e, d) d'entiers compris entre 0 et $\phi(N)$ tel que $e \times d \equiv 1 \text{ modulo } (\phi(N))$, on sait que $M^{ed} \equiv M \text{ modulo } (N)$.

La méthode de déchiffrement est alors simplement une exponentiation à la puissance d . Pour retrouver le message d'origine, il suffit d'élever le message crypté C à la puissance d , puisque $C^d = (M^e)^d = M^{(ed)} \equiv M \text{ modulo } (N)$. L'entier d est appelé la clé privée associée à la clé publique (N, e)

A.1 Rappel

Définition : modulo

$a \text{ mod } (N)$ correspond au reste de la division euclidienne de a par N .

La démonstration du protocole RSA repose sur le petit théorème de Fermat.

Théorème : Petit théorème de fermat Soit p un nombre premier, et x un entier compris entre 1 et $p - 1$, alors $x^{p-1} \equiv 1 \text{ mod } (p)$.

B. Exemple

Nous allons illustrer par un exemple la méthode RSA, et voir comment en tant qu'espion nous pouvons espérer intercepter et comprendre les informations échangées.

En cryptographie, les personnages principaux sont généralement Alice, Bob et Eve. Alice et Bob sont les personnes souhaitant communiquer de manière secrète, et Eve est l'espionne voulant comprendre quel terrible secret ils peuvent bien s'échanger. Plaçons nous dans le cas où Bob souhaite envoyer un message secret à Alice. Alice transmet à Bob sa clé publique, qui est un couple (N_A, e_A) . Lorsque Bob veut envoyer son message M à Alice, il applique le protocole de chiffrement et calcule donc $C = M^{e_A} \text{ mod } (N)$. Il envoie ensuite C à Alice sans avoir besoin de le cacher. Pour le décrypter, Alice va utiliser sa clé privée d_A . Où d_A est l'inverse de $e_A \text{ mod } (\phi(N))$. Alice calcule $C^{d_A} \text{ mod } (N)$ pour retrouver le message d'origine.

Eve souhaite savoir de quoi il retourne, elle connaît le couple (N_A, e_A) , et a la possibilité d'intercepter le message C . Si elle parvient à trouver d elle va pouvoir retrouver le message secret. Elle pourrait lancer une recherche exhaustive sur tous les entiers compris entre 0 et $\phi(N_A)$, pour obtenir d_A mais cela est bien trop long. La sécurité du protocole RSA repose en effet sur la difficulté à retrouver p et q connaissant $N_A = p \times q$, et la connaissance de p et q est indispensable pour calculer $\phi(N) = (p - 1)(q - 1)$.

Elle peut alors tenter de récupérer la clé privée d_A en piratant l'appareil servant à décrypter. Même dans le cas où elle parviendrait à mettre la main dessus, Alice aura été assez maligne pour cacher la clé derrière tout un tas de bouclier. Eve doit donc renoncer à retrouver la clé de manière frontale. Qu'à cela ne tienne, elle va utiliser les fuites d'informations qui se font lors du calcul de C^d . Ces informations sont appelées canaux cachés, il s'agit, du temps mis pour calculer, ou de la consommation de courant faite lors de l'exponentiation. Nous allons décrire l'attaque SPA, acronyme anglais signifiant Simple Power Analysis, soit attaque par analyse simple de consommation de courant.

IV. ATTAQUE SPA

L'attaque SPA est une attaque simple à canaux cachés. Eve peut retrouver des informations par l'analyse de la courbe représentative de la consommation de courant lors de l'exponentiation à la puissance d_A . Cette courbe de consommation est différente suivant les instructions exécutées et les données manipulées. La méthode la plus utilisée pour élever à une puissance donnée et l'algorithme d'exponentiation rapide.

L'algorithme est le suivant :

Algorithme d'exponentiation rapide

On veut calculer C^{d_A}

- 1 - Calculer la décomposition binaire de d_A ,
 $d_A = \overline{d_n d_{n-1} \dots d_1 d_0}^2$
- 2 - $T \leftarrow C$
- 3 - **Pour** $i = n - 1$ à 0 **faire**
- 4 - $T \leftarrow T \times T$
- 5 - **Si** $d_i = 1$ **Alors** $T \leftarrow T \times C$
- 6 - **Renvoyer** T

Lorsque Eve analyse de près l'algorithme, elle s'aperçoit qu'il n'y a jamais deux multiplications consécutives, en effet, à chaque itération, un carré est calculé, mais la multiplication n'est pas faite pour toute. Il y a clairement deux comportements différents de la courbe, la courbe de consommation suit un tracé régulier, qui traduit les opérations effectuées pendant l'exponentiation, soit un carré, soit une multiplication. Comme ces deux opérations ne nécessitent pas le même nombre d'opération élémentaires, la consommation pour effectuer un carré et une multiplication n'est pas la même. Utilisant le fait qu'il ne peut y avoir deux multiplications consécutives, Eve peut déterminer les parties de la courbe correspondant à un carré, et les parties de la courbe correspondant à une multiplication. Ce qui directement lui permet de savoir si le bit de l'exposant lors de l'itération lue est à 0 ou à 1.

Par une lecture attentive de la courbe de consommation, Eve peut déterminer l'exposant utilisé, c'est à dire retrouver d_A . Illustration :

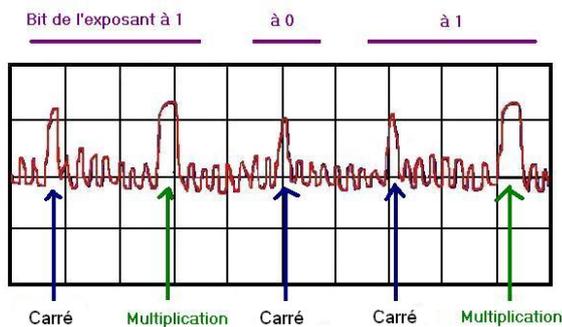


Fig. 1. Consommation de l'exponentiation rapide décryptée

V. CONTRE MESURES

Les contre mesures sont les modifications à apporter à l'algorithme pour rendre l'attaque décrite inefficace.

L'attaque SPA que nous venons de décrire utilise une faille cachée dans la courbe de consommation de courant.

Afin d'empêcher cette attaque, la parade est simple. Il suffit de faire en sorte que la courbe de consommation de courant ne puisse apporter d'information sur l'exposant. Ce qui équivaut à ce que la courbe de consommation soit uniforme. Si la consommation de courant est la même pour toute les itérations, il ne sera plus possible de différencier les itérations entre celle où un carré seul est fait, et celle où un carré et une multiplication sont faits. Ceci est très facilement réalisable en ajoutant des opérations factices

dans l'algorithme d'exponentiation rapide. Ce qui donne l'algorithme suivant :

Algorithme d'exponentiation rapide sécurisé contre SPA

On veut calculer C^{d_A}

- 1 - Calculer la décomposition binaire de d_A ,
 $d_A = \overline{d_n d_{n-1} \dots d_1 d_0}^2$
- 2 - $T \leftarrow C, V \leftarrow C$
- 3 - **Pour** $i = n - 1$ à 0 **faire**
- 4 - **Si** $d_i = 0$ **Alors** $T \leftarrow T \times T$ & $V \leftarrow T \times C$
- 5 - **Si** $d_i = 1$ **Alors** $T \leftarrow T \times T$ & $T \leftarrow T \times C$
- 6 - **Renvoyer** T

À chaque itération de cet algorithme, un carré et une multiplication sont faits. Lors de l'étape 5, la multiplication faite est une multiplication inutile pour l'exponentiation, elle est donc stockée dans une variable muette V qui n'intervient pas lors des calculs nécessaires à l'exponentiation. La courbe de consommation devient alors la suivante, la lecture de cette courbe signifiera que tous les bits de l'exposant sont 1...

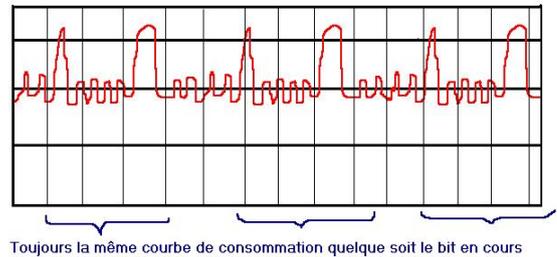


Fig. 2. Consommation de l'exponentiation rapide avec contre mesures

VI. CONCLUSION

À travers l'exemple de l'algorithme RSA, nous venons de voir que la cryptographie est l'art de transmettre des messages cryptés sans qu'une tierce personne ne puisse les intercepter et les comprendre. Cependant, les secrets attirent suffisamment les convoitises pour qu'il existe toujours des tierces personnes tentant de découvrir les secrets échangés. Ces espions plus ou bien intentionnés s'arrangent toujours pour trouver une faille dans l'algorithme et à l'exploiter afin de décoder des messages interceptés. La cryptographie s'apparente à une course entre la mise au point de protocoles sûrs et le cassage de ces protocoles.

À travers l'exemple de l'algorithme RSA, nous avons introduit l'attaque SPA, qui consiste simplement à lire attentivement la courbe de consommation de l'appareil effectuant le décryptage, pour retrouver la clé secrète utilisée. Une contre mesure simple mais efficace permet de rendre l'attaque SPA inutilisable. Les espions n'étant jamais à court d'imagination, des méthodes pour casser un algorithme sécurisé par l'ajout d'opérations factices ont été mises au point. Il s'agit des attaques par injection de fautes. Ces attaques trouvent leur source dans l'adage, "comment apprendre de ses erreurs".

À suivre donc : "Comment apprendre de ses erreurs".

RÉFÉRENCES

- [1] H.Cohen and G.Frey *Elliptic and Hyperelliptic Curve Cryptography* Editions Chapman and hall, 2006.

DOCISS 2008

DagMap : Exploration de hiérarchies

Pierre-Yves KOENIG¹ (Doctorant 2e année), Guy MELANÇON² (Directeur de thèse)

LIRMM UMR 5506
161 Rue Ada 34392 Montpellier Cedex 5, France

¹Pierre-Yves.Koenig@lirmm.fr, ²Guy.Melancon@labri.fr

Résumé— Cet article présente une nouvelle technique pour explorer des données hiérarchiques, qu’il s’agisse d’arbres ou de graphes acycliques orientés (DAGs). Grâce à la combinaison d’une vue noeuds-liens, d’une TreeMap adaptée et d’un mécanisme de filtrage, l’utilisateur peut naviguer dans les données à différents niveaux de détails.

Mots-clés— Visualisation d’information, TreeMap, DAG, hiérarchie, focus+context.

I. INTRODUCTION

Les données hiérarchiques sont de plus en plus présentes dans nos applications : que ce soit pour décrire le contenu d’un disque dur avec une arborescence de fichiers ou encore dans la classification des espèces. Dans cet article, nous nous intéressons à des hiérarchies plus complexes : on permet à un élément de la hiérarchie d’avoir plusieurs antécédents (pères) donnant lieu à une hiérarchie sous forme de DAG - au lieu d’un arbre. Ce type de structures caractérise les relations de classes dans la programmation orientée objet (une classe peut être la spécialisation d’une ou de plusieurs classes, héritant des propriétés de ses classes parentes). Les relations d’héritage apparaissent aussi dans la description de relations entre entités dans d’autres domaines d’application. Par exemple, les relations entre des sociétés et leurs filiales (à tous niveaux). Une société peut avoir différentes filiales et une filiale peut être contrôlée par plusieurs sociétés “mères”.

Différentes techniques permettent de représenter ces données. Les représentations “noeuds-liens” dessinent le plus souvent les sommets sur différents niveaux (cf. Figure 1). Les entités les plus générales - sources - correspondant aux sommets sans ancêtre sont dessinées en haut de la hiérarchie, tandis que les autres sommets sont dessinés suivant leur distance aux sommets sources. La qualité et la lisibilité de ces représentations noeuds-liens sont le plus souvent mesurées en fonction de leur capacité à éviter le croisement d’arêtes. Les TreeMaps sont des techniques de visualisation pour représenter des hiérarchies d’information sur un espace 2D [7]. Les TreeMaps suivent une approche de pavage de l’espace (“space-filling”) représentant les sommets feuilles d’un arbre sur des aires contiguës du plan, avec différents artifices visuels rendant compte d’attributs des données. [L’aire elle-même peut être calculée selon un des attributs des feuilles.] Cette approche diffère radicalement des représentations classiques noeuds-liens des arbres (voir [1], [4]) où l’accent est mis sur la position relative des

sommets reflétant la structure de la hiérarchie, par opposition à la sémantique des données (des sommets feuilles). La méthode proposée ici combine les deux techniques afin de mettre l’accent sur la structure de la hiérarchie et sur des attributs des données. Après avoir exposé notre cas d’étude, nous allons expliquer comment étendre les TreeMaps aux DAGs. Une navigation par niveau de détails sera alors exposée.

II. CAS D’ÉTUDE

Dans le cadre du projet ANR SPANGEO, les géographes s’intéressent aux stratégies de gouvernance d’entreprises européennes et leurs filiales. Les sociétés Fiat, Peugeot (ou encore Nestlé, Danone et Bongrain) - appelées maisons mères - possèdent de nombreuses filiales en Europe et dans le monde. Ces filiales possèdent elles-mêmes des sous-filiales. On note aussi qu’une société peut être contrôlée par plusieurs sociétés mères. Ces relations entre une société et ses filiales peuvent être représentées par un DAG. Une vue

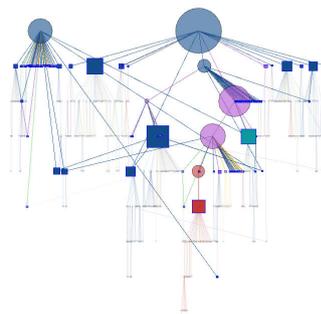


Fig. 1. Représentation du DAG de gouvernance de Fiat avec l’algorithme de Sugiyama.

classique du DAG (cf. Figure 1) permet aux géographes de voir la structure de la hiérarchie. Le DagMap (cf. Figure 2) permet de comparer les sociétés suivant différents attributs. L’aire des rectangles représente la somme des capitaux de la société et de ses filiales pondérée par la participation. La couleur représente la localisation de la société (couleur attribuée par les géographes suivant le continent d’appartenance : bleu pour l’Europe, rouge pour l’Amérique du Nord, vert pour l’Amérique du Sud, jaune pour l’Asie, marron pour l’Afrique et enfin fuchsia pour les paradis fiscaux). Des stratégies économiques et géographiques ont pu être mises en évidence. Une stratégie économique, visant à localiser les

filiales financières d'une entreprise dans des territoires à fiscalité privilégiée, est caractérisée par la présence de la couleur fuchsia (cf. Figure 2). Une comparaison des différentes

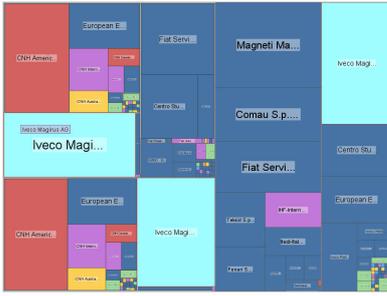


Fig. 2. DagMaps de la société Fiat. En cliquant sur une cellule, on sélectionne un élément et toutes ses occurrences dans le DagMap - sélection mise en évidence par la couleur turquoise. Une aide contextuelle (tooltips) permet de visualiser un attribut de la cellule.

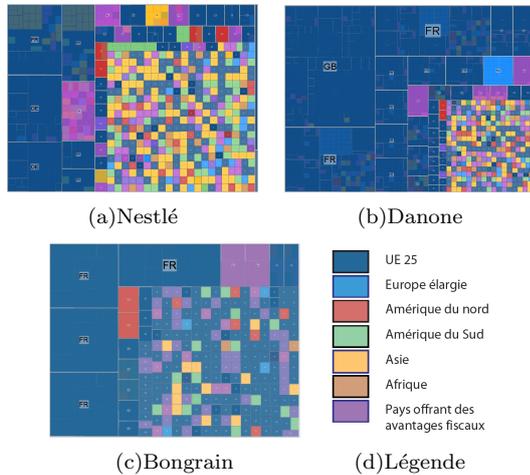


Fig. 3. DAG de gouvernance de Nestlé, Danone et Bongrain avec un DagLevel de 2.

compagnies est possible suivant différents niveaux. Ainsi, nos géographes (utilisateurs experts) ont pu déterminer des stratégies communes aux entreprises Nestlé, Danone et Bongrain. Ces sociétés, bien qu'européennes, possèdent des filiales à l'extérieur de l'Europe. Celles-ci sont alors directement contrôlées par la maison mère au plus haut niveau montrant une certaine frilosité des compagnies européennes à déléguer le contrôle des filiales hors de leur territoire. Cela se traduit par la présence d'un nombre important de compagnies non européennes visibles à un niveau élevé (cf. Figure 3).

III. ÉTENDRE LES TREEMAPS AUX DAGS

Les TreeMaps ne permettent de visualiser que des arbres. Afin de visualiser un DAG, nous allons le transformer en arbre en dupliquant les sous-arbres de sommets ayant plusieurs prédécesseurs (cf. Figure 4). La taille de l'arbre obtenu du DAG peut être potentiellement très large, dépendant du nombre d'arêtes transversales. Ce problème n'est pas gênant quand on traite des DAGs peu denses (ce qui est notre cas). Il est à noter, comme c'est le cas dans DynaDags [6], que la plupart des techniques traitant

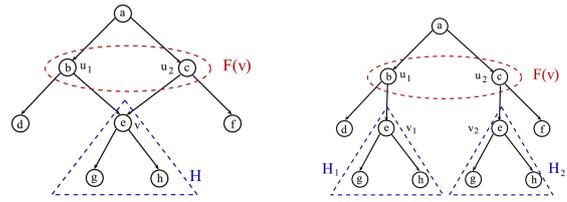


Fig. 4. Un arbre est obtenu d'un DAG en dupliquant les noeuds (et étiquettes) ayant plusieurs ancêtres.

des DAGs souffrent de cette limitation (voir aussi [5], [3]). Cette limitation est contournée par le fait que l'arbre n'est calculé qu'à partir d'une partie du DAG (voir section IV), permettant de garder la taille de l'arbre sous contrôle. Cette duplication de sommets s'accompagne de différentes interactions permettant à l'utilisateur de retrouver la structure du DAG. En cliquant sur une cellule dans le DagMap, les différentes cellules ayant la même étiquette sont colorées rendant ainsi compte de l'importance relative de l'élément (cf. Figure 2).

IV. NAVIGATION PAR NIVEAU DE DÉTAILS

La visualisation globale d'une hiérarchie de grande taille est difficile. En effet dans une vue noeuds-liens, on ne peut afficher à l'écran l'ensemble des éléments de la hiérarchie sans éviter les chevauchements de sommets et croisements d'arêtes pénalisant sa lisibilité. La visualisation du DagMap est elle aussi plus difficile si la hiérarchie s'étale sur trop de niveaux. Une vue abstraite de cette hiérarchie permet d'en avoir une meilleure compréhension. L'utilisateur peut, dans cette vue abstraite lui donnant un contexte, vouloir localement plus de détails. Le travail fondateur de Furnas[2] a introduit la gestion des niveaux de détails en exploitant les paramètres structuraux de la hiérarchie sous-jacente aux données (un arbre dans son cas, encodant l'imbrication des blocs de code source C). Grâce à un indice à priori d'intérêt (*API*) et un centre d'intérêt défini par l'utilisateur, la hiérarchie est filtrée. Les éléments trop éloignés du centre d'intérêt sont abstraits, seule reste l'instruction du début du bloc d'instructions C (ex : `while {...}`). Cette technique et celle de Van Wijk et Van Ham[8] (voir ci-dessous) correspondent à déterminer et afficher une antichaine dans la hiérarchie, qui peut être vue comme un ensemble ordonné. Dans le cas de Furnas, les éléments affichés sont ceux de l'antichaine et tous ceux qui les précèdent. En effet, si une instruction est affichée, les instructions de plus hauts niveaux le sont aussi. Van Wijk et Van Ham[8] proposent une méthode qui généralise l'approche de Furnas. Les sommets d'un graphe peuvent être regroupés en grappes (clusters) qui peuvent être elles-mêmes regroupés en grappes ... constituant un arbre de clustering. Ils étudient le cas où l'ensemble ordonné correspond à l'arbre de clustering d'un graphe. A chaque élément de l'ensemble ordonné (cluster) est associé une mesure de disparité d , mesure qui peut refléter un attribut des données (ex : le capital d'entreprise). Cette méthode consiste à n'afficher que les sommets k respectant l'inégalité : $d_k \leq d_{root} \times DOA < d_{fk}$ avec fk le prédécesseur de k , d_{root} la disparité maximale et $DOA \in [0, 1]$ le degré d'abstraction souhaité. Un sommet f de la hiérarchie peut être pris comme centre

d'intérêt, l'inégalité est alors : $d_k \leq d_{root} \times DOA(|k - f|)$ et $d_{root} \times DOA(|fk - f|) < d_{fk}$ où $|k - f|$ est la distance (euclidienne ou dans le graphe) entre les sommets k et f . Ces mécanismes développés pour les arbres fonctionnent aussi lorsque la hiérarchie est un DAG. Le DAG pouvant être vu comme un ensemble ordonné, on peut lui appliquer cette méthode. Toutefois, dans un DAG ces inégalités ne produisent pas forcément une antichaine, on parle alors de "coupe". Une "coupe" est une sélection de certains éléments du DAG se trouvant à des niveaux contigus (mais pas forcément situés sur un même niveau) et couvrant toute sa largeur. On se propose de combiner les deux approches. Sur un DAG, différentes mesures topologiques peuvent être calculées et jouer le rôle d'indice de disparité. L'utilisateur ayant déterminé le niveau de détails souhaité, on peut calculer une coupe avec les inégalités ci-dessus. On peut alors afficher les sommets de cette coupe ainsi que tous leurs prédécesseurs - comme c'est le cas dans la méthode de Furnas - ou au contraire n'afficher que les successeurs. Nous avons travaillé à concevoir un artifice visuel capturant la combinatoire sous-jacente au mécanisme de Furnas et de Van Wijk et Van Ham. (cf. Figure 5)

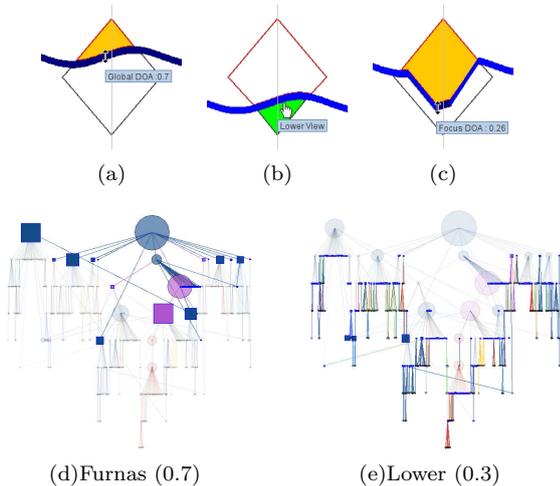


Fig. 5. Curseur contrôlant la vue noeuds-liens et le DagMap. Différentes interactions : tooltips, curseur, sélection de régions

V. EXPLORATION DE LA STRUCTURE HIÉRARCHIQUE À TRAVERS LE DAGMAP

Notre expérience avec des utilisateurs finaux a clairement montré la nécessité de visualiser la structure de la hiérarchie directement sur le DagMap. Nous permettons donc à l'utilisateur de visualiser comment les sommets ancêtres couvrent les cellules dans le DagMap. Après avoir sélectionné une coupe, l'utilisateur peut définir une "bande" rassemblant tous les éléments à une certaine distance au-dessus de la coupe. Ce faisant, l'utilisateur peut visualiser comment les éléments dépendent de leurs ancêtres.

