



HAL
open science

Détection et inspection par réseaux de neurones dans le cadre de la scattérométrie ellipsométrique

Godi Tchere Moustapha

► **To cite this version:**

Godi Tchere Moustapha. Détection et inspection par réseaux de neurones dans le cadre de la scattérométrie ellipsométrique. Optique / photonique. Université Jean Monnet - Saint-Etienne, 2023. Français. NNT : 2023STET0010 . tel-04540660

HAL Id: tel-04540660

<https://theses.hal.science/tel-04540660>

Submitted on 10 Apr 2024

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 : 2023STET006

**THÈSE de DOCTORAT
DE L'UNIVERSITÉ JEAN MONNET DE SAINT-ÉTIENNE**

Membre de l'Université de LYON

**École Doctorale N° 488
Sciences Ingénierie Santé (SIS)**

Discipline : Optique, Photonique et Hyperfréquences

Soutenue publiquement le 20 janvier 2023, par :

MOUSTAPHA Godi Tchéré

**Détection et inspection par réseau de neurones
dans le cadre de la scattérométrie ellipsométrique**

Devant le jury composé de :

Cécile GOURGON	Directrice de recherche, CNRS-LTM, Grenoble	Présidente, Examinatrice
Manuel FLURY	Maître de conférences HDR, INSA Strasbourg/Laboratoire ICube, Strasbourg	Rapporteur
Gérard GRANET	Professeur, Université Clermont Auvergne, Institut Pascal, Clermont-Ferrand	Rapporteur
Bernard BAYARD	Professeur, Université Jean Monnet, LaHC, Saint-Étienne	Directeur de thèse
Stéphane ROBERT	Maître de conférences, Université Jean Monnet, LaHC, Saint-Étienne	Co-directeur

Dédicace

À ma grande famille

À tous mes amis et connaissances

Remerciements

Ce travail est effectué au Laboratoire Hubert Curien CNRS 5516 de l'Université Jean Monnet de Saint-Étienne. Ce fut une expérience particulièrement enrichissante sur le plan professionnel que personnel.

Je tiens tout d'abord à remercier Monsieur Bernard Bayard, mon directeur de thèse pour l'aide précieuse qu'il a su m'apporter et pour la confiance qu'il m'a accordé tout le long de la thèse. Ses remarques pertinentes m'ont permis d'exploiter au mieux le problème rencontrés.

Je tiens également à remercier Monsieur Stéphane Robert, mon encadrant. J'exprime ma reconnaissance pour l'attention particulière portée à l'élaboration de ce travail, pour la disponibilité d'encadrement, la pertinence et surtout la patience montré tout le long de la thèse. Ses qualités tant scientifiques qu'humaines ont montré son excellence à bien mener un travail d'encadrement.

J'exprime également toute ma gratitude à l'ensemble des personnes présentes dans le jury pour avoir juger ce travail :

Monsieur Gérard Granet et Monsieur Manuel Flury pour le temps qu'ils ont consacré pour rapporter sur mes travaux.

Madame Cécile Gourgon qui m'a fait honneur de présider ce jury. Je la remercie également pour l'intérêt qu'elle a porté à mon travail tout au long de ces années notamment pour la fabrication des échantillons et les mesures réalisées.

Ma reconnaissance va également à Jean Jacques Rousseau de m'avoir accueilli au sein du laboratoire et pour ses multiples conseils et encouragements ; à Damien Jamon pour les mesures réalisées ; à Julie Dutems pour la fabrication des échantillons et les mesures AFM réalisées et sans oublié Zaki Sabit pour l'accompagnement durant le travail de thèse.

Je souhaiterais également remercier tous les collègues du laboratoire : chercheurs, post-doctorants et doctorants et stagiaires algériens, tchadiens, français, marocains, senegalais et libanais pour les bons moments de solidarité et de sympathie.

A tous les membres de ma famille et connaissances pour leur soutien permanent au cours de ces années.

Table des matières

Dédicace	2
Remerciements	3
Introduction générale	8
1 Scattérométrie	11
1.1 Intérêt de la métrologie	11
1.1.1 Généralité	11
1.1.2 Procédés de fabrication	12
1.1.3 Structure sous test : le réseau de diffraction	15
1.1.3.1 Définition	15
1.1.3.2 Modélisation des profils	18
1.1.3.3 Diffraction d'une onde électromagnétique	20
1.2 Les techniques microscopiques classiques	22
1.2.1 Le microscope électronique à balayage	23
1.2.2 Le microscope à force atomique	24
1.3 Les méthodes optiques	26
1.3.1 Positionnement du problème	26
1.3.2 Etat de l'art sur la scattérométrie	27
1.3.3 Résolution du problème inverse de la diffraction	28
1.3.3.1 Modélisation mathématique du problème inverse	28
1.3.3.2 Résolution mathématique du problème inverse	28
1.3.4 Scattérométrie ellipsométrique	32
1.3.4.1 Définition de l'ellipsométrie	32
1.3.4.2 Appareil de mesure	34
1.3.4.3 Signature ellipsométrique	36
1.3.5 Estimation des performances de la caractérisation	39
1.4 Conclusion	41

2	Classification et Reconnaissance de formes	42
2.1	Introduction à la reconnaissance de formes	42
2.1.1	Principe général	42
2.1.2	Applications	44
2.1.3	Prétraitement des données	45
2.1.4	Notion de proximité	45
2.2	Les méthodes classiques de classification	47
2.2.1	Les méthodes de la classification indirecte	47
2.2.1.1	Généralités et définitions	47
2.2.1.2	Les méthodes paramétriques	48
2.2.1.3	Les méthodes non paramétriques	48
2.2.2	Les méthodes de la classification directe	50
2.2.2.1	Généralités et définitions	50
2.2.2.2	Les Séparateurs à Vaste Marge	50
2.3	Les Réseaux de Neurones	54
2.3.1	Des outils mathématiques puissants	54
2.3.1.1	Historique des R.N	54
2.3.1.2	Applications	55
2.3.2	Description et fonctionnement général	56
2.3.2.1	Définition du neurone formel	56
2.3.2.2	Architecture des R.N	58
2.3.2.3	Apprentissage des R.N	59
2.3.3	Les réseaux de Kohonen	61
2.3.3.1	Les Cartes Auto-Organisatrices	61
2.3.3.2	Learning Vector Quantization	64
2.3.4	Le Perceptron Multicouche	65
2.3.4.1	Architecture du PMC	66
2.3.4.2	Principe de fonctionnement	67
2.3.4.3	Apprentissage du PMC	68
2.3.5	Caractérisation par R.N régresseur classique	71
2.4	Conclusion	74
3	Prise en compte des défauts de fabrication dans le processus de caractérisation optique	76
3.1	Cadre de l'étude	76
3.1.1	Contexte de fabrication	76
3.1.2	Échantillons testés	77
3.1.3	Caractérisation par scattérométrie «classique»	79

3.1.4	Hypothèse de forme géométrique inhérente à toute méthode de scattérométrie	80
3.2	Méthode de pondération de profils	82
3.2.1	Principe de la méthode	82
3.2.2	Résultats par scattérométrie classique	83
3.2.3	Résultats par pondération de profils	86
3.2.3.1	Résultats théoriques	86
3.2.3.2	Résultats expérimentaux	89
3.2.4	Limite de la méthode	93
3.3	Classification de profils	97
3.3.1	Détection de la forme cible	97
3.3.1.1	R.N classifieur tout ou rien	98
3.3.1.2	Cas du réseau de forme rectangulaire	99
3.3.1.3	Cas du réseau de forme sinusoïdale	102
3.3.1.4	Étude de la frontière	111
3.3.2	Détermination du type de profil géométrique	114
3.3.2.1	R.N classifieur à 3 classes	114
3.3.2.2	R.N classifieur à 4 classes	128
3.3.2.3	Bilan sur la détermination du type de profil géométrique	132
3.4	Méthode complémentaire dédiée aux réseaux de forme sinusoïdale	133
3.4.1	Généralités sur les signaux périodiques	133
3.4.2	Décomposition d'un profil géométrique quelconque	136
3.4.3	Reconstruction d'un profil	138
3.4.4	Critères de détection	139
3.4.4.1	Rapport entre les deux premiers harmoniques	140
3.4.4.2	Rapport entre les 4 premiers harmoniques	142
3.4.5	Méthode de détection de forme sinusoïdale	144
3.4.5.1	Performances théoriques après apprentissage	145
3.4.5.2	Résultats théoriques	146
3.4.5.3	Résultats expérimentaux	149
3.4.5.4	Bilan sur la méthode de détection de forme sinusoïdale	152
3.5	Conclusion	153
	Conclusion générale _____	154
	Bibliographie _____	157
	Table des figures _____	170

Liste des tableaux	176
--------------------	-----

Introduction générale

L'émergence remarquable des industries de la microélectronique ces dernières années est rendue possible grâce aux avancées technologiques. Cette évolution s'articule autour de la miniaturisation du transistor qui est l'élément de base de tous circuits électroniques intégrés. En effet, la réduction de la taille de ce composant a permis de rendre plus performantes les puces électroniques, à moindre coût, économique en termes d'énergie et à hautes densités d'intégration. Les procédés de fabrication se trouvent alors confrontés à cette course perpétuelle de la miniaturisation. En effet, les dimensions des composants à l'échelle nanométrique deviennent alors le plus grand défi des procédés de fabrication. L'enjeu crucial est de développer des nouvelles techniques capables de franchir les futurs nœuds technologiques. Il est donc essentiel d'accompagner ces avancées par des outils de métrologie performants, rapides et précis afin de mieux contrôler la chaîne de fabrication.

Les méthodes classiques de caractérisation largement employées sont essentiellement microscopiques. Les plus utilisées sont le microscope électronique à balayage (MEB) et le microscope à force atomique (AFM). Le MEB permet d'obtenir une image réelle de la structure sans connaître sa forme au préalable. Elle est considérée comme une technique de référence pour la caractérisation aux dimensions critiques fabriquées. Cependant, elle commence à atteindre ses limites en termes de résolution et est, la plupart du temps, destructive. L'AFM est, quant à elle, limitée par la forme et l'usure de la pointe qu'elle utilise pour scruter la surface de l'échantillon. Par ailleurs, ces techniques permettent seulement une caractérisation locale et, de part leur lenteur, sont inefficaces pour le contrôle en ligne de production. Par conséquent, en raison de ces limites, des nouvelles techniques plus adaptées comme la scattérométrie ont été développées.

La scattérométrie est une technique de métrologie optique non destructive basée sur l'analyse de la lumière diffractée sur une structure périodique. Son fonctionnement est décrit en deux étapes : la mesure et le traitement. La première étape consiste à mesurer la signature optique de l'échantillon par un appareil adapté (ellipsomètre, réflectomètre ou autre). La seconde étape permet d'extraire les paramètres géométriques de la structure à partir de la mesure par résolution du problème inverse. Des nombreuses méthodes numériques sont employées pour la résolution du problème inverse dont les réseaux de neurones artificiels. Ces outils remarquables s'imposent aujourd'hui dans beaucoup de domaines de la physique et notamment en métrologie grâce à leur rapidité et leur efficacité de traitement de l'information. Ils peuvent

aussi être utilisés comme outil d'aide à la décision tout au long du processus de caractérisation scattérométrique.

Cette thèse s'inscrit dans le cadre de la détection et de l'identification de formes géométriques de la structure à tester à partir de la même signature optique utilisée lors d'un processus de caractérisation scattérométrique. L'objectif de l'étude consiste à mettre en place un réseau de neurones en mode classifieur en amont de l'étape de caractérisation proprement dite. Ce système peut ainsi permettre de détecter ou d'identifier la forme de l'échantillon choisi. En effet, pour garantir l'unicité de la solution, la caractérisation scattérométrique se base sur un certain nombre d'hypothèses comme la connaissance préalable de la forme géométrique de l'échantillon. Il s'agit ici de « l'hypothèse de forme ». Ainsi, l'idée est de certifier, par un classifieur, le modèle géométrique sur lequel va fonctionner la caractérisation. Les applications visées sont de deux types : tout d'abord, la détection d'un défaut lors d'un processus de fabrication ou le suivi d'une déviation par rapport à une forme cible. Ensuite, il est possible d'utiliser de tels outils dans l'extraction qualitative d'information sur la structure et notamment sur la forme géométrique de l'échantillon. Les travaux sont effectués au sein du Laboratoire Hubert Curien qui a développé depuis plusieurs années un savoir-faire dans l'utilisation des réseaux de neurones pour la caractérisation scattérométrique.

Ainsi, ce manuscrit est constitué principalement de trois chapitres.

Le chapitre 1 rappelle les concepts fondamentaux et les principales techniques utilisées en scattérométrie. Nous aborderons le contexte général de la métrologie et plus particulièrement les différentes étapes des procédés de fabrication. Les différents profils géométriques des structures testées seront définis ainsi que la méthode numérique utilisée pour simuler leur comportement optique. Les techniques microscopiques classiques seront brièvement abordées. Ensuite, nous présenterons en détail les différentes méthodes couramment utilisées dans la résolution du problème inverse. Pour finir, nous discuterons des critères mis en place pour qualifier les performances de la caractérisation.

Le chapitre 2 traite des principes généraux de la reconnaissance de formes et plus particulièrement les méthodes de la classification automatique. Les différentes méthodes existantes seront décrites. Ensuite, nous aborderons plus en détail les réseaux de neurones utilisés dans ce contexte. Le principe de fonctionnement d'un réseau de neurones déjà utilisé dans le cadre de la caractérisation scattérométrique sera décrit mais nous insisterons sur la particularité de celui ci dans le cadre de la classification utilisée dans cette thèse.

Enfin, dans le dernier chapitre, nous étudierons les potentialités des outils mis en place dans le cadre précis de la classification pour répondre aux besoins de la scattérométrie. Il contient les différents résultats obtenus lors de nos travaux. Tout d'abord, nous étudierons le comportement du réseau de neurones dans le cadre d'une caractérisation par scattérométrie classique afin de mettre en évidence le besoin d'information complémentaire. Dans un premier temps, nous

proposerons une méthode par pondération de profils pour caractériser les formes présentant une déviation minime par rapport à la cible. Ensuite, on se focalisera sur l'apport du réseau de neurones en tant que classifieur utilisé en amont de l'étape de caractérisation proprement dite. Une première application concerne la capacité du réseau de neurones à détecter une forme géométrique fixée tout au long du processus de fabrication. La seconde présente, quant à elle, la potentialité du réseau de neurones à identifier une forme structurelle parmi un nombre limité de profils. Pour finir, nous présenterons une nouvelle approche basée sur le calcul des harmoniques spécifiquement adaptée pour la détection de réseaux de forme sinusoïdale.

Scattérométrie

La miniaturisation des composants électroniques a poussé le monde de la microélectronique à développer des techniques performantes pour mieux contrôler et valider les procédés de fabrication. En effet, avec la limite des méthodes de caractérisation classiques, il est indispensable d'avoir recours à des techniques de métrologie capables de déterminer avec précision et dans un temps raisonnable les propriétés géométriques d'une structure au cours de sa fabrication.

Dans ce chapitre, nous présentons les concepts fondamentaux et les différentes techniques de scattérométrie. Après avoir passé en revue le contexte de la métrologie, nous évoquerons les différentes étapes du procédé de fabrication. Nous définirons ainsi les différents modèles géométriques de structure mis en place. La méthode numérique retenue dans cette thèse pour simuler le comportement optique de ces structures sera ensuite présentée. Puis nous évoquerons brièvement les techniques microscopiques classiquement utilisées en métrologie. Ensuite, nous détaillerons le cœur de la méthode scattérométrique qui constitue le cadre de notre étude en abordant les différentes méthodes couramment usitées dans la littérature scientifique. Enfin, nous détaillerons le principe de la scattérométrie ellipsométrique qui est celle retenue dans nos travaux. Nous en profiterons pour discuter des critères mis en place pour estimer les performances de la caractérisation.

1.1 Intérêt de la métrologie

1.1.1 Généralité

L'essor technologique de l'industrie de la microélectronique repose essentiellement sur la fabrication des puces électroniques. Cette évolution est liée à l'invention du transistor qui est l'élément de base de tous circuits intégrés. Les puces électroniques sont de plus en plus répandues dans notre quotidien : transport, téléphone portable, ordinateur, etc. La réduction de la taille du transistor a permis d'améliorer la performance de ces circuits intégrés, moins coûteux, à haute densité d'intégration et surtout autonomes en terme d'énergie. L'accroissement exponentielle du nombre de transistors par puce a été prédit par Gordon Moore en 1965 [1]. Cette prédiction est marquée par le doublement de la densité de puces électroniques tous les 18 mois. On parle

de nœuds technologiques.

Cependant, les dimensions caractéristiques visées nécessitent une maîtrise accrue des techniques de fabrication. Les procédés technologiques de fabrication tels que le dépôt, la lithographie et la gravure se trouvent alors confrontés à des nouveaux défis. La problématique est de contrôler parfaitement ces composants à l'échelle nanométrique afin de mieux maîtriser la chaîne de fabrication. En effet, il est important que ces avancées technologiques soient accompagnées par des techniques de caractérisation robustes, par exemple, le passage à l'émergence de l'empilement 3D où on superpose un transistor sur un autre pour ainsi augmenter la performance de la puce électronique.

La caractérisation est ainsi une étape importante qui vise à déterminer les dimensions géométriques du futur dispositif. Elle nécessite des outils de métrologie fiables et précis pour valider les résultats. Dans ce contexte la technique de la métrologie dimensionnelle est cruciale et doit répondre aux attentes par une amélioration constante pour faire face aux avancées technologiques. Il est également important de développer des techniques plus performantes et faciles à intégrer en ligne de production afin d'optimiser les processus de fabrication.

1.1.2 Procédés de fabrication

La production des puces électroniques à haute densité d'intégration est essentiellement réalisée par des techniques de fabrication reposant sur la lithographie. Nous allons présenter dans cette partie, les différentes techniques de lithographies les plus couramment utilisées.

La lithographie est une technique qui consiste à transférer le motif d'une structure à travers un masque de résine sur un matériau à graver afin d'obtenir la forme souhaitée. Une représentation schématique des différentes étapes nécessaires est illustrée sur la figure 1.1.

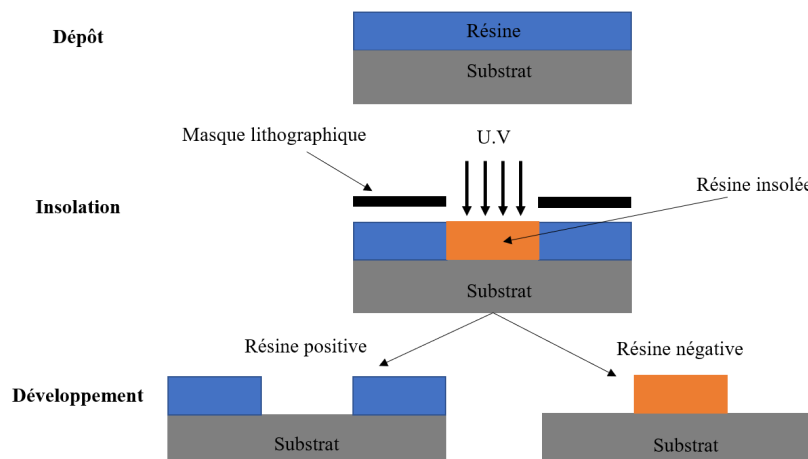


FIGURE 1.1 – Schéma de principe des différentes étapes d'un procédé de lithographie.

Lithographie par voie optique : Cette technique [2] nécessite le dépôt préalable d'une couche de résine sur le substrat à graver. Un faisceau lumineux est ensuite envoyé à travers un masque pour reproduire le motif sur la résine par insolation. Cette interaction avec le rayon modifie les propriétés de la solubilité de la résine. Ensuite, après révélation dans un développeur (solution

basique), la zone exposée va subsister (résine négative) ou être éliminée (résine positive) (figure 1.1). Cette technique a été primordiale pour la réduction de la dimension des motifs dans l'industrie de la microélectronique (figure 1.2).

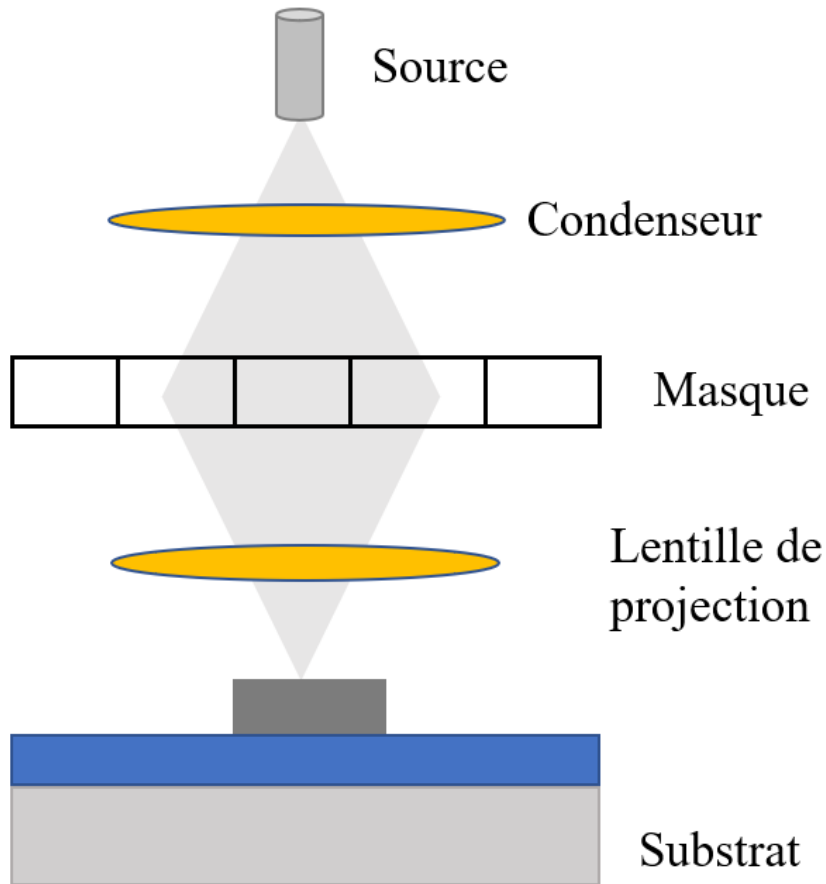


FIGURE 1.2 – Principe de la lithographie optique par projection.

La résolution notée R obtenue par lithographie optique, peut alors être déterminée grâce à l'équation de Rayleigh :

$$R = k_1 \frac{\lambda}{ON} \quad (1.1)$$

λ est la longueur d'onde d'insolation, k_1 une constante positive (de 0.6 à 0.8) qui dépend de la qualité du masque, des éléments optiques et de la résine. ON est l'ouverture numérique du système optique, donnée par :

$$ON = n \times \sin(\theta_{max}) \quad (1.2)$$

Où n est l'indice de réfraction du milieu traversé par la lumière UV et θ_{max} le demi-angle maximal d'ouverture de l'objectif.

De nos jours, la majorité des techniques de la lithographie optique possède des longueurs d'onde de 193 nm [3, 4].

La technique de fabrication par holographie [5], consiste, quant à elle, à faire interférer deux ondes lumineuses sur la couche de la résine. Cette technique permet d'obtenir un réseau

en résine de forme sinusoidale [6]. Cependant, il est difficile d'avoir un réseau qui débouche directement sur le substrat. L'intérêt de cette technique réside dans sa facilité à obtenir à la fois des réseaux de faibles et de grandes dimensions. Une partie de nos travaux porte sur ces structures.

Il existe d'autres techniques de lithographie optique : la lithographie extrême ultraviolet (EUV) [7] fonctionnant sur une longueur d'onde de 13.4 nm, la lithographie par immersion [8] et également la lithographie par faisceau d'électron [9] qui est adaptée pour la fabrication de motifs de très faible dimension.

Lithographie par nanoimpression : C'est une technique [10] adaptée pour la fabrication de structures nanométriques. Elle consiste à dupliquer sur une résine ou sur un polymère déposé(e) sur le substrat à graver des motifs contenus dans un moule. Ce dernier est fabriqué par technique de lithographie standard. Il y a deux types des techniques de nanoimpression : la nanoimpression thermique et la nanoimpression assistée par l'ultraviolet (UV-NIL pour Ultra-Violet curing NanoImprint Lithography) [11].

Pour la nanoimpression thermique, le moule qui contient le motif est pressé dans le polymère chauffé au-delà de sa température de transition visqueuse. Ainsi, le polymère devient visqueux et se déforme. Ensuite, une fois refroidi, le moule est retiré et le motif est transféré dans le polymère. Pour éviter tout contact entre le moule et le substrat, une couche résiduelle est laissée au fond des motifs. On retrouve cette couche résiduelle comme dans le cas de la lithographie par holographie. Cette dernière doit ainsi être prise en compte dans tout processus de caractérisation. Dans la suite, une partie des échantillons testés présentera une telle couche.

