



HAL
open science

Classification par réseaux de neurones dans le cadre de la scattérométrie ellipsométrique

Sabit Fawzi Philippe Zaki

► **To cite this version:**

Sabit Fawzi Philippe Zaki. Classification par réseaux de neurones dans le cadre de la scattérométrie ellipsométrique. Optique / photonique. Université de Lyon, 2016. Français. NNT : 2016LYSES070 . tel-02078679

HAL Id: tel-02078679

<https://theses.hal.science/tel-02078679>

Submitted on 25 Mar 2019

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

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



N°d'ordre NNT :

THESE de DOCTORAT DE L'UNIVERSITE DE LYON

opérée au sein de

L'UNIVERSITÉ JEAN MONNET DE SAINT ÉTIENNE

**Ecole Doctorale N° 488
Science-Ingénierie-Santé**

Spécialité de doctorat :
Discipline : Optique photonique Hyperfréquences

Soutenue publiquement le 12 Décembre 2016 par :
ZAKI SABIT FAWZI PHILIPPE

Classification par réseaux de neurones dans le cadre de la scattérométrie ellipsométrique

Devant le jury composé de :

Cécile Gourgon Directrice de recherche à Grenoble (CNRS-LTM), Présidente, Rapporteur

Gérard Granet Professeur à l'Université Blaise Pascal, Rapporteur

Amir Mougache Maitre Assistants, Université de Ndjamen, Examineur

Bernard Bayard HDR Université Jean Monnet, Directeur de thèse

Stéphane Robert Maitre de conférences, Laboratoires Hubert Curien, Encadrant

Laboratoire Hubert Curien, UMR.CNRS.5516.Saint Étienne



Dédicace

A

Mon père, Sabit Philippe

Ma mère Awadia Ali

Mes tantes, Sabit Naima et Mabrouka Ali

Aux défunts, Hissein Koty Yacoub et Mahamat Moussa Oumar

A ceux qui ont cru en l'accomplissement de ce travail

A celui animé par le désir d'apprendre et de découvrir

Remerciement

Cette thèse est réalisée au Laboratoire Hubert Curien de l'université Jean Monnet de Saint Etienne en partenariat avec l'INSTA du Tchad. Ce fut pour moi une expérience particulièrement enrichissante aussi bien sur le plan professionnel que sur le plan personnel.

Je tiens à remercier mon Directeur de thèse Bernard Bayard d'avoir accepté de diriger ce travail. Mes remerciements s'adressent en particulier à Stéphane Robert, Maître de conférences au LaHC pour la disponibilité d'encadrement, la pertinence et surtout la patience montrée tout le long de la thèse. Je ne pourrais avoir meilleur encadrement.

Je remercie Pr. Gérard Granet, professeur de l'Université Blaise Pascal et Cécile Gourgon, directrice de recherche CNRS Grenoble d'avoir accepté de rapporter ce travail sans oublier également Dr Amir Mougache d'avoir accepté d'évaluer ce travail.

Je ne saurais citer : Pr. Mahamoud Youssouf, Directeur du CNRD pour la confiance en m'ouvrant les voies de la thèse ; Dr. Malloum Soultan Directeur de l'INSTA pour la sagesse et qui a toujours eu la certitude de l'accomplissement de ce travail ; Pr. Jean Jacques Rousseau de m'avoir accueilli dans le laboratoire. Dr. Ismail Elkalyoubi pour les mesures réalisées et l'accompagnement durant le travail de thèse.

Ma reconnaissance va aux frères et amis Dr. Mahamat Borgou et Dr. Idriss Saleh pour le soutien inconditionnel durant cette thèse.

Je ne saurais oublier toute ma famille et ma fiancée pour l'encouragement et le soutien inconditionnel dans la poursuite de mes études et surtout d'avoir pu supporter toutes ces années d'absence.

Mes remerciements s'adressent également :

A Toute l'équipe du laboratoire et en particulier les permanents J.P Chatelon, F. Royer et notre secrétaire Nathalie P. toujours aussi serviables ;

Aux Docteurs et Doctorants Abderahim Awat, Mahamat H. B, Khamis Y, Badreddine O, Billal B., Arafat O, Ouzer N, Mahamout B, Mahamat Taha, Ahmat Idriss G, Koularambaye, Danoumbé, Eric D., Boukhari M., Djido A., Basma B. pour les moments d'ambiances dans le laboratoire ;

Ma petite sœur Mariam Djamila qui a porté une attention particulière à la lecture de ce document ;

A Sonia Safar du SCAC ainsi qu'à toute l'équipe de Campus France pour leur soutien

A tous les professeurs dont les enseignements m'ont permis d'entamer avec aisance ce travail ;

Enfin j'ai également le souvenir inoubliable des jeunes actifs de ST-Etienne en particulier Ousmane, Claire, Charlotte et Rodolphe, car les années de thèse ne se vivent pas uniquement au laboratoire.

Que tous ceux qui de près ou de loin ont contribué à la réussite de ce travail trouvent à travers ces lignes l'expression de ma profonde gratitude.

Introduction Générale	6
Chapitre I Scattérométrie	9
I.1 Contexte de la métrologie optique.....	9
I.1.1 Les méthodes microscopiques	9
I.1.1.1 Le Microscope Electronique à Balayage.....	9
I.1.1.2 Microscopie à Force Atomique	11
I.1.2 Une alternative aux méthodes classiques : la scattérométrie	13
I.1.2.1 Pourquoi la scattérométrie ?.....	13
I.1.2.2 Une méthode de mesure indirecte	14
I.1.2.3 Etat de l'art	14
I.1.3 Exemple de la lithographie optique	16
I.1.3.1 Détection d'un défaut identifié.....	18
I.1.3.2 Reconnaissance de type de profils géométriques.....	19
I.2 Fonctionnement de la scattérométrie ellipsométrique	22
I.2.1 La structure périodique test.....	22
I.2.2 Signature ellipsométrique	25
I.2.2.1 Principe de l'ellipsométrie classique.....	25
I.2.2.2 Appareil de mesure.....	27
I.2.3 Problème direct de la diffraction.....	31
I.2.3.1 Etat de l'art des méthodes	31
I.2.3.2 Une méthode performante : la Multilayer modal method by fourier expansion	32
I.2.3.3 Modélisation des formes géométriques utilisées.....	34
I.2.4 Problème inverse en scattérométrie	42
I.2.4.1 Définition du problème inverse	42
I.2.4.2 Méthodes d'optimisation classiques.....	43
I.2.4.3 Les méthodes utilisées en scattérométrie	45
Chapitre II Classification Statistique des Formes.....	51
II.1 Le domaine de la reconnaissance de formes	51
II.1.1 Généralité	52
II.1.2 Prétraitement des données.....	54
II.1.3 Une mesure de proximité : La distance.....	55
II.2 Approche indirecte de classification	58
II.2.1 Estimation paramétrique des densités de probabilité : Le classifieur gaussien	59
II.2.2 Estimation non paramétrique des densités de probabilité : La méthode des K plus proches voisins (K-ppv)	60
II.3 Approche directe de classification	62
II.3.1 Les Séparateurs à Vastes Marges.....	62
II.3.1.1 Principe des SVM linéaires.....	62
II.3.1.2 SVM non linéaire	66
II.3.2 Méthode connexionniste : Les Réseaux de Neurones.....	68
II.3.2.1 Généralités sur les réseaux de neurones.....	68
II.3.2.1.1 Historique des R.N	69
II.3.2.1.2 Applications des R.N.....	69
II.3.2.1.3 Définition du neurone artificiel	70
II.3.2.1.4 La mise en réseau	73
II.3.2.1.5 Apprentissage des R.N	75
II.3.2.2 Le perceptron Multicouche	77
II.3.2.2.1 Présentation et architecture	77
II.3.2.2.2 Apprentissage du PMC.....	79
II.3.2.2.3 Le PMC dans le cadre de la classification supervisée.....	83
II.3.2.3 Réseau de Kohonen.....	85
II.3.2.3.1 Les cartes auto organisatrices.....	86
II.3.2.3.1.1 Architecture d'une SOM et principe	86
II.3.2.3.1.2 Apprentissage d'une SOM	88
II.3.2.3.2 Learning Vector Quantization	92
II.3.2.3.2.1 Architecture du LVQ.....	92
II.3.2.3.2.2 Apprentissage du LVQ.....	94

Chapitre III	Le réseau de neurones comme outil de détection	96
III.1	Détection d'une couche résiduelle	96
III.1.1	Cadre de l'étude	96
III.1.1	Visualisation par projection des signatures optiques	97
III.1.1.1	SOM sans classification	97
III.1.1.2	SOM comme outil d'aide à la décision	101
III.1.1.3	Optimisation de la carte SOM	104
III.1.1.3.1	Etude de différentes architectures	104
III.1.1.3.2	Tests sur des échantillons théoriques	106
III.1.1.4	Test sur échantillon réel	108
III.1.2	Utilisation de la classification neuronale	110
III.1.2.1	Critères de performance pour la classification	110
III.1.2.2	Classification par LVQ	112
III.1.2.3	Classification par PMC	116
III.1.2.4	Test sur échantillon réel	119
III.2	Identification de forme de profil géométrique pour la construction d'un modèle scatterométrique	120
III.2.1	Cadre de l'étude	120
III.2.1.1	Domaine d'application visée	120
III.2.1.2	Définition des différentes classes : les profils géométriques retenus	121
III.2.2	Etude du défaut de modélisation	122
III.2.2.1	Hypothèse du créneau parfait	124
III.2.2.2	Hypothèse du modèle trapézoïdal	130
III.2.2.3	Simulation d'un profil test complexe	135
III.2.3	Projection des signatures	139
III.2.3.1	SOM sans classification	140
III.2.4	Méthode basique de détermination du profil géométrique	149
III.2.4.1	Comparaison des signatures reconstruites	149
III.2.4.2	Méthode de pondération de profil	150
III.2.4.3	Utilisation d'un profil complexe	151
III.2.4.3.1	Paramètres du modèle générique	151
III.2.4.3.2	Caractérisation scatterométrique classique fonctionnant sur modèle de profil complexe	152
III.2.5	Identification de type de profil par LVQ	157
III.2.6	Identification de type de profil par PMC	160
III.2.6.1	Résultats de la classification à 3 et 4 classes	160
III.2.6.2	Résistance de la classification PMC au bruit de mesure	164
III.2.7	Mesure sur échantillon réel	165
III.2.7.1	Identification expérimentale du type de profil	166
III.2.7.2	Caractérisation scatterométrique	168
III.2.7.3	Comparaison des différentes caractérisations scatterométriques	169
Conclusion générale		173
Bibliographie		176
ANNEXE A : Compléments à l'étude de défaut de modélisation de profil		184
ANNEXE B : Etude du corpus B de 3 classes		190
ANNEXE C : Etude d'un PMC fonctionnant sur une signature UV		193

Introduction Générale

La microélectronique et la nanotechnologie ont connu des avancées remarquables ces dernières années. Les puces électroniques figurent dans tous les appareils de notre quotidien (téléphonie mobile, électroménager, transport, médecine, ...). Cette évolution est étroitement liée à la miniaturisation des composants. Il faut fabriquer des circuits à forte densité d'intégration, économiques en termes de coût et d'énergie et surtout plus performants.

Ainsi, l'industrie de la microélectronique suit les recommandations de l'ITRS (International Technology Roadmap for Semiconducteur) sur les dimensions critiques à atteindre. En effet, le rapport dicté par ce dernier sert de feuille de route pour les procédés de fabrication ainsi que les techniques de métrologie associées. Le contrôle des dimensions à l'échelle nanométrique devient donc le plus grand défi des méthodes de caractérisation. Ainsi, il faut s'assurer que toutes les étapes mises en œuvre dans un procédé de fabrication s'enchaînent correctement et que les structures correspondent bien aux résultats prévus.

Les méthodes utilisées jusqu'alors sont de nature microscopique. On peut citer la Microscopie à Force Atomique (AFM) et la Microscopie Electronique à Balayage (MEB ou SEM pour Scanning Electron Microscope). L'AFM est limitée par la taille de la pointe qu'elle utilise pour imager la surface de l'échantillon. Elle est ainsi confrontée à la problématique de la fabrication de celle-ci dans des dimensions de plus en plus fiables. Le MEB, quant à lui, est destructif bien qu'il puisse fournir une image réelle de l'échantillon. La qualité de la caractérisation est ainsi liée à la qualité de l'image obtenue. Par ailleurs, ces deux techniques sont locales et incompatibles avec des applications temps réel. Le développement de nouvelles techniques de caractérisation a ainsi vu le jour depuis plusieurs années. Les méthodes optiques telles que la scattérométrie sont aujourd'hui une alternative prometteuse pour répondre aux besoins imposés par cette course à la miniaturisation. Cette dernière technique qualifiée d'indirecte implique la mesure du spectre de diffraction d'une structure périodique associée à la résolution d'un problème inverse par des méthodes numériques. Dans ce domaine, notons que les réseaux de neurones se sont peu à peu imposés comme l'un des outils mathématiques incontournables grâce notamment à la rapidité du traitement de l'information. Toutefois, l'ensemble de ces méthodes scattérométriques nécessitent un certain nombre d'hypothèses pour assurer la résolution du problème inverse et notamment la connaissance de la forme géométrique de la structure à tester. Le modèle de structure supposé conditionne la qualité même de la caractérisation. Ainsi, si l'hypothèse de forme est fautive et ne correspond pas à la

structure étudiée, on parle mathématiquement de problème mal posé et cela peut induire des résultats erronés, et ce, quelle que soit la qualité de la méthode utilisée pour la résolution du problème inverse. Ceci reste le point faible de toute méthode scattérométrique. Dans ce cadre, nous invitons le lecteur à découvrir dans cette thèse un nouvel apport original des réseaux de neurones comme outils d'aide à la décision en amont de toute méthode de caractérisation.

Cette thèse s'inscrit donc dans le cadre de la reconnaissance par réseaux de neurones des formes géométriques de l'échantillon à tester par la simple signature optique utilisée dans toute étape de caractérisation scattérométrique. Les travaux sont réalisés au laboratoire Hubert Curien dans l'équipe « surface complexe » qui depuis quelques années a développé un savoir-faire dans l'utilisation de méthodes mathématiques empruntées au domaine de l'intelligence artificielle comme les réseaux de neurones pour le contrôle non destructif et plus particulièrement dans le domaine de la scattérométrie.

Le travail présenté apparaîtra aux lecteurs de diverses natures : physique, informatique, mathématiques appliquées. Ce rapport est constitué principalement de 3 chapitres.

Le chapitre I est entièrement consacré à la méthode scattérométrique avec une partie consacrée spécifiquement au besoin en termes de caractérisation dans le milieu industriel. D'abord, il situe la technique dans le cadre de la métrologie optique puis détaille les différents principes de fonctionnement. Ainsi, nous définissons les grandeurs mesurées et la manière selon laquelle elles peuvent être simulées numériquement à partir d'un modèle géométrique (problème direct). Enfin, nous présentons les méthodes mathématiques utilisées dans la résolution du problème inverse permettant de retrouver les caractéristiques géométriques d'une structure diffractante.

Le chapitre II est consacré aux principes généraux et aux bases de la classification automatique. Nous décrivons les méthodes les plus connues dans la littérature. Ensuite, nous nous focaliserons sur l'utilisation des réseaux de neurones dans ce domaine. Nous définissons le fonctionnement général d'un neurone ainsi que l'architecture neuronale déjà utilisée classiquement en scattérométrie. Puis, nous insisterons davantage sur les spécificités des réseaux dédiés à la classification qui constitue le principal outil utilisé dans cette thèse.

Le chapitre III concerne l'utilisation des réseaux de neurones comme outil de détection dans le cadre précis du contrôle optique non destructif. Il contient les résultats obtenus lors de nos travaux. Le chapitre est scindé en deux parties. La première concerne la détection automatique d'un défaut lithographique particulier lié à la présence d'une couche résiduelle de

résine au fond des sillons. La seconde partie, quant à elle, relate les résultats obtenus dans le cadre de la sélection de modèles géométriques en amont d'un processus classique de caractérisation scattérométrique.

Chapitre I : Scattérométrie

Ce premier chapitre présente les fondements théoriques de la scattérométrie. Il est composé de deux grandes parties : la première partie dresse un état de l'art de cette méthode et détaille les différentes applications dans le domaine du contrôle non destructif. Ainsi nous abordons les besoins industriels et les apports potentiels de la classification dans un procédé de fabrication et de caractérisation. Dans la seconde partie, nous nous focaliserons essentiellement sur la technique utilisée dans nos travaux. Ainsi, nous présentons la mesure ellipsométrique et la méthode numérique que nous avons choisie pour résoudre le problème direct de la diffraction. Enfin, nous passons en revue les méthodes de résolution de problème inverse qui permettent d'extraire l'information recherchée.

I.1 Contexte de la métrologie optique

Avec la miniaturisation, il est indispensable pour l'industrie micro électronique d'avoir recours à des techniques de métrologie rapides et performantes qui soient capables de déterminer avec précision les dimensions critiques afin d'assurer un contrôle rigoureux des composants fabriqués. Les techniques classiques de métrologie telles que l'AFM et le MEB sont jusqu'à présent largement utilisées dans le milieu industriel. Nous allons décrire brièvement le principe de ces deux méthodes.

I.1.1 Les méthodes microscopiques

I.1.1.1 Le Microscope Electronique à Balayage

Le MEB [1] [2] repose sur l'interaction matériau électron. Le principe consiste simplement à envoyer un faisceau d'électrons sur la surface à analyser. Des électrons « rétrodiffusés » de différentes natures sont émis et la détection de ces derniers offre la possibilité de connaître la typologie de la surface insolée. Au niveau du faisceau d'émission, les électrons, chargés négativement, sont extraits d'un filament chauffé ou d'une pointe portée à un très haut potentiel négatif. Ces derniers sont ensuite accélérés à forte énergie et focalisés par des bobines magnétiques. Le choc entre ces électrons et la surface du matériau crée des électrons secondaires ainsi que des électrons rétrodiffusés. C'est la détection des électrons secondaires qui permet de déterminer la composition chimique ou typographique de la

structure.

Lors de l'exposition du matériau au faisceau d'émission on observe une accumulation de charges dont l'énergie chauffe l'échantillon, entraînant des modifications des propriétés géométriques du motif. Généralement, la mise en œuvre d'une métallisation de l'échantillon permet d'éviter cet effet thermique. Cette observation est d'autant plus vraie dans le cas des résines. En effet, ces dernières sont constituées de polymères et donc de nature isolante.

On distingue 2 catégories de MEB : le CD - SEM et le XSEM (X pour cross section).

Le CD-SEM effectue une vue de dessus et permet également de s'affranchir de l'étape de métallisation et du clivage du matériau. L'échantillon peut être réutilisé par la suite, ce qui rend la méthode très populaire dans l'industrie pour le contrôle en ligne des sites de production. La mesure sur l'échantillon se fait de manière automatisée et le CD est extrait à partir de techniques de traitement d'images. Le CD-SEM permet également d'obtenir des informations sur la rugosité du motif. La rapidité de la mesure dépend du nombre d'images souhaité car le traitement associé nécessite des algorithmes qui consomment du temps. La mesure ne peut donc être obtenue rapidement. Cette méthode ne peut convenir pour des caractérisations ayant des contraintes temporelles importantes comme par exemple le contrôle dynamique.

Le XSEM s'effectue sur une section de matériau. Cette méthode est destructive, car il faut sectionner la plaque à l'endroit de la mesure. L'étape de métallisation est également nécessaire pour éviter le réchauffement et cela rend l'échantillon inutilisable par la suite. Contrairement au CD-SEM, le XSEM permet une caractérisation verticale de l'échantillon rendant notamment possible la caractérisation de couche sous-jacente et de profils.

La figure I.1 montre l'exemple d'images MEB d'un même réseau en silice de période 565 nm provenant de 2 appareils différents. La délimitation du profil n'est pas nette et l'on peut déceler une zone d'incertitude plus ou moins prononcée suivant la qualité de l'image.

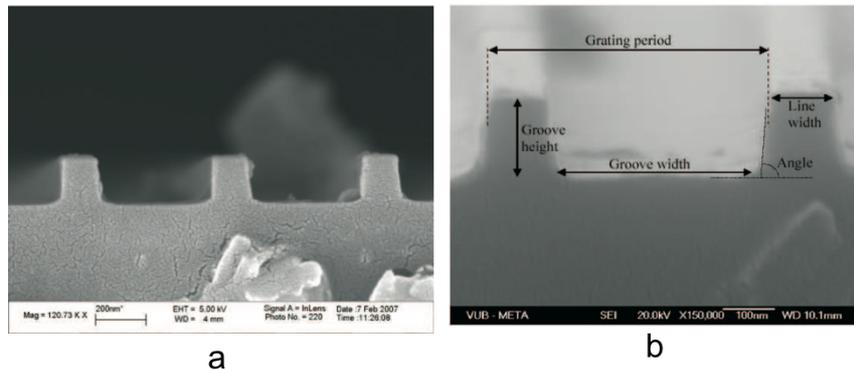


Figure I. 1 Exemple d'images d'un même réseau en silice provenant de deux MEB différents notés (a) (Department of Physics, Université de Joensuu, Finlande) et (b) (Department of Applied Physics and Photonics, Vrije Universiteit Brussel, Bruxelles) [3].

Malgré le fait qu'il fournisse une image réelle de l'échantillon sans connaissance a priori sur la structure, le MEB reste une méthode qui nécessite un équipement lourd et complexe, qui peut s'avérer dans certains cas, destructif et dont l'extraction du profil est fortement liée au traitement d'images associé.

I.1.1.2 Microscopie à Force Atomique

L'AFM permet de caractériser différents types d'échantillons (conducteur, semi-conducteur, isolant ou biologique) et peut être utilisée dans plusieurs types d'environnement (air, ultravide, liquide). La technique repose sur le principe d'attraction et de répulsion entre une pointe placée à l'extrémité d'une poutre et la surface du matériau à caractériser. Un système de photodiodes permet de capter la réflexion d'un laser focalisé sur la face arrière de la pointe. Le mouvement de la pointe est ainsi retranscrit sur un plan à 2 dimensions et permet de reconstituer le profil [4].

Il faut noter que la pointe présente l'élément essentiel qui détermine la qualité d'une mesure AFM. Il existe différents types de pointes en matière de composition selon le type d'utilisation. Plusieurs modes de fonctionnement permettent également d'effectuer une mesure AFM : contact, tapping ou non-contact. Pour le premier, la pointe est maintenue en contact permanent avec l'échantillon. Dans le deuxième, le contact est intermittent, la pointe oscille à une faible amplitude avec une fréquence proche de sa fréquence de résonance. Cela permet d'éviter la déformation de certains échantillons ou encore l'usure de la pointe qui peut se produire rapidement en mode contact. Enfin pour le troisième mode, la pointe est maintenue à une distance d'au moins 10 nm de l'échantillon et oscille également à une fréquence proche de sa fréquence de résonance.

La caractérisation des profils géométriques qui nous intéressent dans ces travaux est fortement limitée par la forme de la pointe. La figure I.2 illustre une caractérisation AFM par une pointe conique.

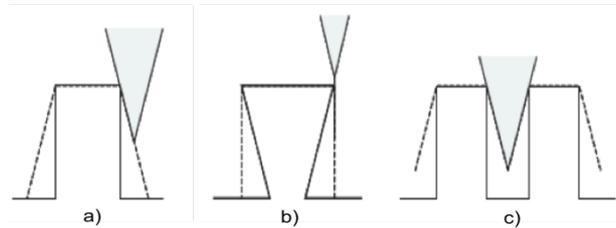


Figure I. 2 : Déformation du profil réel par une pointe AFM. Les traits en pointillés et en continus désignent respectivement le profil mesuré et le profil réel.

On observe que la forme du profil reconstruit est différente de celle du profil réel. Dans ce cas, il est donc nécessaire de réaliser un traitement qui consiste à déconvoluer la mesure en tenant compte de la forme supposée de la pointe à cet instant pour approcher au mieux la forme réelle (figure I.2 a). Pour le profil rentrant (figure I.2 b), la déconvolution n'est pas suffisante et le troisième cas (figure I.2 c) est lié à la résolution spatiale. La taille de la pointe ne permet pas de déboucher au fond des sillons et la déconvolution ne donnerait pas le résultat escompté.

Une variante de la méthode, appelée CD-AFM, utilise une pointe évasée dite en « patte d'éléphant » qui permet de mesurer la topologie des flancs de profil de manière beaucoup plus précise afin de pallier la difficulté mentionnée ci-dessus. Elle donne une information sur la rugosité et les ondulations verticales du motif (figure I.3).



Figure I. 3 : Schématisation d'une reconstruction de profil avec un CD AFM [5].

La verticalité du profil est ainsi reproduite beaucoup plus précisément. Par ailleurs, de la même façon que l'AFM classique il faut déconvoluer la forme de la pointe sur le résultat brut pour obtenir l'allure exacte du profil.

Ces techniques classiques présentées permettent de connaître la forme du profil (CD, hauteur, pente, rugosité, ...) et dans certains cas la hauteur des couches sous-jacentes. Ce sont des méthodes parfois destructives et elles sont assez lourdes d'utilisation et donc difficilement applicables au contrôle rapide en ligne de production. Or, le challenge pour l'industrie microélectronique consiste clairement à mesurer rapidement des dimensions critiques tout en conservant des instruments de métrologie performants facilement intégrables. L'attention se porte alors sur la scattérométrie qui constitue une réelle alternative idéale pouvant relever le défi de caractérisation des futurs nœuds technologiques.

I.1.2 Une alternative aux méthodes classiques : la scattérométrie

I.1.2.1 Pourquoi la scattérométrie ?

La scattérométrie est une technique optique de métrologie non destructive basée sur l'analyse de la lumière issue de la structure à tester. Si cette dernière est périodique (par exemple un réseau de diffraction), la lumière est alors diffractée. L'ensemble des mesures nécessaires à la caractérisation est communément appelé signature optique. La scattérométrie est dite statique si elle se limite à la mesure ponctuelle d'un échantillon et dynamique dans le cas où elle intervient dans le suivi de procédés en temps réel.

Par rapport aux techniques classiques de métrologie utilisées dans l'industrie de la microélectronique, la scattérométrie présente l'avantage d'offrir la mesure en champ lointain, particulièrement adaptée pour un contrôle au cours de la fabrication. De plus, l'instrumentation nécessaire est simple et le traitement rapide répondant aux contraintes temporelles imposées pour un suivi en temps réel. La scattérométrie est capable d'extraire des dimensions latérales dont la résolution est basée sur la sensibilité de la signature aux variations du profil. Enfin, contrairement à l'AFM et au MEB qui sont des méthodes locales, le résultat d'une mesure scattérométrique est une moyenne faite sur l'ensemble des motifs illuminés par le faisceau optique.

Par ailleurs, la mesure scattérométrique permet de contrôler plusieurs étapes dans la fabrication. On peut par exemple commencer par caractériser des couches minces et en second lieu on peut réaliser un contrôle dimensionnel. Différentes applications ont montré, selon la configuration employée, les possibilités de mesure de nombreuses grandeurs physiques (épaisseur de couches, indices optiques, paramètre de forme géométrique [6] [7] [8]). Elle peut également permettre d'effectuer des cartographies afin de contrôler l'homogénéité des échantillons. Ainsi, la scattérométrie s'impose comme une méthode non destructive, non

invasive et non locale offrant également la possibilité de mesures in situ en temps réel.

I.1.2.2 Une méthode de mesure indirecte

La scattérométrie appartient à la famille des techniques métrologiques qualifiées d'indirectes. En effet, contrairement aux techniques microscopiques, elle ne permet pas de mesurer directement la valeur des paramètres recherchés.

Concrètement, elle se déroule en deux étapes : la mesure de la signature correspondant à la structure à tester et le traitement de celle-ci qui consiste à remonter aux paramètres géométriques recherchés par la résolution du problème inverse de la diffraction (figure I.4).

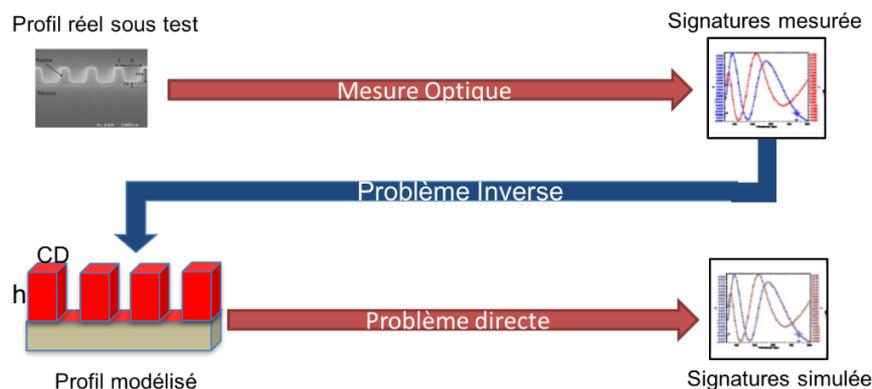


Figure I. 4 : Schéma général de principe de la scattérométrie

On définit le problème direct comme l'obtention de la signature par calcul électromagnétique à partir d'un modèle géométrique supposé de l'échantillon. Le problème inverse quant à lui consiste à déterminer le profil géométrique de l'échantillon à partir de la signature optique.

Ce problème est assez complexe et ne possède pas de solution analytique. Afin de s'assurer de l'unicité de la solution, on fixe un certain nombre d'hypothèses comme la connaissance des indices optiques des matériaux mis en jeu ou le modèle géométrique définissant la forme de la structure à tester. Typiquement, c'est ce qui est réalisé pour une caractérisation ellipsométrique classique dans laquelle on fixe l'hypothèse structurale de l'échantillon (nombre et type de couches, épaisseurs connues, etc ...). Les paramètres géométriques recherchés lors de la résolution du problème inverse sont donc déterminés dans le cadre des hypothèses retenues.

I.1.2.3 Etat de l'art

La signature optique est généralement constituée soit de mesures d'intensité soit de

grandeurs ellipsométriques liées à l'appareil de mesure utilisé. Il existe ainsi différents types de dispositifs expérimentaux pouvant permettre la mesure scattérométrique.

Deux types d'appareils sont les plus largement utilisés : l'ellipsomètre et le réflectomètre. On parlera ainsi de scattérométrie ellipsométrique si l'analyse prend en compte la phase de l'onde réfléchie et de scattérométrie réflectométrique dans le cas où l'analyse se rapporte à une mesure d'intensité de la lumière réfléchie. Les conditions expérimentales sont fixées à l'avance : angle d'incidence fixe (généralement incidence normale), ou variable, mono ou multi longueur d'ondes. Ainsi, on peut trouver des appareils de mesures qui utilisent différentes méthodes expérimentales. On peut par exemple utiliser les intensités de diffraction en fonction de la longueur d'onde ou les intensités diffractées en fonction de l'angle d'incidence.

Pour la réflectométrie à angle variable, la signature optique est constituée d'une série de mesures d'efficacités de diffraction dans l'ordre zéro à une longueur d'onde et une polarisation données en fonction de l'angle d'incidence [9] [10] [11]. Dans ce cas, on parle également de goniométrie ou de scattérométrie 2θ [12] [13]. La signature optique peut également être constituée de mesures d'efficacités en fonction d'un angle azimutal ϕ définissant une incidence conique [14] sur l'échantillon. L'inconvénient de cette technique est la complexité du système de montage mécanique et la difficulté de maintenir un bon alignement optique.

Concernant la réflectométrie à incidence normale, on mesure la variation de réflectivité pour une polarisation donnée en maintenant l'angle d'incidence fixe et en faisant varier les longueurs d'onde [15]. Cette technique est notamment utilisée pour un contrôle *in situ*. En effet, le contrôle d'épaisseur des couches minces déposées se fait grâce à la mesure des interférences dans les couches.

L'ellipsomètre (spectroscopique ou non) mesure, quant à lui, les variations de polarisation résultant de l'interaction entre le faisceau incident et l'échantillon. La signature optique se présente sous la forme d'un spectre en longueurs d'onde ou, plus rarement en angle d'incidence de la quantité mesurée [16] [17] [18] qui dépend du type d'ellipsomètre utilisé. Dans tous les cas, il s'agit de mesurer les variations de l'amplitude complexe de l'onde incidente après réflexion sur l'échantillon. Notons cependant la particularité de l'ellipsométrie de Muller [19] qui utilise un formalisme différent de celui de Jones utilisé en ellipsométrie classique. Ce formalisme dit de Stockes permet de traiter complètement les changements d'état de polarisation de la lumière par de simples mesures d'intensités dans des configurations

polarimétriques précises [20].

En pratique, la mesure de la signature optique est ensuite associée à une méthode de traitement pour retrouver les paramètres de la structure à caractériser. Cette étape est nommée résolution du problème inverse. Elle se fait indépendamment des types de mesures cités ci-dessus. La résolution du problème inverse peut se faire soit par l'emploi d'une méthode d'optimisation (classique de type Levenberg Marquardt ou plus sophistiquée de type métaheuristique comme les algorithmes génétiques) soit par une méthode de régression (linéaire [21] ou non linéaire [15]).

Les méthodes d'optimisation se basent sur la recherche de l'erreur minimale entre une signature théorique simulée et celle mesurée directement sur l'échantillon. Parmi les méthodes couramment utilisées en scattérométrie on trouve classiquement les méthodes non linéaires de second ordre [22] et l'utilisation de bibliothèques [23] [24] qui peuvent s'apparenter à une méthode d'optimisation énumérative. On trouve également l'emploi de métaheuristicques qui couvrent l'ensemble des méthodes d'optimisation empruntées au domaine de l'intelligence artificielle telles que les algorithmes génétiques [25]. Nous détaillerons plus loin le principe de fonctionnement de ces méthodes dans le paragraphe I.2.4.2.

Les méthodes de régression quant à elles se basent sur un nombre fixé d'exemples connus pour retrouver les paramètres géométriques de la structure à tester. Les premières méthodes à être utilisées furent les méthodes de régression linéaire [21] [26]. Ensuite des méthodes de régression non linéaire ont vu le jour [27] [13]. Les performances de celles-ci sont nettement meilleures mais dépendent fortement du modèle non linéaire supposé dans la régression. Les réseaux de neurones, que nous utilisons dans ces travaux, peuvent s'apparenter à une méthode de régression non linéaire qui ne nécessite pas cette connaissance *a priori*.

I.1.3 Exemple de lithographie optique

L'industrie de la microélectronique repose essentiellement sur la maîtrise de la fabrication des transistors. En effet, le transistor MOS est le composant actif le plus utilisé dans les circuits intégrés. Il constitue l'élément de base des dispositifs à haute densité d'intégration. Selon leur niveau d'intégration, les transistors permettent de réaliser des processeurs ou des mémoires RAM (Random Acces Memory). La figure I.5 illustre un schéma basique d'un transistor.

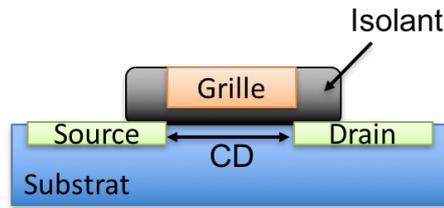


Figure I. 5: Représentation schématique simple d'un transistor MOS

La grille permet de contrôler le courant entre la source et le drain. Les électrons entrent par la source, passent dans un canal commandé par la grille et sortent par le drain. L'isolant a pour rôle d'isoler électriquement la grille du reste du transistor. Si cette dernière n'est pas activée électriquement, l'information est bloquée. Les performances du transistor dépendent de la taille de la grille. Cette dernière correspond à la dimension critique (CD) évoquée précédemment. Plus elle est petite et moins les électrons ont de chemin à parcourir. Le transistor est alors plus rapide. L'accroissement des performances du transistor est donc intimement lié à la miniaturisation.

Dans le processus de fabrication dédié, la lithographie optique est une des méthodes utilisées y compris pour la réduction des dimensions de motifs [28]. Sans entrer dans les détails, nous pouvons nous attarder brièvement sur son fonctionnement. Le système expérimental est généralement composé d'une source, d'un masque, d'un système optique de projection, d'un film de résine photosensible et d'un substrat sur lequel une ou plusieurs couches vont pouvoir être déposées. De nos jours, l'industrie utilise majoritairement une source de photons de longueur d'onde 193 nm, on parle alors de photolithographie 193 nm.

Le principe de la photolithographie consiste d'abord à déposer la résine photosensible sur le substrat (par exemple en silicium) et d'imager sur celle-ci les motifs inscrits sur le masque via l'optique de projection en modifiant ses propriétés de solubilité. Ensuite, après révélation dans un développeur (solution basique) la zone exposée va subsister (résine négative) ou être éliminée (résine positive). Enfin, le transfert du motif de résine à travers le substrat est réalisé par gravure qui peut être de nature sèche ou humide. Cette dernière doit se faire en respectant les dimensions initiales du motif de résine fixées par la lithographie. On distingue différents types de gravures en lithographie. Si la gravure de la résine et du matériau s'effectue à la même vitesse, la gravure est dite non sélective. Sinon elle est qualifiée de sélective. La figure I.6 illustre les différentes étapes subies par la résine lors d'un procédé lithographique.

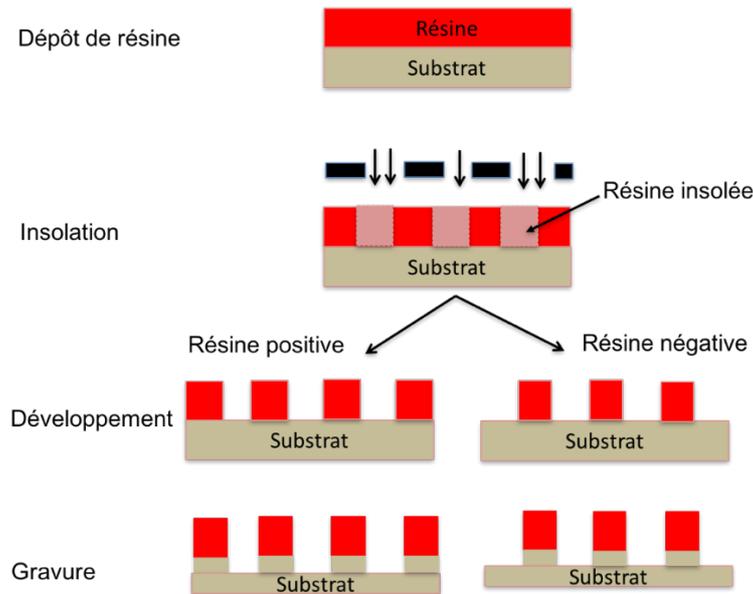


Figure I. 6 : Schéma du principe général des différentes étapes d'un procédé lithographique. Dans la phase de développement l'image de gauche montre un procédé utilisant une résine positif et celle de droite négative.

Dans le processus de fabrication des circuits, l'étape de lithographie est donc cruciale, étant donné que c'est elle qui définit les dimensions des motifs à imprimer dans le substrat. Elle doit être parfaitement maîtrisée et contrôlée afin d'écarter certaines dérives ou défauts qui peuvent survenir en cours de production. Pour valider nos travaux, nous avons choisi de définir 2 exemples typiques de défauts qui peuvent intervenir dans un tel procédé. Toutefois, la méthode développée ici n'est pas limitée à ce cadre et peut facilement s'étendre à d'autres besoins.

I.1.3.1 Détection d'un défaut identifié

L'un des principaux défauts qui peut apparaître lors de l'étape de lithographie est celui de la présence d'une couche résiduelle de résine. Ce défaut peut induire des conséquences catastrophiques sur l'étape de gravure. En effet, tout masque de résine ne débouchant pas sur le substrat est incompatible pour l'étape de gravure ultérieure. Par exemple, dans le cas d'une gravure sélective, la présence d'une couche résiduelle au-dessus du substrat conduit à fausser complètement le transfert de la forme désirée.

Pour ce type de défaut, la caractérisation complète du profil ne relève pas d'une importance capitale. Seul le fait de savoir qualitativement s'il y a présence de résine suffit dans le contrôle du procédé. Une réponse à cette problématique consiste en la création d'un dispositif de détection automatique de ce type de défaut. Une solution est envisagée dans cette thèse en

introduisant la notion de classification dans le processus des étapes photolithographiques. En effet, l'idée est de pouvoir détecter ce défaut à partir de la signature optique mesurée sur la structure lors du processus de fabrication à l'aide d'un réseau de neurones utilisé en mode classifieur afin de pouvoir procéder ou non à l'étape ultérieure de gravure. Le principe du système proposé est illustré sur la figure I.7.

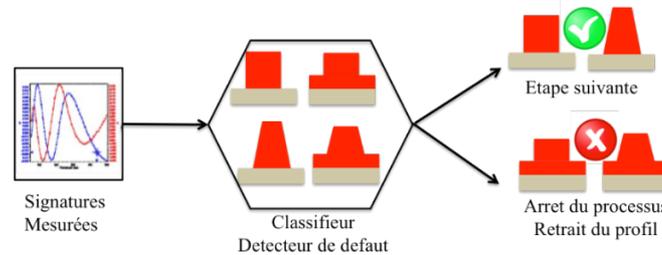


Figure I. 7 : Représentation schématique de la classification pour la détection d'un défaut identifié lors d'un procédé lithographique. Les échantillons possédant une couche résiduelle de résine sont détectés et écartés du processus.

Notons que pour certaines techniques comme la nanoimpression par exemple, la couche résiduelle est laissée volontairement au fond du réseau après pressage du moule dans la résine pour éviter tout contact entre ce dernier et le substrat [29] [30]. Elle doit être constante tout le long de la surface de pressage afin de garantir une gravure uniforme dans le substrat. La couche sera nettoyée ensuite avec des techniques adéquates. Les échantillons utilisés pour la phase expérimentale de nos travaux sont fabriqués par cette technique.

I.1.3.2 Reconnaissance de type de profils géométriques

Un autre problème rencontré fréquemment lors de l'étape de lithographie concerne généralement la déviation de la forme géométrique obtenue par rapport à celle désirée. Le résultat est en effet plus ou moins fidèle à la forme attendue (généralement créneau). Les raisons peuvent être de diverses natures comme un défaut de calibration par exemple mais pas seulement. Nous citons quelques exemples dans les paragraphes qui suivent :

Dans une grande majorité des cas, le substrat (en Silicium) est très réfléchissant aux différentes longueurs d'onde d'expositions actuelles utilisées en lithographie. Des interférences se créent entre les ondes incidentes et les ondes réfléchies pendant l'exposition et provoquent des différences d'intensité lumineuse dans le film de résine, induisant des motifs avec des déformations périodiques après développement. L'ajout d'un matériau absorbant appelé BARC (Bottom Anti reflective Coating) permet d'éviter ce phénomène de réflexion. Toutefois, l'utilisation du BARC, malgré son importance, complique l'empilement des couches et

nécessite des étapes de gravures supplémentaires [31].

La gravure humide est une méthode de gravure sélective qui consiste à tremper le substrat et son masque de résine dans une solution. Cette dernière est choisie en fonction de la nature du substrat, car la résine doit rester inerte à la solution. Ainsi, seules les zones non protégées par la résine vont subir l'attaque chimique. L'inconvénient de cette technique est sa nature isotrope. En effet, la solution chimique a tendance à attaquer le matériau dans toutes les directions, ce qui engendre une déformation des flancs des motifs. Ainsi il est impossible de conserver rigoureusement la forme initiale du masque.

La gravure sèche est le mode de gravure le plus utilisé pour les réseaux submicroniques. Le matériau à graver est soumis à un faisceau d'électrons en présence d'un plasma. Ce dernier est défini comme un milieu gazeux ionisé neutre qui contient des électrons, ions positifs et négatifs ainsi que des espèces neutres. On distingue trois types de gravure plasma selon la nature de l'attaque :

- IBE (Ion Beam Etching), cette attaque est de nature physique. La matière est arrachée par le faisceau d'ions composé d'éléments n'ayant aucune interaction chimique avec le matériau. Généralement on utilise des gaz rares comme l'argon.
- RIBE (Reactive Ion Beam Etching), cette attaque est de nature physico-chimique. En effet, le bombardement des ions ainsi que la réaction du plasma avec le matériau provoquent l'arrachement de la matière.
- RIE (Reactive Ion Etching), l'attaque est de nature chimique pour cette technique. Le faisceau d'ions incident ne sert qu'à apporter l'énergie nécessaire pour que le processus de gravure ait lieu.

La gravure sèche présente l'avantage d'être anisotrope et de graver n'importe quel type de matériau, mais la difficulté est de trouver le plasma et les ions adéquats à une bonne condition de gravure. En effet, l'interaction entre le plasma et les résines reste mal connue et peut engendrer des modifications dans les résines photosensibles. De plus les différents paramètres de gravure tels que la vitesse d'accélération des ions, choix du gaz réactif, ou l'inclinaison du substrat ont une grande influence sur la forme finale du motif.

La figure I.8 illustre quelques défauts de forme rencontrés à l'issue de la gravure.

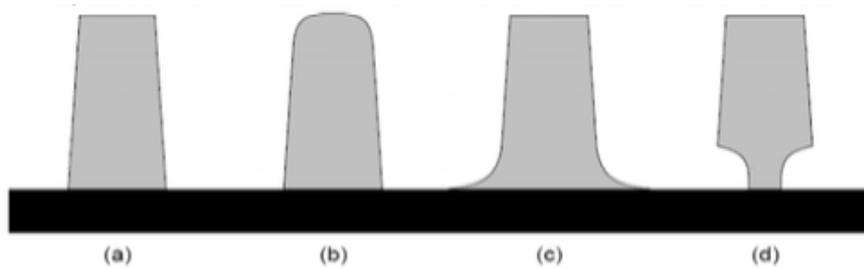


Figure I. 8 : Défauts rencontrés suite à la gravure, **(a)** défaut de pente, **(b)** défaut d'arrondi, **(c)** effet de sous gravure (pied) et **(d)** effet de la sur-gravure (« notch») [32]

Rappelons que dans le cadre d'une caractérisation scattérométrique, la résolution du problème inverse nécessite la connaissance préalable de la forme géométrique de la structure à tester (cf paragraphe I.1.2.2). Cette hypothèse classique de départ consistant à fixer le profil géométrique le plus probable est l'essence même des méthodes d'optimisation paramétrique, quel que soit le niveau de complexité du problème. Les paramètres recherchés sont donc déterminés dans le cadre des hypothèses retenues. Ainsi, pour un motif ayant dérivé de la forme attendue le problème devient mal posé car une des hypothèses n'est plus vérifiée ; ce qui peut conduire à une caractérisation erronée. Ce problème est d'autant plus fréquent que le procédé de gravure est mal maîtrisé ou a subi un dérèglement causé par une dérive systématique ou ponctuelle.

La solution proposée dans cette thèse consiste à utiliser la signature optique pour obtenir une première information qualitative sur les hypothèses structurales à choisir lors de la phase de résolution du problème inverse dans le cadre de la caractérisation scattérométrique classique. La figure I.9 illustre dans le cas général le système proposé dans cette thèse.

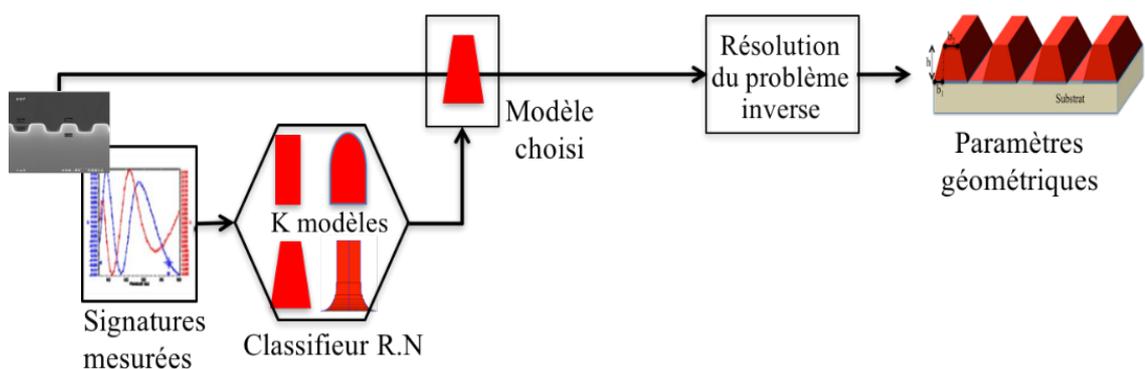


Figure I. 9 : Schéma illustrant le phénomène de classification d'une signature ellipsométrique selon K modèles de profils géométriques distincts.

La signature de l'échantillon utilisée lors de la caractérisation proprement dite est classée parmi K classes prédéfinies qui représentent chacune un modèle de profil géométrique particulier. Le modèle sélectionné par le classifieur sera ensuite utilisé comme hypothèse structurale dans l'étape de caractérisation scattérométrique et fixera ainsi le nombre de paramètres à déterminer.

Dans cette partie nous avons précisé le contexte dans lequel évolue la méthode scattérométrique. Nous avons ainsi mis en évidence les potentialités de cette méthode pour répondre à des besoins ciblés. Ainsi, nous avons présenté en premier lieu l'apport d'une méthode de classification dans la détection de défauts lors d'un processus de fabrication par lithographie et en second lieu nous avons mis en évidence plusieurs problématiques observées lors d'une étape de gravure qui pourraient être détectées par une étape de classification en amont de la caractérisation scattérométrique proprement dite.

I.2 Fonctionnement de la scattérométrie ellipsométrique

Dans cette partie nous allons détailler le principe général de la scattérométrie en s'attardant sur la signature ellipsométrique qui est la plus utilisée dans ce domaine. Les conditions expérimentales particulières utilisées dans cette thèse seront également précisées.

I.2.1 La structure périodique test

Un réseau de diffraction ou réseau optique est une structure périodique (1D ou 2D) de forme géométrique quelconque reposant sur un support matériel : le substrat (figure I.10).

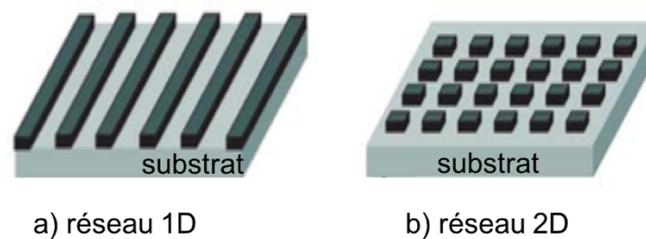


Figure I. 10 : Exemples de modèles de réseaux de diffraction lamellaire 1D (a) et 2D (b) [3]

La caractéristique principale d'un réseau de diffraction est sa période spatiale exprimée en micromètre ou nanomètre. On peut définir également sa fréquence par le nombre de rainures ou traits par millimètre. Ces réseaux ont la capacité de diffracter une lumière incidente dans plusieurs directions privilégiées appelées ordres de diffraction. En effet, chaque rainure se comporte comme un dipôle ou source secondaire de lumière qui émet aussi bien en transmission

(suivant le type de substrat) qu'en réflexion. La mise à profit de cette propriété intéressante offre un large champ d'applications aux réseaux périodiques (télécommunication, capteur, optique intégrée, microélectronique [33] [34]). Ainsi, dans l'industrie microélectronique ils sont souvent utilisés comme structures types pour tester et comparer les méthodes de caractérisation dimensionnelle lors de la fabrication des composants de faibles dimensions.

Il convient donc de rappeler brièvement le fondement théorique de la diffraction de la lumière par un réseau périodique.

Considérons un réseau de diffraction de période d séparant 2 milieux d'indices respectifs n_t et n_r et un faisceau incident de longueurs d'onde λ (figure I.11). Notons par θ_i l'angle d'incidence entre la normale au réseau et le faisceau incident. On appelle équation des réseaux, l'équation donnant les angles orientés des différents ordres de diffraction notés θ_{dm} en fonction de θ_i et d :

$$n_d \sin \theta_{dm} = n_r \sin \theta_i + m \frac{\lambda_0}{d}, m = 0, \pm 1, \pm 2, \dots \quad \text{I.1}$$

où m est un entier positif ou négatif caractérisant l'ordre de diffraction et n_d représente l'indice du milieu dans lequel on étudie la diffraction (pour les ordres réfléchis $n_d = -n_r$ et pour les ordres transmis $n_d = n_t$).

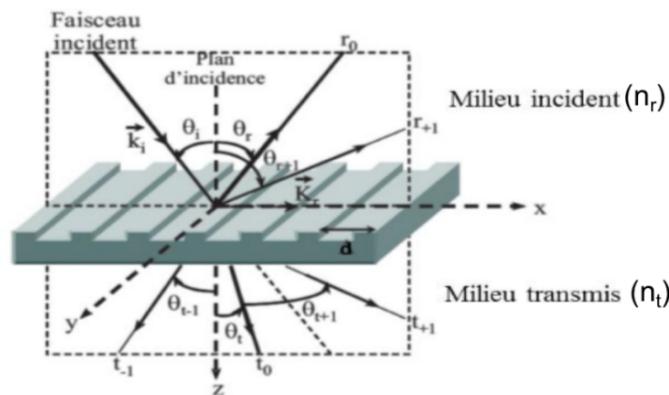


Figure I. 11 : représentation du phénomène de diffraction par un réseau de période d en configuration classique.

Si $m=0$, on retrouve pour $n_d=n_t$ l'équation de Snell-Decartes caractérisant la réfraction à travers une interface séparant deux milieux d'indices différents (Eq I.2) et pour $n_d=-n_r$ on retrouve la loi de réflexion (Eq I.3) :

$$n_t \sin \theta_m = n_r \sin \theta_i \quad \text{I.2}$$

$$\theta_m = -\theta_i \quad \text{I.3}$$

Dans le cas particulier où le milieu émergent est de l'air et que le milieu incident est le même milieu (on a $n_r=n_i=1$), l'équation des réseaux devient :

$$\sin\theta_{dm} = \sin\theta_i + m \frac{\lambda}{d}, m = 0, \pm 1, \pm 2, \dots \quad \text{I.4}$$

Suivant la façon d'éclairer le réseau il existe deux types de diffraction. Soit \mathbf{k}_i le vecteur d'onde incidente défini par la relation suivante :

$$\mathbf{k}_i = n_r \mathbf{k}_0 \quad \text{avec } k_0 = \frac{2\pi}{\lambda_0} \quad \text{I.5}$$

On appelle plan d'incidence le plan formé par \mathbf{k}_i et la normale au réseau. On définit également le vecteur réseau \mathbf{k}_r par le vecteur perpendiculaire au sillon du réseau et dont le module est appelé nombre d'ondes est :

$$k_r = \frac{2\pi}{d} \quad \text{I.6}$$

L'équation des réseaux (Eq. I.1) s'écrit alors :

$$n_d \sin\theta_{dm} = n_r \sin\theta_i + m \frac{k_r}{k_0} \quad \text{I.7}$$

Deux cas sont alors possibles :

- Le cas où le plan d'incidence est perpendiculaire au sillons (\mathbf{k}_r est contenu dans le plan d'incidence). Dans ce cas on parle de diffraction classique. Tous les ordres diffractés sont contenus dans le plan d'incidence. Si le vecteur \mathbf{k}_i est perpendiculaire au plan du réseau on parle de diffraction normale.
- Dans le cas où le plan d'incidence fait un angle φ avec le vecteur réseau \mathbf{k}_r , on parle alors de diffraction conique. Les ordres diffractés ne sont plus contenus dans un plan, mais sont situés dans un cône d'axe \mathbf{k}_r . φ est également appelé angle azimutal.

Dans cette thèse, nous traiterons uniquement le cas d'une incidence classique.

Il est également nécessaire de définir la polarisation de l'onde incidente. Rappelons que l'onde électromagnétique plane incidente est caractérisé par le trièdre direct $(\mathbf{E}, \mathbf{H}, \mathbf{k}_i)$ dans lequel \mathbf{E} est le champ électrique et \mathbf{H} le champ magnétique.

Si l'onde incidente est linéairement polarisée, on distingue alors 2 cas particuliers :

- Le vecteur \mathbf{E} est perpendiculaire au plan d'incidence qui contient donc \mathbf{k}_i et \mathbf{H} . on parle alors d'une polarisation transverse électrique notée TE (ou s).

- Le vecteur champ magnétique \mathbf{H} est perpendiculaire au plan d'incidence qui contient alors \mathbf{k}_i et \mathbf{E} , on parle de polarisation transverse magnétique notée TM (ou p).

Tous les états de polarisation peuvent être envisagés par combinaison linéaire de ces 2 cas particuliers.

Dans la résolution d'un problème de diffraction d'une onde électromagnétique à travers un réseau on cherche à connaître la distribution du champ électrique réparti dans les différents ordres dans les différents milieux en fonction de l'échantillon (période, indice des matériaux, géométrie du profil etc ...) et des conditions d'excitation (angle d'incidence, polarisation etc ...). Chaque échantillon possède donc une répartition ou une signature propre dans des conditions expérimentales fixées. Plusieurs types de mesures sont possibles. Nous nous limiterons à la mesure ellipsométrique que nous détaillons ci-dessous.

I.2.2 Signature ellipsométrique

I.2.2.1 Principe de l'ellipsométrie classique

L'ellipsométrie consiste à déterminer les caractéristiques d'une ou plusieurs couches (indices optiques complexes, épaisseurs, etc ...) composant un échantillon dont la structure est connue (mono ou multicouche par exemple).

Commençons par considérer une onde plane arrivant sur la surface d'un échantillon séparant 2 milieux d'indice n_0 et n_1 . Une partie de l'onde est transmise ou absorbée à travers la surface, une autre partie est réfléchi (figure I.12). Le champ électrique \mathbf{E}_i de l'onde incidente possède a priori deux composantes : l'une (E_{ip}) parallèle au plan d'incidence, l'autre (E_{is}) perpendiculaire au plan d'incidence.

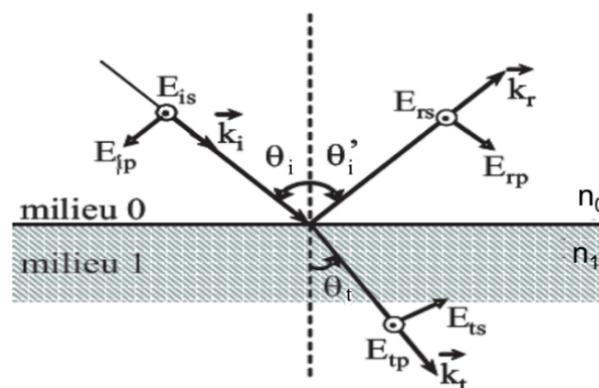


Figure I. 12 : Réflexion et transmission d'une onde plane incidente avec un angle θ_i , sur une surface plane séparant deux milieux semi-infinis, d'indice respectif n_0 et n_1 .

La modification du champ électrique après réflexion sur l'échantillon est définie par deux coefficients agissant sur chacune des composantes du champ :

- Le coefficient de réflexion en amplitude pour une polarisation TM:

$$r_p = \frac{E_{rp}}{E_{ip}} = |r_p| \exp(j\delta_p) \quad \text{I.8}$$

- Le coefficient de réflexion en amplitude pour une polarisation TE :

$$r_s = \frac{E_{rp}}{E_{ip}} = |r_s| \exp(j\delta_s) \quad \text{I.9}$$

$|r_p|$ et $|r_s|$ représentent respectivement la modification apportée en amplitude à la composante p et s du champ incident, δ_p (respectivement δ_s) représente le retard de phase introduit par la réflexion dans le cas d'une polarisation p (respectivement s).

En pratique, la quantité mesurée est le rapport des deux coefficients. Elle est définie comme étant l'équation fondamentale de l'ellipsométrie :

$$\rho = \frac{r_p}{r_s} = \tan(\psi) \exp(j\Delta) \quad \text{I.10}$$

Par identification on a :

$$\psi = \arctan\left(\left|\frac{r_p}{r_s}\right|\right) \quad \text{I.11}$$

avec $0 < \Psi < 90^\circ$

$$\Delta = \delta_p - \delta_s \quad \text{I.12}$$

avec $0 < \Delta < 180^\circ$

Les angles Ψ et Δ sont appelés les angles ellipsométriques accessibles par la mesure.

Notons que l'on peut définir les mêmes quantités pour la transmission de l'onde mais cela est peu utilisé en pratique car ce n'est pas adapté pour tous les types de substrats (en particulier les matériaux absorbants).

Les caractéristiques physiques recherchées de l'échantillon analysé sont ensuite issues d'un traitement mathématique des mesures de Ψ et Δ . Il s'agit dans la plupart des cas d'une optimisation fonctionnant sur une modélisation paramétrique préalable de la structure.

I.2.2.2 Appareil de mesure

L'appareil utilisé pour la mesure est un ellipsomètre. Le montage optique est généralement constitué d'une source, d'un polariseur, d'un analyseur et d'un détecteur. Ce sont les constituants de base auxquels peuvent être ajoutés suivant la méthode de mesure différents éléments comme un modulateur ou un compensateur, etc. Dans le montage présenté sur la figure 1.13, les différents éléments sont arrangés selon un bras source (PSG pour Polarization State Generator) et un bras de détection (PSD pour Polarization State Detector). Les angles de l'axe de l'analyseur, du polariseur et du modulateur avec le plan d'incidence sont respectivement notés θ_A , θ_P et θ_C .

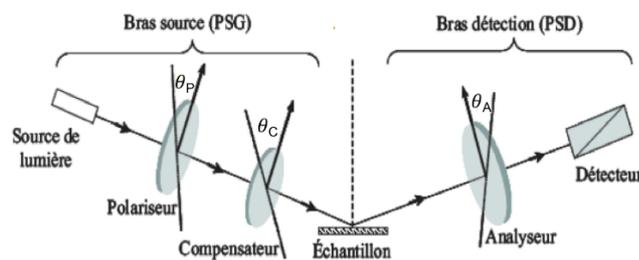


Figure I. 13 : Montage de base d'un ellipsomètre classique

On peut ainsi trouver différents types d'ellipsomètres fonctionnant sur des principes différents :

➤ **Ellipsomètre à annulation** [20]

C'est l'ellipsomètre le plus simple. Sa configuration correspond au montage de la figure I.13. Il est constitué d'une source de lumière, puis successivement, d'un polariseur (P), d'un compensateur (C), d'un analyseur (A) et d'un détecteur. La mesure avec ce type d'ellipsomètre se fait par un ajustement de l'orientation du polariseur et de l'analyseur de façon à ne plus détecter aucun signal en sortie de montage, d'où le terme annulation. Lorsque l'extinction est atteinte, les angles θ_P du polariseur, θ_C du compensateur et θ_A de l'analyseur sont directement reliés aux angles ellipsométriques suivant la relation :

$$\tan(\Psi) \operatorname{exj}(\Delta) = -\tan\theta_A \frac{\tan\theta_C - \tan(\theta_P - \theta_C)}{1 + \tan\theta_C \tan(\theta_P - \theta_C)} \quad \text{I.13}$$

Cette formulation permet par identification, la détermination de Ψ et Δ . Cependant, même automatisé, ce type d'ellipsomètre est lent, car la mesure nécessite la recherche de l'extinction du faisceau. De plus, sa précision dépend du bruit du détecteur car on travaille toujours au minimum du signal.

➤ Ellipsomètre à modulation

L'automatisation des ellipsomètres a permis l'incorporation d'une modulation pour améliorer la qualité de la détection du signal. Cette modulation peut être produite par la rotation d'un élément [35] (ellipsomètre à polariseur tournant (ou RPE pour Rotating Polarizer Element, ellipsomètre à analyseur tournant (ou RAE pour Rotating Analyzer Element) ou ellipsomètre à compensateur tournant (ou RCE pour Rotating Compensator Element)) mais également par modification de la polarisation de l'onde utilisée pour la mesure. Nous détaillerons dans la suite uniquement cette dernière méthode que nous avons utilisée dans nos travaux.

L'ellipsomètre à modulation de phase du Laboratoire Hubert Curien est un UVISEL de Jobin Yvon (figure I.14). Ce type d'ellipsomètre peut effectuer facilement des mesures en mode spectroscopique dans une large gamme spectrale qui s'étend de l'ultraviolet au proche infra-rouge (260 nm — 2100 nm) et éventuellement en mode dynamique.