En faisant varier la hauteur de la bande, l'utilisateur obtient des informations sur les attributs des cellules d'ancêtre et la taille de son voisinage. Cette fonctionnalité s'est révélée très utile sur notre jeu de données, rendant évident le contrôle de filiales par des sociétés jouant le rôle de paradis fiscaux (cellules de couleur fuchsia), comme le montre l'exemple de la figure 6.

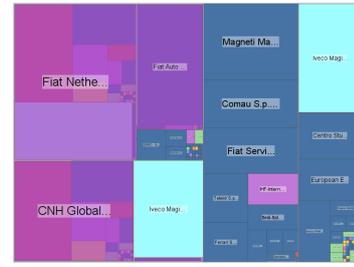


Fig. 6. Le DagMap montré ici est obtenu à partir de celui de la figure 2 auquel on a ajouté une bande. L'emboîtement des filiales devient clair quand on peut visualiser les sociétés mères se trouvant au-dessus d'elles. Le code couleur indique que les sociétés sont contrôlées par des sociétés mères suspectées de jouer le rôle de paradis fiscaux.

VI. REMERCIEMENTS

Ce travail a été réalisé dans le cadre du projet ANR SPANGEO. Nous remercions nos collègues géographes qui ont apporté un oeil expert et critique sur la visualisation proposée. Nous remercions aussi Laurent Perrier qui s'est chargé de la collecte des données.

VII. CONCLUSION

Cet article présente une technique afin de visualiser des hiérarchies suivant différents niveaux de détails. Une représentation nouvelle a été élaborée combinant une vue noeuds-liens et un DagMap qui est l'extension des Tree-Maps aux DAGs. Le travail de conception de l'interface s'est inscrit dans le cadre d'un travail "rapproché" avec les utilisateurs experts. Les manipulations proposées correspondent aux demandes de nos utilisateurs. L'outil fait l'objet d'une utilisation quasi-quotidienne dans le cadre d'un projet pluri-disciplinaire (ANR SPANGEO). Toutefois, nous projetons d'effectuer une expérimentation contrôlée afin de conforter cette validation "sur le terrain".

RÉFÉRENCES

- [1] G. di Battista, Peter Eades, Roberto Tamassia, and Ian G. Tollis. *Graph Drawing : Algorithms for the Visualisation of Graphs*. Prentice Hall, 1998.
- [2] George W. Furnas. Generalized fisheye views. In *Human Factors in Computing Systems CH'86 Conference Proceedings*, pages 16–23, 1986.
- [3] C. Gutwenger and P. Mutzel. An experimental study of crossing minimization heuristics. In B. Liotta, editor, *Graph Drawing 2003*, volume 2912 of *Lecture Notes in Computer Science*, pages 13–24. Springer-Verlag, 2004.
- [4] Michael Kaufmann and Dorothea Wagner, editors. *Drawing Graphs, Methods and Models*, volume 2025 of *Lecture Notes in Computer Science*. Springer, 2001.
- [5] Guy Melançon and Ivan Herman. Dag drawing from an information visualization perspective. In W. de Leeuw and R. van Liere, editors, *Joint Eurographics and IEEE TCVG Symposium on Visualization (Data Visualization '00)*, pages 3–13, Amsterdam, 2000. Springer-Verlag.
- [6] Stephen C. North and Gordon Woodhull. Online hierarchical graph drawing. In *Graph Drawing : 9th International Symposium, GD 2001*, pages 77–81, Vienna, Austria, 2002.
- [7] Ben Shneiderman. Tree visualization with tree-maps : A 2-D space-filling approach. *ACM Transactions on Graphics*, 11, 1992.
- [8] Frank van Ham and Jarke van Wijk. Interactive visualization of small world graphs. In *INFOVIS '04 : Proceedings of the IEEE Symposium on Information Visualization (INFOVIS'04)*, pages 199–206. IEEE Computer Society, 2004.

Conflits entre voisins dans les réseaux mobiles

Hamidou TEMBINE¹ (Doctorant 2e année), Eitan ALTMAN² (Directeur de Thèse), Rachid ELAZOUZI³ (Directeur de Thèse)

Laboratoires de Recherche

¹Laboratoire d'Informatique d'Avignon(LIA), Université d'Avignon

²Groupe MAESTRO, INRIA Sophia Antipolis, France

¹tembine@ieee.org, ²Altman@sophia.inria.fr, ³Rachid.ElAzouzi@univ-avignon.fr

Résumé— Le mécanisme de contrôle d'accès MAC du type ALOHA est très utilisé dans les grands réseaux de mobiles utilisant des communications sans fil et avec un ou plusieurs relais. Ce mécanisme est fondé sur une représentation des interférences (et donc des collisions) par zone de voisinage des récepteurs. Il est donc important pour l'ensemble des mobiles se trouvant dans les voisinages proches de trouver un compromis entre la densité de positionnement des communications et la portée des transmissions pour estimer le taux de transmissions réussies. Le sort des uns ne dépendant pas seulement de leur propre choix mais de celui des autres, il y a des interactions stratégiques locales par zones de voisinage qui s'intersectent. Nous étudions un contrôle d'accès au médium (Medium Access Control – MAC) utilisant un protocole du type ALOHA en modélisant les interactions locales comme des jeux stratégiques et le système global comme un jeu de population. Différentes approches (stratégie évolutionnairement stable, stratégie corrélée, équilibre de Nash) sont proposées pour mieux gérer les risques de collisions des paquets de données transmises.

Mots-clés— théorie des jeux, réseaux mobiles, stratégie corrélée évolutionnairement stable.

I. INTRODUCTION

La théorie des jeux se définit généralement comme l'outil mathématique permettant d'analyser les interactions stratégiques entre joueurs (individus, utilisateurs, mobiles, agents etc), en particulier lorsque ces derniers ont des intérêts divergents ou des ressources à partager. Elle s'intéresse à toutes configurations dans lesquelles la situation de chacun dépend du comportement de tous et analyse donc les comportements stratégiques. Il est donc naturel d'introduire la théorie des jeux pour analyser le contrôle d'accès et d'interférences entre mobiles. L'objet d'étude de la théorie des jeux, à savoir les comportements stratégiques, implique un champ d'application extraordinairement vaste : économie et gestion, mais aussi stratégie militaire, science politique, philosophie, biologie et maintenant l'informatique et les télécommunications. Nous étudions un problème de contrôle d'interférence dans un réseau en modélisant les interactions comme des jeux sous formes stratégiques ou des jeux de populations.

Le reste de ce document est organisé comme suit : Nous revisitons dans la sous-section suivante le célèbre exemple du jeu de la musique qui illustre bien le problème que nous étudions. Dans la sous-section I-B nous décrivons un exemple simple de conflits d'interférence entre quatre mo-

biles autour de trois récepteurs qui illustre bien les interactions non-symétriques. La section II présente des concepts de solutions : équilibre de Nash [5], équilibre corrélé [2], [1] et stratégie évolutionnairement stable [8] et la section III conclut notre étude. Nous calculons explicitement les situations d'équilibres dans les jeux locaux et analysons le système global par l'approche jeu évolutionniste en considérant les (re)transmissions sur le long terme.

A. Un exemple très célèbre : musique ou pas ?

En matière de musique, il arrive parfois que vous n'ayez pas les mêmes goûts que votre voisin pendant le weekend. Il lui arrive souvent d'écouter sa musique à fond. De même il vous arrive (en repréailles si vous continuez à cohabiter) de mettre votre musique à un volume plus que raisonnable. Ce qui a pour conséquences que le lendemain il recommence à nouveau. En dehors de ces périodes agitées, vous appréciez les périodes où aucun de vous ne gêne l'autre. Évaluons votre taux de satisfaction sur une grille : Vous avez une satisfaction de 7 à écouter votre musique à un volume important si votre voisin n'écoute pas la sienne. Votre satisfaction est de 1 lorsque votre voisin met sa musique à fond. Une soirée calme, sans musique vous apporte une satisfaction de 2. Le fait d'écouter simultanément votre musique mêlée à celle de votre voisin, donne une satisfaction de 0. Le tableau ci-dessous représente votre satisfaction en fonction du choix de votre voisin.

		votre voisin	
		écouter à fond	ne pas écouter
vous	écouter à fond	0	7
	ne pas écouter	1	2

Si vous savez ce que votre voisin a eu comme comportement les weekends précédents, que faites-vous ce weekend ?

B. Problème d'interférences : interaction non-réciproques

La figure 1 représente un environnement constitué de quatre mobiles m_1, m_2, m_3, m_4 (schématisés par des cercles) et trois récepteurs (carrés). La transmission d'un mobile se trouvant à distance moins de r d'un récepteur cause une interférence à la transmission de tout autre mobile se trouvant dans ce voisinage. Les interactions $m_1 - m_2$ et $m_3 - m_4$ ne sont pas *réciproques*

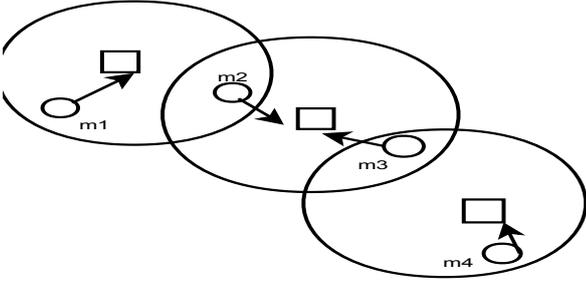


Fig. 1. (non-)réciprocité des interactions.

II. OUTILS D'ANALYSES

A. Jeu d'accès en un coup avec coût de regret

Considérons maintenant $n \geq 2$ mobiles (joueurs) en compétition pour avoir accès à un récepteur. Notons $N = \{1, \dots, n\}$ l'ensemble des joueurs. Chaque joueur a deux actions : transmettre T ou ne pas transmettre S . On note $A = \{T, S\}$. Il y a collision si au moins deux joueurs transmettent en même temps. Le coût d'une transmission est de c et le coût d'une collision est estimé de l'ordre de Δ . Il y a un coût de regret κ si aucun mobile ne transmet dans l'intervalle de temps donné. Ce coût évalue augmentation le délai d'attente pour réussir la transmission. La valeur d'une transmission réussie est de $v > \delta$. Ce jeu fait partie d'une grande classe de jeux appelée *jeux stables*.

Le tableau ci-dessous représente des interactions (réciproques ou non) entre deux mobiles $n = 2$.

$1 \setminus 2$	T	S
T	$(-\Delta - \delta, -\Delta - \delta)$	$(v - \delta, 0)$
S	$(0, v - \delta)$	$(-\kappa, -\kappa)$

La fonction de paiement¹ est définie par $u^i : A^n \rightarrow \mathbb{R}$ est donné

$$u^i(a) = \begin{cases} v - \delta & \text{si } a^i = T \text{ et } a^j = S \forall j \neq i \\ 0 & \text{si } a^i = S \text{ et } \{j \in N \mid a^j = T\} \geq 1 \\ -\Delta - \delta & \text{si } a^i = T \text{ et } \{j \in N \mid a^j = T\} \geq 2 \\ -\kappa & \text{si } a^j = S, \forall j \in N \end{cases}$$

A.1 Equilibre de Nash et optimalité au sens de Pareto

On appelle *stratégie pure*, une action dans A i.e T ou S . Une distribution de probabilité entre T et S est appelée *stratégie mixte*. Un équilibre de Nash est une situation où aucun des participants n'a intérêt à changer de stratégies de façon unilatérale. Le jeu d'accès a $2^n - 1$ équilibres de Nash décrit comme suit :

- Si un seul des participants transmet, on a un équilibre de Nash. Ainsi, il y a exactement n équilibres purs de Nash et tous ces équilibres purs sont optimal au sens de Pareto².
- k ($1 \leq k < n - 1$) des n participants choisissent de ne pas transmettre et les $n - k$ restant jouent la stratégie mixte optimale dans le jeu d'accès avec $n - k$ joueurs. Ce qui donne exactement $2^n - (n + 2)$ équilibres de Nash partiellement mixtes.

¹aussi appelée fonction d'utilité ou encore fitness

²On ne peut augmenter les satisfactions d'un des participants sans en diminuer celles d'un autre.

- un unique équilibre mixte de Nash obtenu lorsque la probabilité de ne pas transmettre est exactement $\left[\frac{\Delta + \delta}{v + \Delta + \kappa} \right]^{\frac{1}{n-1}}$.
- Les allocations de paiements partiellement ou complètement mixtes ne sont pas Pareto optimales.

A.2 Mécanismes de coordination

Un mécanisme de coordination entre les mobiles peut être introduit dans le jeu d'accès pour améliorer les performances des issues. Cela consiste à envoyer des signaux/messages (publics ou privés) à tous les utilisateurs présents dans le voisinage avant qu'ils ne prennent leurs décisions. Ces signaux ne changent pas l'ensemble des choix mais mettent une corrélation [1], [2] sur l'issue. Ils sont éventuellement gérés par un *arbitre* supposé neutre sur l'issue (ce rôle peut être joué par une station de base par exemple).

B. Jeux en réseau ou comment modéliser les interactions non-réciproques ?

Il arrive parfois que les voisinages s'intersectent et un mobile peut bien se trouver dans plusieurs voisinages de récepteurs différents. La position d'un tel mobile cause d'énormes problèmes d'interférences et provoquent des pertes de données. Ce type de problème peut être modélisé par un jeu à participation multiple : jeux en réseau. Un joueur peut alors participer simultanément à plusieurs jeux locaux : pas forcément avec les mêmes d'acteurs (le paiement étant nul si la destination du mobile n'est pas le récepteur concerné). Ainsi dans l'exemple de la figure 1,

- les mobiles $m1$ et $m2$ participent à une seule interaction binaire.
- les mobiles $m2$ et $m3$ participent chacun à une deux interactions binaires. Lorsque le mobile $m2$ décide de transmettre, il interfère avec $m1$ et $m3$. Dans l'interaction $m1 - m2$ le paiement de $m2$ est nul et $m2$ reçoit u^2 dans l'interaction $m2 - m3$ où

$$u^2 :$$

		choix de $m3$	
		transmettre	ne pas transm.
$m2$	transmettre	$-\Delta - \delta$	$v - \delta$
	ne pas transm.	0	$-\kappa$

C. Analyse du système global : approche jeu évolutionnaire

La théorie des jeux évolutionnaire s'est développée à la suite des travaux du biologiste John Maynard Smith [7], [8] dans les années 1970. Cette théorie est vue comme étant une application de la théorie des jeux classique (introduite par von-Neumann et Morgenstern, 1944 puis John Nash, 1950) dans des contextes biologiques. Depuis ce temps, la théorie des jeux évolutionnaire a connu un très grand intérêt notamment dans les sciences économiques et sociales [3].

Dans le domaine réseaux et télécommunications et en particulier les réseaux mobiles, la théorie des jeux évolutionnaire [8], [9] peut donner de nombreuses réponses à des questions particulières de l'évolution. Un grand nombre de spécialistes s'intéressent à cette théorie en espérant obtenir des outils – d'apprentissage, de mises à

jour des stratégies, détection-correction des erreurs, imitation, etc – permettant de compléter la théorie des jeux classique. En outre la théorie des jeux évolutionnaire, en tant que théorie explicitement dynamique [4], [6], fournit un élément important qui n'apparaît pas dans la théorie classique des jeux.

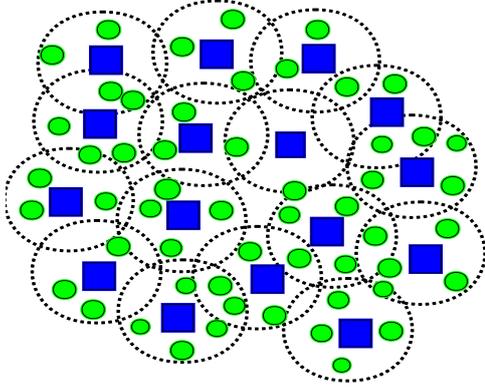


Fig. 2. Contrôle d'interférence.

Considérons un réseau (ou population) constitué d'un grand très nombre de noeuds dans un système du type ALOHA avec plusieurs interactions locales et simultanées entre voisins d'un récepteur (figure 2). Chaque jeu local est décrit sous forme stratégique comme précédemment. Compte tenu de la taille du système, un seul joueur a une influence négligeable sur toute la population et le système évolue selon une certaine dynamique [12], [9] de jeux basée sur les paiements espérés. Ces dynamiques caractérisent l'état du système à tout moment et offrent des algorithmes [6] de convergence (des mécanismes d'apprentissage) vers des compromis ou solutions. Une des notions de stabilité de stratégie proposée est la notion de *stratégie évolutionnairement stable* [7], [8]. Elle caractérise une situation où une fois que la population y est, est indétrônable par petite proportion de noeuds déviants (mutants). On montre que le jeu d'accès a une unique stratégie évolutionnairement stable donnée par $\left(1 - \left(\frac{\Delta+\delta}{v+\Delta+\kappa}\right)^{\frac{1}{n-1}}, \left(\frac{\Delta+\delta}{v+\Delta+\kappa}\right)^{\frac{1}{n-1}}\right)$. Par contre, on obtient une multitude d'équilibres corrélés évolutionnairement stables dans le jeu corrélé. Le modèle de jeu évolutionnaire avec plus de joueurs peut être utilisé pour intégrer la mobilité des noeuds dans les réseaux adhoc (sans infrastructures).

Soit $\Theta^{n-1} = \{b \in A^n \mid \text{numb}_T(b) = 1\}$ avec $\text{numb}_T(b)$ la fonction comptant le nombre d'actions T dans le profil de stratégies b . On définit la distribution de probabilité

$$Q^{n-1}(a) = \begin{cases} \frac{1}{|\Theta^{n-1}|} = \frac{1}{n} & \text{si } a \in \Theta^{n-1} \\ 0 & \text{ailleurs.} \end{cases}$$

Alors Q^{n-1} est une stratégie corrélée évolutionnairement stable qui garantit le paiement $\frac{v-\delta}{n} > 0$. A noter que le paiement d'équilibre évolutionnairement stable est négatif.

Supposons que le nombre d'opposants autour d'un récepteur choisit aléatoirement n'est pas connu et que

les mobiles changent de position d'un moment à l'autre. Désignons par K_t la variable aléatoire désignant le nombre de mobiles se trouvant dans le voisinage d'un mobile arbitraire [11]. Pour k entier naturel fixé, l'évènement $K_t = k$ est réalisé avec la probabilité $P(K_t = k)$. Le théorème ci-dessous donne un paiement d'équilibre corrélé évolutionnairement stable :

Théorème II-D. Soit $\Theta^k = \{b \in A^{k+1} \mid \text{numb}_T(b) = 1\}$ pour un certain k . On définit la distribution de probabilité sur les profils d'actions

$$Q^k(a) = \begin{cases} \frac{1}{|\Theta^k|} = \frac{1}{1+k} & \text{if } a \in \Theta^k \\ 0 & \text{otherwise} \end{cases}$$

Alors la collection des distributions de probabilités $Q = (Q^0, Q^1, Q^2, \dots)$ est une stratégie corrélée évolutionnairement stable qui garantit le paiement $\sum_{k \geq 0} P(K_t = k) \frac{v-\delta}{1+k} > 0$.