Dans l'UV-NIL, la résine est exposée au rayonnement UV après pressage à travers un moule entraînant sa polymérisation. Un des avantages de cette technique est qu'elle n'est pas limitée par la diffraction, la diffusion ou les interférences de l'onde dans le polymère. De la même manière, une couche résiduelle de résine est laissée au fond des motifs [12].

Ainsi, après avoir listé les différentes techniques de lithographie, nous allons à présent décrire l'étape de gravure proprement dite. Cette étape consiste à transférer le motif dans le substrat à partir du masque en résine. Il existe deux cas de gravure : la gravure sélective et non sélective. Dans la gravure sélective, on grave préférentiellement le matériau alors que dans la gravure non sélective, la résine et le matériau sont gravés à même vitesse.

Généralement, deux cas sont possibles pour réaliser la gravure : la gravure humide et la gravure sèche.

Gravure humide : Le substrat contenant le masque de résine est trempé dans une solution chimique. Les zones protégées par la résine subsistent alors que les autres seront éliminées. L'inconvénient de cette technique est qu'elle est de nature isotrope, car la solution chimique peut attaquer le matériau et cela engendrera une déformation des flancs du motif.

Gravure sèche : Cette technique sèche est utilisée pour graver des réseaux de faibles dimensions. Elle utilise un faisceau d'électrons en présence d'un plasma pour graver le matériau. Le plasma est un milieu gazeux ionisé neutre contenant des électrons, des ions positifs et négatifs et

également des particules neutres (molécules, atomes...). L'interaction avec le matériau permet de limiter la déformation du profil. Il est de nature anisotrope. Il existe trois types de gravure sèche selon la nature de l'attaque :

- La gravure IBE (Ion Beam Etching) de nature physique : Il n'y a pas de réaction chimique entre le faisceau d'électron et le matériau à graver ;
- La gravure RIBE (Reactive Ion Beam Etching) de nature physico-chimique : Le bombardement des ions ainsi que la réaction chimique du plasma avec le matériau provoquent l'arrachement de la matière ;
- La gravure RIE (Reactive Ion Etching) de nature chimique : Il dépend de l'interaction entre le matériau à graver et les espèces réactives du plasma.

Dans tous les cas, la forme du profil géométrique dépend de la technique de fabrication utilisée et du choix de leurs différents paramètres. En effet, ces différentes étapes du procédé de fabrication doivent être parfaitement maîtrisées et contrôlées afin d'écartier certaines dérives ou défauts qui peuvent survenir en cours de production.

Le défaut qui peut apparaître lors de l'étape du procédé de fabrication peut être vu comme une déviation de la forme géométrique par rapport à celle désirée. En effet, le résultat obtenu est plus ou moins fidèle à celui attendu. Les raisons peuvent être diverses comme une erreur de calibrage, un défaut de matériel ou liée à l'étape de fabrication. Dans le cadre de notre étude, nous avons choisi de traiter quelques exemples de déviations de forme géométrique qui peuvent survenir lors d'un processus de fabrication (lithographie et gravure) : pente, arrondi, effet de la sous gravure ou la sur-gravure, etc.

Ainsi, la photolithographie et la gravure sont des étapes cruciales pour la fabrication de circuits intégrés.

1.1.3 Structure sous test : le réseau de diffraction

1.1.3.1 Définition

Un réseau de diffraction est une structure périodique selon une ou deux dimensions de forme géométrique quelconque (rectangulaire, trapézoïdale, sinusoidale, etc...) et qui repose sur un support matériel appelé substrat. La caractéristique essentielle du réseau de diffraction est sa période exprimée en nanomètre par la suite. Il a la capacité de séparer ou disperser une lumière incidente dans plusieurs directions privilégiées appelées aussi ordres de diffraction. Cette propriété intéressante est largement mise à profit dans beaucoup de domaines tels que les télécommunications, l'optique ou la photonique [13, 14].

Plus particulièrement en optique guidée ils permettent le couplage de la lumière à l'intérieur des guides d'onde [15]. Ils sont également associés avec un guide d'onde planaire dans le domaine de système de bio-détection [16]. Dans le domaine des lasers, ces réseaux sont utilisés pour réaliser une compression des impulsions femtosecondes [17] comme des filtres à multiplexage à division de longueur d'onde. Ils sont également utilisés dans des cellules solaires organiques [18]

pour augmenter l'absorption du composant.

En scattérométrie, les réseaux de diffraction sont souvent utilisés comme structure sous test à la place des composants submicroniques intégrés. Dans nos travaux, ils servent de références pour tester et comparer les différentes méthodes mises en place.

Rappelons le fondement théorique de la diffraction de la lumière par un réseau périodique. On considère un réseau de diffraction de période T séparant deux milieux d'indices respectifs n_t et n_r et un faisceau incident de longueur d'onde λ_0 . L'angle d'incidence θ_i est défini comme l'angle entre la normale du réseau et le faisceau incident (figure 1.3). L'équation fondamentale du réseau relie, pour chaque ordre m , les caractéristiques de la structure périodique aux différents angles de diffraction θ_{dm} par (équation 1.3) :

$$n_d \sin \theta_{dm} = n_r \sin \theta_i + m \frac{\lambda_0}{T}, m = 0, \pm 1, \pm 2, \dots \quad (1.3)$$

où n_d est l'indice du milieu dans lequel on étudie la diffraction ($-n_r$ pour les ordres réfléchis et n_t pour les ordres transmis)

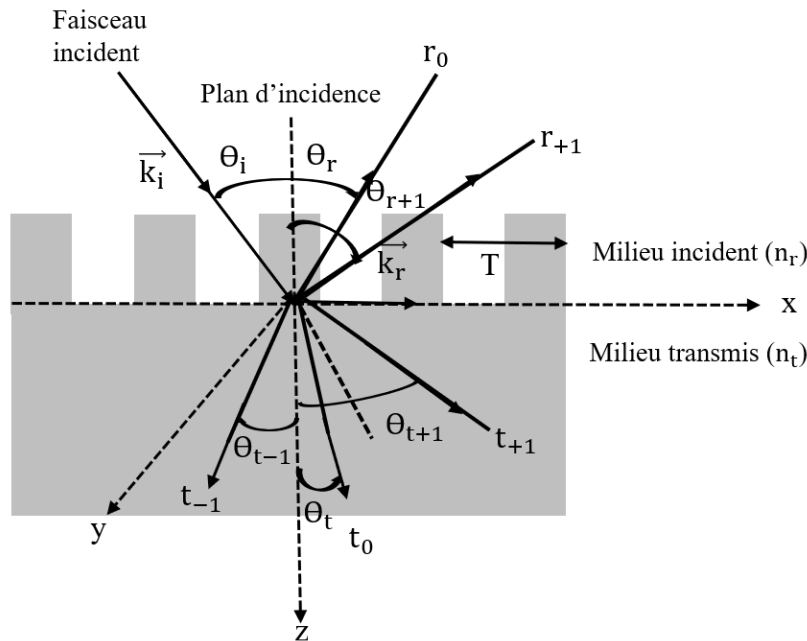


FIGURE 1.3 – Représentation du phénomène de diffraction par réseau.

Lorsque $m = 0$, on retrouve les lois de Snell-Descartes caractérisant la réfraction et la réflexion à travers une interface séparant deux milieux d'indices différents (équation 1.4) :

$$n_t \sin \theta_m = n_r \sin \theta_i \quad n_d = n_t \quad (1.4)$$

$$\theta_m = -\theta_i \quad n_d = -n_t \quad (1.5)$$

Si le milieu émergent est de l'air et qu'il en est de même pour le milieu incident ($n_t = n_r =$

1), alors l'équation du réseau devient (équation 1.6) :

$$\sin\theta_{dm} = \sin\theta_i + m \frac{\lambda_0}{T}, \quad m = 0, \pm 1, \pm 2, \dots \quad (1.6)$$

Il faut cependant noter que plus la période T est faible, plus le nombre d'ordres diffractés diminue.

Il existe deux types de diffractions suivant la manière dont on éclaire le réseau. Soit \mathbf{k}_i le vecteur d'onde incidente défini sur la figure 1.3 tel que :

$$k_i = n_r k_0 \quad \text{avec} \quad k_0 = \frac{2\pi}{\lambda_0} \quad (1.7)$$

Le plan formé entre le vecteur \mathbf{k}_i et la normale au réseau est appelé plan d'incidence. Le vecteur réseau \mathbf{k}_r est le vecteur perpendiculaire au sillon du réseau dont le module appelé nombre d'ondes et donné par (équation 1.8) :

$$k_r = \frac{2\pi}{T} \quad (1.8)$$

L'équation du réseau se résume à (équation 1.9) :

$$n_d \sin\theta_{dm} = n_r \sin\theta_i + m \frac{k_r}{k_0} \quad (1.9)$$

Deux cas se présentent alors :

- Dans le cas où le plan d'incidence est perpendiculaire au sillon (\mathbf{k}_r est contenu dans le plan d'incidence), on parle de la diffraction classique et les ordres de diffraction sont contenus dans le plan d'incidence. Lorsque le vecteur \mathbf{k}_i est perpendiculaire au plan du réseau, on parle de la diffraction normale.
- Dans le cas où le plan d'incidence fait un angle ϕ (angle azimutal) avec la normale au réseau \mathbf{k}_r , on parle de la diffraction conique par opposition à la configuration classique.

Dans le cadre de cette thèse, nous traiterons seulement le cas d'une incidence classique.

Il est également nécessaire de définir la polarisation de l'onde incidente caractérisée par la direction du champ électrique \mathbf{E} .

On distingue deux cas si l'onde incidente est linéairement polarisée :

- Le vecteur champs électrique \vec{E} est perpendiculaire au plan d'incidence, on dit que la polarisation est transverse électrique et notée TE ou s (figure 1.4(a)) ;
- Le vecteur champs électrique \vec{E} est inclus dans le plan d'incidence. On parle alors de polarisation transverse magnétique et notée TM ou p (figure 1.4(b)).

Les deux états de polarisation de la lumière sont représentés dans la figure 1.4.

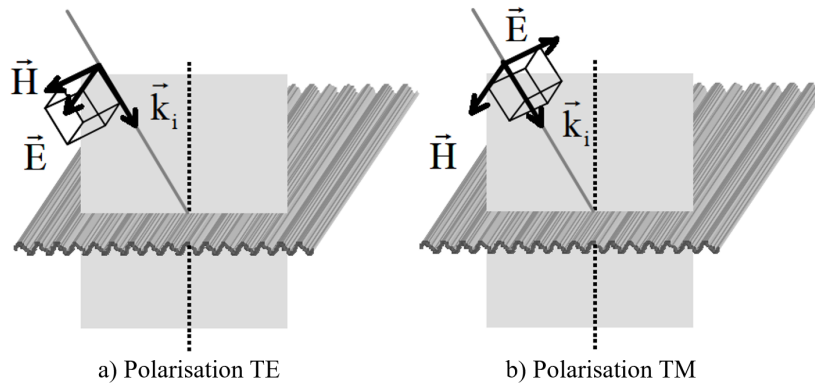


FIGURE 1.4 – Représentation des deux états de polarisation de la lumière a)TE et b)TM [19].

1.1.3.2 Modélisation des profils

Dans ce paragraphe, nous présentons les formes de profils géométriques les plus utilisées et rencontrées en scattérométrie afin de modéliser au mieux les échantillons à tester. Ces formes de réseaux représentent les profils cibles et leurs possibles défauts connus occasionnés par la méthode de fabrication.

Le profil le plus simple est constitué d'un motif rectangulaire (figure 1.5(a)). Il est défini par deux paramètres : la dimension critique CD et la hauteur h ($h > 0$ et $0 < CD < T$). Il est généralement considéré comme le profil cible. Ainsi, les profils qui représentent différents types de défauts fréquemment rencontrés en lithographie sont :

- La forme trapézoïdale symétrique impliquant 3 paramètres (figure 1.5(b)). Elle est la plus employée en scattérométrie pour inclure une déviation particulière de la forme rectangulaire. Les paramètres géométriques de ce profil sont la projection de la pente b_1 , la dimension critique b_2 et la hauteur h . Ils répondent aux conditions suivantes : $2 \times b_1 + b_2 < T$, avec $b_1, b_2 > 0$ et $h > 0$.
- Le profil rectangulaire arrondi en haut impliquant un paramètre supplémentaire r_h (figure 1.5(c)). Il est défini par 3 paramètres, l'arrondi au sommet r_h , la dimension critique CD_h et la hauteur h . Ce profil géométrique existe exclusivement pour $0 < r_h < \min(h, \frac{CD_h}{2})$, $0 < CD_h < T$ et $h > 0$. Cependant, le paramètre r_h est remplacé par le coefficient normalisé suivant δ ($0 < \delta < 1$) pour garantir une plage de variation indépendante.

$$\delta = \frac{2 \times r_h}{CD_h} \quad (1.10)$$

- Enfin, le profil rectangulaire arrondi en bas qui implique également un paramètre supplémentaire (figure 1.5(d)) par rapport à la forme rectangulaire. Il est caractérisé par trois paramètres ; le rayon de la courbure r_b , la dimension critique CD_b et la hauteur h , et définie selon les conditions suivantes : $0 < r_b < h$ et $2 \times r_b + CD_b < T$.

Nous avons également modélisé des réseaux de forme sinusoidale (figure 1.6(a)) considérée comme le profil cible pour certains procédés de fabrication par exemple la technique par holographie (paragraphe 1.1.2). Il est défini par un seul paramètre : l'amplitude (ou hauteur) h .

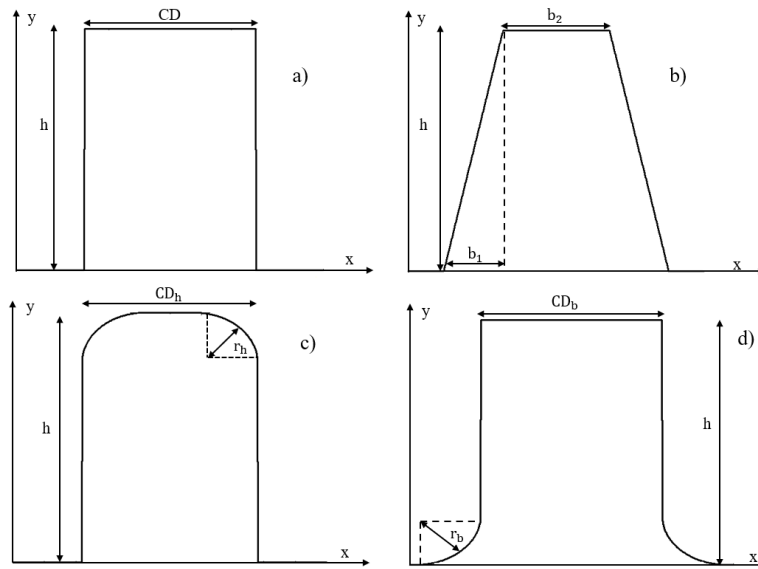


FIGURE 1.5 – Modélisation paramétrique des profils géométriques largement utilisés en scattérométrie : profil rectangulaire (a), profil symétrique trapézoïdal (b), profil rectangulaire arrondi en haut (c) et profil rectangulaire arrondi en bas (d).

En général, il est difficile d'obtenir des profils sinusoidaux parfaits. La forme du profil souhaité peut être plus ou moins fidèle à celle attendue à l'issue des différentes étapes technologiques d'un processus de gravure. Ainsi, nous définissons un ensemble de profils permettant de modéliser les défauts occasionnés lors de la fabrication :

- La forme gaussienne (figure 1.6(b)) définie par 2 paramètres : l'écart type σ et la hauteur h . La condition d'existence de ce profil est : $\sigma > 0$ et $h > 0$.
- Le profil sinus déformé (figure 1.6(c)) qui est une combinaison de 2 profils, dont l'un est un profil sinus parfait et l'autre un profil sinus réduit sur sa période. Il est défini par 2 paramètres : la hauteur h et un coefficient dénommé ρ permettant de réduire la forme sur la période. Le profil est défini par :

$$f(x) = h \times (f_1(x) - \rho \times f_2(x)) \quad (1.11)$$

$$\text{où } f_1(x) = \frac{\cos(\frac{2\pi x}{T})+1}{2} \text{ et } f_2(x) = -\frac{\cos(2(\frac{2\pi x}{T})) + 1}{2}$$

Cette forme existe pour $h > 0$.

- Le profil sinus tronqué (figure 1.6(d)) qui est une forme sinusoidale tronquée sur sa partie positive. Le profil est défini par 3 paramètres : le coefficient α , la projection b du sinus et l'amplitude h . Le paramètre α spécifie la hauteur de troncature par un pourcentage de la hauteur initiale du sinus. Il permet à la fois de moduler les flancs du profil mais également de définir le rapport ligne/espace. Les conditions d'existence du profil sont : $0 \leq \alpha \leq 1$, $0 < b \leq T$ et $h > 0$.

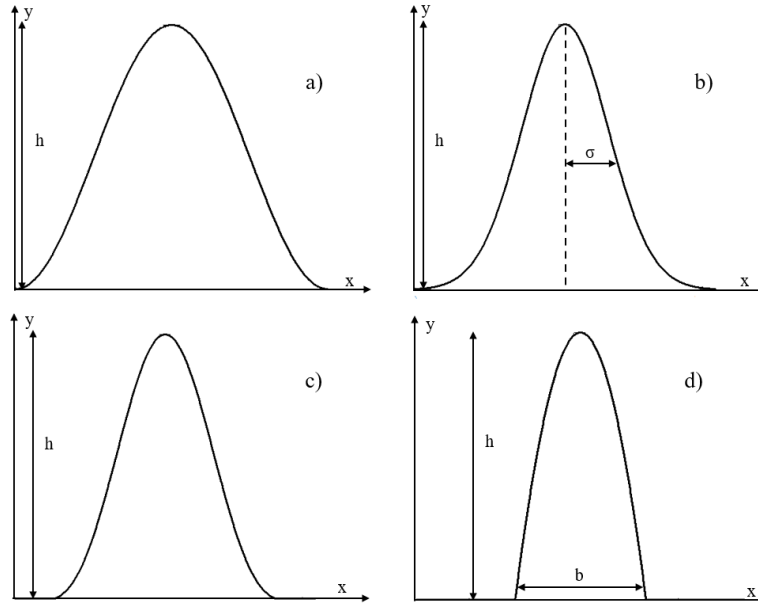


FIGURE 1.6 – Modélisation paramétrique des profil géométriques utilisés en scattérométrie : profil sinusoïdal (a), profil gaussien (b), profil sinus déformé (c) et profil sinus tronqué (d).

1.1.3.3 Diffraction d'une onde électromagnétique

Résoudre le problème de diffraction d'une onde électromagnétique à travers un réseau périodique de profil connu consiste à connaître la distribution du champ électrique dans les différents milieux.

Considérons un phénomène physique F décrit par des paramètres \mathbf{x} . La résolution du problème de diffraction consiste à modéliser F par M afin de trouver les effets \mathbf{y} engendrés par les paramètres \mathbf{x} . Après l'obtention du modèle M , nous pouvons estimer $F(\mathbf{x})$ par $M(\mathbf{x})$.

En scattérométrie, la résolution du problème direct de diffraction consiste donc à simuler des signatures optiques à partir de modèles de l'échantillon. Ces signatures optiques sont déterminées à l'aide d'outils de simulations électromagnétiques.

Il existe deux méthodes pour résoudre le problème de diffraction : les méthodes approchées (théorie scalaire, Rayleigh...) et les méthodes rigoureuses (méthode intégrale, différentielle, méthode modale par expansion de série de Fourier ...).

Les méthodes approchées présentent un avantage indéniable de rapidité de calcul mais estiment des résultats plus ou moins fidèles au comportement physique du phénomène. La théorie scalaire de Fresnel-Kirchoff [20] est l'une de ces méthodes. Elle traite la propagation de l'onde comme un phénomène scalaire sans prendre en compte sa nature vectorielle. Bien qu'elle soit simple et facile à interpréter, elle reste très peu employée dans le domaine des réseaux optiques. Elle permet également de résoudre analytiquement le problème de diffraction dans des cas limités, lorsque l'angle d'incidence θ_i est proche de la normale et les dimensions du réseau sont largement supérieures à la longueur d'onde λ . Dans le même ordre d'idée, l'hypothèse de Rayleigh traite le problème de diffraction par la voie électromagnétique en résolvant un système d'équation linéaire [21, 22]. Toutefois, cette méthode reste à valider dans le cas où le réseau est

profond. Par exemple pour des réseaux sinusoïdaux, il a été démontré [23] que cette théorie est bien adaptée à condition que $h/T < 0.15$.

Les méthodes rigoureuses consistent à résoudre les équations de Maxwell sans aucune approximation sur le plan théorique. Parmi ces méthodes, la méthode intégrale [24, 25] se base sur le théorème de Green en exprimant le champ électrique en chaque point de l'espace par une intégrale sur la surface du réseau. L'onde incidente génère un courant de surface et le champ diffracté est obtenu par rayonnement de ce courant de surface. La méthode des éléments finis [26], quant à elle, représente la surface par un maillage. Le champ électrique est exprimé comme une somme des fonctions élémentaires sur toutes les cellules du maillage. Le problème est réduit à un système d'équations linéaires algébriques. Une autre méthode intégrale dite des sources fictives [27] considère qu'un champ électrique est généré par des sources fictives choisies de manière à vérifier les équations de Maxwell.

La méthode différentielle [28, 29] exprime les équations de Maxwell sous leur forme différentielle. Le champ électrique et la permittivité sont représentés en série de Fourier. Dans la méthode C [30, 31], on effectue un changement de coordonnées simples pour séparer l'espace en deux milieux homogènes.

Enfin, l'approche que nous allons utiliser dans le cadre de nos travaux est la méthode modale par expansion de série de Fourier (MMFE) [32, 33].

Développée par Moharam et Gaylor sous le nom de la théorie rigoureuse des ondes couplées (en anglais RCWA pour Rigorous Coupled Wave Analysis) [32]. La MMFE consiste à exprimer les solutions des équations de Maxwell sous la forme de mode de propagation dans le réseau à analyser. La MMFE fonctionne sur un réseau rectangulaire et pour étendre son utilisation aux profils quelconques, on découpe ce dernier en tranches (soit en un empilement de réseaux rectangulaires) (figure 1.7). La permittivité à l'intérieur du réseau est représentée par une série de Fourier. Le nombre M d'harmoniques est le nombre de termes de la série de Fourier. Numériquement, on définit également un nombre limité d'harmoniques M pour effectuer le calcul.

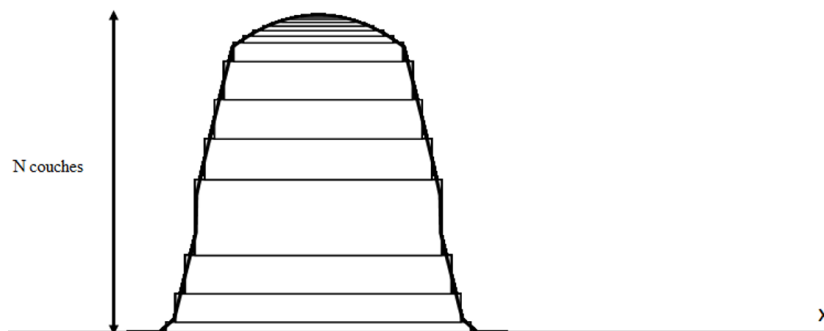


FIGURE 1.7 – Exemple de découpe d'un profil géométrique quelconque en couches rectangulaires d'épaisseurs différentes.

La MMMFE peut s'appliquer à des réseaux 1D ou 2D, absorbants ou métalliques [34] en configuration classique ou conique. Le code de calcul de la MMMFE que nous utilisons est développée par l'équipe « Electromagnétisme » de l'institut Pascal (UMR 6602-CNRS,

Université Blaise Pascal-Clermont Ferrand) [34].

Le principal inconvénient de cette méthode est qu'il faut approcher la forme du réseau par des empilements de couches rectangulaires. Pour cela, nous avons développé au sein du laboratoire une méthode de découpe automatique de profils géométriques quelconques à condition que ces derniers soient symétriques. Cette méthode prend en compte le nombre de couches désirées et ajuste automatiquement les hauteurs de chaque couche afin d'être le plus fidèle possible au motif réel.