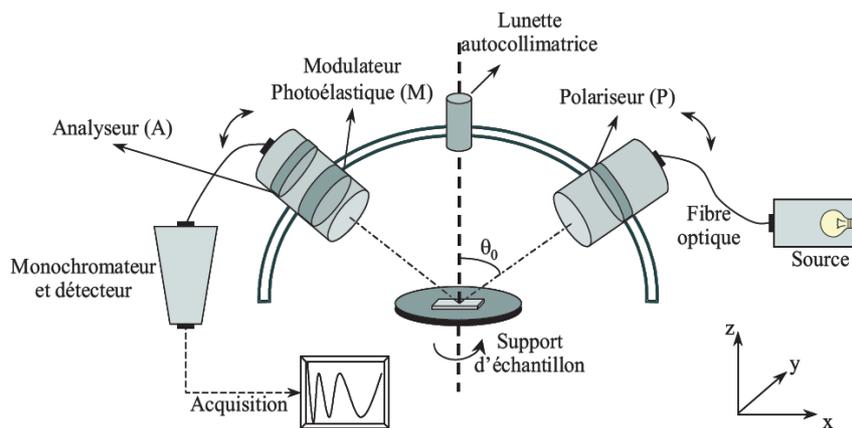


Figure I. 14 : Représentation schématique de l'ellipsomètre spectroscopique à modulation de phase (UVISEL de Jobin Yvon) [3]

La durée d'une mesure dépend du temps d'intégration. Cette dernière est généralement comprise entre 100 et 500 ms. Il faudrait également noter que plus le temps d'intégration est faible et plus la mesure sera bruitée. Dans nos travaux nous avons considéré un bruit de mesure correspondant à un temps d'acquisition équivalent à 500 ms.

Le mode spectroscopique permet d'effectuer des mesures en balayant successivement toutes les longueurs d'onde d'un spectre défini à l'avance en imposant un pas fixe entre deux acquisitions. Notons que plus l'échantillonnage est important plus le temps de mesure est grand (environ 2 minutes [36] pour un spectre dans la gamme 300 - 800 nm avec un pas de 0.2 nm).

Le mode dynamique quant à lui permet d'effectuer plusieurs mesures d'un nombre limité de longueurs d'onde en simultané. L'échantillonnage dépend du nombre de photorécepteurs présents en sortie du système. Dans ce cas, le spectre est faiblement plus résolu, mais obtenu rapidement. Aussi le mode dynamique est mieux adapté pour le contrôle en temps réel.

Ce type d'ellipsomètre est équipé d'un modulateur photoélastique situé sur le bras de détection entre l'analyseur et l'échantillon (classiquement entre le polariseur et l'échantillon pour les autres équipementiers). Ce dernier introduit un déphasage sinusoïdal dynamique entre les deux composantes du champ qui le traverse, noté :

$$\delta(t) = \delta_0 \sin(\omega t)$$

avec δ_0 l'amplitude de modulation et ω la pulsation du modulateur.

La fréquence de modulation est typiquement de 50 Hz. La matrice de transfert du modulateur peut être décrite par le formalisme de Jones :

$$M = \begin{pmatrix} \exp j\delta(t) & 0 \\ 0 & 1 \end{pmatrix} \quad \text{I.14}$$

Le champ électrique sur le détecteur peut être déterminé en utilisant le formalisme de Jones par le calcul matriciel suivant :

$$\begin{pmatrix} E_p \\ E_s \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \cos\theta_A & -\sin\theta_A \\ \sin\theta_A & \cos\theta_A \end{pmatrix} \begin{pmatrix} r_p & 0 \\ 0 & r_s \end{pmatrix} \begin{pmatrix} \cos\theta_M & -\sin\theta_M \\ \sin\theta_M & \cos\theta_M \end{pmatrix} \begin{pmatrix} \exp j\delta(t) & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \cos(\theta_p - \theta_M) & -\sin(\theta_p - \theta_M) \\ \sin(\theta_p - \theta_M) & \cos(\theta_p - \theta_M) \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} E_x \\ E_y \end{pmatrix} \quad \text{I.15}$$

On obtient donc en intensité :

$$I(t) = |E_p|^2 + |E_s|^2 = I_0 + I_s \sin\delta(t) + I_c \cos\delta(t) \quad \text{I.16}$$

avec

$$I_0 = B(1 - \cos 2\psi \cos 2\theta_A + \cos 2(\theta_p - \theta_M) \cos 2M (\cos 2\theta_A - \cos 2\psi) + \sin 2\theta_A \cos \Delta \cos 2(\theta_p - \theta_M) \sin 2\psi \sin 2M) \quad \text{I.17}$$

$$I_s = B[\sin 2(\theta_p - \theta_M) \sin 2\theta_A \sin 2\psi \sin \Delta] \quad \text{I.18}$$

$$I_c = B \sin 2(\theta_p - \theta_M) [(\cos 2\psi - \cos 2\theta_A) \sin 2\theta_M + \sin 2\theta_A \cos 2\theta_M \sin 2\psi \cos \Delta] \quad \text{I.19}$$

où

$$B = \frac{E_0^2}{4|r_p^2+r_s^2|} \quad \text{I.20}$$

La connaissance des intensités I_s et I_c pour une configuration angulaire donnée permet donc de calculer Ψ et Δ . La plupart du temps, le polariseur et le modulateur (ou l'analyseur et le modulateur dans certaines configurations) sont rendus solidaires de telle sorte que : $\theta_p - \theta_M = \pm 45^\circ$ (ou $\theta_A - \theta_M = \pm 45^\circ$). Ainsi deux configurations particulières permettent d'exprimer I_s et I_c uniquement en fonction des angles ellipsométriques Ψ et Δ :

- **Configuration I** : $\theta_M = 0^\circ$, $P = 45^\circ$ et $\theta_p - \theta_M = \pm 45^\circ$ donne,

$$I_s = \sin 2\Psi \sin \Delta$$

$$I_c = \sin 2\Psi \cos \Delta$$

- **Configuration II** : $\theta_M = 45^\circ$, $P = 45^\circ$ et $\theta_p - \theta_M = \pm 45^\circ$ donne les expressions suivantes

$$I_s = \sin 2\Psi \sin \Delta$$

$$I_c = \cos 2\Psi$$

Les intensités I_s et I_c varient entre -1 et 1. L'utilisation de l'une ou de l'autre des deux configurations comporte des ambiguïtés dans la détermination de Ψ et Δ . En effet dans la configuration I, Ψ est indéterminé entre Ψ et $90^\circ - \Psi$ par la seule connaissance de $\sin 2\Psi$. Il existe une incertitude sur Ψ alors que Δ est parfaitement déterminé. A l'inverse dans la configuration II, Δ est indéterminé entre 2 valeurs Δ et $180^\circ - \Delta$. Cette fois-ci il résulte une incertitude sur Δ par la seule connaissance de $\sin \Delta$ alors que Ψ est déterminé parfaitement. Pour éviter ces problèmes il faut donc déterminer Ψ et Δ systématiquement dans les deux configurations (Δ dans la configuration I et Ψ dans la configuration II).

Dans la suite, nous travaillerons directement avec les mesures « brutes » I_s et I_c , sans passer par la connaissance explicite de Ψ et Δ ; ce qui nous permet de nous affranchir de ce type de considération.

Dans ce paragraphe nous avons présenté la signature ellipsométrique utilisée par la suite dans le cadre d'une caractérisation scattérométrique. Elle est composée de couples de mesures d'intensités I_s et I_c pour un spectre de longueurs d'onde données. Dans le cadre de la résolution du problème inverse permettant de remonter à l'estimation des paramètres à caractériser, ces intensités devront être simulées par des outils numériques que nous détaillons ci-dessous.

I.2.3 Problème direct de la diffraction

D'un point de vue mathématique, considérons un phénomène F décrit par des paramètres variables \mathbf{x} . La résolution du problème direct consiste en une modélisation M de F afin de retrouver les effets observables \mathbf{y} engendrés par les paramètres \mathbf{x} . Une fois le modèle M obtenu, on peut alors estimer $F(\mathbf{x})$ par $M(\mathbf{x})$. Ce problème possède une solution unique.

En scattérométrie, résoudre le problème direct consiste donc à déterminer la signature optique de l'échantillon correspondant à des conditions expérimentales fixées. Dans notre cas, il s'agit des mesures I_s et I_c dépendant directement des paramètres ellipsométriques Ψ et Δ et donc des composantes du champ électrique de l'onde. La plupart du temps on a recours à des outils de simulations électromagnétiques.

I.2.3.1 Etat de l'art des méthodes

Deux familles de méthodes existent pour calculer les intensités des champs diffractés par un réseau périodique : les **méthodes approchées** et les **méthodes rigoureuses** qualifiées d'exactes.

Les méthodes approchées présentent l'intérêt d'être rapides dans le calcul, mais les résultats sont plus ou moins fidèles au comportement physique du phénomène. Parmi ces méthodes nous pouvons citer la théorie scalaire de Fresnel-Kirchoff [37] qui considère la propagation de l'onde comme un phénomène scalaire sans prendre en compte la dimension vectorielle de l'onde électromagnétique. Une autre approche qui se base sur l'hypothèse de Rayleigh considère le problème de diffraction par voie électromagnétique où on résout un système d'équations linéaires [38] [39].

Les méthodes rigoureuses quant à elles consistent à résoudre les équations de Maxwell sans une approximation sur le plan théorique. Elles sont constituées de deux familles.

On trouve ainsi, les méthodes intégrales [40] [41] qui sont basées sur le théorème de Green. Ce dernier exprime le champ électrique en chaque point de l'espace par une intégrale sur la surface du réseau. Un courant de surface est généré par l'onde incidente et le champ diffracté est obtenu par rayonnement de ce courant sur la surface. On cite également la méthode des éléments finis [42] qui utilise des méthodes d'intégrales différentes. La surface est représentée par un maillage et le champ électrique est exprimé comme une somme des fonctions élémentaires sur chaque cellule du maillage. Le problème est ramené à un système d'équations linéaires algébriques. Une autre méthode intégrale présente dans la littérature est celle des

« sources fictives » [43]. On considère qu'un champ électrique est généré par des sources fictives choisies minutieusement de manière à vérifier les équations de Maxwell.

On trouve également les méthodes différentielles qui considèrent, quant à elles, les équations de Maxwell sous leur forme différentielle. Le champ électrique et la permittivité sont représentés en série de Fourier. Nous pouvons citer la méthode C [44] [45] dans laquelle on effectue un changement de coordonnées simple séparant l'espace en deux milieux homogènes. Dans cette thèse nous avons choisi d'utiliser la méthode modale par développement en série de Fourier dite MMFE [46] [47] pour le calcul électromagnétique des signatures optiques.

I.2.3.2 Une méthode performante : la Multilayer modal method by fourier expansion

L'origine de cette méthode remonte aux années 80. En effet, la première version est connue sous le nom de méthode rigoureuse des ondes couplées publiée par Moharam et Gaylor [47] (en anglais RCWA pour Rigorous Coupled Wave Analysis). La Multilayer Modal Method by Fourier Expansion (MMFE) exprime les solutions des équations de Maxwell sous la forme de modes de propagation dans la structure à analyser. Elle s'applique intrinsèquement à un réseau de forme lamellaire. Toutefois, une extension à une forme plus complexe est possible en considérant une juxtaposition de réseaux simples rectangulaires comme approximation du profil géométrique désiré. Ainsi, la forme géométrique de la structure est découpée en N niveaux lamellaires ou couches. Chaque couche se présente comme un rectangle de matériau d'indice n_1 baignant dans un milieu d'indice constant n_0 nommé le superstrat (partie 1 sur la figure I.15). L'ensemble des couches repose sur un substrat d'indice n_s constant (partie 3 sur la figure I.15). Chaque couche, composant le réseau est donc découpé en trois parties permettant de distinguer les zones de changement d'indice (partie 2 sur la figure I.15). Le champ électromagnétique global est donné par une combinaison linéaire de tous les modes calculés indépendamment dans chaque couche. Les amplitudes de ces modes sont ensuite déterminées par l'écriture des relations de continuité. Numériquement un nombre limité d'harmoniques M est fixé pour le calcul.

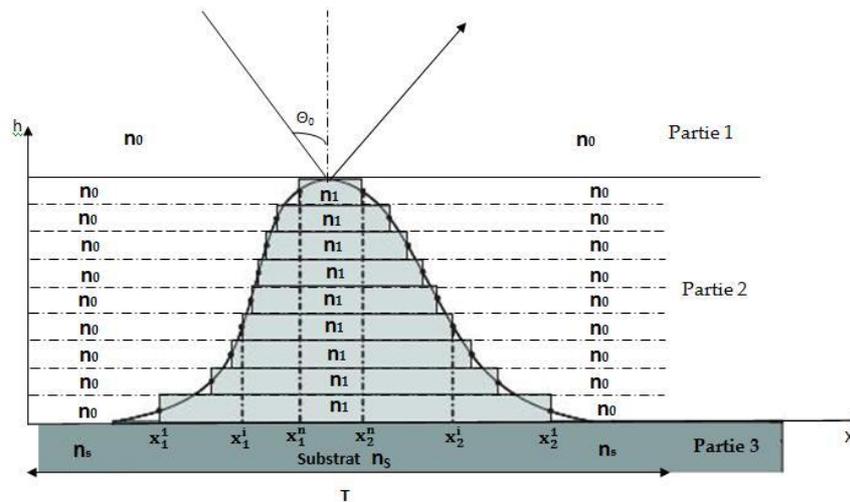


Figure 1.15 : Découpe d'un profil géométrique quelconque en couches lamellaires

La MMFE est très utilisée dans le calcul direct en scattérométrie. Elle peut s'appliquer à des réseaux quelconques (1D ou 2D), absorbants ou métalliques [48], en configuration classique ou conique. Les codes de calcul MMFE que nous avons utilisés dans cette thèse sont ceux utilisés depuis plusieurs années par l'équipe « surface complexe » développés initialement par l'équipe « Electromagnétisme » de l'institut Pascal (UMR 6602 — CNRS, Université Blaise Pascal — Clermont Ferrand).

Un des inconvénients de la MMFE qui nous intéresse particulièrement est lié au fait que le profil réel doit être approché par un empilement de couches. Pour cela, nous avons développé au sein du laboratoire une méthode de découpe de profils géométriques adaptés à chaque modélisation paramétrique en prenant en compte le nombre de couche désirées par l'utilisateur.

Ainsi, le programme MMFE calcule la signature ellipsométrique d'un réseau défini par les paramètres géométriques générés par la fonction de découpe à partir des paramètres M et N et dans les conditions expérimentales connues (longueur d'onde λ , angle d'incidence θ_i , ...).

Mathématiquement, F représente le phénomène de diffraction par le réseau périodique. Le vecteur \mathbf{x} constitue les paramètres géométriques définissant le profil et \mathbf{y} correspond à la signature optique du réseau. Ainsi nous avons F modélisée par :

$$M: \mathbf{x} \rightarrow M_a(\mathbf{x}) = \mathbf{y}$$

où le vecteur \mathbf{a} définit les conditions des configurations expérimentales de mesure et l'ensemble des paramètres mathématiques utilisés.

I.2.3.3 Modélisation des formes géométriques utilisées

Nous présentons dans ce paragraphe les différentes formes géométriques de réseau traitées dans notre travail. Nous définirons pour chacune d'entre elle le nombre de paramètres permettant une caractérisation unique et nous présenterons des exemples de découpe en couches lamellaires nécessaires pour l'emploi de la MMFE.

➤ Le profil créneau

Le profil créneau ou rectangulaire ou encore lamellaire est le plus élémentaire. Il se présente sous la forme de créneaux et est parfaitement représenté par 2 paramètres : la longueur du plateau notée CD et la hauteur h (figure I.16). Ce type de profil présente l'avantage d'être simple et de ne pas nécessiter d'approximation pour le calcul direct (cf paragraphe précédent)

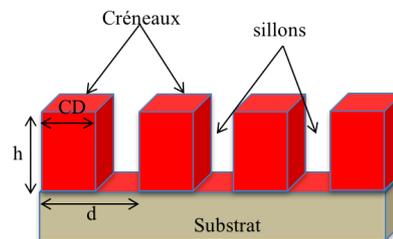


Figure I. 16 : Modélisation paramétrique d'un profil créneau défini par 2 paramètres CD et h

Dans le cas de la présence d'une couche résiduelle au fond des sillons, on ajoute simplement un paramètre supplémentaire noté h_r caractérisant la hauteur de cette dernière.

Notons cependant que le profil créneau est le moins probable physiquement bien qu'étant la plupart du temps celui désiré. Les profils réels sont généralement plus complexes. Il faut donc envisager un autre type de modèle de profil géométrique.

➤ Le profil trapézoïdal

Le profil trapézoïdal symétrique est la forme la plus usitée en scattérométrie pour caractériser un réseau. Le profil est défini par trois paramètres : b_1 (projection de la pente), b_2 (longueur du plateau) et h (hauteur du profil) (figure I.17). Mathématiquement, ce profil n'existe que dans les conditions suivantes :

$$b_1 \geq 0, b_2 > 0 \text{ et } h > 0$$

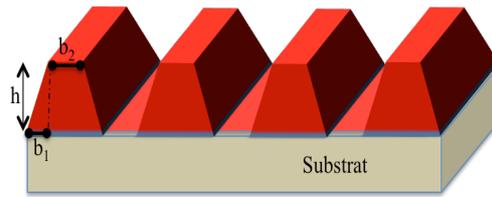


Figure I. 17 : Modélisation paramétrique d'un profil trapézoïdal défini par 3 paramètres b_1 , b_2 et h

La figure I.18 montre l'exemple d'un profil de forme trapézoïdale approché par la procédure de découpe mise en place.

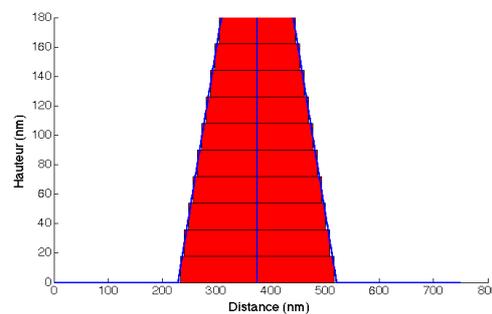


Figure I. 18 : Profil trapézoïdal symétrique défini par $b_1=80$ nm, $b_2=132$ nm, $h=180$ nm de période de réseau $d=750$ nm approché par $N=10$ couches.

Il faut noter que le profil trapézoïdal inclut également le profil créneau dans le cas où la projection de la pente b_1 est nulle.

➤ **Profil sinusoïdal**

Dans ce cas, le profil est un simple sinus (figure I.19) pour lequel on ne conserve que la partie positive. Il est défini par deux paramètres : h (amplitude du sinus), b (base du sinus) :

$$\begin{aligned} 0 < b &\leq T \\ h &> 0 \end{aligned} \tag{I.21}$$

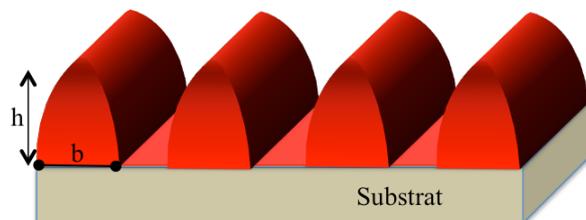


Figure I. 19 : Modélisation paramétrique d'un profil sinusoïdal défini par son amplitude h et la longueur de sa base b

D'un point de vue pratique, l'ajout d'un paramètre α permet de moduler la pente du profil sinusoïdal. On définit ainsi un sinus tronqué par 3 paramètres : la base b , l'amplitude h et le paramètre α ($0 < \alpha \leq 1$) spécifiant la hauteur de troncature par un pourcentage de la hauteur initiale du sinus.

Nous présentons dans la figure I.20 un exemple de profil sinus tronqué défini par les paramètres $b=252,6$ nm, $h=90$ nm et $\alpha=0.7$. La période du sinus étant de $1 \mu\text{m}$.

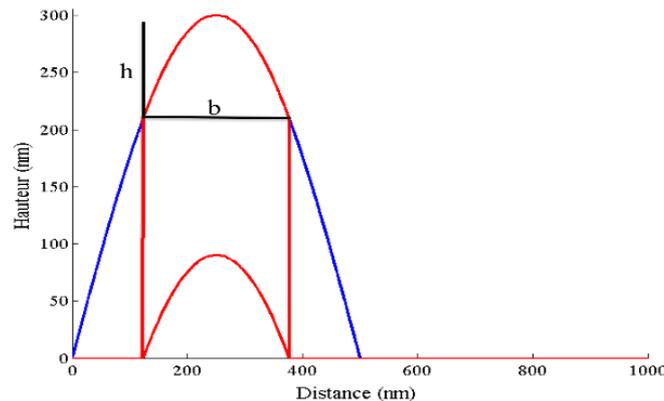


Figure I. 20 : Exemple de sinus tronqué défini par $b=252.6$ nm et $h=90$ nm et $\alpha=0.7$ (rouge) à partir d'un sinus initial de période $d=1 \mu\text{m}$ (bleu).

La figure I.21 montre ce même profil approché par une découpe en couches en considérant $N=15$ couches.

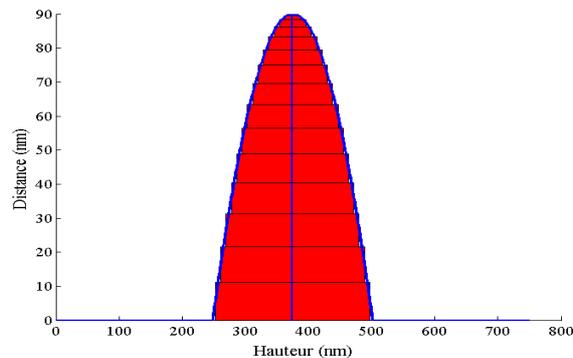


Figure I. 21 : Profil sinusoidal tronqué de période $d=750$ nm défini par $b=252.6$ nm, $h=90$ nm et $\alpha=0.7$ approché par $N=15$ couches.

➤ Profil rectangulaire à bas arrondi

C'est un créneau présentant des arrondis au fond des sillons définis mathématiquement par des quarts de cercles. Le profil est ainsi défini par 3 paramètres : CD (longueur du plateau), h (hauteur totale) et r (rayon de la partie arrondie) (figure I.22) assujettis aux contraintes suivantes :

$$\begin{aligned}
 0 < r &\leq h \\
 2r + CD &\leq d \\
 b > 0 \text{ et } h > 0
 \end{aligned}
 \tag{I.22}$$

La figure I.22 montre un profil rectangulaire à bas arrondis défini par $b=320$ nm, $h=500$ nm, $r=200$ nm découpé en 10 tranches.

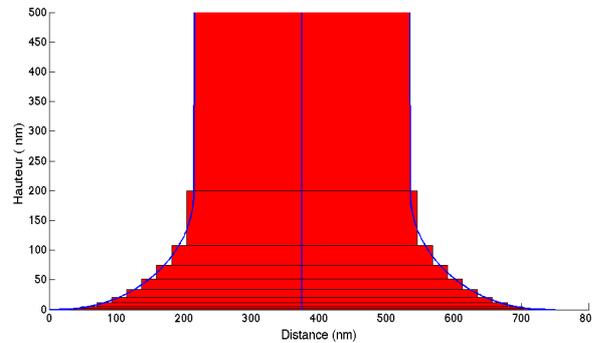


Figure I. 22 : Profil rectangulaire à bas arrondi période $d=750$ nm défini par $b=320$ nm, $h=500$ nm, $r=200$ nm de période de réseau $d=750$ nm approché par $N=10$ couches.

Notons que pour approcher au mieux le profil réel, l'épaisseur des couches varie en fonction de la courbure du profil.

➤ Profil rectangulaire arrondi au sommet

Ce profil est équivalent au précédent avec cette fois-ci des sommets arrondis. 3 paramètres définissent le profil : h (hauteur totale), b (longueur de la base) et r le rayon de courbure de la partie arrondie (figure I.23). On retrouve des contraintes sur les paramètres similaires aux précédentes :

$$\begin{aligned}
 0 \leq r &\leq h \text{ et } 0 \leq r \leq \frac{b}{2} \\
 0 \leq b &\leq d \\
 h &> 0
 \end{aligned}
 \tag{I.23}$$



Figure I. 23 : Modélisation paramétrique d'un profil rectangulaire arrondi au sommet défini par le rayon de courbure r , la longueur du plateau b et sa hauteur h

Nous présentons dans la figure I.24 un exemple de profil rectangulaire arrondi au sommet de période $d=750$ nm défini par $b=320$ nm, $h=500$ nm et $r=100$ nm approché par 10 couches de hauteur variables.

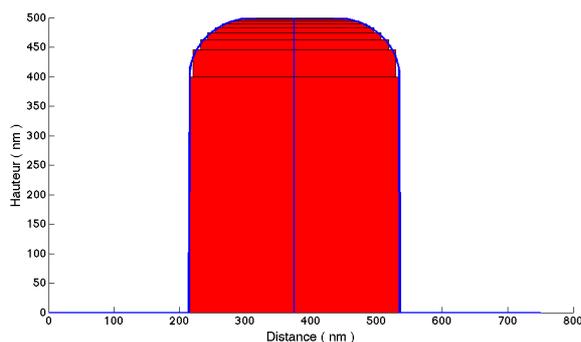


Figure I. 24 : Profil rectangulaire arrondi au sommet de période $d=750$ nm défini par $b=320$ nm, $h=500$ nm, $r=100$ nm découpé en $N=10$ couches.

L'ajout d'un quatrième paramètre permet d'ajuster le niveau de courbure sur le haut de profil. Le profil est alors caractérisé par 4 paramètres. La partie arrondie est modélisée par un quart d'ellipse de demi-axe horizontal r_1 et de demi-axe vertical r_2 (figure I.25). Ce profil est défini dans les conditions suivantes :

$$\begin{aligned} 0 < r_2 &\leq h \\ r_1 &\leq \frac{b}{2} \\ h > 0 \text{ et } 0 < b < d \end{aligned} \quad \text{I.24}$$

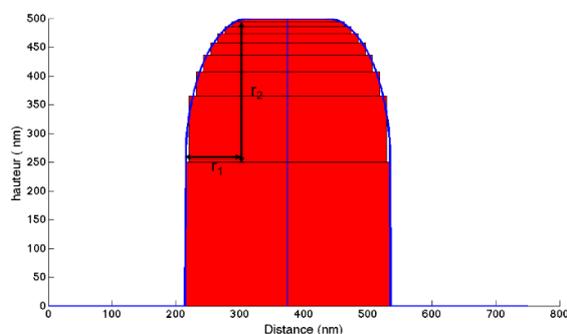


Figure I. 25 : Profil rectangulaire arrondi au sommet de période $d=750$ nm défini par $b=320$ nm, $h=500$ nm, $r_1=100$ nm et $r_2=250$ nm découpé en $N=10$ couches.

➤ Profil rectangulaire arrondi en bas et au sommet

Ce type de profil est une combinaison des 2 profils précédents. Il présente donc une forme arrondie en bas et au sommet des sillons. Les parties arrondies peuvent être définies soit par des quarts de cercle soit par des quarts d'ellipses ou une combinaison des deux. Ainsi le profil peut être défini au minimum par 3 paramètres et au maximum par 6 paramètres suivant la complexité désirée. A titre d'exemple la Figure I.26 présente un profil rectangulaire arrondi

en bas et au sommet de période $d=750\text{nm}$ défini par $b=320\text{ nm}$, $h= 500\text{ nm}$, $r_{\text{bas}}=150\text{ nm}$ et $r_{\text{haut}}=100\text{ nm}$.

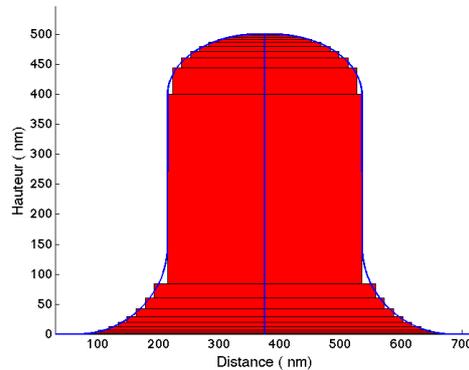


Figure I. 26 : Profil rectangulaire arrondi en bas et au sommet de période $d=750\text{ nm}$ défini par $b=320\text{ nm}$, $h= 500\text{ nm}$, $r_{\text{bas}}=150\text{ nm}$ et $r_{\text{haut}}=100\text{ nm}$ découpé en $N=20$ couches.

Notons que la précision sur l'approche du profil dépend du nombre de couches N . La figure I.27 montre un profil trapézoïdal approché par une découpe en couches en considérant différentes valeurs de N .

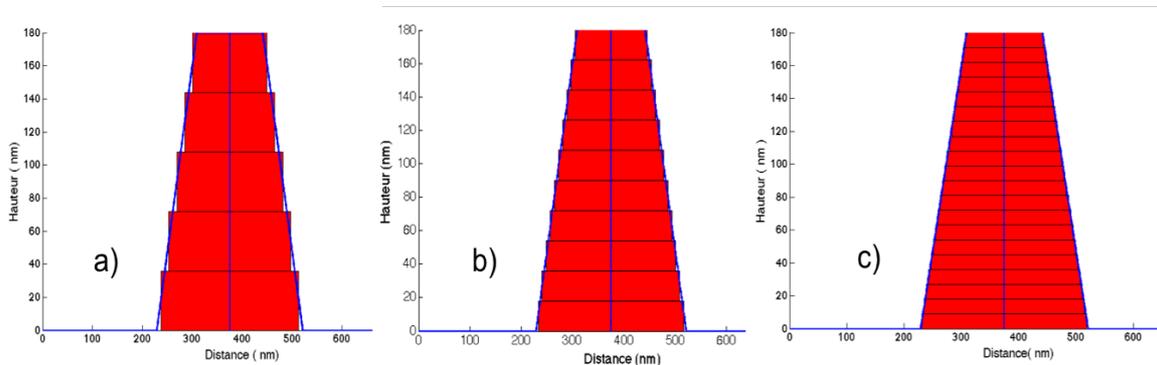


Figure I. 27 : Profil trapézoïdal défini par $b_1=80\text{ nm}$, $b_2=132\text{ nm}$, $h=180\text{ nm}$ de période $d=750\text{ nm}$ approché par $N=5$ couches (a), $N=10$ couches (b) et $N=20$ couches (c) avec $\epsilon = \frac{b_1}{N}$

Plus N est élevé plus l'erreur sur la modélisation du profil en couche est faible. La signature calculée par MMFE sera donc d'autant plus proche de la signature réelle que le nombre N est élevé. Toutefois, il faut tenir compte du temps de calcul du champ diffracté qui croît avec N . Il est donc préférable de réaliser une étude de convergence permettant de déterminer une valeur optimale de N pour chaque type de profil.

Le paramètre de troncature M joue un rôle similaire dans l'approximation des modes de propagation et conditionne la qualité de la signature simulée par la MMMFE.

Ainsi, le couple M et N doit garantir la convergence de la solution numérique tout en restant le plus petit possible afin d'éviter des temps de calculs trop longs qui sont susceptibles d'impacter les performances de la méthode de résolution du problème inverse. Les valeurs optimales de ces paramètres sont déterminées empiriquement et dépendent de plusieurs critères comme la complexité du profil et les variations d'indices entre la couche du réseau et le substrat et/ou le superstrat.

Pour fixer les idées, on définit C_M (respectivement C_N) comme étant la somme des dérivées partielles de I_s et I_c par rapport au paramètre M (respectivement N) :

$$C_M = \left(\left| \frac{\partial I_s}{\partial M} \right| + \left| \frac{\partial I_c}{\partial M} \right| \right) \quad \text{I.25}$$

$$C_N = \left(\left| \frac{\partial I_s}{\partial N} \right| + \left| \frac{\partial I_c}{\partial N} \right| \right) \quad \text{I.26}$$

Prenons l'exemple d'un réseau en résine (NEB22) déposée sur un substrat de silicium de forme trapézoïdal de période 750 nm caractérisé par $b_1=80$ nm, $b_2=132$ nm et $h=180$ nm. Les figures I.28 et I.29 représentent respectivement les courbes de convergence correspondantes de C_M et C_N .

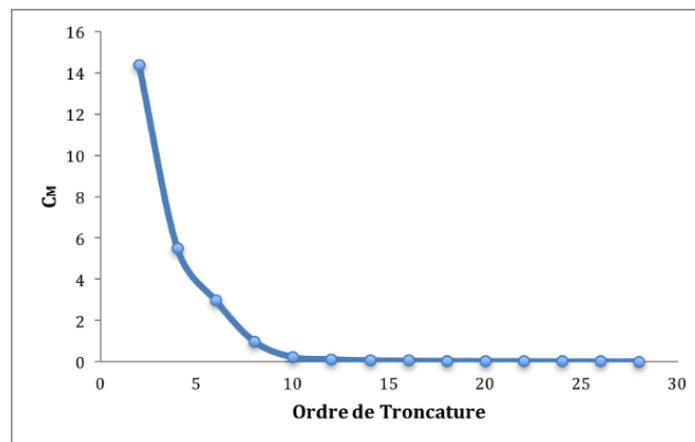


Figure I. 28 : Courbe de convergence selon l'ordre de troncature M ($N=30$) concernant un réseau en résine (NEB22) déposée sur un substrat de silicium de profil trapézoïdal de période $d=750$ nm ($b_1=80$ nm, $b_2=132$ nm, $h=180$ nm). I_s et I_c sont calculées dans la gamme $[1.5 \text{ } 6.5]$ eV pour un angle d'incidence de 60° .

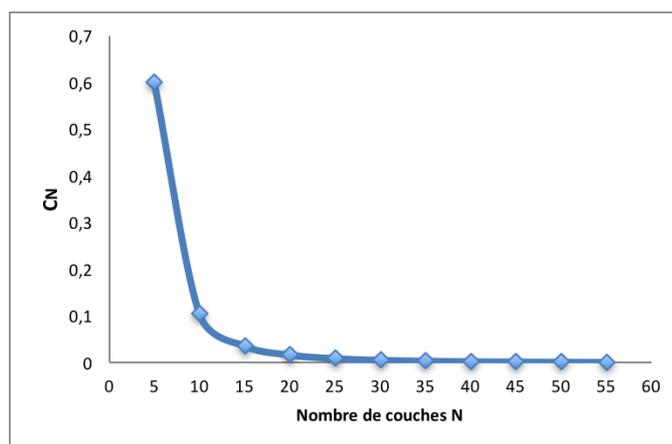


Figure I. 29 : Courbe de convergence selon l'ordre de troncature N ($M=15$) concernant un réseau en résine (NEB22) déposée sur un substrat de silicium de profil trapézoïdal de période $d=750$ nm ($b_1=80$ nm, $b_2=132$ nm, $h=180$ nm). I_s et I_c sont calculées dans la gamme $[1.5 \text{ } 6.5]$ eV pour un angle d'incidence de 60° .

Pour ce profil, les valeurs optimales sont $M=10$ et $N=20$, car à partir de ces valeurs, les intensités I_s et I_c ne semblent plus évoluer.

Ainsi, une fois les profils découpés, les signatures optiques sont calculées par MMFE. Dans le cadre de notre travail, rappelons que la signature est une représentation du couple (I_s , I_c) en fonction de la longueur d'onde (λ) de la lumière incidente. La figure I.30 montre un exemple du couple de signatures (I_s , I_c) calculées par MMFE pour le profil trapézoïdal de la figure I.18 sur un spectre de 26 longueurs d'onde variant entre 1.5 à 6.5 eV pour les paramètres $M=10$ et $N=20$.

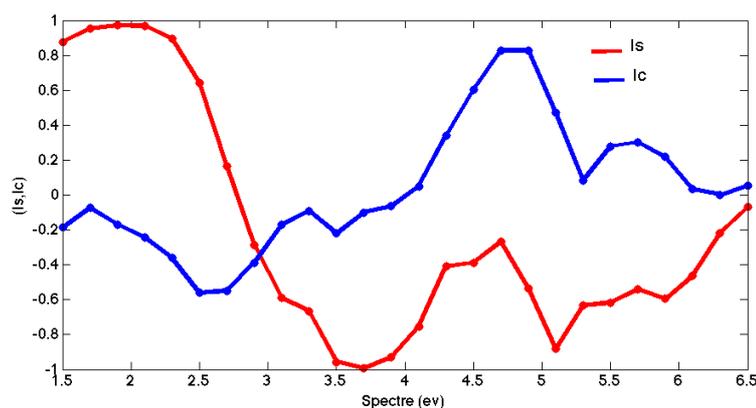


Figure I. 30 : Exemple de simulation des mesures ellipsométriques par MMFE ($N=20$ et $M=10$) représentées par le couple (I_s, I_c) dans le cas d'un réseau de période $d=750$ nm défini par un profil trapézoïdal ($b_1=80$ nm, $b_2=132$ nm, $h=180$ nm) dans la gamme $[1.5 \text{ } 6.5]$ eV pour un angle d'incidence fixe de 60° .

I.2.4 Problème inverse en scattérométrie

I.2.4.1 Définition du problème inverse

Nous avons montré précédemment que le problème direct est accessible par une modélisation M d'un phénomène physique F . Cette fois-ci nous cherchons inversement à résoudre le problème inverse qui consiste à estimer au mieux les paramètres \mathbf{x} qui produisent les mesures \mathbf{y} . Mathématiquement, on cherche une approximation L de la fonction F^{-1} définie par $\mathbf{x}=F^{-1}(\mathbf{y})$ telle que :

$$\mathbf{x} = L(\mathbf{y}) \quad \text{I.27}$$

Ce problème est généralement non trivial sauf dans les cas simples où une résolution analytique est envisageable. La plupart du temps, la résolution est généralement abordée par simulation numérique.

En scattérométrie, le problème inverse est l'étape qui consiste à déterminer les valeurs des paramètres géométriques définissant la forme d'une structure (cf paragraphe I.2.3.3) à partir de sa signature optique mesurée (cf paragraphe I.2.2.2).

La question de l'unicité de la solution peut alors se poser : deux structures différentes peuvent-elles conduire à la même signature optique ? Pour écarter cette possibilité, le nombre de mesures est volontairement augmenté (par exemple en utilisant des configurations angulaires différentes, en augmentant le nombre de longueurs d'onde dans le spectre mesuré, etc). En pratique, on considère des connaissances *a priori* sur la structure. On parle ainsi d'hypothèses structurelles. Ainsi, certaines informations sur l'échantillon à tester sont supposées connues et fixes durant tout le processus de la caractérisation (généralement la période, l'indice optique des matériaux, la forme géométrique...). Ainsi, le nombre de solutions potentielles est fortement réduit, et le problème est plus simple donc moins sujet à l'existence de solutions multiples. On parle en mathématique de régularisation.

Dans notre cas, on cherche le jeu de paramètres \mathbf{x} qui engendre une signature optique \mathbf{y} la plus proche possible de celle mesurée notée $\tilde{\mathbf{y}}$. Pour cela, on se base généralement sur la mesure de l'écart quadratique entre ces 2 signatures. La méthode de résolution se base généralement sur la recherche de l'annulation du χ^2 défini par :

$$\chi^2 = \sum_{i=1}^m (\mathbf{y}_i(\mathbf{x}) - \tilde{\mathbf{y}}_i)^2 \quad \text{I.28}$$

dans laquelle m représente le nombre total de mesures composant la signature optique.

Cependant, il faut noter que la présence de bruit sur les mesures, l'existence d'hypothèses de départ plus ou moins bien vérifiées ne peuvent assurer l'annulation du χ^2 . La solution optimale \mathbf{x}^* correspondra donc à une valeur minimale de χ^2 . Il s'agit donc clairement d'un problème de minimisation.

I.2.4.2 Méthodes d'optimisation classiques

Mathématiquement, la solution doit vérifier deux conditions dites d'optimalité. La première consiste en l'annulation du gradient en \mathbf{x}^* :

$$\nabla\chi^2(\mathbf{x}^*) = 0 \quad \text{I.29}$$

La deuxième impose que la matrice des dérivées secondes appelée matrice Hessienne évaluée en \mathbf{x}^* soit définie positive :

$$\nabla^2\chi^2(\mathbf{x}^*) > 0 \quad \text{I.30}$$

Les méthodes d'optimisation itératives classiquement employées pour rechercher le \mathbf{x}^* sont toutes basées sur la validation de l'une et/ou de l'autre de ces conditions suffisantes d'optimalité. Par exemple l'algorithme de descente du gradient utilise exclusivement la première condition alors que celui de Gauss Newton exploite également la matrice Hessienne [49].

Ainsi à chaque itération k on s'assure que :

$$\chi^2(\mathbf{x}^{k+1}) < \chi^2(\mathbf{x}^k) \quad \text{I.31}$$

avec \mathbf{x}^k et \mathbf{x}^{k+1} les valeurs de \mathbf{x} estimées lors des itérations k et $k+1$.

On peut alors définir le pas entre \mathbf{x}^{k+1} et \mathbf{x}^k par :

$$\mathbf{s} = \mathbf{x}^{k+1} - \mathbf{x}^k = \alpha^k \mathbf{d}^k \quad \text{I.32}$$

avec \mathbf{d}^k la direction de descente et α^k le pas utilisé à l'itération k .

Notons que α^k est un scalaire assurant la relation (I.31). Il est généralement choisi de façon à satisfaire les conditions de Wolfe [50].

Dans la suite nous présenterons brièvement le principe de quelques méthodes classiquement utilisées dans le cadre de l'optimisation paramétrique.

➤ **Algorithme de descente du gradient**

Dans cette méthode la direction de recherche \mathbf{d}^k est donnée par la direction de descente maximale fixée par l'opposé du vecteur gradient calculé au point courant à l'itération k :

$$\mathbf{d}^k = -\nabla\chi^2(\mathbf{x}^k)$$

Ainsi l'équation I.32 devient :

$$\mathbf{s}^k = -\alpha^k \nabla\chi^2(\mathbf{x}^k) \quad \text{I.33}$$

Cette méthode est assez fonctionnelle mais relativement lente.

➤ **Méthode de Gauss Newton**

Cette méthode dite du second ordre, utilise la matrice Hessienne notée \mathbf{H} dans un modèle d'approximation quadratique de la fonction coût en \mathbf{x}^k donné par la formule de Taylor :

$$\chi^2(\mathbf{x}^{k+1}) = \chi^2(\mathbf{x}^k) + \nabla\chi^2(\mathbf{x}^k)^T \mathbf{s}^k + \frac{1}{2} \mathbf{s}^k \mathbf{H} \mathbf{s}^k \quad \text{I.34}$$

Il faut alors déterminer le pas \mathbf{s}^k qui permet d'atteindre le minimum de ce modèle donné par $\nabla\chi^2(\mathbf{x}^{k+1}) = 0$. Il vient donc :

$$\mathbf{H} \cdot \mathbf{s}^k = -\nabla\chi^2(\mathbf{x}^k) \quad \text{I.35}$$

ainsi

$$\mathbf{s}^k = -\mathbf{H}^{-1} \nabla\chi^2(\mathbf{x}^k) \quad \text{I.36}$$

Les performances de la méthode sont ici conditionnées par l'approximation quadratique plus ou moins précise de la fonction à optimiser.

➤ **La méthode du simplexe**

Contrairement aux méthodes présentées précédemment, la méthode du simplexe n'utilise pas les dérivées de la fonction coût. Présentée par Nelder et Mead en 1965 [51], elle fait appel à un polyèdre de $n+1$ cotés dans un espace de dimension n pour n paramètres composant \mathbf{x} . Ce polyèdre est appelé simplexe. La méthode évalue la fonction coût (χ^2 dans notre cas) en chacun des sommets ainsi construits. Dans le cadre d'une minimisation, celui correspondant à la plus forte évaluation est supprimé et un nouveau sommet est défini par des transformations

géométriques du polyèdre. Ainsi la fonction de coût tend à diminuer jusqu'à atteindre un critère d'arrêt défini à l'avance (nombre d'itérations, volume minimum du polyèdre ...).

La méthode du simplexe a l'avantage de ne pas être gourmande en mémoire, de ne pas être contrainte par les conditions d'existence des dérivées et d'être moins dépendante du point de départ car fonctionnant sur plusieurs solutions potentielles. Par contre, le temps de calcul est important, car elle nécessite parfois un grand nombre d'évaluations de la fonction à optimiser. Comme la plupart des techniques d'optimisation classique, cette méthode ne peut éviter le risque de convergence prématurée dans un minimum local.

La principale difficulté des méthodes d'optimisation classique présentées ici, est liée à la présence de minima locaux dans la fonction coût qui peut conduire à une convergence prématurée vers une solution non optimale. Il est donc souhaitable, lorsque c'est possible, de choisir une méthode capable de s'en affranchir.

I.2.4.3 Les méthodes utilisées en scattérométrie

➤ La méthode de Levenberg Marquart

L'algorithme proposé par Levenberg et Marquart [52] [53] est une technique standard utilisée dans la résolution des problèmes de minimisation dans plusieurs disciplines scientifiques. Cette méthode est une combinaison des méthodes de descente de gradient et de Gauss-Newton (cf paragraphe précédent I.2.4.2). Elle fait donc appel au calcul de dérivées premières et secondes pour récolter une information locale permettant de déplacer le point courant. C'est une méthode dite de trajectoire car elle nécessite la donnée d'une valeur initiale pour les paramètres recherchés. Pour faire simple, elle est similaire au gradient en début d'optimisation et devient proche de la méthode de Newton en fin d'optimisation si le minimum se rapproche d'une forme quadratique.

Ainsi, la direction de \mathbf{s} à l'itération k est donnée par l'expression :

$$[\mathbf{C}]' \mathbf{s} = \boldsymbol{\beta} \quad \text{I.37}$$

avec $\boldsymbol{\beta} = -\frac{1}{2} \nabla \chi^2$ et la matrice $[\mathbf{C}]'$ est donné par la relation :

$$[\mathbf{C}]' = [\mathbf{C}] + \lambda \mathbf{I}_d \quad \text{I.38}$$

dans la quelle $[\mathbf{C}] = \frac{1}{2} \mathbf{H}$ est appelée matrice de courbure.

Ainsi, si λ est très grand $[C]'$ est à diagonale dominante et l'équation I.37 devient similaire à l'équation I.33, on a :

$$\mathbf{s}_j = \frac{1}{\lambda C_{jj}} \mathbf{b}_{jj} = -\frac{1}{2\lambda C_{jj}} \nabla \chi^2 \quad \text{I.39}$$

A l'inverse lorsque λ tend vers 0, l'équation I.37 équivaut à celle de Gauss-Newton (Eq. I.35)

L'algorithme de Levenberg-Marquart est largement utilisé en scattérométrie pour résoudre le problème inverse [54] [55]. La méthode est rapide et efficace mais n'évite pas les minima locaux.

➤ **Algorithme génétique** [56]

C'est une métaheuristique qui se base sur l'évolution des espèces découverte par Charles Darwin. Les Algorithmes Génétiques (AG) [56] sont des méthodes qui permettent à une population de solutions potentielles qualifiées d'individus d'évoluer par l'intermédiaire d'opérateurs stochastiques vers un état qui optimise l'ensemble de leur aptitude défini par la fonction coût (pour nous χ^2). Cette évolution est fondée sur la sélection naturelle et la génétique (survie des bonnes solutions potentielles et échanges pseudo aléatoires entre individus).

Les AG travaillent sur plusieurs solutions potentielles en simultané et n'utilisent pas, comme le simplexe, d'information sur la dérivée. Ils sont réputés pour éviter les minima locaux et converger vers l'optimum global. Par contre, ils sont beaucoup plus lents que les méthodes classiques.

En scattérométrie, les AG ont montré une capacité remarquable dans la caractérisation optique et géométrique des couches minces à partir des données ellipsométriques [25]. Les travaux présentés dans [57] ont également montré l'efficacité des AG pour la caractérisation des réseaux de diffraction comparée à la méthode de Levenberg Marquardt. L'inconvénient majeur des AG, réside dans le fait qu'ils nécessitent de nombreuses fois le calcul direct qui, dans notre cas, peut prendre un temps non négligeable. Cette méthode n'est donc pas adaptée pour une caractérisation en temps réel par exemple.

➤ La méthode des bibliothèques

Elle consiste à construire une base de données regroupant des signatures scattérométriques. Chaque signature est issue de la méthode de simulation (par exemple la MMFE) et est indexée par le jeu unique de paramètres géométriques correspondants. La résolution du problème inverse consiste alors à comparer la signature mesurée à l'ensemble des signatures de cette base par l'évaluation du χ^2 . Ce fonctionnement est illustré sur la figure I.31.

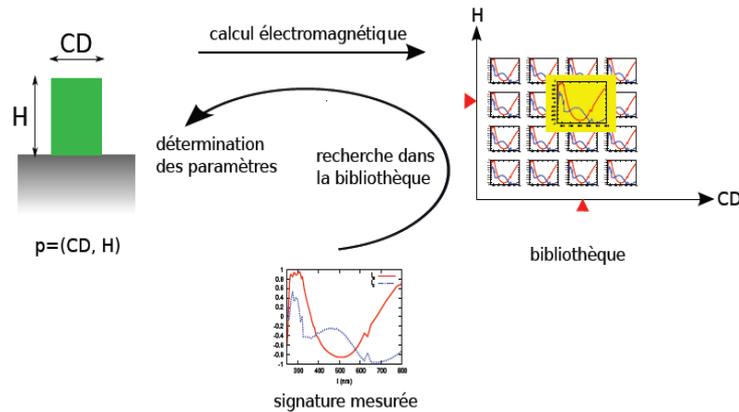


Figure I. 31 : Illustration de la méthode des bibliothèques pour un modèle de profil géométrique créneau (défini par 2 paramètres : largeur CD et hauteur h) [58]

Cette méthode qui peut s'apparenter à une méthode d'optimisation énumérative consiste à évaluer l'ensemble des solutions possibles et d'en extraire celle qui correspond au minimum du critère choisi. On extrait donc les paramètres géométriques qui possèdent la signature simulée de la base la plus proche de celle mesurée. Les paramètres géométriques variant de façon continue, il est impératif de fixer une variation minimale (un pas) pour chacun d'eux. Ce dernier conditionne grandement la précision de la méthode. En effet, le pas d'échantillonnage donne directement l'erreur minimale théorique que l'on peut atteindre sur chaque paramètre.

Cette méthode est performante, car elle évite les minima locaux si le pas est suffisamment faible. Cependant elle peut s'avérer lente, car elle exige un temps non négligeable pour scruter la table. L'utilisation de processeur graphique GPU peut réduire considérablement ce temps [58] [59]. La méthode avec GPU a été appliquée avec succès pour le suivi de procédés dynamiques comme le fluage de résine [60] [61] ou la réduction de cote de résine dans une application de gravure [62] [63].

Par ailleurs, la méthode des bibliothèques peut donc être efficacement utilisée pour les deux modes d'acquisition de l'ellipsomètre : c'est à dire en statique ou en dynamique. Ainsi elle est parfaitement adaptée pour une caractérisation en temps réel.

Par contre, la gestion de grande bibliothèque est assez lourde et nécessite une quantité de mémoire disponible importante. Ainsi, pour un changement de bibliothèque en cours de suivi d'un processus dynamique le temps de chargement de la nouvelle table peut s'avérer pénalisant. Les bibliothèques atteignent leurs limites lorsque le modèle géométrique devient complexe.

➤ **Les Réseaux de Neurones**

Comme évoqué précédemment, cette méthode s'apparente à une méthode de régression non linéaire. Le principe est simple. Un réseau de neurones (R.N), constitué d'une famille d'unités mathématiques simples (les neurones) apprend la relation liant la signature optique aux différents paramètres définissant le profil géométrique d'une structure afin de modéliser la fonction F^{-1} définie précédemment. Son vecteur d'entrée est l'ensemble des mesures y composant la signature optique et les sorties sont les paramètres géométriques x recherchés. Le réseau de neurones peut être défini mathématiquement comme une fonction algébrique, *a priori* non linéaire qui nécessite une étape d'apprentissage. En effet, on a besoin d'un grand nombre de données calculées préalablement qui constitue le corpus d'apprentissage. Cette étape est similaire à la construction d'une bibliothèque de signatures. La phase d'apprentissage peut prendre quelques minutes. Par contre, une fois cette étape terminée, le RN est capable de répondre presque instantanément à une nouvelle entrée inconnue (ici une signature mesurée) en effectuant une simple série de calculs internes. Notons qu'une fois le RN entraîné, il n'est pas nécessaire de conserver la base d'apprentissage. Les besoins en termes de capacité de stockage sont donc extrêmement limités.

Leur grande résistance au bruit de mesures et le fait de ne pas travailler sur la même fonction coût que les méthodes d'optimisation classique leur permettent d'être relativement moins sensibles aux minima locaux. La grande différence avec la méthode des bibliothèques réside dans le fait qu'ils ne sont pas limités à fournir un résultat présent dans la base d'apprentissage, mais sont capables d'interpoler entre 2 résultats proches.

Par contre, à l'inverse des méthodes d'optimisation et à l'instar de la méthode des bibliothèques, ils ne peuvent être efficacement utilisés que sur le domaine de variation des paramètres géométriques fixé dans la base d'apprentissage. Notons enfin que les méthodes d'optimisation nécessitent le calcul du problème direct (MMFE) à chaque itération alors que pour la bibliothèque et les RN le calcul est effectué en amont lors de la construction de la base. Cet avantage peut être efficacement exploité pour réduire fortement le temps de traitement d'une mesure ellipsométrique.

Plusieurs travaux ont montré l'efficacité des R.N dans la résolution du problème inverse en scattérométrie [9] [10] [64]. Les travaux présentés dans [13] ont montré que les R.N ont une meilleure robustesse au bruit de mesure par rapport à la technique des moindres carrés. Les résultats des R.N ont été confrontés à ceux obtenus avec différentes techniques morphologiques (AFM, XSEM) dans [11]. Dans [17] et [65] la signature ellipsométrique est associée aux R.N pour la caractérisation des réseaux de diffraction possédant différents profils géométriques. D'autres études ont montré que les R.N peuvent être efficacement adaptés à de structures géométriques plus complexes [66]. Les RN ont également permis l'étude qualitative et quantitative de la signature optique mesurée. En effet, certaines mesures présentes dans la signature peuvent présenter une information redondante ou peu sensible, voire indépendante des paramètres recherchés. La méthode HVS (pour Heuristic Variable Sélection) est alors utilisée pour permettre au RN de sélectionner à la fois la signature et l'architecture neuronale optimale [67]. Cela conduit à une réduction significative du nombre d'entrées et, par la même occasion, du temps nécessaire pour la mesure. Son application peut rendre alors plus rapide le contrôle de l'homogénéité d'une structure [68].

Au sein du laboratoire, l'équipe « surface complexe » continue à développer un savoir-faire dans le domaine de la résolution de problème inverse et notamment en caractérisation scattérométrique par RN [11] [17] [64] [65] [67] [68]. Deux thèses ont validé l'utilisation des RN dans le cadre d'une caractérisation statique [57] [3]. Une autre a étudié les potentialités des R.N dans la caractérisation en temps réel [36]. En marge de la scattérométrie, d'autres études ont montré que le RN pouvait constituer un outil rapide et efficace pour la détermination des indices de réfraction et des épaisseurs dans le cas de la caractérisation de couches minces [69].

Nous détaillerons plus loin dans le chapitre qui va suivre la structure des RN utilisés dans la caractérisation ellipsométrique au sein du laboratoire. Il faut noter que leur utilisation en scattérométrie s'est presque exclusivement concentrée sur leur capacité de régresseur. Le but de cette thèse consiste en partie à mettre à profit un autre aspect de ces outils mathématiques : leur aptitude à la classification.

Dans ce chapitre, nous avons montré les fondements de base de la caractérisation par scattérométrie. Nous avons présenté le contexte de la métrologie optique, le principe de la mesure ellipsométrique et le type de traitement utilisé dans la résolution d'un problème inverse en se focalisant sur celui utilisé dans le domaine. La méthode utilisée pour le calcul direct des signatures optiques a également été présentée. Par ailleurs, nous avons mis en évidence certains

besoins typiques en termes de caractérisation dans un processus de fabrication lithographique. Les deux exemples cités dans ce chapitre (détection de défauts, reconnaissance de type de profils géométriques) vont nous servir de référence dans ces travaux de thèse pour étudier l'apport d'une méthode de classification par RN dans un processus de scattérométrie. Le chapitre suivant présente donc tout naturellement un état de l'art des méthodes de classification automatique.

Chapitre II Classification Statistique des Formes

L'être humain possède une immense capacité à percevoir et à interpréter l'environnement. En effet, il arrive à reconnaître, des visages, des sons, des signaux temporels, des objets et ceci indépendamment de leur variabilité et des conditions d'observation. La classification est une discipline de la Reconnaissance De Formes (RDF) qui vise à reproduire ces capacités de l'homme. Elle regroupe les méthodes visant à identifier des motifs à partir de données brutes ou aidant la prise de décision d'attribution de ces derniers à un groupe. Elle fait largement appel aux techniques d'apprentissage automatique, de statistique et même d'intelligence artificielle. Le mot « *forme* » est très général, car les objets à reconnaître peuvent être de natures très diverses : code barre, visage, empreinte digitale, son issu de la parole ou image médicale.

Les techniques de classification sont très nombreuses. Dans nos travaux, les formes à identifier sont les signatures optiques issues des structures périodiques énoncées au chapitre I et la technique utilisée est neuronale. Toutefois, les étapes inhérentes à l'identification sont communes, quelle que soit la méthode de classification mise en place. Ainsi, il convient d'abord de rappeler brièvement les principes de la classification automatique, puis de détailler celle que nous utiliserons dans nos travaux. Le présent chapitre est structuré comme suit : nous présenterons d'abord les généralités de la classification automatique en fournissant au lecteur les bases pour aborder les différentes approches existantes. Nous décrirons ensuite deux approches dans la classification statistique : la première dite directe est basée sur le théorème de Bayes. Elle estime les probabilités *a posteriori* connaissant les probabilités *a priori* et les densités de probabilité. Nous ferons la distinction entre les méthodes paramétriques et celle non paramétrique. La seconde quant à elle qualifiée d'indirecte, calcule les probabilités *a posteriori* d'appartenance aux classes sans passer par la phase d'estimation des densités de probabilité. Enfin, dans la dernière partie du chapitre, nous nous attarderons plus en détail sur les réseaux de neurones qui seront développés et appliqués à la scattérométrie dans le chapitre III.

II.1 Le domaine de la reconnaissance de formes

Dans ce qui suit, nous considérons qu'une forme est représentée par un vecteur $\mathbf{X} = \{x_1, x_2, \dots, x_n\}$ de n composantes correspondant à la mesure de n caractéristiques distinctes. Tous les vecteurs qui représentent l'ensemble des objets peuvent alors être positionnés par un point dans l'espace euclidien \mathbb{R}^n . Ces derniers peuvent être regroupés en « amas » associé chacun à une

classe particulière. Une classe peut donc être définie comme un ensemble de vecteurs ayant des caractéristiques semblables. Classifier un ensemble de vecteurs consiste à regrouper ces vecteurs en K classes C_k . On parle alors de **clustering ou regroupement** si les classes ne sont pas définies à l'avance et de **classification (ou classement ou encore discrimination)** dans le cas contraire.

II.1.1 Généralité

On ne peut concevoir directement un classifieur capable de prédire instantanément la classe d'un objet inconnu. En général, on construit *a priori* une famille de fonctions paramétrées par un certain nombre de poids dont l'ajustement ultérieur permettra une adaptation particulière aux données à traiter. Ceci impose par conséquent une phase d'apprentissage sur la base d'un ensemble d'échantillons donnés appelé également corpus. L'apprentissage est nécessairement de nature statistique et les critères de ce dernier sont souvent assortis de certaines notions d'optimisation.

On distingue l'apprentissage supervisé, qui s'opère dans le cas où les classes des échantillons du corpus sont connues, de l'apprentissage non supervisé, qui s'effectue dans le cas du clustering où on cherche à organiser les données selon leur ressemblance sans connaissance *a priori*.

De façon pratique, on cherche à vérifier les deux propriétés suivantes :

- **Compacité** : les points X appartenant à une classe donnée sont plus proches entre eux que des points de toutes les autres classes.
- **Séparabilité** : les classes sont bornées et il n'y a pas de recouvrement entre elles.

La séparabilité peut être linéaire ou non linéaire (figure II.1).

En pratique, on ne peut respecter scrupuleusement ces deux propriétés en partie à cause du bruit et la distorsion des formes.

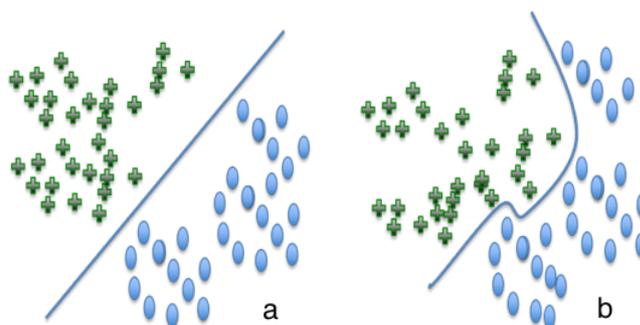


Figure II.1: Séparabilité linéaire (a) et non linéaire (b) entre 2 classes définies dans un espace à 2 dimensions C_1 (représentée par des « + ») et C_2 (représentée par des « o »)

Il existe diverses méthodes de classification automatique. Beaucoup d'ouvrages ont été publiés sur le sujet [70]. Nous nous limiterons à énoncer les principes généraux communs à toutes ces méthodes.

Dans la littérature, les méthodes de regroupement sont réparties en deux grandes catégories : les méthodes hiérarchiques et les méthodes non hiérarchiques.

Historiquement, les méthodes de regroupement hiérarchique sont les premières à être développées. Elles sont inspirées de la taxonomie biologique. Ces méthodes sont soit divisives soit agglomératives. Les premières partent d'un groupe formé de N échantillons complets et le subdivisent récursivement en sous-groupes plus spécifiques. Les secondes quant à elles procèdent dans le sens ascendant, en effectuant un regroupement des échantillons en petits agrégats puis en les fusionnant à des niveaux supérieurs. Les résultats d'un regroupement hiérarchique sont généralement présentés au moyen d'une structure arborescente appelée dendrogramme (figure II.2).

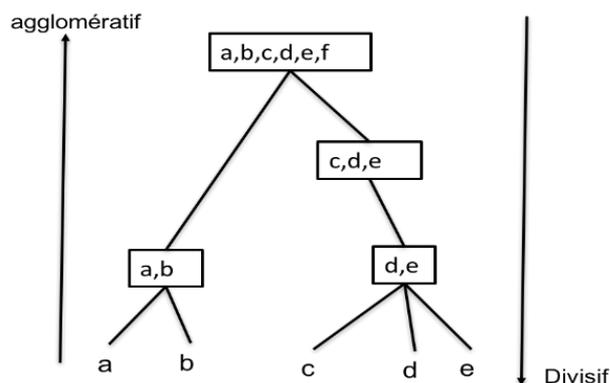


Figure II.2: Exemple de structure d'un dendrogramme hiérarchique pour différents échantillons notés (a , b , c , d , e et f)

L'avènement des ordinateurs a fait perdre une certaine popularité de ces méthodes au profit des méthodes non hiérarchiques qualifiées aussi de méthodes de partitionnement. Contrairement aux méthodes hiérarchiques, le nombre de classes doit être défini à l'avance. Par contre, chacune des classes n'est pas identifiée. Elles fournissent ainsi, une partition unique des éléments à classer. Le principe consiste à déterminer sur un ensemble d'échantillons à l'aide d'opérateurs adéquats une partition représentant au mieux les divers regroupements pouvant exister au sein de ce dernier. Notons qu'une des méthodes neuronales présentées dans cette thèse fait partie des méthodes non hiérarchiques : les réseaux de Kohonen (cf paragraphe II.3.2.7).

Quant aux méthodes de classification, elles diffèrent entre elles par le choix du critère de ressemblance et de la façon de mesurer ce dernier entre les classes. Ainsi la décision d'attribuer un vecteur à une classe est fondée sur la notion de « proximité ». Le plus souvent, un opérateur de mesure est utilisé pour évaluer cette dernière. Il peut ainsi s'agir de la mesure de la ressemblance entre les formes ou inversement la dissemblance entre les formes (par un simple calcul de distance par exemple [71]). Cette notion est mise à profit afin de formuler une décision par l'intermédiaire d'une fonction de discrimination. Cette dernière détermine, parmi un ensemble fini, la classe à laquelle appartient un objet donné. Elle doit être en mesure de modéliser au mieux les frontières qui séparent les classes les unes des autres. Il existe deux approches pour la classification statistique : l'approche indirecte et l'approche directe.

La première permet de prendre la décision en se basant sur la formule de Bayes d'où également le nom d'approche bayésienne. On cherche d'abord à établir un modèle statistique des données qui estime la probabilité $p(C_k/\mathbf{X})$ d'appartenance à la classe C_k connaissant le vecteur \mathbf{X} puis on s'appuie sur ces probabilités *a posteriori* pour formuler la décision. Les méthodes utilisant ces approches se divisent en méthodes paramétriques et non paramétriques. Les premières supposent que le modèle de distribution des données est connu et on cherche à en estimer les paramètres tandis que les secondes ne font pas cette d'hypothèse et se basent uniquement sur des propriétés de convergence pour effectuer la décision [72].