III. CONCLUSIONS

Nous avons proposé plusieurs approches du jeu d'accès au medium. Nous avons montré que le jeu statique entre n mobiles comporte $2^n - 1$ équilibres de Nash (purs ou mixtes). Le jeu de population (approche globale) a une unique stratégie évolutionnairement stable. Nous avons proposé des mécanismes de coordination entre les mobiles pour améliorer les performances du système (en particulier pour les interactions non-réciproques) et calculé des stratégies corrélées évolutionnairement stables. Nous terminons notre étude en faisant l'observation suivante : lorsque les fonctions de paiement ne sont pas instantanées c'est à dire lorsqu'une transmission prend un certain temps, le paiement des joueurs sont retardés et un équilibre peut être instable lorsque le délai est grand.

REMERCIEMENT

Nous remercions Julien CHAMP ainsi les quatre rapporteurs anonymes pour leurs suggestions

RÉFÉRENCES

- [1] R. Aumann. 1974, « Subjectivity and Correlation in Randomized Strategies », Journal of Mathematical Economics 1, 67-96.
- [2] R. Aumann. 1987, « Correlated Equilibrium as an Expression of Bayesian Rationality », Econometrica 55, 1-18.
- [3] D. Friedman, « On economic applications of evolutionary game theory », Journal of Evolutionary Economics, 8(1) :1543, 1998.
- [4] J. Hofbauer and K. Sigmund, « Evolutionary game dynamics », American Mathematical Society, Vol 40 No. 4, pp. 479-519, 2003.
- [5] J. Nash, « Equilibrium Points in N-person Games », dans Proceedings of the National Academy of Sciences (USA) (1950)
- [6] J. B. Rosen, « Existence and uniqueness of equilibrium points for concave N-person games », Econometrica 33, 520-534, 1965.
- [7] J. Maynard Smith and GR. Price. « The logic of animal conflict » Nature (1973) 246, 15-18.
- [8] Maynard Smith, J. (1982d), « Evolution and the Theory of Games » Cambridge University Press.
- [9] P. Taylor and L. Jonker, « Evolutionary stable strategies and game dynamics » Mathematical Biosciences, 16 :76-83, 1978.
- [10] H. Tembine, E. Altman, R. ElAzouzi, Y. Hayel, « Multiple Access Game in Ad Hoc Networks » Proc. of GameComm'2007, Nantes, October 2007.
- [11] H. Tembine, E. Altman, R. ElAzouzi, Y. Hayel, « Evolutionary games with random number of interacting players applied to access control » In Proc. of WiOpt'08, Berlin, 2008.
- [12] T. L. Vincent and T. L. S. Vincent, « Evolution and control system design » IEEE Control Systems Magazine, Vol 20 No. 5, pp 20-35, 2000.

Influence de l'architecture des plantes sur l'interception de la lumière, une approche multi-échelles

Da SILVA David[†], BOUDON Frédéric^{2,†}, GODIN Christophe^{2,†}, SINOQUET Hervé^{3,‡}

¹ CIRAD, Avenue Agropolis, 34398 Montpellier Cedex 5, France

² INRIA, 2004 route des lucioles BP 93, 06902 Sophia Antipolis, France

³ INRA, Site de Crouël, 234 avenue du Brézet, 63100 Clermont-Ferrand, France

[†] Virtual Plants, UMR DAP, TA A-96/02, 34398 Montpellier Cedex 5, France

[‡] UMR PIAF, Site de Crouël, 234 avenue du Brézet, 63100 Clermont-Ferrand, France

Résumé— Ce document présente un modèle d'interception de la lumière utilisant une description multi-échelles des plantes. Il permet en outre de faire différentes hypothèses pour l'estimation de l'interception de la lumière à chaque échelle. Sa modularité permet d'unifier différents modèles avec le même formalisme. La comparaison de différentes approches permet à la fois une analyse globale et échelle par échelle de l'influence de la structure d'une plante sur l'interception de la lumière. Une étude est réalisée en utilisant un ensemble de 8 plantes digitalisées en trois dimensions à l'échelle des feuilles.

Mots-clés— Multi-échelles, interception de la lumière, modélisation

I. INTRODUCTION

L'interception de la lumière est un processus essentiel pour les plantes, elle fournit l'énergie nécessaire à la photosynthèse et à la transpiration, facteur essentiel du transport d'eau et de nutriments [8]. Elle influe sur la température à l'intérieur de la canopée, facteur de maturation des fruits, mais également sur la température au niveau du sol et par conséquent sur la disponibilité en eau et l'activité biologique de la litière. L'interception directe de la lumière, i.e. quand les phénomènes de diffusion ne sont pas pris en compte, peut être défini simplement comme l'intersection entre la trajectoire des photons venant d'une direction particulière et les organes de la plante, en général les feuilles. Dans ce cadre, l'interception de la lumière ne dépend que de la distribution spatiale du feuillage, i.e. sa structure, et de sa géométrie, i.e. taille et orientation des feuilles. Il existe globalement deux types d'approches pour calculer ce type d'interception. Une première catégorie est celle du lancer de rayons ou Z-buffer [5] qui a l'avantage d'être très précise, mais nécessite une représentation 3D détaillée de la plante et ne permet pas d'appréhender l'influence des paramètres géométriques de structure sur l'interception de la lumière. Le second type d'approches consiste à assimiler la plante à un milieu *turbide*, i.e. un milieu constitué de particules de feuillage de taille infinitésimale réparties de manière aléatoire et homogène. Sous ces hypothèses, la probabilité, p_0 , pour un photon de traverser le feuillage sans être intercepté peut s'exprimer en utilisant la loi de *Beer-Lambert*

$$p_0 = \exp^{-G.LAD.L}, \quad (1)$$

où G est le coefficient d'extinction du feuillage pour une direction, Ω , donnée i.e. le rapport moyen entre la surface d'une feuille et sa surface projetée sur un plan orthogonal à la direction Ω . LAD (Leaf Area Density) est la densité de surface foliaire et L est la longueur du trajet du photon dans la couronne. La probabilité de zéro-interception, p_0 , correspond à la *porosité* du feuillage. La quantité de lumière interceptée est proportionnelle à $p = 1 - p_0$, i.e. la probabilité d'interception ou encore l'*opacité* du feuillage.

Malheureusement, les hypothèses du milieu *turbide* sont rarement vérifiées, en particulier dans le cas d'arbres isolés [18] [2]. L'introduction d'un paramètre de dispersion [11], μ , dans la loi de *Beer-Lambert* est un moyen simple de prendre en compte la distribution non-homogène du feuillage

$$p_0 = \exp^{-G.\mu.LAD.L}. \quad (2)$$

Ce paramètre est un indicateur de l'écart à une distribution homogène, il peut être vu comme un modificateur du LAD afin d'obtenir une canopée homogène ayant des propriétés optiques équivalentes. Un μ inférieur à 1 rend compte des phénomènes d'agrégation du feuillage diminuant sa capacité d'interception tandis qu'un μ supérieur à 1 exprime une plus grande efficacité d'interception, ce qui peut se traduire par une répartition plus régulière du feuillage. Finalement, μ est égal à 1 dans le cas d'une distribution aléatoire et homogène des feuilles. Ce paramètre reste encore empirique et il n'existe pas de relation claire entre μ et la structure du feuillage, néanmoins une relation entre μ et les variations spatiales du LAD a été mise en évidence dans [14]. L'objectif de ce travail est de montrer comment, à l'aide d'un modèle multi-échelles d'interception de la lumière, il est possible d'analyser l'influence de l'architecture d'une plante sur l'interception de la lumière et par conséquent d'améliorer la compréhension de la relation entre μ et la structure du feuillage.

II. APPROCHE MULTI-ÉCHELLES

Dans cette section, nous rappelons comment l'organisation multi-échelles d'une plante peut être définie puis utilisée dans un modèle d'interception de la lumière à base d'enveloppes poreuses. Enfin nous étendrons la définition de μ afin de l'utiliser comme un descripteur de l'organisation de la plante.

A. Description multi-échelle d'une plante

L'organisation multi-échelles d'une plante est décrite par un MTG (Multiscale Tree Graph)[6]. A chaque échelle i ($i = 1 \dots n$) la plante est considérée comme un ensemble de composants botaniques. Chacun des composants est un noeud de l'arbre (graphe) enraciné décrivant la topologie de la plante. Chaque noeud de l'échelle i est composé d'éléments de l'échelle $i + 1$. L'ensemble des noeuds de l'échelle i formant une partition des noeuds de l'échelle $i + 1$. La première échelle ($i = 1$) correspond à la plante entière, Fig. 1 a., tandis que la dernière ($i = n$) correspond à celle des feuilles Fig. 1 d. . Chaque noeud est associé à une représentation géométrique qui est l'enveloppe convexe des représentations géométriques de ses sous-composants Fig. 1 a., b. et c., à l'exception de la dernière échelle où les feuilles sont représentées par des polygones.

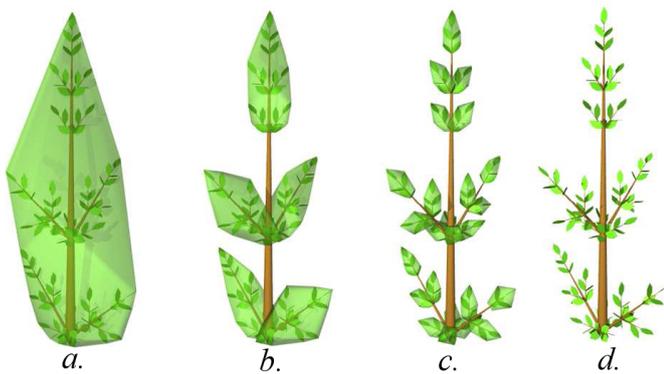


Fig. 1. Décomposition d'une plante virtuelle en 4 échelles pouvant représenter les échelles botaniques suivantes : plante entière (a.), branches charpentières (b.), pousses (c.) et feuilles (d.).

B. Modèle multi-échelles d'interception de la lumière

Le modèle utilisé en association avec la représentation multi-échelles d'une plante est basé sur l'utilisation d'enveloppe poreuse à l'instar de [7] et est présenté dans [3]. Dans ce modèle, l'opacité d'un composant pour une direction donnée, Ω , est définie comme étant fonction de 2 facteurs indépendants :

1. l'opacité de ses sous-composants,
2. la distribution spatiale de ses sous-composants, i.e. le degré de recouvrement de leurs projections selon la direction Ω .

De plus la longueur du trajet d'un photon dans un composant dépend de son point d'entrée dans l'enveloppe, par conséquent l'opacité d'un composant j à l'échelle i , $p_{i,j}$, est estimée en utilisant un ensemble de rayons, b , régulièrement espacés et suffisamment dense pour discrétiser l'espace de projection. L'opacité du composant étant la moyenne des opacités des rayons

$$p_{i,j} = \frac{1}{B_{i,j}} \sum_{b=1}^{B_{i,j}} p_{i,j}^b, \quad (3)$$

où $p_{i,j}^b$ est l'opacité du composant j à l'échelle i pour le rayon b et $B_{i,j}$ le cardinal de l'ensemble des rayons interceptés par ce composant. Pour l'opacité d'un rayon donné,

le premier facteur se traduit par

$$p_{i,j}^b = 1 - \prod_{k=1}^{n_j} [1 - p_{i+1,k}^b], \quad (4)$$

avec n_j représentant le nombre de sous-composants d'échelle $i + 1$ pour le composant j à l'échelle i . Le second facteur modifie l'expression de $p_{i+1,k}^b$ en fonction de l'option choisie pour la distribution spatiale des sous-composants, ces expressions sont détaillées dans [3]. A chaque échelle, le modèle permet de choisir entre 2 options pour estimer l'opacité d'un composant. La première (option A pour *actual*) prend en compte la position réelle des sous composants tandis que la seconde (option R pour *random*) utilise l'analogie au milieu *turbide* et considère une distribution aléatoire et homogène des sous-composants. Une chaîne de caractères de longueur $n - 1$ décrit la succession des options choisies définissant un scénario. Une échelle peut être négligée, elle est remplacée par un “-” dans la chaîne de caractères. Ainsi, si l'on considère une plante organisée en 3 échelles (plante entière, branches et feuilles), illustrée Fig.2, le scénario AA décrit la représentation réelle de la plante, le scénario RR décrit l'organisation où les feuilles sont distribuées de manière aléatoire dans les branches, elles-même distribuées aléatoirement dans la couronne [12]. Enfin le scénario -R signifie que l'on considère une distribution homogène et aléatoire des feuilles directement dans la couronne, i.e. milieu *turbide* à l'échelle de la plante utilisé dans plusieurs modèles [1], [9]. Ces 2 facteurs sont pris en compte

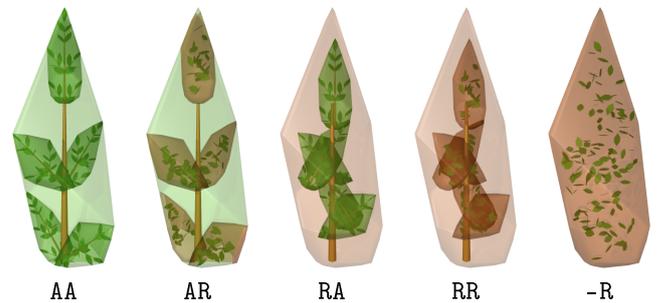


Fig. 2. Exemples de scénarios pour un objet composé de 3 échelles, couronne, branches et feuilles. Une chaîne de caractères de longueur 2 représente les options de distributions inter-échelles. AA décrit la position réelle des feuilles dans la couronne. Dans le scénario AR la vraie position des branches est utilisée mais l'opacité de leurs enveloppes est estimée en utilisant l'analogie au milieu *turbide* tandis qu'avec le scénario RA l'analogie au milieu *turbide* se fait à l'échelle des branches mais leurs opacités sont calculées en utilisant la vraie position des feuilles. Le scénario RR utilise l'analogie au milieu *turbide* à toutes les échelles tandis que le scénario -R prend en compte une distribution aléatoire et homogène des feuilles directement au sein de la couronne.

par l'estimation récursive de l'opacité qui s'arrête lorsque l'échelle la plus fine ($i = n$) est atteinte. Dans la suite, nous nous intéresserons uniquement à l'estimation de l'opacité à l'échelle de la plante entière mais pour différents scénarios afin d'étudier l'influence de la structure. La notation indiquée précisant le composant et son échelle sera remplacée par la notation faisant apparaître le scénario utilisé, ainsi, en reprenant l'exemple précédent, l'opacité *réelle* de la plante entière, i.e. scénario AA, sera notée $p_{<AA>}$ tandis que l'opacité pour l'hypothèse *turbide* à l'échelle de la plante sera notée $p_{<-R>}$.

C. μ comme un descripteur de structure

Originellement μ exprime l'écart entre la porosité réelle d'une couronne et celle estimée avec l'hypothèse *turbide* à l'échelle de la plante. Nous étendons ici cette définition pour exprimer l'écart d'une opacité, p , à une opacité de référence, $p_{<ref>}$. En utilisant l'analogie, introduite dans [15], avec le formalisme de la loi de *Beer-Lambert*, $p_0 = p_{0<ref>}^\mu$ et avec la relation de complémentarité entre la porosité et l'opacité, μ peut s'exprimer en fonction des opacités :

$$\mu = \frac{\log(1 - p)}{\log(1 - p_{<ref>})}. \quad (5)$$

En calculant μ à partir de scénarios spécifiques, il est possible d'analyser l'organisation spatiale du feuillage dans sa globalité ou échelle par échelle. En effet, en reprenant l'exemple illustré Fig. (2), le μ *original* est obtenu en comparant la situation AA à celle -R utilisée comme référence, un tel μ sera noté $\mu(\text{AA}/\text{-R})$. Dans le cas général, n échelles, il sera noté $\mu(\text{all-A}/\text{turbid})$, où **all-A** est le scénario de la représentation réelle de la plante et **turbid**, celui de la distribution homogène et aléatoire des feuilles directement dans la couronne. De manière analogue, le scénario où toutes les échelles sont considérées avec l'hypothèse *turbide* sera noté **all-R**. L'analyse échelle par échelle est réalisée en comparant 2 scénarios qui ne diffèrent que d'un caractère, celui de l'échelle étudiée. Ainsi, la distribution des branches est caractérisée par $\mu(\text{AA}/\text{RA})$. Par ailleurs, afin de mieux appréhender la relation entre le μ exprimant l'écart à une plante *turbide* et l'architecture d'une plante, nous utiliserons une décomposition particulière de $\mu(\text{all-A}/\text{turbid})$

$$\begin{aligned} \mu(\text{all-A}/\text{turbid}) &= \frac{\log(1 - p_{<AA>})}{\log(1 - p_{<-R>})} \\ &= \frac{\log(1 - p_{<all-A>})}{\log(1 - p_{<all-R>})} \cdot \frac{\log(1 - p_{<all-R>})}{\log(1 - p_{<turbid>})} \\ &= \mu(\text{all-A}/\text{all-R}) \cdot \mu(\text{all-R}/\text{turbid}). \end{aligned} \quad (6)$$

$\mu(\text{all-A}/\text{all-R})$ exprime l'écart global entre la réalité et l'hypothèse de milieu *turbide* à toutes les échelles et $\mu(\text{all-R}/\text{turbid})$ exprime l'écart dû au nombre d'échelles, dans le cas où il n'y a que 2 échelles, les scénarios **turbid** et **all-R** sont équivalents.

III. EXPÉRIMENTATION

Le modèle a été utilisé sur une base de données de 8 plantes digitalisées à l'échelle des feuilles. Les opacités à l'échelle de la plante ont été calculées pour les 46 directions représentant la discrétisation de la voute céleste proposé par [4]. Une valeur intégrée de l'opacité a ensuite été obtenue par pondération en utilisant les coefficients standard SOC (Standard OverCast) [10].

A. Plantes digitalisées

La base de données de plantes est composée de 4 pêchers de 4 ans digitalisés en mai 2001 à Balandran-CITFL (France) [16] et de 4 manguiers de 3 ans digitalisés en mars 2001 à Saint-Paul (La Réunion) [17]. La digitalisation 3D est effectuée grâce à un digitaliseur magnétique [13] composé d'un générateur de champ magnétique et d'un pointeur permettant d'enregistrer la position ainsi que les

angles d'azimuth, d'élévation et de rotation autour de son axe. Les manguiers sont organisés suivant 4 échelles botaniques : plante entière, branches charpentières, pousses annuelles et feuilles. Une échelle supplémentaire est présente chez les pêchers : les pousses d'un an entre les branches charpentières et les pousses annuelles.

B. Résultats

Les 2 premières lignes des résultats présentés Table I permettent d'analyser la structure globale de la plante en la comparant à 2 types de scénarios utilisant l'hypothèse simplificatrice du milieu *turbide*. Les lignes suivantes servent à l'analyse de la distribution, pour chaque échelle, de ses sous-composants, i.e. l'organisation des feuilles dans les pousses, des pousses dans les branches et ainsi de suite.

TABLE I
 μ GLOBAL ET ÉCHELLE PAR ÉCHELLE POUR LES PÊCHERS ET LES MANGUIERS.

	Peach1	Peach2	Peach3	Peach4
$\mu(\text{all-A}/\text{turbid})$	0.5482	0.6127	0.5734	0.6082
$\mu(\text{all-A}/\text{all-R})$	1.0368	1.2582	1.211	1.1599
$\mu(\text{all-A}/\text{AAAR})$	1.071	1.1445	1.1036	1.1331
$\mu(\text{all-A}/\text{AARA})$	0.9987	1.01	1.0096	1.021
$\mu(\text{all-A}/\text{AAA})$	0.9281	0.8971	0.9154	0.904
$\mu(\text{all-A}/\text{RAAA})$	1.0868	1.2071	1.1782	1.1257

	Mango1	Mango2	Mango3	Mango4
$\mu(\text{all-A}/\text{turbid})$	0.7039	0.7315	0.8112	0.7982
$\mu(\text{all-A}/\text{all-R})$	1.2157	1.2552	1.2809	1.3077
$\mu(\text{all-A}/\text{AAR})$	1.064	1.0886	1.0859	1.1019
$\mu(\text{all-A}/\text{ARA})$	0.9366	0.9391	0.9964	1.1240
$\mu(\text{all-A}/\text{RAA})$	1.2465	1.3346	1.2596	1.2163

Pour les pêchers et les manguiers, l'organisation réelle du feuillage est clairement agrégée par rapport à l'hypothèse d'une distribution homogène des feuilles dans la couronne, $\mu(\text{all-A}/\text{turbid}) < 1$. Cette agrégation est sensiblement plus prononcée pour les pêchers que pour les manguiers. A l'inverse, le scénario **all-R** où toutes les échelles sont considérées avec l'hypothèse *turbide* génère une agrégation supérieure à la réalité, $\mu(\text{all-A}/\text{all-R}) > 1$. Les deux espèces présentent également les mêmes tendances d'organisation régulière des feuilles dans les pousses et des branches charpentières dans les couronnes, $\mu(\text{all-A}/\text{AAAR})$ et $\mu(\text{all-A}/\text{AAR}) > 1$, et $\mu(\text{all-A}/\text{RAAA})$ et $\mu(\text{all-A}/\text{RAA}) > 1$, respectivement. L'organisation régulière des branches charpentières est plus nette pour les manguiers, la situation étant inversée dans le cas des feuilles. L'organisation des pousses quant à elle apparaît clairement aléatoire pour les pêchers et les manguiers avec une légère tendance à l'agrégation pour 2 de ces derniers, $\mu(\text{all-A}/\text{AARA})$ et $\mu(\text{all-A}/\text{ARA}) \simeq 1$. Finalement l'organisation des pousses d'un an au sein des branches charpentières chez les pêchers semble légèrement agrégative, $\mu(\text{all-A}/\text{ARAA}) < 1$. Afin de simuler la distribution régulière des branches charpentières, nous avons également décomposé $\mu(\text{all-A}/\text{turbid})$ en utilisant un scénario de type **all-R** en négligeant cette échelle, i.e. -RRR pour les pêchers et -RR pour les manguiers. L'utilisation de ces scénarios dans le calcul des μ

est noté par **all-R*** dans les résultats Table II.