Soit une fonction $Y(x)$ (avec $0 < x < T$) définie comme étant un profil géométrique. Le profil est d'abord centré par rapport à l'origine du repère. On note X la distance par rapport à l'abscisse qui correspond à $Y(X) = 0$. Nous définissons également un pas ε dépendant du nombre de couches désirées. La position X est incrémentée à chaque itération de ε entre $-\frac{T}{2}$ et 0 par : $X^{i+1} = X^i + \varepsilon$. L'abscisse du début de couche est donnée par : $x_1^i = X^{i+1} - \frac{\varepsilon}{2}$ et celui de la fin de couche par $x_2^i = -(X^{i+1} - \frac{\varepsilon}{2})$. L'épaisseur de couche est donnée par $h_i = Y(X^{i+1}) - Y(X^{i+1} - \varepsilon)$. Ainsi, l'épaisseur de chaque couche dépend de la forme à une autre (pente, courbure, etc.) du profil présenté et du nombre de couches demandé (figure 1.8).

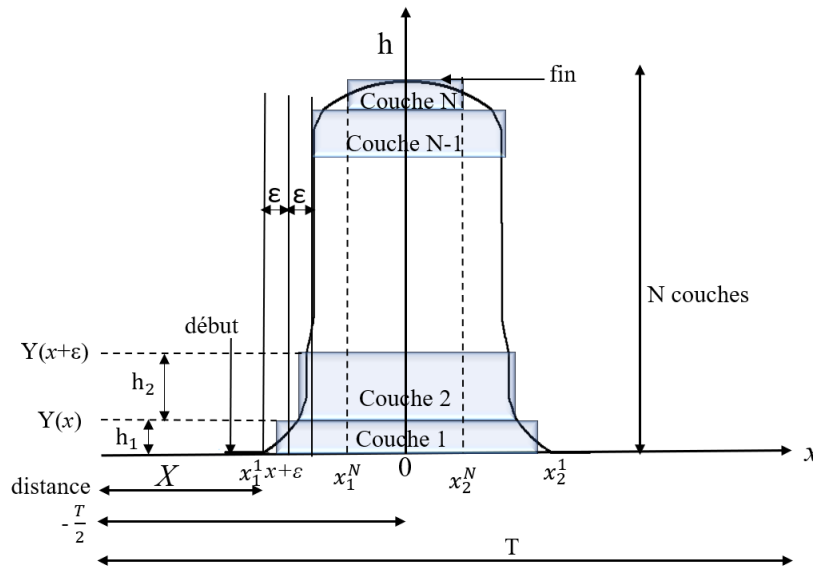


FIGURE 1.8 – Procédure de découpe d'un profil quelconque.

1.2 Les techniques microscopiques classiques

Les méthodes microscopiques telles que le microscope électronique à balayage (MEB) et le microscope à force atomique (AFM) sont largement utilisées dans le domaine de recherche et en milieu industriel pour déterminer les dimensions caractéristiques d'un échantillon. Nous allons présenter rapidement le principe de chacune d'entre elles ainsi que leurs avantages et inconvénients.

1.2.1 Le microscope électronique à balayage

Le fonctionnement du MEB [35, 36] est basé sur l'interaction entre un faisceau d'électrons et un matériau constituant la surface à analyser. Des électrons « rétrodiffusés » et secondaires de différentes natures sont ainsi émis. Les électrons secondaires détectés permettent de constituer la composition chimique ou topographique de la structure. On distingue deux catégories de MEB : le CD-SEM (pour Critical Dimension Scanning Electron Microscopy) et le X-SEM (X pour cross-section).

Le CD-SEM permet de vérifier la largeur des lignes fabriquées par une vue de dessus (figure 1.9). Cette technique est non destructive et cela permet à l'échantillon d'être réutilisé par la suite. Elle est très utilisée par l'industrie pour le contrôle en ligne dans les zones de production. Elle a aussi l'avantage de donner une information sur la rugosité du motif. Elle peut, dans certains cas, nécessiter un algorithme de traitement d'image chronophage rendant notamment la technique inadaptée pour une caractérisation en temps réel. Il faut noter également qu'elle ne donne pas la forme du motif.

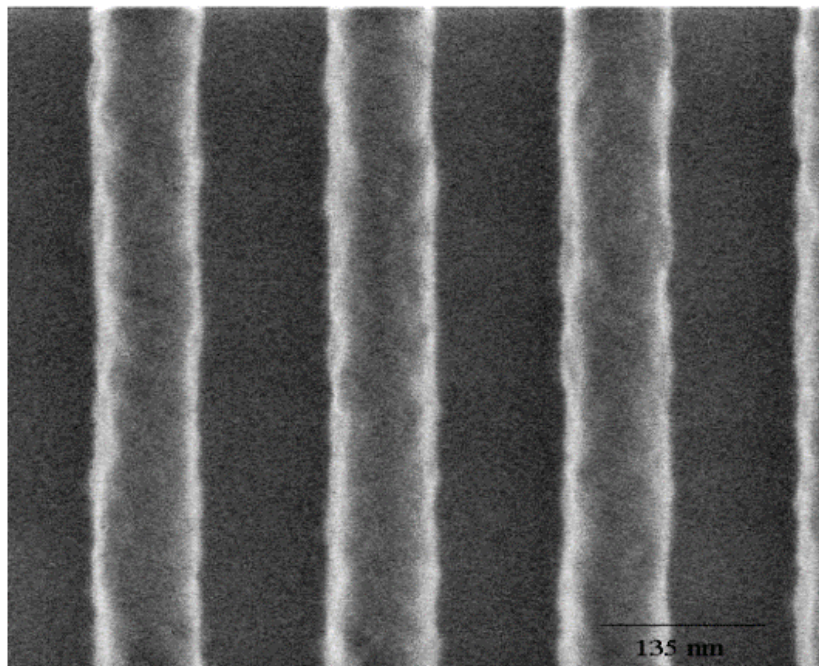


FIGURE 1.9 – Image CD-SEM d'une ligne de résine sur Silicium de période $T=189$ nm [37].

La technique de X-SEM consiste à observer la coupe d'un échantillon. C'est une technique destructive, car elle nécessite la découpe de la plaque à l'endroit de la mesure. Pour éviter le réchauffement de la plaque, il est donc indispensable de passer par l'étape de métallisation. Cette dernière rend l'échantillon inutilisable par la suite. Par contre, à la différence du CD-SEM, le X-SEM permet de caractériser la forme géométrique des motifs et aussi potentiellement des couches sous-jacentes. La figure 1.10 montre l'image d'un réseau en silice par un MEB.

Le MEB permet ainsi d'avoir une image réelle de l'échantillon sans connaître l'information préalable sur la forme de la structure. Cependant, il est difficile d'obtenir une caractérisation avec une détermination précise des paramètres du profil. En effet, l'image fournie par le MEB

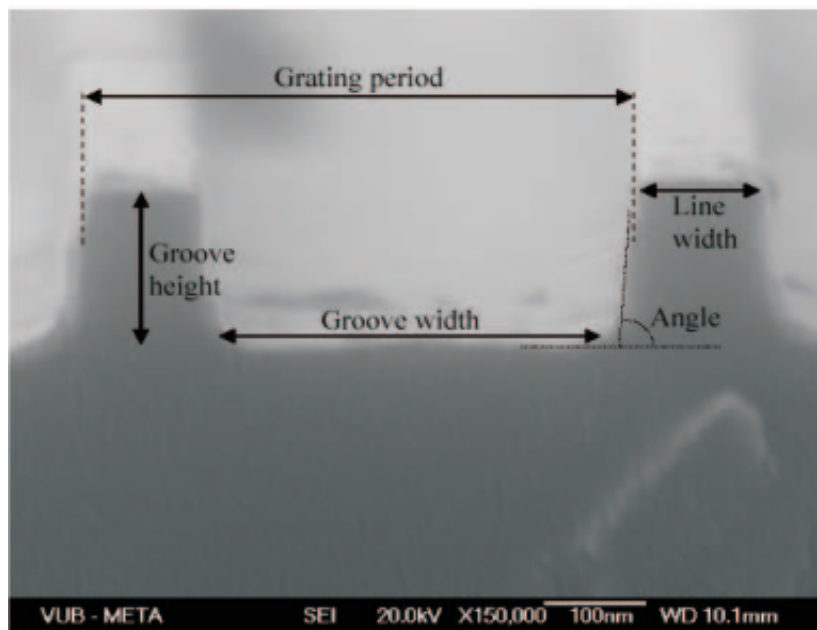


FIGURE 1.10 – Exemple d’image d’un réseau en silice provenant du X-SEM (Department of Applied Physics and Photonics, Vrije Universiteit Brussel, Bruxelles) [38].

comporte différents niveaux de gris et cela complique l’extraction du profil. C’est un outil lourd et complexe qui nécessite de nombreux réglages. Il peut nécessiter des algorithmes pour traiter les images et surtout, nécessite un temps significatif pour l’obtention d’une image fiable. Il faut également préciser que cette technique peut être destructive, car l’obtention d’une image en vue de coupe implique la destruction de l’échantillon par suite de clivage.

1.2.2 Le microscope à force atomique

Introduite en 1986 par G. Binnig et ses collaborateurs pour succéder au microscope à effet tunnel (Scanning Tunneling Microscope ou STM), l’AFM permet de caractériser différents types d’échantillons (conducteur, semi-conducteur, isolant, ...). Il a l’avantage aussi d’être utilisé dans différents types d’environnements comme l’air, l’ultravide ou en milieu liquide. Il fonctionne suivant la loi d’attraction et de répulsion entre une pointe placée à l’extrémité d’une poutre et la surface du matériau à analyser. Sur la face arrière de la pointe, un système de photodiode est placé pour capter la réflexion du faisceau envoyé par un laser. Ainsi, le mouvement de la pointe est transcrit sur un plan à deux dimensions et permet de reconstituer le profil [39]. La figure 1.11 donne un aperçu d’une image brute obtenue par un AFM classique.

Il existe différents modes de fonctionnement qui permettent d’effectuer une mesure AFM : contact, tapping ou non-contact. Dans le mode contact, la pointe est maintenue en contact permanent avec le motif mais cela peut engendrer l’usure de la pointe. Pour pallier cet inconvénient, d’autres techniques ont vu le jour. D’abord, le contact est intermittent, il oscille à une faible amplitude suivant une fréquence proche de sa fréquence de résonance. Enfin, le mode non-contact ou résonant qui est caractérisé par une pointe distante de 10 *nm* du motif qui oscille également suivant une fréquence proche de sa fréquence de résonance.

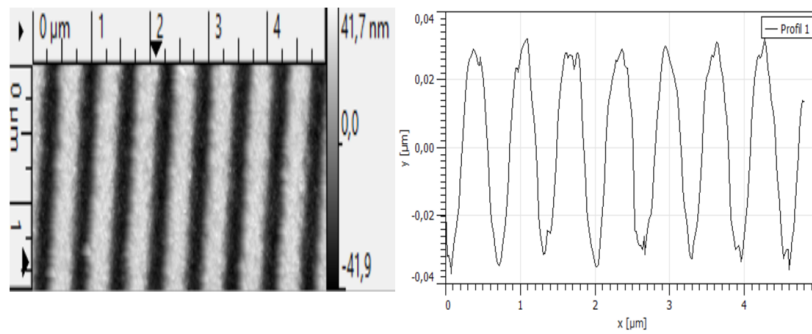


FIGURE 1.11 – Exemple d’une image obtenue par AFM classique

Les différents types de pointes (coniques ou évasées) conditionnent la qualité d’une mesure AFM. Ainsi, la forme du profil mesuré est directement liée à l’état de la pointe (forme, taille et usure). Parfois, il est nécessaire de réaliser une étape de traitement afin de déconvoluer le signal obtenu avec la forme de la pointe pour approcher au mieux le profil réel. Par ailleurs, l’AFM est peu ou pas adapté pour la caractérisation de réseaux très profonds. En effet, la taille de la pointe étant limitée par sa forme, il est donc difficile d’atteindre le fond de sillons. La figure 1.12 montre une caractérisation par l’AFM de différents profils géométriques par des pointes coniques ainsi que les artefacts de mesures associés qui nous intéressent dans le cadre de ces travaux.

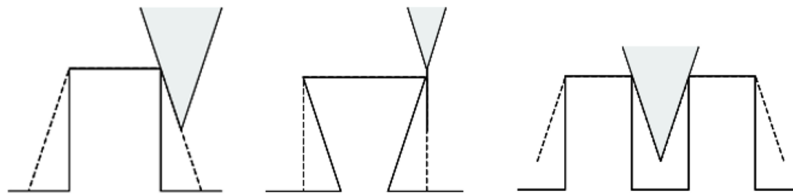


FIGURE 1.12 – Influence de la forme de la pointe AFM sur la mesure du profil géométrique. Les traits en pointillés et continus désignent respectivement le profil mesuré et le profil réel [38].

Il existe une variante de l’AFM appelée CD-AFM [40] qui utilise une pointe évasée dite en « patte d’éléphant » qui permet de mesurer la forme des flancs de la structure de manière plus précise que l’AFM classique. Elle permet également la caractérisation de la rugosité des lignes du réseau et les ondulations du motif (figure 1.13).



FIGURE 1.13 – Schématisation d’une reconstruction de profil avec un CD-AFM [41].

Les méthodes présentées ci-dessus permettent d’obtenir la forme du profil (CD, hauteur, pente, rugosité, etc.) et ne nécessitent pas la connaissance à priori sur la structure. Malgré

les avantages qu'elles possèdent, ces techniques restent assez complexes et offrent une caractérisation locale de l'échantillon. Elles sont également, pour certaines, limitées en termes de résolution. Elle peuvent parfois être destructives et nécessitent un temps de mesure significatif. Ainsi, l'exigence des divers nœuds technologiques a provoqué le développement des techniques alternatives plus adaptées comme la scattérométrie.

1.3 Les méthodes optiques

1.3.1 Positionnement du problème

La scattérométrie est une méthode de métrologie optique non destructive basée sur l'analyse de la lumière diffractée sur une structure périodique. Elle est qualifiée de méthode indirecte c'est-à-dire qu'elle ne mesure pas directement les paramètres géométriques du profil mais les calcule à partir d'un ensemble de mesures réalisées sur l'échantillon appelé signature optique. C'est une technique qui peut être utilisée aussi bien pour la mesure ponctuelle d'un échantillon ou pour le contrôle des procédés en temps réel contrairement aux méthodes microscopiques.

Le principe de fonctionnement de la scattérométrie peut être décrit en deux étapes : la mesure et le traitement (figure 1.14) :

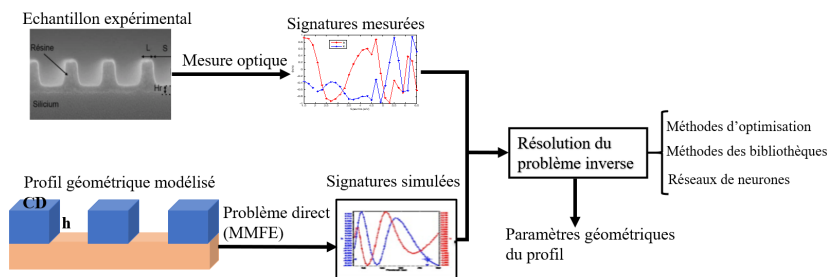


FIGURE 1.14 – Schéma illustrant le principe de la scattérométrie.

La première étape consiste à mesurer la signature optique émise par la structure à tester en utilisant un appareil adapté (ellipsomètre ou réflectomètre). Pour les traitements, on aura besoin de simuler des signatures optiques dans les mêmes conditions que la mesure précédente. On parle de calcul direct. Pour cela on utilise une méthode numérique exacte (paragraphe 1.1.3.3). Ensuite, on détermine les paramètres géométriques de la structure à partir de la comparaison des signatures simulées et réelles. On parle alors de résolution du problème inverse. Ce problème est défini sous certaines conditions expérimentales (connaissance des indices optiques des matériaux, de l'angle d'incidence, de la longueur d'onde, etc.) et l'hypothèse préalable notamment sur la structure de l'échantillon (modèle géométrique, période, l'épaisseur, etc.).

1.3.2 Etat de l'art sur la scattérométrie

Comme nous l'avons vu, le principe de la scattérométrie n'est pas basé directement sur la mesure de la structure mais sur le traitement de son interaction avec une onde électromagnétique. Cette mesure est directement liée aux paramètres géométriques du motif et constitue de ce fait une signature optique de ce dernier. Il en existe différents types. Cependant la plupart du temps elle est issue d'une mesure en réflexion afin de s'affranchir des problèmes éventuels liés à la traversée de matériaux plus ou moins absorbants. Elle peut, par exemple, être constituée soit de différentes intensités (scattérométrie réflectométrique) soit de mesures ellipsométriques (scattérométrie ellipsométrique). La mesure expérimentale est assujettie dans ce cas à des conditions fixées à l'avance : angle d'incidence classique fixe ou variable, mono ou multi longueur d'onde.

Dans le cas de la réflectométrie mono-longueur d'onde à angle variable, on parle de scattérométrie goniométrique [42, 43, 44]. La signature optique est composée d'une série de mesures d'intensité diffractée dans l'ordre zéro principalement dans une polarisation fixe en fonction de l'angle d'incidence θ_i . On parle également de scattérométrie 2θ [45, 46]. Il existe également d'autres techniques qui utilisent comme signature optique des mesures effectuées en configuration coniques (angle azimutal ϕ variable) [47].

Dans la réflectométrie à incidence normale, on mesure cette fois-ci la variation de la réflectivité pour une polarisation donnée [48] en faisant varier la longueur d'onde.

En scattérométrie ellipsométrique, la variation de polarisation résultant de l'interaction entre la lumière incidente et la structure est prise en compte dans la mesure. La signature optique est présentée sous la forme d'un spectre en longueur d'onde [49, 50, 51]. La quantité mesurée dépend du type d'ellipsomètre utilisé. En ellipsométrie classique, fondée sur le formalisme de Jones [52], il s'agit de deux mesures de paramètres caractéristiques pour une longueur d'onde donnée (par exemple I_s et I_c pour un ellipsomètre à modulation de phase et α et β pour l'ellipsomètre à polariseur tournant). En ellipsométrie de Mueller [53] utilisant le formalisme de Stokes [54], la signature optique est simplement représentée par les éléments de la matrice de Mueller de l'échantillon.

Cependant, il faut noter que la majorité des études réalisées en scattérométrie concerne de formes géométriques de types trapézoïdales [42, 55]. Ainsi, des études de caractérisation par réseau de neurones (R.N) [44, 51] sont effectuées sur ce type de réseau. Ce dernier est employé comme profil cible dans le cadre de la classification de formes géométriques par réseau de neurones [56] et également par Séparateurs à Vaste Marge (SVM) [57, 58]. Des études sur la reconstruction d'un réseau multicouche de type trapézoïdal dans le cadre de gravure en profondeur [59, 60] tout en prenant en compte l'erreur sur la pente ont été également effectuées. D'autres [61, 62] sont effectuées pour détecter par réseau de neurones la présence ou non d'une couche résiduelle de résine. Il existe également différentes formes dérivées du profil trapézoïdal utilisées dans la littérature, dont certaines possèdent des arrondis au sommet [37, 63] et d'autres prennent en compte les arrondis en haut et bas [64]. En effet, ces études sont effectuées pour

ainsi détecter certains défauts de fabrication. Par ailleurs, l'étude concernant d'autres formes de réseau tel que la forme sinusoïdale est encore faible [65]. Toutefois, dans notre étude, on tient compte de plusieurs formes possibles rarement exploitées en pratique.

Dans le cadre de nos travaux, nous utilisons la scattérométrie ellipsométrique pour déterminer les signatures optiques.

1.3.3 Résolution du problème inverse de la diffraction

1.3.3.1 Modélisation mathématique du problème inverse

Dans le cas du problème direct de diffraction, nous avons défini la modélisation M du phénomène physique F . Maintenant, nous cherchons à déterminer les paramètres \mathbf{x} qui produisent les mesures \mathbf{y} . Mathématiquement, on cherche à modéliser la fonction F^{-1} définie par (équation 1.12) :

$$\mathbf{x} = F^{-1}(\mathbf{y}) \quad (1.12)$$

De façon générale, la résolution de ce problème se fait soit analytiquement ou numériquement. On cherche alors les paramètres \mathbf{x} qui induisent une signature optique \mathbf{y} la plus proche possible de celle mesurée notée $\tilde{\mathbf{y}}$. Pour cela, on cherche à minimiser l'écart noté χ^2 entre les deux signatures suivantes (équation 1.13) :

$$\chi^2 = \sum_{i=1}^m (F_i(\mathbf{x}) - \tilde{\mathbf{y}}_i)^2 \quad (1.13)$$

avec m le nombre total de mesures composant la signature optique

La résolution du problème inverse en scattérométrie consiste à déterminer les paramètres géométriques définissant la forme de la structure à partir de mesures optiques effectuées sur ce dernier.

1.3.3.2 Résolution mathématique du problème inverse

Il existe de nombreuses méthodes pour la résolution de ce problème. Citons notamment les méthodes d'optimisation, la méthode des bibliothèques ou les réseaux de neurones qui peuvent être présentés comme une méthode de régression non linéaire.

1.3.3.2.1 Méthode d'optimisation

La méthode d'optimisation adaptée à la scattérométrie consiste, à chaque itération, à diminuer le critère χ^2 exprimant l'écart entre une signature optique simulée obtenue à partir d'un jeu ajustable de paramètres géométriques et la signature optique expérimentale. Au début les valeurs initiales des paramètres géométriques du modèle sont fixées en fonction de la connaissance de la structure. Ensuite, ces valeurs sont modifiées itérativement suivant des règles

précises impliquant les informations récoltées au cours de l'optimisation conduisant à la diminution du critère choisi (ou fonction coût). A l'issue de la convergence de la méthode, on relève les paramètres géométriques correspondants à une signature optique simulée la plus proche de la mesure. L'algorithme est ainsi stoppé lorsque le critère à minimiser a atteint un minimum ou devient inférieur à une valeur prédéfinie choisi par l'utilisateur.

L'objectif de la méthode d'optimisation est bien évidemment de trouver le minimum global de la fonction χ^2 et de ne pas converger prématurément vers un minimum local conduisant à une solution non optimale. Malheureusement, sauf sous certaines conditions restrictives, la plupart des méthodes d'optimisation qualifiées de classiques ne permettent pas d'assurer cet objectif.

L'algorithme de descente du gradient : Le principe de fonctionnement de l'algorithme de descente de gradient [66] consiste à calculer à chaque itération k , que nous noterons \mathbf{x}^{k+1} par (équation 1.14) :

$$\mathbf{x}^{k+1} = \mathbf{x}^k + \alpha^k \mathbf{d}^k \quad (1.14)$$

où \mathbf{d}^k représente la direction de descente donnée par $\mathbf{d}^k = -\nabla\chi^2(\mathbf{x}^k)$ et α^k le pas tel que : $\chi^2(\mathbf{x}^{k+1}) < \chi^2(\mathbf{x}^k)$.

L'équation devient (équation 1.15) :

$$\mathbf{x}^{k+1} = \mathbf{x}^k - \alpha^k \nabla\chi^2(\mathbf{x}^k) \quad (1.15)$$

La direction de recherche est donc donnée par l'opposé du gradient évalué au point \mathbf{x}^k . Cette méthode est fonctionnelle mais l'inconvénient majeur est sa lenteur de convergence.