La seconde approche indirecte (ou également discriminante) permet, quant à elle, de calculer directement la fonction de décision qui retourne la classe C_k d'appartenance pour un vecteur \mathbf{X} fourni sans faire intervenir la formule de Bayes. Notons qu'une des méthodes neuronales présentées dans cette thèse utilise cette approche : le perceptron multi couche (cf paragraphe II.3.2.6).

II.1.2 Prétraitement des données

La construction d'un système classifieur est généralement précédée d'une étape préparatoire de traitement des données à analyser. Cette dernière est primordiale dans la classification et conditionne les performances de la méthode. En effet, les données d'entrée peuvent parfois être de nature variée ou présentant des éléments de différents ordres de grandeur pouvant introduire un biais dans l'analyse. Le vecteur d'entrée \mathbf{X} peut également posséder des informations redondantes qui peuvent altérer les décisions de classement. C'est pourquoi il est préférable d'effectuer un prétraitement afin de limiter ces problèmes.

Par ailleurs, l'aspect qualitatif de \mathbf{X} doit également être pris en compte. En effet, il est

important de considérer uniquement les informations susceptibles d'influencer la décision d'affectation à une classe. Une extraction et/ou une sélection des caractéristiques de \mathbf{X} les plus pertinentes pour les classes s'avèrent nécessaires. Pour cela, on utilise des techniques d'analyse de données [73]. Parmi celles-ci, l'Analyse en Composantes Principales (ACP) [74] et l'analyse discriminante factorielle [75] sont les méthodes linéaires les plus utilisées. Pour des corrélations de données non linéaires, nous pouvons citer l'utilisation de la projection de Sammon [76]. Les cartes auto-organisatrices citées dans cette thèse sont également considérées comme des outils de projection non linéaire et peuvent être utilisées dans ce but.

Dans nos travaux, rappelons que les données d'entrée \mathbf{X} correspondent à des signatures optiques (cf Chapitre I). Elles sont donc de nature semblable tout en présentant cependant un caractère bruité. Ainsi, nous effectuons par la suite uniquement une simple normalisation. Toutes les données seront centrées et réduites (moyenne proche de 0 et écart type de 1) avant d'être exploitées. Ainsi pour une composante quelconque x_i , on effectue un changement de variable en x_i' par la transformation suivante :

$$x_i' = \frac{x_i - \mu_i}{\sigma_i} \quad \text{II.1}$$

où μ_i et σ_i sont respectivement la moyenne et l'écart type calculés sur chacune des composantes x_i du vecteur \mathbf{X} .

Nous ne traiterons pas la sélection des mesures les plus pertinentes pour la discrimination car la composition de la signature ellipsométrique est fixée par l'étape de caractérisation proprement dite.

A l'issue du prétraitement nous pouvons procéder à la construction du classifieur. Avant de détailler les 2 grandes approches sur lesquelles sont fondés ces outils, il est nécessaire de définir un critère de proximité.

II.1.3 Une mesure de proximité : La distance

Dans cette thèse, nous utilisons la distance comme le critère permettant d'évaluer la proximité entre 2 échantillons. C'est l'une des notions les plus simples utilisées en classification automatique. Ainsi, 2 vecteurs sont d'autant plus proches que la distance entre eux est plus faible.

D'un point de vue mathématique, une distance (ou métrique) est une application d définie de l'espace des données $E \times E$ dans \mathbb{R}^+ satisfaisant les propriétés suivantes :

- Séparabilité : $\forall \mathbf{X}, \mathbf{Y} \in E^2, \mathbf{X} \neq \mathbf{Y} \Rightarrow d(\mathbf{X}, \mathbf{Y}) > 0$

- Réflexivité : $\forall \mathbf{X} \in E, d(\mathbf{X}, \mathbf{X}) = 0$
- Symétrie : $\forall \mathbf{X}, \mathbf{Y} \in E^2, d(\mathbf{X}, \mathbf{Y}) = d(\mathbf{Y}, \mathbf{X})$
- Inégalité triangulaire : $\forall \mathbf{X}, \mathbf{Y}, \mathbf{Z} \in E^3, d(\mathbf{X}, \mathbf{Z}) \leq d(\mathbf{X}, \mathbf{Y}) + d(\mathbf{Y}, \mathbf{Z})$

La définition de la distance n'est pas unique et dépend de la forme à traiter. Il existe plusieurs possibilités d'exprimer la distance entre deux points. Ainsi, dans le cas de variable quantitative, la plus connue est la mesure de Minkowski ou p distance.

L'expression de la distance entre 2 vecteurs $\mathbf{W}=(w_1, w_2, \dots, w_n)$ et $\mathbf{X}=(x_1, x_2, x_n)$ de n composantes dans un ensemble E est donnée par :

$$d_p(\mathbf{W}, \mathbf{X}) = \left[\sum_{i=1}^n |w_i - x_i|^p \right]^{\frac{1}{p}} \quad \text{II.2}$$

Ainsi avec $p=1$, on retrouve la distance de Manhattan connue aussi sous le nom de distance « city-block » :

$$d_1(\mathbf{W}, \mathbf{X}) = \sum_{i=1}^n |w_i - x_i| \quad \text{II.3}$$

Pour $p= 2$, on retrouve la distance euclidienne qui est la plus utilisée en pratique :

$$d_2(\mathbf{W}, \mathbf{X}) = \sqrt{\sum_{i=1}^n |w_i - x_i|^2} \quad \text{II.4}$$

Lorsque p est infini, l'expression correspond à la distance de Chebychev:

$$d_\infty(\mathbf{W}, \mathbf{X}) = \max_i |w_i - x_i| \quad \text{II.5}$$

La figure II.3 illustre les différentes distances de la métrique de Minkowski entre 2 points A et B de coordonnées respectives (x_A, y_A) et (x_B, y_B) dans un espace E à 2 dimensions.

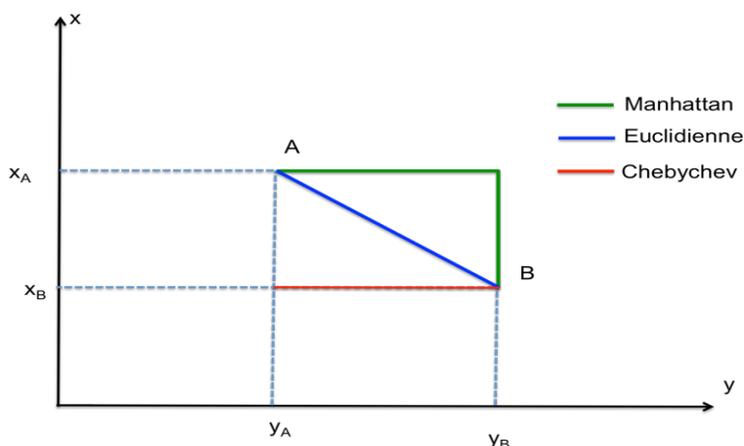


Figure II.3: Exemple de 3 métrique de Minkowski entre 2 points A et B

Il faut noter que le choix de la distance conditionne la forme des classes obtenues. Par exemple, la distance euclidienne contribue à la formation de classes hypersphériques alors que celle de Chebychev implique des formes hypercubiques. La figure II.4 illustre la forme de l'ensemble des points équidistants centrés sur un point fixe dans un espace tridimensionnel, dans le cas des différentes distances citées.

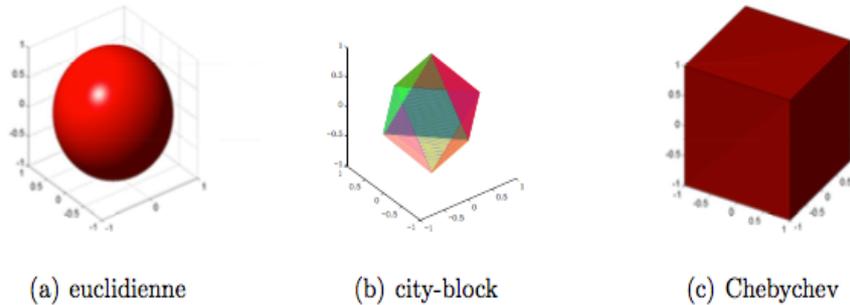


Figure II.4: Exemples de forme obtenue suivant le choix de la métrique employée dans le cas tridimensionnel [77]

On peut également tenir compte de l'importance relative de chacune des composantes caractéristiques i des vecteurs de l'espace E . Ainsi l'expression de la distance de Minkowski définie ci-dessus peut être réécrite comme suit :

$$d_p(\mathbf{W}, \mathbf{X}) = \left[\sum_{i=1}^N \alpha_i |w_i - x_i|^p \right]^{\frac{1}{p}} \quad \text{II.6}$$

où α_i est le poids associé à la composante i .

Une autre distance intéressante utilisée dans la classification est la distance de Mahalanobis. Elle diffère des autres distances citées précédemment par le fait qu'elle prend en compte la variance et la corrélation des données. Elle s'exprime par la relation suivante :

$$d_M(\mathbf{W}, \mathbf{X}) = \sqrt{(\mathbf{W} - \mathbf{X})^T \Sigma^{-1} (\mathbf{W} - \mathbf{X})} \quad \text{II.7}$$

où Σ est la matrice de covariance associée aux vecteurs \mathbf{W} et \mathbf{X} . Elle est estimée par la relation suivante :

$$\Sigma = \frac{1}{N} \sum_{i=1}^N (\mathbf{W}_i - \boldsymbol{\mu}_w)(\mathbf{X}_i - \boldsymbol{\mu}_x)^T \quad \text{II.8}$$

Ainsi si \mathbf{W} et \mathbf{X} sont indépendants, Σ correspond à la matrice unité et d_M est simplement égale à la distance euclidienne. La distance de Mahalanobis est utilisée dans le cas où les différentes composantes des vecteurs ne sont pas forcément homogènes. La forme des classes

produites en utilisant cette distance est de type ellipsoïdal. Notons également que l'utilisation de cette distance peut poser des problèmes quant à l'inversion de la matrice de covariance.

II.2 Approche indirecte de classification

Dans cette approche de la classification, les composantes des vecteurs \mathbf{X} et les classes C_k du corpus d'apprentissage sont respectivement définies par des variables aléatoires continues et par des variables aléatoires discrètes. On définit alors une densité de probabilité conditionnelle pour chaque vecteur \mathbf{X} de la classe C_k par $f(\mathbf{X}/C_k)$. Chaque classe C_k est définie par une probabilité *a priori* notée $p(C_k)$ supposée connue qui est la probabilité de tirer au hasard un échantillon de C_k parmi la population d'objets à classer.

Les méthodes de classification indirectes utilisent le théorème de Bayes pour calculer les vraisemblances entre les classes. Ce dernier permet effectivement de calculer les probabilités *a posteriori* $p(C_k/\mathbf{X})$ d'obtention de la classe C_k alors que \mathbf{X} est réalisé par la relation suivante :

$$p(C_k | \mathbf{X}) = \frac{p(C_k)f(\mathbf{X}|C_k)}{\sum_{k=1}^K p(C_k)f(\mathbf{X}|C_k)} \quad \text{II.9}$$

où K est le nombre total de classes.

La règle de décision de Bayes consiste à affecter à l'objet la classe dont la probabilité *a posteriori* est la plus grande :

$$\mathbf{X} \in C_k \Leftrightarrow p(C_k|\mathbf{X}) > p(C_j|\mathbf{X}) \quad \forall k \neq j \quad \text{II.10}$$

La décision bayésienne est considérée optimale. En effet, le risque d'erreur de classement n'est pas nul mais, il a été montré qu'il est alors minimisé [72]. Cette règle est géométriquement illustrée sur la figure II.5 dans laquelle on représente les probabilités $p(C_1/\mathbf{X})$ et $p(C_2/\mathbf{X})$ correspondants respectivement à deux classes distinctes C_1 et C_2 . Le seuil de classification donné par la règle de décision de Bayes est marqué par le trait vertical rouge à partir duquel le choix d'affectation passe de la classe C_1 à la classe C_2 .

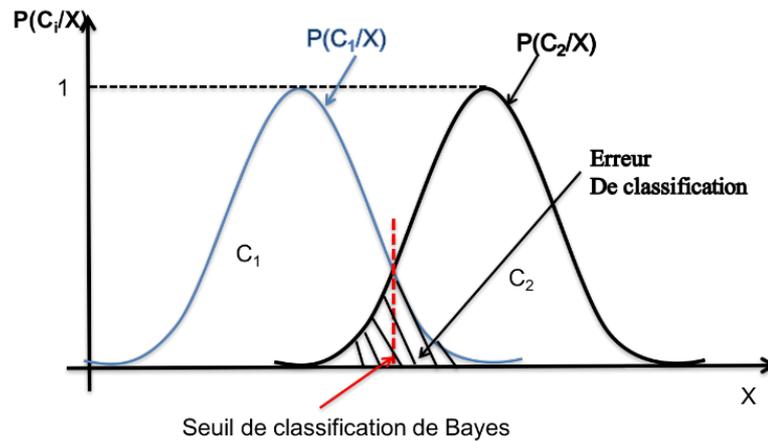


Figure II.5: Interprétation géométrique de la règle de décision de Bayes concernant le classement d'un vecteur X en 2 classes distinctes (C_1 et C_2) à partir des probabilités a posteriori.

Toutefois, il est malheureusement difficile de pouvoir estimer de manière précise les densités de probabilité $f(\mathbf{X}/C_k)$ des classes C_k . Ainsi, ces dernières peuvent être estimées par deux types de méthodes différentes : les méthodes paramétriques et celles non paramétriques. Les premières supposent que les lois de probabilités obéissent à un modèle de distribution connu (par exemple gaussien) dont les paramètres peuvent être déterminés à partir des exemples disponibles composant le corpus de départ. Les secondes quant à elles n'utilisent aucune connaissance ou hypothèse a priori sur les distributions (méthode des noyaux, méthode des plus proches voisins).

II.2.1 Estimation paramétrique des densités de probabilité : Le classifieur gaussien

Le classifieur gaussien suppose que les éléments de chaque classe suivent une loi de distribution gaussienne multivariée. Les paramètres à estimer sont donc les moyennes $\boldsymbol{\mu}_k$ et les écarts types σ_k correspondant à chaque élément de \mathbf{X} pour la classe C_k . Les densités des probabilités alors peuvent être estimées comme suit :

$$f(\mathbf{X}|C_k) = \frac{1}{(2\pi)^{\frac{n}{2}}(\det\Sigma_k)^{\frac{1}{2}}} \exp\left[-\frac{1}{2}(\mathbf{X} - \boldsymbol{\mu}_k)^T \Sigma^{-1}(\mathbf{X} - \boldsymbol{\mu}_k)\right] \quad \text{II.11}$$

avec Σ_k est la matrice de covariance associé a \mathbf{X} et $\boldsymbol{\mu}_k$

II.2.2 Estimation non paramétrique des densités de probabilité : La méthode des K plus proches voisins (K-ppv)

Les méthodes non paramétriques permettent de converger vers la distribution des classes sans aucune hypothèse *a priori*. Le principe de l'estimation de $f(\mathbf{X}|C_k)$ repose d'abord sur la délimitation d'une région R autour d'un vecteur \mathbf{X} fixé. Ensuite on compte le nombre d'exemples d'individus dans le volume V défini par R . Enfin on estime la densité de probabilité comme suit :

$$\hat{f}(\mathbf{X}) = \frac{K_N}{NV} \quad \text{II.12}$$

avec N le nombre total d'exemples et K_N le nombre d'exemples contenu dans la région R

Trois conditions doivent garantir la convergence de $\hat{f}(\mathbf{X})$ vers $f(\mathbf{X})$:

$$\lim_{N \rightarrow \infty} V = 0 \quad \text{II.13}$$

$$\lim_{N \rightarrow \infty} K_N = 0 \quad \text{II.14}$$

$$\lim_{N \rightarrow \infty} \frac{K_N}{N} = 0 \quad \text{II.15}$$

Il existe deux méthodes très répandues qui utilisent ce principe de fonctionnement : les noyaux de Parzen [78] (pour lesquels V est fixé et on cherche K_N) et la méthode des K_N -plus proches voisins [79] (où les K_N individus sont fixés et on recherche le volume nécessaire pour les contenir). En guise d'exemple, nous décrirons uniquement la méthode des K_N -plus proches voisins.

La méthode des K_N -plus proches voisins (K-PPV) consiste à assigner une classe C_k à un objet \mathbf{X} si celle-ci est la plus représentée parmi les K_N plus proches voisins de \mathbf{X} . On choisit généralement des volumes réguliers centrés en \mathbf{X} (hypercubes, hypersphère) d'où l'utilisation de la notion de distance (cf paragraphe II.1.3) pour définir le type de voisinage.

Supposons que le corpus d'apprentissage de taille N comprend N_k individus dans une classe C_k et que $K_N = \sqrt{N}$, on a donc $\sum_{k=1}^K N_k = N$. Pour classer un nouveau vecteur \mathbf{X} , on considère donc une hypersphère centrée en \mathbf{X} . Soit V le volume de cette sphère qui contient K_k objets appartenant à la classe C_k parmi les K_N fixés. D'après la relation II.12 on a :

$$\hat{f}(\mathbf{X}|C_k) = \frac{K_k}{N_k V} \quad \text{II.16}$$

De même, le nombre d'individus N_k de chaque classe C_k nous donne :

$$p(C_k) = \frac{N_k}{N} \quad \text{II.17}$$

Ainsi nous pouvons combiner les deux relations II.16 et II.17 en appliquant la formule de Bayes II.9 pour donner la probabilité *a posteriori* :

$$p(C_k|\mathbf{X}) \cong \frac{K_k}{\sum_{k=1}^K K_k} \cong \frac{K_k}{K_N} \quad \text{II.18}$$

La décision d'affectation est alors prise de la même manière que précédemment en sélectionnant la classe qui possède la probabilité *a posteriori* la plus élevée (Eq.II.10).

Devijver et Kittler [80] ont montré que pour un nombre d'échantillons N tendant vers l'infini le taux d'erreur converge vers le minimum absolu qui est celui assuré par le classifieur de Bayes.

Une variante de la méthode consiste à introduire un seuil de réponse S ($K_N/2 < S \leq K_N$). Ensuite la décision d'attribuer \mathbf{X} à la classe C_k est effective si au moins S parmi ses K_N plus proches voisins appartiennent à C_k . Si aucune classe ne vérifie ce critère alors une décision de rejet est prise.

En pratique la méthode K_N -PPV nécessite un grand nombre N d'échantillons constituant la base de départ. Elle est également très coûteuse en temps de calcul. En effet, pour trouver les plus proches voisins d'un vecteur \mathbf{X} , il faut comparer ce dernier à tous les échantillons de la base. Le calcul de N distances est ainsi effectué dans un espace de n dimensions pour prendre la décision.

Dans la littérature, on retrouve quelques variantes accélérées de la méthode. En effet, plusieurs auteurs ont proposé des algorithmes permettant de réorganiser le corpus de départ en vue de réduire le temps de calcul. Ainsi, Delannoy développe un algorithme fondé sur l'idée de pavage où l'espace est divisé en plusieurs cellules. La recherche des voisins est alors limitée à quelques une d'entre elles seulement [81]. Fukunaga et Narendra quant à eux ont proposé un algorithme fondé sur la notion de hiérarchie, où les échantillons de la base de départ sont regroupés à différents niveaux [82] [83]. Kim et Park quant à eux ont mis au point un algorithme basé sur l'idée du tri des objets de la base sur chaque axe, ce qui limite la recherche à quelques dimensions [84] [85] [86]. Ils existent également d'autres algorithmes basés sur la notion de condensation visant également à réduire l'espace total en un sous-espace [87]. L'efficacité de tous ces algorithmes dépend de la dimension n de l'espace des formes. En effet, cette dernière décroît lorsque n augmente [88] [85]. Miclet et Dabouz [89] ont montré aussi que le choix de la

métrique à également beaucoup d'influence sur le taux d'erreur et de rejet pour un nombre N d'échantillons fixés.

II.3 Approche directe de classification

Dans ce paragraphe, nous présentons l'approche directe dans la résolution des problèmes de classification supervisée. Nous exposons les fondements théoriques des méthodes les plus utilisées dans la littérature à savoir les séparateurs à vastes marges SVM basés sur l'utilisation des noyaux et les réseaux de neurones fondés sur un modèle connexionniste que nous utiliserons par la suite.

II.3.1 Les Séparateurs à Vastes Marges

Les SVM sont des techniques de classification binaire fondées sur l'apprentissage supervisé. Le principe repose sur l'existence d'un classifieur linéaire dans un espace approprié qui fournit des solutions optimales dites à « marge maximale ». Dans ce paragraphe, nous présentons le cas classique des SVM [90] pour un problème de classification à 2 classes.

II.3.1.1 Principe des SVM linéaires

Considérons un corpus d'apprentissage $D = \{(\mathbf{X}_1, Y_1), \dots, (\mathbf{X}_N, Y_N)\}$ linéairement séparable en 2 classes C_1 et C_2 étiquetées comme suit :

$$D = \{(\mathbf{X}_1, Y_1), \dots, (\mathbf{X}_N, Y_N)\}, \mathbf{X}_i \in \mathbb{R}^n, Y_i \in \{-1, +1\}$$

$$C_1 = \{\mathbf{X}_i \text{ tel que } Y_i = -1\} \text{ et } C_2 = \{\mathbf{X}_i \text{ tel que } Y_i = +1\}$$

En munissant l'espace \mathbb{R}^n du produit scalaire, Les SVM visent à séparer les exemples de chacune des 2 classes C_1 et C_2 par un hyperplan d'équation :

$$\langle \mathbf{W}, \mathbf{X} \rangle + w_0 \tag{II.19}$$

tel que :

$$\forall \mathbf{X} \in C_2 \quad \langle \mathbf{W}, \mathbf{X} \rangle + w_0 > 0 \tag{II.20}$$

$$\forall \mathbf{X} \in C_1 \quad \langle \mathbf{W}, \mathbf{X} \rangle + w_0 < 0 \tag{II.21}$$

avec $\mathbf{W} \in \mathbb{R}^n$, $w_0 \in \mathbb{R}$ et $\langle \mathbf{W}, \mathbf{X} \rangle = \mathbf{W}^T \mathbf{X}$, le produit scalaire entre les vecteurs \mathbf{W} et \mathbf{X} ;
Ainsi, le vecteur \mathbf{W} définit la pente de l'hyperplan et le terme w_0 quant à lui permet de translater l'hyperplan parallèlement à lui-même. En général, il existe une infinité d'hyperplans

vérifiant ces conditions. L'objectif est donc de chercher l'hyperplan qui sépare au mieux les deux nuages de points en garantissant une marge maximale, c'est-à-dire une distance maximale entre une frontière de séparation et les exemples les plus proches de celle-ci appartenant à chaque classe. La marge ainsi définie correspond donc à la distance euclidienne entre l'exemple \mathbf{X} et sa projection sur un hyperplan particulier défini par l'équation II.19. Elle est normalisée et s'exprime par :

$$\frac{|\langle \mathbf{W}, \mathbf{X} \rangle + w_0|}{\|\mathbf{W}\|} \quad \text{II.22}$$

L'Hyperplan optimal H défini par (\mathbf{W}^*, w_0^*) est celui qui maximise la marge. Il est solution de l'équation suivante :

$$\arg \max_{\mathbf{W}^*, w_0^*} (\min_i \{\|\mathbf{X} - \mathbf{X}_i\| \text{ tel que } \mathbf{X} \in \mathbb{R}^n, (\mathbf{W}^T \mathbf{X} + w_0) = 0, i = 1, \dots, N\}) \quad \text{II.23}$$

Les données étant linéairement séparables, on peut contraindre les exemples de chaque classe à satisfaire les équations suivantes pour un couple \mathbf{W}, w_0 particulier :

$$\langle \mathbf{W}, \mathbf{X} \rangle + w_0 \geq 0 \quad \text{si } Y_i = +1 \quad \text{II.24}$$

$$\langle \mathbf{W}, \mathbf{X} \rangle + w_0 \leq 0 \quad \text{si } Y_i = -1 \quad \text{II.25}$$

Ces deux inégalités peuvent être combinées en une seule inégalité :

$$Y_i \langle \mathbf{W}, \mathbf{X} \rangle + w_0 \geq 0 \quad \forall i \quad \text{II.26}$$

On définit alors les deux hyperplans particuliers suivants appelés hyperplans canoniques contenant les échantillons à la marge de chaque classe :

$$H_1 : \langle \mathbf{W}^*, \mathbf{X} \rangle + w_{01} = +1 \quad \text{II.27}$$

$$H_2 : \langle \mathbf{W}^*, \mathbf{X} \rangle + w_{02} = -1 \quad \text{II.28}$$

D'après ces équations il n'existe aucun échantillon \mathbf{X} situé entre les deux plans. Les points du corpus d'apprentissage se trouvant sur H_1 et H_2 sont appelés les vecteurs supports. La distance entre H_1 et H_2 équivaut à : $2/\|\mathbf{W}^*\|$. En effet tout \mathbf{X} appartenant à H_1 vérifie l'équation II.27 donc la marge vaut d'après l'équation II.22 $1/\|\mathbf{W}^*\|$. Il en est de même pour ceux appartenant à H_2 et vérifiant l'équation II.28, et pour lesquels la marge vaut également $1/\|\mathbf{W}\|$ (figure II.6).

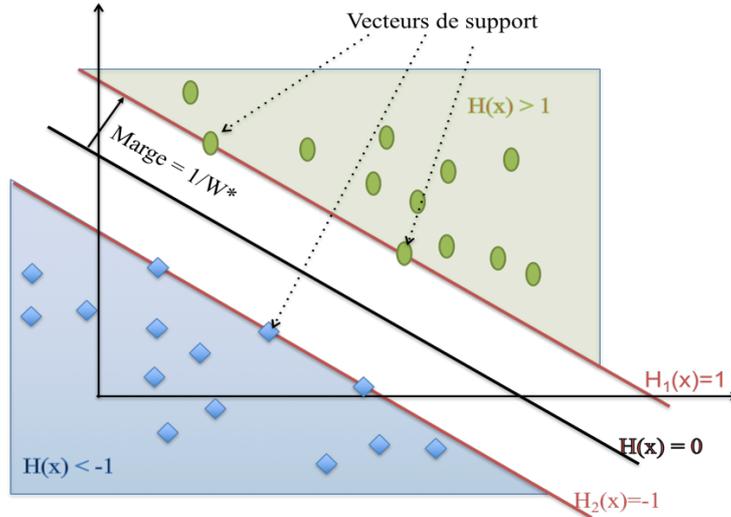


Figure II.6: Définition de l'hyperplan optimal et de la marge d'un classifieur SVM dans le cas d'une classification en 2 classes linéairement séparable dans un espace à 2 dimensions

Ainsi, le problème posé ne dépend que des vecteurs supports, car si les autres points sont éliminés la solution reste la même. Ainsi, trouver l'hyperplan optimal revient à chercher la marge maximale donc à résoudre le problème d'optimisation suivant :

$$\begin{cases} \text{minimiser } \tau(\mathbf{W}) = \frac{1}{2} \|\mathbf{W}\|^2 \\ \text{sous les contraintes } Y_i(\langle \mathbf{W}, \mathbf{X}_i \rangle + w_0) - 1 \geq 0 \quad \forall i = 1, \dots, N \end{cases} \quad \text{II.29}$$

Cette écriture est appelée formulation primale des SVM. Elle implique la détermination de $N+1$ paramètres.

Pour résoudre ce type d'optimisation avec contraintes, on utilise une fonction appelée Lagrangien définie comme la somme de la fonction objective et d'un terme correspondant à une combinaison linéaire des contraintes, dont les coefficients notée $\alpha_i \geq 0$ sont appelés multiplicateurs de Lagrange :

$$L(\mathbf{W}, w_0, \alpha_i) = \frac{1}{2} \|\mathbf{W}\|^2 - \sum_{i=1}^N \alpha_i [Y_i(\langle \mathbf{W}, \mathbf{X}_i \rangle + w_0) - 1] \quad \text{II.30}$$

avec $\boldsymbol{\alpha} = [\alpha_1, \dots, \alpha_N]^T$

Cette dernière doit être minimisée par rapport aux paramètres \mathbf{W} et w_0 et maximiser par rapport à $\boldsymbol{\alpha}$. La résolution est possible avec les méthodes classiques d'optimisation décrite au

paragraphe I.2.4.2 (chapitre I) pour n assez petit, mais devient inenvisageable pour des valeurs de n très élevé.

Ainsi, il existe une autre transformation de ce problème connue sous le nom de formulation duale. En effet, le théorème de Kuhn-Tucker [91] montre que le problème primal (Eq. II.29) et sa formulation duale (Eq. II.30) possèdent la même solution. Pour obtenir cette dernière il faut écrire la condition d'optimalité du premier ordre (annulation du gradient) sur les variables primales (\mathbf{W} et w_0).

Il en découle donc :

$$\sum_{i=1}^N \alpha_i Y_i = 0 \quad \text{II.31}$$

$$\sum_{i=1}^N \alpha_i Y_i \mathbf{X}_i = \mathbf{W} \quad \text{II.32}$$

De plus, tous les points \mathbf{X}_i qui ne sont pas des vecteurs supports c'est-à-dire ceux qui ne vérifient pas l'équation $[Y_i(\langle \mathbf{W}, \mathbf{X}_i \rangle + w_0) - 1] = 0$ sont donc associés à des α_i nuls (théorème de Karush-Kuhn-Tucker). Ainsi, on retrouve que l'hyperplan optimal ne dépend que de N_s vecteurs de support ($N_s \leq N$) :

$$\mathbf{W} = \sum_i^{N_s} \alpha_i Y_i \mathbf{X}_i \quad \text{II.33}$$

En substituant les équations II. 31 et II.32 dans l'équation II.30, le Lagrangien prend la forme d'une fonction à maximiser en fonction des seules variables duales α_i . Il s'écrit :

$$L(\alpha) = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i,j}^{N_s} \alpha_i \alpha_j Y_i Y_j \langle \mathbf{X}_i, \mathbf{X}_j \rangle \quad \text{II.34}$$

Le problème ne dépend plus de la dimension n de l'espace d'entrée, mais de la taille $N_s < N$.

L'hyperplan séparateur a donc pour solution :

$$\langle \mathbf{W}, \mathbf{X} \rangle + w_0$$

avec $\mathbf{W} = \sum_i^{N_s} \alpha_i^* Y_i \mathbf{X}_i$

et w_0 est obtenu à partir d'un des vecteur support \mathbf{X}_i par $w_0 = Y_i - \mathbf{W} \mathbf{X}_i$

Notons que dans la pratique, on peut également assouplir les contraintes du problème en introduisant de nouvelles variables ξ_i , appelées variables d'écart positives.

$$\forall i \quad \xi_i \geq 0 \quad \text{II.35}$$

$$\forall \mathbf{X}_i \in C_2, \langle \mathbf{W}, \mathbf{X}_i \rangle + w_0 \geq 1 - \xi_i \quad \text{II.36}$$

$$\forall \mathbf{X}_i \in C_1, \langle \mathbf{W}, \mathbf{X}_i \rangle + w_0 \leq 1 + \xi_i \quad \text{II.37}$$

D'après ces inéquations, si un exemple \mathbf{X}_i de la base d'apprentissage est mal classé alors il faut que la variable ξ_i correspondante soit supérieure à 1. Alors $1 - \xi_i$ ou $1 + \xi_i$ change de signe. Ainsi la somme $\sum \xi_i$ correspond à une borne supérieure d'erreurs de classification sur la base d'apprentissage. Dans ce cas la fonction objective de la formulation primale par exemple devient :

$$\tau(\mathbf{W}, \xi_i) = \frac{1}{2} \|\mathbf{W}\|^2 + C \left(\sum_{i=1}^{N_s} \xi_i \right) \quad \text{II.38}$$

où C est un paramètre positif qui permet de contrôler le compromis entre le fait de minimiser les erreurs de classification commises sur l'ensemble d'apprentissage et maximiser la marge. Le classifieur SVM est alors appelé SVM à marge souple, car les contraintes sur la marge sont assouplies pour certains exemples de la base d'apprentissage. On peut alors suivre le même principe de formulation duale pour résoudre ce problème. Le lecteur intéressé par les détails de calculs sur les SVM peut se référer aux références [92] [93] pour plus de détails.

Dans la formulation duale (II.34), la résolution du problème d'optimisation ne dépend que du produit scalaire entre les objets \mathbf{X} de la base d'apprentissage. Ce constat a motivé l'utilisation des noyaux qui confère aux SVM une extension non linéaire.

II.3.1.2 SVM non linéaire

En pratique, les données ne sont pas toujours linéairement séparables. Il faut donc doter les SVM d'un mécanisme permettant de produire des frontières de décision non linéaires. La solution consiste à transformer l'espace des données. L'idée est donc de projeter les points d'apprentissage de l'espace de départ dans un espace de plus grande dimension nommé espace de Hilbert et noté \mathcal{H} dans lequel les données peuvent être linéairement séparables (figure II.7).

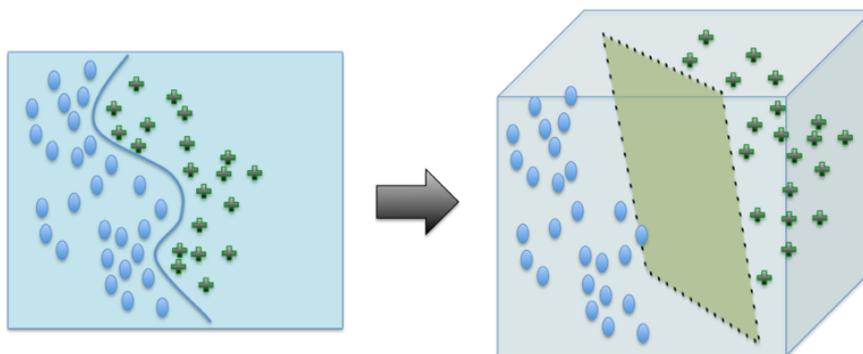


Figure II.7: Exemple schématique de transformation d'un problème de séparation non linéaire en 2 classes en une séparation linéaire dans un nouvel espace de dimension supérieure.

Pour cela, il suffit de définir une application non linéaire Φ dans l'algorithme SVM appliquée aux données \mathbf{X} dans l'espace \mathcal{H} :

$$\begin{aligned} \mathbb{R}^n &\xrightarrow{\phi} \mathcal{H} \\ \mathbf{X} &\rightarrow \Phi(\mathbf{X}) \end{aligned} \quad \text{II.39}$$

La transformation non linéaire $\Phi(\mathbf{X})$ permet donc de produire des surfaces de décision planes dans \mathcal{H} . Etant donné que les données apparaissent dans tous les calculs sous forme de produit scalaire $\langle \mathbf{X}_i, \mathbf{X}_j \rangle$ dans ce qui précède, il suffit d'utiliser une fonction noyau notée k symétrique définie positive satisfaisant les conditions de Mercer [94] [95] qui permet de calculer efficacement le produit $\langle \Phi(\mathbf{X}_i), \Phi(\mathbf{X}_j) \rangle$:

$$k(\mathbf{X}_i, \mathbf{X}_j) = \langle \Phi(\mathbf{X}_i), \Phi(\mathbf{X}_j) \rangle \quad \text{II.40}$$

Parmi les noyaux les plus utilisés dans la littérature on peut citer :

- Noyau polynomial d'ordre p [96] [97]:

$$k(\mathbf{X}_i, \mathbf{X}_j) = (\langle \mathbf{X}_i, \mathbf{X}_j \rangle + 1)^p \quad \text{II.41}$$

- Noyau gaussien de largeur de bande σ [98] [99]

$$k(\mathbf{X}_i, \mathbf{X}_j) = \exp\left(-\frac{\|\mathbf{X}_i - \mathbf{X}_j\|^2}{2\sigma^2}\right) \quad \text{II.42}$$

Dans ce cas, tout le développement présenté dans le paragraphe II.3.1.1 peut être appliqué dans le nouvel espace. Il suffit de remplacer le produit scalaire $\langle \mathbf{X}_i, \mathbf{X}_j \rangle$ par $k(\mathbf{X}_i, \mathbf{X}_j)$.

Cependant, le choix de noyaux est difficile et conditionne la qualité des résultats. La recherche des paramètres (W^*, w_0^*) est très coûteuse en temps de calcul.

Concernant les problèmes multi classes différents algorithmes ont été proposés ; On trouve notamment la méthode « un contre tous » [100] qui traite directement le cas multi classes ou celle « un contre un » [101] qui consiste à décomposer le problème en un ensemble de problèmes à deux classes.

Les SVM ont montré leur efficacité dans différents types de tâches comme le traitement d'images [102] ou encore le diagnostic médical [103]. Ils ont également été utilisés en scattérométrie pour l'identification des profils géométriques pour des cas simples [104]. Dans cette étude la méthode est combinée à la méthode des bibliothèques pour la résolution du problème inverse.

II.3.2 Méthode connexionniste : Les Réseaux de Neurones

II.3.2.1 Généralités sur les réseaux de neurones

En informatique, le connexionnisme est défini comme un ensemble de traitement d'information constitué de processeurs élémentaires simples à fortes connectivités organisés selon une certaine topologie. Dans notre cas, le processeur élémentaire est un neurone formel désigné ainsi, car son fonctionnement est inspiré de celui des cellules neuronales biologiques. Etant donné que la plupart des modèles actuels de connexionnisme reposent sur cette notion, nous nous limiterons uniquement à ce qu'on appelle communément les réseaux de neurones (RN).

Le calcul par RN repose sur la propagation d'information entre neurones. Si chacun d'entre eux ne peut effectuer que des calculs simples, leur interaction permet par contre d'effectuer un calcul global beaucoup plus complexe. Grâce à l'apprentissage, le RN est optimisé pour répondre à une tâche clairement identifiée soit par l'intermédiaire d'exemples fournis soit en exploitant des règles internes.

Dans ce paragraphe, nous abordons les RN en donnant dans un premier temps une présentation générale de ces outils mathématiques tout en définissant les termes essentiels à la compréhension de la suite du document. Ensuite, nous présentons en détail ceux utilisés dans le cadre de nos travaux de thèse. Nous invitons le lecteur désirant approfondir ce domaine à se référer à la thèse de Yann le Cun [105].

II.3.2.1.1 Historique des R.N

En 1943, Mac Culloch et Pitts [106] ont proposé un modèle de neurone formel mimant les neurones biologiques et capable de mémoriser des fonctions booléennes simples. Ce type de neurones est initialement conçu pour réaliser des fonctions arithmétiques et logiques simples. La connexion de ces entités entre elles permet de concevoir des RN. Le physiologiste Donald Hebb propose en 1949 dans son ouvrage « the organization of behaviour » une règle d'apprentissage connue aujourd'hui sous le nom de règle de Hebb [107] qui est basée sur le renforcement ou la dégénérescence des liaisons entre neurones. Ensuite, Frank Rosenblatt introduit en 1957 le perceptron [108]. Ce dernier constitue un modèle simplifié du système de vision. Il est considéré comme le premier modèle de neurones capable de reconnaître des formes. En 1960 Widrow et Hoff développent le modèle ADALINE (pour Adaptive Linear Neuron) qui est devenu par la suite le modèle de base des réseaux multicouches. Très vite, le perceptron est critiqué, Minsky et S. Paper [109] ont montré son incapacité à généraliser ses performances sur des problèmes non linéaires. En parallèle de ces travaux, T. Kohonen, propose le concept de carte auto organisée [110] ayant des propriétés d'auto organisation et d'auto associativité permettant ainsi de donner une représentation simplifiée d'un espace multidimensionnel. Il propose ainsi des applications à la reconnaissance de formes dans le cadre de classification non supervisée. J. Hopfield [111] [112] explore cet aspect en mettant au point un réseau dynamique qui évolue jusqu'à atteindre un état de stabilité par un comportement autonome. D'autres algorithmes d'apprentissage ont été développés au cours des années 80 dont les plus connus sont l'algorithme de retro propagation de l'erreur qui apporte alors une solution au problème soulevé par Minsky et S. Paper ou encore celui de Boltzmann rendant plus performants les réseaux multicouches. Ces travaux ont montré que les R.N sont capables de résoudre des problèmes d'optimisation complexes dans des domaines aussi différents que variés.

II.3.2.1.2 Applications des R.N

De façon générale, les RN sont fortement sollicités pour le traitement d'information qui devient de plus en plus complexe au fil des années. Ils sont ainsi utilisés dans tous les domaines de la science en particulier dans le domaine de la reconnaissance de formes. L'un des premiers modèles dans cette thématique a été développé aux laboratoires AT&T Bell [113] permettant la reconnaissance des codes postaux fournis par les services postaux américains. Ils ont également été utilisés pour la reconnaissance de caractères manuscrits [114]. On les retrouve

aussi dans des applications de reconnaissance et d'identification des ferromagnétiques marins [115] ou encore dans la reconnaissance de formes pour des systèmes de guidage automatisés [116]. Dans le cadre médical, nous pouvons citer en guise d'exemple la reconnaissance d'arythmie cardiaque [117]. Dans d'autres domaines, comme la sécurité des réseaux informatiques, ils ont permis de développer des applications pour détecter des menaces d'intrusions [118] [119]. Les R.N sont également indispensables dans la fouille de données (ou data mining) [120] où le nombre d'applications ne cesse d'augmenter ces dernières années. Ils sont également utilisés en marketing pour le traitement rapide d'informations [121]. En aéronautique, les R.N ont permis d'améliorer la trajectoire d'urgence d'un avion dans le cas d'une panne totale de moteur [122]. Mais ils trouvent également des applications dans des secteurs de la vie quotidienne, par exemple l'environnement, où ils sont utilisés pour le contrôle de systèmes de collecte des eaux pluviales [123]. La liste d'applications présentée ici est loin d'être exhaustive.

Nous rappelons que les RN sont également étudiés depuis plusieurs années au laboratoire et ont fait preuve de performances et de robustesse dans de nombreux travaux liés à la résolution de problèmes inverses notamment en métrologie optique [65] [17] [68]. Cette thèse s'inscrit donc dans la continuité des travaux précédemment réalisés tout en développant un potentiel des RN non encore utilisé en scattérométrie : la classification et l'identification d'objets.

II.3.2.1.3 Définition du neurone artificiel

Par analogie avec la biologie, le neurone artificiel reçoit l'information provenant de ses n entrées x_i ($i=1, \dots, n$) par l'intermédiaire des connexions dont on affecte à chacune d'elles un poids dit synaptique w (en anglais weights) pondérant l'information reçue. Ces dernières sont regroupées dans un vecteur \mathbf{X} . La valeur calculée par le neurone j à partir de \mathbf{X} est appelée sa "sortie" et noté s_j .

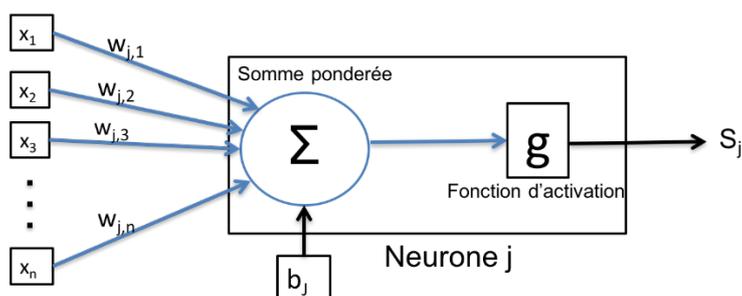


Figure II.8: Représentation d'un neurone mathématique particulier j . Les variables x_i désignent les n entrées et la sortie est notée s_j . Le biais b_j est une caractéristique propre du neurone. Le poids synaptique de chaque connexion reliant l'entrée i au neurone est désigné par $w_{j,i}$.

Le traitement de l'information par le neurone conditionne l'action réalisée par celui-ci. Ainsi, à titre d'exemple nous détaillons le comportement du neurone dit sommateur qui est l'un des plus utilisés (Figure II.8). Ce dernier fonctionne distinctement en deux étapes :

La première phase consiste en un prétraitement des données collectées par le calcul du potentiel v_j du neurone j :

$$v_j = b_j + \sum_{i=1}^n w_{j,i} x_i \quad \text{II.43}$$

où $w_{j,i}$ désigne le poids de la connexion liant le neurone j à l'entrée x_i et b_j est un terme propre au neurone appelé biais.

Notons que ce dernier peut également être considéré comme le poids d'une nouvelle entrée fixe notée x_0 égale à 1. Ainsi, on a plus simplement :

$$v_j = \sum_{i=0}^n w_{j,i} x_i \text{ avec } w_{j,0} = b_j \quad \text{II.44}$$

Dans la deuxième phase, une fonction de transfert g appelée également fonction d'activation calcule la valeur de sortie s_j du neurone j à partir de la valeur du potentiel v_j :

$$s_j = g(v_j) = g\left(\sum_{i=0}^n w_{j,i} x_i\right) \quad \text{II.45}$$

La nature de cette fonction d'activation est prépondérante suivant l'application visée. Les plus utilisées sont :

- La fonction seuil (figure II.9 (a)) :

$$g(x) = \begin{cases} 0 & \text{si } x < 0 \\ 1 & \text{si } x \geq 0 \end{cases} \quad \text{II.46}$$

- La fonction identité dite linéaire (figure II.9 (b)) :

$$g(x) = x, \forall x \quad \text{II.47}$$

Les neurones utilisant ce type de fonction sont appelés des automates linéaires.

- La fonction sigmoïde (figure II.9 (c)) dont les formes les plus utilisées sont la fonction logistique (Eq. II.48) et la fonction tangente hyperbolique (Eq. II.49) :

$$g(x) = \frac{1}{1+e^{-kx}} \quad \text{II.48}$$

$$g(x) = \frac{e^{kx} + 1}{e^{kx} - 1} = \tan\left(\frac{k}{2}x\right) \quad \text{II.49}$$

La valeur du paramètre k permet de modérer la pente à l'origine. Elle est généralement prise égale à 1.

- La **fonction gaussienne** d'écart type unité (figure II.9 (d)) :

$$g(x) = e^{-x^2} \quad \text{II.50}$$

Elle permet notamment de s'assurer que les sorties du neurone sont toutes positives.

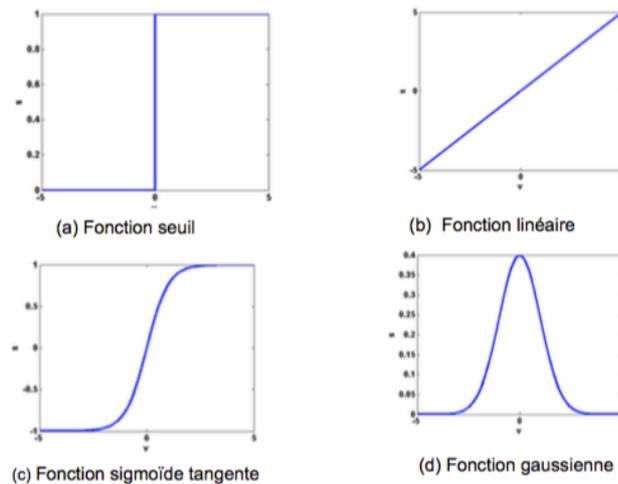


Figure II.9: Exemples de fonctions de transfert g : (a) seuil, (b) linéaire, (c) sigmoïde (logistique $k=1$) et (d) Gaussienne

Un neurone formel j peut donc être vu comme une fonction algébrique non linéaire et bornée des variables \mathbf{X} paramétrées par l'ensemble des poids synaptiques w_{ji} .

Notons que si on considère la classification supervisée, ce type de neurones sommateurs se comporte comme un séparateur linéaire (Eq. II.45). Ainsi, dans un RN sommateur, les interconnexions entre ces neurones vont permettre de combiner les discriminations linéaires pour obtenir une fonction discriminante beaucoup plus complexe.

II.3.2.1.4 La mise en réseau

Par définition, un RN est donc constitué de plusieurs neurones interconnectés entre eux. L'association de ces neurones en réseau permet la composition des fonctions réalisées par chacun d'eux. Les connexions entre les neurones qui composent le réseau décrivent ce qu'on appelle la topologie ou l'architecture du modèle. Il existe plusieurs possibilités selon lesquelles les neurones peuvent être connectés pour former un réseau. Nous citons parmi celles-ci les plus connues :

- **Architecture multicouche** : Les neurones constituant ce type de réseau sont arrangés par couches verticales. Il n'y a pas de connexion entre neurones d'une même couche. Chaque neurone est exclusivement relié aux neurones de la couche précédente (figure II.10) et suivante (s'il y en a). On peut ainsi définir un sens de parcours de l'information (de l'activation) au sein d'un réseau depuis les entrées vers les sorties. La couche d'entrée est constituée de l'ensemble des N entrées du réseau regroupées dans le vecteur X . La couche de sortie quant à elle est composée des neurones effectuant les calculs finaux. Ces derniers composent le vecteur S . Les couches internes effectuant des calculs intermédiaires sont appelées couches cachées. Ce type d'architecture est très utilisé pour des tâches de prédiction ou de classification et fonctionne généralement sur la base d'un apprentissage supervisé. Nous détaillerons son fonctionnement dans les paragraphes suivants. C'est également l'architecture la plus utilisée dans la littérature dans le cadre de la scattérométrie classique.

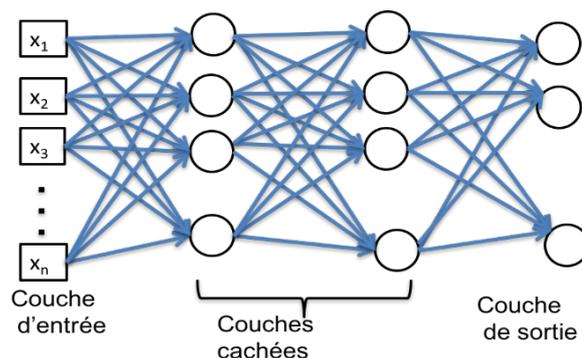


Figure II.10: Exemple d'architecture d'un réseau multicouche à 2 couches cachées

- **Architecture à connectivités locales** : Il s'agit d'une structure possédant une topologie à l'image de la rétine. Chaque neurone entretient des relations localisées avec un nombre fixe, mais réduit de neurones voisins (figure II.11). Les connexions sont donc moins nombreuses que dans le cas d'un réseau multicouche classique. L'intérêt d'utiliser des

connexions locales est lié à l'exploitation des possibilités de regroupement et de traitement localisé des données fournies en entrée. Cela dépend de la spécificité du problème à résoudre. Les cartes auto organisatrices de Kohonen que nous détaillerons dans la suite de nos travaux appartiennent à ce groupe. Ils sont principalement utilisés en traitement d'image, dans des applications de discrimination, dans des problématiques d'aide à la décision ou encore dans des applications nécessitant la résolution d'un problème d'optimisation.

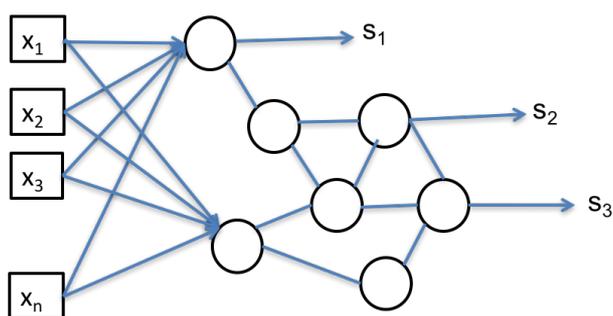


Figure II.11: Exemple d'architecture d'un réseau de neurones à connexions locales

- **Architecture à connectivités récurrentes** : les connexions récurrentes ramènent l'information en arrière par rapport au sens de propagation classiquement défini de gauche à droite. Ces connexions sont le plus souvent locales (figure II.12) et permettent de modéliser des comportements dynamiques de systèmes physiques. Grâce à ces connexions récurrentes, cette architecture possède des mémoires internes. Les R.N ainsi construits sont utilisés principalement dans la reconnaissance ou la reproduction séquentielle comme, par exemple, la reconnaissance d'un mot dans un message vocal.

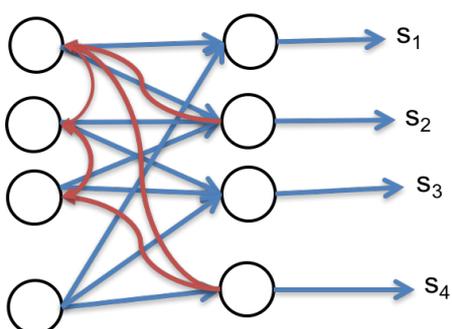


Figure II.12: Exemple d'architecture d'un réseau de neurones à connexions récurrentes. Les flèches bleues amènent l'information d'un neurone vers le suivant et les flèches rouges ramènent l'information vers le neurone précédent.

- **Architecture à connectivité complète (ou réseaux de Hopfield) :** C'est la structure d'interconnexion la plus générale (figure II.13). Chaque neurone est connecté à tous les neurones du réseau et également à lui-même. Ce type de réseau possède le nombre maximal de connexions pour un nombre fixé de neurones. Ce genre d'architecture est une mémoire adressable par son contenu. Les réseaux de ce type peuvent être notamment utilisés pour des tâches de mémorisation d'état.

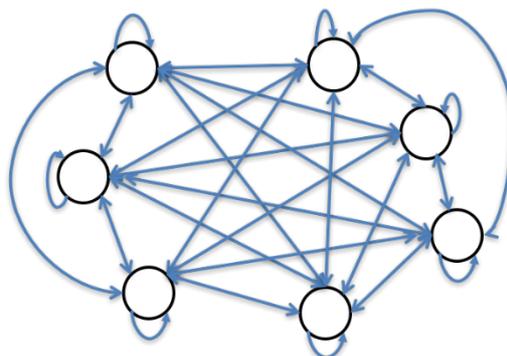


Figure II.13: Exemple d'architecture d'un réseau de Hopfield comprenant 7 neurones

II.3.2.1.5 Apprentissage des R.N

L'apprentissage est l'étape primordiale pour l'utilisation des RN. C'est une phase de développement durant laquelle le comportement du réseau est modifié et ajusté itérativement suivant la tâche qu'il doit réaliser à partir d'un certain nombre d'échantillons composant le corpus d'apprentissage. On ajoute ainsi à la structure topologique du réseau décrite précédemment un algorithme d'apprentissage. Les paramètres intrinsèques du réseau (à savoir les poids w_{ji} des connexions) sont modifiés itérativement pendant cette phase. La modification du poids w_{ij} à l'itération t peut alors s'écrire simplement par:

$$w_{j,i}(t + 1) = w_{j,i}(t) + \Delta w_{j,i} \quad \text{II.51}$$

En pratique, on distingue deux méthodes d'apprentissage selon la manière dont sont fournies les données du corpus d'apprentissage : la méthode adaptative selon laquelle les poids des neurones sont modifiés de manière séquentielle après chaque passage d'une entrée particulière \mathbf{X} et celle dite non adaptative (batch training en anglais) selon laquelle la mise à jour des poids est effective après le passage de l'ensemble des N données d'apprentissage. Ce dernier est beaucoup plus rapide et, de ce fait, plus adapté à une grande quantité de données d'entraînement.

Par ailleurs, comme cela a déjà été mentionné plus haut, l'apprentissage d'un R.N peut s'effectuer de façon supervisée sur la base d'exemples de comportements connus ou en suivant

des règles internes permettant d'optimiser un critère évalué sur l'ensemble des données d'apprentissage.

Dans le cas de l'apprentissage supervisé, le corpus d'entraînement est composé de N entrées \mathbf{X} pour lesquelles on connaît les sorties désirées \mathbf{Y} . Ainsi le vecteur \mathbf{X} est fourni en entrée du R.N et ce dernier doit estimer en sortie les valeurs désirées \mathbf{Y} connues par l'intermédiaire de son vecteur \mathbf{S} résultant d'une série de calcul interne (figure II.14). Ce type d'apprentissage est une métaphore du « professeur » qui donne des « exemples » et fournit une correction par le calcul de l'« erreur de prédiction » permettant au réseau de « corriger » itérativement la valeur de ses poids. Ainsi on cherche à minimiser l'erreur $\mathbf{E}=\mathbf{S}-\mathbf{Y}$. L'apprentissage supervisé des R.N est donc formulé comme un problème d'optimisation paramétrique dans lequel on cherche les valeurs optimales des poids des connexions du réseau pour un comportement précis fourni par les exemples du corpus d'apprentissage.

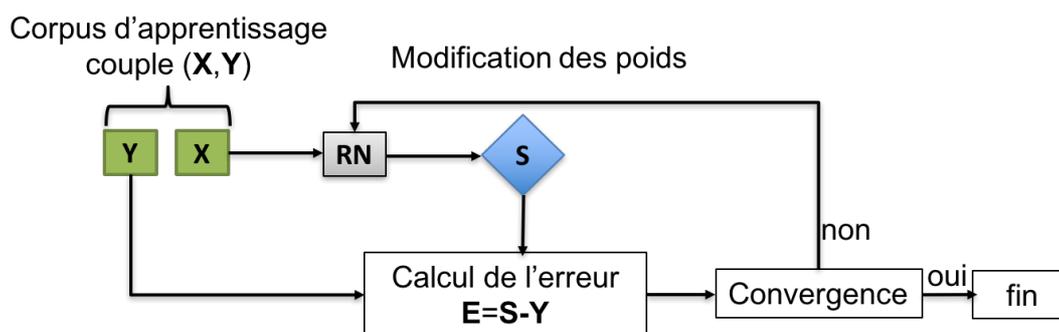


Figure II.14: Schéma explicatif de l'apprentissage supervisé

Parmi les réseaux qui fonctionnent avec un apprentissage supervisé, nous pouvons citer le perceptron multicouche et le Learning Vector Quantization que nous utilisons dans nos travaux.

Dans le cas de l'apprentissage non supervisé (ou apprentissage par compétition), les valeurs d'entrée \mathbf{X} composant généralement un ensemble de données de grande taille sont fournies au RN qui cherche à les regrouper selon des critères de ressemblance qui sont *a priori* inconnus. On laisse le RN s'auto-organiser à partir de lois locales qui régissent l'évolution itérative des poids synaptiques. Les réseaux de Kohonen et plus particulièrement les cartes auto organisatrices utilisent ce type d'apprentissage.

La correction des poids à chaque itération donnée par Δw_{ji} est guidée par un certain nombre de règles. Les plus connues dans la littérature sont la règle de Hebb [107] [124], la règle de Widrow Hoff [125], la règle Delta, la règle de Grossberg [126] ou encore la retro-propagation du gradient de l'erreur. Elles sont spécifiques à un type d'architecture ou à un type

d'apprentissage donné (apprentissage supervisé ou compétitif). Ainsi, nous ne détaillerons dans la suite que celles relatives aux structures mises en place dans nos travaux.

II.3.2.2 Le perceptron Multicouche

II.3.2.2.1 Présentation et architecture

Le PMC est un réseau à couches dont les fonctions d'activation d'une même couche sont identiques. Généralement les neurones des couches cachées sont à fonction de transfert sigmoïdale. Les neurones constituant la couche de sortie présentent une fonction d'activation linéaire dans le cas d'une utilisation en tant que prédicteur et non linéaire (par exemple sigmoïdale) dans le cas d'une application en tant que classifieur. La couche de sortie possède autant de neurones que de sorties désirées. Dans ce qui suit nous noterons par N_s le nombre de neurones de sortie et N_c celui des neurones de la couche cachée. Le poids reliant le neurone j de la couche cachée à l'entrée i est noté w_{ji} et celui relatif à la connexion du neurone m de la couche de sortie au neurone j de la couche cachée est nommé $z_{m,j}$.

A titre d'exemple, considérons le R.N de la figure II.15 qui est connu pour être un approximateur universel parcimonieux.

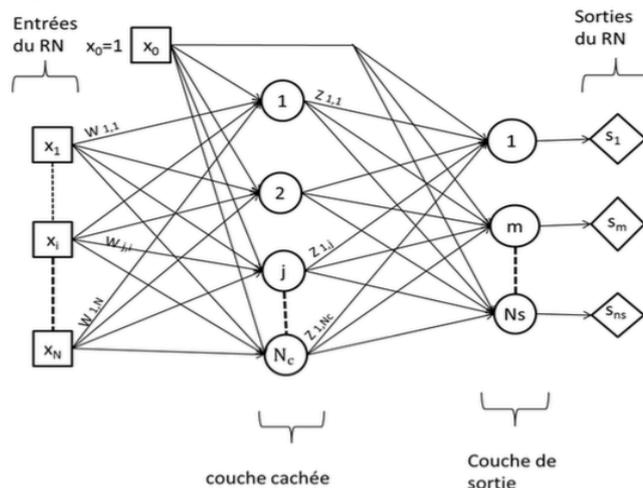


Figure II.15: Structure d'un PMC à n entrées et N_s sorties composé d'une couche cachée de N_c neurones à fonction d'activation sigmoïdale et d'une couche de sortie à fonction d'activation linéaire

Ce type de RN est déjà largement utilisé au laboratoire dans le cadre de la régression pour la résolution du problème inverse scattérométrique.

Si on appelle g la fonction d'activation des neurones de sortie et f celle des neurones cachés, la $m^{\text{ième}}$ sortie du PMC est définie par :

$$s_m(\mathbf{X}, \mathbf{W}, \mathbf{Z}) = g\left(\sum_{j=0}^{N_c} z_{m,j} \left[f\left(\sum_{i=0}^n w_{j,i} x_i\right) \right]\right) \quad \text{II.52}$$

La fonction de sortie g étant linéaire, l'expression peut se simplifier en :

$$s_m(\mathbf{X}, \mathbf{W}, \mathbf{Z}) = \sum_{j=0}^{N_c} z_{m,j} \left[f\left(\sum_{i=0}^n w_{j,i} x_i\right) \right] \quad \text{II.53}$$

Ainsi, la sortie s_m peut être définie comme une fonction algébrique de \mathbf{X} à valeurs bornées paramétrée linéairement par les poids \mathbf{Z} et non linéairement par les poids \mathbf{W} .

Ce type de RN possède des propriétés remarquables. En effet, la propriété d'approximateur universel peut être énoncée de la manière suivante : *toute fonction bornée suffisamment régulière peut être approchée avec une précision arbitraire, dans un domaine fini de l'espace de ses variables, par un RN comportant une couche de neurones cachés en nombre fini, possédant tous la même fonction d'activation, et un neurone de sortie linéaire* [127]. La propriété d'approximation n'est pas seulement spécifique aux RN. Il existe également d'autres de fonctions possédant cette propriété (ondelettes, fonctions splines,...).

La spécificité des RN réside dans le caractère parcimonieux de l'approximation : *à précision égale, les RN nécessitent moins de paramètres ajustables (les poids des connexions) que les approximateurs universels linéaires couramment utilisés. Plus précisément, le nombre de poids varie linéairement avec le nombre de variables de la fonction à approcher, alors qu'il varie exponentiellement pour la plupart des autres approximateurs* [128].

En général, on cherche à modéliser un processus physique à partir de mesures expérimentales. Dans ce cas, les données disponibles sont présentes sous la forme d'un grand nombre N de couples (\mathbf{X}, \mathbf{Y}) où \mathbf{Y} sont les observations mesurables liées aux caractéristiques physiques \mathbf{X} . Le but du RN est de trouver la relation liant \mathbf{Y} à \mathbf{X} à partir de ces données dites d'apprentissage afin d'être capable de généraliser cette relation sur de nouvelles entrées \mathbf{X} . Cette dernière sera bien entendu valable uniquement dans les conditions de mesure utilisées et dans une gamme de variation donnée pour les entrées \mathbf{X} :

$$\mathbf{X} \mapsto \mathbf{S}(\mathbf{X}, \mathbf{W}, \mathbf{Z}) \text{ qui est une estimation de } \mathbf{Y} \quad \text{II.54}$$

La plupart du temps, il faut prendre en compte le fait que les observations \mathbf{Y} sont entachées de bruit de mesure. Ainsi, il est commun de formuler cette réponse du PMC par une approche statistique. Si on suppose que les mesures \mathbf{Y} correspondant aux caractéristiques \mathbf{X} sont des réalisations de variables aléatoires de densité de probabilité gaussienne d'écart type σ ,

on peut considerer la sortie \mathbf{S} du RN comme :

$$\mathbf{S}(\mathbf{X}, \mathbf{W}, \mathbf{Z}) = \mathbf{E}(\mathbf{Y}/\mathbf{X}) \quad \text{II.55}$$

Notons que dans le cas particulier de la scattérométrie, les RN sont utilisés pour résoudre un problème inverse. Ainsi, le vecteur \mathbf{X} représente la signature ellipsométrique et le vecteur \mathbf{Y} caractérise la forme géométrique de la structure correspondante. Le R.N est donc utilisé pour modéliser la relation inverse liant \mathbf{X} à \mathbf{Y} .