TABLE II

DÉCOMPOSITION DE μ POUR LES PÊCHERS ET LES MANGUIERS AVEC PUIS SANS L'ÉCHELLE DES BRANCHES CHARPENTIÈRES (*).

	Peach1	Peach2	Peach3	Peach4
μ (all-A/turbid)	0.5482	0.6127	0.5734	0.6082
μ (all-A/all-R)	1.0368	1.2582	1.211	1.1599
μ (all-R/turbid)	0.5287	0.4870	0.4736	0.5243
μ (all-A/all-R*)	0.8117	0.9554	0.8835	0.8280
μ (all-R*/turbid)	0.6753	0.6413	0.6491	0.7345

	Mango1	Mango2	Mango3	Mango4
μ (all-A/turbid)	0.7039	0.7315	0.8112	0.7982
μ (all-A/all-R)	1.2157	1.2552	1.2809	1.3077
μ (all-R/turbid)	0.5791	0.5828	0.6333	0.6104
μ (all-A/all-R*)	0.8683	0.8577	0.9828	0.9825
μ (all-R*/turbid)	0.8107	0.8529	0.8253	0.8124

Pour les deux espèces, la distribution réelle du feuillage est plus régulière que celle simulée par le scénario all-R mais plus agrégée que celle du scénario all-R*, $\mu(\text{all-A/all-R}) > 1$ et $\mu(\text{all-A/all-R*}) < 1$.

IV. DISCUSSION

Un des principaux intérêts de cette étude est de permettre l'identification des scénarios faisant le meilleur compromis entre une description la plus simple possible, les mesures détaillées étant très coûteuses, et précision de l'estimation sans avoir à introduire un paramètre de dispersion, μ , qui ne peut pas être mesuré. Par ailleurs, ces résultats montrent que dans le cas des pêcheurs et des manguiers, si la distribution réelle du feuillage n'est pas parfaitement simulée par un scénario de type all-R, elle se trouve entre le cas prenant en compte les branches charpentières et celui les négligeant. L'utilisation du scénario all-R* permet une meilleure approximation, en effet $\mu(\text{all-A/all-R*})$ est plus proche de 1 que $\mu(\text{all-A/all-R})$. Ce qui signifie que l'écart entre la distribution réelle des feuilles et l'hypothèse d'une plante *turbide*, i.e. $\mu(\text{all-A/turbid})$, est en grande partie dû à l'organisation en échelles : lorsque la différence entre le feuillage réel et celui simulé par un modèle de type all-R est faible, e.g. $\mu(\text{all-A/all-R*}) \simeq 1$ dans Table II, alors $\mu(\text{all-A/turbid}) \simeq \mu(\text{all-R*/turbid})$. La structure agrégative d'un modèle de plante peut ainsi être modifiée par l'ajout ou la suppression d'échelles. La modularité en échelles du modèle présenté est la propriété principale qui permet d'unifier différents types de modèles avec un même formalisme et ainsi de les comparer aisément. Cette nouvelle approche permet une meilleure analyse de l'influence de l'architecture d'une plante sur l'interception de la lumière et ouvre les perspectives d'une meilleure compréhension de son fonctionnement et des stratégies de déploiement de ses organes.

V. REMERCIEMENTS

Nous souhaiterions remercier Gabriela Sonohat, Didier Combes et Vengadessan Kulandaivelu pour leur précieuse aide à la digitalisation des pêcheurs. La digitalisation des manguiers a été financée par les programmes AIP et PFI de l'INRA et ATP Manguiers du CIRAD.

RÉFÉRENCES

- [1] D. A. Charles-Edwards and J. H. M. Thornley. Light interception by an isolated plant. a simple model. *Ann Bot*, 37(4) :919–928, September 1973.
- [2] S. Cohen, P. Mosoni, and M. Meron. Canopy clumpiness and radiation penetration in a young hedgerow apple orchard. *Agricultural and Forest Meteorology*, 76(3-4) :185–200, October 1995.
- [3] D. Da Silva, F. Boudon, C. Godin, and H. Sinoquet. Multiscale framework for modeling and analyzing light interception by trees. *Multiscale Modeling and Simulation*, In Review.
- [4] J. A. Den Dulk. *The interpretation of Remote Sensing, a feasibility study*. PhD thesis, Wageningen university, 1989.
- [5] J. D. Foley, A. van Dam, S. K. Feiner, and J. F. Hughes. *Computer Graphics : Principles and Practice in C*. Addison-Wesley Professional, second edition, August 1995.
- [6] C. Godin and Y. Caraglio. A multiscale model of plant topological structures. *Journal of Theoretical Biology*, 191(1) :1–46, March 1998.
- [7] J. E. Jackson and J. W. Palmer. A simple model of light transmission and interception by discontinuous canopies. *Annals of Botany*, 44 :381–383, September 1979.
- [8] H. G. Jones. *Plants and Microclimate : A Quantitative Approach to Plant Physiology*. Cambridge University Press, 1992.
- [9] J. E. Mann, G. L. Curry, and P. J. H. Sharpe. Light interception by isolated plants. *Agricultural Meteorology*, 20(3) :205–214, June 1979.
- [10] P. Moon and D.E. Spencer. Illumination from a non-uniform sky. *Transactions of the Illumination Engineering Society*, (37), 1942.
- [11] T. Nilson. A theoretical analysis of the frequency of gaps in plant stands. *Agricultural Meteorology*, 8 :25–38, 1971.
- [12] P. Oker-Blom and S. Kellomaki. Effect of grouping of foliage on within-stand and within-crown light regimes : comparison of radiation and grouping canopy models. *Agricultural Meteorology*, 28 :143–155, 1983.
- [13] Polhemus Inc. Colchester, VT, USA. *3SPACE FASTRAK User's Manual, Revision F.*, 1993.
- [14] H. Sinoquet, G. Sonohat, J. Phattaralerphong, and C. Godin. Foliage randomness and light interception in 3-d digitized trees : an analysis from multiscale discretization of the canopy. *Plant, Cell & Environment*, 28(9) :1158–1170, September 2005.
- [15] H. Sinoquet, J. Stephan, G. Sonohat, P. E. Lauri, and P. Monney. Simple equations to estimate light interception by isolated trees from canopy structure features : assessment with three-dimensional digitized apple trees. *New Phytologist*, 175(1) :94–106, July 2007.
- [16] G. Sonohat, H. Sinoquet, V. Kulandaivelu, D. Combes, and F. Lescourret. Three-dimensional reconstruction of partially 3d-digitized peach tree canopies. *Tree Physiology*, 26 :337–351, 2006.
- [17] L. Urban, X. Le Roux, H. Sinoquet, S. Jaffuel, and M. Jannoyer. A biochemical model of photosynthesis for mango leaves : evidence for the effect of fruit on photosynthetic capacity of nearby leaves. *Tree Physiology*, 23 :289–300, 2003.
- [18] D. Whitehead, J. C. Grace, and M. J. Godfrey. Architectural distribution of foliage in individual pinus radiata d. don crowns and the effects of clumping on radiation interception. *Tree Physiology*, 7(1 2 3 4) :135–155, December 1990.

Approche probabiliste pour la gestion du dialogue Homme-Machine

Florian PINAULT¹ (Doctorant 1ère année),
Renato DE MORI², Fabrice LEFÈVRE³ (Directeurs de Thèse)

Laboratoire d'Informatique d'Avignon
339 Chemin de Meinajariès BP 1228
84 911 Avignon Cedex 9 - France

¹Florian.Pinault ²Renato.Demori ³Fabrice.Lefevre@univ-avignon.fr

Résumé— Le traitement automatique du langage couvre des applications variées, concernant la langue orale ou écrite. Le dialogue homme-machine est une de ces applications. Un gestionnaire de dialogue prend des décisions d'actions selon une stratégie et utilise une modélisation du dialogue évoluant selon des états de dialogue. Les modèles étudiés dans le cadre de ma thèse sont de nature probabiliste, pour permettre une gestion stochastique du dialogue.

Mots-clés— Langage, Langues, Dialogue Homme-Machine.

I. INTRODUCTION

Le langage peut être étudié à travers différentes approches. Nous le présentons ici sous le point de vue de l'Informatique Linguistique (*Computational Linguistic*) dont les objectifs sont pragmatiques plus que descriptifs : bien que les modèles développés puissent être utiles aux évolutions des théories des anthropologues, linguistes ou philosophes, le critère de validité retenu est celui des résultats obtenus aux campagnes d'évaluations sur des tâches précises.

La partie 2 donne une vue d'ensemble des recherches actuelles sur le traitement automatique du langage qui constitue le contexte de l'étude du dialogue homme-machine.

La partie 3 se focalise sur les phénomènes étudiés durant ma thèse. Elle présente le formalisme utilisé pour modéliser le dialogue puis montre son application pour une situation particulière (information touristique, réservation d'hôtels).

II. LE DIALOGUE HOMME-MACHINE DANS LE TRAITEMENT AUTOMATIQUE DES LANGUES

Nous allons dans cette section décrire comment situer l'étude du Dialogue Homme-Machine dans le domaine du Traitement Automatique de la Langue Naturelle (TALN).

A. Définitions

Le terme de *Langue Naturelle* réfère ici aux langues humaines (français, anglais, tagalog, langues des signes, etc.) par opposition aux langages informatiques ou mathématiques.

Ces derniers sont en particulier exempts d'ambiguïté et sont aisés à traiter par des machines car conçus dans

cette perspective. Les méthodes d'analyse (*parsing*) adaptées aux langages de programmation se sont révélées inapplicables pour les langages naturels, la résolution des références et des métaphores nécessitant d'utiliser des notions sémantiques et pragmatiques trop élaborées.

Le terme *Modèle* d'un objet, concept ou phénomène réfère à l'objet mathématique associé.

Par exemple un modèle possible de l'objet Terre est la sphère. Un texte peut-être modélisé par des n -grammes de mots.

B. Applications en langue écrite.

Les **correcteurs orthographiques** reposent sur des modélisations statistiques pour reconnaître les constructions grammaticales correctes ainsi que sur des règles implémentées explicitement.

La **fouille de texte**, ou **recherche d'information**, consiste à rechercher une information dans un ensemble de textes étendu. On distingue les moteurs de recherche généralistes tels que **google** des outils spécialisés développés pour des entreprises. Les rapports, factures, documentation, courrier de toutes les branches d'une entreprise constituent un ensemble de données dans lequel chercher une information. Les systèmes de **Questions-Réponses** sont une alternative à l'utilisation de mot-clés et visent à une compréhension plus profonde du sens des informations traitées.

Le **résumé automatique** et la **synthèse de documents** permettent la veille technologique en condensant une information disparate et constamment remise à jour.

Les outils de **traduction assistée** (de texte à texte), bien que loin des performances humaines, permettent de proposer des bribes de texte pré-traduites pertinentes et supprimer l'usage du dictionnaire.

C. Applications en langue orale.

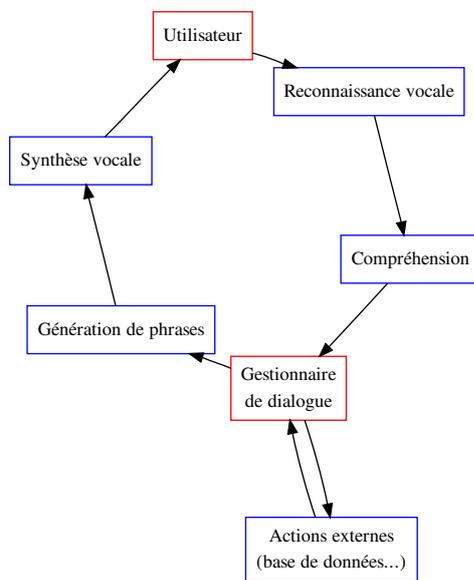
Les progrès en **reconnaissance** et **synthèse de la parole** permettent la dictée automatiques, voire de palier certains handicaps physiques. Les nouvelles technologies en font usage pour des fonctionnalités telles que les kit mains-libres ou la transcription audio de SMS.

Avec une modélisation sémantique adaptée, on cherche à présent à réaliser l'**indexation automatique de documents audio ou audio-visuels** (utilisable pour la recherche de vidéos sur internet comme pour les écoutes téléphoniques).

Les systèmes de **dialogue homme-machine** est efficace dans le cadre de standards automatiques, (la langue de l'interlocuteur peut-être détectée et sa demande traitée automatiquement), la réservation de billets d'avion par téléphone est à présent étudiée de façon approfondie et les gestionnaires actuels, bien qu'assez rigides, sont déjà utilisés dans des systèmes réels.

On trouvera un exposé plus détaillé de l'état de l'art dans [1] et [2].

D. Système de dialogue homme-machine



A chaque prise de parole de l'**Utilisateur**, le module de **Reconnaissance vocale** transforme le signal sonore en un texte, celui-ci est traité par le module de **Compréhension** pour en obtenir une représentation informatique.

Le **Gestionnaire de Dialogue** proprement dit traite cette information, choisit en conséquence la ou les **Actions** appropriées, et génère une représentation informatique d'une réplique.

Cette représentation est transformée en phrases (**Génération de phrases**), puis en signal acoustique (**Text-to-Speech**) retourné à l'**Utilisateur**.

III. SUJET DE THÈSE

A. Formalisme

Cette section présente une modélisation possible du dialogue et la façon dont elle sera utilisée dans ma thèse.

Les données d'entrée (observation o) fournies par le module de compréhension consistent en une représentation informatique de ce qu'a dit l'utilisateur. Le système utilise une variable interne pour représenter l'état du dialogue

(état s). Il possède une liste d'actions possibles de deux types :

- Actes de dialogue (questionner, accepter, refuser, donner/demander une information, *etc.*)
- Actes externes (accès à une base de données, envoi d'un courriel, commande d'un service, d'un engin, *etc.*)

Le comportement du système de dialogue est géré par une stratégie π qui lui permet de choisir une action a en fonction de l'état actuel du dialogue et des données d'entrée.

$$a = \pi(o, s)$$

Définir les états s , les actions a , ainsi que la façon de représenter les données d'entrée o sont le préalable indispensable pour modéliser le dialogue.

L'étape suivante consiste à comprendre la *dynamique du dialogue*, c'est-à-dire la façon dont le dialogue passe d'un état à un autre, en fonction des entrées et des actions choisies.

Il est ensuite possible de définir la notion de *stratégie optimale* de dialogue π .

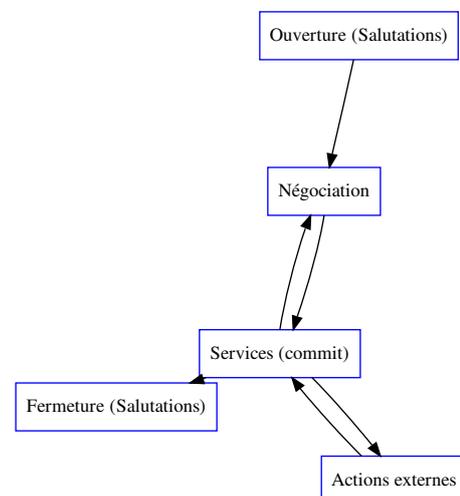
Les *paramètres* de ce modèle du dialogue peuvent être estimés statistiquement en fonction d'un corpus (un ensemble de dialogues dans une situation fixe) pour l'appliquer à une situation réelle.

Le processus d'estimation des paramètres ne pose pas de véritables problèmes calculatoires. En revanche, la recherche de stratégie optimale doit être résolue par des algorithmes (méthodes de calcul) très efficaces. En effet, l'exploration de l'arbre des stratégies possibles semble inévitable lorsqu'on rend les modèles probabilistes et donne lieu à des calculs de complexité NP (bien connu dans le domaine de l'optimisation d'algorithmes, les problèmes NP, non polynomial, sont décrits dans [6]).

B. Modélisation de la dynamique du dialogue

On peut modéliser un dialogue par un parcours aléatoire sur un graphe d'état.

Voici pour cette situation un modèle résumé de la succession des états de dialogue, le nombre d'états est réduit à quatre dans l'espace d'états résumé, mais chaque état représenté ci-dessous contient plusieurs états de dialogue :



La situation de dialogue considérée est la suivante : Après une phase d'*Ouverture*, une étape de *Négociation* (comportant des questions et des réponses) détermine quel service est demandé et, lorsque la requête de l'utilisateur est comprise, le service est fourni. Le dialogue peut ensuite évoluer vers une autre étape de *Négociation* de service, ou vers la phase de *Fermeture*.

Les transitions d'un état à un autre lors des prises de parole o de l'utilisateur sont représentées par les flèches et s'effectuent de manière probabiliste. Les probabilités affectées à chaque transition sont les paramètres du modèle et sont estimées statistiquement à l'aide d'un corpus de dialogue annoté.

Les actions a ne sont pas représentées ici par souci de simplicité.

Ce type de graphe modélise l'évolution du dialogue au cours du temps. L'intérêt d'étudier la dynamique du dialogue est multiple :

- permettre de prévoir le comportement de l'utilisateur et de choisir (automatiquement) une action.
- faciliter la construction de modèles d'utilisateur pour entraîner automatiquement les automates de dialogues.

C. Situation d'application réelle : Informations touristiques

Une situation de dialogue homme-machine réelle est présentée dans cette section.

woz :	...reservation d' hotel souhaitez vous ecouter le guide
spk :	non
woz :	quelle information voulez vous obtenir
spk :	je voudrais reserver un hotel
woz :	donnez le nom de la ville ou vous souhaitez aller
spk :	je voudrais aller a Evreux
woz :	a dreux a quelle periode souhaitez vous y séjourner
spk :	euh c' est pas Dreux c' est Evreux
woz :	a evreux a quelle periode souhaitez vous y séjourner
spk :	euh les quatre premiers jours de septembre
woz :	a evreux les quatre premiers jours de septembre dans quel quartier desirez vous séjourner
spk :	euh n' importe lequel de preference le centre ville
woz :	merci de patienter nous procedons a votre recherche a evreux dans le centre_ville deux hotels correspondent a votre demande confort hotel evreux et hotel mercure evreux voulez vous des details ou reserver dans l' un de ces hotels
spk :	je voudrais reserver dans l' hotel Mercure
woz :	combien de chambres souhaitez vous reserver

TABLE I

EXTRAIT D'UN DIALOGUE DU CORPUS MEDIA [4].

Le corpus des données utilisé est le corpus MEDIA. Les dialogues ont été enregistrés lors de la campagne d'évaluation EVALDA [4] en utilisant la technique du magicien d'Oz : un opérateur humain (woz) joue le rôle de la machine lors de l'interaction avec l'utilisateur (spk).

Le but du système de dialogue est de fournir des informations touristiques et de permettre à l'utilisateur d'effectuer des réservations d'hôtels.

Chaque tour de parole, ou réplique, est transformé en concepts élémentaires comme l'illustre l'exemple suivant :

Transcription texte :	euh les quatre premiers jours de septembre
Transcription en concepts :	
c_1 = nombre-temps(valeur = 4)	
c_2 = temps-axeTps(valeur = premier)	
c_3 = temps-unite(valeur = jour)	
c_4 = temps-mois(valeur = 9)	

TABLE II

EXEMPLE DE TRANSCRIPTION EN CONCEPTS D'UNE RÉPLIQUE DE L'UTILISATEUR

Les données d'entrée o sont les concepts : $o = (c_1, \dots, c_n)$.

Les états du dialogue s sont nombreux et ne sont pas définis explicitement mais sont déterminés selon des algorithmes de regroupement (*clustering*) ou des réseaux de neurones.

IV. CONCLUSION

L'étude statistique du langage donne lieu à de nombreuses applications. L'une d'entre elles est la conception de systèmes de dialogue Homme-Machine. Les modèles rigides conçus jusqu'à aujourd'hui ne donnaient pas la possibilité à l'utilisateur de s'écarter du schéma de communication prédéfini. L'utilisation de modèles probabilistes comportant un grand nombre d'états définis d'après des dialogues réels permettra sans doute une plus grande souplesse d'utilisation.