L'algorithme de Gauss-Newton : L'algorithme de Gauss-Newton appelé également méthode du second ordre utilise la matrice Hessienne \mathbf{H} pour approcher la fonction coût en \mathbf{x}^k . On utilise ainsi le modèle d'approximation quadratique donné par la formule de Taylor (équation 1.16) :

$$\chi^2(\mathbf{x}^{k+1}) = \chi^2(\mathbf{x}^k) - \nabla\chi^2(\mathbf{x}^k)^T \mathbf{s}^k + \frac{1}{2} \mathbf{s}^{kT} \mathbf{H} \mathbf{s}^k \quad (1.16)$$

où $\mathbf{s}^k = \mathbf{x}^{k+1} - \mathbf{x}^k = \alpha^k \mathbf{d}^k$

On cherche alors à déterminer le pas \mathbf{s}^k qui permet d'atteindre le minimum de ce modèle défini par $\nabla\chi^2(\mathbf{x}^{k+1}) = 0$. La solution de ce problème est donc :

$$\mathbf{H} \mathbf{s}^k = -\nabla\chi^2(\mathbf{x}^k) \quad (1.17)$$

Soit :

$$\mathbf{s}^k = -\mathbf{H}^{-1} \nabla\chi^2(\mathbf{x}^k) \quad (1.18)$$

Cette méthode permet d'obtenir une convergence rapide sous certaines conditions notam-

ment lorsque la fonction coût est quadratique. Malgré cela, le fait d'utiliser la matrice Hessienne indispensable pour le calcul du pas à chaque itération rend la méthode très longue et conditionnée par l'existence même de cette dernière. Ainsi, une variante nommée Quasi-Newton ne calcule pas exactement la matrice Hessienne à chaque itération mais utilise une approximation de celle-ci. La plus connue est la méthode BFGS (pour Broyden, Fletcher, Goldfarb et Shanno) [67]. Le temps de convergence s'en trouve significativement amélioré. L'expression de la matrice Hessienne à l'itération $k + 1$ est donnée par la formule suivante :

$$\mathbf{H}^{-1^{k+1}} = \mathbf{H}^{-1^k} - \frac{\mathbf{H}^{-1^k} \mathbf{s}^k (\mathbf{H}^{-1^k} \mathbf{s}^k)^T}{\mathbf{s}^{kT} \mathbf{H}^{-1^k} \mathbf{s}^k} + \frac{q^k q^{kT}}{q^{kT} \mathbf{s}^k} + \phi^k [\mathbf{s}^{kT} \mathbf{H}^{-1^k} \mathbf{s}^k] v^k v^{kT} \quad (1.19)$$

avec $v^k = [\frac{q^k}{q^{kT} \mathbf{s}^k} - \frac{\mathbf{H}^{-1^k} \mathbf{s}^k}{\mathbf{s}^{kT} \mathbf{H}^{-1^k} \mathbf{s}^k}]$ et $q^k = \nabla f(\mathbf{x}^{k+1}) - f(\mathbf{x}^k)$
 $\phi \in [0, 1]$

L'algorithme de Levenberg-Marquardt (LM) : L'algorithme proposé par Levenberg-Marquardt [68, 69] est une combinaison des deux méthodes présentées précédemment en mettant à profit les avantages de chacune d'entre elles. Cette méthode est similaire à la méthode de descente de gradient au début de l'optimisation et se comporte comme la méthode de Gauss-Newton à la fin de l'optimisation si le minimum se rapproche d'une forme quadratique. Ainsi, le pas \mathbf{s}^k à l'itération k est conditionnée par un paramètre λ par (équation 1.20) :

$$[\mathbf{C}]' \mathbf{s} = -\frac{1}{2} \nabla \chi^2 \quad (1.20)$$

avec $[\mathbf{C}]' = [\mathbf{C}] + \lambda \mathbf{I}_d$

où $[\mathbf{C}] = \frac{1}{2} \mathbf{H}$ appelée matrice de courbure, \mathbf{I}_d est la matrice identité de dimension $N \times N$. Si λ est très grand, $[\mathbf{C}]'$ est à diagonale dominante, la matrice Hessienne devient inversible et le pas dans la direction de descente de gradient est donné par (équation 1.21) :

$$\mathbf{s}_{k,j} = -\frac{1}{2\lambda \mathbf{C}_{jj}} \nabla \chi^2 \quad (1.21)$$

Si λ tend vers 0, l'équation équivaut à celui de Gauss-Newton (équation 1.16).

Cette méthode est fonctionnelle, rapide et efficace pour identifier un minimum.

Par contre, toutes ces méthodes ne peuvent identifier, à des rares exceptions près, si le minimum est local ou global. En pratique, il est donc nécessaire de relancer plusieurs simulations à partir de différents points de départ pour contourner ce problème.

L'algorithme de Levenberg-Marquardt (LM) est largement utilisé en scattérométrie dans la résolution du problème inverse [70, 71]. Dans [57], les auteurs proposent notamment une méthode combinant l'algorithme de LM avec les Séparateurs à Vaste Marge (SVM) pour extraire les paramètres d'un profil géométrique.

L’algorithme génétique (AG) : Ce terme regroupe des techniques qui consistent à imiter le comportement et l’évolution de l’espèce humaine [72, 73]. Les AG permettent à une population d’individus (solutions potentielles) d’évoluer par l’intermédiaire d’un certain nombre d’opérateurs stochastiques vers un état qui optimise l’ensemble de leur aptitude ; dans notre cas la fonction notée χ^2 . L’évolution est fondée sur le mécanisme de la sélection naturelle et de la génétique.

Ces algorithmes sont faciles à adapter aux différents types de problème. Leur principal avantage est qu’ils ont la capacité à localiser rapidement le minimum global et permettent ainsi d’éviter les minima locaux. En revanche, ils sont plutôt lents, essentiellement dû à la complexité du problème direct. Les AG sont également utilisés en scattérométrie, ils ont montré une bonne efficacité dans la caractérisation optique et géométrique des couches minces à partir des données ellipsométriques [74]. Dans [19], l’auteur a démontré l’efficacité des AG pour la caractérisation des réseaux de diffraction comparée à la méthode de Levenberg-Marquardt.

1.3.3.2 Méthode des bibliothèques

Le principe de la méthode des bibliothèques appliqué à la scattérométrie consiste à construire une base de données des signatures optiques issues d’une simulation (MMMFE par exemple) indexée chacune par un jeu de paramètres uniques correspondant à une forme géométrique fixée à l’avance. Lors de la caractérisation d’un échantillon, on cherche la signature optique de la base de données la plus proche de celle mesurée sur ce dernier. La recherche de la signature optique dans la base de données est, dans sa forme la plus simple, énumérative : la signature optique mesurée est comparée avec toutes les signatures optiques composant la bibliothèque. La fonction χ^2 (équation 1.13) est souvent le critère qui permet d’évaluer l’écart entre deux signatures optiques. On extrait alors les paramètres géométriques correspondant à la signature optique de la base présentant la valeur minimum de χ^2 .

Cette méthode permet d’éviter les minima locaux si le pas utilisé est très faible. Cependant elle peut s’avérer lente, car elle exige un temps non négligeable pour explorer l’ensemble de la bibliothèque. Pour réduire ce temps de recherche dans la bibliothèque, Soulan [75] a proposé une solution en associant le principe des k plus proches voisins (k -ppv) avec les processeurs graphiques (GPU) pour parcourir la bibliothèque de façon vectorielle. Cette méthode a été testée avec succès pour le suivi de procédés dynamiques comme le fluage de résine [76] ou la réduction de côte de résine dans une application de gravure [77]. D’autres travaux portant sur l’amélioration de la précision des mesures [78] et ceux mettant à profit les SVM dans la méthode des bibliothèques pour l’identification et la reconstruction de profils géométriques [58].

1.3.3.2.3 Les réseaux de neurones artificiels

Le réseau de neurones artificiel (R.N) est un outil qui peut être utilisé pour effectuer une régression non linéaire. Dans le cadre de la scattérométrie, le principe de la méthode consiste à créer un corpus d’apprentissage composé de couples entrée/sortie notés précédemment \mathbf{y}/\mathbf{x} (paragraphe 1.3.1.2). Les entrées sont composées de signatures optiques et les sorties des

paramètres géométriques correspondants à la structure sur laquelle sont effectuées les mesures. Ce jeu de données (signatures/paramètres) est obtenue par simulations numériques (MMMFE par exemple). Pendant la phase d'entraînement, le R.N apprend la relation liant les entrées aux sorties. Enfin, une fois l'apprentissage réalisé, le R.N peut répondre à la sollicitation d'une nouvelle entrée (signature optique) de manière quasi instantanée.

L'une de particularité du R.N est sa résistance au bruit de mesure ; ce qui le rend moins sensible aux minima locaux. Le R.N est capable d'interpoler entre deux résultats proches alors que la méthode des bibliothèques est limitée à fournir un résultat présent dans la base de données. En revanche, le R.N, comme la méthode des bibliothèques, ne peut fournir des résultats efficaces que sur un domaine de variation des paramètres géométriques fixé en amont lors de la création de la base d'apprentissage.

Contrairement à la méthode des bibliothèques, l'avantage réside dans le fait que le traitement de l'information est rapide une fois l'apprentissage réalisé. Les méthodes d'optimisation nécessitent quant à elles le calcul du problème direct (MMMFE) à chaque itération ; ce qui conduit inévitablement à l'augmentation du temps de traitement. Par contre, elles ne sont pas limitées à une plage de variation de paramètres géométriques fixée à l'avance.

De nombreuses études ont montré la performance et l'efficacité du R.N dans la résolution du problème inverse en scattérométrie [42, 44]. Les RNs sont utilisés pour détecter des motifs présentant des rugosités [79]. Ils sont également employés dans le cadre de la classification pour détecter une couche résiduelle considérée comme un défaut de masque en lithographie [61, 62].

Le Laboratoire Hubert Curien (LHC) a depuis quelques années développé un savoir-faire dans le domaine de la résolution du problème inverse plus particulièrement la caractérisation scattérométrique par R.N [56, 63, 80]. Trois thèses ont été validées au sein du laboratoire sur l'utilisation du R.N dans le cadre de la caractérisation scattérométrique. Premièrement, deux thèses [19, 38] ont validé l'apport du R.N dans le cadre de la caractérisation statique. Ensuite, les travaux dans [81] portent sur l'utilisation du R.N en tant que classifieur comme outil d'aide à la décision avant l'étape de la caractérisation proprement dite. Une autre thèse en collaboration avec le Laboratoire de Technologie Micro-électronique (LTM) a étudié la potentialité du R.N dans le cadre d'une caractérisation en temps réel [37].

Il faut noter que le R.N est un outil efficace, précis et bien adapté pour une caractérisation scattérométrique des réseaux en statique ou en dynamique. Ces outils mathématiques ont également montré leur aptitude dans le cadre particulier de la classification et reconnaissance de formes. Le chapitre suivant y est consacré.

1.3.4 Scattérométrie ellipsométrique

1.3.4.1 Définition de l'ellipsométrie

Le principe de l'ellipsométrie classique consiste à mesurer le changement de polarisation d'une onde après réflexion sur la surface d'une structure définie et connue en amont (mono ou

multicouche).

Considérons une onde plane qui arrive sur une surface d'un échantillon avec un angle θ_i , séparant deux milieux d'indice n_0 et n_1 . Une partie de l'onde incidente est transmise ou absorbée à travers la surface et l'autre est réfléchi (figure 1.15). Le champ électrique \mathbf{E}_i peut être décrit par ses deux composantes parallèle au plan d'incidence (E_{ip}) et perpendiculaire au plan d'incidence (E_{is}).

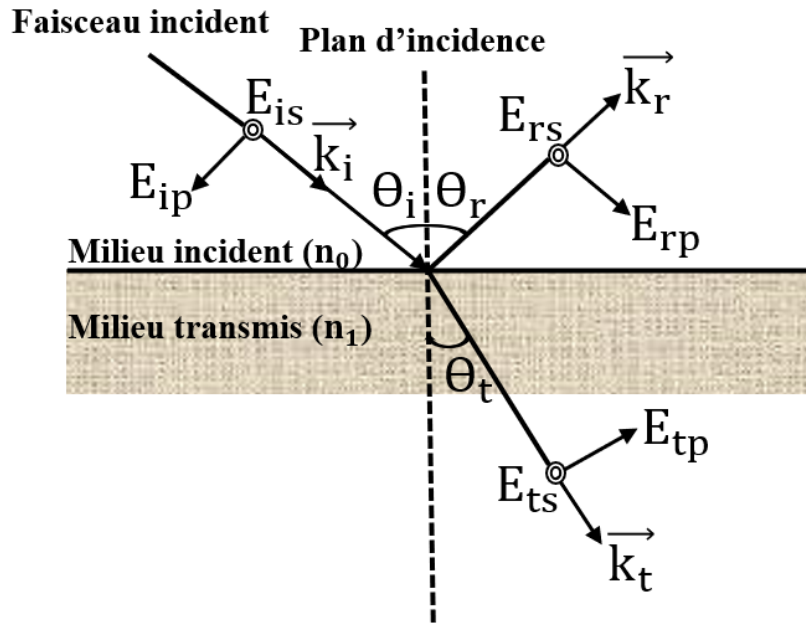


FIGURE 1.15 – Représentation d'une onde plane incidente en réflexion et transmission avec un angle θ_i et d'indice n_0 et n_1 .

Le champ électrique de l'onde incidente \mathbf{E}_i se définit alors (équation 1.22) :

$$\vec{E}_i = \vec{E}_{ip} + \vec{E}_{is} \quad (1.22)$$

Le champ réfléchi s'écrit (équation 1.23) :

$$\vec{E}_r = r_p \vec{E}_{ip} + r_s \vec{E}_{is} \quad (1.23)$$

r_p et r_s sont les coefficients de réflexion en amplitude pour une polarisation *TM* et *TE* tel que (équations 1.24 et 1.25) :

$$r_p = \frac{E_{rp}}{E_{ip}} = |r_p| e^{j\delta_p} \quad (1.24)$$

$$r_s = \frac{E_{rs}}{E_{is}} = |r_s| e^{j\delta_s} \quad (1.25)$$

où r_p et r_s sont les modules des coefficients de réflexion, δ_p et δ_s les retards de phase induits par la réflexion.

En pratique, on ne mesure que le rapport des deux coefficients de réflexion défini par

l'équation fondamentale de l'ellipsométrie suivante (équation 1.26) :

$$\rho = \frac{r_p}{r_s} = \tan\Psi e^{i\Delta} \quad (1.26)$$

Les angles ellipsométriques sont donnés par (équations 1.27 et 1.28) :

$$\Psi = \arctan\left|\frac{r_p}{r_s}\right| \quad 0 < \Psi < 90^\circ \quad (1.27)$$

$$\Delta = \delta_p - \delta_s \quad 0 < \Delta < 180^\circ \quad (1.28)$$

Ψ et Δ dépendent de la structure et sont directement accessibles par la mesure. Le principe de la scattérométrie ellipsométrique consiste à remonter à la forme géométrique de la structure à partir de la mesure directe de Ψ et Δ ou à partir d'autres quantités mesurables dépendants de ces paramètres.

1.3.4.2 Appareil de mesure

L'appareil permettant de mesurer Ψ et Δ est un ellipsomètre. Il se compose d'une source, d'un polariseur (éventuellement d'un compensateur), d'un analyseur et d'un détecteur. La figure 1.16 représente le système d'ellipsomètre classique, où les différents éléments sont arrangés selon un bras source (PSG pour Polarization State Generator) et un bras détection (PSD pour Polarization State Detector). Les angles de l'axe de l'analyseur, du polariseur et celui du compensateur avec le repère fixe sont respectivement θ_A , θ_P et θ_C .

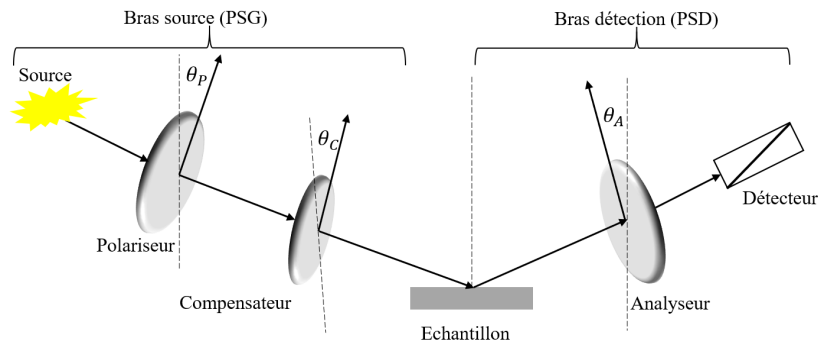


FIGURE 1.16 – Montage de base d'un ellipsomètre classique.

Il existe plusieurs techniques permettant de mesurer des angles ellipsométriques mises à profit par les différents constructeurs. Ainsi, on retrouve l'ellipsomètre à annulation [82], l'ellipsomètre à élément tournant [83] et l'ellipsomètre à modulation de phase [84] que nous utilisons dans le cadre de nos travaux.

Ellipsomètre à modulation de phase : Dans ce type de technique, on introduit un modulateur photoélastique juste après le polariseur (par rapport à la figure 1.16). Un modulateur crée un déphasage périodique sinusoïdal entre les deux composantes (p et s) du champ électrique et donné par (équation 1.29) :

$$\delta(t) = \delta_0 \sin(\omega t) \quad (1.29)$$

δ_0 est l'amplitude de modulation et ω la pulsation du modulateur. La fréquence de modulation est de 50 Hz et la matrice de transfert du modulateur est décrite par (équation 1.30) :

$$M = \begin{pmatrix} e^{j\delta(t)} & 0 \\ 0 & 1 \end{pmatrix} \quad (1.30)$$

En utilisant le formalisme de Jones, on pourra exprimer les composantes du champ électrique sur le détecteur par le calcul matriciel suivant (équation 1.31) :

$$\begin{pmatrix} E_p \\ E_s \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \cos\theta_A & -\sin\theta_A \\ \sin\theta_A & \cos\theta_A \end{pmatrix} \begin{pmatrix} r_p & 0 \\ 0 & r_s \end{pmatrix} \begin{pmatrix} \cos\theta_M & -\sin\theta_M \\ \sin\theta_M & \cos\theta_M \end{pmatrix} \begin{pmatrix} e^{j\delta(t)} & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \cos(\theta_P - \theta_M) & -\sin(\theta_P - \theta_M) \\ \sin(\theta_P - \theta_M) & \cos(\theta_P - \theta_M) \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} E_x \\ E_y \end{pmatrix} \quad (1.31)$$

L'intensité obtenue est de la forme (équation 1.32) :

$$I(t) = |E_p|^2 + |E_s|^2 = I_0 + I_s \sin\delta(t) + I_c \cos\delta(t) \quad (1.32)$$

avec

$$I_0 = B(1 - \cos 2\Psi \cos 2\theta_A + \cos 2(\theta_P - \theta_M) \cos 2M (\cos(2\theta_A - \cos 2\Psi) + \sin 2\theta_A \cos \Delta \cos 2(\theta_P - \theta_M) \sin 2\Psi \sin 2M)) \quad (1.33)$$

$$I_s = B[\sin 2(\theta_P - \theta_M) \sin 2\theta_A \sin 2\Psi \sin \Delta] \quad (1.34)$$

$$I_c = B \sin 2(\theta_P - \theta_M) [(\cos 2\Psi - \cos 2\theta_A) \sin 2\theta_M + \sin 2\theta_A \cos 2\theta_M \sin 2\Psi \cos \Delta] \quad (1.35)$$

dans laquelle :

$$B = \frac{E_0^2}{4|r_p^2 + r_s^2|} \quad (1.36)$$

En mesurant les intensités I_s et I_c , on peut calculer avec précision Ψ et Δ en fixant l'orientation des angles θ_P , θ_M et θ_A suivant deux configurations particulières notées :

— Configuration I ($\theta_M = 0^\circ$, $\theta_P = 45^\circ$ et $\theta_P - \theta_M = \pm 45^\circ$) :

$$I_s = \sin 2\Psi \sin \Delta \quad (1.37)$$

$$I_c = \sin 2\Psi \cos \Delta \quad (1.38)$$

— Configuration II ($\theta_M = 45^\circ$, $\theta_P = 45^\circ$ et $\theta_P - \theta_M = \pm 45^\circ$) :

$$I_s = \sin 2\Psi \sin \Delta \quad (1.39)$$

$$I_c = \cos 2\Psi \quad (1.40)$$

Les valeurs des intensités I_s et I_c varient donc entre -1 et 1 et sont directement liées aux paramètres géométriques recherchés par l'intermédiaire de Ψ et Δ . Ces intensités constituent ce qu'on appelle les signatures optiques de l'échantillon. Notons que la détermination de Ψ et Δ suivant les deux configurations comporte des ambiguïtés :

- Dans la configuration I, Ψ est indéterminé entre Ψ et $90^\circ - \Psi$ par la seule connaissance de $\sin 2\Psi$. Par contre Δ est parfaitement déterminé.
- Dans la configuration II, Δ est indéterminé entre deux valeurs (Δ et $180^\circ - \Delta$). Cette fois ci, Ψ est parfaitement déterminé.

Pour éviter une telle incertitude, généralement on mesure Ψ et Δ dans les deux configurations. Dans notre étude, on mesure directement les intensités I_s et I_c à plusieurs longueurs d'onde sans nécessairement connaître les valeurs de Ψ et Δ ; ce qui nous permettra de s'affranchir de ce genre de configuration.

Nous utilisons dans le cadre de cette thèse, un ellipsomètre à modulation de phase de marque UVISEL de Jobin Yvon (figure 1.17), il est équipé d'un modulateur photo élastique situé, contrairement au cas classique décrit précédemment, sur le bras de détection entre l'analyseur et l'échantillon. Il peut mesurer en mode spectroscopique dans une large gamme qui s'étend de l'ultraviolet et au proche infrarouge (278 nm - 1771 nm). Par ailleurs, le faisceau optique qui éclaire la surface du motif a une taille d'environ 1 mm de section.

Le bruit inhérent à cette mesure peut être réduit en augmentant le temps d'intégration.

1.3.4.3 Signature ellipsométrique

Dans cette partie, nous allons définir la signature ellipsométrique utilisée lors d'une caractérisation scattérométrique. En effet, comme nous venons de le voir, la signature ellipsométrique est composée de mesures d'intensités I_s et I_c pour un spectre de longueurs d'onde donné. Il faut noter que le choix de la signature optique n'est pas anodin et influe grandement sur le résultat de la caractérisation.

Par ailleurs, elle doit pouvoir être simulée efficacement en utilisant les outils numériques.

Ainsi, comme il a été mentionné dans le paragraphe 1.1.3.3, l'utilisation de la MMMFE nécessite la définition préalable de deux paramètres intrinsèques : le nombre de troncature M

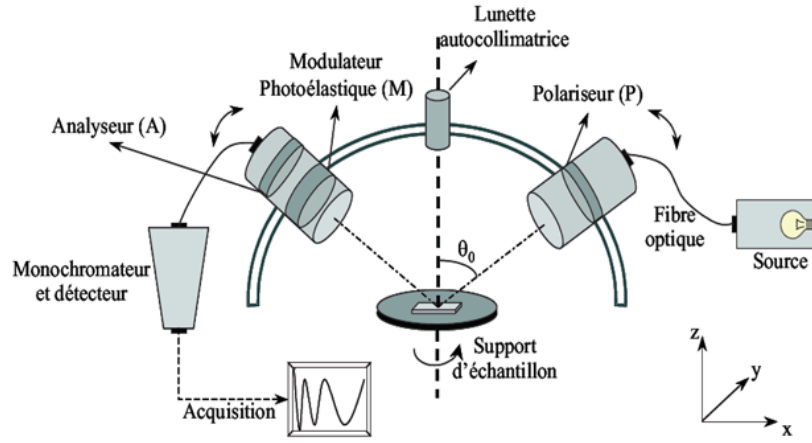


FIGURE 1.17 – Représentation schématique de l'ellipsomètre spectroscopique à modulation de phase (UVISEL de Jobin Yvon) [38].

et le nombre de couches N permettant la modélisation géométrique de l'échantillon. Ce couple conditionne la qualité de la signature optique simulée par la MMMFE. Il faut noter également que pour M et N élevés, l'obtention de signatures optiques consomme beaucoup du temps. Il est ainsi préférable de faire une étude de convergence systématique pour déterminer la valeur optimale de ces paramètres dans le cadre d'une étude fixée. Ces dernières dépendent de plusieurs paramètres notamment la complexité de la forme du profil et les variations d'indices optiques de la structure (couche du réseau, superstrat et substrat). La convergence de M et N peut être étudiée par le calcul de la somme des dérivées de I_s et I_c :

$$C_M = \sum_{\lambda} \left(\left| \frac{\partial I_s}{\partial M} \right| + \left| \frac{\partial I_c}{\partial M} \right| \right) \quad (1.41)$$

$$C_N = \sum_{\lambda} \left(\left| \frac{\partial I_s}{\partial N} \right| + \left| \frac{\partial I_c}{\partial N} \right| \right) \quad (1.42)$$

Considérons deux réseaux en résine ($M78Y$) de période $T = 500 \text{ nm}$ de profils différents. Le premier est un profil trapézoïdal de paramètres géométriques suivants : $b_1 = 30 \text{ nm}$, $b_2 = 180 \text{ nm}$ et $h = 390 \text{ nm}$ et le second est un profil rectangulaire arrondi en haut de paramètres géométriques suivants : $\delta = 0.6$, $CD_h = 210 \text{ nm}$ et $h = 390 \text{ nm}$. Les figures 1.18 et 1.19 représentent respectivement les coefficients C_M en fonction de l'ordre de troncature et C_N en fonction du nombre de couches.

On constate qu'à partir d'une certaine valeur optimale de M et N , la convergence semble largement atteinte. Ces valeurs optimales ne sont pas systématiquement les mêmes pour ces deux profils. On note pour le profil trapézoïdal, les valeurs optimale sont autour de $M = 10$ et $N = 10$ alors que pour le profil rectangulaire arrondi en haut, elles sont de l'ordre de $M = 10$ et $N = 20$. Les intensités I_s et I_c varient peu. On peut donc utiliser ces valeurs comme paramètres de la MMMFE pour ce type de structure. Cette étude de convergence est nécessaire en préambule de chaque construction de corpus de signatures optiques.