II.3.2.2 Apprentissage du PMC

L'apprentissage du PMC est un apprentissage de type supervisé. Il se base donc sur un algorithme de correction de l'erreur (paragraphe II.3.2.1.5). La fonction coût que l'on cherche à minimiser est généralement mise sous la forme d'une erreur quadratique moyenne notée :

$$C(\mathbf{W}, \mathbf{Z}) = \sum_{q=1}^N \left[\sum_{m=1}^{N_s} (y_m^q - s_m^q(\mathbf{X}, \mathbf{W}, \mathbf{Z}))^2 \right] \quad \text{II.56}$$

où s_m est la $m^{\text{ième}}$ sortie du PMC, y_m la $m^{\text{ième}}$ valeur du vecteur \mathbf{Y} du corpus d'apprentissage correspondant à l'entrée \mathbf{X} et N le nombre de total de couples d'apprentissage

Cette erreur peut se décomposer en somme de coûts partiels C^q correspondant chacun à l'erreur quadratique sur chaque entrée particulière \mathbf{X}^q . On a donc :

$$C(\mathbf{W}, \mathbf{Z}) = \sum_{q=1}^N C^q(\mathbf{W}, \mathbf{Z}) \text{ avec } C^q = \sum_{m=1}^{N_s} (y_m^q - s_m^q)^2 \quad \text{II.57}$$

Notons par u_j et v_m les potentiels respectifs des neurones cachés j et de sortie m possédant chacun les fonctions d'activation f et g . On a donc :

$$u_j^q = \sum_{i=0}^N w_{j,i} x_i^q \text{ et } v_m^q = \sum_{j=0}^{N_c} z_{m,j} f(u_j^q) \quad \text{II.58}$$

En se référant à l'équation II.47 la sortie du neurone m vaut $s_m^q = v_m^q$ dans le cas où g est linéaire.

Comme mentionné précédemment, on cherche le jeu de paramètres \mathbf{W} et \mathbf{Z} qui permettent de minimiser la fonction coût $C(\mathbf{W}, \mathbf{Z})$ afin d'obtenir un comportement optimal du RN. Ce problème peut donc être résolu en utilisant une des méthodes classiques d'optimisation paramétrique classiques présentées au chapitre I. Ces dernières nécessitent la connaissance de la valeur des dérivées premières et éventuellement secondes de la fonction coût par rapport à

chacun des paramètres intrinsèques du RN. Si cela ne pose pas de problème pour les paramètres \mathbf{Z} des neurones de sorties, cela n'est pas le cas pour les poids \mathbf{W} des neurones cachés. En effet, la sortie désirée de ces derniers n'est pas connue pour un couple d'observations donné. Pour contourner cette difficulté, l'algorithme le plus connu est celui de la retro-propagation du gradient de l'erreur [129] qui permet d'estimer de manière récursive chaque dérivée partielle $\frac{\partial C}{\partial w_{j,i}}$.

De façon détaillée, l'influence des poids $z_{m,j}$ sur l'erreur quadratique C peut être exprimée par :

$$\left(\frac{\partial C^q}{\partial z_{m,j}}\right) = \left(\frac{\partial C^q}{\partial v_m^q}\right) \left(\frac{\partial v_m^q}{\partial z_{m,j}}\right) = \delta_m^q f(u_j^q) \quad \text{II.59}$$

avec

$$\delta_m^q = \left(\frac{\partial C^q}{\partial v_m^q}\right) = -2(y_m^q - s_m^q) \quad \text{II.60}$$

on a donc :

$$\left(\frac{\partial C^q}{\partial z_{m,j}}\right) = -2f(u_j^q) (y_m^q - s_m^q) = \delta_m^q f(u_j^q) \quad \text{II.61}$$

Ensuite, l'influence des poids $w_{j,i}$ sur l'erreur quadratique C peut s'exprimer comme suit :

$$\left(\frac{\partial C^q}{\partial w_{j,i}}\right) = \left(\frac{\partial C^q}{\partial u_j^q}\right) \left(\frac{\partial u_j^q}{\partial w_{j,i}}\right) = \left[\sum_{m=1}^{N_s} \left(\frac{\partial C^q}{\partial v_m^q}\right) \left(\frac{\partial v_m^q}{\partial u_j^q}\right) \right]_{x_i^q} = \left[\sum_{m=1}^{N_s} \delta_m^q \left(\frac{\partial v_m^q}{\partial u_j^q}\right) \right]_{x_i^q} \quad \text{II.62}$$

or

$$\left(\frac{\partial v_m^q}{\partial u_j^q}\right) = z_{m,j} \left(\frac{\partial f}{\partial u_j^q}\right) \quad \text{II.63}$$

on a donc :

$$\left(\frac{\partial C^q}{\partial w_{j,i}}\right) = \left[\sum_{m=1}^{N_s} \delta_m^q z_{m,j} \left(\frac{\partial f}{\partial u_j^q}\right) \right]_{x_i^q} = x_i^q \left(\frac{\partial f}{\partial u_j^q}\right) \sum_{m=1}^{N_s} \delta_m^q z_{m,j} \quad \text{II.64}$$

Ces dérivées permettent de mettre en œuvre efficacement une méthode d'optimisation classique pour la phase d'entraînement. La méthode de Levenberg Marquardt [52] est celle privilégiée par la suite pour les avantages présentés au chapitre I.

Notons que si le réseau est bien entraîné, il doit être capable de répondre convenablement à l'introduction d'une nouvelle entrée inconnue. Cette aptitude du réseau est

appelée généralisation. L'optimisation décrite précédemment appliquée seule n'est pas suffisante pour garantir un fonctionnement optimal. En effet, la méthode peut conduire à un sur-apprentissage des données utilisées. Plusieurs méthodes existent pour améliorer la phase d'apprentissage du RN pour aboutir à une généralisation optimale. Parmi ces méthodes, nous pouvons citer celle qui supprime les modèles dont la matrice Jacobienne n'est pas de rang plan [130] (matrice dont toutes les lignes et les colonnes sont linéairement indépendantes) ou encore la méthode LOCL (Local Overfitting control via Laverages) [131] qui utilise l'influence de chaque exemple sur le modèle. D'autres méthodes comme la validation croisée [132] estiment les performances de plusieurs RN entraînés sur différents corpus d'apprentissage.

La méthode que nous utilisons par la suite est une méthode assez populaire : celle de l'arrêt prématuré (early stopping en anglais). Comme son nom l'indique, cette technique consiste à arrêter l'apprentissage prématurément sans attendre la convergence vers le minimum absolu de la fonction coût. Le principe consiste à subdiviser l'ensemble des données disponibles en 3 lots distincts :

- **Lot d'entraînement** : Ce corpus sert à l'apprentissage proprement dit du RN. L'erreur calculée sur ce corpus notée E_{entr} (à partir de l'équation II.56) est directement utilisée pour la correction des valeurs des poids \mathbf{W} et \mathbf{Z} des neurones suivant les règles établies par la méthode d'optimisation.
- **Lot de validation** : Ce corpus sert à valider le bon fonctionnement du réseau en termes de généralisation. Il contrôle l'arrêt de l'apprentissage. En effet, l'erreur quadratique sur cet ensemble notée E_{val} est calculée à chaque itération. Au début de l'apprentissage, elle décroît itérativement de la même manière que E_{entr} . Puis, dès que celle-ci commence à croître, la performance en généralisation se détériore et l'apprentissage est stoppé (figure II.16).
- **Lot de test** : Il sert à tester les performances sans biais du RN notées E_{rms} . Il faut noter que ce lot ne sert pas au processus d'apprentissage proprement dit et n'est présenté au réseau qu'une fois l'entraînement terminé. Il permet donc de vérifier si la généralisation est correcte sur des données non employées lors de la phase d'entraînement.

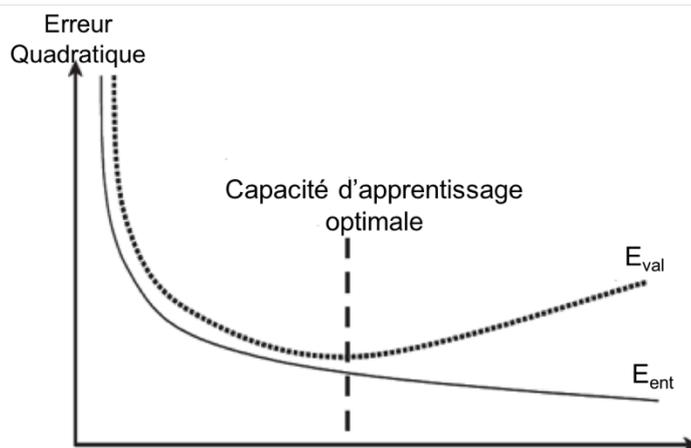


Figure II.16: Evolution schématique des différentes performances en termes d'erreurs quadratiques calculées sur les lots d'entraînement (E_{ent}) et de validation (E_{val}) lors de l'apprentissage d'un RN par la méthode d'arrêt prématuré

Dans nos travaux, la méthode du « early stopping » est mise en place comme suit : 70 % des données disponibles sont retenues pour l'apprentissage, 15 % pour la validation et les 15 % restant pour le test.

La qualité de l'apprentissage du PMC est généralement évaluée en traçant les valeurs du vecteur \mathbf{S} estimées par le RN en fonction de celles désirées \mathbf{Y} sur tous les échantillons composant un corpus prédéfini (généralement le lot de test). Une régression linéaire est ensuite réalisée sur ces données (normalisées ou non). Plus le coefficient de régression R ($0 \leq R \leq 1$) sera proche de 1 plus l'apprentissage sera identifié comme correct. A l'inverse, un coefficient R faible impliquera un apprentissage défaillant.

Les performances globales du R.N sont ensuite estimées sur le lot de test en calculant pour N_{test} échantillons :

$$E_{rms} = \sqrt{\frac{1}{N_{test}} \sum_{q=1}^{N_{test}} \left[\sum_{m=1}^{N_s} (y_m^q - s_m^q(\mathbf{X}, \mathbf{W}, \mathbf{Z}))^2 \right]} \quad \text{II.65}$$

Dans le cas de la scattérométrie, elle s'exprime en nanomètre et représente l'erreur globale réalisée par le RN moyennée sur l'ensemble des sorties et des N_{test} échantillons. On peut également définir une performance propre à chaque sortie m du PMC par :

$$E_{rms}^m = \sqrt{\frac{1}{N_{test}} \sum_{q=1}^{N_{test}} [(y_m^q - s_m^q(\mathbf{X}, \mathbf{W}, \mathbf{Z}))^2]} \quad \text{II.66}$$

II.3.2.2.3 Le PMC dans le cadre de la classification supervisée

Dans le domaine de la classification supervisée, les R.N sont fréquemment utilisés. Leur propriété de classifieur universel dérive de celle d'approximateur universel énoncée dans le paragraphe précédent. De ce fait, l'architecture conserve donc les mêmes caractéristiques que dans le cas de la régression non linéaire. Ainsi, le caractère non linéaire de la fonction d'activation sigmoïdale des neurones cachés, permet ainsi au PMC de générer des fonctions discriminantes non linéaires. La seule modification notable concerne les neurones de sortie et plus particulièrement leur fonction d'activation. En effet, cette dernière est généralement sigmoïdale ou à seuil et non plus linéaire.

Dans nos travaux, nous utilisons un PMC constitué de neurones à fonctions d'activation exclusivement sigmoïdales (couche cachée et couche de sortie) (figure II.17).

D'un point de vue pratique, il est nécessaire d'introduire un codage pour représenter les différentes classes. Ainsi, dans un problème de classification à K classes, on fait correspondre à chaque entrée un vecteur $\mathbf{Y} = (y_1, y_2, \dots, y_K)$ pour lequel les composantes y_k ne peuvent prendre que 2 valeurs a ou b. Généralement on choisit 1 et 0. Ainsi le vecteur \mathbf{Y} représentant la classe C_k , possède toute ses composantes égales à 0 à l'exception de la i-ième composante qui prend la valeur 1.

Prenons le cas d'une classification en 4 classes dans un espace des observations de dimension 3. Le vecteur d'entrée \mathbf{X}_1 composé de 3 éléments et assigné à la classe 3 sera codé par \mathbf{Y}_1 tel que :

$$\left\{ \mathbf{X}_1 = \begin{bmatrix} 2 \\ -1 \\ 0 \end{bmatrix}, \mathbf{Y}_1 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} \right\} \quad \text{II.67}$$

Ainsi, le PMC qui devra réaliser la tâche de classification doit être composé de 4 neurones de sortie. Le nombre de classes détermine donc directement le nombre N_s de neurones de sortie.

Ainsi, la $k^{\text{ième}}$ sortie s_k peut s'exprimer comme:

$$s_k = E(\mathbf{Y}|\mathbf{X}) = a \times p(C_k|\mathbf{X}) + b(1 - p(C_k|\mathbf{X})) \quad \text{II.68}$$

Dans le cas du codage particulier où $a=1$ et $b=0$, la sortie s_k du PMC est donc une approximation de la probabilité *a posteriori* pour X d'appartenir à la classe C_k :

$$s_k = p(C_k|\mathbf{X})$$

II.69

Les sorties nécessairement positives et bornées sur $[0, 1]$ représentent donc les probabilités d'appartenance aux différentes classes pour une entrée \mathbf{X} donnée. Dans ce cas, le PMC permet ainsi d'estimer la fonction discriminante directement sans passer par la formulation de Bayes définie dans le paragraphe II.2.

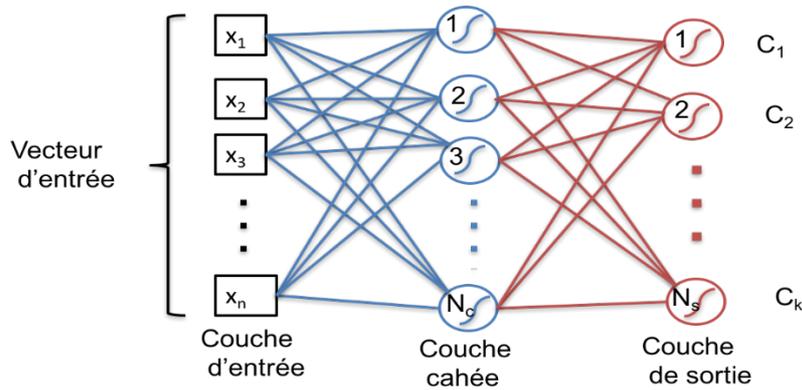


Figure II.17: Structure d'un PMC utilisée dans le cadre de la classification supervisée à n entrées composé d'une couche cachée à fonction d'activation sigmoïdale et d'une couche de sortie de N_s neurones présentant la même fonction d'activation. Le nombre N_s de neurones de sortie correspond au nombre de classes différentes.

L'apprentissage s'effectue de la même manière que précédemment (cf paragraphe II.3.2.2.2). On minimise la fonction coût des moindres carrés donnée par l'équation II.57. Ainsi les poids \mathbf{W} et \mathbf{Z} des neurones composant le PMC sont estimés à partir d'un ensemble d'apprentissage formé de couples $\{\mathbf{X}, \mathbf{Y}\}$ connu où chaque observation \mathbf{X}_i est associée au vecteur \mathbf{Y}_i qui code la classe correspondante.

L'erreur commise par le PMC en cours d'apprentissage peut donc facilement être estimée sur les exemples connus et exploitée par la méthode d'optimisation utilisée dans le cadre d'un entraînement supervisé (cf paragraphe II.3.2.1.5). Cette fois-ci, comme la fonction d'activation g du neurone m de sortie n'est pas plus linéaire la quantité δ_m^q vaut :

$$\delta_m^q = \left(\frac{\partial C^q}{\partial v_m^q} \right) = -2(y_m^q - s_m^q) \left(\frac{\partial g}{\partial v_m^q} \right) \quad \text{II.70}$$

Ainsi l'influence des paramètres $z_{m,j}$ et $w_{j,i}$ sur l'erreur quadratique est donnée par :

$$\left(\frac{\partial C^q}{\partial z_{m,j}} \right) = -2f(u_j^q) (y_m^q - s_m^q) \frac{\partial g}{\partial v_m^q} \quad \text{II.71}$$

et

$$\text{II.72}$$

$$\left(\frac{\partial C^q}{\partial w_{j,i}}\right) = x_i^q \left(\frac{\partial f}{\partial u_j^q}\right) \sum_{i=1}^{N_s} \left[-2z_{m,j} (y_m^q - s_m^q) \frac{\partial g}{\partial v_m^q} \right]$$

Le PMC ainsi décrit a largement fait ses preuves pour classer efficacement des objets selon différents critères connus à l'avance. Nous avons pu montrer que son architecture et son comportement sont similaires à celui utilisé dans le cadre de la régression non linéaire. Il sera donc d'autant plus facile de l'adapter pour mettre à profit ses capacités de classifieur dans le cadre de la scattérométrie.

II.3.2.3 Réseau de Kohonen

L'appellation de réseaux de Kohonen regroupe des réseaux fonctionnant avec un apprentissage compétitif. Ces types de R.N sont généralement constitués de neurones « distances » (figure II.18).

Le potentiel d'activation d'un neurone j est lié à la distance euclidienne entre le vecteur X de ses entrées et le poids W de ses connexions :

$$v_j = \sum_{i=0}^N (w_{j,i} - x_i)^2 \quad \text{II.73}$$

Pour ce type de neurone, la fonction d'activation est généralement de type gaussienne (cf paragraphe II.3.2.1.3).

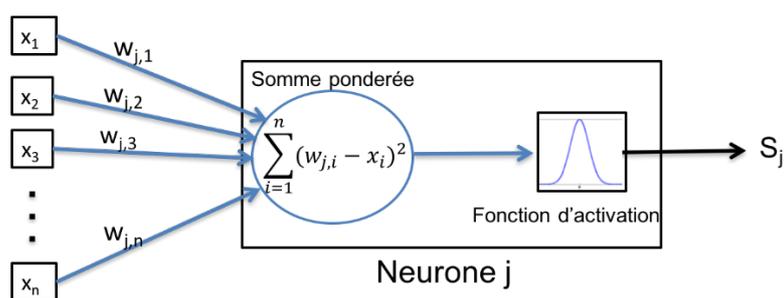


Figure II.18: Représentation d'un neurone distance j utilisé dans un réseau de Kohonen. Les variables x_i désignent les n entrées et la sortie est notée s_j . Le poids synaptique de chaque connexion reliant l'entrée particulière i au neurone est désigné par $w_{j,i}$.

Dans ce qui suit, nous traiterons tout particulièrement le cas des cartes auto organisatrices et des réseaux nommés Learning Vector Quantization (LVQ) que nous avons utilisés dans nos travaux.

II.3.2.3.1 Les cartes auto organisatrices

Introduites dans les années 80 par le statisticien Teuvo Kohonen [133], les cartes auto organisatrices (ou SOM pour Self Organizing Map) sont des réseaux de neurones artificiels fondés sur un apprentissage non-supervisé. Ce sont des structures intelligentes développées pour réaliser des tâches d'analyse de données ou de clustering. Elles sont inspirées de fonctions cérébrales biologiques faisant correspondre au monde extérieur multidimensionnel des représentations internes réduites à 1, 2 voire 3 dimensions. En effet, il a été montré qu'il existe des zones spécifiques dans le cerveau correspondant à la parole, au toucher, à la vision, à l'audition, etc... Par exemple, la région corticale assurant le traitement des informations auditives est composée de zones spécifiques organisées suivant les différentes caractéristiques fréquentielles des sons captés. Les zones géométriquement voisines possèdent ainsi des propriétés cognitives voisines. Il existe donc une organisation interne qui évolue au fil du temps sous l'effet de l'apprentissage. C'est ce comportement qui est à l'origine de la conception des SOM.

II.3.2.3.1.1 Architecture d'une SOM et principe

Une SOM fait correspondre par projection non linéaire à un espace d'entrée généralement de grande dimension une visualisation 1D (2D ou 3D) définie par un ensemble de neurones. Chacun d'eux se spécialise au cours de l'apprentissage pour représenter un type d'entrée donné.

Ainsi, une SOM est un R.N constitué d'une couche de S neurones « distances », répartis linéairement pour une représentation sur une seule dimension, répartis sur une carte de $S=mxn$ neurones pour une représentation en 2D ou bien répartis dans le volume d'un parallélépipède rectangle contenant $S=mxnxq$ neurones pour une représentation en 3D.

L'arrangement des neurones définit ce qu'on appelle la topologie de la carte. Si la carte est unidimensionnelle, chaque neurone possède seulement 2 voisins immédiats. Pour une carte bidimensionnelle, les neurones sont arrangés suivant une méthode particulière. La plus simple est la configuration rectangulaire (figure II.19 a) où chaque neurone possède 8 voisins immédiats. La plus utilisée est la topologie hexagonale où chaque neurone compte 6 voisins immédiats (figure II.19 b) situés à égale distance de ce dernier ; ce qui n'est pas le cas de la

topologie rectangulaire. Cette notion de voisinage est parfois représentée sous la forme d'une connectivité latérale.

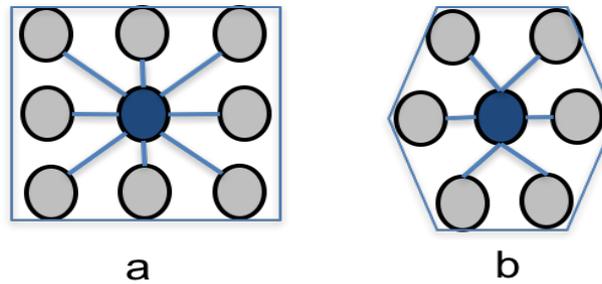


Figure II.19: Exemples de topologie de carte bidimensionnelle : (a) rectangulaire et (b) hexagonale

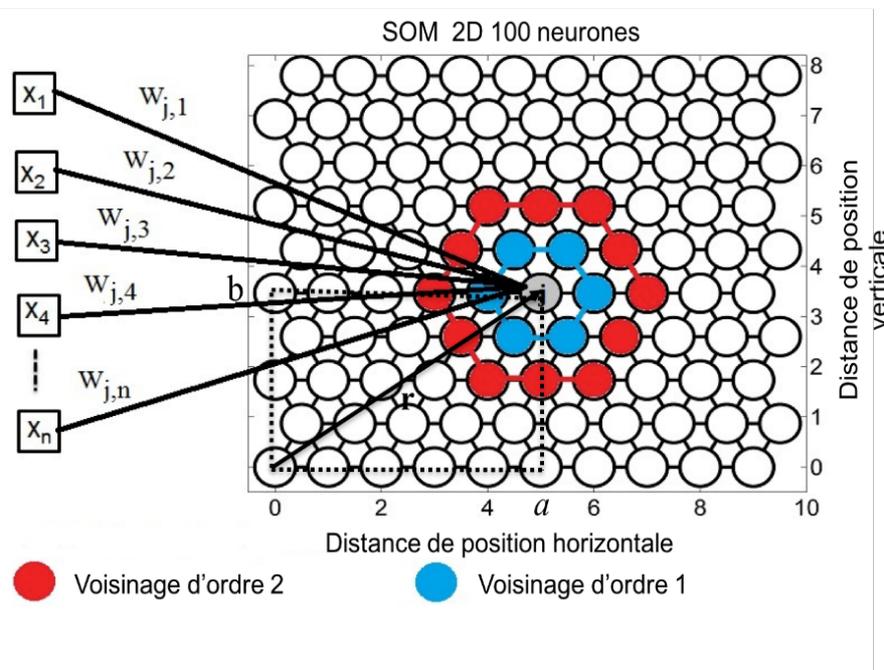


Figure II.20: Architecture d'une SOM 2D composée de n entrées et $S=10*10$ neurones de sortie réparties suivant une topologie hexagonale. Les neurones bleus appartiennent au voisinage d'ordre 1 du neurone j notée $V_j(1)$ et ceux en rouge au voisinage $V_j(2)$

L'entrée du RN est constituée du vecteur \mathbf{X} de n composantes (x_1, x_2, \dots, x_n) et correspondant à l'observation à classer. Chaque composante x_i du signal d'entrée est connectée à l'ensemble des neurones de la carte qui sont considérés comme des neurones de sortie. Ainsi, à chaque neurone j de la carte est associé un vecteur référent appelé aussi vecteur prototype $\mathbf{W}_j = (w_{j,1}, w_{j,2}, \dots, w_{j,n})$ composé des poids $w_{j,i}$ des connexions le reliant à toutes les entrées. L'activation d'un neurone de la carte dépend non seulement du vecteur d'entrée, mais également de son voisinage. Cette constatation nécessite la définition d'une distance entre deux neurones. Chaque neurone est identifié par sa position sur la carte. Par exemple l'emplacement d'un neurone j est désigné par un vecteur position \mathbf{r}_j de coordonnées a et b (figure II.20) tel que

$$\mathbf{r}_j = \begin{pmatrix} a \\ b \end{pmatrix}$$

où a désigne l'abscisse et b l'ordonnée dans le repère défini dans l'espace de la carte.

Pour toute paire de neurones i, j de la carte C , on définit donc une distance discrète entre deux neurones i et j notée δ_{ij} comme étant le plus court chemin sur la carte entre ces deux neurones (figure II.20). Ainsi le voisinage d'ordre p du neurone j par rapport à un neurone i noté $V_j(p)$ est défini comme l'ensemble des neurones de la carte située à une distance δ de ce dernier inférieure à p (figure II.20).

$$V_j(p) = \{r \in C, \delta_{ij} \leq p\} \quad \text{II.74}$$

II.3.2.3.1.2 Apprentissage d'une SOM

Lors de l'apprentissage, les données d'entrée sont soumises une à une à la carte auto organisatrice. Chaque neurone de sortie entre en compétition avec ses voisins respectifs pour représenter le vecteur d'entrée dans l'espace défini par la carte. Celui qui possède le vecteur référent le plus proche de \mathbf{X} est alors appelé neurone gagnant et noté G . On parle de la méthode « winner takes all ».

$$G = \arg \min_j \|\mathbf{X} - \mathbf{W}_j\|^2 \text{ pour } j=1, \dots, S \quad \text{II.75}$$

Le neurone G est l'unité la plus active de la carte. Sa réponse est fixée à 1 tandis que les sorties des autres neurones sont nulles.

Une fois le neurone gagnant déterminé. Les poids \mathbf{W}_G associé à ce dernier ainsi que ceux des autres neurones de la carte \mathbf{W}_j sont simultanément mis à jour selon la règle suivante :

$$\mathbf{W}_j(t+1) = \mathbf{W}_j(t) + \alpha(t)h(\delta_{Gj})[\mathbf{X} - \mathbf{W}_j(t)] \text{ pour } j=1, \dots, S \quad \text{II.76}$$

où \mathbf{X} désigne l'entrée appartenant au corpus d'apprentissage

$\alpha(t)$ désigne le taux d'apprentissage

et $h(\delta_{Gj})$ est une fonction de voisinage quantifiant l'importance de la modification à apporter au neurone j selon sa distance au neurone gagnant G . Pour le neurone G la fonction h retourne la valeur maximale.

Généralement, on utilise des fonctions noyaux symétriques définies positives telles que : $\lim_{\delta \rightarrow \infty} h(\delta) = 0$. La fonction la plus utilisée est de type : gaussienne (figure II.21):

$$h(\delta_{Gj}) = \exp\left(-\frac{\delta_{Gj}^2}{\sigma_t^2}\right) \quad \text{II.77}$$

où σ_t représente un paramètre constant dans la plupart des cas qui permet de moduler l'influence d'un neurone sur ses voisins.

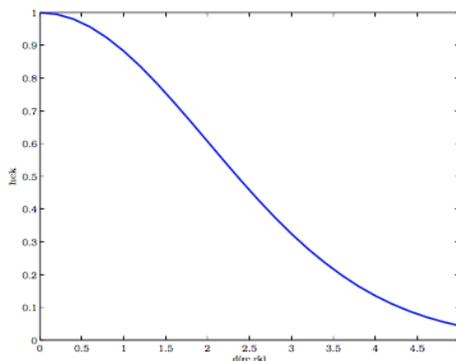


Figure II.21: Exemple de fonction de voisinage gaussienne qui quantifie l'importance de la modification des poids d'un neurone en fonction de la distance de ce dernier au neurone gagnant G

Le taux d'apprentissage $\alpha(t)$ est généralement une fonction monotone décroissante comprise dans l'intervalle $[0,1]$. Le plus souvent, on utilise une décroissance linéaire du type :

$$\alpha(t) = \alpha_0 \left(1 - \frac{t}{T}\right) \quad \text{II.78}$$

où α_0 est la valeur initiale et T le nombre maximal d'itération.

De la même manière, on peut également réduire le paramètre σ_t de la gaussienne au cours de l'apprentissage en utilisant la relation :

$$\sigma_t = \sigma_0 \exp\left(-\frac{t}{T}\right) \quad \text{II.79}$$

Le choix des valeurs initiales α_0 et σ_0 peut influencer les résultats d'une SOM. Ainsi, l'apprentissage se fait en deux phases notamment pour les cartes à grande dimension : une phase d'organisation dans laquelle α et σ sont initialisés avec de grandes valeurs et décroissent rapidement dans le but d'organiser les vecteurs prototypes \mathbf{W} . La deuxième phase de convergence a pour but d'ajuster les vecteurs prototypes vers leur position finale en partant des résultats de la première phase. Les paramètres α et σ sont alors initialisés avec de petites valeurs.

Une version plus rapide de l'apprentissage des SOM est l'apprentissage par lot (ou batch training). Cette version ne nécessite pas l'utilisation du taux d'apprentissage α . Tout le lot des données d'entrées est présenté à la fois au RN. La mise à jour se fait ensuite en utilisant la moyenne des vecteurs d'entrée pondérés par la fonction de voisinage.

$$\mathbf{W}_j = \frac{\sum_{i=1}^N h(\delta_{G(x_i)j}) \mathbf{X}_i}{\sum_{i=1}^N h(\delta_{G(x_i)j})} \quad \text{II.80}$$

On peut donc dire que l'algorithme d'apprentissage des cartes auto organisatrices optimise la fonction coût suivante [134] [135] :

$$C(W) = \frac{1}{2N} \sum_i^N \sum_j^S h(\delta_{G(x_i)j}) \|\mathbf{X}_i - \mathbf{W}_j\|^2 \quad \text{II.81}$$

où $G(x_i)$ est la fonction d'affectation du neurone gagnant G à une entrée particulière \mathbf{X}_i .

La procédure d'apprentissage des SOM conduit à une représentation bien ordonnée des entrées dans l'espace de sortie décrit par les neurones de la SOM. La carte est ainsi ordonnée topologiquement dans le sens où un neurone dans l'espace des sorties (composant la carte) correspond à une région particulière de l'espace des entrées. Ainsi, 2 neurones proches sur la carte représentent 2 régions adjacentes dans l'espace des entrées. Une SOM reflète également les variations statistiques des distributions des entrées. Ainsi, les régions de l'espace des entrées contenant beaucoup de stimuli seront affiliées à davantage de neurones que les régions plus éparées.

Hormis cette notion d'ordre topologique, la carte auto-organisatrice de Kohonen se comporte comme un quantificateur vectoriel. Ainsi, elle effectue simplement une compression des données d'entrée dans un espace codé de dimension réduite. Ce quantificateur n'est pas directement utilisable pour une tâche de classification. Le processus d'apprentissage n'étant pas supervisé. Pour cela, il convient donc de déterminer les neurones les plus sensibles à chaque stimulus composant une classe connue. On peut alors procéder à une opération d'étiquetage des neurones : l'ensemble des vecteurs caractéristiques de chaque classe est à nouveau présenté à la carte déjà entraînée et le nombre de fois où chaque neurone est activé est comptabilisé. Ainsi, chaque neurone se voit associé, suivant un seuil d'affectation préalable, à la classe pour laquelle il est majoritairement activé. Il arrive que certains neurones ne soient jamais activés lors de la phase d'apprentissage. Dans ce cas, ces neurones ne sont pas étiquetés et contribuent

éventuellement à renforcer les frontières entre classe. A l'issue de cette opération d'étiquetage, la SOM peut alors être utilisée pour la classification. Ainsi, pour un vecteur inconnu présenté en entrée de la SOM, la classe retenue est celle attribuée au neurone qui s'active.

Pour illustrer cette notion d'étiquetage nous utilisons l'exemple simple de la base de données iris de Fisher présente dans le logiciel Matlab (MathWorks®). L'ensemble de la base d'apprentissage comporte 150 échantillons regroupés en 3 classes de fleurs : iris setosa notée C_1 , iris virginica notée C_2 et iris versicolor notée C_3 (50 échantillons par classes). 4 paramètres mesurés constituent les caractéristiques de chaque échantillon. Par conséquent le vecteur d'entrée du R.N est composé de 4 éléments : la longueur et la largeur de sépale, la longueur et la largeur de pétale de fleur. Nous effectuons un apprentissage d'une carte SOM 2D constituée de $S=8*8$ neurones. Après apprentissage les données correspondant à chacune des classes sont présentées séparément en entrée du R.N. Nous présentons les projections obtenues pour chacune des classes sur la figure II.22.

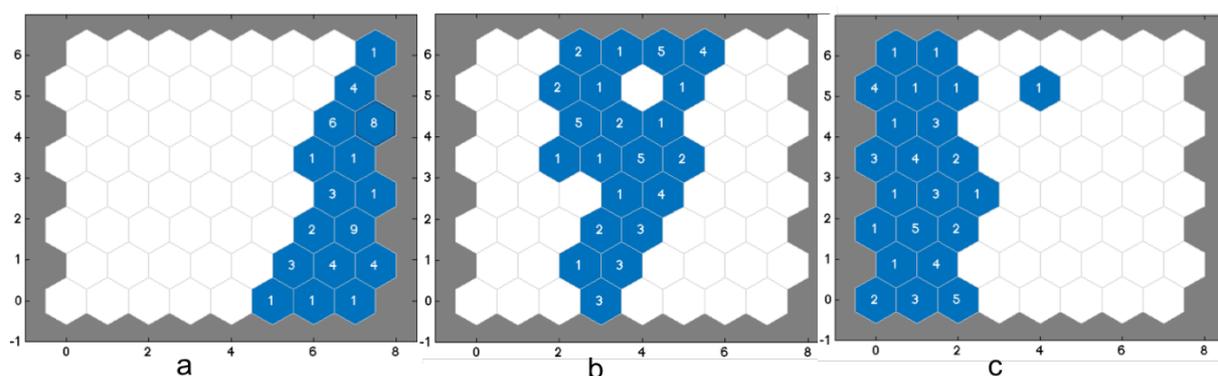


Figure II.22 : Projection des 3 classes C_1 (a), C_2 (b) et C_3 (c) sur une carte SOM de $S=8*8$ neurones dans le cadre de la base de données Iris de Fischer. Les chiffres à l'intérieur de chaque cellule donnent le nombre d'activation du neurone en question.

Chaque cellule de l'espace représente un neurone. Les cellules bleues correspondent aux neurones qui réagissent aux données d'entraînement de la classe particulière. Les chiffres à l'intérieur de chaque neurone donnent le nombre d'activations. Les cellules blanches quant à elles correspondent aux neurones non activés.

On observe un regroupement de chacune des classes par des neurones bien spécifiques de l'espace décrit par la SOM. On remarque que les échantillons de type C_1 activent uniquement les neurones dans la zone droite de la carte. Par ailleurs, ils ne sont jamais activés pour un échantillon de classe C_2 et C_3 . Ces neurones peuvent donc être étiquetés sans ambiguïté comme

sensibles à C_1 . On retrouve les mêmes conclusions pour les classes C_2 et C_3 . La figure II.23 représente la carte de neurones ainsi étiquetés. Notons également la présence de neurones n'ayant jamais été activés par un vecteur d'entrée qui renforce la frontière entre la classe C_1 et C_2 .

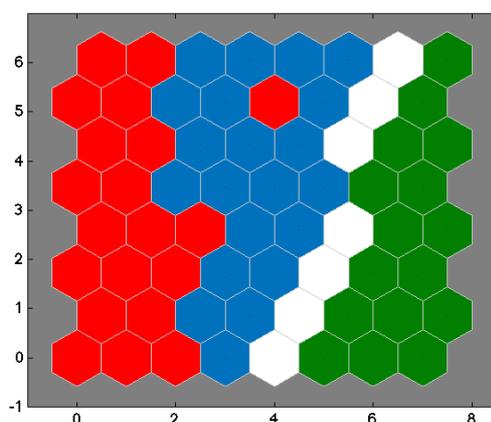


Figure II.23: Carte SOM de $S=8*8$ neurones étiquetée pour la classification dans le cadre de la base de données Iris de Fischer. Les neurones verts, bleus et rouges sont respectivement sensibles à C_1 , C_2 et C_3 . Les neurones blancs n'ont jamais été activés lors de la phase d'apprentissage

Cette opération d'étiquetage a permis de proposer un outil d'aide à la décision à partir d'une carte SOM. Notons cependant que l'apprentissage de la carte SOM ne prend pas en compte les classes des échantillons. Les poids des neurones sont ajustés de manière à minimiser une erreur de quantification. La carte est donc figée lors de la procédure d'étiquetage. Dans le cadre de nos travaux (cf chapitre I), il est facile de construire théoriquement un corpus d'apprentissage en reliant chaque entrée (signatures ellipsométriques) à une classe (modèle de profil géométrique). Il paraît donc intéressant d'utiliser cette information lors du processus d'apprentissage de la carte pour affiner les frontières. Le paragraphe suivant présente donc une variante des SOM s'appuyant sur un apprentissage semi-supervisé.

II.3.2.3.2 Learning Vector Quantization

II.3.2.3.2.1 Architecture du LVQ

Proposée également vers la fin des années 80 par Teuvo Kohonen [136] [137], le Learning Vector Quantization (LVQ) est un RN compétitif à apprentissage supervisé destiné à la classification.

D'un point de vue architectural, le LVQ est composé de 3 couches (figure II.24) : une couche d'entrée de n éléments, une couche cachée dite de compétition fonctionnant de façon similaire à une SOM suivant le principe du « winner takes all » et une couche de sortie composée de N_s neurones linéaires à valeurs binaires correspondant aux K classes connues.

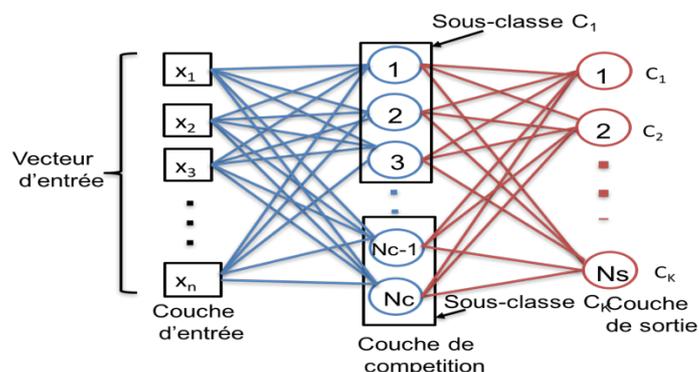


Figure II.24: Exemple d'architecture de LVQ utilisé dans le cadre de la classification semi-supervisé à n entrées composé d'une couche de compétition de N_c neurones et d'une couche de sortie de N_s neurones correspondant à chacune des classes connues.

La couche de neurones cachés possède N_c neurones dont chacun est connecté en entrée à toutes les mesures x_i mais également en sortie à tous les neurones de la couche de sortie. Les poids des connexions de chaque neurone caché j aux entrées sont regroupés dans un vecteur appelé vecteur prototype \mathbf{W}_j représentatif des données d'entrée : $\mathbf{W}_j = (w_{j,1}, w_{j,2}, \dots, w_{j,n})$ ($j=1, \dots, N_c$). La couche de sortie permet de prendre des décisions quant à l'attribution d'une classe en interprétant les données projetées sur la couche de compétition par l'intermédiaire des vecteurs prototypes. On définit également un deuxième vecteur de poids synaptique $\mathbf{Z}_j = (z_{j,1}, z_{j,2}, \dots, z_{j,N_s})$ ($j=1, N_s$) pour les connexions reliant le neurone j de la couche de sortie aux neurones cachés.

Le LVQ fonctionne de la manière suivante : lorsqu'un vecteur \mathbf{X} est présenté en entrée du réseau, le neurone gagnant G de la couche de compétition est sélectionné selon le principe de l'équation II.75. Sa réponse est égale à 1 et toutes les sorties des autres neurones de compétition sont fixées à 0. La couche cachée regroupe ainsi les vecteurs d'entrée en N_c sous classes. L'information est transmise à la couche de sortie à fonction de transfert linéaire afin de déterminer l'appartenance du vecteur d'entrée \mathbf{X} à l'une des N_s classes définies par l'utilisateur. Par exemple si les neurones 1, 2, 3 de la couche de compétition représentent tous des sous classes de l'espace d'entrée qui font partie de la classe C_1 , alors les poids des connexions reliant

ces neurones au premier neurone de sortie (soit $z_{1,1}$, $z_{1,2}$ et $z_{1,3}$) seront renforcés alors que tous les autres poids les reliant aux autres neurones de sortie seront dégradés. Ainsi le premier neurone de sortie tend à être activé lorsqu'un de ces 3 neurones cachés gagne la compétition pour une entrée \mathbf{X} donnée.

II.3.2.3.2.2 Apprentissage du LVQ

L'apprentissage du LVQ est de type supervisé comme défini dans le paragraphe II.3.2.1.5. Il s'effectue donc sur un ensemble de couples $\{\mathbf{X}^q, \mathbf{Y}^q\}$ ($q = 1, \dots, N_t$) où chaque vecteur de sortie cible \mathbf{Y} correspond à la classe à laquelle appartient l'entrée \mathbf{X} correspondante et ce parmi les K (donc N_s) classes possibles. Ce vecteur possède donc K composantes dont une seule est non nulle et égale à « 1 ».

Avant l'apprentissage, on procède à l'étape d'initialisation qui se déroule en 2 phases. D'abord, chacun des poids $w_{j,i}$ est initialisé en prenant la valeur moyenne de toutes les entrées x_i disponibles dans le corpus d'apprentissage. Ensuite, il faut assigner les neurones de la couche de compétition à chacune des classes. Pour cela, l'algorithme d'initialisation étudie la distribution des sorties désirées \mathbf{Y}^q pour toute les entrées. Ainsi si $p\%$ des vecteurs \mathbf{Y}^q de la base d'apprentissage appartiennent à la classe C_k associée à la sortie S_k alors $p\%$ des m neurones de la couche cachée seront associés à des sous-classes de C_k . Ainsi dans l'exemple précédent, si les neurones 1, 2 et 3 sont désignés pour former des sous-classes de la classe 1 les poids $z_{1,1}$, $z_{1,2}$ et $z_{1,3}$ prendront la valeur 1 et tous les autres poids $z_{1,j}$ ($j=4, \dots, N_s$) la valeur 0. La matrice \mathbf{Z} contenant les poids $z_{m,j}$ est donc fixée avant de commencer l'apprentissage.

Une fois l'initialisation terminée, l'apprentissage peut commencer. On utilise la matrice \mathbf{Z} et les vecteurs des sorties désirées \mathbf{Y}^q pour rechercher les vecteurs de sorties désirées \mathbf{Y}_c pour la couche cachée. On a :

$$\mathbf{Y}_c^q = \mathbf{Z}^{-T} \mathbf{Y}^q \quad \text{II.82}$$

Il s'agit ici de l'équivalent de la retropropagation du gradient de l'erreur pour les PMC (cf paragraphe II.3.2.2.2).

L'apprentissage du réseau LVQ consiste simplement à positionner au mieux les vecteurs prototypes en employant la règle de Kohonen (cf paragraphe II.3.2.3.1.2). Ainsi, pour chaque vecteur \mathbf{X} présenté en entrée du LVQ, seul le poids du neurone gagnant est modifié. Contrairement à la carte SOM, il n'y a pas de notion de voisinage pour ce type de R.N.

Si la sortie du neurone gagnant de la couche cachée correspond à la valeur attendue (c'est à dire 1), les poids \mathbf{W}_j de ce neurone sont rapprochés de l'entrée comme suit :

$$\mathbf{W}_G(t + 1) = \mathbf{W}_G(t) + \alpha(t)[\mathbf{X} - \mathbf{W}_G(t)] \quad \text{II.1}$$

Dans le cas contraire, lorsque le neurone vainqueur n'est pas celui attendu, c'est à dire que la sortie désirée vaut 0 (alors le neurone gagnant ne correspond pas à une sous classe de la classe attendue). Il s'agit simplement d'éloigner son vecteur référent de l'entrée en question afin qu'un autre neurone associé à la bonne classe puisse gagner lors de la prochaine itération. La règle de Kohonen devient alors :

$$\mathbf{W}_G(t + 1) = \mathbf{W}_G(t) - \alpha(t)[\mathbf{X} - \mathbf{W}_G(t)] \quad \text{II.2}$$

L'apprentissage du LVQ s'arrête lorsque les vecteurs prototypes sont stabilisés ou le nombre d'itérations maximal fixé par l'utilisateur est atteint.

Les LVQ sont largement utilisés dans la littérature scientifique pour de nombreuses applications. A titre d'exemple, une étude menée dans [138] montre les performances obtenues dans la classification des signaux. Il s'agit de classer les signaux encéphalographiques d'un être humain en bonne santé en deux catégories dans le but de déterminer son état de somnolence et de vigilance. Les résultats du LVQ sont, dans ce cas, bien meilleures que ceux du PMC.

Dans ce chapitre nous avons énoncé les principes généraux de la classification automatique ainsi que les méthodes les plus couramment utilisées dans ce cadre. Plusieurs modèles de classifieurs ont été présentés prenant en compte la complexité des frontières de décision allant de la simple séparation linéaire (SVM) à la séparation non linéaire (RN, SVM non linéaires, KPPV) d'ordre élevé. Parmi les classifieurs neuronaux nous avons notamment détaillé le PMC et les réseaux de Kohonen (SOM et LVQ). La littérature scientifique montre que les résultats d'une classification neuronale est similaire à ceux des autres méthodes de classification non linéaire. L'avantage des R.N réside dans la facilité du déploiement, la robustesse des solutions et la résistance aux bruits de mesure. Dans le chapitre suivant, nous allons exploiter ce potentiel des R.N. pour répondre à la nécessité de classification dans le processus de fabrication et de caractérisation scattérométrique.

Chapitre III Le réseau de neurones comme outil de détection

Dans le chapitre II nous avons présenté les R.N dans la classification. Ce chapitre traite de l'utilisation de ces derniers pour répondre aux besoins de classification dans la métrologie optique mis en évidence au chapitre I. En effet, nous proposons plus particulièrement l'utilisation de RN pour deux types d'applications adressées par le présent chapitre : la première concerne la détection d'anomalie en photolithographie et la seconde la sélection de modèles géométriques pour la caractérisation scatterométrique classique. Pour chacune de ces applications, nous analyserons les projections de type SOM des signatures ellipsométriques dans le but de montrer la faisabilité de la méthode. Ensuite, nous utiliserons les R.N de type LVQ et PMC présentés dans le chapitre II pour effectuer la tâche d'aide à la décision par l'intermédiaire d'une classification supervisée. Enfin, des mesures sur des cas réels seront proposées et analysées.

III.1 Détection d'une couche résiduelle

III.1.1 Cadre de l'étude

L'étude est réalisée dans le but de pouvoir détecter les défauts lithographiques liés à la présence d'une couche résiduelle de résine ou autre (Cf. Chapitre I). En effet, un masque de résine ne débouchant pas directement sur le substrat au fond des sillons peut, par exemple, être considéré comme un défaut incompatible pour l'étape de gravure ultérieure. C'est cet exemple qui nous servira de fil conducteur pour cette partie du chapitre. L'intérêt de l'étude est de pouvoir détecter le défaut avant la gravure afin d'interrompre le processus pour des raisons évidentes de coût de fabrication et de temps et pouvoir recycler la plaque.

La structure périodique diffractante retenue pour l'étude possède une période de 750 nm. Le réseau périodique est constitué de résine NEB22 déposée sur un substrat de Silicium. Nous considérons deux classes de profil : l'une, notée P_1 , présentant une forme trapézoïdale en résine et l'autre, notée P_2 , présentant la même forme avec une couche résiduelle de résine supplémentaire entre la structure périodique et le substrat (Fig.III.1).

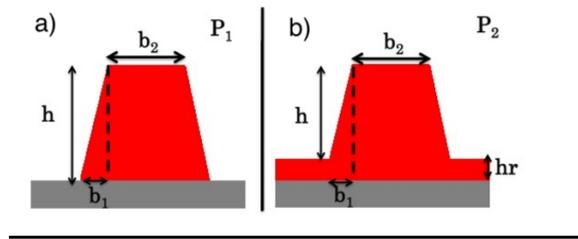


Figure III. 1: Description paramétrique des formes géométriques de profil P_1 (a) et P_2 (b) utilisées pour la détection automatique de couche résiduelle

Le profil trapézoïdal étant défini par 3 paramètres (b_1 , b_2 et h) et la couche résiduelle par sa seule épaisseur h_r , la modélisation du profil P_1 , respectivement P_2 , nécessite 3, respectivement 4, paramètres géométriques (figure III.1). Pour notre étude, 6000 profils sont générés de façon aléatoire (3000 pour chaque classe) dans les plages suivantes :

$$0 \leq b_1 \leq 70 \text{ nm},$$

$$100 \leq b_2 \leq 350 \text{ nm}$$

$$170 \leq h \leq 250 \text{ nm}$$

$$1 \leq h_r \leq 100 \text{ nm}$$

III.1.1 Visualisation par projection des signatures optiques

La carte SOM a la propriété intéressante de donner une visualisation par projection non linéaire sur un espace visualisable de l'ensemble des données d'entrée composé des signatures $\{I_s, I_c\}$ correspondant aux profils P_1 et P_2 . Ces dernières sont calculées par MMMFE pour des longueurs d'onde comprises entre 1.5 eV et 6.5 eV avec un pas de 0.2 eV (soit 52 valeurs) dans des conditions expérimentales fixes (angle d'incidence : 70°). Une signature du même type est utilisée en scattérométrie classique. Afin de simuler des mesures, un bruit gaussien ($\sigma=0,003$) est pris en compte. Les 6000 signatures bruitées sont donc normalisées puis fournies en entrée d'une carte SOM.

III.1.1.1 SOM sans classification

La carte SOM (cf. chapitre I) utilisée est constituée de 400 neurones composant une grille 2D (20 x 20 neurones). Elle est entraînée avec les 6000 signatures de profils présentés plus haut. Cette phase a duré environ 16 min sur un ordinateur Intel Xeon CPU E5 1620 @3.60 GHz 12 Go RAM. Une visualisation de la projection des signatures est alors possible. Chaque neurone de la carte représente un ou plusieurs types de profils et peut donc être étiqueté selon la technique décrite dans le chapitre II

paragraphe II.3.2.3.1.2. Ensuite, la SOM peut répondre à une nouvelle signature (inconnue de la phase d'apprentissage) par l'activation d'un neurone de la carte. Ce dernier regroupe un ensemble de profils du corpus d'apprentissage qui peuvent être considérés comme « similaires » à celui de l'échantillon présenté.

Ainsi, la Figure III.2 présente la répartition sur la carte de tous les neurones activés par les signatures de profils P_1 à l'issue de l'apprentissage (neurones en bleu). Les chiffres inclus dans chaque cellule correspondent au nombre de profils P_1 qui activent le neurone en question. On remarque que les neurones concernés sont distribués dans une zone bien distincte de la carte.

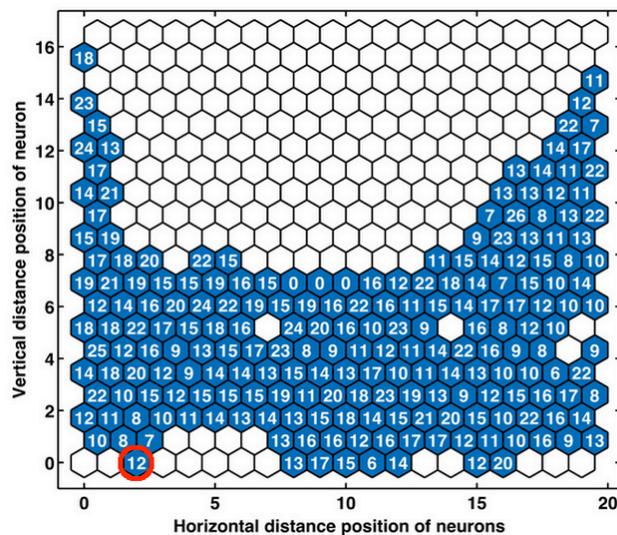


Figure III. 2: Résultats de la projection sur une carte SOM (20×20 neurones) des 3000 signatures correspondant à un profil P_1^*

Considérons à présent les signatures de profils P_2 (fig. III.3). Pour la majorité (2607 signatures soit 86.9 % des signatures P_1), on se retrouve dans le même cas de figure que précédemment. En effet, ces profils activent des neurones spécifiques (en vert) répartis sur la carte dans une zone clairement identifiée. Par contre, une faible part d'entre eux (393 signatures soit 13.1 %) stimulent des neurones déjà activés par des signatures P_1 (bleu). Il faut cependant noter, si on se focalise sur le nombre de signatures concernées dans les 2 cas, que ces neurones sont préférentiellement activés par des profils P_1 .

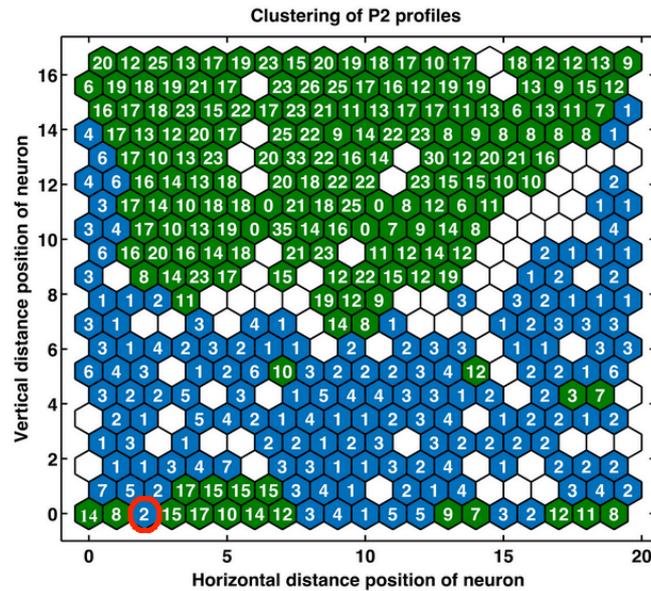


Figure III. 3: Résultats de la projection sur une carte SOM (20×20 neurones) des 3000 signatures correspondant à un profil P_2 . Les neurones verts sont exclusivement activés par les profils P_2 et les neurones bleus sont ceux activés à la fois par P_2 et P_1 .

Il est intéressant de s’arrêter sur les caractéristiques de ces 393 profils P_2 . Nous avons représenté sur la figure III.4 l’épaisseur hr correspondante.

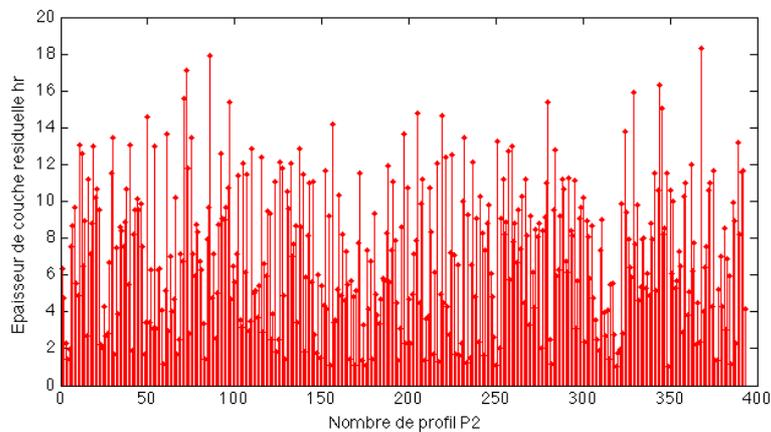


Figure III. 4: Epaisseurs hr des 393 profils P_2 ayant activé les neurones bleus de la figure III.3.

Ces signatures correspondent à des profils, P_2 présentant une faible épaisseur résiduelle (moyenne d’environ 7 nm et écart type de 3,76 nm). La plus forte épaisseur détectée est 18 nm. Cette valeur constitue la limite d’épaisseur de résine détectable sans ambiguïté.

Considérons par exemple le neurone identifié par le vecteur $\mathbf{r}(0^2)$ (3^{ième} neurone de la première ligne entouré de rouge sur les figures III.2 et III.3). Il a été activé par 12 signatures P_1

(fig. III.2) et 2 signatures de profil P_2 (fig. III.3) à l'issue de l'apprentissage. Nous avons tracé toutes ces signatures dans la figure III.5.

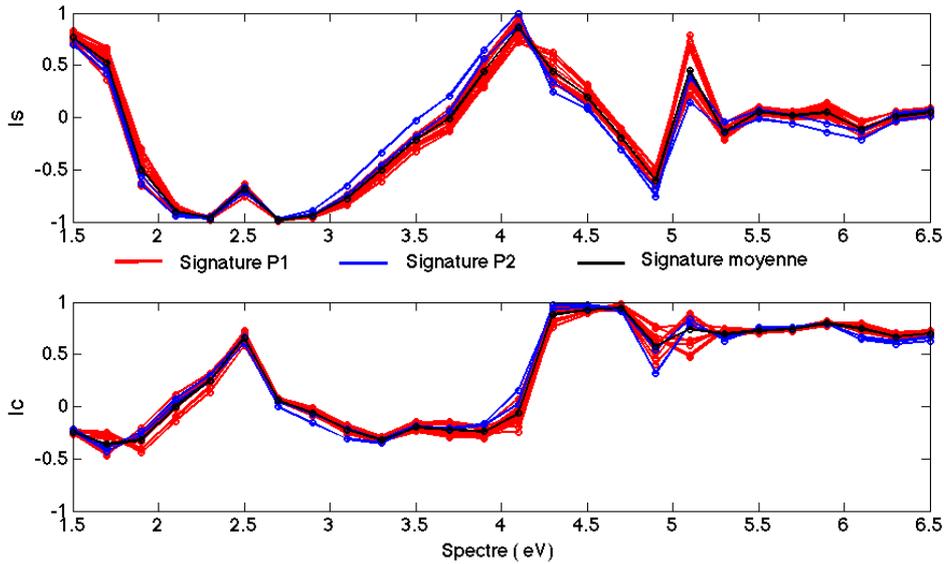


Figure III. 5 : Signatures (I_s et I_c) des profils P_1 (bleu), P_2 (rouge) et signature moyenne (noir) ayant activé le neurone par $\mathbf{r}(\mathbf{0})$ de la carte SOM 20*20.

On peut constater que les signatures sont similaires et assez proches les unes des autres pour stimuler le même neurone. Pour s'en convaincre, on peut également définir l'écart moyen de chaque signature à la moyenne caractérisant l'« empreinte » du neurone activé par :

$$E_{I_{sIc}} = \sqrt{\frac{1}{M} [I_{smoy} - I_s]^2 + [I_{cmoy} - I_c]^2} \quad \text{III.1}$$

avec $M=52$ le nombre de mesures d'intensité I_s et I_c compris dans la signature et I_{cmoy} et I_{smoy} les signatures moyennes de I_s et I_c

Les tableaux III.1 et III.2 montrent $E_{I_{sIc}}$ calculée pour chacune des 14 signatures.

Tableau III. 1 Ecart moyen $E_{I_{sIc}}$ à la signature moyenne du neurone $\mathbf{r}(\mathbf{0})$ pour chacune des signatures P_1 de la figure III.5

Signatures P_1	S_1	S_2	S_3	S_4	S_5	S_6
$E_{I_{sIc}}$	0.0643	0.0427	0.0706	0.0853	0.0868	0.0649
Signatures P_1	S_7	S_8	S_9	S_{10}	S_{11}	S_{12}
$E_{I_{sIc}}$	0.0315	0.0634	0.0545	0.0499	0.0509	0.0733

Tableau III. 2 Ecart moyen E_{Islc} à la signature moyenne du neurone $\mathbf{r}_0^{(2)}$ pour chacune des signatures P_2 de la figure III.5

Signatures P_2	S_1	S_2
E_{Islc}	0.1023	0.0726

On remarque que les erreurs des signatures P_2 (tableau III.2) sont légèrement plus élevées que la plupart des signatures P_1 (tableau III.1), mais restent du même ordre de grandeur.

Les profils correspondants sont représentés sur la figure III.6. On peut remarquer que les dimensions géométriques sont relativement proches. La couche résiduelle des profils de type P_2 est dans ce cas faible (environ 8 nm).

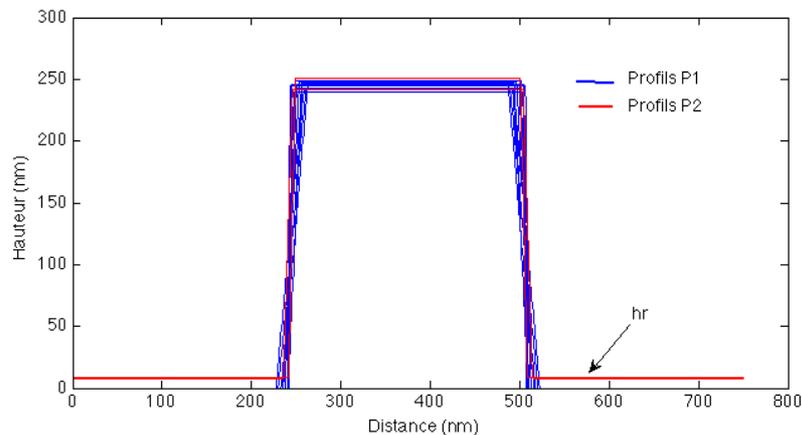


Figure III.6: Représentation des Profils géométriques P_1 et P_2 ayant activé le neurone $\mathbf{r}_0^{(2)}$ de la carte SOM 20*20.

Au vu de ces résultats, on peut considérer que chaque neurone de la carte représente un ensemble de profils qui possèdent des signatures proches d'une signature de référence.

III.1.1.2 SOM comme outil d'aide à la décision

Intrinsèquement, la carte SOM ne permet pas d'effectuer une classification, mais donne une visualisation graphique de la répartition des signatures à distinguer. Nous procédons donc à un étiquetage des neurones de la carte SOM entraînée précédemment sur la base d'un calcul de probabilité afin de fournir des éléments d'aide à la décision pour la détection de défauts. Ainsi, la probabilité qu'un neurone spécifique de la carte noté j représente un profil de type P_i ($i=1, 2$) est définie par :

$$p_j(P_1) = \frac{k}{m} \quad \text{III.2}$$

m étant le nombre total d'échantillons ayant activé le neurone j et k le nombre de profils P_1 concernés lors de l'apprentissage.

Ainsi, la probabilité $p_j(P_2)$ de tous les neurones colorés en bleu sur la figure III.3 est très faible et proche de zéro. Par contre, les probabilités $p_j(P_2)$ de tous les neurones colorés en vert est égale à 1. Le neurone particulier r_2^0 évoqué plus haut possède les probabilités suivantes $p(P_1) = \frac{12}{14} = 0.86$ et $p(P_2) = \frac{2}{14} = 0.14$.

Afin d'assurer une détection automatique, un seuil d'acceptabilité peut être ensuite fixé sur la probabilité p_j . Le cas le plus simple consiste à choisir la valeur 0.5. Ainsi, chaque neurone j de la carte possédant une probabilité $p_j(P_2)$ respectivement si $p_j(P_1)$ supérieur à 0.5 sera étiqueté comme de type P_2 respectivement P_1 . Sur la base de ce seuil, le neurone r_2^0 sera étiqueté sensible au profil P_1 . La figure III.7 montre l'ensemble de la carte neuronale ainsi étiquetée. Les cellules vertes sont les neurones sensibles aux signatures P_2 et les cellules bleues aux signatures de profils P_1 . Les cases blanches sont des neurones qui n'ont jamais été activés lors de l'apprentissage.