RÉFÉRENCES

- [1] G. Sabah *Compréhension des langues et interaction*. Hermès-Sciences Lavoisier, 2006.
- [2] R. De Mori *Spoken Dialogue With Computers*. Academic Press, 1998.
- [3] Technolanguae. *Le portail des technologies de la langue*. [en ligne]. [Page consultée le 16 février 2008]. <<http://www.technolanguae.net>>.
- [4] L. Devillers, H. Maynard, S. Rosset, P. Paroubek, J.Y. Antoine, F. Béchet, C. Bousquet, O. Bontron, L. Charnay, K. Choukri, J. Gouliani, K. McTait, L. Romary, M. Vergnes, N. Vigouroux, J. Villaneau. *The French MEDIA/EVALDA project : the evaluation of the understanding capability of Spoken Language Dialogue Systems*
- [5] H. Maynard, H. Rosset. *A Semantic representation for spoken dialogs*, EUROSPEECH 2003.
- [6] C. Papadimitriou. *Computational Complexity, 1st edition, Addison Wesley*. ISBN 0-201-53082-1. Chapter 9 : NP-complete problems, pp.181-218. 1994.

AMELIORER LE RENDEMENT GRACE AUX STRUCTURES TOLERANTES AUX FAUTES

JULIEN VIAL (DOCTORANT 2E ANNEE), ALBERTO BOSIO, PATRICK GIRARD, CHRISTIAN LANDRAULT, SERGE PRAVOSSODOVITCH, ARNAUD VIRAZEL

Laboratoire d'Informatique, de Robotique et de Microélectronique de Montpellier
161 rue Ada, 34392 Montpellier Cedex 5, France

{ VIAL, BOSIO, GIRARD, LANDRAUL, PRAVO, VIRAZEL }@LIRMM.FR
HTTP://WWW.LIRMM.FR/~W3MIC/

Résumé— Les technologies nanométriques, malgré les avantages liés à leur miniaturisation, comportent un certain nombre d'inconvénients liés à leur processus de fabrication. Pour résoudre ces inconvénients, les structures tolérantes aux fautes pourraient être utilisées, dans le futur, pour augmenter le rendement en tolérant certains défauts de fabrication. Ces structures seraient alors détournées de leur objectif actuel qui est de tolérer des défauts en ligne (pendant le fonctionnement du circuit). Dans cet article, l'architecture des circuits est modifiée pour obtenir une structure tolérante aux fautes. Cette structure est ensuite étudiée et testée pour savoir dans quelles mesures, elle permet d'augmenter le rendement de fabrication.

Mots clés — Architectures tolérantes aux fautes ; rendement ; test

I. INTRODUCTION.

Les technologies nanométriques permettent d'envisager la réalisation de structures numériques de taille considérable regroupant plusieurs centaines de millions de transistors sur une même puce. Cette miniaturisation croissante de l'électronique rend les procédés de fabrication de plus en plus complexes et de moins en moins fiables. Les imperfections des processus de fabrication engendrent des défauts de fabrication et il est de plus en plus difficile de fabriquer un circuit sans aucun défaut.

Les circuits tolérants aux fautes peuvent être une solution pour augmenter le rendement. Ainsi, pour tolérer les défauts de fabrication, on envisage d'avoir recours à des structures tolérantes aux fautes qui permettront, au détriment d'une surface silicium plus grande, de s'affranchir de certains défauts et donc de « remonter » le rendement effectif.

Dans les mémoires, on utilise déjà la tolérance aux défauts de fabrication. Des cellules supplémentaires sont présentes et permettent de remplacer des cellules que le test a révélé défectueuses. Cependant, les techniques utilisées dans des structures régulières comme les mémoires ne peuvent pas servir pour des structures numériques composées de divers blocs logiques.

La structure tolérante aux fautes étudiée dans cet article est la structure TMR (« Triple Modular Redundancy »). Elle est basée sur la redondance matérielle et est composée de 3

modules identiques M1, M2 et M3 (Figure 1). Les sorties de ces modules sont ensuite comparées et la sortie finale du système est déterminée par un vote de majorité (V).

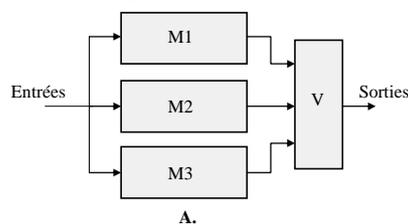


Fig. 1. Architecture d'une structure TMR

Dans ce type de structure, un défaut de fabrication présent dans un module sera toléré. En effet, si ce défaut propage une valeur erronée en sortie de son module, la présence du voteur permettra de masquer le module défectueux. De manière générale, les sorties erronées d'un module, à cause d'un ou de plusieurs défauts de fabrication, sont masquées grâce à la présence du voteur et des deux autres modules non fautifs. Ainsi, la structure TMR permet de tolérer au moins un défaut présent dans un des trois modules [2]. Par contre, si un défaut de fabrication est présent dans le voteur, la structure TMR peut être fautive. Pour éviter ce type de problème, le voteur peut être réalisé avec des techniques de conception plus robustes [3].

II. ETUDE DES STRUCTURES TMR.

Dans la suite de cette partie, le voteur est supposé sans défaut. En conséquence, les défauts de fabrications touchent uniquement les modules de la structure TMR.

II.1. Combien de défauts peut tolérer la structure TMR ?

Comme vu précédemment, la structure TMR peut tolérer un défaut mais en pratique peut en tolérer plus. Par exemple, en présence de deux défauts, la structure TMR peut fonctionner correctement en fonction de la nature et l'emplacement de ces défauts. En effet, deux défauts peuvent être tolérés s'ils ne sont pas observables en sortie du voteur. Pour être observable, deux défauts de fabrication doivent se trouver dans deux modules différents et doivent

propager une erreur jusqu'à deux sorties de module identiques. La Figure 2 montre deux exemples (le voteur n'a pas été représenté).

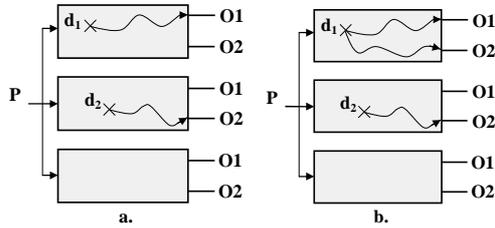


Fig. 2. Observabilité de deux défauts de fabrication
a) non observables b) observables

Dans les deux cas, le même vecteur P est appliqué en entrée des trois modules. Chaque module a deux sorties O1 et O2. Si une de ces sorties est erronée plus d'une fois, alors les défauts sont observables en sortie de la structure TMR.

Par exemple dans le cas présenté sur la Figure 2.a, les sorties O1 et O2 sont chacune fautive une fois et correctes deux fois. Le vote de majorité permet de masquer les sorties de modules erronées. Les défauts ne sont donc pas observables. Dans le cas présenté sur la Figure 2.b, la sortie O2 est fautive deux fois : dans le premier et le deuxième module. Le voteur choisit alors pour O2 la valeur erronée présente à la sortie des modules 1 et 2. Les défauts sont alors observables.

Lorsqu'il y a plus de deux défauts de fabrications, ces défauts peuvent être regroupés en couple de défauts. Par exemple, trois défauts de fabrication (d1, d2, d3) peuvent être regroupés en trois couples de défauts (d1, d2), (d1, d3), (d2, d3). Si nous considérons qu'il n'y pas de phénomènes de masquage entre les différents couples de défauts, chaque couple de défauts peut être étudié individuellement. Ainsi, grâce à cette hypothèse, les défauts de fabrications ne sont pas observables et ainsi tolérés si tous les couples de défauts ne sont pas observables.

II.2. Amélioration du rendement

La fabrication de structures TMR est intéressante si, pour une même surface de silicium, il y a plus de structures TMR qui fonctionnent que de circuits sans redondance qui fonctionnent. Pour les calculs qui suivent, on considère la taille du voteur comme nulle (négligeable devant la taille des trois modules). Par conséquent, les structures TMR prennent trois fois plus de places que les structures sans redondance. Sur une surface de silicium S, on peut donc fabriquer :

- soit N structures TMR avec un rendement η_{TMR} ($\eta_{TMR} \times N$ fonctionneront)
- soit 3N structures sans redondance (non tripliquées) avec un rendement η_c ($\eta_c \times 3N$ fonctionneront)

Pour qu'il y ait intérêt à fabriquer des TMR, il faut que :

$$\eta_{TMR} \times N > \eta_c \times 3N \Leftrightarrow \eta_{TMR} > 3 \times \eta_c$$

η_c peut être calculé à partir de la loi de Poisson en fonction d'un nombre moyen de défauts λ . η_{TMR} peut être

calculé toujours avec la loi de Poisson en fonction d'un nombre moyen de défaut 3λ (la structure étant trois fois plus grande) et de la probabilité R qu'une paire de défauts ne soit pas observable. Après calculs, on arrive à exprimer η_{TMR} en fonction de η_c .

$$\eta_{TMR} = e^{-3\ln \eta_c} \times \left(1 - 3\ln \eta_c + \sum_{i=2}^{\infty} R^{C_i^2} \times \frac{(-3\ln \eta_c)^i}{i!} \right)$$

Pour compenser la surface perdue en fabriquant des structures TMR, le rendement doit être trois fois supérieur par rapport au rendement de fabrication du circuit sans redondance. La Figure 3 montre η_{TMR} en fonction de η_c pour différentes valeurs de R. La ligne en pointillé représente la condition pour laquelle $\eta_{TMR} > 3\eta_c$. D'après cette figure, il apparaît que cette condition peut être vraie pour des valeurs de R supérieures à 93%.

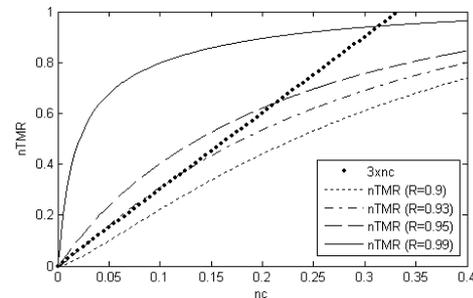


Fig. 3. Cas où le rendement est amélioré en utilisant des structures TMR

Pour résumer, si le pourcentage de couples de défauts non observables est supérieur à 93%, alors pour certains rendements η_c , il est plus intéressant de fabriquer des structures TMR. Pour connaître ce que représentent en pratique 93% de couples de défauts non observables, nous avons étudié et calculé le pourcentage R pour les circuits ISCAS'85 et ITC'99.

III. TEST DES STRUCTURES TMR.

Le test d'un circuit consiste à mettre en évidence son éventuel mauvais fonctionnement. Nous avons réalisé un test structurel qui consiste à considérer chaque ligne du circuit pouvant être collée à '1' ou '0'. L'objectif du test est donc de trouver des mots à appliquer en entrée (vecteur de test) pour pouvoir visualiser en sortie ces fautes de collage. On utilise un ATPG (Automatic Test Pattern Generator) qui permet de trouver des vecteurs de test permettant de vérifier l'absence d'un type de faute dans le circuit. Le modèle de faute que l'on veut utiliser est la paire de faute. Comme aucun ATPG n'existe avec ce type de faute, nous avons adapté un ATPG existant.

III.1. Les paires de fautes

Il faut au moins deux fautes pour que ces dernières puissent mettre en défaut la structure et être observable en sortie. C'est le cas d'une paire de fautes qui est constituée de deux fautes de collage.

III.2. Caractéristiques

Pour être observable dans une structure TMR, les deux fautes de collage constituant une paire de fautes doivent être présentes dans deux modules différents et se propager vers au moins une sortie de module commune.

Si chaque module contient n fautes de collage, alors le nombre total de fautes dans les trois modules est $3n$. Le nombre total Ψ de paires de fautes est alors égal à :

$$\Psi = C_{3n}^2 = \frac{3n!}{(3n-2)! \times 2!} = \frac{9n^2 - 3n}{2}$$

IV. ETUDE DES CIRCUITS ISCAS'85 et ITC'99.

Les circuits ISCAS'85 et ITC'99 sont dans un premier temps transformés en structure TMR. Pour savoir si fabriquer des structures TMR pour tolérer des défauts de fabrication a un intérêt, il faut que le pourcentage de paires de fautes non observables soit supérieur à 93%. Le test de chaque circuit permet de connaître ce pourcentage. Un ATPG avec un modèle de paires de fautes a été lancé pour chaque circuit. L'ATPG permet de connaître les paires de fautes détectables (observables) avec le vecteur de test mettant la paire de faute en évidence. L'ATPG permet aussi de connaître le nombre de paires de fautes non observables.

Le Tableau 1 présente les résultats de l'ATPG. La première colonne liste les circuits testés. La deuxième colonne désigne le nombre de paires de fautes pour chaque structure TMR. L'ATPG permet de connaître le pourcentage de fautes indétectables (colonne 3), le pourcentage de fautes détectées (colonne 4) ainsi que le nombre de vecteurs de test (colonne 5). On remarque que la somme des pourcentages des fautes indétectables et des fautes détectées n'est pas toujours égale à 100%. Cela signifie que l'ATPG n'a pas toujours réussi à déterminer la non testabilité d'une faute ou le vecteur de test permettant de détecter la faute.

Circuits	# de paires de fautes	Après l'ATPG		
		% paires de fautes indetectables	% paires de fautes détectées	#vecteurs de test
C2670	36.6M	75.95%	17.51%	4987
C3540	53.2.M	54.09%	23.75%	1756
C5315	111M	93.20%	3.75%	642
C6288	176M	38.02%	54.79%	4532
C7552	249M	84.92%	11.71%	8745
B04	9.81M	84.32%	12.70%	3390
B05	29.5M	88.65%	8.98%	4699
B11	7.7M	74.50%	21.16%	1551
B12	34.7M	95.45%	4.09%	1274
B13	3.14k	96.94%	3.06%	167

Tableau 1. Résultats de l'ATPG

Le pourcentage de paires de fautes est supérieur à 93% pour trois circuits (c3540, b12, b13).

Nous avons vu précédemment que modifier l'architecture d'un circuit en le transformant en une structure TMR est intéressant si $\eta_{TMR} > 3\eta_c$. Sur la Figure 4 est représenté $\eta_{TMR} - 3\eta_c$. Si $\eta_{TMR} - 3\eta_c > 0$ alors transformer un circuit en

TMR est une bonne solution pour tolérer des défauts de fabrication et ainsi augmenter le rendement.

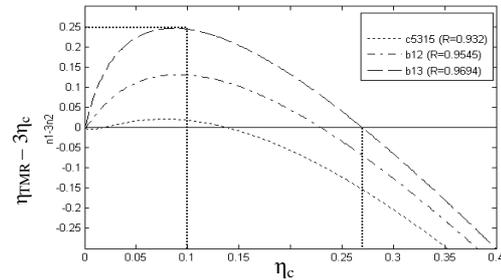


Fig. 4. Intérêts de fabriquer des structures TMR

Par exemple, considérons la courbe du circuit b13 sur la Figure 4. Si la fabrication de circuits b13 se fait avec un rendement $\eta_c < 27\%$, alors il est plus intéressant de tripler le circuit pour former une structure TMR. Si $\eta_c = 10\%$ alors $\eta_{TMR} - 3\eta_c = 0,25$ donc $\eta_{TMR} = 0,55$. Dans ce cas, pour une même surface de silicium, on fabrique $\frac{\eta_{TMR} - 3\eta_c}{3\eta_c} = 83,3\%$ plus de structures TMR qui fonctionnent que de circuits b13 qui fonctionnent.

V. CONCLUSIONS.

Cet article traite de l'utilisation de structures tolérantes aux fautes pour augmenter le rendement. De nos jours, les problèmes de fabrication impliquent des rendements de plus en plus faible. En conséquence, les structures tolérantes aux fautes pourront être utilisées et permettront d'accepter des circuits qui auront un fonctionnement correct malgré la présence de défauts de fabrication.

Nous avons étudié les structures TMR et nous avons calculé dans quelles conditions et sous quel rendement, transformer l'architecture en TMR devient intéressant. Ainsi, malgré une surface de silicium trois fois plus importante et malgré un nombre moyen de défauts de fabrication trois fois plus grand, une structure TMR permet d'augmenter le rendement de fabrication pour certains circuits étudiés.

Compte tenu des résultats, les structures TMR ne sont pas toujours appropriées pour tolérer des défauts de fabrications. Lors de travaux futurs, d'autres types de tolérance seront étudiées avec des structures moins coûteuses en silicium ou/et qui tolèrent plus de défauts.

REFERENCES

- [1] I. Koren and C. M. Krishna, "Fault Tolerant Systems", Ed. Morgan Kaufmann, 2007
- [2] L. Fang and M. S. Hsiao, "Bilateral Testing of Nano-scale Fault Tolerant Circuits", IEEE Int. Symp. on Defect and Fault-Tolerance in VLSI, pp.309-317, 2006
- [3] J. M. Cazeaux, D. Rossi and C. Metra, "New High Speed CMOS Self-Checking Voter", Proc. of IEEE Int. On-Line Testing Symposium pp.58-63, 2004.

Amplificateur Faible-Bruit Dédié à l'Enregistrement d'ENG à partir d'une Electrode Cuff Hexagonale

Lionel Gouyet, Guy Cathébras, Serge Bernard, Fabien Soulier, David Guiraud
and Yves Bertrand

LIRMM, Université Montpellier II - CNRS - INRIA, 161 rue Ada, 34392 Montpellier, France
Email: lastname@lirmm.fr

Résumé - En cas de dysfonctionnement du système nerveux, la stimulation électrique fonctionnelle qui consiste à provoquer des contractions musculaires artificielles par stimulations électriques, a prouvé son efficacité dans l'aide à la restauration de fonctions neurales. Les informations sensorielles provenant de capteurs naturels et se propageant sur les fibres nerveuses afférentes, pourraient être utilisées comme informations de retour pour un système de contrôle en boucle fermée de stimulation électrique fonctionnelle. Dans le but de faire l'acquisition de ces informations, nous proposons une nouvelle structure d'électrode cuff multipolaire ainsi qu'un système d'amplification adapté.

I. INTRODUCTION

Le système nerveux étant le centre de régulation et de communication de l'organisme, les pathologies empêchant son fonctionnement optimal entraînent des troubles tels que l'hémiplégie ou la paraplégie. La technique de Stimulation Electrique Fonctionnelle (SEF) consistant à compléter ou remplacer le système nerveux en stimulant électriquement les nerfs ou muscles est l'une des solutions les plus probantes à ces pathologies. L'intégration d'informations neurales (force, contact, etc.) fournies par les biocapteurs naturels permettrait d'obtenir un système de stimulation contrôlé en boucle fermée et donc, plus efficace. Ces informations sensorielles sont propagées sur les fibres nerveuses afférentes et sont appelées « potentiels d'action » (PA : vus comme les différences de potentiels entre l'intérieur et l'extérieur de la membrane d'un axone). Les signaux nerveux enregistrés sont très nombreux et de très faible amplitude comparativement à celles des signaux parasites d'origine musculaire (électromyogrammes EMG). Ces signaux ont en effet un ordre 1000 fois plus grand que celui des signaux

nerveux (électroneurogrammes ENG). Ainsi, en plus de pouvoir discriminer les informations sensorielles, le dispositif d'acquisition de signaux nerveux doit être doué d'une bonne réjection des signaux parasites.

Nous débutons ce papier en discutant des avantages apportés par l'électrode cuff hexagonale, électrode ayant déjà été présentée dans [1]. Nous développons ensuite la structure d'un amplificateur adapté à la structure de notre électrode. Il est important de souligner ici que la bande de fréquence des signaux que l'on cherche à mesurer est définie à partir du même modèle, déjà utilisé pour évaluer les sensibilités externes et internes de notre électrode. Ce modèle considère un axone myélinisé de 10µm de diamètre avec des nœuds de Ranvier espacés de 1mm. Finalement, nous donnerons nos remarques conclusives ainsi que nos perspectives.

II. ELECTRODE CUFF HEXAGONALE

Les électrodes cuff ont été les plus employées durant ces dix dernières années [2]-[3]. Elles présentent l'avantage d'être facilement implantable, de ne pas être invasive pour le nerf et la stabilité de leur implantation offre la possibilité d'expérimenter en chronique. Ces électrodes enregistrent les ENG comme la différence de potentiels créée sur les pôles par les charges associées aux potentiels d'action se propageant sur les fibres nerveuses. La plus répandue de ces électrodes, l'électrode cuff tripolaire, est constituée de trois anneaux séparés d'une distance égale (fig1.a). Avec ce type d'électrode, une méthode classique pour rejeter les signaux parasites consiste à calculer la moyenne des différences de potentiels entre le pôle central et les autres pôles extérieurs [4]-[5]:

$$V_{rec} = \frac{(V_0 - V_1) + (V_0 - V_2)}{2} = V_0 - \frac{V_1 + V_2}{2} \quad (1)$$

Afin d'obtenir des mesures plus localisées dans le nerf, nous avons proposé une structure comportant de nombreux pôles dans une configuration à motifs hexagonaux (fig1.b : 42 pôles) [1].