En considérant les valeurs optimales de M et N obtenues par l'étude de convergence, la

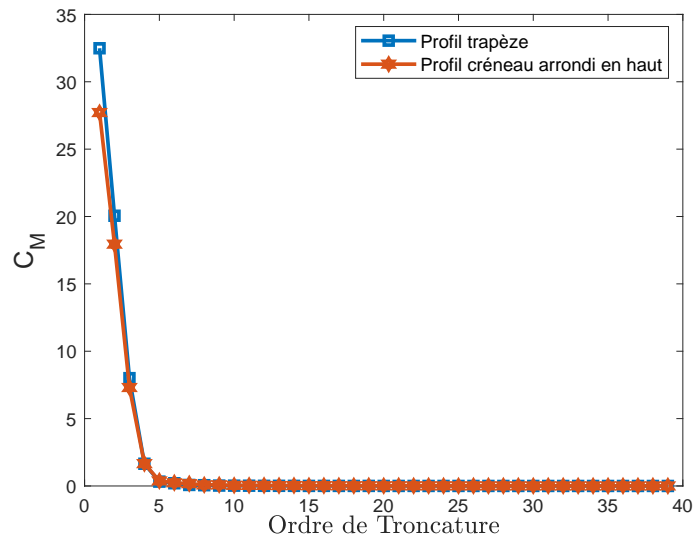


FIGURE 1.18 – Courbe de convergence selon l'ordre de troncature M pour $N = 40$ pour deux réseaux en résine définis par un profil trapèze ($b_1 = 30$ nm, $b_2 = 180$ nm et $h = 390$ nm) et un profil rectangulaire arrondi en haut ($\delta = 0.6$, $CD_h = 210$ nm et $h = 390$ nm).

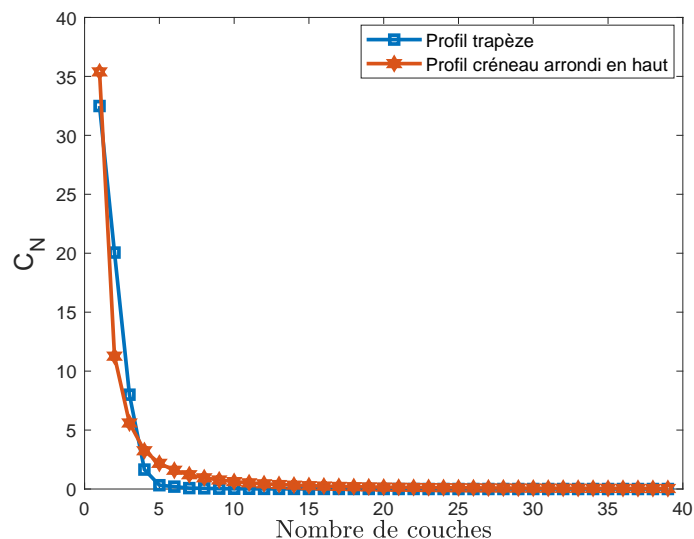


FIGURE 1.19 – Courbe de convergence selon le nombre de couche N pour deux réseaux en résine définis un profil trapèze ($b_1 = 30$ nm, $b_2 = 180$ nm et $h = 390$ nm avec $M = 10$) et un profil rectangulaire arrondi en haut ($\delta = 0.6$, $CD_h = 210$ nm et $h = 390$ nm avec $M = 10$).

signature optique du profil trapézoïdal peut ainsi être calculée théoriquement par la MMMFE. Ensuite, on tient compte de la simulation de la mesure en introduisant un bruit gaussien ($3\sigma = 0.01$) correspondant à la perturbation d'une mesure expérimentale. Ainsi, la figure 1.20 représente la signature optique mesurée et celle simulée théoriquement à partir d'un profil trapézoïdal supposé connu en fonction de la longueur d'onde.

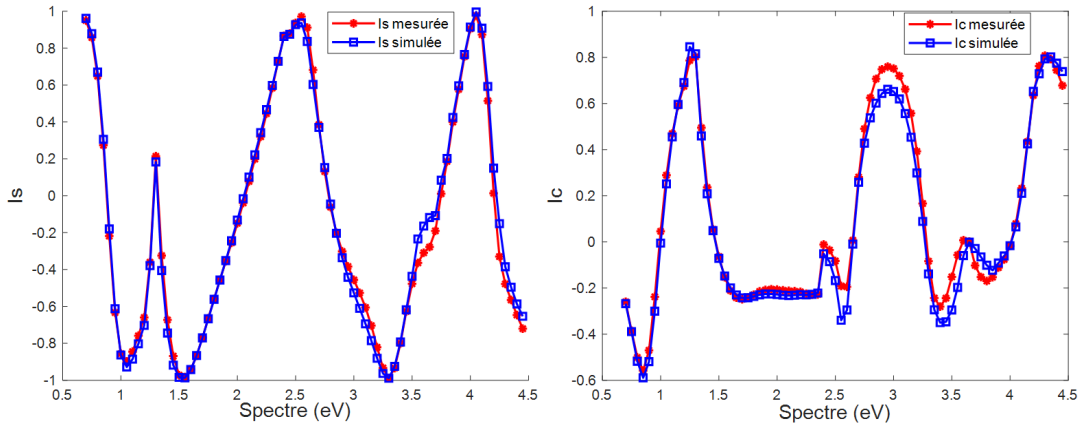


FIGURE 1.20 – Représentation des signatures optiques mesurées (rouge) et simulées à partir d'un profil trapézoïdal ($b_1 = 5$ nm, $b_2 = 200$ nm et $h = 300$ nm) (bleu).

Il apparaît clairement que la signature optique simulée est proche de la signature optique mesurée. Cependant, sur certaines parties du spectre, certains écarts sont significatifs. Cela peut provenir d'une modélisation imparfaite de la structure géométrique, la mauvaise connaissance des indices ou encore de la présence de rugosité.

1.3.5 Estimation des performances de la caractérisation

En scattérométrie, la performance de la caractérisation est donnée par l'estimation de la qualité du résultat obtenu.

En théorie, il existe une méthode directe pour contrôler la qualité de la caractérisation. Il suffit de calculer la différence géométrique entre la forme réelle et la forme reconstruite par :

$$E_{geo} = \int_0^T |h(x) - l(x)| dx \quad (1.43)$$

où $h(x)$ et $l(x)$ représentent respectivement le profil réel et celui reconstruit sur une période T

Dans le cas où on veut normaliser cet écart par rapport à la surface du réseau obtenu, on définit l'écart géométrique normalisé $E_{geonorm}$ par :

$$E_{geonorm} = \frac{E_{geo}}{E_{geoprofil}} * 100 \quad (1.44)$$

avec

$$E_{geoprofil} = \int_0^T |l(x)| dx$$

En guise d'exemple, nous avons calculé la signature optique par MMMFE d'un échantillon fictif de profil trapézoïdal ($b_1 = 30 \text{ nm}$, $b_2 = 180 \text{ nm}$ et $h = 390 \text{ nm}$). Dans tout ce qui suit un bruit gaussien correspondant à la perturbation d'une mesure expérimentale ($3\sigma = 0.01$) sera introduit. La signature obtenue est ainsi utilisée comme mesure expérimentale dans un processus de résolution du problème inverse en fonctionnant sur un modèle rectangulaire. La figure 1.21 montre la représentation du profil reconstruit par rapport au profil initial.

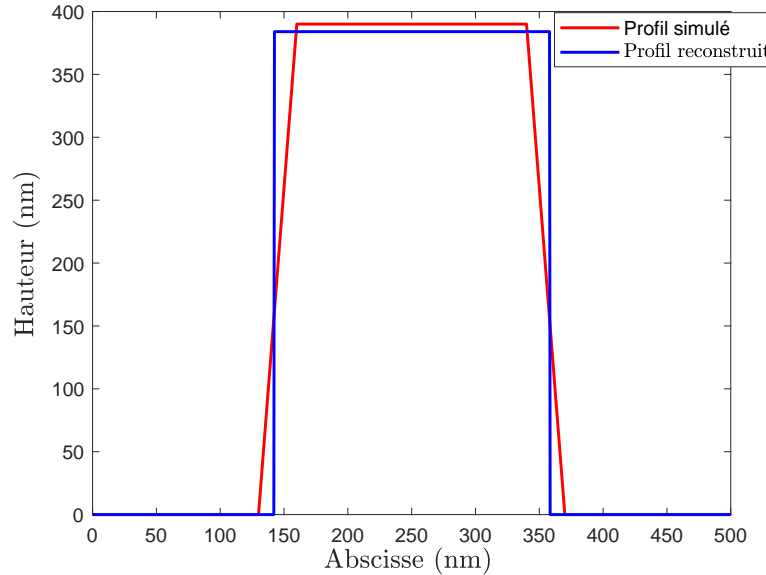


FIGURE 1.21 – Représentation graphique d'un profil trapézoïdal simulé ($b_1 = 30 \text{ nm}$, $b_2 = 180 \text{ nm}$ et $h = 390 \text{ nm}$) (rouge) et du profil reconstruit (bleu) à partir de la signature optique théorique bruitée.

On constate alors une différence géométrique notable entre les deux profils (simulé et reconstruit). Il faut préciser également que le profil simulé présente des pentes qui ne peuvent être prises en compte par le modèle de caractérisation choisi.

Malheureusement, ce critère d'écart géométrique E_{geo} n'est calculable que sur une forme théorique et ne peut être utilisé en pratique car on ne connaît pas le profil réel.

Ainsi, on utilise généralement comme alternative, l'écart entre les signatures optiques simulées ($I_{simulées}$) à partir de l'estimation de la structure et les signatures optiques mesurées ($I_{mesurées}$). On calcule alors l'erreur quadratique moyenne (E_{I_s, I_c}) [63, 78, 85] donnée par :

$$E_{I_s, I_c} = \sqrt{\frac{1}{2N_m} \sum_{i=1}^{N_m} ((I_{smesurées} - I_{ssimulées})^2 + (I_{cmesurées} - I_{csimulées})^2)} \quad (1.45)$$

où N_m représente le nombre de longueurs d'ondes composants la signature

Ce critère d'erreur E_{I_s, I_c} est également utilisé comme la fonction coût à minimiser pour déterminer les paramètres géométriques optimaux notamment par optimisation classique (équation 1.13). Ainsi, il est d'usage de ne pas calculer cette erreur sur le même jeu de données pour ne pas biaiser l'estimation des performances.

1.4 Conclusion

Dans ce chapitre, nous avons présenté le contexte de la métrologie en microélectronique ainsi que les différentes techniques employées. Nous avons passé en revue les méthodes optiques indirectes en insistant plus particulièrement sur les différentes méthodes de traitement numériques permettant la résolution du problème inverse. Nous avons également détaillé le principe de la scattérométrie ellipsométrique largement utilisée dans ces travaux.

Le chapitre suivant est consacré aux méthodes de classification automatique et plus particulièrement aux techniques neuronales mises à profit dans le cadre particulier de la scattérométrie.

Classification et Reconnaissance de formes

La classification automatique est un domaine de la reconnaissance de formes qui est inspirée de la capacité humaine à percevoir l'environnement. Elle a la capacité de discerner, distinguer et interpréter des données à partir d'une observation. Cette discipline utilise notamment les techniques d'apprentissage automatique et statistique pour traiter les données.

Dans ce chapitre, nous présentons le principe général de la reconnaissance de formes en exposant plus particulièrement le domaine de la classification automatique. Nous décrivons les différentes techniques de la classification existantes en passant en revue les méthodes de classification directe et indirecte. Enfin, nous aborderons plus en détail le réseau de neurones utilisé dans le cadre de cette thèse.

2.1 Introduction à la reconnaissance de formes

La reconnaissance de formes (RDF) est l'élaboration des théories et des méthodes de classification des observations $\mathbf{X} = (x_1, x_2, x_3, \dots, x_i)$. La classification est une discipline qui consiste à identifier la classe d'un objet à partir des observations issues de ce dernier. Une classe représente un ensemble des données ayant des caractéristiques semblables. On parle de classification (ou discrimination) si les classes sont définies à l'avance et de regroupement (ou clustering) dans le cas contraire.

2.1.1 Principe général

Le but de la RDF est d'identifier le modèle d'un motif associé à chaque classe. En effet, la forme désigne une observation quelconque. Elle peut être de diverse nature telle qu'une image, un son, une empreinte digitale, un visage, etc. Cette identification requiert le recours à des techniques de classification automatique [86] permettant de regrouper par apprentissage, ces objets par rapport à des critères de similarités. La conception d'un système de la RDF nécessite en général plusieurs étapes (acquisition, prétraitement, extraction, classification ou clustering, décision, etc.) pour prédire la classe d'un objet (figure 2.1). Ainsi, l'apprentissage constitue le cœur de la plupart des méthodes. Pour cela il utilise un ensemble de données appelé également corpus.



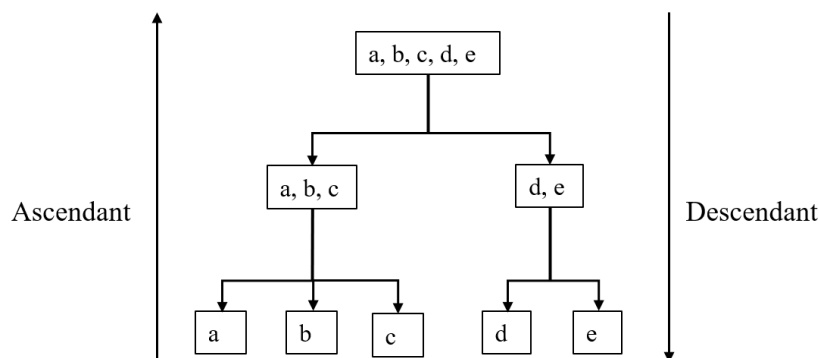
FIGURE 2.1 – Schéma des différentes étapes du système de classification d'un objet.

On distingue deux types d'apprentissage : l'apprentissage supervisé et l'apprentissage non supervisé. L'apprentissage supervisé est réalisé dans le cas où les classes sont connues au préalable. Pour l'apprentissage non supervisé, on cherche à former ou regrouper les données selon des ressemblances sans connaissance *a priori*. En pratique, on cherche à vérifier 2 propriétés :

- compacité : les données doivent être les plus similaires possibles au sein d'une classe.
- séparabilité : les classes doivent également être aussi dissemblables que possible.

Les méthodes de regroupement sont basées sur la notion de distance. Elles sont subdivisées en deux catégories : les méthodes hiérarchiques et les méthodes non hiérarchiques.

Les méthodes hiérarchiques [87] sont les premières techniques de regroupement qui furent développées. Elles permettent d'identifier des groupes d'observations similaires dans un jeu de données. Ces groupes créés sont présentés sous la forme d'une structure arborescente représentant la hiérarchie (figure 2.2). Ces méthodes sont soit ascendantes soit descendantes. Pour l'approche ascendante, les observations sont regroupées en petits agglomérats puis fusionnées progressivement en un seul groupe pour aboutir à la construction d'une structure arborescente ou d'un dendrogramme. Quant à l'approche descendante, elle procède en sens inverse par subdivisions progressives de l'ensemble de données en des petits groupes. L'avantage de ces méthodes est de permettre une lecture sous la forme d'un dendrogramme avec un nombre optimal de classes. Par contre, elles consomment beaucoup du temps en termes de calcul.

FIGURE 2.2 – Exemple d'une structure arborescente hiérarchique pour différentes observations (*a*, *b*, *c*, *d* et *e*).

Les méthodes non hiérarchiques ou méthodes de partitionnement [88] ont connues leur essor avec l'avènement des ordinateurs. Le principe de la méthode consiste à partitionner un ensemble d'observations en classes disjointes. On cherche alors d'une part, à rassembler des observations semblables, et d'autre part, à éviter que les classes différentes soient proches l'une de l'autre. Cette méthode de partition présente l'avantage non négligeable de pouvoir classer rapidement un grand nombre d'observations.

Les méthodes de classification ont pour objectif d'identifier les classes auxquelles appartiennent des observations parmi un ensemble prédéfini. En effet, l'attribution d'une classe est fondée sur la notion de proximité permettant d'évaluer la distance entre deux vecteurs [89]. On note ainsi deux catégories de méthodes de classification : les méthodes de classification indirecte et les méthodes de classification directe.

Les méthodes de classification indirecte fonctionnent sur le principe du théorème de Bayes. Elles sont basées sur un modèle statistique qui cherche à estimer la probabilité *a posteriori* $p(C_k/\mathbf{X})$ d'appartenance à la classe C_k sachant l'observation \mathbf{X} . Cette probabilité est celle qui permet de prendre la décision d'affectation. On distingue alors les méthodes paramétriques et non paramétriques. L'approche paramétrique suppose que les lois de probabilités obéissent à un modèle de distribution connu et on cherche à déterminer ses paramètres. Alors que l'approche non paramétrique est basée sur le principe de convergence pour estimer la probabilité *a posteriori* d'une classe [90].

Les méthodes de classification directe n'utilisent pas la règle de décision de Bayes. Elles estiment directement la probabilité *a posteriori* d'appartenance à la classe C_k sans passer par la phase des densités de probabilité. A noter que, le réseau de neurones, utilisé en tant que classifieur dans le cadre de ces travaux, fait partie de ces méthodes.

2.1.2 Applications

La RDF est un domaine d'applications et de recherche dans lesquels des algorithmes prennent des décisions quant à l'appartenance d'un objet à une classe en fonction des observations fournies en entrée. Elle a notamment trouvé une place prépondérante dans le domaine du traitement d'images tel que l'imagerie biomédical. Citons par exemple la détection des cellules cancéreuses du cerveau [91], la détection du cancer du sein [92] ou, plus généralement l'aide au diagnostic médical [93]. Elle est également très utilisée pour la reconnaissance des caractères manuscrits [94, 95], le tri automatique du courrier [96] ou l'analyse de visage [97]. On l'emploie également dans le domaine de la sécurité informatique pour la détection d'anomalies dans les réseaux [98], dans la bio-informatique pour la détection de l'homologie des protéines [99], en automatique, pour diagnostiquer le défaut des rouleaux de tramway [100]. Elle est également présente dans d'autres domaines comme l'acoustique [101], le marketing [102], la finance [103], etc.

En scattérométrie, on cherche également à transférer ce savoir-faire pour contrôler la qualité des différentes étapes d'un procédé de fabrication [61, 62]. L'objectif de cette thèse s'inscrit dans ce cadre. En effet, on cherche à identifier la forme géométrique d'une structure périodique à partir d'observations mesurables au cours des différentes étapes de fabrication.

2.1.3 Prétraitement des données

Le prétraitement des données est une étape clé du processus de classification. Il conditionne les performances globales du système. En effet, les données d'entrées contiennent des informations brutes, bruitées ou incomplètes susceptibles d'influencer la prise de décision. Il est donc nécessaire d'effectuer un prétraitement en amont. L'idée est d'extraire ou de sélectionner les informations les plus pertinentes à la détection de classes tout en supprimant celles n'ayant aucune influence supposée.

Pour cela, les techniques d'analyse des données [104] sont largement utilisées. Parmi les plus courantes, citons les méthodes linéaires comme l'Analyse en Composante Principale (ACP) [105]. Développée pour l'étude de l'analyse statistique, l'ACP est plus généralement employée pour l'analyse des données multidimensionnelles. Elle est utilisée pour réduire la dimension du vecteur d'entrées afin de simplifier la construction du modèle en aval. Son principe de fonctionnement est basé sur la représentation de l'ensemble des données dans un sous espace de dimension plus petite. L'ACP permet ainsi de réduire le nombre d'entrées tout en conservant la répartition des données. Cette technique se présente comme un outil de projection linéaire qui maximise l'inertie du nuage de points des données.

Notons cependant que l'ACP diminue le nombre d'entrées sans les supprimer. En effet, chaque nouvelle entrée est une combinaison linéaire des anciennes. La réduction du nombre d'entrées permet cependant de réduire les dimensions du problème conduisant à simplifier l'architecture du processus de classification qui suit. Ceci permet de réduire le risque de convergence prématurée et donc d'accroître les performances en généralisation.

Il existe de nombreuses variantes à cette méthode comme l'analyse en composante curviligne (ACC) [106] ou les cartes auto-organisatrices [107] qui peuvent être vues comme une extension non linéaire de l'ACP.

Cependant, ces techniques ne seront pas mises à profit dans la suite de nos études. Dans le cadre de nos travaux, les données d'entrées utilisées sont des signatures ellipsométriques issues de la mesure sur une structure périodique. Ces données bruitées seront simplement centrées (moyenne nulle) et réduites (écart type de 1) avant d'être traitées. Chaque élément x_i du vecteur des observations \mathbf{X} sera transformé en x'_i par :

$$x'_i = \frac{x_i - \mu_i}{\sigma_i} \quad (2.1)$$

avec μ_i la moyenne et σ_i l'écart type calculés sur chacune des entrées i

2.1.4 Notion de proximité

Dans le domaine de la classification, la mesure de proximité est une notion essentielle. Considérons deux vecteurs $\mathbf{X} = (x_1, x_2, \dots, x_n)$ et $\mathbf{Y} = (y_1, y_2, \dots, y_n)$ définis dans un ensemble E . La proximité entre ces deux observations est qualifiée par la distance f définie dans l'espace

des données $E \times E$ dans \mathbb{R}^+ . Elle doit vérifier les propriétés suivantes :

- **Symétrie** : $\forall \mathbf{X}, \mathbf{Y} \in E^2, f(\mathbf{X}, \mathbf{Y}) = f(\mathbf{Y}, \mathbf{X})$
- **Inégalité triangulaire** : $\forall \mathbf{X}, \mathbf{Y}, \mathbf{Z} \in E^3, f(\mathbf{X}, \mathbf{Z}) \leq f(\mathbf{X}, \mathbf{Y}) + f(\mathbf{Y}, \mathbf{Z})$
- **Séparabilité** : $\forall \mathbf{X}, \mathbf{Y} \in E^2, \mathbf{X} \neq \mathbf{Y} \Rightarrow f(\mathbf{X}, \mathbf{Y}) > 0$

L'expression de la distance est donnée par :

$$f_p(\mathbf{Y}, \mathbf{X}) = \left[\sum_{i=1}^n |y_i - x_i|^p \right]^{\frac{1}{p}} \quad (2.2)$$

- Si $p = 1$, on définit la distance de city-block ou Manhattan :

$$f_1(\mathbf{Y}, \mathbf{X}) = \sum_{i=1}^n |y_i - x_i| \quad (2.3)$$

- Si $p = 2$, on retrouve la distance euclidienne (la plus utilisée) :

$$f_2(\mathbf{Y}, \mathbf{X}) = \sqrt{\sum_{i=1}^n |y_i - x_i|^2} \quad (2.4)$$

- Si $p = \infty$, on parle de la distance de Chebychev ou distance max :

$$f_\infty(\mathbf{Y}, \mathbf{X}) = \max_i |y_i - x_i| \quad (2.5)$$

Le choix de la distance joue un rôle important sur la topologie des classes obtenues. Par exemple, la distance euclidienne contribue à la formation de formes hypersphériques alors que celle de Chebychev implique de formes hypercubiques. La figure 2.3 illustre la forme de l'ensemble de points équidistants pour différentes distances utilisées dans un espace tridimensionnel.

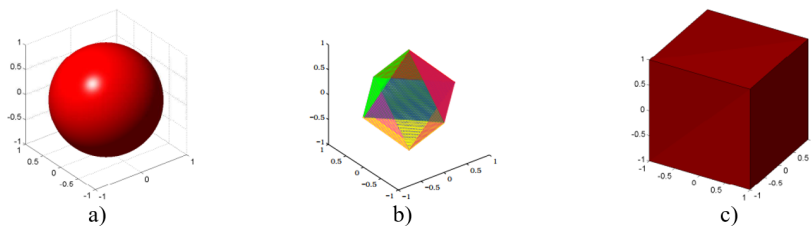


FIGURE 2.3 – Exemple de topologie de classe obtenue suivant le choix de la distance utilisée dans un espace tridimensionnel : a) euclidienne, b) city-block et c) Chebychev [108].

Il existe également une autre distance utilisée dans la classification appelée distance de Mahalanobis [109]. Elle permet d'obtenir des formes ellipsoïdales, et prend en compte la variance et la corrélation de données :

$$f_M(\mathbf{Y}, \mathbf{X}) = \sqrt{(\mathbf{Y} - \mathbf{X})^T \Sigma^{-1} (\mathbf{Y} - \mathbf{X})} \quad (2.6)$$

où Σ est la matrice de covariance associée aux vecteurs \mathbf{Y} et \mathbf{X} et définie par :

$$\Sigma = \frac{1}{n} \sum_i^n (\mathbf{Y}_i - \boldsymbol{\mu}_y)(\mathbf{X}_i - \boldsymbol{\mu}_x)^T \quad (2.7)$$

Lorsque les vecteurs \mathbf{Y} et \mathbf{X} sont indépendants, alors Σ devient égale à l'unité et f_M correspond à la distance euclidienne. Il faut noter que l'utilisation de cette distance est conditionnée par les conditions d'existence de Σ^{-1} .