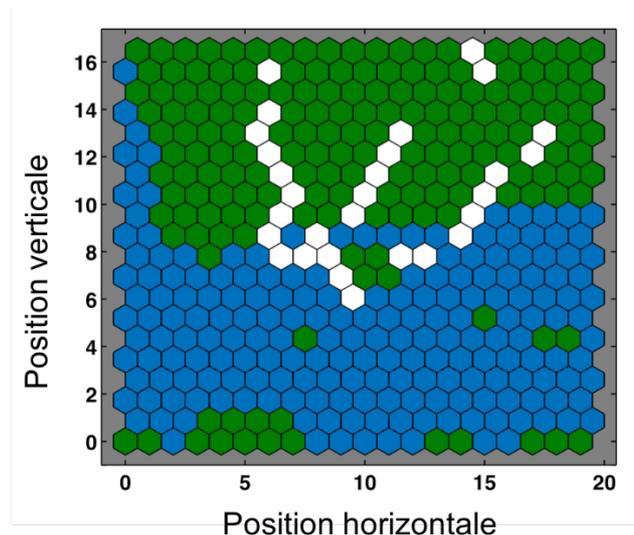


Figure III. 7: Carte SOM (20×20 neurones) étiquetée pour la détection de couche résiduelle avec un seuil de probabilité d'appartenance fixé à 0.5. Les neurones verts sont sensibles aux signatures de profils P_2 et les neurones bleus aux signatures de profils P_1 . Les neurones blancs n'ont jamais été activés lors de l'apprentissage.

Le seuil de probabilité de 0.5 permet de décider sur tous les neurones de la carte, mais nous avons vu que dans certains cas de figure le risque est élevé. Pour réduire ce risque de décision, il peut être intéressant de fixer un seuil plus élevé, par exemple 0.9. La figure III.8 présente la carte neuronale nouvellement étiquetée.

Le même code de couleur est utilisé : les neurones verts sont les neurones sensibles aux signatures P_2 (avec $p(P_2) > 0.9$), les neurones bleus ceux sensibles aux signatures P_1 (avec $p(P_1) > 0.9$). Parmi les neurones majoritairement activés par des signatures P_1 on observe que certains présentent des probabilités $p(P_1)$ inférieures à 0.9. Ceux-ci sont colorés en orange sur la figure III.8.

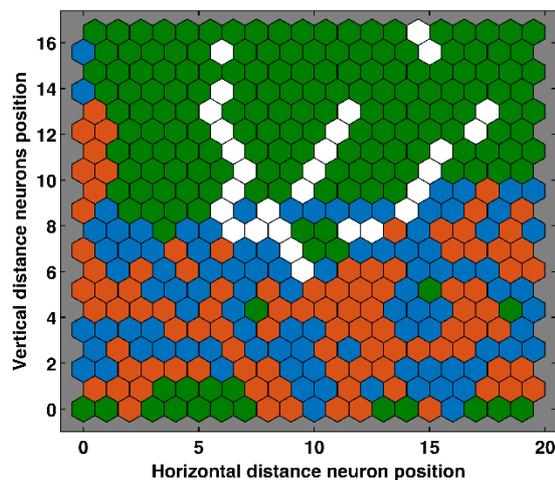


Figure III. 8 : Carte SOM (20×20 neurones) étiquetée pour la détection de couche résiduelle avec un seuil de probabilité d'appartenance fixé à 0.9. Les neurones verts sont sensibles aux signatures de profils P_2 , les neurones bleus aux signatures de profils P_1 , les neurones oranges présentent une probabilité $\max(p(P_1), p(P_2))$ inférieure au seuil. Les neurones blancs n'ont jamais été activés lors de l'apprentissage.

Les résultats décrits sont bien entendu propres au cas particulier étudié. Ainsi, les étiquetages sont liés au nombre et au type d'échantillons composant le corpus ainsi qu'à la structure de la SOM. Si on suppose que la base d'apprentissage est suffisante pour décrire l'espace des entrées, le nombre de neurones « indécis » est directement lié à l'architecture du RN. Si on veut réduire ce nombre, il est essentiel de déterminer le nombre de neurones optimal composant la carte.

III.1.1.3 Optimisation de la carte SOM

III.1.1.3.1 Etude de différentes architectures

Nous effectuons l'apprentissage de différentes architectures de cartes SOM 2D en modifiant le nombre de neurones de la façon suivante : 20*20, 25*25, 30*30, 35*35, 40*40, 45*45, 50*50, 60*60, 65*65, 70*70, 75*75 et 80*80. L'apprentissage est, par ailleurs, réalisé dans les mêmes conditions que précédemment (même corpus d'apprentissage, même nombre d'échantillons). On a représenté sur la figure III.9 les épaisseurs h_r moyennes et maximales des profils P_2 injustement classés P_1 en fonction du nombre de neurones constituant la carte. Dans le même temps, on trace sur la figure III.10 le nombre de signatures P_2 ayant activé des neurones préférentiellement sensibles aux signatures P_1 pour chaque architecture étudiée. La figure III.11 présente alors le pourcentage de neurones actifs (inversement non actifs c'est-à-dire n'ayant jamais été activés lors de l'apprentissage) en fonction du nombre de neurones.

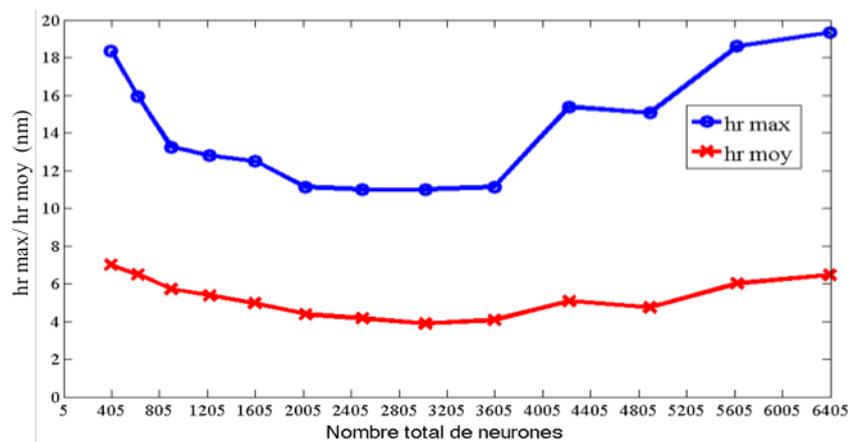


Figure III. 9 : Représentation des épaisseurs h_r moyennes (x en rouge) et maximales (o en bleu) des profils P_2 injustement classés P_1 (en fonction de la dimension de la carte)

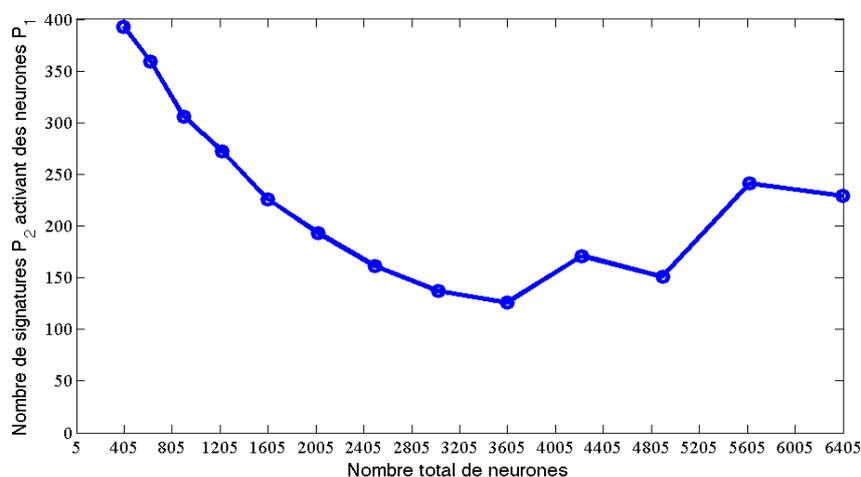


Figure III. 10 : Représentation du nombre de signatures P_2 activant des neurones étiquetés P_1 en fonction de la dimension de la carte

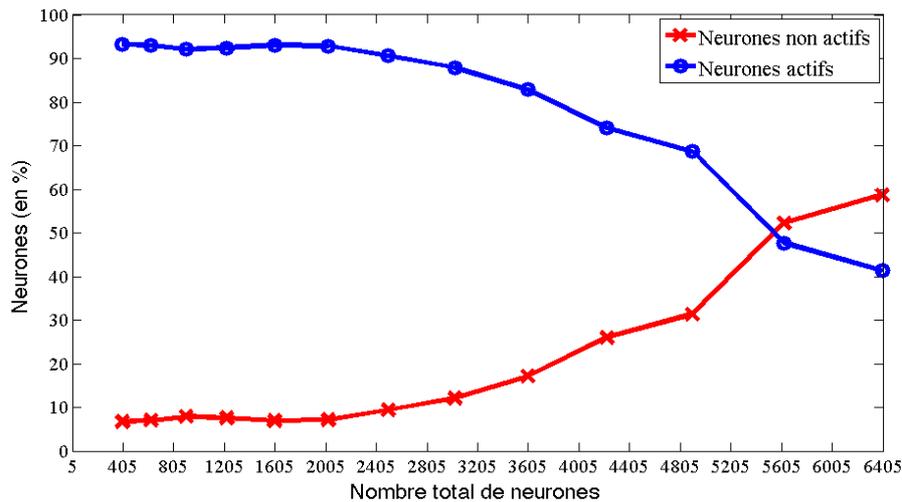


Figure III. 11 : Représentation du pourcentage de neurones actifs (o en bleu) et non actifs (x en rouge) en fonction de la dimension de la carte

On retrouve pour 400 neurones les valeurs décrites dans le paragraphe précédent ; à savoir ; un seuil maximal sur hr de 18 nm et 393 signatures P_2 qui activent des neurones préférentiellement sensibles à P_1 . Le nombre de neurones actifs est de 93.25%. Ces résultats montrent clairement une architecture optimale présentant une réduction des valeurs de hr (jusqu'à 11 nm pour la valeur seuil et 4.5 en moyenne). Cette dernière correspond logiquement à la carte présentant un minimum de signatures P_2 confuses (125 pour la carte SOM 60*60 neurones).

La carte 400 neurones reste celle avec le plus faible pourcentage de neurones non actifs (6.75 %) mais avec le seuil sur hr le plus haut. Ce pourcentage augmente avec l'augmentation du nombre de neurones de la carte. Ces neurones non activés ne peuvent être étiquetés. Ils permettent de marquer plus profondément les frontières entre les classes. Une signature inconnue activant un de ces neurones sera systématiquement rejetée comme n'étant dans aucune des 2 classes. Notons que l'augmentation trop importante du nombre de neurones inactifs conduit à une diminution des performances. Ceci peut provenir du fait que notre étude est réalisée à nombre d'échantillons fixe.

La SOM constituées de 2025 (45*45) neurones peut être retenue. Cette dernière présente un pourcentage de neurones non actif de 7.2 %, seuil d'épaisseur détectable sans ambiguïté de l'ordre de 11 nm (en moyenne 4.5 nm) et présentant 193 signatures P_2 confuses soit 6.43 %.

Le choix du nombre adéquat de neurones est important car il modifie la granularité de l'espace de sortie et permet de réajuster la distribution des signatures d'entrée pour une meilleure répartition dans l'espace de sortie. Rappelons que dans cette étude, nous n'avons pas

modifié la qualité de la signature fournie en entrée. Cette dernière peut être également optimisée pour améliorer les performances ci-dessus.

III.1.1.3.2 Tests sur des échantillons théoriques

Afin d'évaluer les performances de la carte neuronale en situation réelle, nous allons considérer 3 échantillons distincts de profil P_2 avec différentes épaisseurs de couche résiduelle h_r :

- Echantillon 1: $b_1 = 0.5$ nm, $b_2 = 131.4$ nm, $h = 182.6$ nm, $h_r = 31.6$ nm
- Echantillon 2: $b_1 = 3.9$ nm, $b_2 = 103.3$ nm, $h = 238,1$ nm, $h_r = 65.6$ nm
- Echantillon 3: $b_1 = 58.1$ nm, $b_2 = 176.2$ nm, $h = 220.2$ nm, $h_r = 23.3$ nm

Pour chaque échantillon, nous simulons la signature mesurée puis toutes celles correspondant à ce même échantillon pour lequel on réduit itérativement l'épaisseur h_r de 2 nm jusqu'à atteindre une épaisseur nulle et donc un profil P_1 . Ensuite, toutes ces signatures simulées sont fournies en entrée de la carte SOM précédemment étiquetée (figure III.7). Les activations des neurones G sont présentées sur la figure III.12 et les probabilités $p_G(P_2)$ correspondantes sont représentées dans la figure III.13 en fonction de l'épaisseur h_r de l'échantillon.

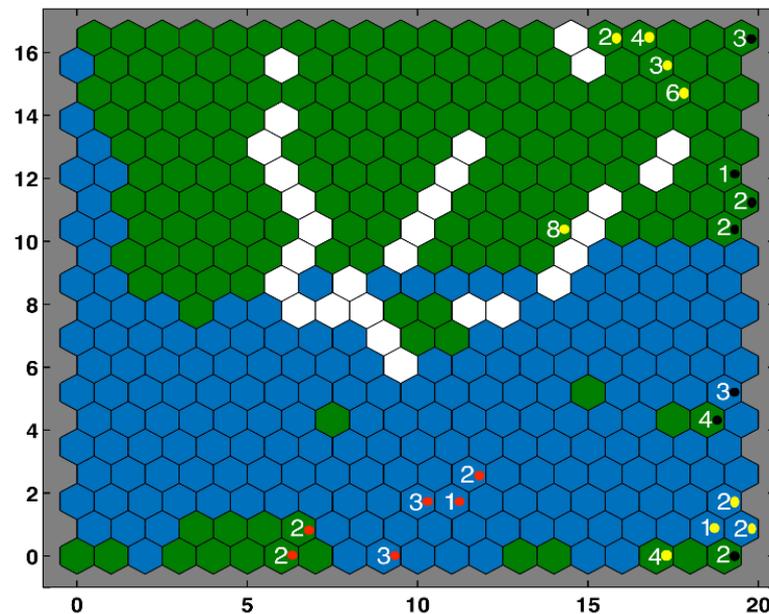


Figure III. 12 : Projection des signatures pour 3 échantillons distincts sur la carte SOM 20×20 . Les neurones marqués en noir sont stimulés par les signatures de l'échantillon 1, ceux marqués en jaune respectivement en rouge sont stimulés par les signatures d'échantillons 2 respectivement 3.

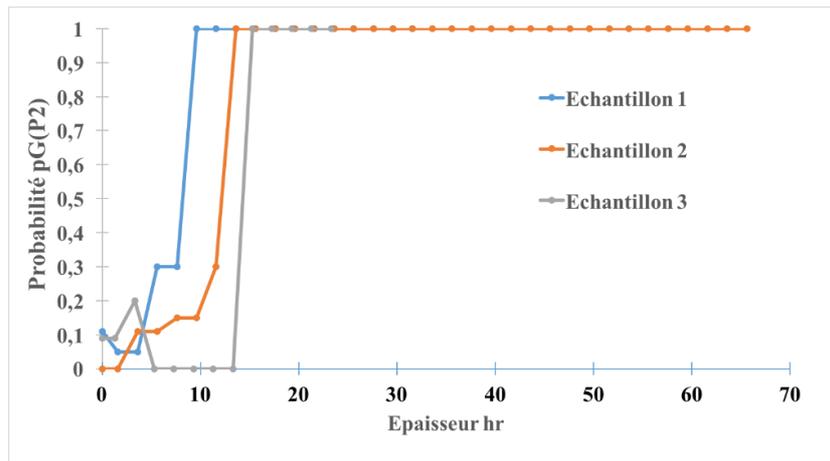


Figure III. 13 : Probabilité $p_G(P_2)$ du neurone gagnant de la carte SOM (20x20 neurones) pour 3 échantillons distincts pour lesquels on réduit artificiellement l'épaisseur hr

On remarque que le comportement est sensiblement différent d'un échantillon à un autre. La probabilité reste néanmoins assez contrastée. En effet, il existe peu de valeur intermédiaire proche de 0.5. Pour un seuil de probabilité fixé à 0.5, on observerait une limite de détection sans ambiguïté entre 8 et 13 nm sur ces 3 échantillons. La limite de détection dépend de l'échantillon considéré (valeurs de b_1 , b_2 , et h) et donc de la sensibilité du neurone activé. Ceci peut conduire à des frontières plus ou moins « abruptes » séparant les 2 classes sur la carte.

Ensuite, les mêmes signatures des 3 échantillons définies précédemment sont fournies en entrée de la SOM 45*45 neurones. Les probabilités $p_G(P_2)$ des neurones gagnants sont représentées sur la figure III.14.

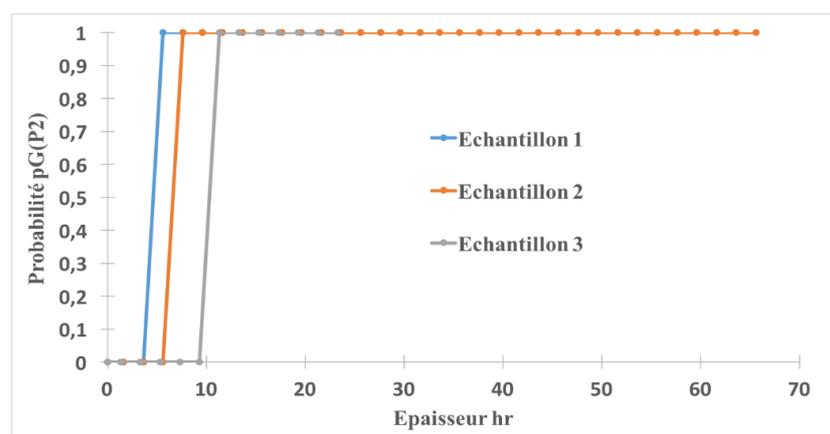


Figure III. 14 : Probabilité $p_G(P_2)$ du neurone gagnant de la SOM (45x45 neurones) pour 3 échantillons distincts pour lesquels on réduit artificiellement l'épaisseur hr

La limite de détection sans ambiguïté est réduite (de 4 à 11 nm) conformément à ce qui a été montré dans le paragraphe précédent. Notons également que la frontière est beaucoup plus abrupte entre les 2 classes.

III.1.1.4 Test sur échantillon réel

Pour valider les résultats théoriques, nous disposons d'un échantillon fabriqué par nanoimpression. Dans ce type de fabrication, la couche résiduelle est laissée cette fois-ci volontairement au fond des sillons après pressage du moule dans la résine pour éviter tout contact entre le moule et le substrat. L'échantillon est composé d'une couche de résine NEB22 déposée sur un substrat de silicium. La forme géométrique du profil est supposée trapézoïdale et la période est de 750 nm. L'estimation des paramètres géométriques par une caractérisation MEB (figure III.15) est $b_1=40$ nm, $b_2=314$ nm, $h=180$, $h_r=70$ nm.

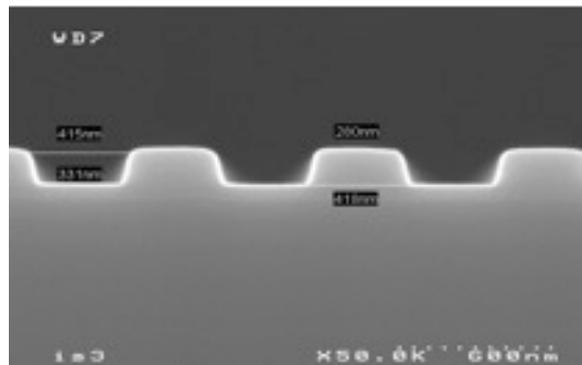


Figure III. 15 : Caractérisation MEB de l'échantillon fabriqué par nanoimpression : structure de forme trapézoïdale en résine NEB22 de période 750 nm sur Silicium présentant une couche résiduelle de 70 nm

La signature optique est mesurée à l'aide d'un ellipsomètre à modulation de phase (UVISEL de Horiba Jobin Yvon) puis fournie aux 2 cartes étudiées précédemment (SOM 20*20 neurones et SOM 45*45 neurones). La signature active dans les 2 cas un neurone entièrement sensible aux signatures de profils P_2 (entouré en rouge sur les figures III.16 (a et b)). Ceci est bien conforme aux résultats attendus présentés ci-dessus. La signature du profil P_2 est donc identifiée sans aucune ambiguïté par les 2 cartes neuronales.

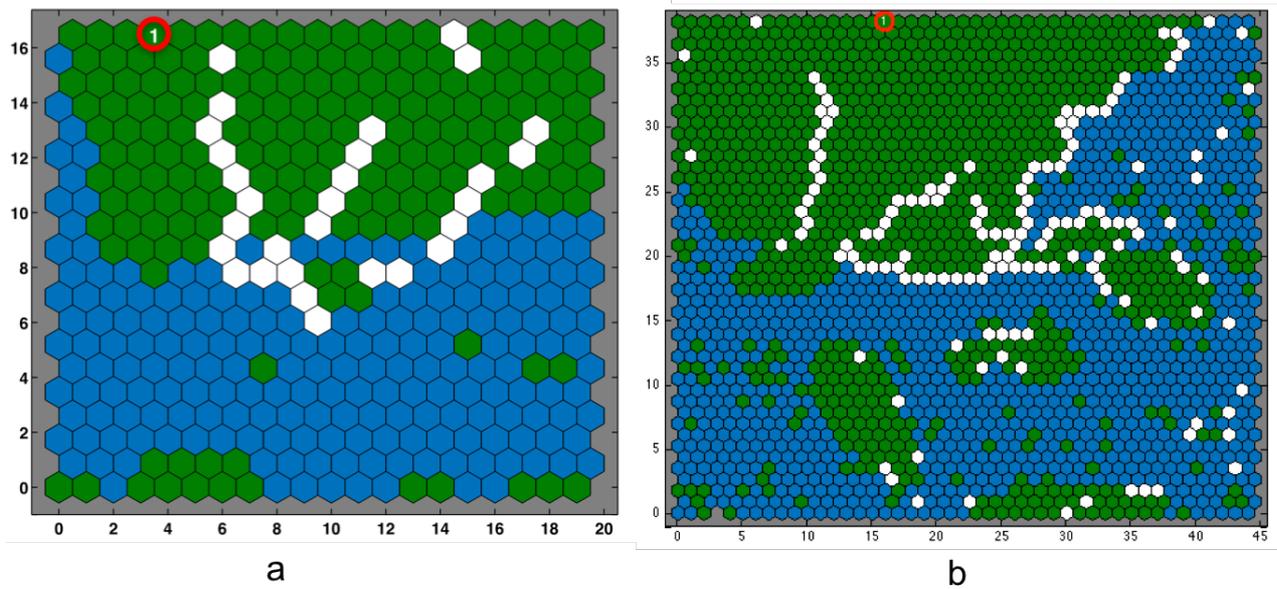


Figure III. 16 : Projection de la signature expérimentale sur les cartes SOM 20*20 (a) et SOM 45*45(b). Le neurone activé est totalement sensible aux signatures P_2 (entouré en rouge)

Ce premier résultat semble confirmer la bonne concordance entre le modèle théorique de signatures (utilisé pour l'apprentissage de la SOM) et la mesure sur l'échantillon. Toutefois, il est évident qu'un seul échantillon n'est pas suffisant et il faut poursuivre les tests sur une série de mesures pour valider ces résultats.

L'étude de la projection sur une carte SOM nous a permis de montrer qu'il existe une certaine ressemblance entre les signatures ellipsométriques d'un même profil. Il est donc nécessaire d'entrevoir une méthode pour les discriminer. La carte SOM peut en constituer une en procédant à un étiquetage des neurones permettant d'estimer la sensibilité de ce dernier à un profil précis. Cette estimation offre la possibilité d'effectuer une inspection systématique. Toutefois, nous avons montré que pour une épaisseur de couche résiduelle faible (environ 10 nm suivant les cas), cette tâche s'avère difficile pour la carte car les signatures correspondantes sont relativement proches de celles de certains profils P_1 .

Rappelons que la SOM n'est pas un classifieur en tant que tel. Elle ne propose qu'une projection particulière des signatures sur laquelle on propose une discrimination. Dans notre cas, il serait dommage de se priver de la connaissance *a priori* de la classe de chaque échantillon composant

la base d'apprentissage. Ainsi, il est intéressant de mettre en œuvre un R.N purement dédié à la classification supervisée.

III.1.2 Utilisation de la classification neuronale

Dans ce paragraphe nous effectuons une classification neuronale supervisée pour l'identification des deux types de profils P_1 et P_2 . Le classifieur neuronal apprend la relation entre la signature et la classe de profil connue. Après avoir reformulé le problème et défini les critères de bon classement, nous montrerons les résultats obtenus avec un R.N de type LVQ puis un PMC.

III.1.2.1 Critères de performance pour la classification

Nous utilisons les mêmes signatures simulées correspondant aux 2 classes P_1 et P_2 ayant servi à la visualisation de la carte SOM énoncée au paragraphe précédent (6000 signatures). L'apprentissage est effectué conformément à la méthode du « early stopping » mentionnée au chapitre II paragraphe II.3.2.2.2. Ainsi 75 % (soit 4200 signatures) constituent le lot d'apprentissage, 15 % (soit 900 signatures) celui de validation et les 15 % restants (donc 900 signatures) sont retenues pour le test. La subdivision des signatures est faite manuellement afin de garder la même composition des signatures qui serviront pour le test des RN. L'entrée du système est constituée de 52 mesures (I_s, I_c) et, cette fois-ci de 2 sorties pour chacune des 2 classes de profil géométrique. La figure III.17 illustre le système de détection de couche résiduelle par un classifieur neuronal.

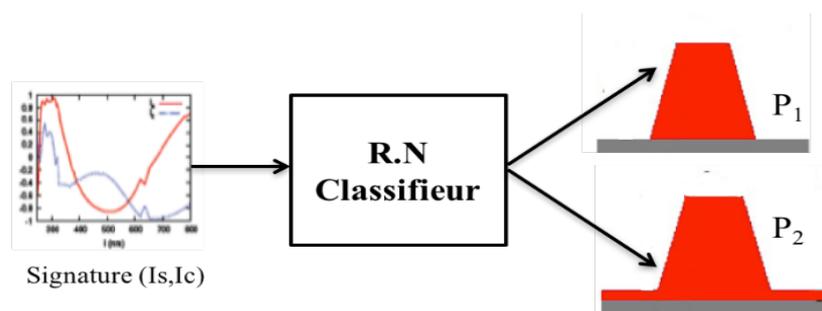


Figure III. 17 : Schéma du système de détection de couche résiduelle par un classifieur neuronal

Dans le cadre d'une classification supervisée à 2 classes, on distingue 2 types d'erreurs : fausse acceptation et faux rejet [139]. Ainsi, pour une classe donnée (P_1 par exemple) et un échantillon fixé, 4 cas peuvent se présenter :

- L'échantillon appartient à la classe P_1 et le classifieur ne se trompe pas. Le résultat est donc un vrai positif (TP pour True Positive). C'est une bonne acceptation du classifieur.
- L'échantillon appartient à la classe P_1 mais, le classifieur le rejette faussement et le classe donc en P_2). Le résultat négatif à tort du classifieur est dit un faux négatif (FN pour False Negative).
- L'échantillon n'appartient pas à la classe P_1 mais le classifieur la lui attribue quand même. Le résultat est un **faux positif** (False Positive : FP) car c'est une fausse acceptation du classifieur.
- L'échantillon n'appartient pas la classe P_1 et le classifieur le rejette très justement pour le classer en P_2 ... Le résultat est un rejet correct du classifieur donc un vrai négatif (TN pou True Negative).

Ces 4 types de résultats sont généralement résumés dans une matrice dite de confusion (tableau III.3). L'étude de cette dernière permet de mesurer les performances du classifieur. Dans le cas d'un problème multi-classe, on peut toujours se ramener à un problème de 2 classes décrit par une matrice de confusion qui consiste à distinguer une classe particulière de toutes les autres.

Ces performances sont calculées sur l'ensemble des classes en comptant le nombre d'erreurs de classification commises sur un ensemble fini de données. Ce nombre est généralement restreint, car les données sont souvent fastidieuses à collecter et à traiter. Dans notre cas, il correspond à une partie du corpus d'apprentissage nommée corpus test qui n'intervient pas dans la phase d'apprentissage proprement dite (cf chapitre II paragraphe II.3.2.2.2).

Tableau III. 3: Matrice de confusion résumant les 4 possibilités de résultat d'un classifieur chargé de détecter P_1 . Les cases vertes donnent les classifications correctes (TP et TN) et les cases rouges celles erronées (FN et FP).

	P_1 réel	autre classe réelle
P_1 estimé	TP	FP
autre classe estimée	FN	TN

Chaque ligne de la matrice de confusion donne le nombre d'occurrences d'une classe estimée par le classifieur et chaque colonne le nombre d'occurrences d'une classe réelle connue.

Un système de classification est d'autant meilleur que sa matrice de confusion s'approche d'une matrice diagonale. Nous en retenons les 3 critères suivants :

- **Précision** : c'est le rapport entre le nombre de vrais positifs et la somme des vrais positifs et des faux positifs :

$$\text{Précision} = \frac{TP}{TP + FP} \quad \text{III.3}$$

Elle mesure la capacité du classifieur à rejeter les classes différentes de celle considérée (P_1). Si la précision vaut 1, cela veut dire que tous les exemples classés positifs le sont vraiment. Il est relié au degré de confiance qu'on accorde à une acception P_1 .

- **Sensibilité** : Elle mesure la capacité du système à donner tous les bons classements pour la classe considérée (P_1). Si ce dernier vaut 1, cela signifie que tous les exemples positifs sont trouvés donc tous les rejets sont également corrects. Il est relié au degré de confiance qu'on accorde au résultat pour une classe P_1 réelle.

$$S = \frac{TP}{TP + FN} \quad \text{III.4}$$

- **Exactitude** : l'exactitude (accuracy) représente le taux global de bonnes classifications :

$$\text{Exactitude} = \frac{TN + TP}{FP + TN + TP + FN} \quad \text{III.5}$$

III.1.2.2 Classification par LVQ

La détection de couche résiduelle de résine est dans un premier temps abordée par l'utilisation d'un LVQ dans les conditions décrites au paragraphe III.1.1. Nous utilisons un R.N constitué de 10 neurones cachés et 2 neurones de sortie... L'apprentissage a duré environ 3 h sur un ordinateur Intel Xeon CPU E5 1620 @3.60 GHz 12 Go RAM. Le tableau III.4 montre les résultats sous la forme de la matrice de confusion calculée sur le corpus test à l'issue de l'apprentissage. Nous avons ajouté la précision pour chacune des classes sur la dernière colonne, la sensibilité sur la dernière ligne du tableau et l'exactitude.

Tableau III. 4: Matrice de confusion du LVQ (10 neurones cachés) calculée sur le corpus test pour la détection de couche de résine

	P ₁ réels	P ₂ réels	Précision
P ₁ estimés	426	65	86.8 %
P ₂ estimés	24	385	94.1 %
Sensibilité	94.7 %	85.6 %	Exactitude : 90.1%

De façon générale, le classifieur affiche de bons résultats avec une exactitude de 90.1 %. De façon plus détaillée, la précision de la seconde sortie du RN est meilleure que la première avec 94.1 %. Par ailleurs, sur les 450 profils P₁ composant le corpus test, le LVQ en a classé correctement 94.7 %. Ce résultat peut surprendre car les fausses estimations de signatures P₁ dans la classe P₂ n’ont pas été observées lors de la visualisation de la carte SOM des signatures. Ceci provient du fait que le LVQ ne se base pas directement sur la visualisation de la carte SOM mais effectue sa propre projection menant ainsi à d’autres règles de discriminations. La sensibilité sur les profils P₂ est légèrement moins bonne. Ceci s’explique par le fait que certains profils présentant un h_r relativement faible peuvent être faussement classés à cause de leur signature relativement proche de certains profils P₁ (fig. III.5).

Afin de pouvoir comparer avec la carte SOM, nous avons représenté, comme sur la figure III.4, l’épaisseur résiduelle h_r des profils correspondant aux 65 signatures P₂ estimées P₁ (soit 7.2% des échantillons testés) (figure III.18).

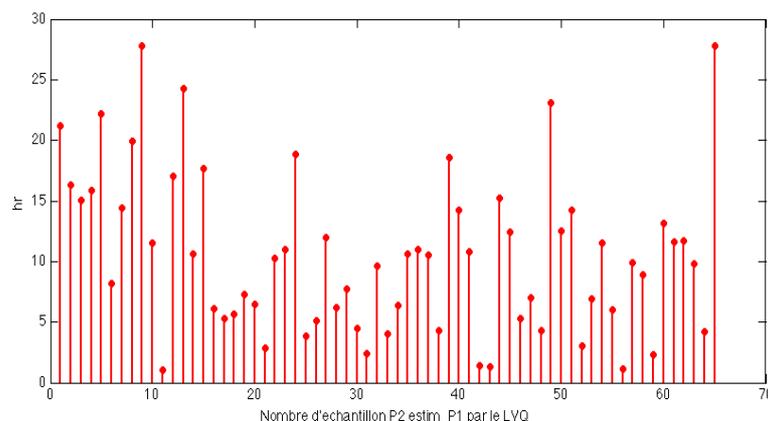


Figure III. 18 : Epaisseurs h_r des 64 profils P₂ identifiés P₁ par le LVQ (10 neurones cachés).

La limite d’épaisseur de résine détectable sans ambiguïté pour le profil P₂ est 28 nm (moyenne d’environ 10.5 nm et écart type de 6.5 nm) contre 11 nm pour la carte SOM.

Afin de confirmer ces constatations, nous avons repris les 3 échantillons précédents du paragraphe III.1.1.3.2 pour réaliser la même étude de sensibilité en fonction de l'épaisseur h_r . La figure III.19 illustre les résultats alors obtenus.

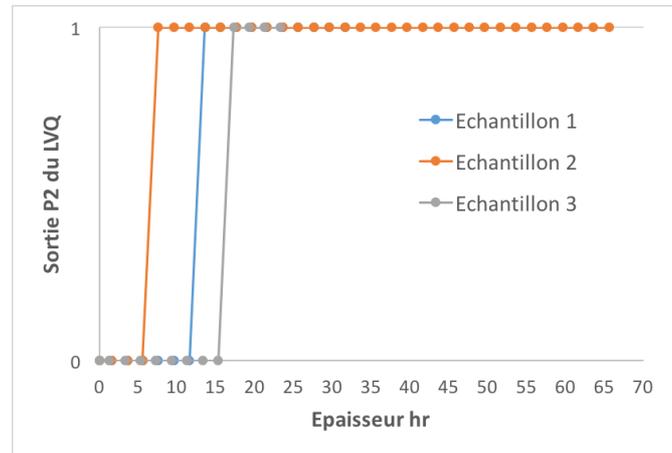


Figure III. 19 : Valeurs de la seconde sortie du LVQ (10 neurones cachés) pour 3 échantillons distincts pour lesquels on réduit artificiellement l'épaisseur h_r

On retrouve le fait que la limite de détection sans ambiguïté dépend de l'échantillon. Cette dernière est comprise entre 5 et 15 nm contre 4 à 11 nm pour la carte SOM 2025 neurones (figure III.14). Il faut également noter que le LVQ fournit en sortie un résultat binaire et ne donne pas une probabilité pouvant aider à la décision finale.

En vue d'optimiser les performances du LVQ nous étudions l'importance du nombre de neurones de la couche cachée. Le tableau III.5 résume les performances obtenues pour chaque architecture présentant respectivement 20, 30, 40, 50 neurones cachés.

Tableau III. 5: Performances du classifieur LVQ en fonction du nombre de neurones cachés pour la détection de couche.

Nombre de neurones cachés	10	20	30	40	50
Nombre de profils P_1 confus	24	8	25	22	31
Nombre de profils P_2 confus	65	73	61	60	59
hr max	28	28	28.5	28	28
Exactitude (en %)	90,1	91	89,7	90.9	90

Les performances du classifieur sont presque constantes et indépendantes du nombre de neurones. L'architecture qui peut être retenue est celle du LVQ constitué de 20 neurones cachés. Cette dernière montre une exactitude légèrement supérieure. Elle permet ainsi d'obtenir le nombre minimal de signatures P_1 estimées P_2 . On se rapproche ainsi des performances de la carte SOM. Par contre, on observe 73 signatures de profils P_2 qui sont estimés P_1 ; ce qui constitue la valeur la plus forte.

Toujours dans le cadre de la comparaison avec la carte SOM, nous pouvons tester le LVQ sur les 14 signatures de la figure III.5 ayant activé le même neurone $r_{(0)}^2$ de la carte 20*20. Les tableaux III.6 et III.7 présentent les résultats obtenus pour P_1 réciproquement P_2 .

Tableau III. 6: Classification par LVQ (20 neurones cachés) des 12 signatures (S_i avec $i=1, \dots, 12$) correspondant à un profil P_1 activant le neurone $r_{(0)}^2$ de la carte SOM 20*20 neurones.

Signatures P_1	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8	S_9	S_{10}	S_{11}	S_{12}
Sortie P_1	1	1	1	1	1	1	1	1	1	1	1	1
Sortie P_2	0	0	0	0	0	0	0	0	0	0	0	0

Tableau III. 7: Classification par LVQ (20 neurones cachés) des 2 signatures S_1 et S_2 correspondant à un profil P_2 ayant activé le neurone $r_{(0)}^2$ de la carte SOM 20*20 neurones.

Signatures P_2	S_1	S_2
Sortie P_1	0	1
Sortie P_2	1	0

Les 12 signatures P_1 sont cette fois-ci classées (tableau III.6) sans ambiguïté. Pour les signatures P_2 , il est difficile de conclure, car une des 2 reste mal classée. Sur ces 14 signatures, le LVQ affiche clairement des résultats différents (ici meilleurs) que la carte SOM. Ceci peut s'expliquer par le fait que ce ne sont pas les mêmes algorithmes régissant la projection.

En conclusion, la classification LVQ affiche des résultats satisfaisants, mais à l'image de la carte SOM, il existe toujours des signatures pour lesquelles le LVQ reste confus. Du fait d'une projection différente, ce ne sont pas les mêmes signatures qui sont en cause. Notons cependant que les résultats relatifs à la carte SOM avec un seuil de 0,5 permettent d'afficher une sensibilité à la classe P_1 et une précision sur P_2 de 100 %. En effet, l'expérimentateur peut avoir une entière confiance à un résultat de la carte estimé P_2 . De la même façon, un profil P_1 sera toujours bien classé. Ceci provient du fait que la SOM possède la particularité d'avoir certains neurones exclusivement sensibles aux signatures de profil P_1 . Par contre, la limite

d'épaisseur de résine détectable sans ambiguïté pour le profil P_2 reste de l'ordre de quelques dizaines de nanomètres (10.5 nm en moyenne pour LVQ et 4.5 nm pour SOM). Il est donc intéressant de procéder à une classification supervisée fonctionnant autrement. Le paragraphe suivant présente l'utilisation du PMC.

III.1.2.3 Classification par PMC

La détection de couche résiduelle de résine est à présent abordée par l'utilisation d'un PMC à une couche cachée toujours dans les conditions décrites au paragraphe III.1.1. L'architecture de ce R.N comprend 5 neurones cachés et 2 neurones de sorties. Les neurones de sorties donnent une estimation de la probabilité d'appartenance aux classes P_1 et P_2 respectivement. L'entraînement a duré 10 seconde sur un ordinateur Intel Xeon CPU E5 1620 @3.60 GHz 12 Go RAM.

Rappelons que, contrairement au LVQ, le PMC ne donne pas directement l'appartenance à la classe, mais une estimation de celle-ci par l'intermédiaire de la probabilité *a posteriori*, pouvant constituer une aide à la décision pour l'expérimentateur. Les sorties du PMC correspondant aux probabilités $p(P_2)$ des 900 signatures tests sont représentées dans la figure III.20.

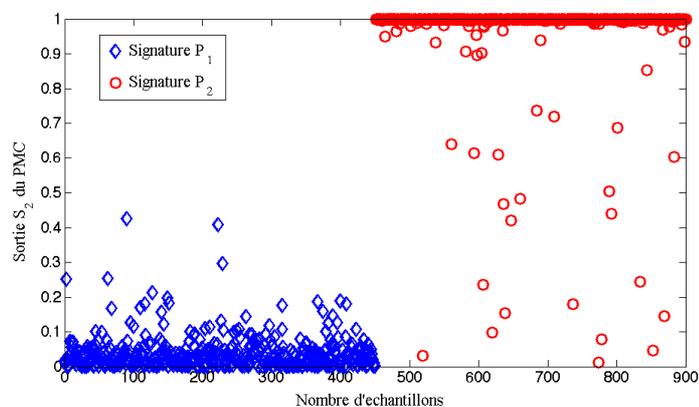


Figure III. 20 : Seconde sortie S_2 du PMC (5 neurones cachés) correspondant à la probabilité *a posteriori* $p(P_2)$ d'appartenance à la classe de profil P_2 calculée sur l'ensemble des 900 signatures du corpus test

Si on fixe un seuil de décision à 0.5, on peut représenter les performances du PMC sur le corpus test après apprentissage sous la forme d'une matrice de confusion (tableau III.8)

Tableau III. 8 : Matrice de confusion du PMC à une couche cachée (5 neurones cachés) calculée sur le corpus test pour la détection de couche de résine.

	P ₁ réelles	P ₂ réelles	Précision
P ₁ estimées	450	14	97 %
P ₂ estimées	0	436	100 %
Sensibilité (%)	100 %	96.9 %	Exactitude : 98.4 %

Les performances du PMC sont excellentes avec une exactitude de 98.4%. Les signatures de type P₁ sont toutes correctement identifiées avec une sensibilité de 100% et on retrouve une précision sur la sortie 2 égale à 100%. Aucune signature P₁ n'a été confondue. Ces résultats sont similaires à ceux identifiés avec la carte SOM (seuil à 0.5). Toutefois, dans le cas du PMC, le nombre de signatures mal classées est fortement réduit à 1.5% contre 13.1% pour la SOM, et 9.8% pour le LVQ. La précision sur S₁ est ainsi fortement améliorée jusqu'à 97%.

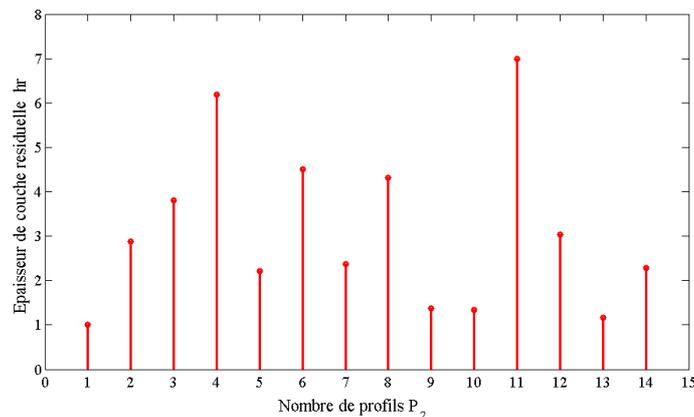


Figure III. 21: Épaisseurs h_r des 14 profils P_2 identifiés P_1 par le PMC à une couche cachée (5 neurones cachés).

On observe que ces profils sont caractérisés par une couche résiduelle h_r en dessous de 7.5 nm (moyenne 3.10 nm 1.86 nm et écart type). Cette épaisseur est largement en dessous de celle observée par le LVQ (28 nm) et la carte SOM (11 nm).

Afin d'améliorer la limite de détection du PMC, nous étudions l'influence du nombre de neurones dans la couche cachée. Le tableau III.9 résume les performances obtenues pour chaque architecture.

Tableau III. 9: Performances du classifieur PMC à une couche cachée en fonction du nombre de neurones N_C cachés pour la détection de couche résiduelle.

Nombre de neurones N_C	5	10	20	30	40	50	60
Nombre de P_2 confus	14	13	13	14	9	19	25
Nombre de P_1 confus	0	2	2	1	0	1	2
hr max	7,5	5,5	5.5	6,5	4,5	7	12
exactitude	98.4	98.3	98.3	98.3	99	97.8	97

Il faut noter que l'ensemble des résultats reste assez proche. Dans certains cas, le nombre de signatures de profils P_1 faussement interprétées devient non nul, mais reste faibles (1 ou 2 signatures). Le PMC avec 40 neurones s'avère meilleur avec une limite de hr de 4.5 nm (moyenne de 2.5 nm).

Pour finir l'étude du PMC dans le cadre de la détection de couche, nous présentons à nouveau les signatures de la figure III.5 (ayant activé le neurone $r_0^{(2)}$ de la carte SOM 20*20 neurones) au RN optimisé ci-dessus. Les résultats sont présentés dans les tableaux III.10 et III.11.

Tableau III. 10: Classification par PMC à une couche cachée (40 neurones) des 12 signatures (S_i avec $i=1, \dots, 12$) correspondant à un profil P_1 activant le neurone $r_0^{(2)}$ de la carte SOM 20*20 neurones

Signatures P_1	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8	S_9	S_{10}	S_{11}	S_{12}
Probabilité P_1	0.91	0.88	0.92	0.96	0.91	0.79	0.90	0.90	0.91	0.96	0.85	0.97
Probabilité P_2	0.09	0.12	0.08	0.04	0.09	0.21	0.11	0.10	0.09	0.04	0.15	0.03

Tableau III. 11 Classification PMC à une couche cachée (40 neurones) pour les 2 signatures S_1 et S_2 correspondant à un profil P_2 activant le neurone $r_0^{(2)}$ de la carte SOM 20*20 neurones

Signatures P_2	S_1	S_2
Probabilité P_1	0.01	0.02
Probabilité P_2	0.99	0.98

Le PMC distingue parfaitement toutes les signatures de classes P_1 et P_2 sans aucune ambiguïté avec de fortes probabilités d'appartenance (supérieures à 79 %). Le PMC confirme

clairement de meilleurs résultats par rapport à la classification réalisée par la carte SOM et également par la classification supervisée par LVQ. La version optimisée à 40 neurones a permis d'améliorer les résultats, mais notons que l'architecture réduite à 5 neurones cachés peut être suffisante suivant la précision demandée.

III.1.2.4 Test sur échantillon réel

La signature mesurée de l'échantillon expérimental décrit dans le paragraphe III.1.1.4 est ensuite présentée aux 2 architectures optimales des 2 types de classifieurs supervisés présentés. Le tableau III.12 décrit les résultats obtenus : la signature est classée dans les 2 cas comme de profil P_2 . Notons que le PMC à 40 neurones cachés affiche une probabilité d'appartenance de 100 %.

Tableau III. 12: Résultats de la classification de la signature expérimentale par le LVQ (20 neurones) et le PMC à une couche cachée (40 neurones)

Résultat LVQ	Résultat PMC
Sortie $P_1=0$	$p(P_1)=0.00$
Sortie $P_2=1$	$p(P_2)=1.00$

Bien que ces résultats doivent être confirmés par d'autres mesures et tests expérimentaux, nous pouvons supposer que les outils mis en place sont susceptibles de répondre correctement aux objectifs fixés de détection systématique.

Dans ce paragraphe, nous avons ainsi montré la visualisation de signatures ellipsométriques à l'aide de cartes SOM pour la mise en évidence d'informations qualitatives sur les différents types de profil. Pour nos tests, nous avons pris l'exemple typique de la détection d'un défaut du masque de résine en lithographie pour fournir un outil pouvant assister la prise de décision. La classification a montré dans ce cas des résultats satisfaisants et encourageants qui doivent être validés définitivement sur de nouvelles mesures expérimentales. Le classifieur PMC s'est avéré meilleur que le LVQ en étant capable de détecter théoriquement avec certitude la présence de couche résiduelle au-dessus de 4.5 nm sur les signatures simulées.

III.2 Identification de forme de profil géométrique pour la construction d'un modèle scatterométrique

Dans ce paragraphe, l'objectif pourrait s'apparenter à de la détection de défaut de modélisation dans le cadre de la caractérisation scatterométrique classique, voire même ellipsométrique simple. Le but de l'étude consiste donc à déterminer le modèle géométrique servant de base lors de la résolution du problème inverse à partir de la signature ellipsométrique utilisée lors de la caractérisation. Mathématiquement, on cherche à supprimer une des hypothèses sous-jacentes, réduisant ainsi le risque de résoudre un problème mal posé. Aucun travail similaire n'a été constaté dans la littérature.

III.2.1 Cadre de l'étude

III.2.1.1 Domaine d'application visée

Le domaine d'applications visées concerne ici le contrôle de l'étape de gravure par la formation du profil souhaité. Prenons le cas de la photolithographie qui constitue la première étape de plusieurs processus de gravure. Le motif ainsi obtenu peut-être plus ou moins proche de la forme finale de résine souhaitée. Par exemple, la figure III.22 montre l'image XSEM d'un profil de période 90 nm réalisé par lithographie par immersion ($\lambda = 193$ nm) sur une couche métallique [140].

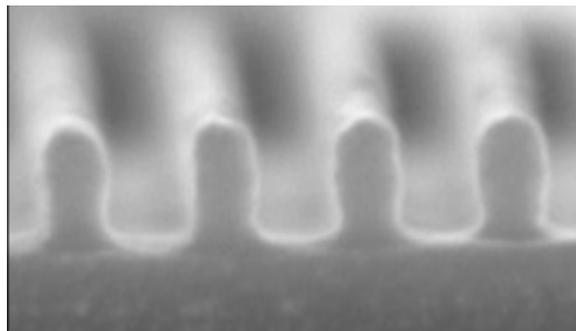


Figure III. 22 : Exemple d'image MEB d'un profil crête de résine présentant des arrondissements aux sommets et aux bases [140].

La forme attendue étant rectangulaire, on remarque une déviation conséquente du motif désiré. En effet lors du processus il survient souvent des formes arrondies au sommet et une perte d'épaisseur de résine par rapport à l'épaisseur initiale (30 % dans ce cas) au cours du

processus. L'effet le plus problématique non souhaité concerne l'apparition d'un profil « en pied de champignon ».

Pour cet échantillon, l'hypothèse du modèle créneau conduirait clairement à un problème mal posé et induirait une caractérisation scattérométrique erronée. L'utilisation d'un modèle tenant compte des arrondis serait bien entendu mieux adaptée au traitement du problème scattérométrique.

III.2.1.2 Définition des différentes classes : les profils géométriques retenus

Les structures périodiques utilisées par la suite dans cette étude sont composées d'un réseau de lignes de période $T=140$ nm réalisé dans la résine IM5010, reposant sur un substrat de Silicium.

Nous retenons quelques formes de profils fréquemment rencontrées en lithographie :

- le créneau parfait (figure III.23.a) caractérisé par 2 paramètres (CD et h) définis pour $h > 0$ et $0 < CD < T$;
- Le profil trapézoïdal (fig. III.23.b) défini par 3 paramètres : b_1 (projection de la pente), b_2 , et h qui doivent vérifier la relation $2b_1 + b_2 < T$ avec $b_1, b_2 > 0$ et $h > 0$;
- Le créneau arrondi au sommet (fig. III.23.c) similaire au premier, mais présentant un arrondi au sommet, il est défini par 3 paramètres : h , CD et r . Le profil existe pour $0 < r < \min(h, \frac{CD}{2})$ et $0 < CD < T$ et $h > 0$;
- Le profil sinusoïdal défini par 2 paramètres b et h (fig. III.23.d). Ce dernier existe pour $0 < b < T$ et $h > 0$;
- Le profil créneau arrondi au bas des sillons, défini par 3 paramètres, dont le rayon du bas r_b , le plateau CD et la hauteur h (fig. III.23.e). Ces derniers doivent vérifier les conditions suivantes : $h > 0$, $0 < r_b \leq h$ et $2 \times r_b + CD \leq T$;
- Le créneau arrondi au bas et au sommet des plots parfaitement défini par 4 paramètres : le rayon de courbure du bas et du haut défini respectivement par r_b et r_h , le plateau CD et la hauteur h (fig. III.23.f). Ces paramètres doivent respecter les conditions $r_b + r_h < h$, $2 * r_h \leq \frac{CD}{2}$ et $2 \times r_b + CD \leq T$ avec r_b, r_h, b et $h > 0$.

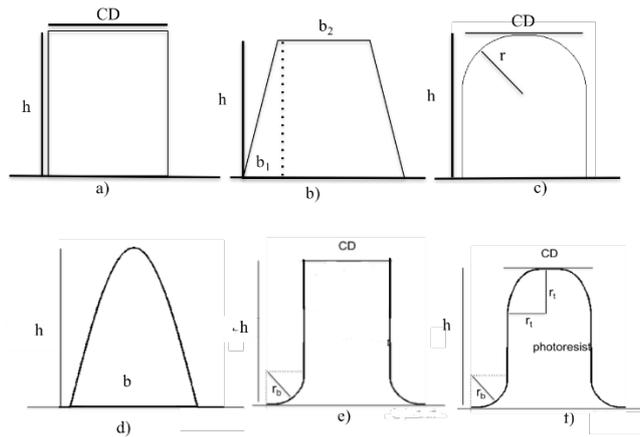


Figure III. 23 : Formes géométriques des profils périodiques retenues dans cette étude : créneau parfait a), trapèze b) créneau arrondi au sommet c), sinus d), créneau à bas arrondi et créneau arrondi au bas et au sommet d)

Le cadre qui nous intéresse ici implique la discrimination de profils de formes géométriques différentes. En termes de classification, chaque classe représente une forme géométrique particulière. Le but de l'étude consiste donc à identifier la forme de profil associée à une signature mesurée. Le R.N est alors utilisé en classifieur dans ce but selon le principe schématisé sur la figure III.24.

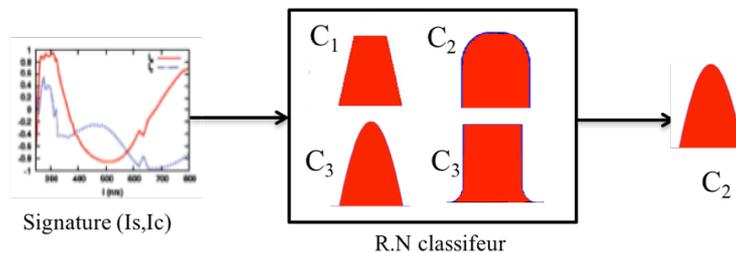


Figure III. 24: Schéma de principe pour l'identification de forme de profil géométrique par l'utilisation d'un RN classifieur

Il faut noter que cette fois-ci, la classification induit la discrimination de plusieurs classes contrairement à l'étude précédente qui consistait seulement à identifier ou non la présence d'un défaut.

III.2.2 Etude du défaut de modélisation

L'étude dans ce paragraphe consiste à simuler l'hypothèse d'un faux profil dans le modèle de traitement de la signature par scattérométrie. Dans le but de montrer l'importance de la connaissance de la forme du profil dans un processus de caractérisation scattérométrique,

nous testons 4 modèles de PMC dédiés à une caractérisation scattérométrique classique fonctionnant chacun exclusivement sur un seul type de profil géométrique pour le traitement du problème inverse.

Ainsi, dans ce qui suit le modèle de R.N noté PMC_a est associé uniquement au profil créneau parfait (fig. III.23.a), celui noté PMC_b exclusivement aux profils de forme trapézoïdale (fig. III.23.b), le PMC_c aux profils de forme créneau arrondi aux sommets (fig. III.23.c) et enfin le PMC_d aux profils de forme sinus (fig. III.23.d).

Toutes les signatures sont simulées par MMFE pour une incidence $\theta=62^\circ$ et un spectre de longueur d'onde compris entre 1.5 et 6.5 eV (suivant un pas de 0.1 soit 102 valeurs). Un bruit correspondant à celui estimé lors de la mesure ($\sigma=0.003$) est également pris en compte.

Afin de ne pas alourdir le paragraphe, l'apprentissage sur 3000 échantillons de chacun des R.N n'est pas décrit, mais est bien entendu optimisé conformément à la méthode du early stopping mentionnée au chapitre II paragraphe II.3.2.2.2 ($N_{entr}=2100$ échantillons, $N_{val}=450$ échantillons et $N_{test}=450$ échantillons). La qualité de l'apprentissage d'un modèle de PMC est évaluée par le calcul de l'erreur quadratique moyenne réalisée sur chaque paramètre géométrique p associé au profil notée E_{rms}^p conformément à l'équation II.66 (cf. chapitre II paragraphe II.3.2.2.2)

Le critère utilisé permettant de comparer les performances des caractérisations est l'erreur quadratique moyenne E_{IsIc} calculée entre les signatures I_{recaI} du profil estimé par le R.N et celle correspondante à la mesure I_{mes} :

$$E_{IsIc} = \sqrt{\frac{1}{M} [I_{recaI} - I_{mes}]^2} \quad \text{III.6}$$

avec M le nombre de mesures (Is et Ic) composant la signature ellipsométrique (ici 102 valeurs).

Enfin, notons que l'erreur minimale de mesure sur Is et Ic est d'environ 0,01 (voir chapitre I paragraphe I.2.2.2). Elle servira de référence pour la suite et sera notée E_{ref} .

III.2.2.1 Hypothèse du créneau parfait

Commençons par utiliser l'architecture la plus simple (PMC_a) constituée de 10 neurones cachés et 2 neurones de sorties correspondant aux valeurs des paramètres CD et h du profil créneau parfait (cf figure 23.a). Le R.N est entraîné avec 3000 signatures de type créneau dans les plages suivantes :

$$30 \text{ nm} < \text{CD} < 90 \text{ nm}$$

$$40 \text{ nm} < h < 130 \text{ nm}$$

La figure III.25 montre les sorties estimées (Output) en fonction des sorties désirées (Target) ainsi que la courbe de régression correspondante pour les paramètres CD et h normalisés de l'ensemble du corpus test à l'issue de l'apprentissage (Cf chapitre II paragraphe II.3.2.2.2).

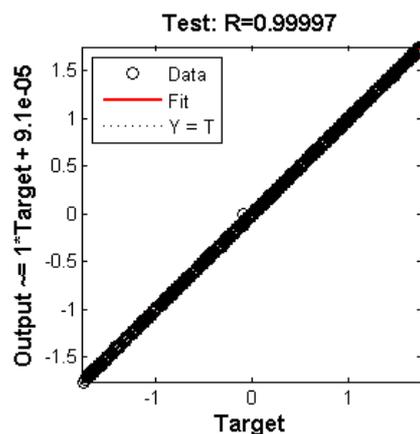


Figure III. 25 : Performance du PMC_a à l'issue de l'apprentissage sur les paramètres CD et h normalisés : courbe de régression évaluée sur N_{test} échantillons

Les erreurs théoriques calculées pour les paramètres CD et h normalisés sont largement inférieures au nanomètre : $E_{rms}^{CD} = 0.093 \text{ nm}$ et $E_{rms}^h = 0.091 \text{ nm}$ et le coefficient R est proche de 1.

Pour illustrer les bonnes performances théoriques du PMC_a le tableau III.13 montre les résultats obtenus pour 4 échantillons pris au hasard parmi le corpus test. On pourra supposer que les erreurs E_{rms} constituent une bonne approximation des barres d'erreurs moyennes sur les valeurs de chaque paramètre calculé par le RN :

$$CD_{est} \cong CD \pm E_{rms}^{CD} \text{ et } h_{est} \cong h \pm E_{rms}^h$$

Tableau III. 13 : Caractérisation par PMC_a (entraîné sur un modèle créneau) de 4 échantillons de type forme de profil créneau

Profil Créneau	Estimation PMC_a
Exemple 1 CD=69.8 nm, h=41.2 nm	$CD_{est}=69.9\pm 0.093$ nm, $h=41.3\pm 0.091$ nm $E_{Is,lc}=0.0003$
Exemple 2 CD=37.5 nm, h=101.5 nm	$CD_{est}=37.6\pm 0.093$ nm, $h=101.6\pm 0.091$ nm $E_{Is,lc}=0.002$
Exemple 3 CD=68.0 nm, h=65.3 nm	$CD_{est}=68.1\pm 0.093$ nm, $h=65.3\pm 0.091$ nm $E_{Is,lc}=0.0004$
Exemple 4 CD=54 nm, h=100.6 nm	$CD_{est}=54\pm 0.093$ nm, $h=100.6\pm 0.091$ nm $E_{Is,lc}=0.0003$

Les valeurs de CD et h estimées sont toutes comprises dans les plages d'apprentissage et très proches des valeurs des paramètres du profil réel. L'erreur sur les signatures reconstruites $E_{Is,lc}$ est toujours très proche de 0 et donc largement inférieure à E_{ref} . Dans ce cas, les paramètres estimés sont bien cohérents avec ceux attendus. Le modèle supposé est le modèle réel.

Nous considérons maintenant le cas où la signature fournie au R.N ne correspond pas au modèle utilisé pour l'apprentissage de ce dernier. Des signatures simulées à partir de différents motifs géométriques sont fournies au PMC_a . Pour ne pas alourdir la lecture, nous présentons juste quelques cas particuliers parmi les différents motifs géométriques définis précédemment (figure III.23).

Ainsi 4 exemples pris au hasard de signatures simulées de profil trapézoïdal sont dans un premier temps fournis au PMC_a . Les résultats sont regroupés dans le tableau III.14 en tenant compte des barres d'erreur.

Tableau III. 14 : Caractérisation par PMC_a (entraîné sur un profil créneau) de 4 échantillons de forme de profil trapézoïdal

Profil Trapèze	Estimation PMC_a
Exemple 1 : b ₁ =4.1 nm, CD=36.6 nm, h=101.7 nm	CD _{est} =39.7 ±0.093 nm, h _{est} =101.2 ±0.091 nm, E _{Is, Ic} =0.043
Exemple 2 : b ₁ =10.2 nm, CD=45.2 nm h=61.2 nm	CD _{est} =54.2 ±0.093 ; h _{est} =60.2 ±0.091 nm E _{Is, Ic} =0.0325
Exemple 3 : b ₁ =14.6 nm, CD=50.5 nm, h=117.5 nm	CD _{est} =62.2 ±0.093 nm, h _{est} =110.7 ±0.091 nm, E _{Is, Ic} =0.064
Exemple 4 : b ₁ =16.9 nm, CD=81.8 nm, h=91.5 nm	CD _{est} =91 ±0.093 nm, h _{est} =81.1 ±0.091 nm, E _{Is, Ic} =0.0708

Le RN estime les 2 paramètres (CD_{est} et h_{est}) correspondant au modèle d'apprentissage. Ainsi, comme on pouvait s'y attendre, les résultats sont moins bons ; et ce d'autant plus que le profil réel s'écarte du modèle supposé. La valeur estimée du CD peut même parfois dépasser les plages d'intervalle utilisées pour l'apprentissage du R.N notamment pour des valeurs élevées du paramètre b₁ du profil réel (exemple 4). Dans ce cas, l'erreur E_{Is, Ic} peut atteindre jusqu'à 7 fois la valeur de E_{ref}. A l'inverse, le paramètre h ne semble pas ou peu affecté par le mauvais profil supposé et la valeur estimée par le PMC_a est relativement proche de celle désirée. Il est difficile de comparer les paramètres qui ne représentent pas les mêmes motifs, mais on peut remarquer que la valeur du paramètre b₁ semble compensée préférentiellement par la valeur du CD de la façon suivante :

$$CD + b_1 = CD_{est}$$

Il est clair que les barres d'erreurs estimées sur les paramètres ne correspondent plus à la réalité de la caractérisation et ne peuvent être efficacement prises en compte.

Pour donner une représentation plus visuelle des résultats, la figure III.26 représente le profil géométrique recherché et le profil estimé par le RN.

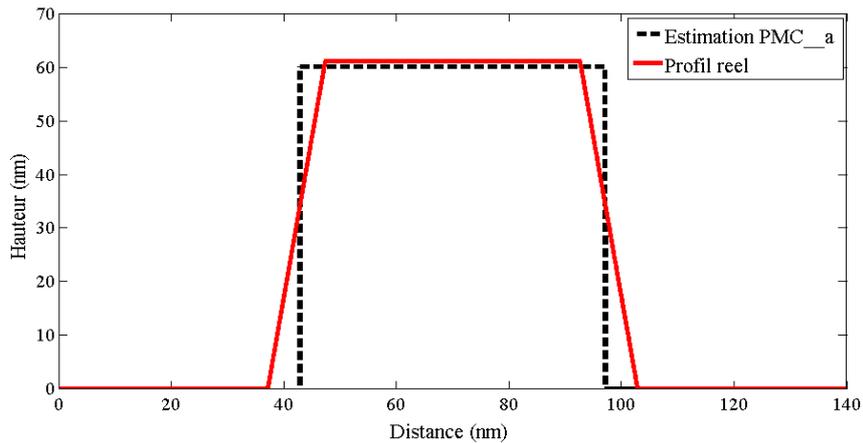


Figure III. 26 : Représentation graphique du profil trapèze de l'exemple 2 du tableau III.14 ($b_1=10.2$ nm, $CD=45.2$ nm $h=61.2$ nm) (rouge) et du profil estimé par le PMC_a (noir) ($CD_{est}=54.2 \pm 0.093$, $h_{est}=60.2 \pm 0.091$, $E_{Is,lc}=0.0325$)

Ensuite, on réalise la même opération, mais avec des signatures de profils créneau arrondi au sommet. Les résultats pour 4 profils pris au hasard sont présentés dans le tableau III.15 et la figure III.27 illustre géométriquement le résultat concernant l'exemple 4.