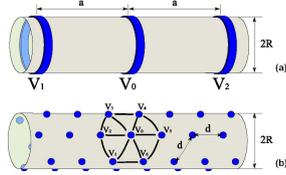


Figure 1. Electrode cuff tripolaire (a) et hexagonale (b)

Sur chacun des motifs, nous appliquons la technique de réjection des signaux parasites présentée plus haut en calculant la moyenne des différences de potentiels entre le pôle central et chacun des pôles extérieurs :

$$V_{rec} = \frac{1}{6} \sum_{i=1}^6 (V_0 - V_i) = V_0 - \sum_{i=1}^6 \frac{V_i}{6} \quad (2)$$

Pour évaluer les performances de cette électrode, nous avons utilisé un modèle du champ électrique extracellulaire créé par un potentiel d'action et nous invitons le lecteur à se reporter à [1] afin d'approfondir les calculs ayant fourni ce modèle ainsi que les figures illustrant nos résultats théoriques. Ces résultats montrent cependant que l'électrode cuff hexagonale offre une meilleure sensibilité, une meilleure sélectivité spatiale ainsi qu'une meilleure réjection des signaux parasites qu'une électrode cuff tripolaire classique. Une électrode cuff hexagonale a été fabriquée (Cf. figure 2) pour pouvoir expérimenter nos recherches. Un amplificateur a été développé afin de réaliser la fonction de l'équation (2), et nous vous le présentons dans la section suivante.

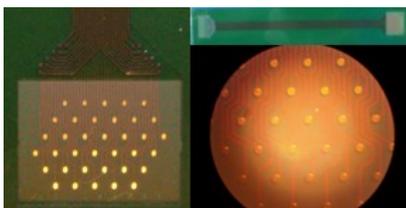


Figure 2. Photographie de l'électrode cuff hexagonale

III. AMPLIFICATEUR MOYENNEUR.

Afin de réaliser la fonction de l'équation (2), nous avons dimensionné un système d'amplification faible-bruit capable d'effectuer la différence entre l'enregistrement du pôle central d'un motif hexagonal et la moyenne des enregistrements de ses six pôles voisins. Un préamplificateur intègre cette fonction au niveau transistor, évitant ainsi une structure classique à bases de résistances et d'AOP qui rendraient difficile l'optimisation en surface, en bruit et en consommation. Ce préamplificateur, destiné essentiellement à améliorer le rapport signal sur bruit des ENG tout en traitant le signal afin de rejeter les signaux parasites, est suivi d'un amplificateur d'instrumentation améliorant le gain du système.

A. Préamplificateur

Le préamplificateur est réalisé à partir d'une paire différentielle classique dans laquelle nous avons « coupé » l'entrée inverseuse en six transistors six fois plus petits (Cf. figure 3). Cette astuce nous permet de dimensionner une structure à moins de 20 transistors tandis que nous aurions besoin d'au moins 8 AOP et de résistances pour la même fonctionnalité. En utilisant le modèle « petits-signal » du transistor, nous pouvons facilement montrer que :

$$V_{out1} - V_{out2} = \frac{g_{m_p}}{g_{m_h}} \sum_{i=1}^6 (V_0 - V_i) \quad (3)$$

Soit la fonctionnalité recherchée.

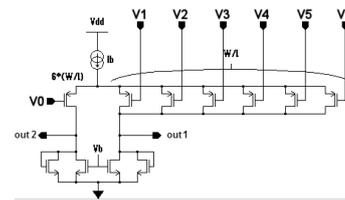


Figure 3. Schématique du préamplificateur

Les signaux intéressants sont de très faible amplitude (typiquement 1-5µV), Nous avons donc porté attention à l'optimisation en bruit de notre structure. Pour cela, afin d'évaluer la bande de fréquence de nos mesures, nous avons considérés les signaux provenant de deux nœuds de Ranvier successifs sur un même axone. Etant donné l'écart de ces nœuds de Ranvier (1mm) et la vitesse d'un potentiel d'action sur un

axone myélinisé (inférieure ou égale à 100m.s^{-1}), le délai entre ces signaux est supérieur à $10\mu\text{s}$. Pour enregistrer correctement ce passage, nous avons besoin d'une bande passante de 50kHz . Les vitesses des potentiels d'action pouvant être très inférieures à 100m.s^{-1} , les signaux basse fréquence en résultant risquent davantage de disparaître dans le bruit de flicker que dans le bruit thermique de la structure. Typiquement, les transistors de type P ont une moins grande contribution en bruit de flicker que les transistors de type N et leur utilisation en entrée d'une structure d'amplification permet d'optimiser le slew-rate et la fréquence gain-unité [6]. Ainsi, les entrées du préamplificateur sont réalisées à partir de transistors de type P afin d'optimiser les performances en bruit du système.

1) Simulations

Afin de simuler notre circuit, nous avons choisi d'exprimer les sept signaux d'entrées du préamplificateur comme les combinaisons de sept composantes fondamentales $U_d, U_c, U_{d1}, U_{d2}, U_{d3}, U_{d4}$ et U_{d5} . U_d et U_c sont respectivement les entrées différentielles et de mode commun. Les tensions U_{d1} à U_{d5} sont les composantes différentielles caractérisant la répartition des tensions entre les entrées de V_1 à V_6 . Les équations (4) et (5) définissent la transformations liant les U_{di} aux signaux V_i .

$$U_m = U_c - U_d \quad (6)$$

$$\begin{pmatrix} U_m \\ U_{d1} \\ U_{d2} \\ U_{d3} \\ U_{d4} \\ U_{d5} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & -1 & -1 & -1 \\ 1 & -1 & 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 & 1 & -1 \\ 1 & -1 & -1 & -1 & 1 & 1 \\ 1 & 1 & -1 & -1 & -1 & 1 \end{pmatrix} \begin{pmatrix} V_1 \\ V_2 \\ V_3 \\ V_4 \\ V_5 \\ V_6 \end{pmatrix} \quad (7)$$

En utilisant cette transformation, nous pouvons simuler la fonction de transfert entre chacune des composantes U_{xx} et la tension différentielle de sortie du préamplificateur. La figure 4 donne l'amplitude de ces fonctions de transfert dans le domaine fréquentiel.

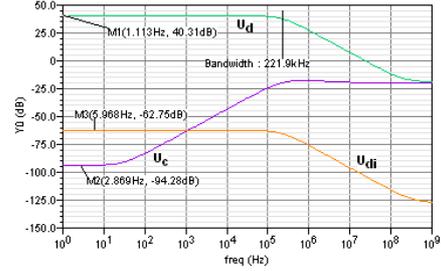


Figure 4. Fonction de transfert en fréquence du préamplificateur

En plus du gain de 40dB pour U_d , les trois résultats principaux de cette simulation sont un taux de réjection de 60dB pour chacune des U_{di} par rapport à U_d sur toute la bande utile, une bande passante de 200kHz (donc supérieure à nos besoins estimés à 50kHz) et plus de 90dB de rejection du mode commun en très basse fréquence (plus de 25dB à la limite de la bande utile). La figure 5 montre les fonctions de transfert en statique. Cette analyse montre que le gain différentiel est linéaire entre -5mV et 5mV autour du point de polarisation. Cette plage paraît mince, cependant elle reste largement suffisante pour l'amplitude des signaux que l'on souhaite mesurer.

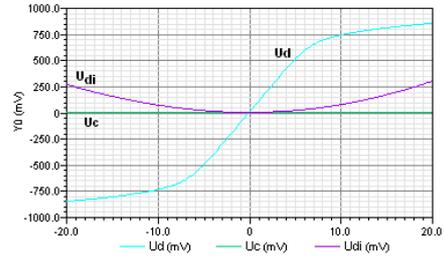


Figure 5. Fonction de transfert statique du préamplificateur

B. Amplificateur d'instrumentation

Un gain de 40dB serait insuffisant pour permettre l'enregistrement d'ENG à partir d'une telle électrode. Nous avons donc complété le préamplificateur moyeneur par un amplificateur d'instrumentation à gain variable afin de pouvoir ajuster le gain total du système en fonction des enregistrements obtenus. L'optimisation du rapport Signal sur Bruit de ce second étage est moins critique que celui du premier étage. En effet, le préamplificateur fournit un signal suffisamment élevé comparativement au bruit généré à l'entrée de l'amplificateur d'instrumentation. Les gains de cet amplificateur sont programmables numériquement à travers un jeu d'interrupteurs et de résistances.

Les gains de cet amplificateur sont donnés dans le tableau 1.

C. Circuit complet fabriqué

Les performances de ce circuit ont été évaluées par analyses Monte Carlo et les résultats obtenus sont reportés dans le tableau 1.

Surface active (7 canaux)	1.16 mm ²
Alimentation	3.3 V
Gain (Preamp)	40 dB
Gain (Inst amp)	6 dB ≤ G ≤ 80 dB
Gain (Amp complet)	46dB ≤ G ≤ 120dB
Bruit ramené en entrée (Preamp)	0.672 μVrms
Bruit ramené en entrée (Amp complet)	0.677 μVrms
Bande passante (Amp complet)	≈ 220 kHz

Tableau 1

Ces résultats sont cohérents avec nos spécifications puisque nous avons un gain de $46\text{dB} \leq G \leq 120\text{dB}$, une bande passante supérieure à 50kHz et un bruit ramené en entrée sous le μV_{rms} dans cette bande. Le circuit (Cf. figure 6) a été réalisé en technologie CMOS AMS 0,35μm.

IV. PERSPECTIVES

Dans un avenir proche, nous prévoyons trois expérimentations ayant pour but de mesurer l'efficacité de notre électrode. Durant ces trois expérimentations, nous chercherons à caractériser la sensibilité, la sélectivité spatiale et la réjection des signaux parasites de notre électrode.

V. CONCLUSION

Dans ce papier, nous avons présenté les deux premiers éléments d'une chaîne d'acquisition de signaux physiologiques nerveux. L'électrode cuff hexagonale propose des caractéristiques de sensibilité et de réjection des signaux parasites plus intéressantes que celles de l'électrode cuff tripolaire classique. Le circuit d'amplification faible-bruit développé par notre équipe permet, au niveau transistor, de moyenniser les signaux enregistrés au voisinage d'un pôle d'enregistrement et donc de reprendre la technique classique utilisée pour s'affranchir des signaux parasites extérieurs.

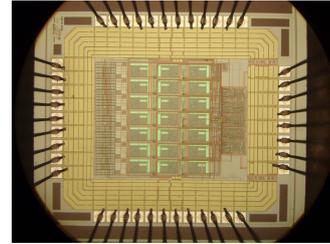


Figure 6. Microphotographie du prototype d'amplification

REFERENCES

- [1] L. Gouyet, G. Cathébras, S. Bernard, D. Guiraud, Y. Bertrand, "A Cuff Electrode dedicated to ENG Recording with Multipolar Configuration for both Efficient Sensitivity and High Rejection of EMG Parasitic Signals", 9th Vienna International Workshop on Functional Electrical Stimulation, Krems, Austria, 2007.
- [2] Zoran M. Nikolić, Dejan B. Popović, Richard B. Stein and Zoltan Kenwell, "Instrumentation for ENG and EMG Recordings in FES Systems", IEEE Transactions on Biomedical Engineering, Vol 41 n°7, July 1994.
- [3] P.B. Yoo and D.M. Durand, "Selective Fascicular Recording of the Hypoglossal Nerve Using a Multi-Contact Nerve Cuff Electrode", Proceedings of the 25th Annual International Conference of the IEEE EMBS, Sept 17-21, 2003.
- [4] Ch.Pflaum, R.R. Riso and G. Wiesspeiner, "Performance of Alternative Amplifier Configurations for Tripolar Nerve Cuff Recorded ENG", 18th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Amsterdam 1996.
- [5] Andreas Demosthenous, John Taylor, Iasonas F. Triantis, Robert Rieger and Nick Donaldson, "Design of an Adaptive Interface Reduction System for Nerve Cuff Electrode Recording", IEEE Transactions on Circuits and Systems ; Regular Papers, Vol 51 n°4, April 2004.
- [6] D.A. Johns and K. Martin, "Analog Integrated Circuit Design", John Wiley & Sons, Inc., 1997.

DOCTISS 2008

Structures élémentaires pour la conception de circuits reconfigurables à l'aide de mémoires magnétiques

Yoann GUILLEMENET¹ (Doctorant 1er année), Lionel TORRES² (Directeur de Thèse)

LIRMM
Université Montpellier 2, UMR CNRS 5506
161, Rue ADA
34392 Montpellier, France

¹guillemene@lirmm.fr, ²torres@lirmm.fr

Résumé—Cet article présente l'étude et la conception d'un FPGA à base de cellules mémoires magnétiques à écriture assistée thermiquement (TAS-MRAM). La non volatilité est réalisée à travers l'utilisation de jonctions tunnel magnétiques (MTJs). Une écriture assistée thermiquement est utilisée afin de réduire la consommation pendant l'opération d'écriture. De plus, cette non-volatilité permet de réduire le temps de configuration nécessaire à chaque mise sous-tension du circuit en comparaison avec les FPGAs classiques basés sur la mémoire SRAM. Nous proposons un micro-FPGA utilisant la cellule TAS-MRAM, permettant de réaliser une reconfiguration dynamique tout en présentant une architecture microélectronique simple.

Mots-clés—FPGA, mémoire magnétique, reconfiguration dynamique, TAS-MRAM.

I. INTRODUCTION

Les FPGAs (Field Programmable Gate Arrays) sont actuellement basés sur une technologie SRAM, impliquant l'utilisation d'une mémoire permanente externe pour le chargement de la mémoire de configuration à chaque mise sous tension. Cependant, avec l'évolution des technologies, et notamment des technologies émergentes, il est de plus en plus envisageable d'utiliser des points mémoire non volatiles afin de gérer la configuration de composants FPGA.

En effet, dans les systèmes embarqués, l'utilisation d'une mémoire non-volatile permet au système d'être déconnecté lorsqu'il n'est pas utilisé afin de réduire sa consommation d'énergie. Cela permet aussi de s'affranchir de la mémoire externe et du temps de latence à la mise sous tension. Dans cet article nous allons nous intéresser aux technologies MRAMs (Magnetic RAM) et plus spécifiquement à la TAS-MRAM. Nous montrerons comment nous pourrions utiliser cette technologie dans le cadre de la conception d'un FPGA.

Dans la section 2, nous décrivons les mécanismes de lecture/écriture des MTJs pour des cellules TAS-MRAM. Dans la section 3 nous présentons le schéma électrique de la structure TAS-MRAM de conversion magnétique/électrique. Les résultats de simulation seront quant à eux présentés dans la section suivante. La Section 5 décrit les principaux blocs élémentaires d'un FPGA à base de TAS-MRAM. Finalement, la dernière section décrit l'architecture du prototype μ FPGA.

II. MÉCANISMES DE LECTURE/ÉCRITURE DES MTJs

Une MTJ est constituée de 2 couches ferromagnétiques séparées par une couche d'oxyde tunnel. L'orientation relative des 2 couches ferromagnétiques va permettre de distinguer 2 valeurs de résistance aux bornes de la jonction selon que les orientations magnétiques des 2 couches sont parallèles (R_p) ou antiparallèles (R_{ap}). Les propriétés magnétiques des 2 couches sont différentes, l'une d'entre elle est dite libre car elle peut être polarisée avec un champ magnétique plus faible que la couche dite figée. La rémanence des matériaux ferromagnétique confère à cette mémoire la non volatilité. La seconde caractéristique importante de ces jonctions est le ratio entre les valeurs de la résistance selon l'orientation des 2 couches (R_{ap} a une résistance moyenne plus élevée que R_p). Ce rapport est donné par :

$$TMR = \frac{\Delta R}{R} = \frac{R_{ap} - R_p}{R_p} = \frac{R_{max} - R_{min}}{R_{min}}$$

Il peut varier de 40 à 220% [2].

A. Structure TAS-MRAM

Cette évolution technologique combine un chauffage local de la jonction (par un faible courant) et un seul champ magnétique d'amplitude faible [3] pour faire commuter la direction de la magnétisation de la couche libre.

La phase d'écriture exige plusieurs étapes qui sont décrites dans la Figure 1. Quand la jonction est chauffée au-dessus de la température de blocage par un courant circulant à travers la jonction, la magnétisation de la couche ferromagnétique est libérée et peut être changée complètement sous l'action d'un champ magnétique de faible puissance (donc avec un courant plus faible). Le champ magnétique est maintenu après la fin de la phase de chauffe afin que la jonction refroidisse sous l'influence du champ, ce qui permet d'assurer la magnétisation stockée dans la couche libre.

Le problème est que l'information est ici stockée sous forme magnétique, or pour la conception de circuits hybrides

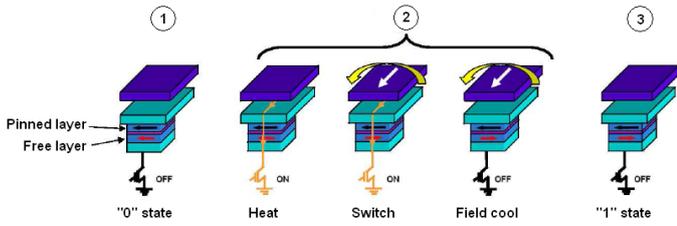


Fig. 1. Écriture d'une MTJ (TAS-MRAM)

CMOS/Magnétique, il sera nécessaire de transformer l'information magnétique en information électrique. La structure que nous proposons est présentée dans la section suivante.

III. STRUCTURE TAS-MRAM DE CONVERSION MAGNÉTIQUE / ÉLECTRIQUE

La structure (Figure 2) doit être aussi petite que possible car elle va être dupliquée pour chaque point mémoire du FPGA. Cette structure est composée de 2 inverseurs rebouclés formant

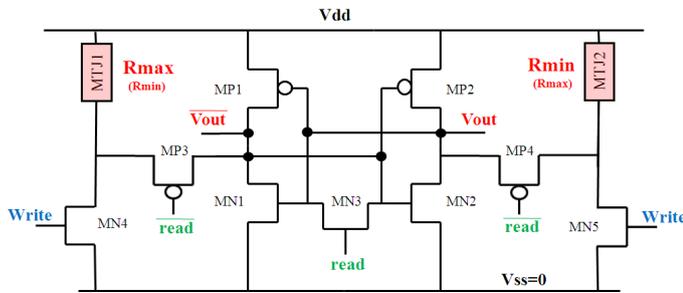


Fig. 2. Cellule TAS-MRAM

un latch. Les transistors MP3 et MP4 permettent d'isoler la structure lors de la phase d'écriture, MN4 et MN5 autorisent ou non la circulation du courant de chauffe dans les jonctions tunnels magnétiques sélectionnées. Le transistor MN3 permet de court-circuiter l'entrée et la sortie, ce qui positionne la structure dans un état métastable [1]. Les 2 MTJs contiennent des informations complémentaires qui vont déséquilibrer la structure et faciliter son basculement dans un état ou dans l'autre. Les transistors ont été dimensionnés de telle manière que la structure fonctionne en prenant en compte les dispersions technologiques (tensions de seuil, mismatch des transistors, etc ...). Ce sont ces dispersions qui imposent l'utilisation de 2 MTJs. En effet, on aurait pu se limiter à 1 MTJ et une résistance de référence, mais une variation de 0,1nm sur une couche d'oxyde de 1,3nm induit une variation de la valeur de résistance moyenne de la MTJ d'environ 100%. Par conséquent, une structure différentielle est nécessaire.

Cette structure autorise un nouveau type de reconfiguration : la reconfiguration masquée ou cachée. Dans un premier temps, on écrit l'information dans les MTJs grâce au courant de chauffe et à la ligne de champ. Ensuite, l'information est chargée dans le latch grâce au transistor de court-circuit (MN3). Lorsque le signal sur le transistor de court-circuit est relâché, il y a redondance de l'information. Celle-ci est stockée sous forme électrique dans le latch et sous forme magnétique dans les MTJs.

De plus ces 2 informations sont indépendantes l'une de l'autre puisque le signal Read sur le transistor de court-circuit a été

relâché. On peut donc venir écrire dans les MTJs pendant que le système utilise l'information dans le latch.

IV. RÉSULTATS DE SIMULATIONS

Les simulations des cellules TAS-MRAM ont été effectuées à l'aide d'un modèle électrique Spectre, ce modèle a été développé par le laboratoire SPINTEC. Il est possible de jouer sur plusieurs paramètres comme la taille de la jonction, le Ra utilisé (Résistance surfacique de la MTJ exprimé en $\Omega \cdot \mu m^2$) qui est fixé en partie par l'épaisseur de la barrière tunnel.

Pour définir ces valeurs, nous avons utilisé des jonctions de forme circulaire de diamètre 350nm, un Ra de $30\Omega \cdot \mu m^2$ et un TMR pessimiste de 20% de manière à s'assurer une commutation de la structure dans un cas extrêmement défavorable. Les valeurs correspondantes à ces paramètres sont pour Rmin 310Ω et 379Ω pour Rmax.

Des simulations de Monte Carlo ont été réalisées pour vérifier la robustesse de la structure face aux variations et aux dispersions liées au process technologique, ces simulations sont présentées dans la figure 3.