2.2 Les méthodes classiques de classification

L'objectif de ce paragraphe est de présenter le fonctionnement des techniques classiques de classification. Nous évoquerons l'approche indirecte de classification où nous passerons en revue les méthodes paramétriques et non paramétriques. Nous exposerons également l'approche directe de classification en présentant le séparateur à vaste marge.

2.2.1 Les méthodes de la classification indirecte

2.2.1.1 Généralités et définitions

L'approche indirecte de la classification est fondée sur la règle de décision de Bayes pour laquelle le problème est résolu en termes de probabilité. L'ensemble des données est composé de vecteurs \mathbf{X} appartenant chacun à une classe C_k . La densité de probabilité conditionnelle pour chaque vecteur \mathbf{X} est définie par $p(\mathbf{X}/C_k)$. La probabilité *a priori* pour chaque classe C_k est définie par $p(C_k)$ qui est supposée connue. Le théorème de Bayes permet de calculer les probabilités *a posteriori* $p(C_k/\mathbf{X})$ suivant :

$$p(C_k/\mathbf{X}) = \frac{p(C_k)p(\mathbf{X}/C_k)}{\sum_{k=1}^{K} p(C_k)p(\mathbf{X}/C_k)} \quad (2.8)$$

où K est le nombre total de classes

La règle de décision de Bayes consiste à affecter à l'observation \mathbf{X} , la classe dont la probabilité *a posteriori* est la plus grande :

$$\mathbf{X} \in C_k \iff p(C_k/\mathbf{X}) > p(C_j/\mathbf{X}) \quad \forall k \neq j \quad \text{pour } j = 1, 2, \dots, K$$

La règle de décision de Bayes et le théorème de Bayes constituent ce qu'on appelle le classifieur de Bayes considéré comme optimal. La figure 2.4 représente une illustration géométrique de ce classifieur dans le cas d'un classement en 2 classes (C_1 et C_2).

Cependant, il est difficile de connaître exactement les densités de probabilités conditionnelles $p(\mathbf{X}/C_k)$ des différentes classes C_k . Il existe des méthodes qui permettent d'estimer ces densités de probabilités en l'occurrence les méthodes paramétriques et non paramétriques.

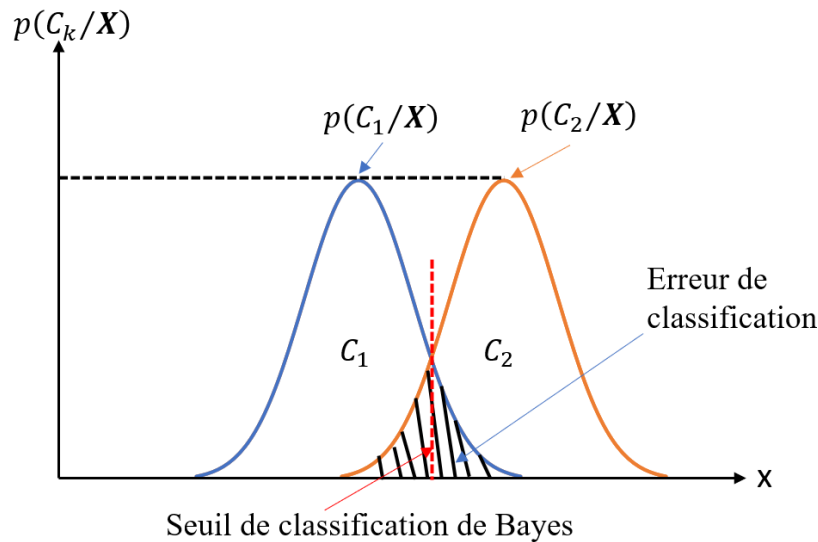


FIGURE 2.4 – Interprétation géométrique de la règle de décision de Bayes sur le classement d'un vecteur \mathbf{X} en 2 classes distinctes (C_1 et C_2).

2.2.1.2 Les méthodes paramétriques

Dans les méthodes paramétriques, on suppose que les lois de probabilités obéissent à un modèle de distribution connue dont les paramètres peuvent être déterminés à partir d'un ensemble des données. Parmi ces approches paramétriques, le classifieur gaussien suppose une densité de probabilité conditionnelle en estimant les moyennes $\boldsymbol{\mu}_k$ correspondant à chaque vecteur \mathbf{X} pour la classe C_k .

$$p(\mathbf{X}/C_k) = \frac{1}{(2\pi)^{\frac{n}{2}} (\det \Sigma_k)^{\frac{1}{2}}} \exp \left[-\frac{1}{2} (\mathbf{X} - \boldsymbol{\mu}_k)^T \Sigma_k^{-1} (\mathbf{X} - \boldsymbol{\mu}_k) \right] \quad (2.9)$$

avec Σ_k la matrice de covariance de dimension $n \times n$ associé à \mathbf{X} et $\boldsymbol{\mu}_k$ de dimension n

Le classifieur gaussien a, par exemple, montré son efficacité pour l'identification des propriétés chimiques des huiles d'olive [110].

2.2.1.3 Les méthodes non paramétriques

Contrairement à la méthode paramétrique, l'approche non paramétrique ne fait aucune l'hypothèse sur les distributions *a priori* des données. L'estimation de la densité de probabilité conditionnelle $p(\mathbf{X}/C_k)$ consiste, dans un premier temps, à délimiter une région R autour d'une observation \mathbf{X} . Puis, le nombre d'observations dans le volume V défini par R est comptabilisé. Enfin, on estime la densité de probabilité par (équation 2.10) :

$$\hat{p}(\mathbf{X}) = \frac{K_N}{NV} \quad (2.10)$$

où N est le nombre d'observations total, K_N le nombre d'exemples contenu dans le volume V

Il existe trois conditions qui peuvent garantir la convergence de $\hat{p}(\mathbf{X})$ vers $p(\mathbf{X})$:

$$\lim_{N \rightarrow \infty} V = 0 \quad (2.11)$$

$$\lim_{N \rightarrow \infty} K_N = 0 \quad (2.12)$$

$$\lim_{N \rightarrow \infty} \frac{K_N}{N} = 0 \quad (2.13)$$

Des nombreuses techniques utilisent ce principe de fonctionnement : les noyaux de Parzen [111] (où on fixe V et on cherche le K_N) et la méthode de K_N plus proches voisins [112] (pour lesquels K_N est fixé et on cherche à déterminer V). Dans ce qui suit, seule la méthode de K_N plus proches voisins sera décrite.

L'approche par la technique des K_N plus proches voisins (K-PPV) consiste à affecter une observation \mathbf{X} à la classe C_k la plus représentée parmi les K-PPV de \mathbf{X} . En pratique, on choisit des volumes réguliers centrés en \mathbf{X} (hypercubes ou hypersphères). On utilise alors la mesure de distance (paragraphe 2.1.4) pour déterminer le type de voisinage.

Ainsi, la densité de probabilité $\hat{p}(\mathbf{X}/C_k)$ de la classe C_k au point \mathbf{X} est estimée par :

$$\hat{p}(\mathbf{X}/C_k) = \frac{K_k}{N_k V} \quad (2.14)$$

où N_k et K_k sont respectivement le nombre d'individus et le nombre d'objets appartenant à la classe C_k parmi le K_N fixé contenu dans le volume V .

Il faut noter que la méthode des K-PPV nécessite beaucoup du temps de calcul. En effet, elle nécessite un grand nombre N d'observations pour estimer de façon fiable les plus proches voisins d'une observation \mathbf{X} . Cela implique donc le calcul de la distance à \mathbf{X} de toutes les observations N . Ainsi, plus N augmente, plus l'approximation de la probabilité est bonne et plus le temps de calcul est élevé.

Par conséquent, plusieurs auteurs ont proposé des algorithmes plus rapides pour éviter une recherche exhaustive du plus proche voisin. Dans [113], l'auteur développe un algorithme où l'espace est divisé en plusieurs cellules. La recherche du plus proche voisin est alors limitée à quelques unes des données seulement. D'autres techniques structurent les données sous la forme d'un arbre en utilisant la notion de regroupement par hiérarchie [114]. Une autre, qualifiée de condensation, vise à réduire l'espace total en un sous-espace [115]. Il existe aussi des algorithmes basés sur l'idée de tri des données du corpus sur chaque axe de projection pour ainsi limiter la région de recherche [116, 117].

Le classifieur du plus proche voisin est une technique qui est également connue pour réaliser des tâches de reconnaissance. Dans [118], cet outil est utilisé pour détecter de possibles défauts dans des systèmes industriels.

2.2.2 Les méthodes de la classification directe

2.2.2.1 Généralités et définitions

L'approche directe de classification n'est pas fondée sur la règle de décision de Bayes. Elle estime directement la probabilité *a posteriori* d'appartenance à une classe. Nous allons exposer les fondements théoriques du Séparateur à Vaste Marge (Machine à Vecteur Support ou SVM) qui est l'une des méthodes les plus utilisées.

2.2.2.2 Les Séparateurs à Vaste Marge

Les SVM sont des techniques d'apprentissage supervisé, conçus pour résoudre un problème de classification. La méthode consiste à projeter les données d'apprentissage sur un espace approprié et à séparer l'espace en deux parties à l'aide d'un hyperplan. Cette distance entre la frontière qui sépare les données des différentes classes est appelée marge ou marge maximale. Les données les plus proches de l'hyperplan sont appelées les vecteurs supports.

2.2.2.2.1 SVM linéaire

Soit D un ensemble des données défini par $(\mathbf{X}_1, Y_1), (\mathbf{X}_2, Y_2), \dots, (\mathbf{X}_N, Y_N)$, linéairement séparable en deux classes C_1 et C_2 .

$$D = \{(\mathbf{X}_1, Y_1), (\mathbf{X}_2, Y_2), \dots, (\mathbf{X}_N, Y_N)\}, \quad \mathbf{X}_i \in \mathbb{R}^n, \quad Y_i \in \{-1, +1\}$$

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

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

$$\langle \mathbf{W}, \mathbf{X} \rangle + b \tag{2.15}$$

tel que :

$$\forall \mathbf{X} \in C_2 \quad \langle \mathbf{W}, \mathbf{X} \rangle + b > 0 \tag{2.16}$$

$$\forall \mathbf{X} \in C_1 \quad \langle \mathbf{W}, \mathbf{X} \rangle + b < 0 \tag{2.17}$$

avec $\langle \mathbf{W}, \mathbf{X} \rangle = \mathbf{W}^T \mathbf{X}$, le produit scalaire entre les vecteurs \mathbf{W} et \mathbf{X}

Le vecteur \mathbf{W} définit ainsi la pente de l'hyperplan, le scalaire b permet de translater l'hyperplan parallèlement à lui-même. Il existe ainsi une multitude d'hyperplans qui vérifient ces conditions. L'objectif est de trouver l'hyperplan qui garantit une distance maximale entre une frontière de séparation et les exemples les plus proches appelées marges maximales. La marge représente donc la distance euclidienne prise entre l'hyperplan et l'exemple \mathbf{X} . Elle s'exprime par :

$$\frac{|\langle \mathbf{W}, \mathbf{X} \rangle + b|}{\|\mathbf{W}\|} \quad (2.18)$$

L'hyperplan optimal noté H défini par les paramètres (\mathbf{W}^*, b^*) est celui qui maximise la marge. Il est ainsi solution de l'équation :

$$\arg \max_{\mathbf{W}^*, b^*} (\min_i \|\mathbf{X} - \mathbf{X}_i\| \quad \text{tel que} \quad \mathbf{X} \in \mathbb{R}^n, (\mathbf{W}^T \mathbf{X} + b) = 0, \quad i = 1, \dots, N) \quad (2.19)$$

Les classes C_1 et C_2 étant linéairement séparables, on peut alors contraindre les points de chaque classe à satisfaire les conditions suivantes :

$$\langle \mathbf{W}, \mathbf{X} \rangle + b \geq 0 \quad \forall i \quad \text{tel que} \quad Y_i = +1 \quad (2.20)$$

$$\langle \mathbf{W}, \mathbf{X} \rangle + b \leq 0 \quad \forall i \quad \text{tel que} \quad Y_i = -1 \quad (2.21)$$

En combinant ces deux équations, on obtient :

$$Y_i \langle \mathbf{W}, \mathbf{X} \rangle + b \geq 0 \quad \forall i = 1, \dots, N \quad (2.22)$$

Ainsi, on définit deux hyperplans H_1 et H_2 contenant les exemples de chaque classe et qui sont appelés hyperplans canoniques par :

$$H_1 : \langle \mathbf{W}^*, \mathbf{X} \rangle + b_1 = +1 \quad (2.23)$$

$$H_2 : \langle \mathbf{W}^*, \mathbf{X} \rangle + b_2 = -1 \quad (2.24)$$

Les points qui se situent sur l'hyperplans H_1 et H_2 sont appelés des vecteurs supports. La distance entre H_1 et H_2 est appelée marge et vaut donc $\frac{2}{\|\mathbf{W}^*\|}$ (figure 2.5). La maximisation de la marge revient donc à résoudre le problème de minimisation suivant :

$$\begin{cases} \text{minimiser} & \tau(\mathbf{W}) = \frac{1}{2} \|\mathbf{W}\|^2 \\ \text{sous les contraintes} & Y_i (\langle \mathbf{W}, \mathbf{X} \rangle + b) \geq 1 \quad \forall i = 1, \dots, N \end{cases} \quad (2.25)$$

Cette écriture est appelée formulation primale des SVMs.

L'optimisation sous contrainte est complexe et nécessite l'emploi de méthodes de Lagrange. Ainsi, on introduit dans la fonction coût un terme correspondant à une combinaison linéaire des contraintes. Dans ce type de cas, on résout en général, la formulation duale du problème.

$$L(\mathbf{W}, b, \boldsymbol{\alpha}_i) = \frac{1}{2} \|\mathbf{W}\|^2 - \sum_{i=1}^N \alpha_i [Y_i (\langle \mathbf{W}, \mathbf{X} \rangle + b) - 1] \quad (2.26)$$

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

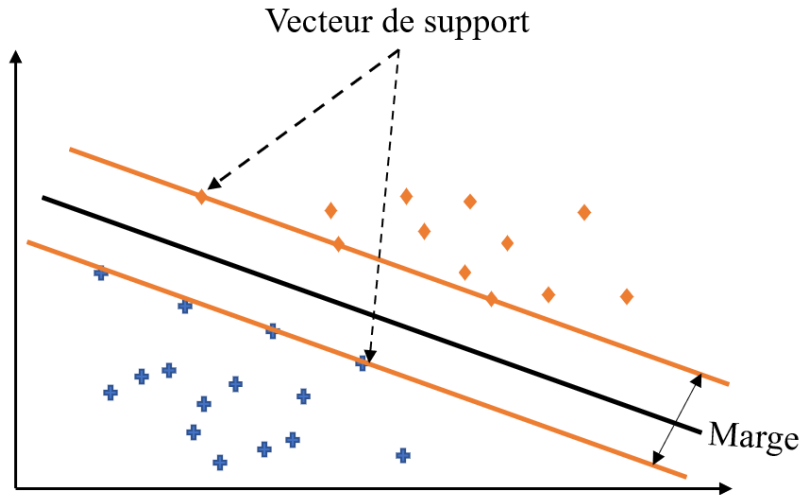


FIGURE 2.5 – Définition de l'hyperplan optimal et de la marge d'un classifieur SVM dans le cas d'une classification en 2 classes linéairement séparables dans un espace à 2 dimensions.

Le Lagrangien L est maximisé par rapport à α et minimisé par rapport aux paramètres \mathbf{W} et b selon les conditions de Kuhn-Tucker [119] :

$$\mathbf{W} = \sum_{i=1}^N \alpha_i Y_i \mathbf{X}_i \quad (2.27)$$

$$\sum_{i=1}^N \alpha_i Y_i = 0 \quad (2.28)$$

Tous les points \mathbf{X}_i qui ne sont pas des vecteurs supports (ne vérifiant pas l'équation $[Y_i(\langle \mathbf{W}, \mathbf{X}_i \rangle + b) - 1] = 0$) sont associés à des α_i nuls.

En remplaçant \mathbf{W} dans l'expression 2.26, le Lagrangien prend la forme d'une fonction à maximiser en fonction des seules variables duales :

$$L(\alpha) = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j Y_i Y_j \langle \mathbf{X}_i, \mathbf{X}_j \rangle \quad (2.29)$$

avec les contraintes $\alpha_i \geq 0$, $\sum_{i=1}^N \alpha_i Y_i = 0$ et $i = 1, \dots, N$

Pour résoudre ce problème, on peut assouplir les contraintes en introduisant les variables d'ajustement ξ_i dans la formulation dans le but de prendre en compte les erreurs de classification. Ainsi, le problème de 2.25 devient alors :

$$\begin{cases} \text{minimiser} & \tau(\mathbf{W}, \xi_i) = \frac{1}{2} \|\mathbf{W}\|^2 + C(\sum_{i=1}^N \xi_i) \\ \text{tel que} & Y_i(\langle \mathbf{W}, \mathbf{X}_i \rangle + b) \geq 1 - \xi_i \quad \forall \xi_i \geq 0 \end{cases} \quad (2.30)$$

Le paramètre supplémentaire C pour "Cost" permet de contrôler le compromis entre l'erreur de classification et la taille de la marge. On qualifie cette approche de classifieur à marge souple car les contraintes sont assouplies sur la marge pour certaines données. Ce principe peut

également être appliqué pour résoudre la formulation duale (équation 2.29). Pour plus de détails sur le fonctionnement des SVM, le lecteur intéressé peut consulter les références [120, 121].

Cependant, les SVM linéaires ont une capacité limitée de classification sur des données non linéairement séparables. En pratique, les données sont rarement séparables par un simple hyperplan.

2.2.2.2.2 SVM non linéaire

Pour contourner ce problème, il est donc préférable de doter les SVM d'un mécanisme permettant de produire des frontières de décision non linéaires. L'idée consiste à projeter les données d'apprentissage sur un espace plus grand nommé espace de Hilbert H dans lequel les données transformées deviennent linéairement séparables.

On définit ainsi une application non linéaire Φ appliquée aux données d'apprentissage définies de \mathbb{R}^n dans l'espace H .

Cette transformation permet d'obtenir un hyperplan séparateur appelé hyperplan optimal généralisé. Cependant, il faut remplacer le produit scalaire $\langle \mathbf{X}_i, \mathbf{X}_j \rangle$ par $\langle \Phi(\mathbf{X}_i), \Phi(\mathbf{X}_j) \rangle$. Ainsi, on peut facilement calculer ce produit à l'aide d'une fonction noyau c définie par :

$$c(\mathbf{X}_i, \mathbf{X}_j) = \langle \Phi(\mathbf{X}_i), \Phi(\mathbf{X}_j) \rangle \quad (2.31)$$

La fonction noyau n'est pas unique mais doit vérifier les conditions de Mercer [122, 123]. En pratique on peut utiliser :

— le noyau linéaire

$$c(\mathbf{X}_i, \mathbf{X}_j) = \mathbf{X}_i \cdot \mathbf{X}_j \quad (2.32)$$

— le noyau polynomial de degré p [124, 125]

$$c(\mathbf{X}_i, \mathbf{X}_j) = (\langle \mathbf{X}_i, \mathbf{X}_j \rangle + 1)^p \quad (2.33)$$

— le noyau gaussien [126, 127]

$$c(\mathbf{X}_i, \mathbf{X}_j) = \exp\left(-\frac{\|\mathbf{X}_i - \mathbf{X}_j\|^2}{2\sigma^2}\right) \quad (2.34)$$

Ainsi, si c'est le cas, les méthodes linéaires décrites précédemment (paragraphe 2.2.2.2.1) peuvent à présent s'appliquer dans le nouvel espace.

Le lecteur intéressé par le fonctionnement du SVM non linéaire peut se référer à la référence [128] pour plus de détails.

Le SVM n'est pas seulement utilisé pour traiter le problème de données à deux classes. Il existe des extensions pour résoudre des problèmes multi classes. Citons notamment la méthode un contre un [129] ou celle un contre tous [122].

Les SVM ont montré leur efficacité et capacité à résoudre des problèmes de traitement d'image [130] ou de diagnostic médical [93]. Ils sont également présent dans la fouille de données (ou data mining), où le SVM est couplé avec des GPU pour traiter un grand ensemble de données [131].

En scattérométrie, les SVM sont également utilisés dans le cadre de la classification de formes de profil. Dans [57, 58], les auteurs combinent la méthode de SVM avec une recherche par bibliothèque pour identifier la déviation du profil géométrique par rapport à un profil trapézoïdal cible ou pour surveiller en temps réel une des dimensions critiques lors des étapes de fabrication [132].

Les SVM présentent l'avantage de fournir une très bonne performance en généralisation. Cependant, ils sont par essence des classifieurs binaires, mais pour la classification multi-classe, il faut utiliser des extensions (un contre un ou un contre tous). Ils nécessitent également beaucoup de temps de calcul pour trouver les meilleurs paramètres qui dépendent du choix de la fonction noyau.

Dans la partie suivante, nous aborderons le R.N comme la méthode de classification directe qui sera utilisée par la suite.

2.3 Les Réseaux de Neurones

Dans cette partie, nous abordons les réseaux de neurones (R.N) en présentant un bref historique ainsi que le principe de base de leur fonctionnement et de leur utilisation. Nous définirons également les différents types de R.N tout en détaillant particulièrement celui que nous utilisons dans le cadre de ce travail : les perceptrons multicouches (PMC).

2.3.1 Des outils mathématiques puissants

Inspiré du principe de fonctionnement du neurone biologique, le R.N artificiel est utilisé comme un outil du traitement de l'information. Il est notamment capable de modéliser un phénomène physique à partir d'un certain nombre d'observations. Cependant, il a besoin d'une phase d'apprentissage préalable pour pouvoir fonctionner correctement. Ainsi, à l'issue de cette phase le R.N est capable de généraliser ce qu'il a appris sur une nouvelle entrée inconnue.

2.3.1.1 Historique des R.N

Les origines du R.N remontent à 1890, où le psychologue W. James présente le concept de mémoire associative. Bien plus tard, en 1943, McCulloch et W. Pitts [133] propose un modèle mathématique appelé neurone formel imitant le principe de fonctionnement du neurone biologique. C'est un neurone binaire qui est capable de réaliser des fonctions logiques simples. En 1949, le physiologiste D. Hebb décrit dans son ouvrage « The organization of behaviour », une règle d'apprentissage connue sous le nom de règle de Hebb [134]. F. Rosenblatt propose en

1957 le perceptron [135] comme le premier modèle de système artificiel capable de reconnaître des formes. En 1960, B. Widrow et T. Hoff proposent le réseau ADALINE (Adaptive Linear Neuron) [136] qui est constitué d'un seul neurone. Vers la fin de l'année 1960, M. Minsky et S. Papert publie un livre intitulé « Perceptron » [137] montrant les limites du perceptron à traiter les problèmes non linéaires ; ce qui a provoqué l'effondrement des publications scientifiques dans ce domaine.

Cependant, certains chercheurs ont décidé de poursuivre leurs travaux comme T. Kohonen et J. Anderson. Ils ont mis au point des réseaux pouvant servir de mémoire associative en développant le concept de cartes auto organisées [107]. En 1982, J. Hopfield [138] publie un article introduisant un réseau dynamique qui évolue jusqu'à atteindre un état de stabilité par un comportement autonome. Enfin, la mise au point de l'algorithme de rétropropagation de l'erreur [139] a permis d'apporter une réponse aux critiques de M. Minsky et S. Papert évoquées plus haut notamment sur l'apprentissage du perceptron multicouche. Cette technique est fortement utilisée encore aujourd'hui et constitue la base de tout apprentissage par correction d'erreur.

C'est ainsi que, plus récemment, « l'apprentissage profond ou Deep Learning » a vu le jour [140] et connaît un essor fulgurant. La particularité de tels réseaux consiste à adjoindre en amont d'un PMC différentes couches de traitement automatique de l'information fournie en entrée. Ces nouvelles architectures sont devenues populaires suite au concours ImagNets de 2012. En effet, ce dernier met en compétition différentes méthodes automatiques de classification d'images selon 100 catégories différentes (L. Fei Fei et al. [141]). Cette année là, le réseau profond AlexNet développé par Alex Krizhevsky [142] remporte haut la main la compétition. Cet essor des RNs profonds est rendu possible grâce à l'augmentation de la puissance de calcul des ordinateurs et de l'accessibilité à de gigantesques bases de données.