Tableau III. 15 : Caractérisation par PMC_a (entraîné sur un profil créneau) de 4 échantillons de forme de profil de type créneau arrondi

Profil Créneau Arrondi	Estimation PMC_a
Exemple 1 $r=6.1$ nm, $CD=69.6$ nm, $h=117.5$	$CD_{est}=69.4 \pm 0.093$ nm, $h_{est}=117.1 \pm 0.091$ nm, $E_{Is,lc}=0.003$
Exemple 2 $r=18.7$ nm, $CD=42.2$ nm, $h=84.2$ nm	$CD_{est}=41.3 \pm 0.093$ nm, $h_{est}=82 \pm 0.09$ nm, $E_{Is,lc}=0.03$
Exemple 3 $r=21.5$ nm, $CD=87.5$ nm, $h=57.3$ nm	$CD_{est}=84.8 \pm 0.093$ nm, $h_{est}=55.9 \pm 0.091$ nm, $E_{Is,lc}=0.01$
Exemple 4 $r=26.6$ nm, $CD=53.8$ nm, $h=77.6$ nm	$CD_{est}=51.7 \pm 0.093$ nm, $h_{est}=74 \pm 0.091$ nm, $E_{Is,lc}=0.03$

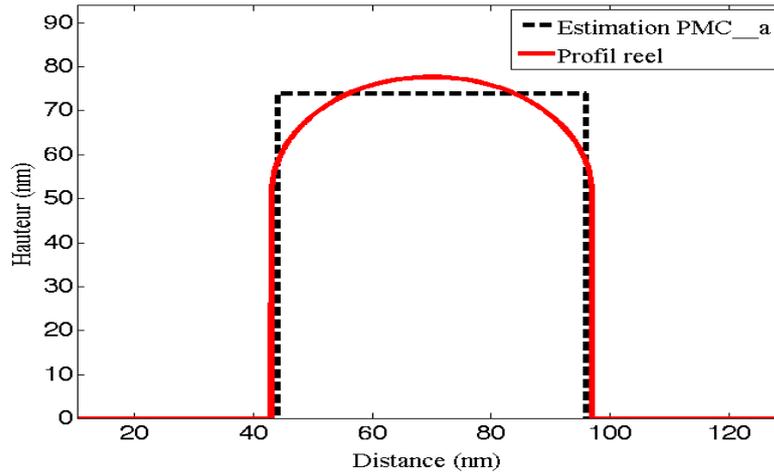


Figure III. 27 : Représentation graphique du profil crénneau arrondi de l'exemple 4 du tableau ($r=26.6$ nm, $CD=53.8$ nm, $h=77.6$ nm) (rouge) et du profil estimé ($CD_{est}=51.7 \pm 0.093$ nm, $h_{est}=74 \pm 0.091$) par PMC_a (noir)

Bien que le modèle de profil soit erroné, les valeurs de CD et h estimés sont cette fois – ci assez proches des valeurs cibles et ce, d'autant plus que la valeur du paramètre r est faible. Dans la plupart des cas, E_{Islc} calculée est plus élevée que E_{ref} mais reste acceptable comparée à l'étude précédente. L'arrondi n'apparaissant pas dans le modèle, le PMC_a ajuste le CD_{est} et la hauteur de sorte à approcher un crénneau parfait sur lequel il a été entraîné. Ainsi, il semble que l'arrondi au sommet soit un paramètre de second ordre affectant peu la signature optique. Pour s'en assurer, nous calculons les signatures de chaque profil du tableau III.15 en faisant varier uniquement le rayon d'arrondi jusqu'à sa valeur maximale (cf. paragraphe III.2.1.2). Ces dernières sont fournies au PMC_a et les résultats sous la forme du calcul de l'erreur E_{Islc} sont représentés sur la figure III.28.

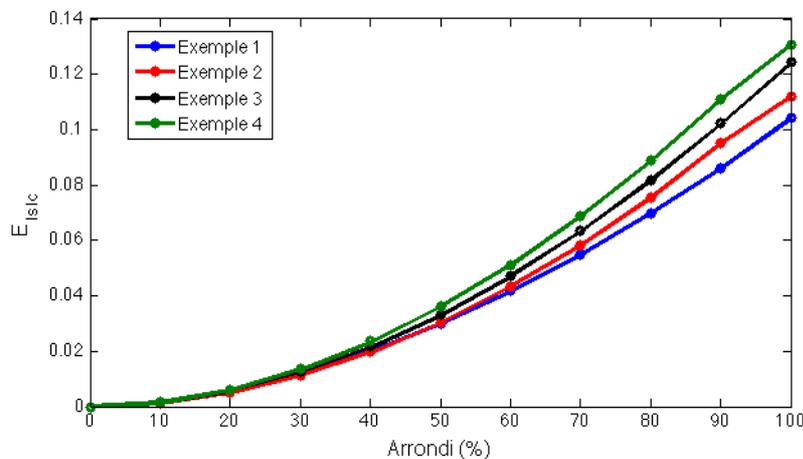


Figure III. 28 : Erreur sur les signatures reconstruites E_{Islc} à partir des résultats du PMC_a pour les 4 échantillons du tableau III.12 en fonction de la valeur d'arrondi sur le profil.

On remarque que l'effet est négligeable sur la signature puisque $E_{Is,lc}$ pour les faibles valeurs d'arrondi (30 %) est en dessous de E_{ref} . Au – delà, l'erreur croît presque linéairement. A la valeur maximale du rayon de courbure, l'erreur est environ 10 fois celle de E_{ref} . Géométriquement, il faut donc un arrondi important c'est à - dire une absence de plateau au-dessus des créneaux pour que l'erreur soit significative. Le paramètre r peut ainsi être qualifié de paramètre de second ordre.

Pour finir, des signatures correspondantes à des profils sinus sont fournies au PMC_a . Les résultats pour 4 profils distincts sont présentés dans le tableau III.16 et la figure III.29 illustre géométriquement le résultat concernant l'exemple 1.

Tableau III. 16 : Caractérisation par PMC_a (entraîné sur un profil créneau) de 4 échantillons de forme de profil sinusoïdal

Profil sinus	Estimation PMC_a
Exemple 1 : $b=76.9\text{nm}$; $h=86.5\text{nm}$	$CD_{est}=50.4\pm 0.093$; $h_{est}=74.4\pm 0.091$; $E_{Is,lc}=0.20$
Exemple 2 : $b=50.9\text{nm}$; $h=99.3\text{nm}$	$CD_{est}=22.9\pm 0.093$; $h_{est}=102.6\pm 0.091$; $E_{Is,lc}=0.72$
Exemple 3 : $b=86.9\text{ nm}$, $h=43.2\text{ nm}$	$CD_{est}=59.1\pm 0.093$; $h_{est}=37.9\pm 0.091$; $E_{Is,lc}=0.42$
Exemple 4 : $b=52.6\text{ nm}$, $h=75.7\text{ nm}$	$CD_{est}=30.7\pm 0.093$; $h_{est}=71.7\pm 0.091$; $E_{Is,lc}=0.75$

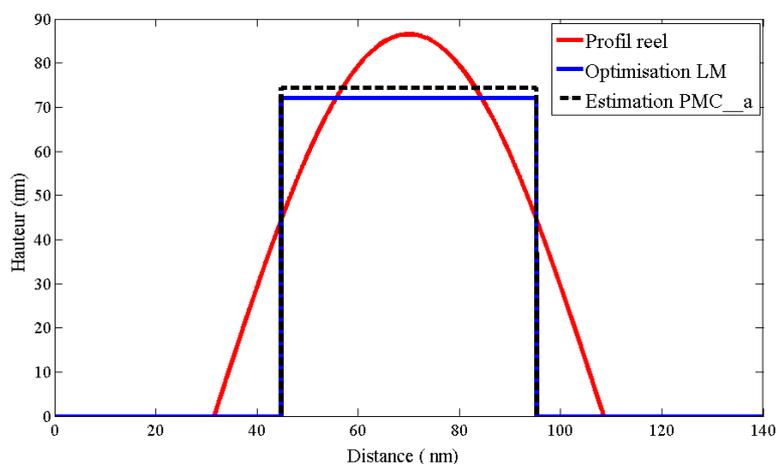


Figure III. 29 : Représentation graphique du profil sinus de l'exemple 1 ($b=76.9$ $h=86.5$) du tableau III.16 (rouge), du profil estimé ($CD_{est}=50.4 \pm 0.093$, $h_{est}=74.4 \pm 0.091$, $E_{Is,lc}=0.20$) par PMC_a (noir) et du profil estimé par une optimisation géométrique de type Levenberg Marquart (bleu)

Cette fois-ci, les paramètres estimés conduisent à un profil fortement éloigné de l'échantillon cible : l'erreur E_{Islc} est conséquente. Le RN tente d'ajuster au mieux les 2 paramètres CD et h de son modèle pour approcher au mieux le profil sinusoïdal. Dans ce cas, le résultat semble se rapprocher du profil obtenu suite à une optimisation géométrique classique (figure III.29) réalisée directement sur la forme géométrique.

Dans certains cas (exemple 2 et 3) la valeur des paramètres estimés (respectivement CD et h) sont en dehors du domaine défini lors de l'apprentissage. Ceci peut constituer une méthode pour détecter un défaut de modélisation. Sinon, sans connaissance a priori sur la forme du profil réel, la seule information disponible est l'erreur E_{Islc} .

III.2.2.2 Hypothèse du modèle trapézoïdal

Le second modèle (PMC_b) présente une architecture de 10 neurones cachés et 3 neurones de sortie correspondant aux paramètres définissant le profil trapèze. Rappelons qu'il s'agit du profil le plus utilisé en scattérométrie. Le R.N est entraîné avec 3000 signatures correspondant à des échantillons dont les valeurs des paramètres géométriques sont réparties dans les plages suivantes :

$$\begin{aligned}0 \text{ nm} < b_1 < 20 \text{ nm}, \\30 \text{ nm} < \text{CD} < 90 \text{ nm}, \\40 \text{ nm} < h < 130 \text{ nm}.\end{aligned}$$

De la même façon que pour le PMC_a (figure III.25), on a représenté sur la figure III.30 les résultats d'apprentissage sous la forme de la courbe de régression globale obtenue sur les données de test normalisées. Le coefficient de régression ($R=0.99652$) confirme la qualité de l'apprentissage. Les erreurs théoriques calculées sur les 3 paramètres b_1 , b_2 et h sont : $E_{\text{rms}}^{b_1} = 1.1 \text{ nm}$, $E_{\text{rms}}^{b_2} = 1.3 \text{ nm}$ et $E_{\text{rms}}^h = 0.6 \text{ nm}$. Notons que ces performances théoriques sont moins bonnes que le PMC_a mais restent relativement correctes (ordre de grandeur 1 nm) et suffisantes. Ceci s'explique par l'utilisation d'un modèle légèrement plus complexe (un paramètre supplémentaire).

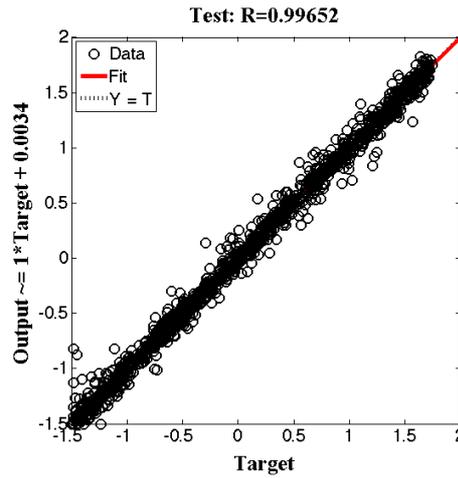


Figure III. 30 : Performance du PMC_b à l'issue de l'apprentissage sur les paramètres b_1 , CD et h normalisés : courbe de régression évaluée sur N_{test} échantillons

Le tableau III.17 confirme la qualité des résultats sur 4 exemples pris au hasard dans la base de données test.

Tableau III. 17 : Caractérisation par PMC_b (entraîné sur un profil trapézoïdal) de 4 échantillons de forme de profil sinusoïdal

Profil Trapèze	Estimation PMC_b
Exemple 1 $b_1=4.1$ nm, $b_2=36.6$ nm, $h=101.7$ nm	$b_{1est}=4.3 \pm 1.1$ nm, $b_{2est}=36.2 \pm 1.3$ nm, $h_{est}=101.8 \pm 0.6$ nm $E_{Islc}= 0.005$
Exemple 2 $b_1=10.2$ nm, $b_2=45.2$ nm $h=61.2$ nm	$b_{1est}=10.4 \pm 1.1$ nm, $b_{2est}=45 \pm 1.3$ nm, $h_{est}=61.3 \pm 0.6$ nm, $E_{Islc}=0.001$
Exemple 3 $b_1=14.6$ nm, $b_2=50.5$ nm, $h=117.5$ nm	$b_{1est}=14.7 \pm 1.1$ nm, $b_{2est}=50.4 \pm 1.3$ nm, $h_{est}=117.6 \pm 0.6$ nm $E_{Islc}=0.001$
Exemple 4 $b_1=16.9$ nm, $b_2=81.8$ nm, $h=91.5$ nm	$b_{1est}=17 \pm 1.1$ nm, $b_{2est}=81.3 \pm 1.3$ nm, $h_{est}=91.6 \pm 0.6$ nm, $E_{Islc} =0.003$

Les paramètres estimés sont bien cohérents avec ceux attendus :

$$b_1 \cong b_{1est} \pm E_{rms}^{b_1} ; b_2 \cong b_{2est} \pm E_{rms}^{b_2} \text{ et } h \cong h_{est} \pm E_{rms}^h.$$

Tous les paramètres estimés sont dans les intervalles utilisés pour l'apprentissage et E_{Islc} reste très en dessous de E_{ref} . Tout concorde pour valider la procédure d'apprentissage.

Pour tester le PMC_b dans le cadre d'une mauvaise connaissance du profil *a priori*, nous avons repris les 3 cas de figure précédemment utilisés pour le test du PMC_a (tableau II.13) ; à savoir 4 profils de forme créneau, puis créneau arrondi puis sinusoïdale.

Le tableau III.18 montre les résultats de caractérisation pour les 4 profils de type créneau et la figure III.31 illustre le résultat relatif à l'échantillon 1.

Tableau III. 18: Caractérisation par PMC_b (entraîné sur un profil trapézoïdal).de 4 échantillons forme de profil créneau

Profil Créneau	Estimation PMC _b
Exemple 1 CD=69.8 nm , h=41.2 nm	$b_{1est}=0.2 \pm 1.1$ nm , $b_{2est}=69.4 \pm 1.3$ nm , $h_{est}=41.2 \pm 0.6$ nm, $E_{Islc}=0.002$
Exemple 2 CD=37.5 nm, h=101.7 nm	$b_{1est}=-0.5 \pm 1.1$ nm , $b_{2est}=37.1 \pm 1.3$ nm , $h_{est}=102 \pm 0.6$ nm $E_{Islc}=0.005$
Exemple 3 CD=68.0 nm, h=65.3 nm	$b_{1est}=-0.1 \pm 1.1$ nm , $b_{2est}=68.2 \pm 1.3$ nm , $h_{est}=65 \pm 0.6$ nm $E_{Islc}=0.0063$
Exemple 4 CD=54 nm, h=100.6 nm	$b_{1est}=-0.1 \pm 1.1$ nm , $b_{2est}=54 \pm 1.3$ nm , $h_{est}=100.7 \pm 0.6$ nm, $E_{Islc} 0.003$

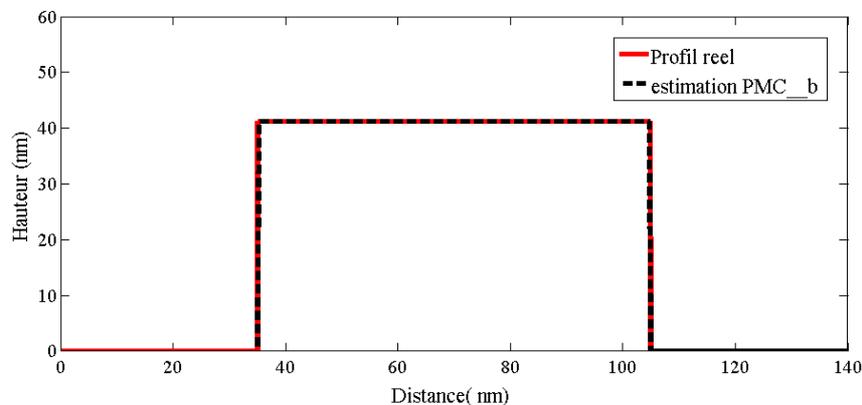


Figure III. 31: Représentation graphique du profil créneau (CD=69.8 nm , h=41.2 nm) de l'exemple 1 tableau III.18 (rouge) et du profil estimé ($b_{1est}=0.2 \pm 1.1$ nm , $b_{2est}=69.4 \pm 1.3$ nm , $h_{est}=41.2 \pm 0.6$ nm, $E_{Islc}=0.002$) par PMC_b (noir)

L'estimation des valeurs des paramètres CD et h est très proche de celles espérées. La valeur de E_{Islc} calculée est également inférieure à E_{ref} . Le modèle PMC_b peut donc caractériser efficacement des profils de type créneau sans aucune ambiguïté. Ceci provient bien entendu du fait que le modèle géométrique du trapèze contient intrinsèquement la forme créneau ($b_1 = 0$). Cette remarque n'est pas anodine, car on pourrait penser qu'il suffit de prendre un modèle suffisamment complexe qui inclut une bibliothèque de profils dégénérés.

Notons cependant que E_{Isc} calculée est bien supérieure à celle réalisée avec le PMC_a (10 fois plus en moyenne). Dans le cas présenté ici, cela n'a que peu d'incidence, car les performances en question restent très bonnes, mais lorsque le nombre de paramètres en jeu dans le modèle devient important cette remarque prend tout son sens.

Concernant les signatures correspondant à des créneaux arrondis au sommet, le comportement du PMC_b est similaire à celui observé sur le PMC_a. Le tableau III.19 résume les résultats pour les 4 profils à créneau arrondi et la figure III.32 illustre le résultat de l'exemple 4.

Tableau III. 19 : Caractérisation par PMC_b (entraîné sur un profil de type trapézoïdal) pour 4 échantillons particuliers de forme de profil de type créneau arrondi

Profil Créneau arrondi	Estimation PMC_b
Exemple 1 $r=6.1$ nm, $CD=69.6$ nm, $h=117.5$ nm	$b_{1est}=0.5 \pm 1.1$ nm, $b_{2est}=69.2 \pm 1.3$ nm, $h_{est}=117.1 \pm 0.6$ nm, $E_{Isc}=0.003$
Exemple 2 $r=18.7$ nm, $CD=42.2$ nm, $h=84.2$ nm	$b_{1est}=1.3 \pm 1.1$ nm, $b_{2est}=39.8 \pm 1.3$ nm, $h_{est}=81 \pm 0.6$ nm, $E_{Isc}=0.028$
Exemple 3 $r=21.5$ nm, $CD=87.5$ nm, $h=57.3$ nm	$b_{1est}=7.7 \pm 1.1$ nm, $b_{2est}=78.4 \pm 1.3$ nm, $h_{est}=56.6 \pm 0.6$ nm, $E_{Isc}=0.013$
Exemple 4 $r=26.6$ nm, $CD=53.8$ nm, $h=77.6$ nm	$b_{1est}=6.5 \pm 1.1$ nm, $b_{2est}=45.9 \pm 1.3$ nm, $h_{est}=74.1 \pm 0.6$ nm, $E_{Isc}=0.029$

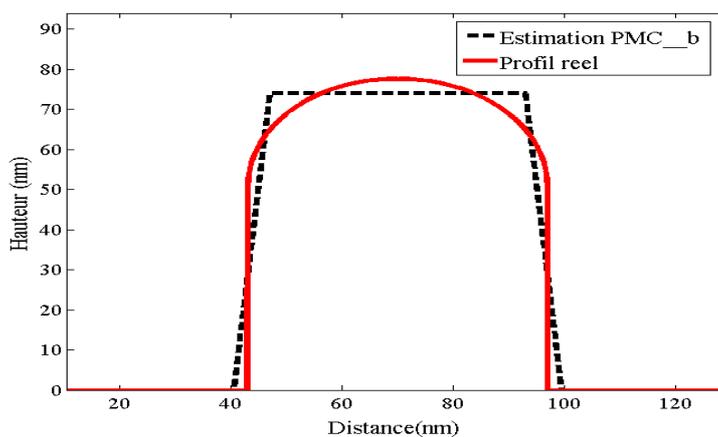


Figure III. 32 : Représentation graphique du profil créneau arrondi de l'exemple 4 ($r=26.6$ nm, $CD=53.8$ nm, $h=77.6$ nm) du tableau III.19 (rouge) et du profil estimé ($b_1=6.5 \pm 1.1$ nm, $b_2=45.9 \pm 1.3$ nm, $h=74.1 \pm 0.6$ nm, $E_{Isc}=0.029$) par PMC_b (noir)

Les résultats sont tous inclus dans les plages de variation des paramètres définie lors de

la phase d'apprentissage. On remarque que $b_{1est} + b_{2est} = CD$. Le RN semble compenser l'arrondi par la création d'une pente b_{1est} non nulle, mais également une réduction du paramètre b_{2est} . Les résultats concernant le paramètre h sont raisonnablement proches de ceux attendus. L'arrondi caractérisé par r apparaît, là encore, comme un paramètre de second ordre. Les valeurs des erreurs E_{Istc} calculées dépassent E_{ref} notamment lorsque l'arrondi est plus prononcé. La précision du réseau est faible par rapport au PMC_a .

Pour finir, les signatures de profils sinus sont présentées au PMC_b . Les résultats sont regroupés dans le tableau III.20 et la figure III.33 pour l'échantillon 4.

Tableau III. 20 : Caractérisation par le PMC_b (entraîné sur un profil de type trapézoïdal) pour 4 échantillons particuliers de forme de profil sinusoidal (gauche)

Profil sinus	Estimation PMC_b
Exemple 1 $b=76.9$ nm, $h=86.5$ nm	$b_{1est}=25.6 \pm 1.1$ nm, $b_{2est}=25.3 \pm 1.3$ nm, $h_{est}=79.6 \pm 0.6$ nm (b_1 et b_2 hors plage)
Exemple 2 $b=50.9$ nm, $h=99.3$ nm	$b_{1est}=27.8 \pm 1.1$ nm, $b_{2est}=11.9 \pm 1.3$ nm, $h_{est}=87.8 \pm 0.6$ nm (b_{1est} et b_{2est} hors plage)
Exemple 3 $b=86.9$ nm, $h=43.2$ nm	$b_{1est}=27.1 \pm 1.1$ nm, $b_{2est}=35.4 \pm 1.3$ nm, $h_{est}=39.9 \pm 0.6$ nm (b_1 et b_2 hors plage)
Exemple 4 $b=52.6$ nm, $h=75.7$ nm	$b_{1est}=16.3 \pm 1.1$ nm, $b_{2est}=17.6 \pm 1.3$ nm, $h_{est}=71.9 \pm 0.6$ nm (b_{2est} hors plage)

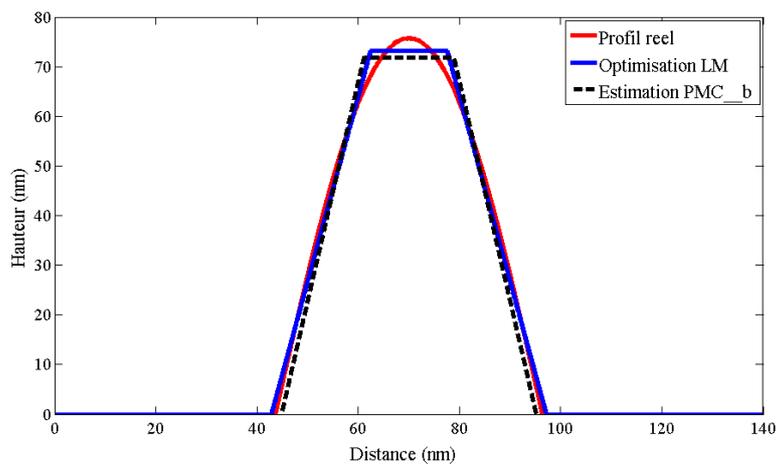


Figure III. 33 : Représentation graphique du profil de l'exemple 4 ($b=52.6$ nm, $h=75.7$ nm) du tableau III.20 (rouge), du profil estimé ($b_1=16.3 \pm 1.1$ nm, $b_2=17.6 \pm 1.3$ nm, $h=71.9 \pm 0.6$ nm) caractérisé par PMC_b et du profil estimé par optimisation Levenberg Marquardt (bleu)

Dans tous les cas de figure présentés ici, un (exemple 4) ou plusieurs paramètres (exemples 1, 2 et 3) sont estimés en dehors des plages utilisées pour l'apprentissage. Les erreurs E_{IsIc} n'ont ainsi pas été calculées. Dans ce cas, il est facile de détecter un problème mal posé. Par contre, aucune indication ne nous permet de statuer sur un défaut de forme ou un défaut de plage de variation pour l'entraînement.

Notons que sur la figure III.33, nous pouvons remarquer que le PMC_b tend à définir une forme trapézoïdale qui « accroche » les pentes des flancs du profil sinus. Cela induit un paramètre b_1 hors des plages de variation. De la même façon, l'absence de plateau sur la partie haute du profil réel conduit à une estimation b_2 relativement faible soit en dessous des plages d'apprentissage. Un apprentissage sur des plages plus larges peut être envisagé. Il conduirait à tendre vers le profil obtenu par optimisation géométrique (courbe bleue).

Afin de ne pas alourdir la lecture, les études relatives aux PMC_c et PMC_d sont décrites en annexe A.

III.2.2.3 Simulation d'un profil test complexe

Dans ce paragraphe nous présentons le cas d'un profil non décrit par la base de données d'apprentissage des 4 PMC précédents. Le profil à analyser présente à la fois un arrondi en bas des sillons et au sommet des plots (figure III.23.f). Il peut être considéré comme un modèle imparfait présentant 2 déformations si on se place dans le cas d'un échantillon cible de forme rectangulaire.

Comme précédemment, nous analysons le résultat concernant 3 profils pris au hasard (tableaux III.21 à III.23). Les modèles géométriques utilisés pour les caractérisations sont respectivement rectangulaires, arrondis au sommet et trapézoïdaux (PMC_a, PMC_b, PMC_c).

Tableau III. 21 : Caractérisation par PMC_a (entraîné sur un profil de type créneau) de 3 échantillons de forme de profil de type créneau présentant des arrondis en haut et bas des plots

Profil créneau haut et bas arrondi	Estimation PMC_a
Exemple 1 : $r_h=7.3$ nm, $CD=80.5$ nm, $h=90.8$ nm, $r_b=2.4$ nm	$CD_{est}=82\pm 0.093$ nm, $h_{est}=94.3\pm 0.091$ nm, $E_{IsIc}=0.0421$
Exemple 2 : $r_h=8$ nm, $CD=86.9$ nm, $h=50.3$ nm, $r_b=4.5$ nm	$CD_{est}=86.2\pm 0.093$ nm, $h_{est}=57.6\pm 0.091$ nm, $E_{IsIc}=0.1009$
Exemple 3: $r_h=18$ nm, $CD=45.3$ nm, $h=50.3$, $r_b=11.6$ nm	$CD_{est}=48.2\pm 0.093$ nm, $h_{est}=116.2\pm 0.091$ nm, $E_{IsIc}=0.5807$

Tableau III. 22 : Caractérisation par PMC_b (entraîné sur un profil de type trapézoïdal) pour 3 échantillons de forme de profil de type créneau présentant des arrondis en haut et bas des plots.

Profil créneau haut et bas arrondi	Estimation PMC_b
Exemple 1 $R_h=7.3$ nm, $CD=80.5$ nm, $h=90.8$ nm, $r_b=2.4$ nm	$b_{1est}=16.2\pm 1.1$ nm, $CD_{est}=72.6\pm 1.3$ nm, $h_{est}=92\pm 0.6$ nm, $E_{IsIc}=0.09$
Exemple 2 $r_h=8$ nm, $CD=86.9$, $h=50.3$ nm, $r_b=4.5$ nm	$b_{1est}=31.3\pm 1.1$ nm, $CD_{est}=46.5\pm 1.3$ nm, $h_{est}=84.6\pm 0.6$ nm (b_1 hors plage)
Exemple 3 $r_h=18$ nm, $CD=45.3$ nm, $h=50.3$ nm, $r_b=11.6$ nm	$b_{1est}=132.9\pm 1.1$ nm, $CD_{est}=-39.4\pm 1.3$ nm, $h_{est}=138.2\pm 0.6$ nm (profil inexistant)

Tableau III. 23: Caractérisation par PMC_c (entraîné sur un profil de type créneau arrondi au sommet) pour 3 échantillons de forme de profil de type créneau présentant des arrondis en haut et bas des plots.

Profil Créneau haut bas Arrondi	Estimation PMC_c
Exemple 1 $R_h=7.3$ nm, $CD=80.5$ nm, $h=90.8$ nm, $r_h=2.4$ nm	$r_{est}=-10.6\pm 2.5$ nm, $CD_{est}=89.8\pm 0.7$ nm; $h_{est}=72.6\pm 0.8$ nm; (profil inexistant)
Exemple 2 $R_h=8$ nm, $CD=86.9$ nm, $h=50.3$ nm, $r_b=4.5$ nm	$r_{est}=28.9\pm 2.5$ nm; $CD_{est}=85.5$ nm; $h_{est}=61.9\pm 0.8$ nm; $E_{IsIc}= 0.1051$
Exemple 3 $R_h=18$ nm, $CD=45.3$ nm, $h=50.3$ nm, $r_b=11.6$ nm	$r_{est}=-40.4\pm 2.5$ nm; $CD_{est}=3.5\pm 0.7$ nm; $h_{est}=134.9\pm 0.8$ nm; (profil inexistant)

Le PMC_a est le seul RN à fournir des valeurs de paramètres compris dans les domaines d'apprentissage. Les erreurs sur les paramètres CD_{est} et h_{est} dépassent celles obtenues pour l'apprentissage du R.N (cf paragraphe III.2.2.1). Les erreurs E_{IsIc} sont cette fois-ci assez importantes (Exemple 1) et augmentent rapidement avec l'amplitude des arrondis. Malgré cela, les valeurs des paramètres CD et h restent cohérentes. En raison de sa forme relativement simple, le profil créneau donne donc un premier aperçu de la forme géométrique, mais peut aussi s'orienter vers une estimation complètement erronée. En effet dans l'exemple 3, la hauteur estimée est le double de la hauteur attendue. Il s'agit clairement ici d'un problème comparable à la présence de minima locaux. En effet, il se trouve dans ce cas que le profil créneau le plus proche en termes de signature optique ne correspond pas à celui qui est le plus proche en termes de géométrie.

Le PMC_b , quant à lui, de forme plus complexe, induit une forte pente (exemple 1) sur le résultat. Les paramètres CD_{est} et h_{est} restent cependant assez cohérents pour des valeurs faibles des arrondis. Lorsque le profil réel devient trop éloigné du modèle, cela peut donc

conduire à un profil non pris en compte dans la phase d'apprentissage ou non physique. Dans ce cas, un défaut de modèle peut être facilement identifié.

Prenons l'exemple 1 et comparons les résultats du PMC avec ceux résultant d'une optimisation géométrique réalisée directement à partir du modèle trapézoïdal sur l'échantillon simulé. On obtient les estimations suivantes : $b_1=0.2\text{nm}$, $b_2=80.3$, $h=90.5$. La figure III.34 présente les différents profils estimés comparés au profil réel.

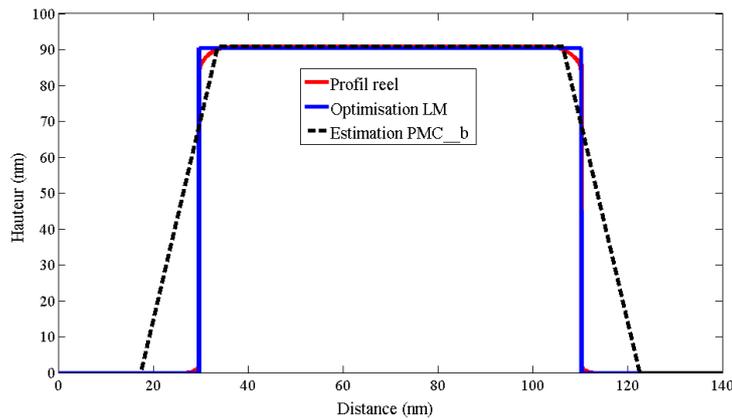


Figure III. 34 : Représentation graphique du profil de l'exemple 1 ($r_h=7.3$ nm, $CD=80.5$ nm, $h=90.8$ nm, $r_b=2.4$) du tableau III.22 (bleu), du profil estimé ($b_{1\text{est}}=16.2$ nm ± 1.1 , $CD_{\text{est}}=72.6$ ± 1.3 nm, $h_{\text{est}}=92$ ± 0.6 nm) caractérisé par PMC_b et du profil estimé par optimisation Levenberg Marquardt (pointillé rouge)

Il apparait clairement que le profil correspondant à l'optimisation paramétrique est plus proche du profil réel que celui donné par le PMC_b. La figure III.35 compare les signatures reconstruites à partir de ces 2 résultats à celle du profil réel.

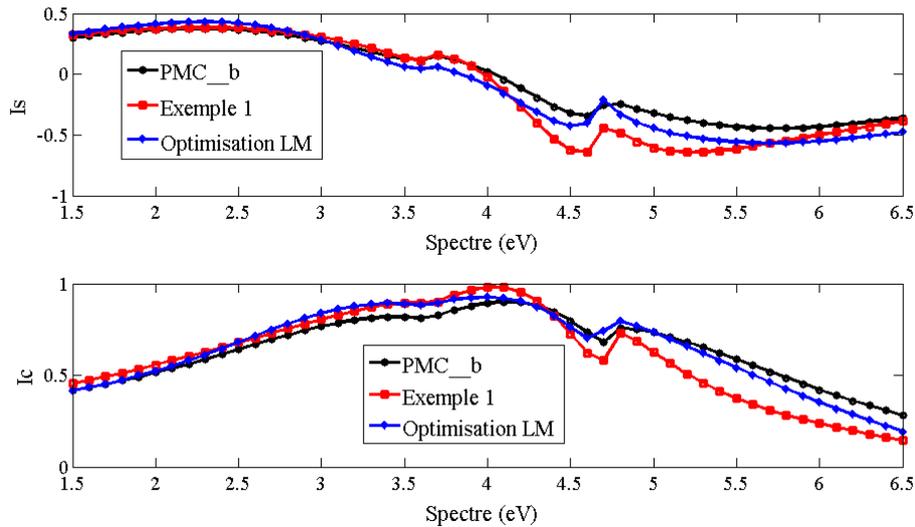


Figure III. 35 : Signature I_s et I_c de l'exemple 1 (rouge) du tableau III.22 et signatures reconstruites à partir des 2 profils estimés par le PMC_b (noir) et celui issu de l'optimisation paramétrique (bleu)

Les erreurs sur les signatures reconstruites sont semblables et sont respectivement de $E_{I_{slc}}=0.090$ pour le PMC_b et $E_{I_{slc}}=0.086$ pour le profil obtenu par optimisation. En fait, comme énoncé précédemment, le R.N estime le profil géométrique trapézoïdal induisant une signature optique proche de celle mesurée sur l'échantillon réel en se basant sur le modèle supposé. L'introduction de pente fictive (inexistante pour le profil réel) lui permet de compenser les influences des arrondis (inaccessibles par le modèle trapézoïdal) sur les signatures. Les 2 optimisations conduisent bien à des résultats optimaux, mais rapportés à chacun de leur critère.

Le PMC_c qui prend en compte l'arrondi en haut des sillons, donne dans la plupart des cas des résultats correspondants à des profils non physiques. On observe que seul l'exemple 2 a pu être caractérisé, mais avec des écarts et une erreur relativement importants ($E_{I_{slc}}=0.1051$). Curieusement, la prise en compte dans le modèle d'un arrondi supplémentaire par rapport au modèle rectangulaire n'apporte pas plus de précision dans les résultats et même conduit à des profils non viables.

L'étude effectuée dans ce paragraphe a permis de mettre en évidence l'existence de paramètres géométriques que l'on peut qualifier de second ordre (r, b_1) et d'autres plus sensibles à la signature optique (h par exemple). Ensuite, l'erreur sur les signatures reconstruites apparaît comme un critère pouvant donner une estimation de la qualité de la caractérisation. Une erreur en dessous de E_{ref} permet de confirmer les hypothèses scattérométriques utilisées dans le problème inverse en particulier la forme du profil. Dans le cas d'une erreur nettement supérieure

à E_{ref} (en pratique au-delà de 5 fois), l'estimation est erronée et les hypothèses structurelles doivent être remaniées. Il en est de même lorsqu'un des paramètres est estimé par le RN en dehors des plages d'apprentissage. Par contre, aucune information n'est fournie sur le nouveau modèle à choisir ou sur les nouvelles plages. Nous avons également montré qu'il était possible d'utiliser une forme plus complexe, disons générique, incluant davantage de profils qu'un modèle unique pour prendre en compte le plus de cas possible. Il s'agit par exemple du cas du profil trapézoïdal ou du profil arrondi en haut incluant tous les deux le profil rectangulaire ($b_1=0$ pour l'un et $r=0$ pour l'autre). Toutefois, il s'avère que la précision théorique du RN pour un profil plus complexe se dégrade par rapport au profil simple. De plus le nombre de minima locaux augmente avec le nombre de paramètres décrivant le modèle. Enfin, nous avons noté dans le cas d'une mauvaise connaissance du modèle (par exemple profil arrondi en haut et en bas des sillons), que les résultats se dégradent rapidement. En effet, si les hypothèses utilisées en scattérométrie ne sont pas vérifiées, en particulier la connaissance du modèle de profil, le problème est alors mal posé. Le PMC cherche alors parmi sa bibliothèque de profils possibles celui qui se rapproche le plus de la signature expérimentale même si ce dernier s'écarte grandement du profil optimal (obtenu par optimisation géométrique).

Une classification pourrait donc permettre de sélectionner le modèle le plus probable afin de conduire à une caractérisation scattérométrique plus fiable. Le paragraphe suivant présente donc la faisabilité de la classification

III.2.3 Projection des signatures

Dans ce paragraphe, nous étudions, comme dans le cas de la détection de défaut, la projection des signatures des différents types de profils afin de nous assurer la possibilité d'une discrimination. La composition des signatures optiques est cette fois-ci composée de 64 intensités (I_s et I_c).

Pour notre étude, nous utilisons 3 différents corpus d'apprentissage.

Le premier noté corpus A est composé de 3 classes définies comme suit :

- Classe C_1 : profils créneaux arrondis au sommet sans plateau (figure III.23.c pour $CD=0$)
- Classe C_2 : profils de type sinusoïdal (figure III.23.d)
- Classe C_3 : profils de type créneau présentant des arrondis au bas des plots (figure III.23.e).

Ainsi, l'identification de la classe C_1 permet de détecter des profils déformés au sommet, la classe C_2 des profils présentant des déviations latérales de type sinusoïdal et la classe C_3 des déformations aux pieds des créneaux. La présence de pente dans le profil n'est pas prise en compte dans ce cas de figure.

Le deuxième corpus noté B contient un profil présentant des flancs non abrupts. Les 3 classes constituant le corpus sont définies comme suit :

- Classe C_1 : profils créneaux arrondis au sommet sans plateau (figure III.23.c pour $CD=0$)
- Classe C_2 : profils de type trapézoïdal (figure III.23.b)
- Classe C_3 : profils de type créneau présentant des arrondis au bas des plots (figure III.23.e).

Ainsi, nous avons substitué le profil sinus du corpus A par des profils trapézoïdaux (fig. III.23.b) pour constituer la classe C_2 permettant l'identification d'une pente dans les profils géométriques à analyser. Notons que la classe C_2 peut inclure le créneau parfait pour $b_1=0$.

Enfin le troisième corpus noté C inclut les 4 classes mentionnées précédemment dans la configuration suivante :

- Classe C_1 : profils créneaux arrondis au sommet (figure III.23.c pour $CD=0$)
- Classe C_2 : profils de type sinusoïdal (figure III.23.d)
- Classe C_2 : profils de type trapézoïdal (figure III.23.b)
- Classe C_4 : profils de type créneau présentant des arrondis au bas des plots (figure III.23.e).

Dans tous les cas, les paramètres géométriques des profils sont choisis de telle sorte que les formes ainsi définies sont toutes imbriquées les unes par rapport aux autres. Ainsi, 3000 signatures sont générées par classe. Les corpus A et B contiennent 9000 signatures alors que le corpus C en compte 12 000.

III.2.3.1 SOM sans classification

Dans un premier temps, nous utilisons le corpus A pour l'apprentissage d'une carte SOM (notée SOM_A). Pour permettre une bonne visualisation des signatures sur la carte SOM nous utilisons 625 neurones repartis sur un espace 2D de 25×25 . Ce nombre est largement

suffisant pour permettre la visualisation des 9000 signatures. L'apprentissage pour cette architecture dure environ 1 h sur un ordinateur Intel Xeon CPU E5 1620 @3.60 GHz 12 Go RAM. La figure III.36 montre la projection des signatures sur la carte SOM_A.

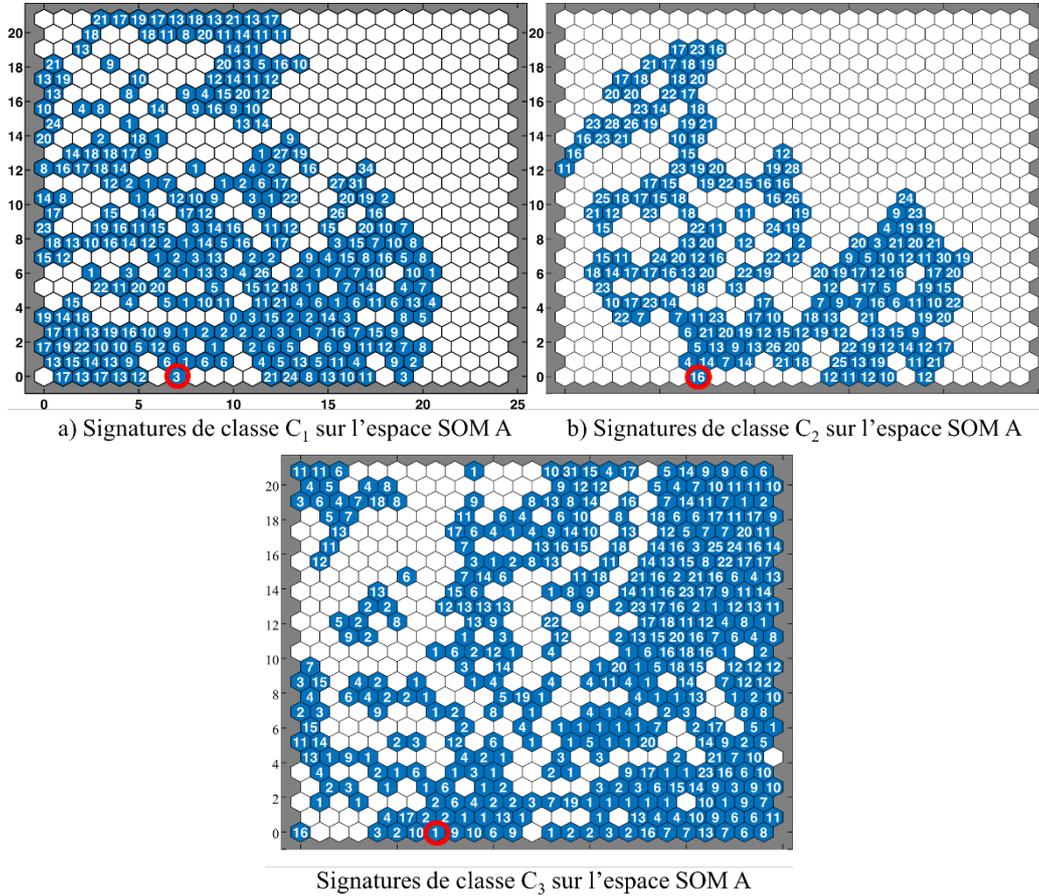


Figure III. 36 : Résultats de la projection sur la carte SOM_A (25×25 neurones) des 9000 signatures correspondant aux 3 classes, C_1 a), C_2 b) et C_3 c)

Les cases bleues représentent les neurones stimulés par les signatures de chaque classe et les cases blanches représentent les neurones jamais activés. Le chiffre à l'intérieur des neurones bleus donne le nombre de fois que ce neurone a été activé.

La zone de regroupement majoritaire de chacune des classes reste bien visible. Dans le détail, on observe ainsi 388 neurones activés uniquement par une seule classe dont 94 pour C_1 (soit 46,7 % des 3000 signatures de type C_1), 55 pour C_2 (soit 34,8 % des signatures C_2) et 239 pour C_3 (soit 87,7 % des signatures C_3). La classe C_3 semble la plus discriminante. Il y a également 42 neurones n'ayant jamais été activés par aucune des classes. Il reste 195 neurones activés à la fois par des signatures de différentes classes (soit 31 % des neurones de la carte).

Prenons le cas particulier du neurone $r \begin{pmatrix} 7 \\ 0 \end{pmatrix}$ (fig.III.36, neurone entouré de rouge) qui est stimulé par 20 signatures de différentes classes dont 3 appartenant à C_1 , 16 pour la classe C_2 et une seule pour C_3 . Ce neurone est donc beaucoup plus sensible aux signatures C_2 c'est-à-dire de profil sinusoïdal ($p(C_1) = 0.15$, $p(C_2) = 0.8$ et $p(C_3) = 0.05$). La figure III.37 représente l'ensemble des signatures ayant activé ce neurone et la figure III.38 illustre les profils géométriques correspondants.

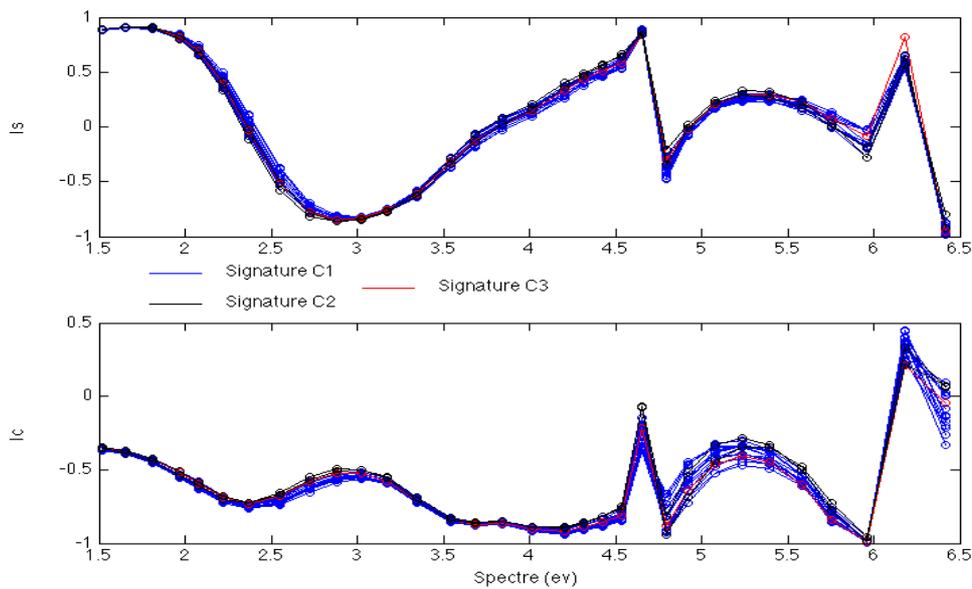


Figure III. 37 : Signatures des profils de classe C_1 , C_2 et C_3 du corpus A ayant activés le neurone $r \begin{pmatrix} 7 \\ 0 \end{pmatrix}$ de la carte SOM_A (25*25 neurones)

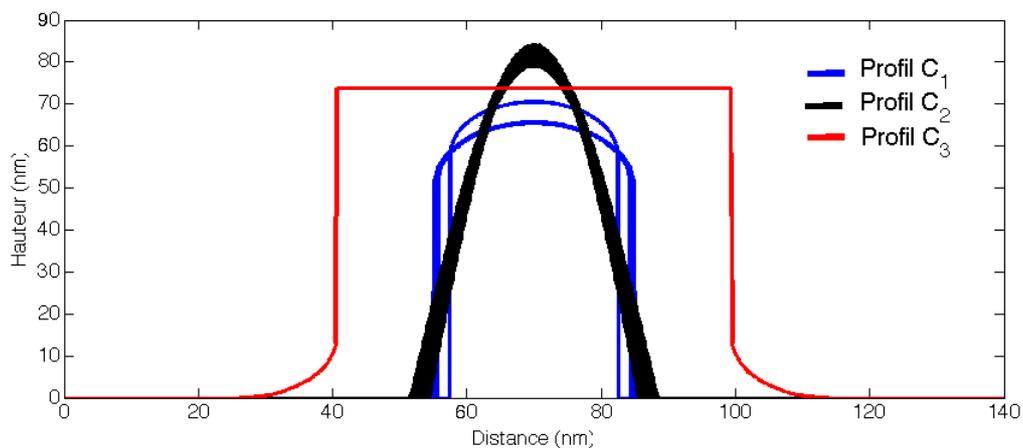


Figure III. 38: Représentation des profils géométriques dont les signatures activent le neurone $r \begin{pmatrix} 7 \\ 0 \end{pmatrix}$ de la carte SOM_A (25*25 neurones)

On remarque que toutes ces signatures sont relativement proches et qu'il est difficile de discerner les classes C_2 et C_3 de la classe C_1 . Cependant, il semble que la partie U.V soit plus discriminante, ce qui amène à penser qu'une optimisation de la qualité de la signature n'est pas à exclure pour améliorer les résultats. Concernant les représentations graphiques, si les profils de type sinus (classe C_2) et créneau arrondi (classe C_1) montrent bien un rapprochement géométrique, ce n'est pas le cas pour le profil à pied arrondi (classe C_3) qui semble assez éloigné des 2 autres. Cela confirme l'hypothèse émise dans le paragraphe précédent ; à savoir que le rapprochement des signatures ellipsométriques n'induit pas systématiquement un rapprochement géométrique des profils. Il s'agit ici d'une limite intrinsèque à toute méthode scattérométrique.

On peut ainsi définir une signature moyenne qui caractériserait le neurone en question. les tableaux III.24 à III.26 montrent l'erreur E_{Islc} par rapport à cette signature moyenne calculée pour chacune des signatures de type C_1 , C_2 et C_3 qui ont activé ce neurone.

Tableau III. 24 : Ecart moyen E_{Islc} à la signature moyenne caractéristique du neurone $\mathbf{r} \begin{pmatrix} 7 \\ 0 \end{pmatrix}$ calculée pour chacune des signatures S_i ($i=1, 2, 3$) de classe C_1 de la figure III.37

Signature C_1	S_1	S_2	S_3	E_{Islc_moy}
E_{Islc}	0.0629	0.0461	0.0348	0,0479

Tableau III. 25: Ecart moyen E_{Islc} à la signature moyenne caractéristique du neurone $\mathbf{r} \begin{pmatrix} 7 \\ 0 \end{pmatrix}$ calculée pour chacune des signatures S_i ($i=1,2,\dots,8$) de classe C_2 de la figure III.37

Signatures C_2	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8	E_{Islc_moy}
E_{Islc}	0.0155	0.0310	0.0473	0.0250	0.0233	0.0404	0.0259	0.0295	0,0285
Signatures C_2	S_9	S_{10}	S_{11}	S_{12}	S_{13}	S_{14}	S_{15}	S_{16}	
E_{Islc}	0.0336	0.0125	0.0465	0.0350	0.0166	0.0067	0.0360	0.0312	

Tableau III. 26: Ecart moyen E_{Islc} à la signature moyenne caractéristique du neurone $\mathbf{r} \begin{pmatrix} 7 \\ 0 \end{pmatrix}$ calculée pour la signature S_1 de classe C_3 de la figure III.37

Signature C_3	S_1
E_{Islc}	0.0627

Les erreurs E_{Islc} sont toutes faibles, mais légèrement supérieures à E_{ref} . En y regardant de plus près, on remarque que les classes C_1 et en particulier C_3 ont les valeurs les plus fortes.

Ainsi, on pourrait penser qu'une représentation plus détaillée de la carte pourrait permettre de dissocier plus finement les classes.

La visualisation des signatures sur la carte SOM_A montre bien une possibilité de discerner en partie les différentes classes.

L'étude du corpus B abouti au même conclusion (cf. annexe B).

Enfin, étudions le Corpus C incluant les 4 classes de profils. Nous utilisons pour la projection une carte légèrement plus grande de 900 neurones repartis sur un espace 2D de 30*30 notée SOM_C. La figure III.39 montre les résultats de la projection.

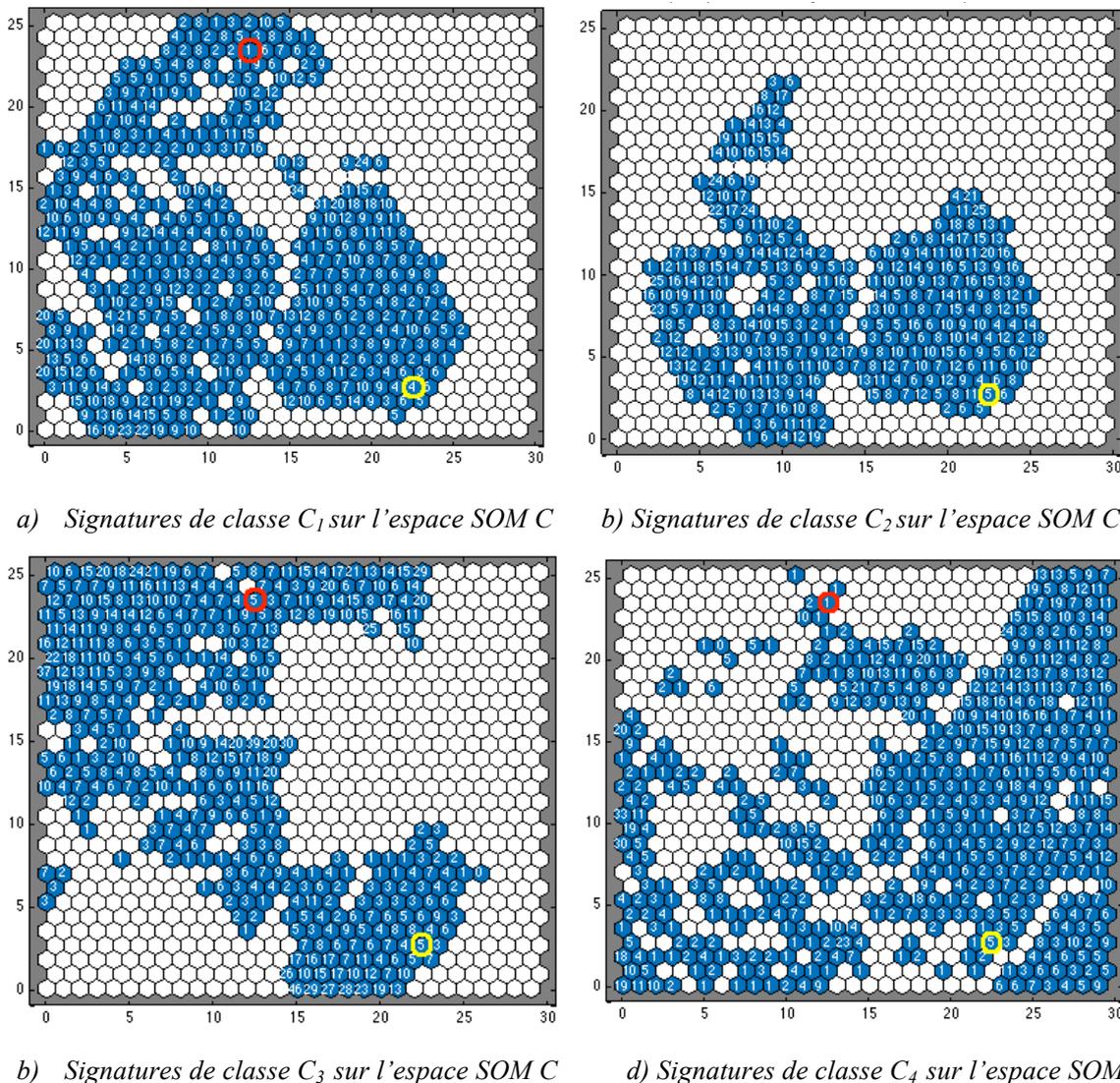


Figure III. 39: Résultat de la projection sur la carte SOM_C (30×30 neurones) des 12000 signatures correspondantes aux 4 classes, C_1 a), C_2 b), C_3 c) et C_4 d)

Il est difficile de définir des zones précises, mais il semble que les classes C_3 , C_4 se regroupent majoritairement dans une zone de neurones bien spécifique alors que les 2 autres se partagent le même espace (C_1 , C_2). De façon plus précise, on observe 371 neurones activés uniquement par une seule classe, dont 17 pour la classe C_1 (soit 7 % des signatures de C_1), 22 pour la classe C_2 (soit 11 % des signatures de C_2), 99 pour la classe C_3 (soit 48,5 % des signatures de C_3) et 234 pour la classe C_4 (soit 85 % des signatures de C_4). La classe C_4 est la classe la plus distinctive. Il y a 61 neurones n'ayant jamais été activés par aucune signature. Le nombre de neurones activés à la fois par des signatures de différentes classes est de 468.

Considérons le cas d'un neurone particulier identifié par la position $\mathbf{r} \begin{pmatrix} 12,5 \\ 24 \end{pmatrix}$ (neurone entouré de rouge sur la figure fig.III.39). Il est activé par 7 signatures au total, dont 1 signature C_1 , aucune signature C_2 , 5 signatures C_3 et 1 signature C_4 . Le neurone peut alors être étiqueté C_3 avec $p(C_3) = 0.72$ contre $p(C_1) = p(C_4) = 0.14$ et $p(C_2) = 0$. Toutes ces signatures sont représentées sur la figure III.40 et les profils géométriques correspondants sont illustrés dans la figure III.41.

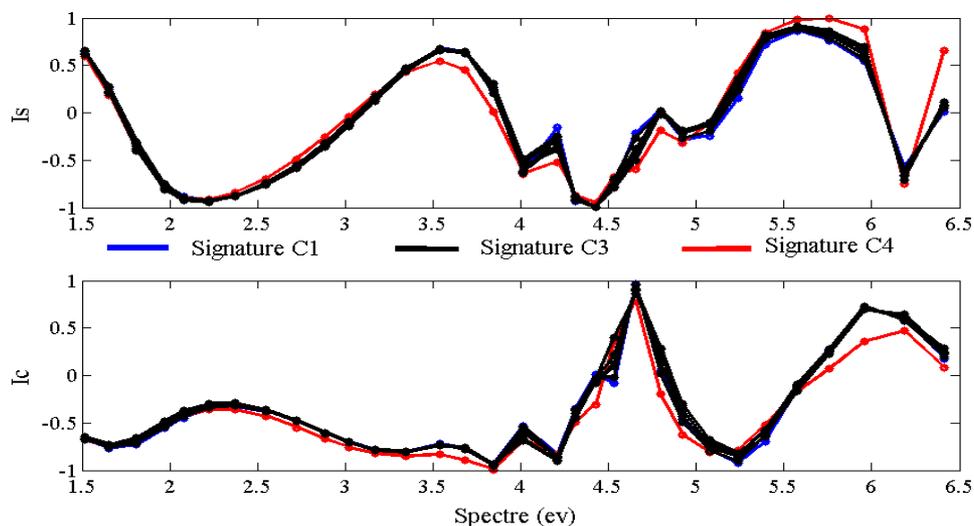


Figure III. 40 : Signatures des profils de classe C_1 , C_3 et C_4 du corpus C ayant activé le neurone $\mathbf{r} \begin{pmatrix} 12,5 \\ 24 \end{pmatrix}$ de la carte SOM_C (30×30 neurones)

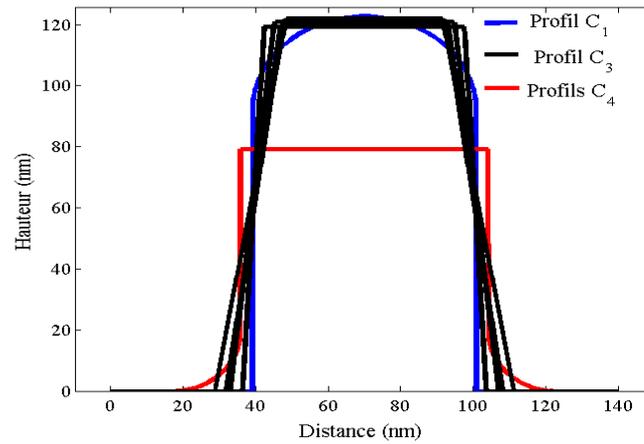


Figure III. 41 : Représentation des profils géométriques dont les signatures activent le neurone

$$\mathbf{r} \begin{pmatrix} 12,5 \\ 24 \end{pmatrix} \text{ de la carte SOM-C (30*30 neurones)}$$

Ces signatures sont relativement proches. Seule celle correspondant à la classe C_4 semble légèrement différente. C'est ce que l'on retrouve sur le tableau III.27, II.28 et III.29 qui affiche les erreurs E_{Islc} calculées dans les mêmes conditions que précédemment

Tableau III. 27 : Ecart moyen E_{Islc} à la signature moyenne caractéristique du neurone

$\mathbf{r} \begin{pmatrix} 12,5 \\ 24 \end{pmatrix}$ calculée pour la signature S_1 de classe C_1 de la figure III.40

Signatures $_1$	S_1
E_{Islc}	0.0644

Tableau III. 28 : Ecart moyen E_{Islc} à la signature moyenne caractéristique du neurone

$\mathbf{r} \begin{pmatrix} 12,5 \\ 24 \end{pmatrix}$ calculée pour chacune des signatures S_i ($i=1, \dots, 5$) de classe C_3 de la figure III.40

Signatures C_3	S_1	S_2	S_3	S_4	S_5	E_{Islc_moy}
E_{Islc}	0.0471	0.0460	0.0428	0.0385	0.0240	0,0396

Tableau III. 29 : Écart moyen E_{Islc} à la signature moyenne caractéristique du neurone

$\mathbf{r} \begin{pmatrix} 12,5 \\ 24 \end{pmatrix}$ calculée pour la signature S_1 de classe C_4 de la figure III.40

Signatures C_4	1
E_{Islc}	0.1276

Les erreurs sur C_3 sont plus faibles que celle sur C_1 et de façon plus prononcée sur C_4 . Il s'agit ici d'un neurone à forte sensibilité pour la classe C_3 . Là encore, il paraît envisageable d'augmenter le nombre de neurones de la carte pour affiner les frontières entre les neurones ;

ce qui permettrait dans ce cas précis d'exclure le profil C_4 du neurone $\mathbf{r} \begin{pmatrix} 12,5 \\ 24 \end{pmatrix}$ qui semble assez éloigné.

Prenons un autre neurone $\mathbf{r} \begin{pmatrix} 22,5 \\ 3,5 \end{pmatrix}$ (neurone entouré de jaune sur la figure III.39) activé par 4 signatures C_1 , 5 signatures C_2 et 5 signatures C_3 et 5 signatures C_4 . Les probabilités des classes sont presque équivalentes et relativement faibles pour chacune d'elle : $p(C_1) = 0.22$, $p(C_2) = p(C_3) = p(C_4) = 0.26$. La figure III.42 représente ces signatures et la figure III.43 les profils géométriques correspondants.