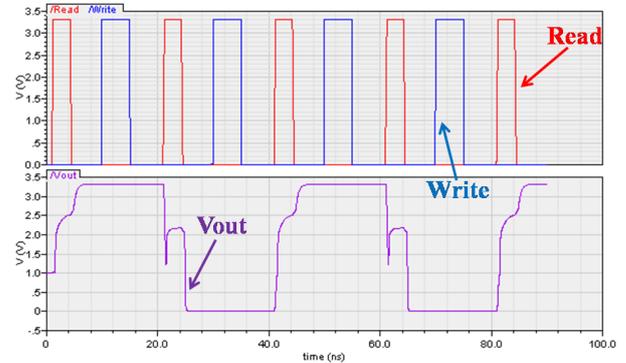


Fig. 3. Simulation de la cellule TAS-MRAM

V. BLOCS ÉLÉMENTAIRES D'UN FPGA MRAM

Sur les dispositifs programmables, le bitstream de configuration est stocké dans des cellules SRAMs afin de configurer les différents blocs logiques permettant de réaliser la fonction souhaitée. Dans les FPGAs, ces blocs sont des CLB (bloc logique configurable) et des matrices d'interconnexions.

Dans cette section, nous proposons une structure à base de TAS-MRAM permettant de réaliser une Look Up Table (LUT) qui est l'élément principal d'un CLB, ainsi qu'une matrice d'interconnexions.

A. LUT et matrice d'interconnexions

Le but d'une Look up Table est de mettre en application des fonctions booléennes. La table de vérité est stockée dans des points mémoire. Un arbre de multiplexage contrôlé par les entrées vient sélectionner un des points mémoires pour placer la valeur du bit stocké en sortie. Les matrices d'interconnexions permettent le lien entre les CLB.

La solution consiste à remplacer les points mémoires standards par des points mémoires TAS-MRAM (Figure 4). Cela permet outre la non volatilité des LUTs et des matrices d'interconnexions de pouvoir amener un nouveau concept de reconfiguration : la reconfiguration masquée et dynamique.

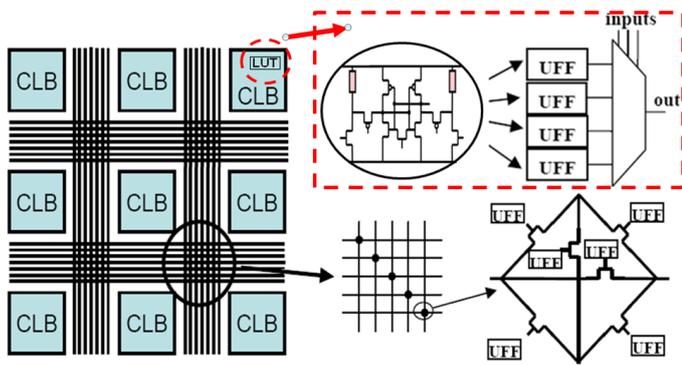


Fig. 4. Architecture d'un FPGA à base de MRAM

VI. CONCEPTION D'UN μ FPGA À BASE DE TAS-MRAM

A. Structures implémentées

Afin d'écrire l'information sur les MTJs, des générateurs de courants sont exigés. Nous avons implémenté dans le démonstrateur deux générateurs de courant bi-directionnel trois états (CG1 et CG2) dont l'un peut fournir $\pm 8\text{mA}$ et l'autre $\pm 14\text{mA}$ de façon à avoir une plus grande liberté lors du test en écriture.

Nous avons aussi intégré en plus de la LUT décrite précédemment, un autre type de LUT que nous appelons "LUT compensée". La différence de celle-ci est que les cellules mémoires TAS-MRAM qui la compose fonctionnent avec une seul MTJ, la deuxième étant émulée par une résistance variable (transistor) contrôlée en tension. Cela nous permet de s'affranchir des dispersions induites sur les MTJs lors du process.

Pour plus de marge de manoeuvre lors du test, nous avons implémenté des cellules mémoires TAS-MRAM indépendantes (compensées : cellules 1 et 3, et non compensées : cellules 2 et 4). Chacune d'elles a été implémentée avec des tailles de transistor différentes. Ainsi nous pourrons déterminer les dimensions optimales de la cellule en se basant sur les performances de chacune d'elles. La cartographie du μ FPGA est présentée en Figure 5.

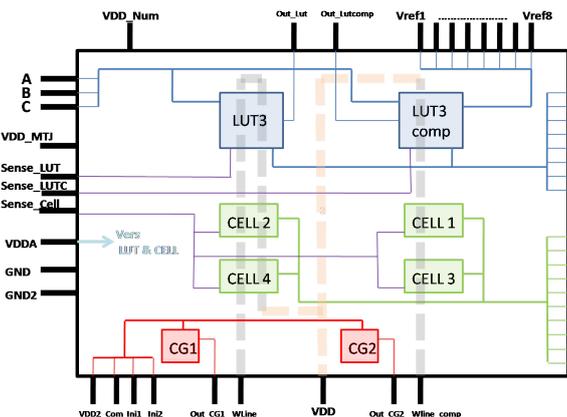


Fig. 5. Cartographie du μ FPGA TAS-MRAM

B. Design et Layout

Pendant la phase de design, les 4 premiers niveaux (Figure 6) de métaux sont utilisés pour la CMOS classique; le dernier étant réservé pour les lignes d'écriture. En effet, le process

magnétique a lieu après le process CMOS. La structure est dite "above IC" car les MTJs sont déposées au-dessus de la partie CMOS, permettant également de limiter la surface utilisée par la mémoire.

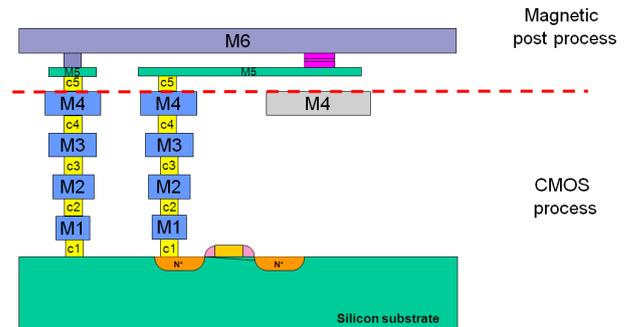


Fig. 6. Les différents niveaux de process

Ce μ FPGA est en cours de conception, il combine à la fois la technologie CMOS $0.35\mu\text{m}$ et le post process magnétique. Le Layout de ce démonstrateur est présenté en Figure 7.

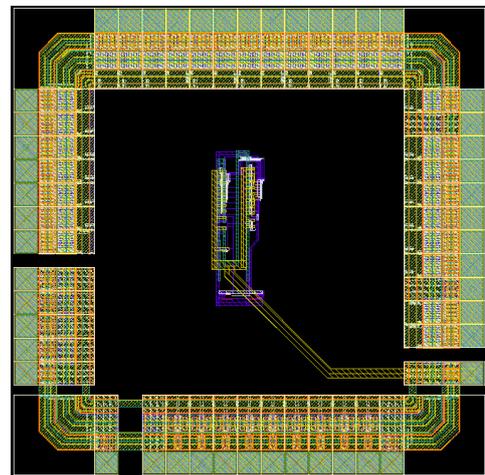


Fig. 7. Layout du μ FPGA TAS-MRAM

VII. CONCLUSION

Dans cet article, nous avons présenté une architecture microélectronique TAS-MRAM pour des cellules élémentaires non-volatiles de FPGA permettant la reconfiguration dynamique. La principale qualité de l'utilisation d'une telle technologie dans l'application visée est sa flexibilité. Grâce à sa mémoire de configuration non volatile, le circuit reconfigurable peut à tout moment être désalimenté pour limiter la consommation et être alimenté lorsqu'il doit être utilisé. De plus, l'utilisation de la TAS-MRAM permet de réduire le temps de configuration puisqu'il n'y a pas besoin de charger les données de configuration dans une mémoire externe au circuit.

RÉFÉRENCES

- [1] W. Black and B. Das. Programmable logic using giantmagneto-resistance and spin-dependent tunneling devices. *J.Appl. Phys.*, 87 :6674–6679, 2000.
- [2] S. P. Parkin and al. Giant tunneling magnetoresistance at room temperature with MgO (100) tunnel barriers. *online publication*, www.nature.com/naturematerials, 2004.
- [3] O. Redon, M. Kerekes, R. Sousa, S. H. Prejbeanu, F. Ponthonnier, A. Persico, and J. Nizieres. Programmable logic using giantmagneto-resistance and spin-dependent tunneling devices. *ICMTD*, pages 113–114, 2005.

MODELISATION DE COURTS-CIRCUITS RESISTIFS EN PRESENCE DE DIAPHONIE

NICOLAS HOUARCHE (DOCTORANT 2E ANNEE), MARIANE COMTE, MICHEL RENOVELL

LIRMM
161 rue Ada, 34392 Montpellier Cedex 5, France

HOUARCHE@LIRMM.FR
MARIANE.COMTE@LIRMM.FR
MICHEL.RENOVELL@LIRMM.FR

Résumé— Cet article propose une analyse du comportement d'un court-circuit résistif en fonction de sa résistance R_b a priori inconnue ; en considérant le court-circuit aggravé par de la diaphonie, d'autres paramètres entrent également en compte comme la valeur de l'écart temporel entre les transitions sur les entrées des portes concernées par le court-circuit. Une équation est finalement proposée afin de déterminer l'intervalle de résistance du défaut, connaissant l'écart temporel et le seuil de détection du chemin.

Mots clés — modélisation ; test en délai ; court-circuit résistif ; diaphonie

INTRODUCTION.

Dans les circuits électroniques actuels, les défauts les plus fréquents sont les défauts causant des courts-circuits et/ou des circuits ouverts et ce sur un ou plusieurs niveaux de métal du composant. Le test s'appuie sur des modèles de faute supposés représenter le comportement du défaut et permettre une génération simple des vecteurs de test grâce à un ATPG (Automatic Test Pattern Generator) et des simulations de fautes. Les modèles de faute classiques ont prouvé leur efficacité mais celle-ci n'est plus suffisante pour couvrir le spectre des défauts des circuits actuels. Il est donc nécessaire de développer des modèles plus précis, notamment ceux représentant des comportements complexes et des défauts ayant une grande probabilité d'occurrence [1] [2]. Les courts-circuits résistifs appartiennent à ces deux catégories puisqu'ils modifient le comportement électrique des connexions et qu'ils sont souvent présents.

Dans cet article nous considérons deux portes logiques dont le court-circuit en sortie est modélisé par une résistance et une capacité en parallèle qui représente le bruit de diaphonie. Il est important de noter que la valeur de cette résistance est un paramètre imprévisible alors que la valeur de la capacité est quant à elle connue d'après la conception (grâce à des extractions de layout). Afin de détecter ce défaut, on se doit donc de l'analyser en tant que fonction du paramètre inconnu : la résistance.

Ce papier est organisé comme suit. La partie 2 est consacrée à l'étude plus simple du cas des courts-circuits seuls, sans considérer la diaphonie. La troisième partie prend en compte la capacité de diaphonie. La partie 4 présente le modèle proposé pour représenter le comportement d'un circuit affecté par un court-circuit résistif et du bruit de diaphonie. Enfin, une dernière partie conclue ce travail.

CAS DES COURTS-CIRCUITS SIMPLES.

La première étape de notre étude est de s'intéresser aux courts-circuits simples, c'est-à-dire sans considérer la présence de la capacité de couplage.

On considère les résistances équivalentes des transistors constituant les portes logiques [3] comme indiqué sur la figure 1 :

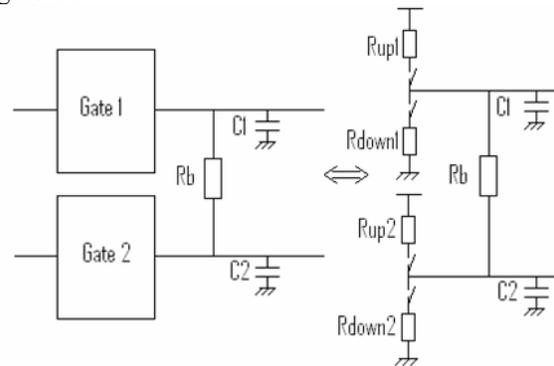


Fig. 1. Court-circuit sans diaphonie.

Si l'on considère un front montant à l'entrée de la porte 1 et un niveau bas à l'entrée de la porte 2, alors seules les résistances équivalentes R_{down2} et R_{up1} sont à considérer. Les travaux menés dans [3] nous permettent alors d'écrire une équation liant le retard aux différents éléments du circuit, notamment la résistance de court-circuit R_b .

$$delay = - \left[1 + \frac{(R_{down}^2)^2}{(R_{down}^2 + R_b)^2} C_2 / C_1 \right] \left[\frac{R_{down}^2 + R_b}{R_{up}^2 + R_{down}^2 + R_b} \right] \log_2 \left(1 - \frac{0.5}{\frac{R_{down}^2 + R_b}{R_{up}^2 + R_{down}^2 + R_b}} \right) d_1 \quad (1)$$

où d_1 est le délai sans défaut.

Le défaut sera détecté si et seulement si le délai induit par celui-ci est supérieur au seuil de détection. La limite de détectabilité (R_{limite}) peut ainsi être aisément définie comme étant la résistance pour laquelle le délai supplémentaire est égal à ce seuil de détection et on peut ensuite définir un intervalle de détectabilité : $[0, R_{limite}]$.

Par la suite, nous allons compléter ce modèle en considérant une capacité parasite en parallèle de la résistance de court-circuit R_b , modélisant le bruit de diaphonie.

COURTS-CIRCUITS AGGRAVES PAR DE LA DIAPHONIE.

Le comportement de deux portes court-circuitées est modifié par la présence de la capacité de diaphonie [4]. Si la valeur de la résistance de court-circuit reste un paramètre inconnu, la valeur de la capacité est, quant à elle, déterminable par une extraction de layout. Le but de cette partie est toujours de définir un intervalle de détectabilité du défaut, en tenant compte de cette capacité.

La figure 2 montre les variations du retard de la ligne victime en fonction de l'écart temporel entre les signaux d'entrées des deux portes.

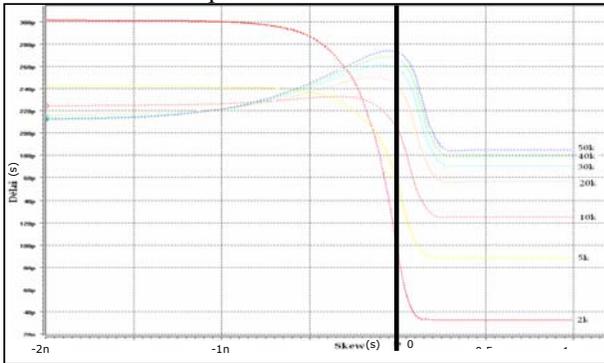


Fig. 2. Variation du retard sur la ligne victime en fonction de l'écart temporel entre les signaux d'entrée.

Sur ce réseau de courbes, on peut distinguer plusieurs régions : - l'écart temporel est négatif et grand en valeur absolue, c'est-à-dire que le signal sur la ligne agresseur arrive longtemps avant celui sur la ligne victime.

- l'écart temporel est proche de 0, c'est-à-dire que les deux signaux arrivent en même temps.
- l'écart temporel est positif.

L'écart temporel est négatif et grand en valeur absolue.

On considère que l'écart temporel entre les deux signaux d'entrées est grand lorsqu'il y a au moins 5 portes de différence sur les chemins de sensibilisation, soit, en considérant un délai moyen de 200ps par porte, environ 1ns.

On suppose que la résistance de court-circuit est faible, le cas où elle est grande n'étant pas intéressant (comportement nominal). Sur la figure 3, on observe le comportement du délai en fonction de la résistance de court-circuit pour différentes valeurs de capacité.

Pour une résistance d'environ 500Ω , et une capacité variant de 0 à 40fF, la variation de résistance pour un délai identique est de 15Ω soit une erreur de 3%.

On peut donc considérer que, pour un écart temporel inférieur à 1ns, il n'est pas nécessaire de prendre en compte la capacité de diaphonie et on peut utiliser le modèle décrit dans la partie précédente.

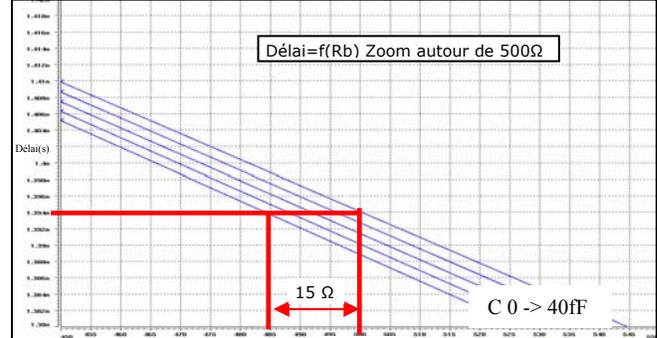


Fig. 3. Variation de la résistance en fonction de la capacité pour un délai fixe.

L'écart temporel est proche de 0

Lorsque l'écart temporel est proche de 0, c'est-à-dire que les signaux arrivent sur les entrées des deux portes en même temps, la capacité et la résistance ont chacun leur effet.

Si la résistance R_b est grande (pas de court-circuit), seule l'action de la capacité est visible. Nous nous intéresserons donc au cas où la résistance de court-circuit est faible.

La simulation du délai en fonction de l'écart temporel pour différentes valeurs de résistance est donnée sur la figure 4 :

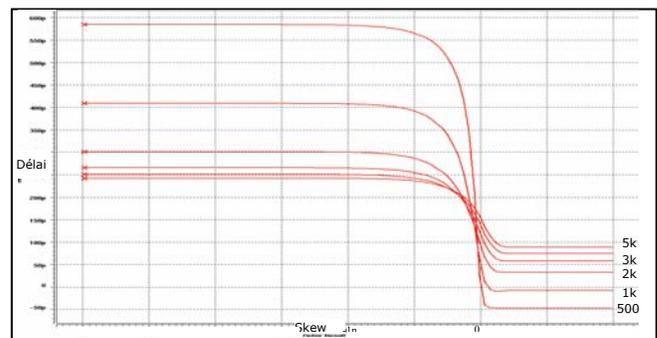


Fig. 4. Variation du retard sur la ligne victime en fonction de l'écart temporel pour des résistances faibles.

On va maintenant chercher à modéliser ce réseau de courbes. Ainsi, en connaissant l'écart temporel z entre les signaux d'entrée des deux portes et le seuil de détection du chemin considéré, on pourra déterminer la résistance limite et l'intervalle de détection.

MODELISATION.

On considère deux familles de points connus sur la courbe :

- $D_i(R_b)$: délai pour $z = -1ns$.
- $d_i(R_b)$: délai pour $z = 0$.

$D_i(R_b)$

On a prouvé dans la partie précédente que lorsque l'écart temporel entre les signaux d'entrée est inférieur à $-1ns$, la

capacité n'a plus d'influence sur le délai. La valeur de $D_i(R_b)$ est donc donnée par la formule de Walker donnée dans la partie II.

$d_i(R_b)$

On cherche à obtenir la valeur du délai d_i en fonction de la résistance de court-circuit R_b lorsque les signaux arrivent sur les entrées des portes en même temps. Afin d'obtenir cette équation, on trace la courbe $délai=f(R_b)$ pour $z=0$.

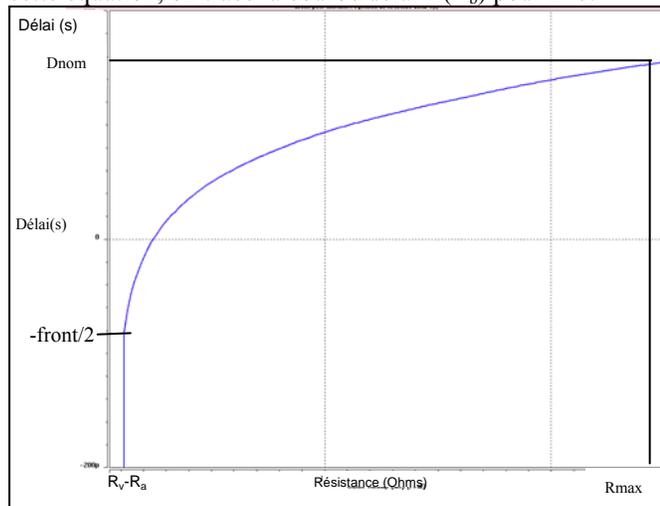


Fig. 5. Simulation du délai en fonction de la résistance R_b , pour un écart temporel nul.

A cette courbe, on peut associer l'équation :

$$d_i(R_b) = \left[\frac{R_b - (R_{down}^2 - R_{up}^1)}{R_{max} - (R_{down}^2 - R_{up}^1)} \right]^{1/3} \left(D_{nom} + \frac{front}{2} \right) - \frac{front}{2} \quad (2)$$

Où R_{max} est la valeur de résistance pour laquelle on obtient le délai nominal et $front$ est la largeur du front des signaux d'entrée.

La connaissance des points D_i et d_i nous permet alors de déterminer l'équation reliant le délai à l'écart temporel z entre les signaux d'entrées des deux portes court-circuitées :

$$Délai = \left[\frac{d_i^4 - D_i^4}{10^{-9}} z + d_i^4 \right]^{1/4} \quad (3)$$

VALIDATION.