2.3.1.2 Applications

Les RNs offrent des solutions dans de nombreux domaines tels que l'informatique, l'électronique, les sciences cognitives, etc. Ils servent à toutes sortes d'applications pour traiter les informations de plus en plus complexes. Il est difficile d'en faire une liste exhaustive ici tant la popularité de tels outils a explosée ces dernières années. Ils sont notamment utilisés dans le domaine de la reconnaissance de formes, des objets ou de visage. Le laboratoire AT&T Bell est notamment le premier à développer un système de reconnaissance des codes fournis par les services postaux américains [143]. Les RNs sont aussi utilisés pour la reconnaissance des chiffres manuscrits [144].

Dans le domaine de la sécurité des réseaux informatiques, le R.N est utilisé pour détecter des menaces d'intrusion [145, 146]. On le trouve également en marketing pour le traitement d'énormes bases de données [102]. Ils sont également employés dans le système d'assainissement pour surveiller le système de collecte des eaux de pluie [147].

A noter également que certains secteurs de l'ingénierie font appel aux RNs pour résoudre certaines problématiques, comme par exemple la robotique, où ils sont utilisés pour le système de

guidage automatique d'un véhicule autonome [148]. On les emploie également en aéronautique pour améliorer la trajectoire d'urgence d'avion en cas d'une panne de moteur [149]. Dans le secteur de la bio-ingénierie, ils sont utilisés pour la prédiction de propriétés pharmacologiques de molécules [150]. En instrumentation, on les emploie pour la calibration des signaux [151] ou pour l'amélioration de la précision de mesure d'un capteur ultrasonique [152].

Ils sont également présents dans le secteur du transport pour prédire la vitesse du trafic [153, 154] ou la détection de voies routières [155, 156].

Ainsi, les RNs ont montré leur capacité à résoudre des problèmes complexes dans plusieurs domaines d'applications. Il est facile de se rendre compte, à travers les différentes publications actuelles, la place prise par ces derniers dans le traitement de données et ce, dans tous les domaines (scientifiques ou non).

Plus spécifiquement au cadre de cette thèse, les RNs ont également suscité l'intérêt de certains chercheurs pour la métrologie optique et plus particulièrement la scattérométrie [43, 44] ou la caractérisation ellipsométrique [157]. Des nombreuses études ont également montré l'efficacité du R.N dans la scattérométrie [51, 55] démontrant notamment leur étonnante résistance au bruit de mesure [46]. Les RNs sont d'ailleurs utilisés au Laboratoire Hubert Curien dans ce cadre depuis plus de 20 ans [19, 37, 38, 81]. Dans [79], l'auteur montre la possibilité de détecter des rugosités de bords d'un motif. Ils sont employés aussi pour détecter toute déviation par rapport à une forme nominale cible [63]. Le R.N est également un atout majeur dans le concept de scattérométrie hybride [158]. Enfin, certaines études issues du laboratoire [61, 62] démontrent la capacité d'un réseau neuronal à détecter une couche résiduelle considérée comme un défaut en lithographie à partir d'une simple signature ellipsométrique.

2.3.2 Description et fonctionnement général

2.3.2.1 Définition du neurone formel

Créé par McCulloch et W. Pitts pour imiter le fonctionnement du neurone biologique, le neurone formel (figure 2.6) est un système mathématique qui a une capacité d'apprentissage de l'environnement extérieur. Il calcule simplement une réponse à partir de n entrées $x_i (i = 1, 2, \dots, n)$ qui lui sont fournies par l'intermédiaire de connexions pondérées par des poids synaptiques w .

Le fonctionnement d'un neurone j se décompose en deux étapes :

La première phase est une étape de prétraitement des données reçues par le calcul du potentiel v_j :

$$v_j = b_j + \sum_{i=1}^n w_{j,i} x_i \quad (2.35)$$

où $w_{j,i}$ le poids synaptique de la connexion reliant le neurone j à l'entrée particulière i , b_j un paramètre interne du neurone appelé biais.

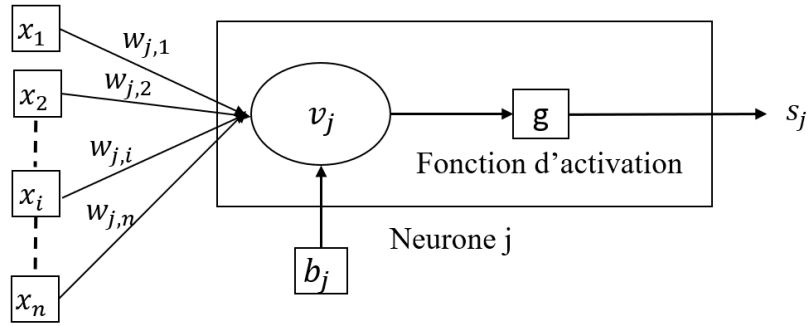


FIGURE 2.6 – Représentation d'un neurone artificiel.

En considérant le biais b_j comme le poids d'une entrée constante x_0 égale à 1, la formule précédente devient simplement :

$$v_j = \sum_{i=0}^n w_{j,i} x_i \quad (2.36)$$

Dans la suite nous utiliserons exclusivement comme fonction d'entrée totale cette somme pondérée des entrées. D'autres fonctions existent mais nous ne les avons pas utilisé dans nos architectures de R.N. Dans la seconde phase, la valeur de la sortie s_j du neurone j est calculée en fonction du potentiel v_j :

$$s_j = g(v_j) = g\left(\sum_{i=0}^n w_{j,i} x_i\right) \quad (2.37)$$

La fonction g est appelée fonction de transfert ou fonction d'activation. Il existe différentes fonction d'activation (figure 2.7) suivant le type d'applications visées. Les fonctions les plus utilisées sont :

— La fonction seuil (figure 2.7(a)) :

$$g(x) = 0 \quad \text{si } x < 0 \quad (2.38)$$

$$g(x) = 1 \quad \text{si } x \geq 0 \quad (2.39)$$

— La fonction identité (linéaire) (figure 2.7(b)) :

$$g(x) = x \quad \forall x \quad (2.40)$$

— La fonction sigmoïde (figure 2.7(c)) dont les formes les plus utilisées sont la fonction logistique et la fonction hyperbolique :

$$g(x) = \frac{1}{1 + \exp^{-Kx}} \quad (2.41)$$

$$g(x) = \frac{\exp^{Kx} + 1}{\exp^{Kx} - 1} \quad (2.42)$$

Le paramètre K permet de modérer la pente à l'origine. Il est généralement pris égal à 1.

- La fonction gaussienne (figure 2.7(d)) d'écart type unité :

$$g(x) = \exp^{-x^2} \quad (2.43)$$

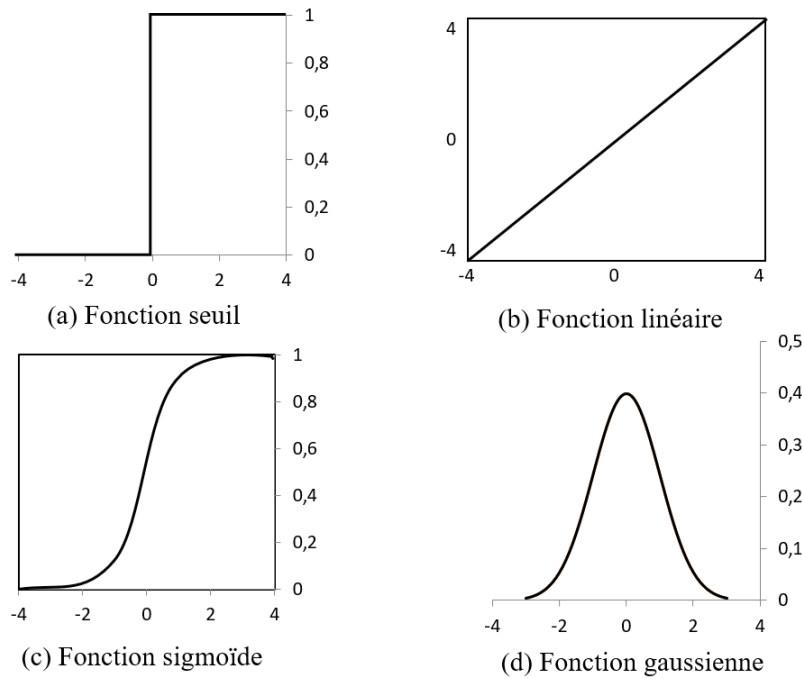


FIGURE 2.7 – Exemple de fonctions d'activations g utilisées pour calculer la sortie d'un neurone.

Le neurone formel tel qu'il a été décrit dans ce paragraphe constitue l'élément de base pour la construction de tout R.N.

2.3.2.2 Architecture des R.N

Le R.N est un ensemble de plusieurs neurones interconnectés entre eux. La mise en réseau de ces neurones permet d'échanger des informations par l'intermédiaire des connexions qui les relient. Suivant le modèle de connexion, on distingue différents types de R.N : les réseaux bouclés (ou encore appelés récurrents) et les réseaux non bouclés.

Les premiers [159] sont principalement utilisés pour la modélisation de processus dynamiques non linéaires. Le calcul de la sortie d'un neurone à un instant donné peut dépendre de la sortie d'autres neurones à des instants différents. Ce type de réseau n'entre pas dans le cadre de notre étude et ne sera pas détaillé par la suite. Dans le second, les informations circulent de l'entrée vers la sortie sans retour en arrière possible. Il n'y a pas de notion de temps, c'est pour cela qu'on les appelle également des réseaux statiques.

Les plus connus sont les réseaux multicouches : Les neurones sont disposés en couches verticales successives. Il n'y a pas des connexions entre les neurones d'une même couche (figure 2.8). Chaque neurone d'une couche est relié, en entrée, aux sorties des neurones des couches

précédentes et, en sortie, aux entrées de chaque neurone de la couche suivante. On définit ainsi une couche d'entrée qui reçoit l'information, une ou plusieurs couches cachées qui réalisent des calculs intermédiaires et une couche de sortie qui effectue le dernier calcul du réseau fournissant alors le résultat final. Nous pouvons ainsi définir un sens de parcours de l'information (activation) au sein d'un réseau depuis les entrées vers les sorties. Ce type de réseau est utilisé pour des tâches de prédiction ou de classification. C'est l'architecture la plus utilisée particulièrement en scattérométrie. Nous détaillerons son principe de fonctionnement par la suite.

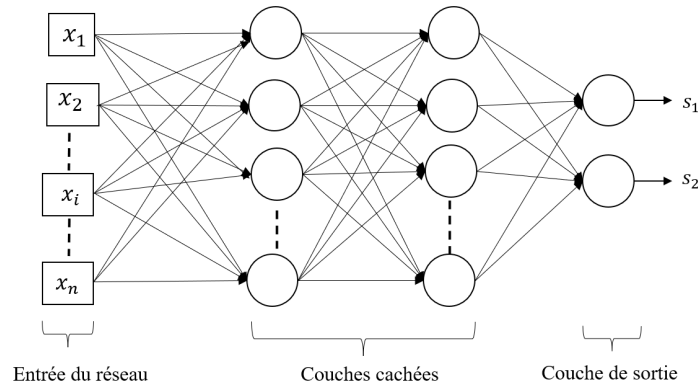


FIGURE 2.8 – Exemple d'architecture de réseau multicouche à deux couches cachées.

Les réseaux à connexions locales constituent un cas particulier d'une structure multicouche. En effet, certaines connexions sont désactivées à l'image de la rétine (figure 2.9). Chaque neurone est relié à un nombre fixe de neurones voisins. Les connexions sont donc moins nombreuses que dans le cas du réseau multicouche. Ce type d'architecture est utilisé en traitement d'image, dans des applications de discrimination [108].

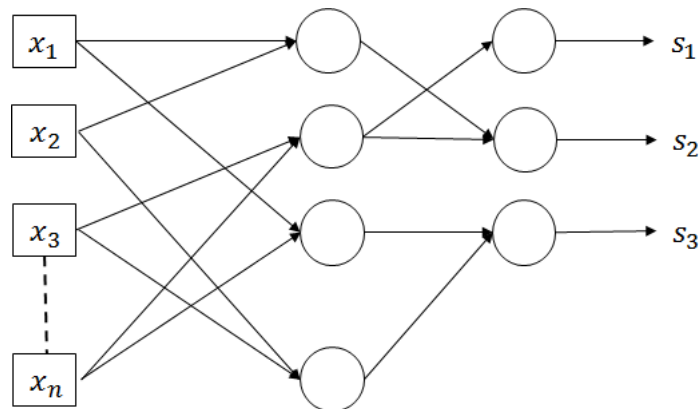


FIGURE 2.9 – Exemple d'architecture de réseau à connexions locales.

2.3.2.3 Apprentissage des R.N

C'est la phase la plus importante pour le fonctionnement du R.N. C'est une étape pendant laquelle le comportement du R.N est modifié d'une manière itérative sur la base d'un certain nombre d'échantillons composant le corpus d'apprentissage. Pendant cette phase d'apprentissage, chaque poids synaptique $w_{j,i}$ est corrigé à l'itération $m + 1$ de la façon suivante :

$$w_{j,i}^{m+1} = w_{j,i}^m + \Delta w_{j,i} \quad (2.44)$$

avec $\Delta w_{j,i}$ la modification apportée à l'itération m

L'apprentissage comprend deux grandes branches : l'apprentissage supervisé et l'apprentissage non supervisé.

Dans le cas supervisé, le corpus d'apprentissage est composé de couples entrée/sortie (\mathbf{X}/\mathbf{Y}). La finalité est de déterminer la sortie \mathbf{Y} à partir d'une entrée \mathbf{X} sur la base d'exemples fournis en nombre limité. Pour cela, à chaque fois qu'un vecteur \mathbf{X} est fourni en entrée du R.N, sa sortie \mathbf{S} résultant d'une série de calcul interne est calculée (figure 2.10). Connaissant la valeur désirée \mathbf{Y} correspondante à cette entrée \mathbf{X} , il est alors possible de calculer l'erreur \mathbf{E} commise par le R.N par $\mathbf{E} = \mathbf{S} - \mathbf{Y}$. L'apprentissage supervisé est donc ramené à un problème d'optimisation paramétrique dans lequel on cherche les poids qui correspondent à une erreur \mathbf{E} minimale calculée sur l'ensemble des données du corpus d'apprentissage. La plupart du temps on utilise les méthodes paramétriques « classiques » (descente de gradient, Gauss Newton type BFGS ou méthode de Levenberg-Marquardt) (paragraphe 1.3.3.2.1). Le Perceptron Multicouche fonctionne avec ce type d'apprentissage.

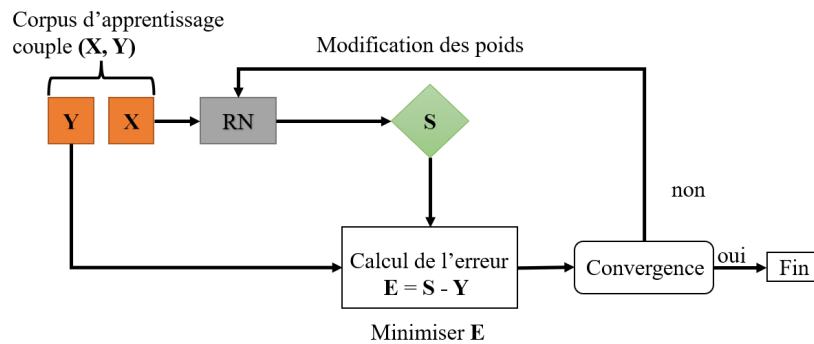


FIGURE 2.10 – Schéma expliquant le fonctionnement de l'apprentissage supervisé.

En apprentissage non supervisé, il n'y a pas d'exemples de sorties connues à l'avance. Seules les entrées \mathbf{X} composant l'ensemble des données, généralement de grande taille, sont fournies au R.N. Ce dernier cherche à les regrouper selon certains critères de ressemblance, *a priori* inconnus, à partir de lois internes fixées à l'avance. La Carte Auto-Organisatrice de Kohonen utilise par exemple ce type d'apprentissage.

Lors de chaque processus d'apprentissage, on distingue également deux modes de fonctionnement : le mode adaptatif où les poids synaptiques sont modifiés après chaque passage d'un couple (\mathbf{X}/\mathbf{Y}) ou le mode non adaptatif (en anglais batch training) où la modification des poids est réalisée après le passage de l'ensemble des données de l'apprentissage. Ce type d'apprentissage est plus rapide et également adapté à une grande quantité des données. C'est celui que nous utiliserons par la suite.

La modification du poids synaptique $\Delta w_{j,i}$ à chaque itération est le cœur de l'apprentissage. Il y a plusieurs règles qui sont spécifiques à un type d'architecture et d'apprentissage donné (supervisé ou non supervisé) : la règle de Hebb [134, 160], la règle de Widrow Hoff [136] ou

la règle de Grossberg [161]. Par exemple la règle de Widrow Hoff est utilisé en apprentissage supervisé, celle de Grossberg en apprentissage non supervisé et la règle de Hebb est employé dans le cas des réseaux non bouclés (modèle de Hopfield). Elles sont les plus connues et les plus simples à utiliser.

Dans les 2 paragraphes suivants nous détaillerons le processus d'apprentissage dans le cas de 2 RNs au fonctionnement distinct.

2.3.3 Les réseaux de Kohonen

Le réseau de Kohonen est un réseau qui fonctionne avec un apprentissage qualifié de compétitif. Il permet de construire une représentation visualisable (1, 2, ou 3 dimensions) des entrées par une sorte de projection non-linéaire de l'espace des entrées qui soit telle que : La topologie initiale soit conservée des données proches (dans l'espace d'entrée) et qu'elles aient des représentations proches dans l'espace de sortie et vont donc être classées dans une même classe (ou classes voisines).

Les réseaux de Kohonen sont utilisés dans beaucoup d'applications. Citons notamment le domaine de la sécurité informatique où ils permettent de détecter les intrusions [146, 162]. Ils sont également présents dans l'analyse de données pour la classification des données symboliques [108]. Dans [163, 164], ils ont montré leurs performances pour classer les signaux encéphalographiques d'un être humain afin de déterminer son état de somnolence et de vigilance.

Parmi les réseaux de Kohonen, on distingue la Carte Auto-Organisatrice (SOM) et le Learning Vector Quantization (LVQ).

2.3.3.1 Les Cartes Auto-Organisatrices

Les Cartes Auto-Organisatrices (en anglais Self Organizing Map ou SOM) sont développées par T. Kohonen vers les années 80 [107]. Ces réseaux fonctionnent avec un apprentissage non supervisé. Ils sont généralement développés pour réaliser une tâche d'analyse de données ou clustering. La SOM est un réseau qui apprend à modéliser l'espace d'entrée sur une carte à une (deux ou trois) dimension(s) définie par un ensemble de neurones. Chaque neurone est associé au vecteur prototype dans l'espace d'entrée. Pendant l'apprentissage, chaque neurone de la carte se spécialise pour représenter une entrée donnée.

2.3.3.1.1 Architecture d'une SOM

La structure de la SOM est constituée de deux couches (figure 2.11). La couche d'entrée est composée de n éléments (x_1, x_2, \dots, x_n) représentant le vecteur d'entrée \mathbf{X} . La couche de sortie ou de compétition est composée de S neurones répartis sur une carte C (1, 2 ou 3 dimensions).

L'emplacement sur C d'un neurone j est déterminé par un vecteur position r_j de coordonnées a et b (figure 2.11) tel que :

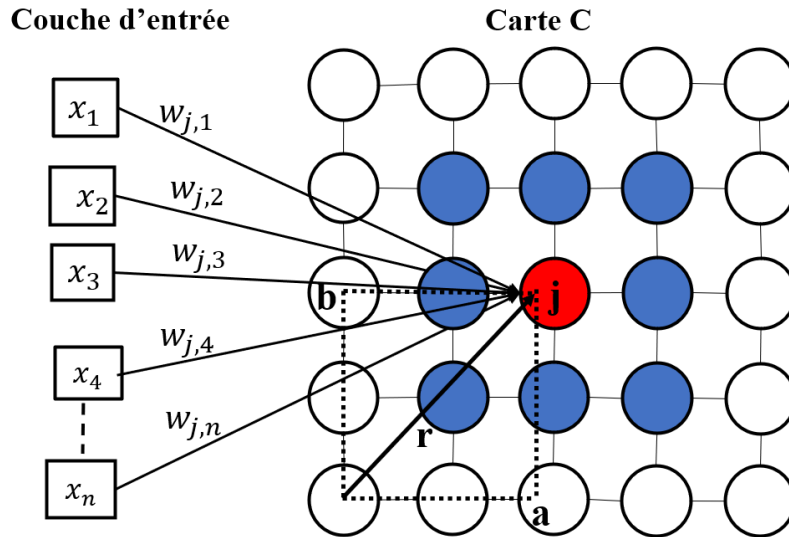


FIGURE 2.11 – Architecture d'une SOM en 2D composée de n entrées et de $5 * 5$ neurones de sortie répartis suivant une topologie rectangulaire.

$$r_j = \begin{pmatrix} a \\ b \end{pmatrix} \quad (2.45)$$

où a est l'abscisse et b est l'ordonnée dans le repère de la carte.

On définit une distance discrète entre deux neurones i et j notée δ_{ij} comme étant la plus courte distance sur la carte entre ces deux neurones. Le voisinage d'ordre p du neurone j noté $V_j(p)$ est donc défini comme l'ensemble des neurones de la carte située à une distance δ de ce dernier inférieure ou égale à p .

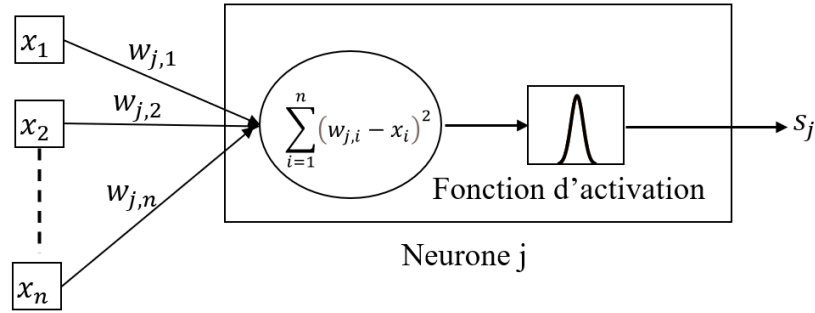
$$V_j(p) = \{r \in C, \delta_{ij} \leq p\} \quad (2.46)$$

Par exemple le voisinage d'ordre 1 du neurone j est représenté sur la figure 2.11.

L'activation d'un neurone de la carte C ne dépend pas seulement du vecteur d'entrée mais également de son voisinage. Ainsi le neurone gagnant sélectionné est celui qui a le vecteur prototype le plus proche du vecteur d'entrée ; d'où le nom des neurones « distances » (figure 2.12).

Le calcul du potentiel v_j d'un neurone j est donc lié à la proximité de chaque \mathbf{X} avec les poids synaptiques de ses connexions. La fonction d'activation qui suit est généralement de type gaussienne (équation 2.47).

$$v_j = \sum_{i=1}^n (w_{j,i} - x_i)^2 \quad (2.47)$$

FIGURE 2.12 – Représentation d'un neurone j utilisé dans un réseau de Kohonen.

2.3.3.1.2 Apprentissage d'une SOM

Le principe de l'apprentissage d'une SOM consiste à projeter les données d'entrées sur la carte C . Ainsi, à chaque connexion d'une entrée \mathbf{X} vers un neurone j est associé un vecteur prototype \mathbf{W}_j composé des poids $w_{j,1}, w_{j,2}, \dots, w_{j,n}$. Le neurone sélectionné est celui qui a le vecteur prototype le plus proche du vecteur d'entrée. Toutefois, l'activation d'un neurone de la carte C ne dépend pas seulement du vecteur d'entrée mais également de son voisinage. En effet, chaque neurone de C entre en compétition avec ses voisinages pour représenter une entrée donnée. Le neurone gagnant est noté G . On parle de « winner takes all ».

$$G = \arg \min_j \|\mathbf{X} - \mathbf{W}_j\|^2 \quad \forall j \quad (2.48)$$

Le poids \mathbf{W}_G associé au neurone gagnant et ceux des autres neurones de la carte \mathbf{W}_j sont ensuite ajustés selon la règle d'apprentissage suivante :

$$\mathbf{W}_j^{m+1} = \mathbf{W}_j^m + \alpha(m)h(\delta_{Gj})[\mathbf{X} - \mathbf{W}_j^m] \quad (2.49)$$

Où $\alpha(m)$ représente le taux d'apprentissage et $h(\delta_{Gj})$ une fonction de voisinage permettant de pondérer la variation à appliquer au poids d'un neurone suivant sa proximité au neurone vainqueur G .