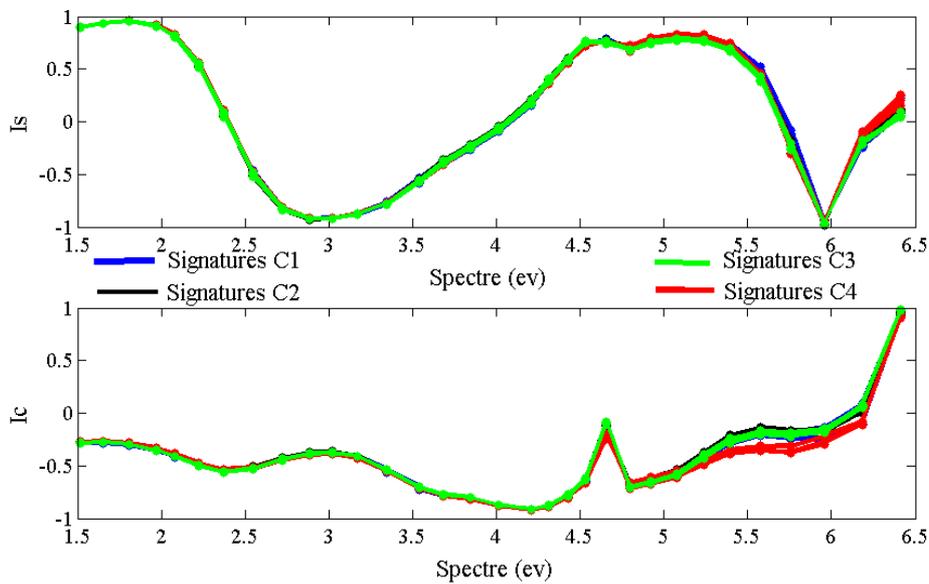


Figure III. 42 : Signatures des profils de classe C_1 , C_2 , C_3 et C_4 ayant activé le neurone $\mathbf{r} \begin{pmatrix} 22,5 \\ 3,5 \end{pmatrix}$ de la carte SOM_C (30*30 neurones)

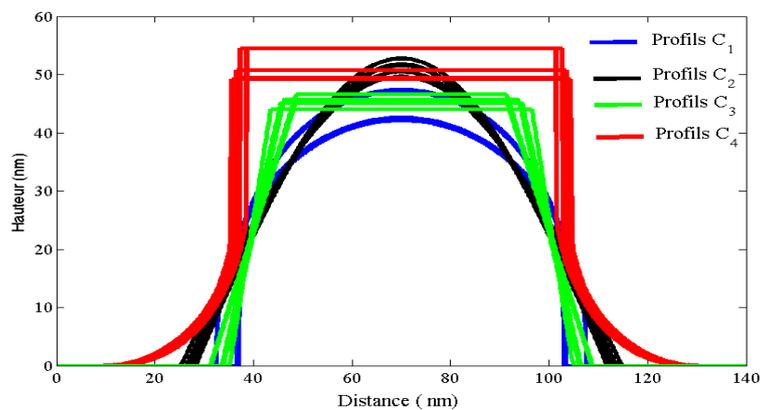


Figure III. 43 : Représentation des profils géométriques dont les signatures activent le neurone $\mathbf{r} \begin{pmatrix} 22,5 \\ 3,5 \end{pmatrix}$ de la SOM_C (30*30 neurones)

Cette fois-ci, les signatures sont quasiment indiscernables sauf sur une petite partie du spectre dans l'UV. Les profils géométriques bien que différents semblent proches en termes de hauteur et d'épaisseur. Les erreurs E_{Islc} calculées pour chaque type de signature par rapport à la moyenne sont regroupées dans les tableaux III.30 à III.33.

Tableau III. 30 : Ecart moyen E_{Islc} à la signature moyenne caractéristique du neurone $r \binom{22,5}{3,5}$ calculée pour chacune des signatures S_i ($i=1, \dots, 4$) de classe C_1 de la figure III.42

Signatures C_1	S_1	S_2	S_3	S_4	E_{Islc_moy}
E_{Islc}	0.0248	0.0220	0.0309	0.0281	0,02645

Tableau III. 31 : Ecart moyen E_{Islc} à la signature moyenne caractéristique du neurone $r \binom{22,5}{3,5}$ calculée pour chacune des signatures S_i ($i=1, \dots, 5$) de classe C_2 de la figure III.42

Signatures C_2	S_1	S_2	S_3	S_4	S_5	E_{Islc_moy}
E_{Islc}	0.0200	0.0205	0.0201	0.0226	0.0241	0,02146

Tableau III. 32 : Ecart moyen E_{Islc} à la signature moyenne caractéristique du neurone $r \binom{22,5}{3,5}$ calculée pour chacune des signatures S_i ($i=1, \dots, 5$) de classe C_3 de la figure III.42

Signatures C_3	S_1	S_2	S_3	S_4	S_5	E_{Islc_moy}
E_{Islc}	0.0235	0.0207	0.0210	0.0205	0.0167	0,02048

Tableau III. 33: Ecart moyen E_{Islc} à la signature moyenne caractéristique du neurone $r \binom{22,5}{3,5}$ calculée pour chacune des signatures S_i ($i=1, \dots, 4$) de classe C_4 de la figure III.42

Signatures C_4	S_1	S_2	S_3	S_4	S_5	E_{Islc_moy}
E_{Islc}	0.0406	0.0317	0.0435	0.0328	0.0296	0,03564

Les valeurs de E_{Islc} obtenues confirment quantitativement le rapprochement des signatures. De plus, les erreurs moyennes sont du même ordre de grandeur que E_{ref} ; ce qui les rend indiscernables à la mesure.

Ce neurone ne permet pas de discriminer clairement un type de profil. Toutefois, une mauvaise interprétation est à exclure, car la connaissance des probabilités d'activation nous indique une insensibilité de ce neurone à un type de classe précis si on se fixe un seuil de 0,5.

La visualisation à l'aide d'une carte SOM des signatures correspondant à plusieurs formes de profils montre bien dans l'ensemble les possibilités de discrimination. Par contre, nous avons mis en évidence des cas particuliers où une même signature (ou groupe de signatures) peut conduire à des géométries de profils assez différentes. La probabilité d'activation de chaque neurone par une classe donne une première idée sur l'appartenance de la signature et reflète le niveau de risque d'affectation. Pour améliorer la discrimination, on peut envisager soit de modifier qualitativement la qualité de la signature (utiliser seulement l'IR) ou changer l'architecture de la carte (augmenter ou diminuer le nombre de neurones). Dans tous les cas il reste des neurones non étiquetés qui ne permettent pas de conclure quant à l'appartenance d'une classe précise.

III.2.4 Méthode basique de détermination du profil géométrique

III.2.4.1 Comparaison des signatures reconstruites

Une première solution de sélection de modèles peut être effectuée sur la base de l'erreur E_{Isc} calculée sur les signatures reconstruites à partir des différents résultats. Un seuil peut être simplement fixé pour une décision de rejet systématique. Sur la base des résultats précédents (paragraphe III.2.2) nous pouvons fixer le seuil autour de 0,10 au-delà duquel les profils géométriques estimés par les R.N montrent une déviation conséquente par rapport à la forme attendue. En dessous, le modèle donnant la plus faible erreur peut être retenu parmi l'ensemble des modèles de traitement utilisés.

Prenons l'exemple du profil complexe (paragraphe III.2.2.3.) correspondant aux paramètres $r_t=7.3$, $CD=80.5$, $h=90.8$, $r_b=2.4$ dont le modèle géométrique est absent des différentes bases d'apprentissages utilisées. La signature de ce profil a été caractérisée par PMC_a et PMC_b avec une erreur sur la signature reconstruite respectivement de $E_{Isc}=0.0421$ (cf. tableau III.21) et $E_{Isc} = 0,090$ (cf. tableau III.22). Les deux solutions sont en dessous du seuil que l'on s'est fixé. Les estimations des paramètres par PMC_a ($CD=82 \pm 0.093$, $h=94.3 \pm 0.091$) sont donc retenues et celles estimées par PMC_b ($b_1=16.2 \pm 1.1$, $b_2=72.6 \pm 1.3$, $h=92 \pm 0.6$) rejetées sur la base de l'erreur la plus faible. Les caractérisations par PMC_c et PMC_d n'ont quant à eux pas donné des profils exploitables.

La solution de sélection sur la base de l'erreur rms s'avère simple et rapide bien qu'elle nécessite systématiquement le calcul de la signature reconstruite. Cette méthode est efficace pour écarter un profil (erreur supérieure à 0.1) ou confirmer le modèle géométrique choisi

(erreur de l'ordre de grandeur de E_{ref}). Pour des valeurs d'erreur comprises entre E_{ref} et 0,1, l'interprétation est délicate.

III.2.4.2 Méthode de pondération de profil

Une autre solution possible s'appuyant sur le critère d'erreur consiste à calculer la moyenne pondérée des résultats des différents profils géométriques estimés par plusieurs modèles neuronaux. Etant donné que l'erreur E_{IsIc} est d'autant plus faible que le modèle de caractérisation est supposé fidèle, on peut apparenter cette erreur à un coefficient de pondération noté α_i pour tout modèle i de profil correspondant à l'erreur du E_{IsIc_i} et calculée comme suit :

$$\alpha_i = 1 - \frac{E_{IsIc_i}}{\sum_i E_{IsIc_i}} \quad \text{III.7}$$

Ainsi la fonction F définissant le profil géométrique peut être donnée par :

$$F(x) = \frac{\sum_i \alpha_i f_i(x)}{\sum_i \alpha_i} \quad \text{III.8}$$

où $f_i(x)$ correspond à la fonction donnant la hauteur du profil géométrique en fonction des abscisses x définis à partir du modèle i .

Par exemple nous effectuons une pondération du profil complexe précédent ($r_t=7.3$, $CD=80.5$, $h=90.8$, $r_b=2.4$). Le PMC_a induit un coefficient $\alpha_a=0.68$ (Tableau III.21). La fonction $f_a(x)$ est la fonction qui délimite la géométrie du créneau à partir des paramètres CD et h du créneau. Le PMC_b donne un coefficient $\alpha_b=0.32$ (Tableau III. 22). La fonction $f_b(x)$ trace la géométrie du trapèze à partir des paramètres (b_1 , b_2 et h). Ainsi la pondération est induite par 68 % de f_a et 32 % de f_b . La figure III.44 présente le profil obtenu par pondération des résultats des PMC_a et PMC_b selon les erreurs E_{IsIc} obtenues pour chacun d'eux.

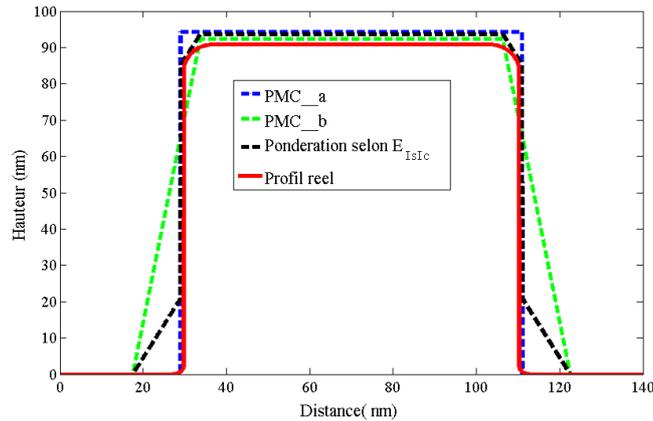


Figure III. 44 : Représentation des profils obtenus par PMC_a (bleu), PMC_b (vert) et par la méthode des pondérations de profil (noir) dans le cadre d'un échantillon présentant un profil créneau arrondi en bas et haut des sillons (rouge)

On remarque que le profil reconstruit par pondération tend à se rapprocher du profil réel. D'un point de vue qualitatif, si on calcule les erreurs quadratiques de ces profils par rapport à la cible on trouve : 9.93 nm pour le créneau, 10.13 nm pour le profil obtenu par pondération et 16.88 pour le trapèze. Ainsi, cette méthode permet de prendre en compte les différentes particularités de chaque profil élémentaire.

Cette méthode met à profit le calcul systématique de l'erreur sur le profil reconstruit. Elle nécessite la caractérisation de l'échantillon sur la base de plusieurs modèles géométriques. L'avantage principal réside dans le fait de proposer un modèle original non inclus dans la bibliothèque de modèles.

Les résultats fournis ici constituent la base d'une nouvelle méthode qui semble envisageable. D'autres études plus poussées doivent être menées pour apprécier la qualité de la solution fournie. On peut notamment envisager d'améliorer la définition des coefficients de pondération α_i (par exemple la prise en compte de valeurs seuil de E_{IsIc_i}), augmenter le nombre de profils élémentaires possibles pour affiner la forme finale, etc ... Ce travail ne sera pas davantage développé dans le manuscrit.

III.2.4.3 Utilisation d'un profil complexe

III.2.4.3.1 Paramètres du modèle générique

Nous avons constaté précédemment que l'utilisation d'un profil complexe contenant intrinsèquement un certain nombre de profils dits élémentaires pouvait conduire à réduire les

erreurs d'interprétation possibles. C'était le cas du profil trapèze qui inclut le modèle rectangulaire pour $b_1=0$ (cf paragraphe III.2.2.2). Sur cette constatation, on considère un modèle de profil géométrique complexe (figure III.45.c) défini par 4 paramètres dont le plateau CD_{Top} , la base CD_{bottom} , la hauteur h et un paramètre nommé degré d'arrondi α . Cette valeur est comprise entre 0 (pas d'arrondi) et 1 (valeur maximale possible soit $CD_{Top}/2$).

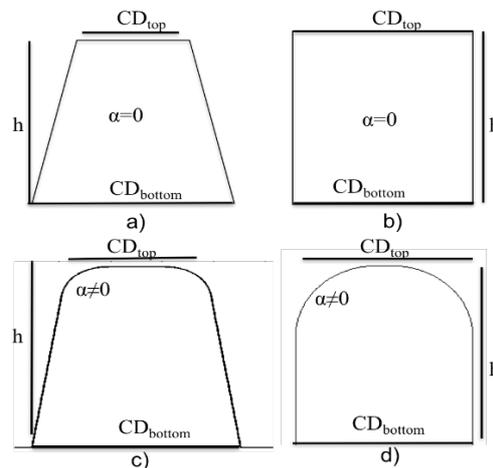


Figure III. 45 : Exemples de formes de profils trapèze a), créneau parfait b), créneau arrondi d) obtenues à partir du modèle générique c) défini par 4 paramètres CD_{top} , CD_{bottom} , h et α

Ce profil générique contient 3 types de motifs géométriques simples décrits précédemment : le trapèze parfait pour $CD_{top} \neq CD_{bottom}$ et $\alpha=0$ (cf figure III.45.a), le créneau parfait pour $CD_{top}=CD_{bottom}$, et $\alpha=0$ (cf figure III.45.b), et le créneau arrondi pour $CD_{top}=CD_{bottom}$, et $\alpha \neq 0$ (cf figure III.45.d). Pour des valeurs quelconques des paramètres géométriques, on peut définir une forme qui s'apparente à un trapèze arrondi en haut des plots (cf figure III.45.c). Les 3 formes inclus dans le profil générique doivent respecter les conditions d'existence mentionnées au paragraphe III.2.1.2. Par conséquent, le profil générique est défini physiquement dans les conditions suivantes : $CD_{top} \leq CD_{bottom}$, $0 \leq \alpha \leq 1$ et $CD_{bottom} < T$. Notons que plus la forme est complexe, plus le nombre de paramètres augmente et plus les conditions sont sévères.

III.2.4.3.2 Caractérisation scattérométrique classique fonctionnant sur modèle de profil complexe

Le nombre important de paramètres géométriques définissant le profil générique implique l'utilisation d'un R.N dont l'architecture est plus complexe que ceux utilisés

précédemment. Ainsi, pour la caractérisation scattérométrique, nous utilisons un PMC composé de 50 neurones cachés et 4 sorties correspondant aux paramètres du profil générique. L'étape d'apprentissage a nécessité 12 000 signatures (identique en nombre et en type d'intensité à celles précédemment utilisées pour les différents PMC du paragraphe III.2.2.).

La création de la base d'apprentissage peut être réalisée de 2 façons différentes :

- Tirage aléatoire des paramètres dans un domaine défini à l'avance. Toutefois, comme il faut tenir compte de toutes les conditions d'existence du profil géométrique, il faut limiter considérablement la plage de variations des paramètres.
- Générer séparément et uniquement les 3 types de profils simples (figure III.45) définis précédemment. Cette méthode présente l'avantage de contrôler la répartition et la représentation de chaque type de profil et d'écartier le problème des conditions d'existence. C'est cette méthode qui sera utilisée par la suite.

Ainsi, nous avons généré plusieurs ensembles de 3000 échantillons pour chaque type de profil dans les mêmes plages de variation que celles utilisées précédemment au paragraphe III.2.2.

Les performances d'apprentissage du R.N sont évaluées sur $N_{\text{test}} = 1800$ signatures simulées (450 de chaque type de profil). L'apprentissage du PMC générique a duré environ 2 min sur un ordinateur Intel Xeon CPU E5 1620 @ 3.60 GHz 12 Go RAM. Les erreurs obtenues sur les paramètres sont $E_{\text{rms}}^{\text{CDtop}} = 1.8 \text{ nm}$, $E_{\text{rms}}^{\text{CDbottom}} = 1.2 \text{ nm}$, $E_{\text{rms}}^{\text{h}} = 2,8 \text{ nm}$ et $E_{\text{rms}}^{\alpha} = 0.1$. Les courbes de régression sur chaque sortie évaluée à l'issue de l'apprentissage sont représentées sur les Figures III.46.

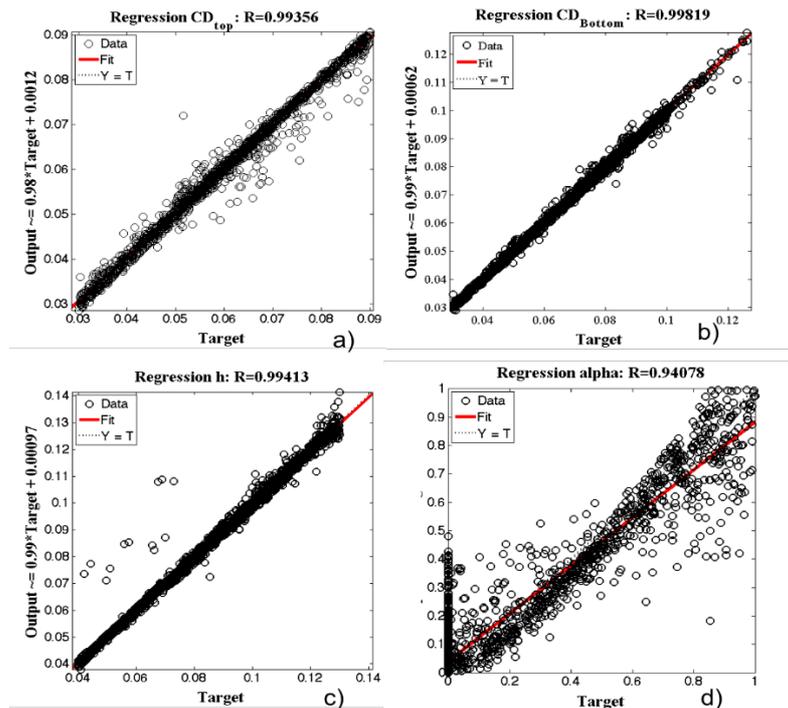


Figure III. 46: Performance du PMC fonctionnant sur un modèle géométrique générique (4 paramètres) à l'issue de l'apprentissage: courbes de régression (évaluée sur les $N_{test}=1800$ échantillons) concernant les paramètres CD_{top} a) CD_{Bottom} b) h c) et α d).

Pour les paramètres CD_{top} , CD_{bottom} et h les courbes de régression sont correctes affichant un coefficient R supérieur à 0,99 (figure III. 46 a, b, c). Par contre, les estimations pour le paramètre α sont nettement moins bonnes avec un coefficient de régression significativement plus faible ($R=0,94$). Ce résultat confirme que l'arrondi se comporte comme un paramètre de second ordre comme mentionné précédemment (cf paragraphe III.2.2.1.) Le RN semble relativement bien résister à l'augmentation du nombre de paramètres mais affiche des performances nettement moins bonnes que celles correspondant à des profils élémentaires.

Pour tester les performances du PMC, commençons par étudier le cas d'échantillons présentant un profil simple, par exemple trapézoïdal. Les mêmes signatures utilisées pour les tests des RN précédents (cf. paragraphe III.2.2) sont fournies au PMC. Le tableau III.34 résume les résultats obtenus pour les 4 profils trapèzes et la figure III.47 donne la représentation graphique concernant l'échantillon particulier numéro 2. Pour faciliter la comparaison, un paramètre b_1 estimé est calculé par l'expression : $b_1 = (CD_{Bottom}-CD_{top})/2$.

Tableau III. 34: Caractérisation par un PMC entraîné sur un profil générique pour 4 échantillons particuliers de forme de profil trapézoïdal

Profil trapèze	Estimation PMC_generique
Exemple 1 $b_1=4.1$ nm, $CD=36.6$ nm, $h=101.7$ nm	$CD_{top}=36.4\pm 1,8$ nm, $CD_{bottom}=44.5\pm 1,2$ nm, $h=101.3\pm 2.8$ nm, $\alpha=-0.0005\pm 0,1$ ($b_{1est}=4.05$ nm) $E_{Is,Ic} = 0.0137$ (calculée pour $\alpha=0$)
Exemple 2 $b_1=10.2$ nm, $CD=45.2$ nm, $h=61.2$ nm	$CD_{top}=45\pm 1,8$ nm, $CD_{bottom}=67.9\pm 1,2$ nm, $h=61.6\pm 2.8$ nm, $\alpha=0.02\pm 0,1$ ($b_{1est}=11.2$ nm) $E_{Is,Ic} = 0.0110$
Exemple 3 $b_1=14.6$ nm, $CD=50.5$ nm, $h=117.5$ nm	$CD_{top}=49.3\pm 1,8$ nm, $CD_{bottom}=80.1\pm 1,2$ nm, $h=117.7\pm 2.8$ nm, $\alpha=-0.09\pm 0,1$ ($b_{1est}=15.4$ nm) $E_{Is,Ic} = 0.0096$
Exemple 4 : $b_1=16.9$ nm, $CD=81.8$ nm, $h=91.5$ nm	$CD_{top}=83.3\pm 1,8$ nm, $CD_{bottom}=114.1\pm 1,2$ nm, $h=92.9\pm 2.8$ nm, $\alpha=-0.011\pm 0,1$ ($b_{1est}=15.35$ nm) $E_{Is,Ic} = 0.0191$ (calculée pour $\alpha=0$)

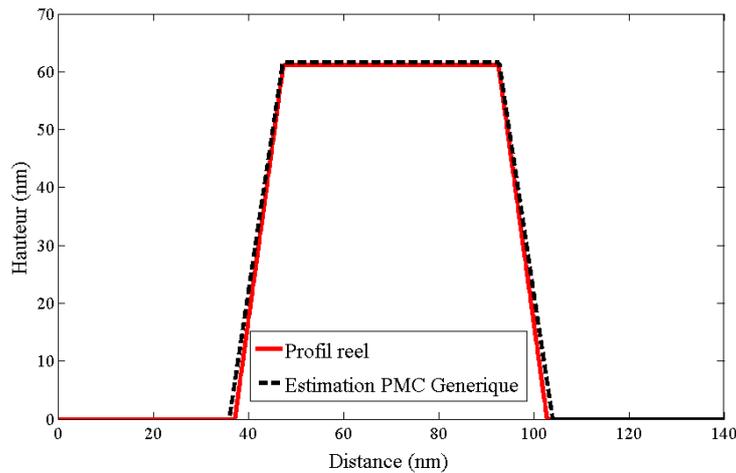


Figure III. 47 : Représentation graphique du profil trapèze de l'exemple 2 du tableau III.35 ($b_1=10.2$ nm, $CD=45.2$ nm, $h=61.2$ nm) (rouge) et du profil estimé ($CD_{top}=45 \pm 1,8$ nm, $CD_{bottom}=67.9 \pm 1,2$ nm, $h=61.6 \pm 2.8$ nm, $\alpha=0.02\pm 0,1$ ($b_{1est}=11.2$ nm)) par le PMC entraîné sur un profil générique (noir)

Dans tous les cas, la forme estimée du profil est assez proche de celle attendue. De plus, les valeurs de α sont très faibles et proches de 0. L'erreur $E_{Is, Ic}$ est ainsi du même ordre de grandeur que E_{ref} . Toutefois elle reste largement supérieure à celle obtenue avec le modèle de PMC_b entraîné exclusivement avec des trapèzes.

Dans le même esprit, les 4 échantillons de type créneau arrondis au sommet sont étudiés en annexe C. Les résultats sont similaires en termes de performances.

Le cas d'échantillons présentant un profil sinusoïdal est ensuite analysé. Le comportement du PMC dans ce cas est intéressant, car il ne s'agit pas ici d'un modèle de profil élémentaire inclus dans le modèle générique. Pour cela, nous utilisons les mêmes échantillons simulés dans le tableau III.20. Le tableau III.35 présente les résultats obtenus et la figure III.48 illustre géométriquement le résultat relatif à l'échantillon numéro 3.

Tableau III. 35 : Caractérisation par un PMC entraîné sur un profil générique pour 4 échantillons particuliers de forme de profil sinusoïdal

Profil sinus	Estimation PMC_generique
Exemple 1 b=76.9 nm, h=86.5 nm	$CD_{top}=37.4\pm 1,8$ nm, $CD_{bottom}=72.2\pm 1,2$ nm, $h=86.9\pm 2.8$ nm, $\alpha=1.01\pm 0,1$, $E_{Is,lc} = 0,200$
Exemple 2 b=50.9 nm, h=99.3 nm	$CD_{top}=51.3\pm 1,8$ nm, $CD_{bottom}=42.4\pm 1,2$ nm, $h=70.3\pm 2.8$ nm, $\alpha=-0.02\pm 0,1$, ($CD_{top}>CD_{bottom}$) profil inexistant
Exemple 3 b=86.9 nm, h=43.2 nm	$CD_{top}=48.7\pm 1,8$ nm, $CD_{bottom}=76.1\pm 1,2$ nm, $h=41.2\pm 2.8$ nm, $\alpha=1.01\pm 0,1$ $E_{Is,lc} = 0.0257$
Exemple 4 b=52.6 nm, h=75.7 nm	$CD_{top}=30.6\pm 1,8$ nm, $CD_{bottom}=40.1\pm 1,2$ nm, $h=81.3\pm 2.8$ nm, $\alpha=-0.2\pm 0,1$ (α hors plage)

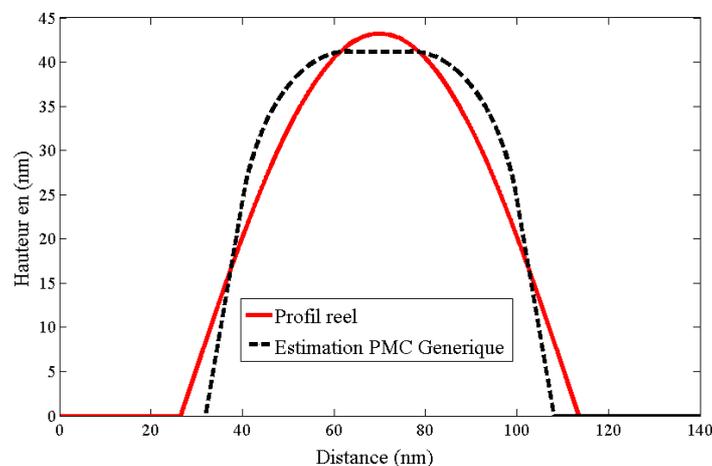


Figure III. 48 : Représentation graphique du profil de l'exemple 3 du tableau III.36 (b=86.9 nm, h=43.2 nm) du tableau (noir) et du profil estimé ($CD_{top}=48.7 \pm 1,8$ nm, $CD_{bottom}=76.1 \pm 1,2$ nm, $h=41.2 \pm 2.8$ nm, $\alpha=1.01\pm 0,1$ $E_{Is,lc} = 0.0257$) caractérisé par PMC entraîné sur un profil générique (rouge).

Dans la plupart des cas, les valeurs calculées par le PMC conduisent soit à un profil non défini physiquement ne validant pas les conditions d'existence définies précédemment (cf paragraphe III.2.1.2). Seul l'exemple 3 a pu être caractérisé avec une erreur sur la signature reconstruite correspondant à $2E_{ref}$. Pour les autres, lorsque le calcul est possible, les valeurs de $E_{Is,lc}$ confirment une déviation conséquente par rapport au profil réel attendu.

En conclusion, le modèle générique permet effectivement de prendre en compte simultanément plusieurs profils simples. La caractérisation des formes dégénérées du modèle donne des résultats corrects. Par contre, lorsque le profil à caractériser ne peut être approché par la fonction générique, on est ramené au cas de profil unique pouvant conduire à des ambiguïtés dans les prises de décision et d'interprétation. Le modèle générique reste donc limité à un nombre restreint de profils.

Par ailleurs, pour assurer des performances fiables (dans les plages de variation), il est contraint à de petites variations autour des valeurs nominales des paramètres.

Enfin, le modèle générique étant également plus complexe (4 paramètres) qu'un modèle simple, les résultats de la caractérisation sont moins précis que ceux obtenus dans le cas d'un modèle élémentaire. La baisse de performance est d'autant plus forte que le nombre de paramètres augmente dans le modèle générique.

III.2.5 Identification de type de profil par LVQ

Reprenons dans un premier temps le corpus A précédemment utilisé lors de la projection sur la carte SOM_A (cf. paragraphe III.2.3.1). Le réseau correspondant à ce modèle de corpus est noté LVQ_A par la suite. Son architecture optimale a nécessité 40 neurones dans la couche cachée avec 3 neurones de sortie correspondant à chaque classe. L'apprentissage pour cette architecture dure environ 1 h sur un ordinateur un Intel Xeon CPU E5 1620 @3.60 GHz 12 Go RAM. 1350 signatures de la base d'apprentissage sont retenues pour le test. Le tableau III.36 résume les performances du LVQ sous la forme de la matrice de confusion.

Tableau III. 36 : Matrice de confusion du LVQ_A (40 neurones cachés) calculée sur le corpus test pour l'identification de profils géométriques. La classe C₁ correspond au créneau arrondi (CD=0), C₂ au sinus et C₃ au créneau à bas arrondi

	C ₁ réelle	C ₂ réelle	C ₃ réelle	Précision
C ₁ estimée	293	105	97	59.2 %
C ₂ estimée	125	304	80	59.7 %
C ₃ estimée	32	41	273	78.9 %
Sensibilité	65.1 %	67.6 %	60.7 %	Exactitude: 64,4%

L'exactitude du classifieur est assez faible (64.4%). On retrouve le fait que la classe C_3 semble plus discriminante (cf carte SOM_A figure III.36) avec une précision proche de 80% sur la troisième sortie du RN. Le LVQ montre une certaine confusion pour les signatures des 3 classes.

Dans le but de comparer avec les performances de la carte SOM_A, les signatures activant le neurone $r \begin{pmatrix} 7 \\ 0 \end{pmatrix}$ de ladite carte (cf. paragraphe III.2.3.1) sont présentées au LVQ_A. Les résultats sont écrits dans les tableaux III.37 à III.39.

Tableau III. 37 : Classification par LVQ_A (40 neurones cachés) des 3 signatures (S_i avec $i=1, \dots, 3$) correspondant à une classe C_1 ayant activé le neurone $r \begin{pmatrix} 7 \\ 0 \end{pmatrix}$ de la carte SOM_A.

Signatures	S_1	S_2	S_3
Sortie C_1	1	0	0
Sortie C_2	0	1	1
Sortie C_3	0	0	0

Tableau III. 38 : Classification par LVQ_A (40 neurones cachés) des 16 signatures (S_i avec $i=1, \dots, 16$) correspondant à une classe C_2 ayant activé le neurone $r \begin{pmatrix} 7 \\ 0 \end{pmatrix}$ de la carte SOM_A.

Signatures C_2	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8	S_9	S_{10}	S_{11}	S_{12}	S_{13}	S_{14}	S_{15}	S_{16}
Sortie C_1	0	0	0	1	1	0	1	0	0	1	1	0	0	0	0	0
Sortie C_2	1	1	1	0	0	1	0	1	1	0	0	1	1	1	1	1
Sortie C_3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Tableau III. 39 : Classification par LVQ_A (40 neurones cachés) de la signature S_1 correspondant à une classe C_3 ayant activé le neurone $r \begin{pmatrix} 7 \\ 0 \end{pmatrix}$ de la carte SOM_A.

Signatures C_3	S_1
Sortie C_1	0
Sortie C_2	1
Sortie C_3	0

Sur le total de signatures présentées au LVQ, 60 % des estimations sont correctes contre 40 % erronées. La confusion constatée dans la projection de la carte n'est visiblement pas levée par une classification LVQ.

Les performances moyennes du LVQ peuvent s'expliquer par le fonctionnement intrinsèque de ce dernier. En effet, la couche cachée du LVQ fonctionne de manière similaire à une SOM c'est-à-dire qu'un neurone particulier de la couche cachée se spécialise pour une classe spécifique. L'organisation des neurones, la mise à jour des poids et l'étiquetage sont cependant différents. On retrouve le fait que des signatures proches de plusieurs profils différents activent le même neurone impliquant ainsi une confusion dans la détermination du profil.

L'étude sur le corpus B conduit aux mêmes conclusions.

Le LVQ dans la forme décrite ici ne répond que partiellement aux objectifs et présente de performances qui ne correspondent pas aux objectifs fixés dans cette thèse.

Une solution pourrait consister à modifier la qualité des signatures. En effet, rappelons que la visualisation des signatures a montré dans les paragraphes précédents (cf. figure III.37 et figure III.40 et III.42) une sensibilité plus importante dans le domaine de l'U.V. Ainsi, il est intéressant d'étudier le comportement du LVQ lorsque la signature fournie est tronquée en dessous de 4.5 eV (soit 42 mesures). Les performances du RN après apprentissage noté LVQ_A' sont affichées dans le tableau III.40.

Tableau III. 40: Matrice de confusion du LVQ_A' (40 neurones cachés) calculée sur le corpus test pour l'identification de profils géométriques. La classe C₁ correspond au créneau arrondi (CD=0), C₂ au sinus et C₃ au créneau à bas arrondi

	C ₁ réelle	C ₂ réelle	C ₃ réelle	Précision
C ₁ estimée	450	0	6	98.7 %
C ₂ estimée	0	435	160	73.1 %
C ₃ estimée	0	15	284	95.0 %
Sensibilité	100 %	96.7 %	63.1 %	Exactitude: 86,6%

Le LVQ_A' affiche une exactitude de 86.6 % contre 64.4 % pour la signature non tronquée. Notons que la sensibilité sur la classe C₁ est de 100 % c'est-à-dire que tous les profils créneaux arrondis seront parfaitement bien identifiés par le classifieur. Les performances sont donc nettement améliorées par la qualité de la signature optique.

Cette simple étude du spectre dans l’UV confirme le fait que la qualité de la signature ellipsométrique est un point clé de toute méthode scattérométrique. Elle conditionne grandement la qualité des résultats. En effet, le principe même de la scattérométrie se base exclusivement sur l’étude de la signature. La composition de cette dernière est donc d’une importance capitale.

III.2.6 Identification de type de profil par PMC

III.2.6.1 Résultats de la classification à 3 et 4 classes

A présent nous effectuons une classification supervisée à l’aide d’un PMC classifieur qui a montré précédemment de bien meilleurs résultats que le LVQ (cf. paragraphe III.1.2.4). Dans un premier temps nous utilisons les signatures composant le Corpus A (cf. paragraphe III.2.3). L’apprentissage optimal conduit à une architecture de 20 neurones cachés et 3 neurones de sortie correspondant chacun à une classe. Ce dernier a nécessité seulement 15 secondes sur un ordinateur Intel Xeon CPU E5 1620 @3.60 GHz 12 Go RAM. Le tableau III.41 résume les performances du PMC ainsi formé nommé PMC_A concernant les 3 classes citées sous la forme de la matrice de confusion.

Tableau III. 41: Matrice de confusion du PMC_A (20 neurones cachés) calculée sur le corpus test pour l’identification de profils géométriques. La classe C₁ correspond au créneau arrondi (CD=0), C₂ au sinus et C₃ au créneau à bas arrondi

	C ₁ réelle	C ₂ réelle	C ₃ réelle	Précision
C ₁ estimée	449	0	4	99.1 %
C ₂ estimée	1	450	1	99.6 %
C ₃ estimée	0	0	445	100 %
Sensibilité	99.8 %	100 %	98.9 %	Exactitude: 99.6%

Le PMC_A affiche de très bonnes performances (exactitude de 99.6 %). Les précisions sur chaque sortie ainsi que les sensibilités aux 3 classes sont respectivement supérieures à 99.1 % et 98.9 %. Seuls 6 profils sont faussement classés (soit 0.4 % des échantillons du corpus test). Ces résultats sont nettement meilleurs que le LVQ.

Pour s’en assurer reprenons les signatures activant le neurone $\mathbf{r} \begin{pmatrix} 7 \\ 0 \end{pmatrix}$ de la carte SOM_A (cf paragraphe III.2.3.1 figure III.37) qui présentent des similitudes évidentes tout en

appartenant à des classes de profils différents. Ces signatures sont fournies au PMC_A à l'issue de l'entraînement et les résultats sont regroupés dans les tableaux III.42 à III.44 pour les 3 classes.

Tableau III. 42 : Classification par PMC_A (20 neurones cachés) des 3 signatures (S_i avec $i=1, \dots, 3$) correspondant à une classe C_1 ayant activé le neurone $\mathbf{r} \begin{pmatrix} 7 \\ 0 \end{pmatrix}$ de la carte SOM_A.

Signatures C_1	S_1	S_2	S_3
p(C1)	1	0.99	0.99
p(C2)	0	0.01	0.009
P(C3)	0	0	0.001

Tableau III. 43 : Classification par PMC_A (20 neurones cachés) des 16 signatures (S_i avec $i=1, \dots, 8$) correspondant à une classe C_2 ayant activé le neurone $\mathbf{r} \begin{pmatrix} 7 \\ 0 \end{pmatrix}$ de la carte SOM_A.

Signature C_2	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8
p(C1)	0.0007	0.0009	0.0013	0.0007	0.0010	0.0036	0.0009	0.0003
p(C2)	0.9989	0.9987	0.9986	0.9987	0.9969	0.9948	0.9983	0.9996
P(C3)	0.0004	0.0004	0.0002	0.0006	0.0021	0.0016	0.0009	0.0001
Signature C_2	S_9	S_{10}	S_{11}	S_{12}	S_{13}	S_{14}	S_{15}	S_{16}
p(C1)	0.0007	0.0009	0.0009	0.0011	0.0008	0.0008	0.0013	0.0012
p(C2)	0.9988	0.9989	0.9973	0.9987	0.9989	0.9989	0.9986	0.9987
P(C3)	0.0005	0.0002	0.0018	0.0002	0.0002	0.0003	0.0001	0.0001

Tableau III. 44: Classification par PMC_A (20 neurones cachés) de la signature $S1$ correspondant à la classe $C3$ ayant activé le neurone $\mathbf{r} \begin{pmatrix} 7 \\ 0 \end{pmatrix}$ de la carte SOM_A.

Signature C_3	S_1
p(C1)	0.0002
p(C2)	0.0064
P(C3)	0.9934

Chacune des signatures a été clairement identifiée avec de fortes probabilités d'appartenance dans la bonne classe. L'ambiguïté que semblait faire apparaître certains

neurones de la SOM_A et du classifieur LVQ_A se trouve levée ici par une classification supervisée par PMC.

L'étude sur le corpus B conduit aux mêmes conclusions. (Annexe B).

Ainsi, contrairement au LVQ les très bons résultats du PMC concernant 3 classes nous permettent d'envisager une étude plus conséquente. Ainsi le cas des 12 000 signatures du corpus C est étudié.

L'architecture optimale est cette fois-ci constituée de 50 neurones cachés et 4 neurones de sorties. Le R.N ainsi formé est noté PMC_C. L'apprentissage de ce dernier a duré environ 30 secondes sur un ordinateur Intel Xeon CPU E5 1620 @3.60 GHz 12 Go RAM. Les résultats du classifieur sur les 1800 échantillons composant le corpus test sont présentés dans le tableau III.45.

Tableau III. 45: Matrice de confusion du PMC_C (50 neurones cachés) calculée sur le corpus test pour l'identification de profils géométriques. La classe C₁ correspond au créneau arrondi (CD=0), C₂ au sinus, C₃ au trapèze et C₄ au créneau a bas arrondi

	C ₁ réelle	C ₂ réelle	C ₃ réelle	C ₄ réelle	Précision
C ₁ estimée	449	5	0	1	98.7 %
C ₂ estimée	0	445	0	3	99.3 %
C ₃ estimée	1	0	450	0	99.8 %
C ₄ estimée	0	0	0	446	100 %
Sensibilité	99.8 %	98.9 %	100 %	99.1	Exactitude: 99,4%

L'exactitude qui résume les performances globales reste élevée à 99,4%. Le perceptron semble très bien adapté à ce problème de classification. Les précisions sur chaque sortie ainsi que les sensibilités aux 4 classes sont respectivement supérieures à 98,7% et 98,9%. Les performances ne semblent pas souffrir de la présence d'une classe supplémentaire. Seuls 10 profils sont faussement classés (soit 0,5% des échantillons du corpus test).

Les signatures activant le neurone $\mathbf{r} \begin{pmatrix} 22,5 \\ 3,5 \end{pmatrix}$ de la SOM_C sont ensuite présentées en entrée du PMC_C. Les résultats du PMC sont regroupés dans les tableaux III.46 à III. 49 pour les 4 classes représentées.

Tableau III. 46 : Classification par PMC_C (50 neurones cachés) des 4 signatures (S_i avec $i=1, \dots, 4$) correspondant à une classe C_1 ayant activé le neurone $r \begin{pmatrix} 22,5 \\ 3,5 \end{pmatrix}$ de la carte SOM_C.

Signature C_1	S_1	S_2	S_3	S_4
p(C1)	0.7725	0.7769	0.8115	0.8303
p(C2)	0.2205	0.2173	0.1415	0.1242
P(C3)	0.0001	0.0001	0.0231	0.0167
P(C4)	0.0070	0.0057	0.0238	0.0288

Tableau III. 47: Classification par PMC_C (50 neurones cachés) des 5 signatures (S_i avec $i=1, \dots, 5$) correspondant à une classe C_2 ayant activé le neurone $r \begin{pmatrix} 22,5 \\ 3,5 \end{pmatrix}$ de la carte SOM_C.

Signature C_2	S_1	S_2	S_3	S_4	S_5
p(C1)	0.1679	0.1705	0.1625	0.1626	0.1822
p(C2)	0.8204	0.8193	0.8256	0.7136	0.8021
P(C3)	0.0111	0.0001	0.0001	0.0000	0.0149
P(C4)	0.0005	0.0100	0.0118	0.1238	0.0008

Tableau III. 48: Classification par PMC_C (50 neurones cachés) des 5 signatures (S_i avec $i=1, \dots, 5$) correspondant à une classe C_3 ayant activé le neurone $r \begin{pmatrix} 22,5 \\ 3,5 \end{pmatrix}$ de la carte SOM_C.

Signature C_3	S_1	S_2	S_3	S_4	S_5
p(C1)	0.0097	0.0172	0.0061	0.0053	0.0369
p(C2)	0.0204	0.0275	0.0172	0.0153	0.0409
P(C3)	0.9698	0.9553	0.9767	0.9794	0.9222
P(C4)	0.0000	0.0000	0.0000	0.0000	0.0000

Tableau III. 49: Classification par PMC_C (50 neurones cachés) des 5 signatures (S_i avec $i=1, \dots, 5$) correspondant à une classe C_4 ayant activé le neurone $r \begin{pmatrix} 22,5 \\ 3,5 \end{pmatrix}$ de la carte SOM_C.

Signature C_4	S_1	S_2	S_3	S_4	S_5
p(C1)	0.0000	0.0000	0.0000	0.0001	0.0001
p(C2)	0.0010	0.0104	0.0005	0.0503	0.0488
P(C3)	0.0000	0.0000	0.0000	0.0000	0.0000
P(C4)	0.9990	0.9896	0.9995	0.9496	0.9511

Le PMC donne, dans tous les cas, la probabilité d'appartenance à la classe cible relativement la plus élevée (supérieure à 77.2 %). Dans certains cas, on observe un rapprochement avec une autre classe (par exemple C_1 dans le tableau III.46 ou C_2 dans le tableau III.47) mais la probabilité n'excède jamais 22 %. Les ambiguïtés soulevées par la carte SOM sont donc levées une nouvelle fois par le PMC.

Comme pour le LVQ, nous avons testé l'importance du spectre UV dans la classification par PMC (cf. annexe C). Les résultats sont sensiblement meilleurs, mais l'apport est moindre comparativement à l'amélioration apportée sur le LVQ. C'est pour cela que nous ne le détaillons pas ici.

La qualité des résultats du PMC peut paraître surprenante comparée à ceux du LVQ. Toutefois, le fonctionnement intrinsèque de chaque méthode est différent. Ainsi, comme cela a été dit lors de l'analyse précédente, le LVQ fonctionne comme une SOM en identifiant un groupe de signatures à un neurone particulier. Ce n'est pas le mode opératoire d'un classifieur par PMC qui modélise, par une régression non linéaire, la relation entre une signature particulière et la classe correspondante. Dans notre cas particulier de détection de profils géométriques, la méthode par PMC semble plus appropriée.

III.2.6.2 Résistance de la classification PMC au bruit de mesure

Pour compléter l'étude sur la classification supervisée par PMC et avant de passer à la caractérisation expérimentale, il est nécessaire d'évaluer la résistance de la méthode de classification au bruit de mesure. Pour cela nous retenons le classifieur PMC_C. Ainsi, 4 échantillons appartenant à chacune des classes C_1 , C_2 , C_3 , et C_4 sont choisis comme références :

- Echantillon 1 : $r=28,6$, $h=103.8$ nm (classe C_1)
- Echantillon 2 : $b=53.4$, $h=60.4$ nm (classe C_2)
- Echantillon 3 : $b_1=16.6$, $b_2=61.9$, $h=126.3$ (classe C_3)
- Echantillon 4 : $r_b=17.5$, $CD=41.4$, $h=85.3$ (classe C_4)

Les signatures correspondantes sont simulées puis bruitées avant d'être traitées par le R.N classifieur. Les différents bruits sont modélisés par une fonction Gaussienne avec 6 écarts types différents ($\sigma=0$, $\sigma=0.003$, $\sigma=0.01$, $\sigma=0.02$, $\sigma=0.03$ et $\sigma=0.09$). La figure III.49 décrit les résultats par la probabilité d'appartenance à la classe cible pour chaque échantillon en fonction du bruit introduit sur les signatures.

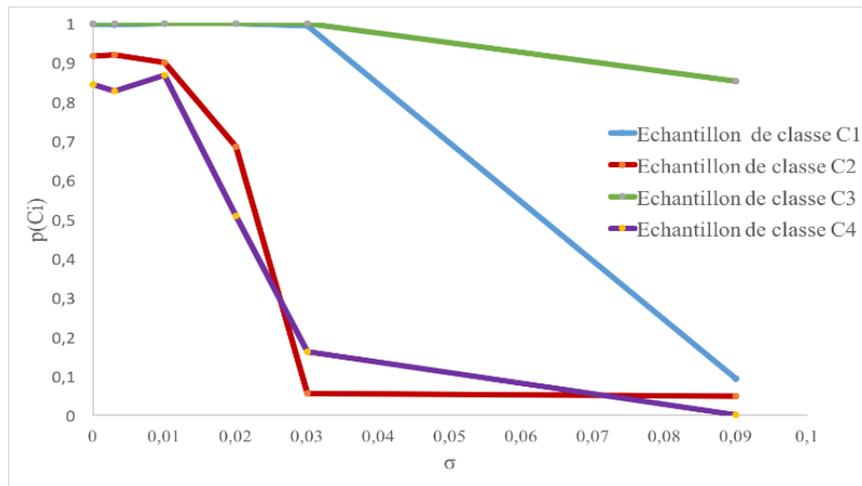


Figure III. 49 : Probabilités d'appartenance à la classe cible données par le PMC_C (50 neurones cachés) en fonction du niveau de bruit introduit sur les signatures (écart type dans le modèle Gaussien) de 4 échantillons présentant des profils de différentes classes (C_1 , C_2 , C_3 et C_4).

La résistance au bruit de mesure est variable d'une classe à l'autre. Globalement, le classifieur montre une bonne résistance jusqu'à une valeur $\sigma=0.02$. Au-delà, les probabilités d'appartenance à certaines classes commencent par chuter énormément et conduisent à une mauvaise affectation (échantillons 1, 2 et 4). Les résultats restent donc viables sur une zone incluant largement le bruit de mesure de l'ellipsomètre (variation maximale fournie par le constructeur sur I_s ou I_c de 0,01, cf. chapitre I paragraphe I.2.2.2).

III.2.7 Mesure sur échantillon réel

Pour finir, les différentes méthodes d'identification de profils sont testées sur un cas réel. L'échantillon est fabriqué par la société STMicroelectronics par lithographie optique par immersion à 193 nm. Les lignes de réseau sont réalisées sur de la résine IM5010 reposant sur du BARC le tout sur un substrat de silicium. La période du motif est de 140 nm. Le profil attendu est de type créneau. La figure III.50 montre les résultats d'une caractérisation AFM qui semble présenter des arrondis au sommet. On observe également la présence de rugosité non négligeable y compris sur les flancs.

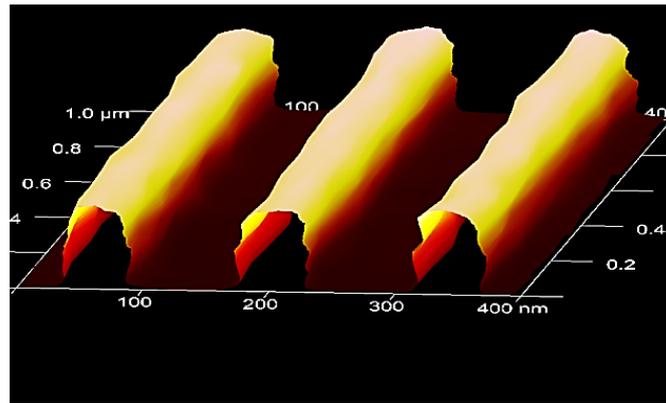


Figure III. 50 : Caractérisation AFM de l'échantillon (période 140 nm) en résine IM5010 reposant sur du BARC sur un substrat de silicium

Les signatures de l'échantillon sont mesurées puis fournies aux différents R.N utilisés pour la classification. Dans ce qui suit, on considérera la couche de BARC fixe et connue (82 nm).

III.2.7.1 Identification expérimentale du type de profil

La signature mesurée est fournie aux différentes cartes étudiées précédemment (SOM_A, SOM_B, SOM_C). Les résultats concernant les sensibilités des neurones activés sur les 3 cartes sont résumés dans le tableau III.50.

Tableau III. 50: sensibilité de neurones activés par la signature expérimentale de l'échantillon (période 140 nm) sur les cartes SOM obtenues à partir des corpus A, B et C (SOM_A, SOM_B, SOM_C)

SOM_A	SOM_B	SOM_C
$P(C_1)=1.00$	$P(C_1)=1.00$	$P(C_1)=0.67$
$P(C_2)=0$	$P(C_2)=0$	$P(C_2)=0.28$
$P(C_3)=0$	$P(C_3)=0$	$P(C_3)=0$
		$P(C_4)=0.05$

Dans tous les cas, la probabilité d'appartenance à la classe C_1 est largement majoritaire par rapport aux autres classes en particulier pour les cartes SOM_A et SOM_B pour lesquelles la précision est de 100 %. Les 3 cartes donnent des performances en conformité avec les résultats attendus. Le neurone activé sur la carte SOM_C a tendance à être sensible à certains profils de type sinusoïdal. Cette ambiguïté est absente de la carte SOM_A.

Les résultats obtenus par les LVQ étudiés précédemment sont résumés dans le tableau III.51.

Tableau III. 51: Résultats de la classification de la signature expérimentale de l'échantillon (période 140 nm) pour les 3 modèles de LVQ (entraînés sur les corpus A, A' et B)

	LVQ_A	LVQ_B	LVQ_A'
Sortie 1	0	1	0
Sortie 2	1	0	1
Sortie 3	0	0	0

Le LVQ entraîné sur le corpus A estime un profil de classe C_2 c'est-à-dire de type sinusoïdal y compris dans le cas de l'utilisation d'une signature dans l'U.V. Les sensibilités de la seconde sortie sont respectivement de 59,7% et 73,1%. Par contre le LVQ_B donne un résultat de classes C_1 soit arrondi au sommet (précision de 58,8%). Nous avons vu précédemment sur les SOM que les classes de profil arrondi au sommet et sinus peuvent être envisagées pour cet échantillon (SOM_C). C'est ce que l'on retrouve dans les résultats du LVQ. Ainsi, le corpus B ne contenant pas la classe de type sinusoïdal, il est normal que le LVQ_B estime le profil de type arrondi au sommet. La faible valeur des précisions de chaque sortie doit cependant inciter à la prudence quant à l'interprétation des résultats.

Pour finir, le tableau III.52 résume les résultats obtenus pour tous les classifieurs de type PMC détaillés dans le paragraphe précédent.

Tableau III. 52: Résultat de la classification de la signature expérimentale de l'échantillon (période 140 nm) pour les 3 modèles de PMC (entraînés sur les corpus A et B et C)

	PMC_A	PMC_B	PMC_C
p(C_1)	1.0000	0.9999	1
p(C_2)	0.0000	0.0001	0.0000
p(C_3)	0.0000	0.0000	0.0000
p(C_4)			0.0000

La classe estimée est dans tous les cas la classe C_1 avec des probabilités a posteriori très proches de 1. Rappelons que la précision sur la première sortie de chaque PMC est respectivement de 99,1%, 98,7% et 98,7% également. Ainsi, tous les critères sont réunis pour

conclure sur la forme supposée du profil de l'échantillon en question : il s'agit d'un profil de type créneau arrondi (CD=0).

Il est difficile de conclure sur la validation des performances des classifieurs sur des cas réels à partir d'un seul échantillon. D'autres études sont bien entendu nécessaires. Toutefois, rappelons que l'échantillon utilisé présente une part non négligeable de rugosité, ce qui n'a pas été pris en compte dans les différentes modélisations géométriques. Cette dernière comme tout autre défaut par rapport à la modélisation (indice optique, connaissance de la hauteur de BARC, etc...) peut être comptabilisée comme un bruit si l'écart par rapport à l'hypothèse structurale n'est pas trop significatif. Ceci peut expliquer en partie les résultats du LVQ.

III.2.7.2 Caractérisation scattérométrique

Le modèle créneau arrondi au sommet (CD=0) est choisi conformément au PMC classifieur de la section précédente. La caractérisation scattérométrique proprement dite peut-être menée dans le cadre de cette hypothèse structurale. Différents PMC régresseurs (cf chapitre II paragraphe II.3.2.2.2) possédant 2 neurones de sortie linéaires sont mis en place. La couche de sortie estime donc les valeurs des paramètres h et r du profil arrondi. Le tableau III.53 montre les résultats obtenus pour 3 architectures de RN différentes. Chacun d'eux est entraîné avec 3000 signatures simulées, dont 950 sont dédiées à la phase de test. La figure III.51 compare les signatures mesurées aux signatures reconstruites à partir des paramètres issus de la caractérisation par le PMC à 10 neurones cachés.

Tableau III. 53: Résultat de la caractérisation scattérométrique de l'échantillon (période 140 nm) par plusieurs PMC (respectivement 10, 20 et 50 neurones cachés) fonctionnant sur un modèle de profil géométrique type créneau arrondi (CD=0) de classe C_1

Paramètres	PMC 10 neurones cachés	PMC 20 neurones cachés	PMC 50 neurones cachés
r	22.4 ± 0,02 nm	25.2 ± 0.016 nm	25.1 ± 0.017 nm
h	109.4 + 0,11 nm	109.7 ± 0,05 nm	108.5 ± 0.05 nm
E_{isc}	0.12	0.14	0.12

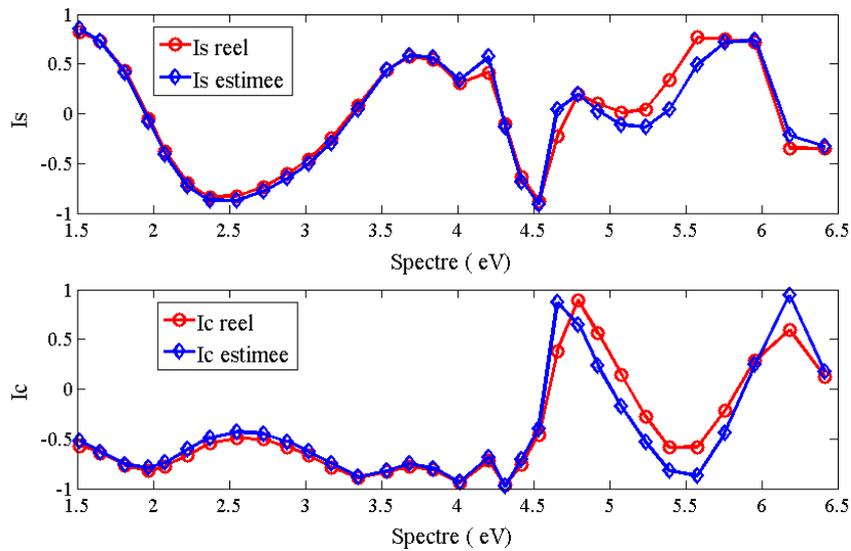


Figure III. 51 : Signature mesurée sur l'échantillon (période 140 nm) et signature reconstruite à partir de la caractérisation par un PMC (10 neurone cachés) fonctionnant sur un modèle de profil géométrique type créneau arrondi ($r=22.4\text{nm}$, $h=109.4\text{ nm}$) de classe C_1

Les 3 PMC donnent des valeurs de paramètres relativement proches les uns des autres. Nous retenons le PMC 10 neurones cachés qui constitue l'architecture la plus simple pour une erreur E_{Islc} la plus faible. On observe sur la figure III.51 des signatures relativement proches avec cependant un léger décalage dans l'U.V visible. Notons que l'erreur E_{Islc} est largement plus grande que E_{ref} . En effet, le profil supposé correspond à une modélisation, plus ou moins fidèle du profil réel qui semble beaucoup plus complexe (cf figure III.50) avec notamment la présence d'arrondis en partie basse de la structure. De plus, la présence de rugosité non prise en compte dans la modélisation peut influencer significativement les mesures et perturber la qualité des résultats.

III.2.7.3 Comparaison des différentes caractérisations scattérométriques

Les résultats des différentes méthodes présentées dans les paragraphes III.2.2, III.2.4.2, et III.2.4.3 sont comparés à une caractérisation AFM.

La mesure ellipsométrique de la signature est fournie directement aux différents modèles de R.N fonctionnant en mode régresseur sur un modèle géométrique unique (paragraphe III.2.2). Les plages de variations des paramètres sont identiques à ceux utilisées précédemment et la signature simulée tient compte de la couche fixe de Barc sous la structure périodique. Les résultats obtenus sont résumés dans le tableau III.54

Tableau III.54: Résultat de la caractérisation scattérométrique de l'échantillon (période 140 nm) par différents PMC fonctionnant en mode régresseur sur un modèle géométrique fixé (profil créneau pour PMC_a, profil trapèze pour PMC_b, profil créneau arrondi au sommet pour PMC_c et profil sinus pour PMC_d et PMC générique)

Modèle PMC	Paramètres estimés
PMC_a	$CD_{est}=45.2 \pm 0.02 \text{ nm}$; $h_{est}=109 \pm 0.019 \text{ nm}$; $E_{Islc}=0.1300$
PMC_b	$b_{1est}=15.8 \pm 0.07 \text{ nm}$; $b_{2est}=28.7 \pm 0.12 \text{ nm}$; $h_{est}=118 \pm 0.10 \text{ nm}$; $E_{Islc}=0,1113$
PMC_c	$r_{est}=77.9 \pm 1.2 \text{ nm}$; $CD_{est}=59.3 \pm 0.33 \text{ nm}$; $h_{est}=119 \pm 0.31 \text{ nm}$ (profil inexistant)
PMC_d	$b_{est}=72.8 \pm 0.02 \text{ nm}$; $h_{est}=127.8 \pm 0.032 \text{ nm}$; $E_{Islc}=0,1493$
PMC_Générique	$CD_{top}=19.9 \pm 1 \text{ nm}$, $CD_{bottom}=77.6 \pm 0.7 \text{ nm}$, $h=110 \pm 1.6 \text{ nm}$ et $\alpha=1.15 \pm 0.10 \text{ nm}$; $E_{Islc}=0.1870$

Seul le PMC fonctionnant sur le modèle créneau présentant des arrondis en haut des plots donne des résultats inexploitable car non physique ($r_{est} > CD_{est}$). Cela peut paraître surprenant car il s'agit d'une forme proche de celle supposée. Toutefois, les résultats montrent bien la tendance du PMC_c à estimer un profil présentant des arrondis conséquents et donc une forme définie aux limites du modèle

Les erreurs E_{Islc} calculées sont au-delà de E_{ref} pour les raisons évoquées plus haut. Le PMC_b entraîné exclusivement sur des trapèzes donne une erreur légèrement faible. Notons qu'il fait apparaître des pentes sur les flancs du motif, inexistantes sur la figure AFM. Comme évoqué précédemment l'augmentation du paramètre b_1 permet de compenser la présence d'arrondis pour atteindre un minimum d'erreur E_{Islc} . On peut également faire la même constatation concernant le profil générique.

A partir de ces résultats, nous appliquons la méthode de construction du profil géométrique par pondération à partir des valeurs de E_{Islc} (cf paragraphe III.2.4.2). Le résultat est comparé aux profils estimés par le PMC fonctionnant sur le modèle générique, obtenu par application d'une méthode d'identification par PMC et enfin à celui mesuré par AFM. Les profils géométriques superposés sont représentés sur la figure III.52.

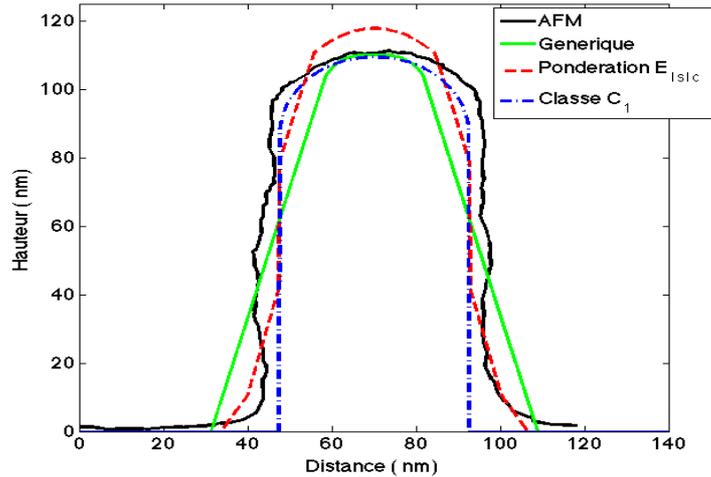


Figure III. 52 : Représentation graphique des résultats de la caractérisation expérimentale de l'échantillon (période 140 nm) par différentes méthodes : mesure AFM (noire), scattérométrie par classifieur neuronal (bleu), scattérométrie classique par utilisation d'un modèle générique (vert) et par pondération de résultats scatterométriques classiques fonctionnant sur plusieurs modèles géométriques (rouge))

Dans tous les cas, ces résultats montrent bien que l'hypothèse du profil créneau mentionnée par le constructeur n'est pas vérifiée. Tous les profils ont tendance à faire apparaître une forme fortement arrondie au sommet. La mesure AFM quant à elle montre des arrondis asymétriques au pied de la structure.

Le profil générique semble le plus éloigné par l'ajout d'une pente qui n'est pas décelée par la caractérisation AFM. On retrouve les limites ici du modèle générique : le profil sera déformé (présence d'une pente) pour compenser une forme qu'il ne peut modéliser (arrondi au pied de la structure).

Le profil créneau arrondi (classe C_1) donné par la PMC classification semble le plus fidèle à la caractérisation AFM. L'erreur E_{Islc} correspondant à ce profil est de 0.12.

La méthode de pondération selon E_{Islc} semble s'adapter au bas du profil alors qu'aucune des formes simples utilisées n'en présentait. Il a également tendance à se rapprocher au mieux à l'AFM.