Afin de valider la pertinence de l'équation (2), nous avons mené une étude visant à comparer des résultats issus de simulation Spice et les résultats de cette équation.

Pour un écart temporel et un seuil de détection donnés, on détermine l'intervalle de détection $[0, R_{limite}]$ tel que $Délai(R_{limite})=Tsl$. Les résultats sont donnés dans le tableau ci-dessous.

Tableau 1. Comparaison des R_{limite} simulées et calculées.

z=-1ns et Tsl=300ps		Rsimulé	Rcalculé
Ra=1900 et Rv=1950	Cm=40fF	770	810
	Cm=70fF	766	986
Ra=2400 et Rv=2600	Cm=40fF	2030	1844
	Cm=70fF	2450	2164

z=-1ns et Tsl=500ps		Rsimulé	Rcalculé
Ra=1900 et Rv=1950	Cm=40fF	222	235
	Cm=70fF	221	332
Ra=2400 et Rv=2600	Cm=40fF	677	574
	Cm=70fF	676	764

z=-0.5ns et Tsl=500ps		Rsimulé	Rcalculé
Ra=1900 et Rv=1950	Cm=40fF	215	148
	Cm=70fF	214	228
Ra=2400 et Rv=2600	Cm=40fF	630	380
	Cm=70fF	631	544

Tableau 2. Comparaison des taux de couverture simulés et calculés grâce au modèle.

z=-1ns et Tsl=500ps		Rsimulé	Rcalculé
Ra=1900 et Rv=1950	Cm=40fF	51%	52%
	Cm=70fF	51%	60%
Ra=2400 et Rv=2600	Cm=40fF	88%	81%
	Cm=70fF	88%	92%

z=-0.5ns et Tsl=500ps		Rsimulé	Rcalculé
Ra=1900 et Rv=1950	Cm=40fF	50%	42%
	Cm=70fF	50%	49%
Ra=2400 et Rv=2600	Cm=40fF	85%	64%
	Cm=70fF	85%	79%

Ce dernier tableau prouve que les taux de couverture, c'est-à-dire la détectabilité du défaut, calculés et simulés sont très proches (en moyenne 4% d'écart), ce qui prouve la validité du modèle.

CONCLUSION.

Cette étude a proposé une modélisation tant comportementale qu'électrique des courts-circuits résistifs aggravés par de la diaphonie. Un modèle mathématique a été proposé, ainsi que les preuves de sa validité.

Grâce à ce modèle, on est maintenant capable de déterminer l'intervalle de détectabilité d'un défaut sur un chemin, connaissant l'écart temporel entre les signaux des portes court-circuitées et le seuil de détection du chemin considéré.

Par la suite, ce modèle sera utilisé dans les calculs de taux de couverture afin de déterminer l'efficacité d'un vecteur de test.

REFERENCES.

- [1] W. Needham, C. Prunty, E. Hong Yech, "High Volume Processor Test Escape, an Analysis of Defect our Test are Missing", International Test Conference, pp. 25-34, 1998.
- [2] C.L. Henderson, J.M. Soden and C.F. Hawkins, "The Behavior and Testing Implications of CMOS IC Open Circuits", International Test Conf., pp. 302-303, 1991.
- [3] Z. Li, X. Lu, W.Qiu, W. Shi, D. Walker, "A circuit level fault model for resistive opens and bridges", Trans. On Design Automation of Elect. System, vol. 8, n°4, pp 546-559, 2003.
- [4] L. Wang, S. Gupta, M. Breuer, "Modeling and simulation for crosstalk aggravated by weak-bridge defects between on-chip interconnects", Asian Test Symposium 2004.

DOCTISS 2008

Modélisation et Simulation de la Stimulation Électro-Fonctionnelle de la Vessie

Jérémy LAFORÊT¹ (Doctorant 2e année), David GUIRAUD² (Directeur de Thèse)
Christine AZEVEDO-COSTE (Co-encadrante)³, David ANDREU (Co-encadrant)⁴

LIRMM

161 rue Ada, Montpellier France

¹laforet@lirmm.fr, ²guiraud@lirmm.fr, ³azevedo@lirmm.fr, ⁴andreu@lirmm.fr

Résumé— Nous présentons un nouveau modèle de muscles lisses pouvant être utilisé pour simuler les effets de la Stimulation Électro-Fonctionnelle (SEF). Il est composé d'un jeu d'équations différentielles dont les paramètres ont un sens physiologique. Ils peuvent donc servir à évaluer de façon objective et quantitative l'état du muscle. De plus, l'entrée du modèle est contrôlée par un signal de SEF de façon à ce qu'il puisse simuler le comportement du muscle sous stimulation artificielle. Nous utilisons ce modèle pour simuler la contraction de la vessie par la stimulation du détrusor. Cela montre que le modèle est capable de prédire la réponse temporelle du muscle, la pression intra-vésicale et le temps nécessaire pour vider la vessie. Les simulations présentent des résultats en accord avec la littérature. Ces résultats préliminaires, après validation par des expériences in-vivo, seront utilisés pour caractériser les vessies ayant besoin de stimulation, par exemple pour les paraplégiques, et ensuite optimiser la stimulation nécessaire lorsqu'une neuroprothèse est utilisée pour rétablir la miction.

Mots-clés— Muscles lisses, modélisation, vessie, simulation, stimulation électro-fonctionnelle.

I. INTRODUCTION

La modélisation des muscles permet de comprendre la manière dont le muscle se contracte, de le caractériser quantitativement pour le diagnostic, et enfin de concevoir des stratégies de contrôle utilisant des systèmes de Stimulation Electro-Fonctionnelle (SEF).

Il existe trois types de muscles : les muscles striés squelettiques qui sont sous le contrôle volontaire et sont utilisés principalement pour la motricité ; les muscles lisses, hors du contrôle volontaire et principalement localisés dans les parois d'organes creux (vaisseaux sanguins, voies respiratoires, systèmes digestif et urinaire) ; et le muscle cardiaque qui forme les parois du coeur.

Nous visons : i) un modèle de connaissance du muscle lisse, ii) un modèle de muscle contrôlé par la SEF. Nous adoptons la même approche que celle qui nous a permis de développer notre modèle de muscles striés [1].

Le principe de la contraction muscle est similaire entre les différents types de muscle. Il repose sur les interactions entre deux protéines organisées en filaments : l'actine et la myosine. Dans un premier temps, la commande neurale fait augmenter la concentration intracellulaire de Ca^{2+} . Cette augmentation du Ca^{2+} initie une chaîne de réactions amenant à la kinase des chaînes légères de myosine (MLCK).

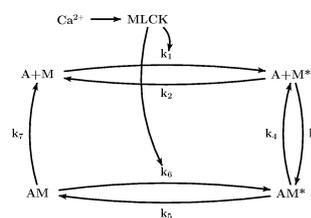


Fig. 1. Modèle de Hai & Murphy

Une fois activée (Fig. 1, état ②), la tête de myosine s'étend et s'accroche à un site libre sur la molécule d'actine. La paire actine/myosine attachée est appelée un pont ou "cross-bridge" (état ③). Ensuite la tête de myosine tourne, générant le mouvement. Finalement le pont se casse (retour à l'état ②) et le cycle se répète.

Il existe un état supplémentaire spécifique au muscle lisse, dans lequel le muscle est capable de maintenir sa contraction avec une consommation d'énergie minimale. Hai & Murphy ont modélisé cet état par le latch-bridge (état ④), un pont déphosphorylé [2]. La déphosphorylation entraîne une diminution de la probabilité de détachement du pont et libère un groupe phosphate qui peut alors être réutilisé. Ce modèle donne les équations cinétiques liant les différents états de la paire actine/myosine (myosine libre, myosine activée, cross-bridge, latch-bridge).

II. MODÈLE

A. Dynamique du Calcium

Pour décrire l'évolution de la concentration de calcium induite par la stimulation, nous nous basons sur le modèle de Koenigsberger et al. [3]. Cependant, nous ne considérons que les effets dus à une stimulation électrique des cellules musculaires. On peut alors déduire la concentration de calcium $[Ca^{2+}]_i$ à partir du potentiel $V_{membrane}$, celui-ci étant imposé par le nerf stimulé. On écrit alors l'évolution de la concentration de calcium dans la cellule et le réticulum sarcoplasmique en fonction des différents courants ioniques :

$$[Ca^{2+}]_i = J_{IP3} - J_{VOCCi} + J_{Na/Ca} - J_{SRuptake} + J_{CICR} - J_{extrusion} + J_{leak} \quad (1)$$

$$[Ca^{2+}]_{SR} = J_{SRuptake} - J_{CICR} - J_{leak} \quad (2)$$

B. Actine & Myosine

Hai et Murphy ont développé un modèle du *latch-bridge* [2], le considérant comme un pont actine/myosine classique déphosphorylé. Ce qui rend son détachement plus difficile et remet en circulation un groupe phosphate. Ce modèle décrit la cinétique de quatre espèces chimiques intervenant dans la contraction du muscle : M la myosine et AM les ponts actine/myosine, chacun pouvant être activé ou non (l'activation est représentée par l'ajout d'une étoile $*$ à la suite du symbole). Les coefficients k_i sont les vitesses des réactions faisant passer d'une espèce à l'autre. Ces différentes réactions sont représentées sur la figure 1. Nous avons modifié ce modèle de manière à prendre en compte la dynamique du calcium ainsi que le mouvement relatif des filaments d'actine et de myosine. Les équations cinétiques deviennent alors :

$$\begin{aligned} \dot{[M]} &= -k_1([Ca^{2+}]) [M] + k_2 [M^*] + k_7(\epsilon_c) [AM] \\ \dot{[M^*]} &= k_4(\epsilon_c) [AM^*] + k_1([Ca^{2+}]) [M] - (k_2 + k_3) [M^*] \\ \dot{[AM^*]} &= k_3 [M^*] + k_6([Ca^{2+}]) [AM] - (k_4(\epsilon_c) + k_5) [AM^*] \\ \dot{[AM]} &= k_5 [AM^*] - (k_7(\epsilon_c) + k_6([Ca^{2+}])) [AM] \end{aligned} \quad (3)$$

C. Raideur & Force

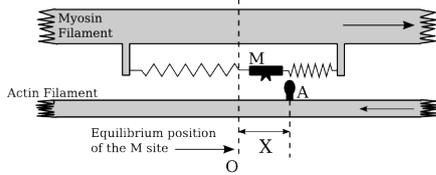


Fig. 2. Modèle de filament glissant de Huxley

Les développements suivants sont basés sur le modèle de filaments glissant de Huxley (figure 2) [4]. Huxley donne la dynamique de $n(x, t)$, la distribution de la proportion attachée des ponts en fonction de la probabilité d'attachement ou de détachement d'un pont.

C.1 Fonctions d'attachement et de détachement (f et g)

Il est possible de définir f et g de manière à ce que $f + g$ soit constant en ξ et d'y intégrer les fractions chimiques. L'avantage de cette méthode est de supprimer le besoin du calcul explicite de la distribution des ponts, tout en étant une approximation acceptable [1]. Les fonctions d'attachement et de détachement peuvent alors être écrites sous la forme suivante :

$$\begin{aligned} f(\xi, t) &= \begin{cases} \frac{[M^*]}{[M] + [M^*]} f_1, & 0 < \xi < 1, \\ 0, & \xi \notin [0; 1] \end{cases} \quad (4) \\ g(\xi, t) &= \begin{cases} \frac{[AM^*]}{[AM] + [AM^*]} g_1 + \frac{[AM]}{[AM] + [AM^*]} g_2 - f(\xi), & \forall \xi \end{cases} \end{aligned}$$

Avec : $f_1 \propto k_3, g_1 \propto k_4, g_2 \propto k_7$.

C.2 Moments de $n(\xi, t)$

En considérant l'introduction de la relation force/longueur au même stade que Yu et al [5]. ainsi que les fonctions f et g définies plus haut, on obtient la dynamique de

la distribution des ponts :

$$\frac{\partial n(\xi, t)}{\partial t} = fl(\epsilon_c) f(\xi, t) - (f(\xi, t) + g(\xi, t)) n(\xi, t) \quad (5)$$

La dérivé des moments d'ordre 1 et 2 obtenu par la méthode de Zahalak deviennent alors après simplification :

$$\dot{M}_0 = fl(\epsilon_c) f(\xi, t) - (f + g)(\xi, t) M_0 \quad (6)$$

$$\dot{M}_1 = \frac{fl(\epsilon_c) f(\xi, t)}{2} - (f + g)(\xi, t) M_1 + \epsilon_c M_0 \quad (7)$$

Il est donc possible d'exprimer les caractéristiques de l'élément contractile à l'aide des équations précédentes. Puisque l'on a $k_c \propto M_0$ et $F_c \propto M_1$.

D. Modèle mécanique cellulaire

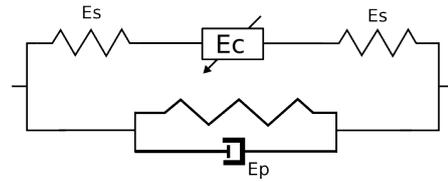


Fig. 3. Modèle Mécanique.

Le modèle mécanique (représenté sur la figure 3) est défini de façon à rendre compte des propriétés visco-élastiques de la vessie le plus simplement. Sa symétrie permet d'homogénéiser les effets des forces et déplacements sur le modèle. Il fait le lien entre la force contractile calculée précédemment et la force réellement générée par le muscle.

E. Modèle Global de la vessie

À l'échelle macroscopique le modèle n'est plus générique ; il devient spécifique à la vessie. Pour modéliser la vessie à cette échelle, nous décrirons sa géométrie et ses propriétés mécaniques.

E.1 Géométrie

Pour des raisons de simplicité, on considère la vessie comme une sphère creuse d'épaisseur finie (comme sur la figure 4). De plus, on considère le volume de la paroi constant.

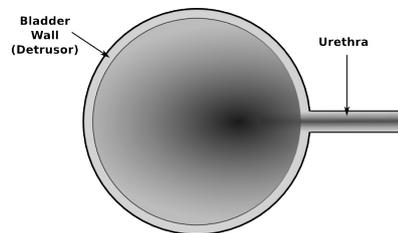


Fig. 4. Géométrie simplifiée de la vessie.

On obtient alors S la surface interne de vessie, V volume correspondant et V_{sh} le volume de la paroi :

$$\begin{aligned} S &= 4\pi r^2, \quad V = \frac{4\pi}{3} r^3 \\ V_{sh} &= \frac{4\pi}{3} ((r_0 + ep)^3 - r_0^3) \end{aligned} \quad (8)$$

Avec r le rayon de la sphère, et ep l'épaisseur de la paroi. En prenant l_{p0} une circonférence de la vessie au repos, on peut exprimer ces grandeurs en fonction de ϵ_p :

$$S = \frac{l_{p0}^2}{\pi} \epsilon_p^2, \quad V = \frac{l_{p0}^3}{6\pi^2} \epsilon_p^3 \quad (9)$$

E.2 Relation Force-Pression

D'après Arts et al., il est possible de relier la pression interne d'une cavité aux efforts dans la paroi [6]. Ces travaux, initialement appliqués au ventricule gauche du coeur, sont généralisés à toute cavité possédant une symétrie de rotation. Dans notre cas le volume de la paroi n'étant pas toujours négligeable, il nous faut utiliser la relation suivante :

$$P = \frac{1}{3} \ln \left(1 + \frac{V_{sh}}{V} \right) \sigma_f \quad \text{avec} \quad \sigma_f = \frac{F}{S} \quad (10)$$

Elle est obtenue par intégration sur un certain nombre de parois fines superposées, en considérant les efforts homogènes à l'intérieur de la paroi complète [6]. Il est alors possible de connaître la pression, à partir de la force exercée tangenciellement et de la géométrie.

E.3 Ecoulement

En se plaçant dans un cas simple (fluide parfait et écoulement irrotationnel) pour utiliser le théorème de Bernoulli, il est possible d'obtenir une approximation des conditions d'écoulement de l'urine hors de la vessie. On applique alors la formule de Torricelli (conservation de l'énergie en tout point de l'écoulement. Le point n°1 se situe à l'intérieur de la vessie et le n°2 dans l'urètre, on obtient alors :

$$\frac{1}{2} \rho v_1^2 + \rho g z_1 + P_1 = \frac{1}{2} \rho v_2^2 + \rho g z_2 + P_2 \quad (11)$$

Les effets gravitationnels seront négligés du fait de la très faible différence d'altitude entre les deux points. En considérant la pression dans l'urètre égale à la pression abdominale, on peut alors écrire :

$$\frac{P_1}{\rho} = \frac{v_2^2}{2} \quad (12)$$

On en déduit le débit Q :

$$Q = s_u \sqrt{2 \left(\frac{P_1}{\rho} \right)} \quad (13)$$

Avec s_u l'aire de la lumière de l'urètre.

III. SIMULATION

Nous avons choisi de simuler le comportement d'un implant de stimulation de la vessie par SEF développé par Brindley et al. [7]. La figure 5 confronte les résultats de simulation avec les mesures expérimentales publiées par Brindley. Pour un même signal de stimulation, on obtient une réponse du muscle comparable qualitativement, aussi bien dans son allure que dans les ordres de grandeurs temporels. On peut voir que le modèle produit des sorties comparables qualitativement avec les mesures publiées.

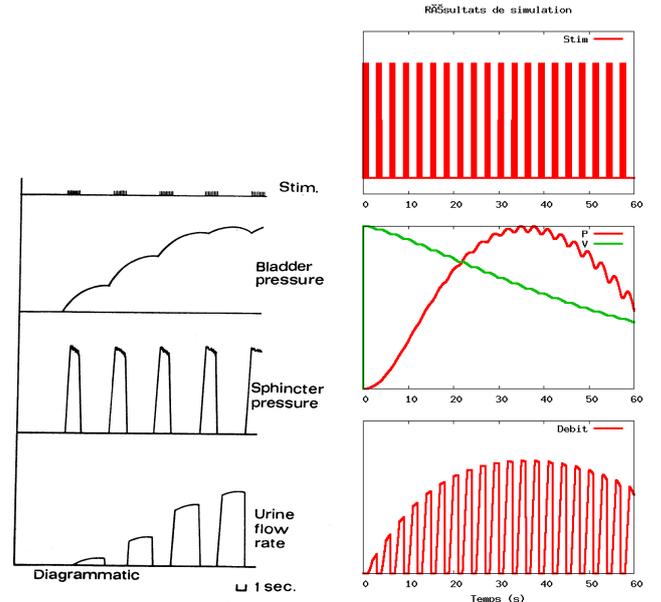


Fig. 5. A gauche les mesures expérimentales [7], et à droite la simulation avec le modèle.

IV. PERSPECTIVES

Pour affiner les prédictions du modèle et s'assurer de leur consistance, il est nécessaire de disposer de paramètres bien définis. Nous réaliserons donc des expérimentations animales et humaines pour poursuivre la validation du modèle et identifier ses paramètres. Cette identification utilisera la méthode récemment mise en oeuvre pour le muscle strié [8]. Nous pensons aussi étendre le modèle pour prendre en compte des méthodes de stimulation sélective en intégrant les travaux de l'équipe Odysée (Inria) [9] ainsi qu'un modèle numérique des nerfs stimulés.

RÉFÉRENCES

- [1] Hassan El Makssoud. *Modélisation et Identification des Muscles Squelettiques sous Stimulation Electrique Fonctionnelle*. PhD thesis, Université Montpellier 2, 2 December 2005.
- [2] C. M. Hai and R. A. Murphy. Cross-bridge phosphorylation and regulation of latch state in smooth muscle. *Am J Physiol*, 254(1 Pt 1) :C99–106, 1988.
- [3] M. Koenigsberger, R. Sausser, M. Lamboley, J.L. Beny, and J.J. Meister. Ca²⁺ dynamics in a population of smooth muscle cells : modeling the recruitment and synchronization. *Biophys J*, 87(1) :92–104, 2004.
- [4] A. F. HUXLEY. Muscle structure and theories of contraction. *Prog Biophys Biophys Chem*, 7 :255–318, 1957.
- [5] S.N. Yu, P.E. Crago, and H.J. Chiel. A nonisometric kinetic model for smooth muscle. *Am J Physiol*, 272(3 Pt 1) :C1025–39, 1997.
- [6] T. Arts, P. H. Bovendeerd, F. W. Prinzen, and R. S. Reneman. Relation between left ventricular cavity pressure and volume and systolic fiber stress and strain in the wall. *Biophys J*, 59(1) :93–102, 1991.
- [7] G. S. Brindley, C. E. Polkey, and D. N. Rushton. Sacral anterior root stimulators for bladder control in paraplegia. *Paraplegia*, 20(6) :365–81, 1982.
- [8] Mitsuhiro Hayashibe, Philippe Poignet, and David Guiraud. Non-linear identification of skeletal muscle dynamics with sigma-point kalman filter for model-based fes. In *ICRA'08 : International Conference on Robotics and Automation*, 2008.
- [9] Sabir Jacquir, Joan Fruitet, David Guiraud, and Maureen Clerc. Computation of the electrical potential inside the nerve induced by an electrical stimulus. In Joan Fruitet, editor, *Proc. 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society EMBS 2007*, pages 1711–1714, 2007.