Cette étude nous a permis de présenter la potentialité des R.N à identifier le type de profil géométrique de structures diffractantes à partir de la signature ellipsométrique utilisée en scattérométrie. La visualisation des signatures optiques sur une carte SOM nous a permis de mettre en évidence le fait qu'elles pouvaient se regrouper en fonction de la forme géométrique de la structure

concernée. Ainsi, nous avons proposé plusieurs méthodes permettant l'identification d'un profil pour un échantillon donné. Nous avons notamment proposé une idée de méthode fonctionnant par pondération de profils élémentaires conduisant à des motifs complexes. Toutefois, l'essentiel de l'étude a porté sur deux types de classification supervisée. Les résultats ont montré que la classification par projection (SOM et LVQ) permet bien d'identifier et de regrouper les signatures optiques suivant une forme de profil. Cependant, dans certains cas, la signature choisie n'est pas suffisante et ne permet pas une distinction systématique. Nous avons montré par l'utilisation d'une information spécifique au domaine UV, qu'un travail sur l'optimisation de la qualité de la mesure doit être réalisé. Par ailleurs, la classification par PMC s'est avérée, comme pour la détection de défaut, nettement plus performante que celle utilisant un LVQ qui, dans l'état actuel de nos travaux, ne semble pas répondre au cahier des charges fixé. Par contre, la probabilité a posteriori donnée par le PMC constitue un véritable estimateur de confiance résistant au bruit de mesure. Il possède l'avantage de donner une probabilité d'appartenance à une classe et donc, laisse la possibilité à l'utilisateur de gérer les risques en fixant un seuil de rejet systématique. Les tests expérimentaux doivent être poursuivis pour valider définitivement la méthode mais les premiers résultats sont encourageants.

Conclusion générale

Ce travail de thèse répond à une problématique de métrologie par voie optique concernant l'identification de formes ou de défauts dans la caractérisation de structures aux dimensions de plus en plus réduites. Le travail mené dans cette thèse s'inscrit particulièrement dans le cadre précis de la scattérométrie. L'objectif de ces travaux consistait à étudier les potentialités des RN, déjà fortement implantés dans le cadre de la scattérométrie, comme outil de détection et d'identification de formes particulières ou de défaut. La récolte de ce type d'information en amont d'une étape de métrologie classique peut permettre de valider certaines hypothèses inhérentes à la caractérisation ou même interrompre, sous conditions établies à l'avance, le procédé de fabrication évitant ainsi une perte de temps et d'argent évidente. La contrainte fixée réside dans le fait que cette inspection doit se faire de manière transparente pour l'expérimentateur. Ainsi, le temps de mesure et de traitement doit être faible, voire instantané. C'est pour cela que l'utilisation des RN a été choisie : d'une part, ils permettent un traitement quasi instantané de la mesure après apprentissage, et d'autre part, ils sont déjà utilisés dans l'étape de caractérisation proprement dite ; ce qui permet d'envisager une continuité dans le traitement de l'information. Ainsi, l'utilisation de la même signature optique a été envisagée.

Pour mener cette étude, nous nous sommes limités à 2 problèmes typiques : la détection de défaut lors d'une étape de photolithographie et l'identification qualitative de la forme de la structure à analyser.

Concernant la première étude, les RN ont été utilisés pour la détection de couche résiduelle de résine. Deux types de profils de forme trapézoïdale avec et sans couche résiduelle ont ainsi été définis. Un « clustering » des signatures simulées par MMMFE des deux types de profils a été effectué à l'aide des cartes SOM. La visualisation dans l'espace 2D des SOM a montré la possibilité de différencier les signatures utilisées. La méthode mise en place à partir des cartes ainsi construites a permis de détecter sans ambiguïté des couches de résine d'une hauteur supérieure à 10 nm. Ensuite, nous avons effectué une classification supervisée avec les classifieurs LVQ et PMC. Le PMC affiche les meilleurs résultats avec une épaisseur limite théorique de 4.5 nm. Un résultat expérimental sur un échantillon fabriqué par nano impression de période 750 nm constitué d'une couche de résine NEB22 déposée sur un substrat de Si possédant une couche résiduelle conséquente de 100 nm a été efficacement détecté par la classification neuronale.

En seconde application, les R.N sont utilisés pour identifier à partir des signatures le type de profil géométrique caractérisant la forme de la structure périodique. Nous avons fait préalablement une étude visant à montrer que le choix du modèle géométrique dans une étape de caractérisation conditionne la qualité des résultats. Nous avons insisté sur les indicateurs qui peuvent guider l'expérimentateur dans la confiance à accorder au résultat (erreur calculée entre la mesure et les signatures reconstruites, résultats en dehors du domaine d'entraînement...). Dans certains cas, aucune information n'est disponible et il est difficile d'accorder du crédit à un résultat. La solution qui consiste à construire un R.N fonctionnant sur un modèle de profil générique incluant un nombre limité de profils élémentaires a été étudiée. Ce modèle s'est avéré efficace sur des formes simples, mais il se trouve limité par le nombre fini de géométries utilisées. Par ailleurs, il ne peut être utilisé efficacement que dans des plages de variation assez restreintes autour de la valeur nominale. De plus, les performances en termes de précisions sont réduites et ce d'autant plus que le nombre de paramètres décrivant la forme augmente. Ainsi, un nouveau « clustering » des signatures de différents profils géométriques a permis de confirmer une certaine discrimination en fonction de la forme de la structure. La différence avec l'étude précédente réside dans le fait que cette fois-ci, le nombre de regroupements est supérieur à 2. Il ne s'agit plus d'une validation/rejet, mais d'un choix de profils à faire parmi plusieurs motifs. Ces inconvénients peuvent être évités dans ce cadre, la classification effectuée par un classifieur PMC s'est avérée encore une fois beaucoup plus performante que le LVQ affichant une exactitude théorique de plus de 99 %. Le PMC est arrivé à séparer efficacement jusqu'à 4 classes de formes de profil différents à partir des signatures optiques. Ce type de classifieur a montré également une robustesse à un niveau de bruit supérieur à celui de la mesure. Un échantillon composé d'une couche de résine IM50 déposée sur du BARC le tout reposant sur un substrat de Si a été examiné. Le résultat de la caractérisation par PMC est cohérent avec la caractérisation AFM.

L'ensemble de ces résultats confirme que le RN, en particulier le PMC, peut répondre à la problématique de détection/identification fixée initialement. Cependant, la validation expérimentale doit être menée sur une plus large série de mesures sur des échantillons présentant différentes configurations.

Dans le même ordre d'idée, différentes études peuvent être menées à court terme : en effet, nous avons observé au cours de nos travaux que la prise en charge uniquement de la partie UV du spectre de mesure composant la signature améliore significativement les résultats. Une étude d'optimisation de la signature doit être menée afin d'en extraire les informations

recherchées. A noter que cette étude doit tenir compte du fait que cette même signature peut être utilisée pour l'étape de caractérisation ultérieure. Ensuite, les résultats concernant les classifieurs LVQ peuvent également être améliorés en utilisant d'autres techniques d'apprentissage qui n'ont pas été exploités dans cette thèse (par exemple LVQ2). Enfin, pour s'assurer de la qualité des résultats des RN, une confrontation avec les méthodes classiques de classification doit être faite.

Cette thèse contribue à la construction d'une solution globale de caractérisation scattérométrique exclusivement neuronale. La figure III. 53 représente l'architecture globale d'un tel système dédié dans ce cas à l'identification de profil, mais on peut envisager d'autres applications par exemple la détection et la prise en compte dans l'étape de caractérisation d'un changement d'indice important. Tout ceci passe bien entendu par une bonne connaissance du problème direct pour connaître l'influence du phénomène à détecter sur la signature optique.

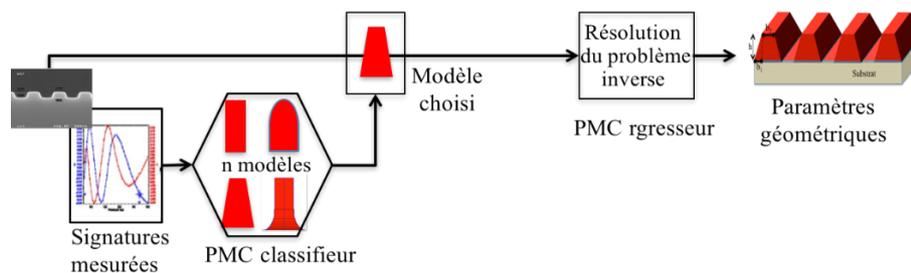


Figure III. 53 : Caractérisation scattérométrique neuronale globale pour l'identification de profils

Ce type de solution est particulièrement performant dans le cas où l'information doit être traitée rapidement. On pense bien entendu au contrôle en ligne en cours de production. Dans ce cadre, il existe des architectures de RN plus complexes dédiées exclusivement à la modélisation de processus physique dynamique. On pourrait ainsi envisager un système à base de RN qui ne serait plus statique, mais capable d'anticiper le changement le profil en fonction d'une mesure de la déviation des signatures au cours du temps

Bibliographie

- [1] L. R. Edward, «Scanning electron microscopy and x-ray microanalysis,» *PTR Prentice Hall Englewood cliffs, N.J.*, 1993.
- [2] C. L. Gressus, «Microscopie Electronique à Balayage,» *Technique de l'ingénieur PE 865*.
- [3] I. Gereige., «Contribution des réseaux de neurones dans le domaine de l'ellipsométrie : Application à la scatterométrie,» 2008 .
- [4] J. C. Rivoal et C. Fretigny, «Microscopie à force atomique,» *Technique de l'ingénieur, r1394.*.
- [5] L. Azarnouche, «Défis liés à la rugosité des motifs de résine photosensible,» 2012.
- [6] W. Yang, J. Hu, R. L. Webb, R. Korlahalli, D. Shivaprasad, H. Sasano, W. Liu et D. L. Miu, «Line profile and critical dimension monitoring using a normal incidence optical cd metrology,» *IEEE Trans. Semicon. Manuf.*, vol. 17, p. 564 572, 2004.
- [7] A. Hettwer, N. Benesch, C. Schneider, L. Pfitzner et H. Ryssel, «Phi scatterometry for integrated linewidth and process control in dram manufacturing,» *IEEE Trans. Semicon. Manuf.*, vol. 15, pp. 470-477, 2002 .
- [8] K. Minhas, S. A. Coulombe, S. S. H. Naqvi et J. R. McNeil, «Ellipsometric scattérométry for the metrology of sub 0.1Um linewidth structures,» *App. Opt.*, vol. 37 , pp. 5112-5115, 1998.
- [9] I. Kallioniemi, J. Saarinen et E. Oja, «Optical scatterometry with neural network model for nondestructive measurement of submicron features,» chez *Proc. SPIE*, 1999.
- [10] I. Kalliomemi, J. Saarinen et E. Oja, «Optical Scatterometry of Subwavelength Diffraction Gratings: Neural-Network Approach,» *Applied Optics*, vol. 37, n° 125, pp. 5830- 5835, 1998.
- [11] S. Robert, A. M. Ravaud, S. Reynaud, S. Fourment, F. Carcenac et P. Arguel, «Experimental characterization of subwavelength diffraction gratings by an inverse-scattering neural method,» *Journal of the Optical Society of America A*, vol. 19, n° 112, p. 2394- 2402, 2002.
- [12] J. Bischoff, H. Truckenbrodt et J. Bauer, «Diffraction analysis based characterization of very fine gratings,» chez *Proc. SPIE*, 1997.
- [13] J. Bischoff, J. Bauer, U. Haak, L. Hutschenreuther et H. Truckenbrodt, «Optical scatterometry of quarter-micron patterns using neural regression,» chez *Proc. SPIE*, 1998.
- [14] N. Benesc, A. Hettwer, C. Schneider, P. L. et R. Heneir, «Phi- scatterometry for integrated linewidth control in DRAM manufacturing,» *Semiconductor Manufacturing Symposium*, pp. 129-132, 2001.
- [15] B. S. Stutzman, H.-T. Huang et F. L. T. Jr., «Two-channel spectroscopic reflectometry for in situ monitoring of blanket and patterned structures during reactive ion

- etching,» *Journal of Vacuum Science & Technology B*, vol. 18, n° 16, p. 2785-2793, 2000.
- [16] J. Hazart, G. Grand, P. Thony, D. Herisson, S. Garcia et O. Lartigue, «Spectroscopic ellipsometric scatterometry: sources of errors in critical dimension control,» chez *Proceedings of SPIE*, 2003.
- [17] I. Gereige, S. Robert, G. Granet, D. Jamon et J. R. J., «Rapid Control of submicrometer periodic structures by a neural inversion from ellipsometric measurement,» *Optics Communications*, vol. 278, n° 12, pp. 270-273, 2007.
- [18] X. Niu et S. Yedur, «Specular Spectroscopic Scatterometry in DUV Lithography,» *SPIE Metrology, inspection and Process control for monolithography*, vol. 3677, n° 11999, p. 159-168. .
- [19] R. M. A. Azzam et N. M. Bashra, *Ellipsometry and polarized Light*, 1987.
- [20] B. Kaplan et T. Novikova, «Characterization of bidimensional gratings by spectroscopic ellipsometry and angle-resolved Mueller polarimetry,» *Applied Optics*, vol. 43, n° 16, 2004.
- [21] R. Krukar, S. Prins, M. K. D., G. Peterson, S. Gaspar, J. McNeil, H. Sohail et D. R. Hush, «Using scattered-light modeling for semiconductor critical dimension metrology and calibration,» chez *Proc. SPIE*, 1993.
- [22] C. J. Raymond, M. E. Littau, A. Chuprin et S. Ward, «Comparison of solutions to the scatterometry inverse problem,» chez *Proc. SPIE, 5375, Metrology, Inspection, and Process Control for Microlithography XVIII*, 2004.
- [23] X. Chena, S. Liua, C. Z. b et J. Zhua, «Improved measurement accuracy in optical scatterometry using fitting error interpolation based library search,» *Measurement*, vol. 46, n° 18, p. 2638–2646 , 2013.
- [24] X. Niu, N. Jakatdar, J. Bao et C. Spanos, «Specular spectroscopic scatterometry,» *EEE Trans. Semicond. Manuf. ,* vol. 14, n° 12, may 2001.
- [25] C. G. et B. R., «Genetic algorithm for ellipsometric data inversion of absorbing layers,» *J. Opt. Soc. Am. A*, vol. 17, p. 129–134, 2000.
- [26] S. S. H. Naqvi, R. Krukar, J. McNeil, J. E. Franke, M. N. T., D. M. Haaland, R. A. Gottsho et A. Kornblit, «Etch depth estimation of large period silicon gratings with multivariate calibration of rigorously simulated diffraction profiles,» *J. Opt. Soc. Am. A*, vol. 11, pp. 2485-2493, 1994.
- [27] R. M. Michael, J. R. Christopher, R. H. Ziad, H. N. S. Sohail et R. M. John, «Developed photoresist metrology using scatterometry,» *Proc. SPIE 2196, Integrated Circuit Metrology, Inspection, and Process Control*, vol. 3, n° 147, may 1994.
- [28] L. F. Thompson, «Introduction to Microlithography,» *ACS Symposium*, 1983.
- [29] G. L. Jay, «Nanoimprint Lithography: Methods and Material Requirements,» *Adv. Mater.*, vol. 19, p. 495–513, 2007.
- [30] S. Y. Chou, P. R. Krauss et P. J. Renstrom, «Nanoimprint lithography,» *JVST B*, 1996.
- [31] A. Bazin, «Analyse de procédés de traitement plasma des résines photosensibles à 193 nm pour le développement de technologies CMOS sub-65 nm,» Grenoble, 2012.
- [32] J. Petit, «control dimensionnel sub-micromique utilisant un appareil goniometrique bidimensionnel rapide,» 2005.

-
- [33] S. W. James et R. P. Tatam, «Optical fibre long-period grating sensors: characteristics and application,» *Measurement Science and Technology*, vol. 14, n° 15, March 2003.
- [34] C. Elachi et C. Yeh, «Periodic structures in integrated optics,» *J. Appl. Phys.*, vol. 44, n° 13146, 1973.
- [35] F. Bernoux, «Ellipsométrie – théorie,» *techniques de l'ingénieur*, 2003.
- [36] I. E. Kalyoubi, «Developpement de la technique de scattérométrie neuronale dynamique,» 2015.
- [37] Goodman J.W., *Introduction to Fourier Optics*, Third edition éd., Stanford University, 2005.
- [38] Millar R.F., «The Rayleigh hypothesis and a related least squares solution to scattering problems for periodic surfaces and other scatteres,» *Radio Sci.*, vol. 8, pp. 785-796,, 1973 .
- [39] P. M. Van den Berg , «Refelection by a grating: Rayleigh methods,» *Journal Society of America*, vol. 71, pp. 1224-1229, 1981.
- [40] R. Petit, «Etude théorique de la diffraction d'une onde plane et monochromatique par un réseau métallique infiniment conducteur,» *App. Opt.*, vol. 4, pp. 1551-1554, 1965.
- [41] D. Maystre, «Sur la diffraction d'une onde plane par un réseau métallique de conductivité finie,» *Opt. Comm.*, vol. 6, pp. 50-54, 1972.
- [42] T. Delort et D. Maystre, «Finite-element method for grating,» *journal of optics society of America A*, vol. 10, pp. 2592-2601, 1993.
- [43] G.Tayeb, «The method of fictitious sources applied to diffraction gratings,» *Appl. Comput. Electromag, Soc.*, vol. 9, pp. 90-100, 1994.
- [44] G. Granet et J. Chandezon, «The method of curvilinear coordinates applied to the problem of scattering from surface-relief gratings defined by parametric equations: application to scattering from a cycloidal grating,» *Pure Appl. Opt.*, vol. 6, pp. 727-740, 1997.
- [45] J. Chandezon, D. Maystre et G. Raoult, «A new theoretical method for diffraction gratings and its numerical application,» . *J. Optics*, vol. 11, pp. 235-241, 1980.
- [46] L.Li et C. Haggans, «Convergence of the coupled-wave method for metallic lamellar diffraction gratings,» *J. Opt. Soc. Am. A*, vol. 10, pp. 1184-1189, 1993.
- [47] M. G. Moharam et T. K. Gaylord, «Rigorous coupled-wave analysis of planargrating diffraction,» *J. Opt. Soc. Am. A* , vol. 71, p. 811–818, 1981.
- [48] G. Granet et G. Guizal, «Efficient implementation of the coupled-wave method for metallic lamellar gratings in TM polarization,» *J. Opt. Soc. Am. A*, vol. 13, n° 15, May 1996.
- [49] K. Madsen, H. Nielsen et O. Tingleff, *Methods for Non-Linear Least Squares Problems*, Technical University of Denmark, 2004.
- [50] J. Nocedal et J. W. S., *Numerical Optimization*, S. Verlag, Éd., NY, 1999.
- [51] J. A. Nelder et R. Mead, «A simplex method for fonction minimization,» *Computer Journal*, vol. 7, p. 308–313, 1965.
- [52] D. Marquardt, «An Algorithm for Least-Squares Estimation of Nonlinear Parameters,» *SIAM Journal on Applied Mathematics*, vol. 11, n° 12, p. 431–441, June 1963.

-
- [53] L. K., «A method for the solution of certain non-linear problems in least squares.,» *Quarterly of Applied Mathematics*, vol. 2, p. 164–168, 1944.
- [54] H.-T. Huang et F. T. Jr., «Erratum to “Spectroscopic ellipsometry and reflectometry from gratings (Scatterometry) for critical dimension measurement and in situ, real-time process monitoring”,» *Thin Solid Films*, vol. 468, pp. 339-346, 2004.
- [55] J. M. Holden, T. Gubiotti, W. A. McGaham, M. Dusa et T. Kiers, «Normal-incidence spectroscopic ellipsometry and polarized reflectometry for measurement and control of photoresist critical dimension,» chez *Proc. SPIE*, 2002.
- [56] J. H. Holland, «Adaptation in natural and artificial systems,» 1975.
- [57] S. Robert, «Caractérisation optique de réseaux de diffraction submicroniques par des techniques génétique et neuronale,» 2003.
- [58] M.Elkodadi, «Développement des techniques de scatterométrie en temps réel pour le suivi des procédés de gravure plasma,» 2010.
- [59] S. Soulan, «Développement de la scattérométrie dynamique pour le suivi en temps réel de procédé. Application à la microélectronique,» 2008.
- [60] S. Soulan, M. Besacier, P. Schiavone et T. Leveder, «In-line etching process control using dynamic scatterometry,» chez *Proc. SPIE 6617, Modeling Aspects in Optical Metrology*, 2007.
- [61] S. Soulan, B. M., T. Leveder et P. Schiavone, «Real-time profile shape reconstruction using dynamic scatterometry,» chez *Proc. SPIE 6518, Metrology, Inspection, and Process Control for Microlithography XXI*, April 05, 2007.
- [62] M. E. Kodadi, S. Soulan, M. Besacier et P. Schiavone, «Real time scatterometry for profile control during resist trimming process,» *Journal of Vacuum Science & Technology B*, vol. 27, n° 16, pp. 3232-3237, 2009.
- [63] M. E. Kodadi, S. Soulan, M. Besacier et P. Schiavone, «Resist trimming etch process control using dynamic scatterometry,» *Microelectronic Engineering*, vol. 86, n° 14, pp. 1040-1042, 2009.
- [64] S. Robert, A. Mure-Ravaud et D. Lacour, «Characterization of optical diffraction gratings by use of a neural method,» *Journal of the Optical Society of America A*, vol. 19, n° 11, p. 24-32, 2002.
- [65] I. Gereige, S. Robert et G. Granet, «Optimal architecture of a neural network for a high precision in ellipsometric scatterometry,» chez *Proc. SPIE*, 2007.
- [66] S. Wei et L. Lifeng, «Measurement of photoresist grating profiles based on multiwavelength scatterometry and artificial neural network,» *Applied Optics*, vol. 47, n° 113, p. 2524-2532, 2008.
- [67] S. Robert, A. M. Ravaud, M. Yacoub, S. Thiria et F. Badran, «Neural selection of the optimal optical signature for a rapid characterization of a submicrometer period grating,» *Optics communications*, vol. 238, n° 14-6, pp. 215-228, August 2004.
- [68] S. Robert et A. Mure-Ravaud, «Control of the homogeneity of an optical grating by a neural characterization,» *Optical Engineering*, vol. 44, n° 13, pp. 033601-033601,, 2005.
- [69] Y. Battie, S. Robert, I. Gereige, D. Jamon et M. Stchakovsky, «Demonstration of the feasibility of a complete ellipsometric characterization method based on an artificial neural network,» *Applied Optics*, vol. 48, n° 128, pp. 5318-5323, October 2009 .

-
- [70] R. O. Duda, P. E. Hart et G. S. David, *Pattern Classification*, Second edition, 2001.
- [71] Jain et Dubes, *Algorithm in clustering data*, 1988.
- [72] R. O. Duda et P. E. Hart., *Pattern classification and science analysis*, John Wiley,, 1973.
- [73] J. Benzécri, *L'Analyse des Données tome 1 : Taxinomie et tome 2 : L'Analyse des Correspondances*, Bordas, Éd., Paris, 1973.
- [74] I. T. Jollife, «Principal component analysis,» *Springer*, 1986.
- [75] M. Crucianu, *Méthodes factorielles pour l'analyse des données : méthodes linéaires et extensions non-linéaires*, Hermes Science Publications , 2004.
- [76] Sammon Jr.J.W., «A nonlinear mapping for data structure analysis,» *IEEE Transactions on Computers* , vol. 18, pp. 401-409 , 1969.
- [77] H. C. Saad, «Cartes auto-organisatrice pour la classification de donnée symbolique mixtes de données de type intervalle et données discretisées,» 2014.
- [78] E. Parzen, «On estimation of a probability density function and mode,» *Ann. Math. Stat.*, vol. 33, pp. 1065-1076, 1962.
- [79] B. Christopher, *Pattern recognition and machine learning*, Computer Science, Springer, , 2007 .
- [80] T. Cover et P. Hart, «Nearest Neighbour Pattern Classification,» *IEEE Transactios Information theory*, vol. 13, pp. 1-27, 1967.
- [81] C. Delannoy, «Un algorithme rapide de la recherche de plus proche voisin,» *RAIRO Informatique*, vol. 14, n° 13, pp. 275-286.
- [82] K. Fukunaga et P. Narendra, «A branch and bound algorithm for computing k-nearest neighbours,» *IEEE transactions on computer* , July 1975 .
- [83] J. H. Friedman, J. L. Bentley et R. A. Finkel, «An algorithm for finding best matches in logarithmic expected time.,» *ACM transaction on Software*, vol. 3, n° 13, 1977 .
- [84] B. Kim et S. B. Park, «A fast k Nearest neighbor finding algorithm based on the ordered partition,» *IEEE Transaction on pattern analysis and machine intelligence PAMI*, vol. 8, n° 16, November 1986 .
- [85] J. H. Friedman, F. Baskettand et L. Shustek, «An algorithm for finding Nearest Neighbors,» *IEEE transaction on computer*, vol. 24, pp. 1000-1006, 1975 .
- [86] T. Yunk, «A technique to identify Nearest neighbors,» *IEEE transactions on systems, Man and cybernetics*, vol. 6, pp. 678-683, 1976.
- [87] T. P. E. Hart, «The condensed Nearest neighbor rule,» *IEEE Transaction information theory*, 14 May 1968.
- [88] I. K. Sethi, «A fast Algorithm for recognizing Nearest Neighbors,» *IEEE transactions on systems, Man and cybernetics*, vol. 11, n° 13, 1981.
- [89] L. Miclet et M. Dabouz, «Approximative fast nearest neighbour recognition,» *Pattern recognition letters*, vol. 1, pp. 277-285, 1983.
- [90] B. Boser, I. Guyon et V. Vapnik, «An training algorithm for optimal margin classifiers,» chez *Proc. 5th Annual Workshop on Computational Learning Theory*,, 1992.
- [91] D. P. Bertsekas, *Non linear programming*, 2 éd., Athena scientific, 1999.
- [92] M. O. Stitson, J. A. E. Weston, A. Gammerman, V. Vovk et V. Vapnik, «Theory Of Support Vector Machines,» 1996.

-
- [93] S. Essid, «Classification automatique des signaux audio-fréquences : reconnaissance des instruments de musique.,» Paris, 2005.
- [94] R. Courant et D. Hilbert, «Methods of mathematical physics,» *Interscience Publishers*, 1954.
- [95] V. Vapnick, «The nature of statistical learning theory,» *Springer-Verlag*, 1995.
- [96] G. Zhu et D. G. Blumberg, «Classification using ASTER data and SVM algorithms;: The case study of Beer Sheva, Israel,» *Remote sensing of Environment*, vol. 80, n° 12, p. 233–240, mai 2002.
- [97] Y. Goldberg et M. Elhadad, «splitSVM: fast, space-efficient, non-heuristic, polynomial kernel computation for NLP applications,» chez *Proceeding HLT-Short '08 Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics on Human Language Technologies: Short Papers*, 2008 .
- [98] W. Wang, Z. Xu, W. Lu et X. Zhang, «Determination of the spread parameter in the Gaussian kernel for classification and regression,» *Neurocomputing*, vol. 55, n° 13-4, p. 643–663, October 2003.
- [99] S. S. Keerthi et C.-J. Lin, «Asymptotic Behaviors of Support Vector Machines with Gaussian Kernel,» *Neural Computation*, vol. 15, n° 17, pp. 1667-1689, July 2003.
- [100] V. Vapnik, «The Nature of Statistical Learning Theory.,» *Springer-Verlag*, 1995.
- [101] U.-G. Kreßel, «Pairwise classification and support vector machines,» *Advances in kernel methods : support vector learning,*, p. 255–268, 1999.
- [102] R. F. E. Osuna et F. Girosi., «Training Support Vector Machines : An Application To Face Detection,» chez *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1997.
- [103] A. Kharrat, N. Benamrane, M. b. messaoud et A. Mariem, «Evolutionary Support Vector Machine for Parameters Optimization Applied to Medical Diagnostic,» chez *International Conference on Computer Vision Theory and Applications (VISAPP)*, 2011.
- [104] J. Zhu, S. Liu, C. Zhang, X. Chen et Z. Dong, «Identification and reconstruction of diffraction structures in optical scatterometry using support vector machine method,» *JM3*, vol. 12, n° 11, 2013.
- [105] Y. L. Cun, «Modèles connexionnistes de l'apprentissage,» *Université Pierre et Marie C, Paris Curie VI,*, 1987.
- [106] M. Culloch et W. Pitts, «A logical calculus of ideas immanent in nervous Activity,» *bulletin of mathematical biophysics*, vol. 5, pp. 115-133, 1943.
- [107] D. Hebb, «The Organization of Behavior,» *Wiley* :, 1949.
- [108] F. Rosenblatt, «Principles of neurodynamics:Perceptrons and the theory of brain mechanisms,» *Spartan*, 1962.
- [109] M. L. Minsky et S. A. Papert, «Perceptrons,» *Cambridge, MIT press*, 1969.
- [110] T. Kohonen, «Self Organized Formation of Topologically correct Feature Maps,» *Biological Cybernetics*, vol. 46, pp. 59-69, 1982.
- [111] J. Hopfield, «neural Networks and physical systems with Emergent Collective Computational Abilities,» *Proc of the Natl Acad. Sci.*, vol. 79, pp. 2554-2558, 1982.

- [112] J. Hopfield, «neuron with graded response have collective computational properties like those of two state neuron,» *Proc. Of the Natl. Acad. Sci.*, vol. 81, pp. 3088-3092, 1984(4e edition).
- [113] Y. LeCun, B. Boser, J.S. Denker, D. Henderson, R. E. Howard, W. Hubbard et L. D. Jackel, « Backpropagation applied to handwritten zip code recognition. *NeuralComputation*,» vol. 1, pp. 541-551, 1989.
- [114] B. GOSELIN, Application de réseaux de neurones artificiels a la reconnaissance automatique de caractères manuscrits, Faculté polytechnique de Mons, 1996.
- [115] E. Radoi, «Contribution à la reconnaissance des objets 3D à partir de leur Signature Ferromagnétique,» 1999.
- [116] I. Rivals, D. Canas, L. Personnaz et G. Dreyfus, «Modeling and control of mobile robots and intelligent vehicles by neural networks,» chez *Proceedings of the IEEE conference on Intelligent Vehicles*, 1994.
- [117] Z. Chalabi, A. Boudjemaoui, L. Saadia et N. Berrached, «Detection et classification automatique d'arythmie cardiaques,» chez *5th International Conference: Sciences of Electronic, Technologies of Information and Telecommunications, SETIT, TUNISIA*, 2009.
- [118] R. S. Naoum et Z. N. A. Sultani, «Learning Vector Quantization (LVQ) and k Nearest Neighbor for Intrusion Classification,» *World of Computer Science and Information Technology Journal (WCSIT) ISSN: 2221 0741* , vol. 2, n° 13, pp. 105-109, 2012 .
- [119] X. Tong, Z. Wang et H. Yu, «A research using hybrid RBF/Elman neural networks for intrusion detection system secure model,» *Computer Physics Communications* , vol. 180, p. 1795–1801, 2009.
- [120] M. W. Craven et J. W. Shavlik, «Using neural networks for data mining,» *Future Generation Computer Systems*, vol. 13, pp. 211-229, 1997.
- [121] D. Enke et S. Thawornwong, «The use of data mining and neural networks for forecasting stock market returns,» *Expert Systems with applications*, vol. 29, n° 14, pp. 927-940, 2005.
- [122] H. WU, «Contribution to Engine-out Aircraft Trajectory Management and Control,» Avril 2013.
- [123] P. Roussel, F. Moncet, B. Barrieu et A. Viola, «Modélisation d'un processus dynamique à l'aide de réseaux de neurones bouclés. Application à la modélisation de la relation pluie-hauteur d'eau dans un réseau d'assainissement et à la détection de défaillances de capteurs,» *Innovative technologies in urban drainage* , vol. 1, pp. 919- 926,.
- [124] B. Widrow et S. Sterns, «Adaptive Signal Processing,» *Prentice-Hall*, 1985.
- [125] B. Widrow et H. M.E., «Adaptative switching circuit,» *1960 IRE WESCON Convention Record*, p. 96–104, 1960.
- [126] S. Grossberg, «Studies of the Mind and Brain,» *Dordrecht, Holland, Reidel Press* , 1982.
- [127] G. Cybenko, «Approximation by superpositions of sigmoïdal function,» *Mathematics of Control, Signals and Systems*, vol. 2, pp. 303_314,, 1989.
- [128] K. Hornik, M. Stinchcombe, H. White et P. Auer, «Degree of approximation results for feedforward networks approximating unknown mappings and their derivatives,» *Neural Computation*, vol. 6, p. 1262 1275, 1994.

-
- [129] D. E. Rumelhart, G. E. Hinton et R. J. Williams, «Learning internal representations by error back-propagation,» *Parallel Distributed Processing : Explorations in the Microstructure of Cognitions*, , p. 318–362, 1986.
- [130] S.Saarinen, R.Bramley et G. Cybenko, «lil-conditioning in neural network training problems,» *SIAM J. Sci. Stat. Comp.*, vol. 14, pp. 693-714, 1993.
- [131] G. Monari, «Sélection de modèles non-linéaires par leave one out; étude théorique et application des réseaux de neurones au procédé de soudage par points,» Paris, 1999.
- [132] M. Stone, «Cross-validatory choice and assessment of statistical predictions,» *Journal of the Royal Statistical Society B*, vol. 8, pp. 349-359, 1974.
- [133] T. Kohonen, «the Self Organising Map,» *Proceedings of the IEEE*, vol. 78, n° 19, Septembre 1990 .
- [134] H. Ritter, T. Martinetz et S. Schulten, « Neural Computation and Self- Organizing Maps : An Introduction,» *Addison-Wesley, Reading, Massachussets,*, 1992..
- [135] H. Ritter et S. Schulten, «Kohonen’s Self-Organizing Maps : exploring their computational capabilities.,» chez *IEEE International Conference on Neural Networks*, July 1988.
- [136] T. Kohonen, «Learning Vector Quantization,» *Neural networks*, vol. 1, n° 11, p. 303, 1988.
- [137] T. Kohonen, *Self organising and Associative memory* 3rd Ed., Springer, Verlag, Germany, 1989.
- [138] A. Vuckovic, V. Radivojevic, A. C. N. Chen et D. Popovic, «Automatic recognition of alertness and drowsiness from EEG by an artificial neural network,» *Medical engineering & Physics*, vol. 24, n° 15, pp. 349-360, juin 2002.
- [139] Z. H. Mamar, «Analyse TempsEchelle et Reconnaissance des Formes pour le Diagnostic du Système de Guidage d’un Tramway sur Pneumatiques,» 2008.
- [140] A.Szucs et al, «Advanced OPC Mask-3D and Resist-3D modeling,» chez *Proc. SPIE 9052, Optical Microlithography XXVII,*, 2014.

ANNEXE A : Compléments à l'étude de défaut de modélisation de profil

➤ Hypothèse du modèle créneau arrondi au sommet

Le modèle PMC_c est constitué de 10 neurones sur la couche cachée et 3 neurones de sorties correspondant aux valeurs r , CD et h du profil arrondi. Le R.N est entraîné avec 3000 signature dans les plages de variation suivante :

$$0 \text{ nm} < r < 45 \text{ nm}$$

$$30 \text{ nm} < CD < 90 \text{ nm}$$

$$40 \text{ nm} < h < 130 \text{ nm}$$

La figure A.1 montre la courbe de régression obtenue pour les données de test. Les performances obtenues sur les paramètres sont $E_{\text{rms}}^r = 2.5 \text{ nm}$, $E_{\text{rms}}^{CD} = 0.7 \text{ nm}$, et $E_{\text{rms}}^h = 0.8 \text{ nm}$. Le tableau A.1 montre le résultat de caractérisation de 4 profils particuliers.

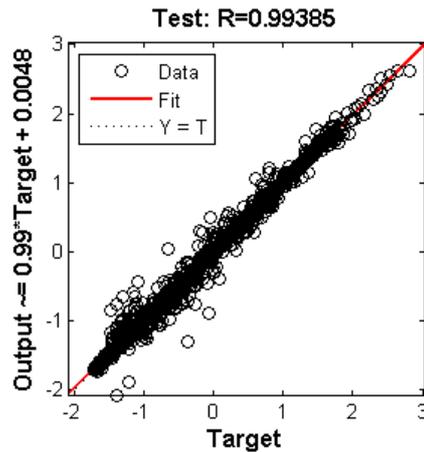


Figure A.1: Performance du PMC_c à l'issue de l'apprentissage sur les paramètres r , CD et h normalisés : courbe de régression évaluée sur $N_{\text{rest}} = 450$ échantillons.

Tableau A.1 Caractérisation par PMC entraîné sur un profil de type créneau arrondi pour 4 échantillons particuliers.

Profil Créneau Arrondi	Estimation PMC_C
Exemple 1 $r=6.1 \text{ nm}$, $CD=69.6 \text{ nm}$, $h=117.5 \text{ nm}$	$r_{\text{est}}=5.3\pm 2.5$, $CD_{\text{est}}=69.5\pm 0.7$; $h_{\text{est}}=117.5\pm 0.8$; $E_{\text{Is,lc}}=0.0009$
Exemple 2 $r=18.7 \text{ nm}$, $CD=42.2 \text{ nm}$, $h=84.2 \text{ nm}$	$r_{\text{est}}=19.3\pm 2.5$; $CD_{\text{est}}=41.8$.; $h_{\text{est}}=84.5\pm 0.8$; $E_{\text{Is,lc}} = 0.0066$
Exemple 3 $r=21.5 \text{ nm}$, $CD=87.5 \text{ nm}$, $h=57.3 \text{ nm}$	$r_{\text{est}}=22\pm 2.5$; $CD_{\text{est}}=88.1\pm 0.7$; $h_{\text{est}}=57.6\pm 0.8$; $E_{\text{Is,lc}}=0.004$
Exemple 4 $r=26.6 \text{ nm}$, $CD=53.8 \text{ nm}$, $h=77.6 \text{ nm}$	$r_{\text{est}}=26.3\pm 2.5$, $CD=53.2\pm 0.7$; $h=77.3\pm 0.8$;

	$E_{Is,lc}=0.01$
--	------------------

Le tableau A.2 montre le résultat de caractérisation pour 4 profils particuliers de type créneau estimés par le PMC_C et la figure A.2 affiche le résultat concernant l'exemple 1.

Tableau A.2 : Caractérisation par un PMC entraîné sur un modèle de profil de type créneau arrondi pour 4 échantillons particuliers de forme de profil créneau.

Profil Créneau	Estimation PMC_C
Exemple 1 : CD=69.8 nm h=41.2 nm	$r_{est}=4.4\pm 2.5$; $CD_{est}=69.9\pm 0.7$; $h_{est}=41.4\pm 0.8$; $E_{Is,lc}=0.001$
Exemple 2: CD=37.5 nm h=101.5 nm	$r_{est}=2.9\pm 2.5$; $CD_{est}=37.6\pm 0.7$; $h_{est}=101.3\pm 0.8$; $E_{Is,lc}=0.010$
Exemple 3 : CD=68.0 nm h=65.3 nm	$r_{est}=3.7\pm 2.5$; $CD_{est}=67.9\pm 0.7$; $h_{est}=65.2\pm 0.8$; $E_{Is,lc}=0.006$
Exemple 4: CD=54 nm, h=100.6 nm	$r_{est}=0.1\pm 2.5$; $CD_{est}=54\pm 0.7$; $h_{est}=100.7\pm 0.8$; $E_{Is,lc}=0.002$

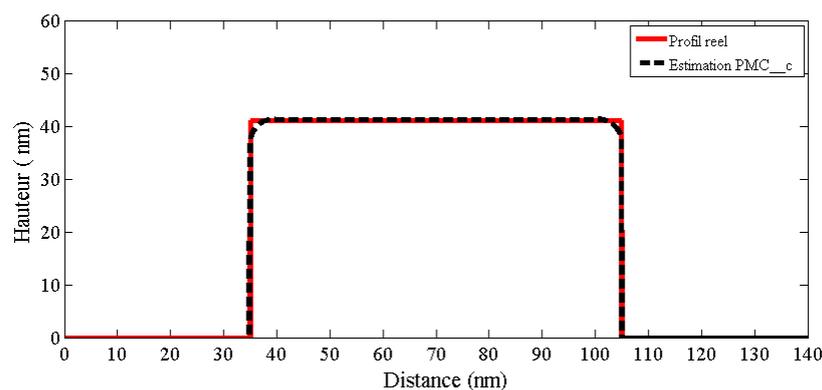


Figure A.2 Représentation graphique du profil créneau de l'exemple 1 (CD=69.8 nm, h=41.2 nm) du tableau A.2 (rouge) et du profil estimé ($r_{est}=4.4$ nm; $CD_{est}=69.9$ nm ; $h_{est}=41.4$ nm) par PMC_c (noir)

Le tableau A.3 montre le résultat de caractérisation de 4 profils particuliers de type trapèze estimés par le PMC_C et la figure A.3 illustre le résultat de l'exemple 2 du tableau.

Tableau A.3 : Caractérisation par un PMC entraîné sur un modèle de profil de type créneau arrondi pour 4 échantillons particuliers de forme trapézoïdale.

Profil Trapèze	Estimation PMC_C
Exemple 1 $b_1=4.1$ nm, $b_2=36.6$ nm, $h=101.7$ nm	$r_{est}=16\pm 2.5$; $CD_{est}=40\pm 0.7$; $h_{est}=105.6\pm 0.7$; $E_{Is,lc}=0.06$
Exemple 2 $b_1=10.2$ nm, $b_2=45.2$ nm, $h=61.2$ nm	$r_{est}=27.5$; $CD_{est}=56.2\pm 0.7$; $h_{est}=63.4\pm 0.8$; $E_{Is,lc}=0.01$

Exemple 3: $b_1=14.6$ nm, $b_2=50.5$ nm, $h=117.5$ nm	$r_{est}=17.4\pm 2.5$; $CD_{est}=60.5\pm 0.7$; $h_{est}=111.7\pm 0.8$; $E_{Is,lc}$ 0.07
Exemple 4 $b_1=16.9$ nm, $b_2=81.8$ nm, $h=91.5$ nm	$r_{est}=22.2\pm 2.5$; $CD_{est}=94\pm 0.7$; $h_{est}=82\pm 0.8$; (CD hors plage)

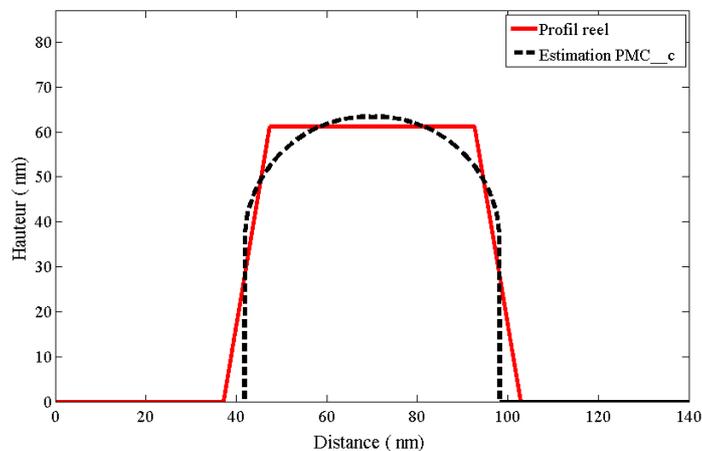


Figure A.3: Représentation graphique du profil trapèze de l'exemple 2 ($b_1=10.2$ nm, $b_2=45.2$ nm, $h=61.2$ nm) du tableau A.3 (rouge) et du profil estimé ($r_{est}=27.5$, $CD_{est}=56.2\pm 0.7$, $h_{est}=63.4\pm 0.8$) par PMC_c (noir)

Enfin, le tableau A.4 montre le résultat de caractérisation de 4 profils particuliers de type sinus estimés par le PMC_C .

Tableau A.4 : Caractérisation par un PMC entraîné sur un modèle de profil de type créneau arrondi pour 4 échantillons particuliers de forme sinusoïdale.

Profile sinus	Estimation PMC_C
Exemple 1: $b=76.9$ nm, $h=86.5$ nm	$r_{est}=55.6\pm 2.5$; $CD_{est}=55.2\pm 0.7$; $h_{est}=83.1\pm 0.8$ (inexistant)
Exemple 2: $b=50.9$ nm, $h=99.3$ nm	$r_{est}=53.1\pm 2.5$; $CD_{est}=24.8\pm 0.7$; $h=100\pm 0.8$ (inexistant)
Exemple 3: $b=86.9$ nm, $h=43.2$ nm	$r_{est}=48.5\pm 2.5$; $CD_{est}=66.5\pm 0.7$; $h_{est}=42.2\pm 0.8$ (inexistant)
Exemple 4: $b=52.6$ nm, $h=75.7$ nm	$r_{est}=78.8\pm 2.5$; $CD_{est}=33.9\pm 0.7$; $h_{est}=83.8\pm 0.8$ (inexistant)

➤ Hypothèse du modèle sinusoïdale

Le modèle PMC_d est constitué de 10 neurones sur la couche cachée et 2 neurones de sorties correspondant aux valeurs h et b du profil sinus. Le R.N est entraîné avec 3000 signatures dans les plages de variation suivante :

$$30 \text{ nm} < b < 90 \text{ nm}$$

$$40 \text{ nm} < h < 130 \text{ nm}$$

L'apprentissage du PMC_d est effectué avec succès (figure A.4). Les performances obtenues sur les paramètres sont : $E_{\text{rms}}^b = 0.043 \text{ nm}$ et $E_{\text{rms}}^h = 0.058 \text{ nm}$.

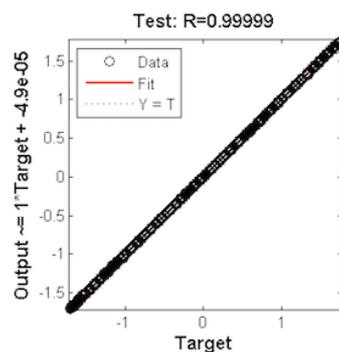


Figure A.4: Performance du PMC_d à l'issue de l'apprentissage sur les paramètres b et h normalisés : courbe de régression évaluée sur $N_{\text{test}} = 450$ échantillons.

Le tableau A.5 présente les résultats de 4 profils particuliers de type sinus caractérisés par PMC_d.

Tableau A.5 : Caractérisation par un PMC entraîné sur un modèle de profil de type sinus pour 4 échantillons particuliers.

Profil sinus	Estimation PMC_d
Exemple 1: b=76.9 nm, h=86.5 nm	$b_{\text{est}}=76.9$, $h_{\text{est}}=86.5$ $E_{\text{Is,lc}}=0.000$
Exemple 2: b=50.9 nm, h=99.3 nm	$b_{\text{est}}=50.9$, $h_{\text{est}}=99.3$ $E_{\text{Is,lc}}=0.0004$
Exemple 3: b=86.9 nm, h=43.2 nm	$b_{\text{est}}=86.8$, $h_{\text{est}}=43.2$ $E_{\text{Is,lc}}=0.0003$
Exemple 4: b=52.6 nm, h=75.7 nm	$b_{\text{est}}=52.5$, $h_{\text{est}}=74.7$ $E_{\text{Is,lc}}=0.0002$

Le tableau A.6 présente les résultats de 4 profils de type créneau particulier caractérisés par le PMC_d et la figure A.5 illustre le résultat de l'exemple 4 du tableau.

Tableau A.6 Caractérisation par un PMC entraîné sur un modèle de profil de type sinus pour 4 échantillons particuliers de forme créneau.

Profil Créneau	Estimation PMC_d
Exemple 1 : CD=69.8 nm h=41.2 nm	$b_{\text{est}}=104.9$, $h_{\text{est}}=49$ (b hors plage)

Exemple 2: CD=37.5 nm h=101.5 nm	$b_{est}=72, h_{est}=124.8 E_{Is,lc}=0.200$
Exemple 3 : CD=68.0 nm h=65.3 nm	$b_{est}=107.5, h_{est}=85.2$ (b hors plage)
Exemple 4: CD=54 nm, h=100.6 nm	$b_{est}=77.4, h_{est}=122.5 E_{Is,lc}= 0.12$

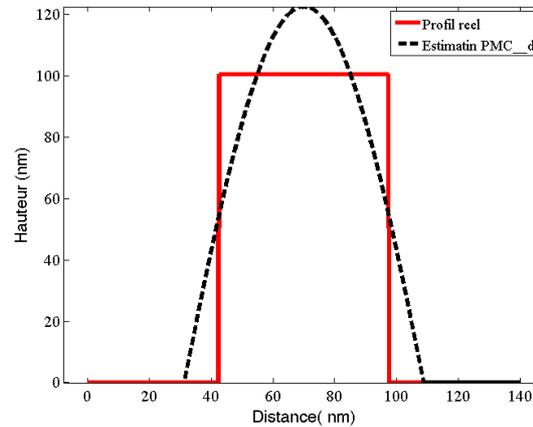


Figure A.5: Représentation graphique du profil trapèze de l'exemple 4 (CD=54 nm, h=100.6 nm) du tableau A.6 (rouge) et du profil estimé ($b_{est}=77.4, h_{est}=122.5$) par PMC_d (noir)

Le tableau A.7 présente les résultats de 4 profils particuliers de type créneau arrondi caractérisés par PMC_d.

Tableau III.7 Caractérisation par un PMC entraîné sur un modèle de profil de type sinus pour 4 échantillons particuliers de forme créneau arrondi.

Profil Créneau Arrondi	Estimation PMC_d
Exemple 1: r=6.1 nm, CD=69.6 nm, h=117.5 nm	$b_{est}=86.5, h_{est}=167.8, E_{Is,lc}=0.3400$
Exemple 2: r=18.7 nm, CD=42.2 nm, h=84.2 nm	$b_{est}=69.2, h_{est}=108.3, E_{Is,lc}= 0.0546$
Exemple 3: r=21.5 nm, CD=87.5 nm, h=57.3 nm	$b_{est}=135.8, h_{est}=73.6, E_{Is,lc}=0.09$ (b hors plage)
Exemple 4: r=26.6 nm, CD=53.8 nm, h=77.6 nm	$b_{est}=80.4, h_{est}=90.6, E_{Is,lc}=0.023$

Le tableau A.8 présente les résultats de 4 profils particulier de type créneau caractérisé par PMC_d et la figure A.6 illustre le résultat de l'exemple 4 du tableau.

Tableau A. 8 Caractérisation par un PMC entraîné sur un modèle de profil de type sinus pour 4 échantillons particuliers de forme trapézoïdale.

Profil Trapèze	Estimation PMC_d
Exemple 1 $b_1=4.1$ nm, $b_2=36.6$ nm, h=101.7 nm	$b_{est}=66.8, h_{est}=119.7 ; E_{Is,lc}=$

	0.0881
Exemple 2 $b_1=10.2$ nm, $b_2=45.2$ nm $h=61.2$ nm	$b_{est}=84.7$, $h_{est}=71.3$; $E_{Is,lc}=0.0176$
Exemple 3 : $b_1=14.6$ nm, $b_2=50.5$ nm, $h=117.5$ nm	$b_{est}=102$, $h_{est}=138$ (b hors plage)
Exemple 4 $b_1=16.9$ nm, $b_2=81.8$ nm, $h=91.5$ nm	$b_{est}=162$, $h_{est}=105.6$ (b hors plage)

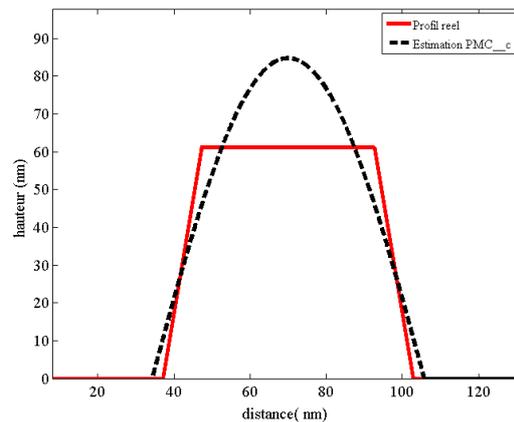


Figure A.6: Représentation graphique du profil trapèze de l'exemple 2 ($b_1=10.2$ nm, $b_2=45.2$ nm $h=61.2$ nm) du tableau A.8 (rouge) et du profil estimé ($b_{est}=84.7$, $h_{est}=71.3$) par PMC_d (noir)

Le tableau A.9 présente les résultats de 3 profils particulier de type créneau arrondi en bas et au sommet estimés par PMC_d . Aucun résultat n'est exploitable.

Tableau A.9 : Caractérisation par un PMC entraîné sur un modèle de profil de type sinus pour 3 échantillons particuliers de forme créneau arrondi en bas et en haut.

Profil Créneau Arrondi (haut et bas)	Estimation PMC_d
Exemple 1 $r_t=7.3$, $CD=80.5$, $h=90.8$, $r_b=2.4$	$b_{est}=157.8$, $h_{est}=135.1$ (hors plage)
Exemple 2 : $r_t=8$, $CD=86.9$, $h=50.3$, $r_b=4.5$	$b_{est}=140.6$, $h_{est}=69.7$ (b hors plage)
Exemple 3: $r_t=18$, $CD=45.3$, $h=50.3$, $r_b=11.6$	$b_{est}=-0.0381$, $h_{est}= 0.1029$ (inexistant)

➤ Modèle générique

Le modèle générique est constitué de 50 neurones sur la couche cachée et 4 neurones de sorties correspondant aux valeurs CD_{top} , CD_{bottom} , h et α du profil générique. Le R.N est entraîné avec 12000 signature. Le tableau A.10 montre le résultat de caractérisation de 4 profils particuliers de forme créneau arrondi au sommet

Tableau A. 10 : Caractérisation par un PMC entraîné sur un modèle de profil générique pour 4 échantillons particuliers de forme créneau arrondi au sommet.

Profil créneau arrondi au sommet	Estimation PMC générique
Exemple 1 $r=6.1$ nm, $CD=69.6$ nm, $h=117.5$	$CD_{top,est}=69.8\pm 1,8$ nm, $CD_{bottom,est}=69.8\pm 1,2$ nm, $h_{est}=117.7\pm 2.8$ nm, $\alpha_{est}=-0.09\pm 0,1$; $E_{Is,lc} = 0.0080$
Exemple 2 $r=18.7$ nm, $CD=42.2$ nm, $h=84.2$ nm	$CD_{top,est}=44.5\pm 1,8$ nm, $CD_{bottom,est}=40.8\pm 1,2$ nm, $h_{est}=85.2\pm 2.8$ nm, $\alpha_{est}=0.84\pm 0,1$ ($CD_{top} > CD_{bottom}$) profil inexistant
Exemple 3 $r=21.5$ nm, $CD=87.5$ nm, $h=57.3$ nm	$CD_{top,est}=86.3\pm 1,8$ nm, $CD_{bottom,est}=89.3\pm 1,2$ nm, $h_{est}=57.7\pm 2.8$ nm, $\alpha_{est}=0.48\pm 0,1$; $E_{Is,lc} = 0.002$
Exemple 4 $r=26.6$ nm, $CD=53.8$ nm, $h=77.6$ nm	$CD_{top,est}=53\pm 1,8$, $CD_{bottom,est}=54.6\pm 1,2$, $h_{est}=74.8\pm 2.8$, $\alpha_{est}=-0.11\pm 0,1$; $E_{Is,lc} = 0.0080$

Le R.N s'adapte à la forme arrondie en calculant une valeur appropriée de α . L'erreur $E_{Is,lc}$ calculée est largement améliorée comparé à celle obtenue par les modèles de PMC_a,t, PMC_b et PMC_d. La précision du modèle générique est induite par le fait de l'existence de la forme arrondie dans la base d'apprentissage.

ANNEXE B : Etude du corpus B de 3 classes

➤ Projection SOM_B

L'apprentissage de la carte (25*25 neurones) avec le corpus B (cf paragraphe III.2.3.1) constitué de 3 classes de profil (C_1 : créneau arrondi au sommet avec un plateau nul, C_2 : trapézoïdal et C_3 : créneau avec des arrondis en bas) a nécessité environ 1 h sur un ordinateur Intel Xeon CPU E5 1620 @ 3.60 GHz de 12 Go RAM.. La carte ainsi obtenue est notée SOM_B. Les figures B .1 montrent la visualisation des signatures pour chaque classe.

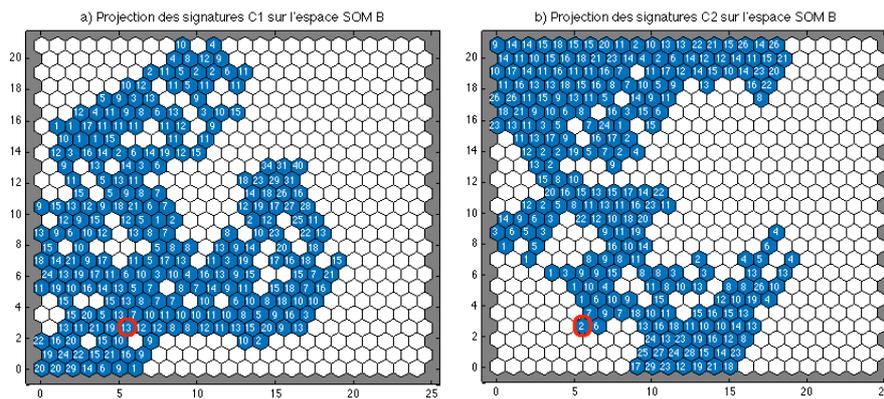
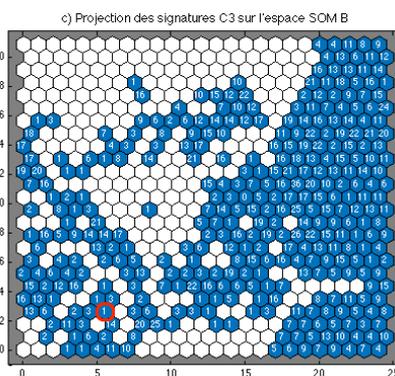
a) Signatures de classe C_1 sur SOM Bb) Signatures de classe C_2 sur SOM Bc) Signatures de classe C_3 sur SOM B

Figure B.1: Résultats de la projection sur la carte SOM_B (25×25 neurones) des 9000 signatures correspondantes aux 3 classes : C_1 a), C_2 b) et C_3 c)

➤ Etude PMC_B

L'architecture optimale du PMC_B entraîné sur le corpus B est constituée de 20 neurones cachés et 3 neurones de sortie. Nous présentons les performances estimées pour 1350 signatures test.

Tableau B.1: Matrice de confusion du PMC_B (20 neurones cachés) calculée sur le corpus test pour l'identification de profils géométriques. La classe C_1 correspond au créneau arrondi (CD=0), C_2 au trapèze et C_3 au créneau à bas arrondi

	C ₁ réelle	C ₂ réelle	C ₃ réelle	Précision
C ₁ estimée	447	5	1	98,7 %
C ₂ estimée	3	445	5	98,2 %
C ₃ estimée	0	0	444	100 %
Sensibilité	99.3 %	99.8 %	98.7 %	Exactitude: 99.0%

Les résultats du PMC B pour les signatures qui ont stimulé le neurone $r \begin{pmatrix} 5.5 \\ 2.5 \end{pmatrix}$ (entouré en rouge sur la figure B.1) sont résumés à titre d'exemple dans les tableaux B.2 à B.4.

Tableau B.2 : Classification PMC_B (20 neurones cachés) des 4 signatures (S_i avec $i=1, \dots, 4$) correspondant à la classe C₁ ayant activé le neurone $r \begin{pmatrix} 5.5 \\ 2.5 \end{pmatrix}$ de la carte SOM_B

Signatures C ₁	S ₁	S ₂	S ₃	S ₄	S ₅	S ₆	S ₇	S ₈
p(C1)	0.9994	0.9988	0.9996	0.9986	0.9954	0.9965	0.9905	0.9996
p(C2)	0.0001	0.0000	0.0001	0.0010	0.0038	0.0003	0.0001	0.0001
P(C3)	0.0005	0.0012	0.0002	0.0005	0.0007	0.0033	0.0094	0.0003

Tableau B.2 : Classification PMC_B (20 neurones cachés) des 2 signatures (S_i avec $i=1,2$) correspondant à une classe C₂ ayant activé le neurone $r \begin{pmatrix} 5.5 \\ 2.5 \end{pmatrix}$ de la carte SOM_B

Signatures C ₂	S ₁	S ₂
p(C1)	0.0283	0.0074
p(C2)	0.9279	0.9692
P(C3)	0.0437	0.0234

Tableau B.4: Classification PMC_B (20 neurones cachés) de la signature S₁ correspondant à une classe C₃ ayant activé le neurone $r \begin{pmatrix} 5.5 \\ 2.5 \end{pmatrix}$ de la carte SOM_B

Signature C ₃	1
p(C1)	0.0247
p(C2)	0.0016
P(C3)	0.9737

ANNEXE C : Etude d'un PMC fonctionnant sur une signature UV

Le PMC entraîné sur le corpus C avec à l'aide de signatures tronquées en dessous de 4.5 eV (soit 32 mesures) est noté PMC_C'. L'architecture optimale est constituée de 50 neurones cachés et 3 neurones de sortie. Le tableau C.1 montre les résultats obtenus pour 1800 signatures composant le corpus de test.

	C ₁ réelle	C ₂ réelle	C ₃ réelle	C ₄ réelle	Précision
C ₁ estimée	450	0	0	1	99,8 %
C ₂ estimée	0	450	0	5	98,9 %
C ₃ estimée	0	0	450	2	99,6 %
C ₄ estimée	0	0	0	442	100%
Sensibilité	100 %	100 %	100	98.2 %	Exactitude : 99.6%

Tableau C.1: Matrice de confusion du PMC_C' (50 neurones cachés) calculée sur le corpus test pour l'identification de profils géométriques. La classe C₁ correspond au créneau arrondi (CD=0), C₂ au sinus, C₃ au trapèze et C₄ au créneau à bas arrondi

Résumé

La miniaturisation des composants impose à l'industrie de la micro-électronique de trouver des techniques de caractérisation fiables rapides et si possible à moindre coût.

Les méthodes optiques telles que la scattérométrie se présentent aujourd'hui comme des alternatives prometteuses répondant à cette problématique de caractérisation. Toutefois, l'ensemble des méthodes scattérométriques nécessitent un certain nombre d'hypothèses pour assurer la résolution d'un problème inverse et notamment la connaissance de la forme géométrique de la structure à tester. Le modèle de structure supposé conditionne la qualité même de la caractérisation.

Dans cette thèse, nous proposons l'utilisation des réseaux de neurones comme outils d'aide à la décision en amont de toute méthode de caractérisation. Nous avons validé l'utilisation des réseaux de neurones dans le cadre de la reconnaissance des formes géométriques de l'échantillon à tester par la signature optique utilisée dans toute étape de caractérisation scattérométrique.

Tout d'abord, le cas d'un défaut lithographique particulier lié à la présence d'une couche résiduelle de résine au fond des sillons est étudié. Ensuite, nous effectuons une analyse de détection de défaut de modèle utilisé dans la résolution du problème inverse. Enfin nous relatons les résultats obtenus dans le cadre de la sélection de modèles géométriques par réseaux de neurones en amont d'un processus classique de caractérisation scattérométrique.

Ce travail de thèse a montré que les Réseaux de neurones peuvent bien répondre à la problématique de classification en scattérométrie ellipsométrique et que l'utilisation de ces derniers peut améliorer cette technique optique de caractérisation.

Mots clé : scattérométrie, ellipsométrie, classification, réseaux de neurones, problème inverse

Abstract

The miniaturization of components in the micro-electronics industry involves the need of fast reliable technique of characterization with lower cost.

Optical methods such as scatterometry are today promising alternative to this technological need. However, scatterometric method requires a certain number of hypothesis to ensure the resolution of an inverse problem, in particular the knowledge of the geometrical shape of the structure under test. The assumed model of the structure determines the quality of the characterization.

In this thesis, we propose the use of neural networks as decision-making tools upstream of any characterization method. We validated the use of neural networks in the context of recognition of the geometrical shapes of the sample under testing by the use of optical signature in any scatterometric characterization process.

First, the case of lithographic defect due to the presence of a resist residual layer at the bottom of the grooves is studied. Then, we carry out an analysis of model defect in the inverse problem resolution. Finally, we report results in the context of selection of geometric models by neural networks upstream of a classical scatterometric characterization process.

This thesis has demonstrated that neural networks can well answer the problem of classification in ellipsometric scatterometry and their use can improve this optical characterization technique.

Keywords: Scatterometry, ellipsometry, classification, neural networks, inverse problem