Le taux d'apprentissage $\alpha(m)$ est une fonction monotone décroissante comprise entre $[0,1]$. On utilise souvent une décroissance linéaire de type :

$$\alpha(m) = \alpha_0 \left(1 - \frac{m}{T}\right) \quad (2.50)$$

où α_0 représente la valeur initiale et T le nombre maximal d'itération fixé pour l'apprentissage

La fonction de voisinage du neurone gagnant est une fonction continue, généralement de forme gaussienne :

$$h(\delta_{Gj}) = \exp\left(-\frac{\delta_{Gj}^2}{\sigma^2(m)}\right) \quad (2.51)$$

où $\sigma(m)$ est un paramètre qui évolue au cours des itérations suivant :

$$\sigma(m) = \sigma_0 \exp\left(-\frac{m}{T}\right) \quad (2.52)$$

Le choix des valeurs initiales de α_0 et σ_0 pourraient avoir une influence sur les résultats obtenus. Généralement, l'apprentissage d'une SOM se passe en deux étapes pour les cartes C à grandes dimensions. Dans la première phase, α et σ sont initialisés avec de grandes valeurs, et décroissent rapidement afin d'organiser grossièrement les vecteurs prototypes \mathbf{W} . La seconde phase appelée également phase de convergence consiste à ajuster les vecteurs prototypes vers leurs positions finales à partir des résultats de la première phase. Les paramètres α et σ sont alors initialisés avec de petites valeurs.

Toutefois, il existe une variante plus rapide de l'apprentissage d'une SOM. C'est l'apprentissage par lot (ou batch training en anglais) qui ne nécessite pas l'utilisation du taux d'apprentissage α . Par contre, l'ensemble de données d'entrées est présenté à la carte. Ensuite, les nouveaux vecteurs poids sont calculés comme suit :

$$\mathbf{W}_j = \frac{\sum_{i=1}^N h(\delta_{G(x_i)j}) \mathbf{X}_i}{\sum_{i=1}^N h(\delta_{G(x_i)j})} \quad (2.53)$$

La fonction coût à optimiser est alors :

$$C(\mathbf{W}) = \frac{1}{2N} \sum_i^N \sum_j^S h(\delta_{G(x_i)j}) \|\mathbf{X}_i - \mathbf{W}_j\|^2 \quad (2.54)$$

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

A titre d'exemple, une étude effectuée dans [62] montre la capacité d'un SOM à détecter une couche résiduelle de résine considérée comme un défaut à partir d'une signature optique fournie en entrée.

Ainsi, l'apprentissage des SOM conduit à une représentation topologique des vecteurs d'entrées dans l'espace de sortie décrit par les neurones de la carte. La SOM se comporte donc comme un quantificateur vectoriel. L'apprentissage étant non supervisé, ce quantificateur ne permet pas d'affecter directement une classe à une entrée. Ainsi, c'est à l'expérimentateur d'affecter une classe suivant le résultat de la projection de la SOM. Si on veut une classification automatique, il faut avoir recours au Learning Vector Quantization (LVQ).

2.3.3.2 Learning Vector Quantization

Développé également par T. Kohonen, le Learning Vector Quantization (LVQ) est un R.N compétitif qui fonctionne avec un apprentissage supervisé. Ce R.N est spécifiquement destiné à la classification des données. Sa structure est composée de trois couches (figure 2.13) : une couche d'entrée qui reçoit les informations par l'intermédiaire des vecteurs d'entrées, une couche cachée ou de compétition qui fonctionne comme une SOM suivant le principe de « winner takes

all » et une couche de sortie qui donne le résultat final sous forme binaire. Chaque neurone de la couche de compétition noté N_c est relié à toutes les entrées ainsi qu'à tous les neurones de la couche de sortie notée N_s .

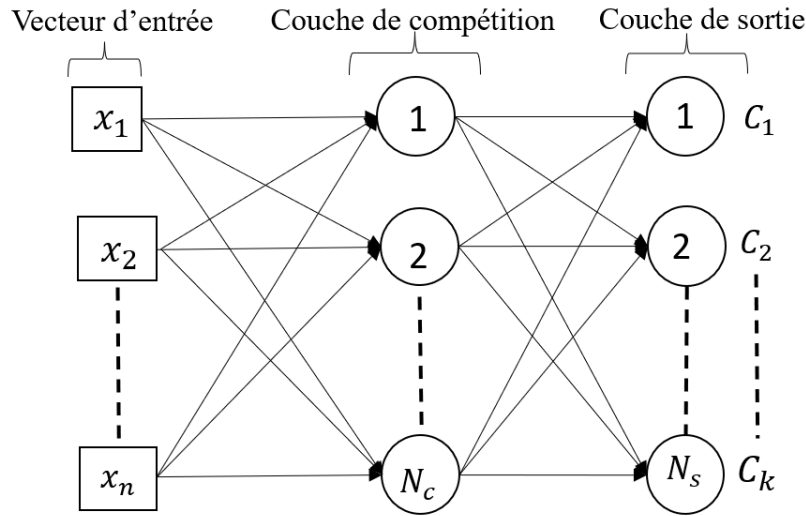


FIGURE 2.13 – Architecture d'un LVQ composée de n entrées, d'une couche de compétition de N_c neurones et d'une couche de sortie de N_s neurones correspondant chacune à une classe connue.

Les données d'entraînement sont composées de couples \mathbf{X}/\mathbf{Y} où \mathbf{Y} la classe correspondante à chaque entrée \mathbf{X} . L'entraînement consiste à renforcer, à chaque itération, les vecteurs prototypes en employant la règle de Kohonen (équation 2.49). Par contre, pour une entrée \mathbf{X} donnée, seul le poids du neurone élu est modifié. Si la sortie du neurone gagnant correspond à celle de la classe attendue, les poids de ce neurone sont rapprochés de l'entrée sinon les deux vecteurs sont éloignés. Le vecteur poids du neurone gagnant est adapté comme suit :

$$\mathbf{W}_G^{m+1} = \mathbf{W}_G^m + \alpha(m)[\mathbf{X} - \mathbf{W}_G^m] \quad \text{si } \mathbf{W}_G = \mathbf{Y} \quad (2.55)$$

$$\mathbf{W}_G^{m+1} = \mathbf{W}_G^m - \alpha(m)[\mathbf{X} - \mathbf{W}_G^m] \quad \text{si } \mathbf{W}_G \neq \mathbf{Y} \quad (2.56)$$

Dans [81], une étude a été réalisée sur le LVQ pour identifier la forme géométrique à partir de la signature optique. Dans le cadre de nos travaux, nous n'allons pas utiliser ces types de R.Ns, le lecteur intéressé pourra se référer dans ces études [165, 166].

2.3.4 Le Perceptron Multicouche

Le Perceptron Multicouche (PMC) est un réseau non bouclé constitué de neurones disposés en couche. Il est utilisé pour des problèmes d'approximation, de prédiction et de classification.

Le PMC a également un large éventail d'applications de classification et de régression dans de nombreux domaines comme par exemple l'énergie, où on l'utilise pour la prédiction du processus de gazéification de la biomasse [167]. Ils sont également présents en robotique, plus particulièrement dans les tâches d'analyse des commandes vocales [168]. En imagerie biomédical, les PMC sont employés pour la détection du cancer de la vessie [169].

2.3.4.1 Architecture du PMC

Le PMC est un type particulier d'un réseau à couche (paragraphe 2.3.2.2). En effet, les fonctions de transfert d'une même couche sont identiques. Les neurones de la couche cachée sont à fonction d'activation sigmoïdale. Les neurones de la couche de sortie ont une fonction de transfert linéaire dans le cas d'une utilisation en tant que prédicteur et non-linéaire dans le cas d'un classifieur. Nous noterons N_c le nombre de neurones de la couche cachée et N_s celui de la couche de sortie. La figure 2.14 montre l'exemple d'une structure d'un PMC composé de n entrées, possédant une seule couche cachée.

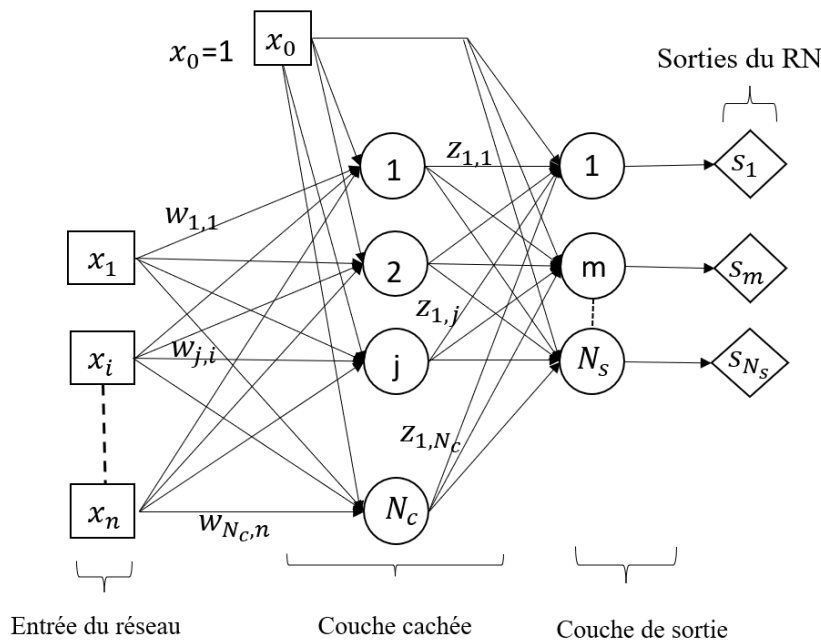


FIGURE 2.14 – Structure d'un réseau de neurones de type perceptron multicouche à une couche cachée.

Le PMC possède la propriété d'approximateur universel [170] et parcimonieux [171]. La propriété d'approximateur est énoncée de la manière suivante : *toute fonction bornée et suffisamment régulière peut être approchée avec une précision arbitraire, dans un domaine fini de l'espace de ses variables, par un R.N comportant une couche de neurones cachés en nombre fini, possédant toute la même fonction d'activation, et un neurone de sortie linéaire.* Cette propriété n'est pas spécifique aux R.Ns. Il existe d'autres fonctions possédant cette propriété (ondelettes, fonctions splines, ...).

L'approximation est également parcimonieuse : *à précision égale, les R.Ns 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.*

Ces propriétés décrites dans le cadre de l'approximateur se généralise bien entendu au classifieur universel.

Dans le cadre spécifique de la classification, on introduit un codage pour représenter les

différentes classes. Dans un problème de classification à k classes, on fait correspondre à chaque entrée \mathbf{X} un vecteur \mathbf{Y} pour lequel les composantes y_k ne peuvent prendre que 2 valeurs a ou b . Le vecteur \mathbf{Y} représentant la classe C_k , seule la k^{ime} composante prend la valeur a et les autres composantes sont égales à b . Ainsi, la k^{ime} sortie du PMC s_k représente :

$$\begin{aligned} s_k &= E(\mathbf{Y}/\mathbf{X}) \\ &= a \times p(C_k/\mathbf{X}) + b(1 - p(C_k/\mathbf{X})) \end{aligned} \quad (2.57)$$

Pour un codage particulier où $a = 1$ et $b = 0$, l'état s_k du k^{ime} neurone de la couche de sortie est donc une approximation de la probabilité *a posteriori*.

$$s_k = p(C_k/\mathbf{X}) \quad (2.58)$$

Les différentes sorties sont nécessairement comprises entre $[0, 1]$, cela montre que le PMC classifieur permet d'estimer la fonction discriminante sans passer par la règle de Bayes (paragraphe 2.2.1).

L'architecture du PMC est donc parfaitement définie en fixant le nombre de couches et le nombre de neurones N_c par couche cachée. Dans le cas de notre étude, les entrées correspondent aux mesures des signatures optiques I_s et I_c à différentes longueurs d'onde et les sorties aux classes définies ou aux paramètres géométriques modélisant la forme de la structure suivant le mode d'utilisation du PMC.

2.3.4.2 Principe de fonctionnement

Considérons un PMC utilisé en mode régresseur composé d'une seule couche cachée de N_c neurones, possède n entrées et une couche de N_s sorties. Nous désignons par f la fonction d'activation de la couche cachée et g celle de la couche de sortie. La m^{ime} sortie du PMC s'écrit alors :

$$s_m(\mathbf{X}, \mathbf{W}, \mathbf{Z}) = g \left(\sum_{j=0}^{N_c} z_{m,j} \left[f \left(\sum_{i=0}^n w_{j,i} x_i \right) \right] \right) \quad (2.59)$$

où \mathbf{W} est la matrice regroupant les poids des connexions entre la couche cachée et les entrées et \mathbf{Z} celle regroupant les poids entre la couche de sortie et la couche cachée.

En général, l'étape d'apprentissage est précédée d'une phase de pré-traitement qui consiste à normaliser les données du corpus (paragraphe 2.1.3). En pratique, il est recommandé de pré-traiter l'ensemble des données afin que ces dernières soient centrées et réduites. Les sorties sont également normalisées de la même façon.

2.3.4.3 Apprentissage du PMC

Le PMC fonctionne avec un apprentissage supervisé, et consiste à adapter les poids synaptiques de manière itérative reliant les entrées \mathbf{X} aux valeurs cibles \mathbf{Y} . Durant cette phase, le réseau va apprendre la relation liant les entrées aux valeurs cibles. L'apprentissage supervisé se base donc sur un algorithme de correction d'erreur. D'une manière générale, on cherche à minimiser (paragraphe 2.3.2.3) la fonction coût représentant l'erreur quadratique moyenne entre la sortie du réseau \mathbf{S} et la valeur cible \mathbf{Y} sur l'ensemble de données d'apprentissage par :

$$C(\mathbf{W}, \mathbf{Z}) = \sum_{q=1}^N \left[\sum_{m=1}^{N_s} (y_m^q - s_m^q(\mathbf{X}, \mathbf{W}, \mathbf{Z}))^2 \right] \quad (2.60)$$

où s_m est la $m^{\text{ème}}$ sortie du PMC, y_m la $m^{\text{ème}}$ valeur du vecteur \mathbf{Y} et N le nombre d'exemples de couple d'apprentissage.

On peut également décomposer cette erreur en somme de coûts partiels C^q qui correspondent chacun à l'erreur quadratique sur chaque entrée \mathbf{X}^q . Nous avons donc :

$$C(\mathbf{W}, \mathbf{Z}) = \sum_{q=1}^N C^q(\mathbf{W}, \mathbf{Z}) \quad (2.61)$$

$$\text{avec } C^q = \sum_{m=1}^{N_s} (y_m^q - s_m^q)^2$$

Les potentiels des neurones cachés j et de sortie m notés respectivement u_j et v_m possédant chacun les fonctions d'activation f et g :

$$u_j^q = \sum_{i=0}^N w_{j,i} x_i^q \quad \text{et} \quad v_m^q = \sum_{j=0}^{N_c} z_{m,j} f(u_j^q) \quad (2.62)$$

Pour résoudre le problème de minimisation de la fonction coût, on utilise les méthodes d'optimisation classiques présentées au chapitre 1. Ainsi, ces méthodes nécessitent nativement la connaissance des dérivées premières et secondes de la fonction coût par rapport aux paramètres qui sont ici les poids du R.N. Si cela ne pose pas de problème pour les poids \mathbf{Z} des neurones de sorties, ce n'est pas le cas pour les poids \mathbf{W} des neurones cachés. En effet, pour un couple d'observations donné la sortie désirée des neurones cachés n'est, a priori, pas connue.

Un algorithme conçu par Rumelhart connu sous le nom de retro-propagation (paragraphe 2.3.1.2) du gradient [139] permet de résoudre ce problème en estimant de manière récursive la dérivée partielle $\frac{\partial C}{\partial w_{j,i}}$.

L'influence des paramètres $z_{m,j}$ sur la fonction coût peut être calculée par :

$$\begin{aligned} \left(\frac{\partial C^q}{\partial z_{m,j}} \right) &= \left(\frac{\partial C^q}{\partial v_m^q} \right) \left(\frac{\partial v_m^q}{\partial z_{m,j}} \right) \\ &= \delta_m^q f(u_j^q) \end{aligned} \quad (2.63)$$

où δ_m^q s'exprime par :

$$\begin{aligned}\delta_m^q &= \left(\frac{\partial C^q}{\partial v_m^q} \right) \\ &= -2(y_m^q - s_m^q)\end{aligned}\quad (2.64)$$

Cela donne :

$$\left(\frac{\partial C^q}{\partial z_{m,j}} \right) = -2f(u_j^q)(y_m^q - s_m^q)^2 \quad (2.65)$$

L'influence des poids $w_{j,i}$ s'exprime alors de la façon suivante :

$$\begin{aligned}\left(\frac{\partial C^q}{\partial w_{j,i}} \right) &= \left(\frac{\partial C^q}{\partial u_j^q} \right) \left(\frac{\partial u_j^q}{\partial w_{j,i}} \right) \\ &= \left[\sum_{m=1}^{N_s} \left(\frac{\partial C^q}{\partial v_m^q} \right) \left(\frac{\partial v_m^q}{\partial u_j^q} \right) \right] x_i^q \\ &= \left[\sum_{m=1}^{N_s} \delta_m^q \left(\frac{\partial v_m^q}{\partial u_j^q} \right) \right] x_i^q\end{aligned}\quad (2.66)$$

or

$$\left(\frac{\partial v_m^q}{\partial u_j^q} \right) = z_{m,j} \left(\frac{\partial f}{\partial u_j^q} \right) \quad (2.67)$$

Ce qui donnera :

$$\begin{aligned}\left(\frac{\partial C^q}{\partial w_{j,i}} \right) &= \left[\sum_{m=1}^{N_s} \delta_m^q z_{m,j} \left(\frac{\partial f}{\partial u_j^q} \right) \right] x_i^q \\ &= x_i^q \left(\frac{\partial f}{\partial u_j^q} \right) \sum_{m=1}^{N_s} \delta_m^q z_{m,j}\end{aligned}\quad (2.68)$$

Ces formules permettent de mettre en œuvre la méthode d'optimisation pour la phase d'apprentissage. La méthode d'optimisation de Levenberg-Marquardt [68, 69] sera privilégiée dans les études qui vont suivre pour les avantages qu'elle présente pour cette phase d'apprentissage (paragraphe 1.3.3.2.1).

Une fois l'apprentissage du PMC terminé, l'étape suivante consiste à évaluer ses performances. Pour cela, on utilise des données qui n'ont pas servi lors de la phase d'entraînement pour éviter tout biais dans l'estimation des performances. Ainsi, l'erreur propre à chaque sortie m du PMC est définie par :

$$E_{test}^m = \sqrt{\frac{1}{N_{test}} \sum_{q=1}^{N_{test}} [(y_m^q - s_m^q(\mathbf{X}, \mathbf{W}, \mathbf{Z}))^2]} \quad (2.69)$$

où N_{test} est le nombre de couples utilisés

On peut également estimer la performance globale du PMC en calculant l'erreur quadratique moyenne sur l'ensemble de N_s sorties :

$$E_{test} = \sqrt{\frac{1}{N_{test}} \sum_{q=1}^{N_{test}} \left[\frac{1}{N_s} \sum_{m=1}^{N_s} (y_m^q - s_m^q(\mathbf{X}, \mathbf{W}, \mathbf{Z}))^2 \right]} \quad (2.70)$$

Toutefois, cette méthode d'optimisation utilisée seule ne peut pas garantir systématiquement un bon fonctionnement de la phase d'apprentissage. En effet, la méthode peut aboutir à un surapprentissage désignant un entraînement conduisant à un R.N incapable de généraliser sur de nouvelles entrées. Plusieurs méthodes permettent d'atteindre une généralisation optimale à l'issue de la phase d'apprentissage. Citons la méthode de validation croisée [172], la méthode de LOCL (Local Overfitting Control via Laverages) [173], la suppression systématique du modèle dont la matrice Jacobienne n'est pas de rang plein [174] ou la méthode d'arrêt prématuré [175]. C'est cette dernière que nous allons détailler par la suite et qui sera utilisée dans nos études. Cette technique se base sur l'évaluation des performances du R.N en cours d'apprentissage.

La méthode d'arrêt prématuré (ou *early stopping* en anglais) consiste à arrêter l'entraînement prématurément pour assurer une bonne généralisation. Pour cela, le corpus d'entraînement est subdivisé en trois lots :

- **Lot d'apprentissage** : Ce lot effectue l'entraînement proprement dit du R.N. En effet, l'erreur quadratique moyenne calculée sur ce dernier notée E_{entr} est directement utilisée pour la correction des poids \mathbf{Z} et \mathbf{W} en suivant la règle de la méthode d'optimisation choisie.
- **Lot de validation** : Il permet d'estimer l'erreur E_{val} réalisée par le R.N à chaque itération sur un jeu de données n'ayant pas servi lors de la phase de modification de poids. Cette erreur décroît de la même manière que E_{entr} au début de l'apprentissage et dès qu'elle commence à croître, l'apprentissage est stoppé (figure 2.15). A ce moment-là, la performance du R.N en termes de généralisation est supposée optimale.
- **Lot de test** : Ce corpus est constitué des données n'ayant pas été utilisées par les deux lots précédents. Il permet de tester les performances globales du R.N sans biais. C'est l'erreur définie en 2.61 (E_{test}^m).

C'est une méthode rapide et facile à mettre en œuvre, mais il est important de procéder à une bonne répartition des données dans les différents lots pour un fonctionnement correct. Dans la suite de nos travaux, toutes les données utilisées pour l'entraînement des R.N sont réparties de la façon suivante : 70% pour l'apprentissage, 15% pour la validation et 15% pour le test.

L'apprentissage du PMC classifieur s'effectue de la même façon que pour la régression en minimisant la fonction coût des moindres carrées (équation 2.60). Les poids \mathbf{W} et \mathbf{Z} du PMC classifieur sont estimés à partir d'un ensemble d'apprentissage où chaque entrée \mathbf{X} est associée au vecteur \mathbf{Y} qui code la classe correspondante.

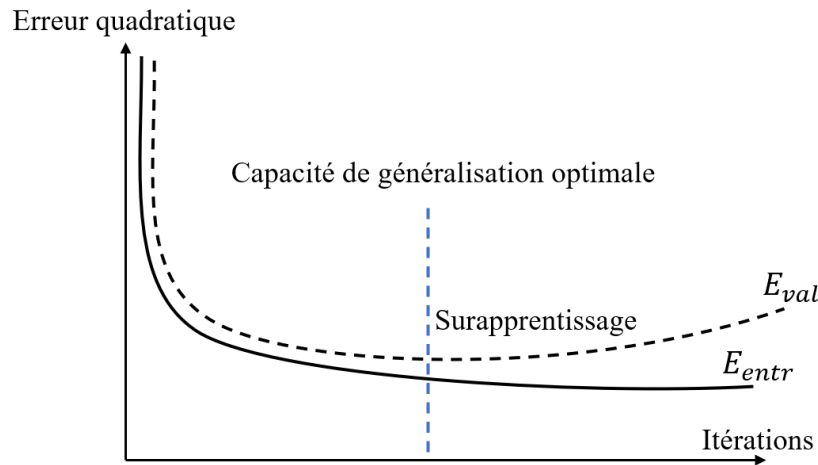


FIGURE 2.15 – Evolution schématique des erreurs quadratiques calculées sur le corpus de données d'entraînement E_{entr} et de validation E_{val} dans le cadre de méthode d'arrêt prématuré.

Ainsi, le PMC présenté ci-dessus peut-être utilisé en mode régresseur ou en tant que classifieur. Il a démontré ses preuves pour identifier ou classer des observations selon différents critères connus à l'avance. Cette capacité d'identification et de classification peut être mise à profit dans le processus de caractérisation scattérométrique. Le mode classifieur sera largement utilisé dans cette thèse. Le paragraphe suivant illustre le mode régresseur dans le cas d'un exemple précis.

Le lecteur intéressé trouvera dans [176, 177, 178] un état de l'art détaillé du PMC.

2.3.5 Caractérisation par R.N régresseur classique

Dans ce paragraphe, nous présentons un R.N régresseur dans le cadre de la caractérisation scattérométrique. La structure périodique étudiée est fabriquée au sein du Laboratoire Hubert Curien par photolithographie optique par interférence de deux ondes lumineuses. Elle est composée d'un réseau en résine (S1805) ne débouchant pas sur le substrat de silicium (figure 2.16). La période du réseau est de 630 nm . D'après l'image obtenue par l'AFM (figure 2.17), on peut en déduire que le profil est proche d'une forme sinusoïdale. Cette dernière sera donc utilisée comme hypothèse de forme pour effectuer la caractérisation.

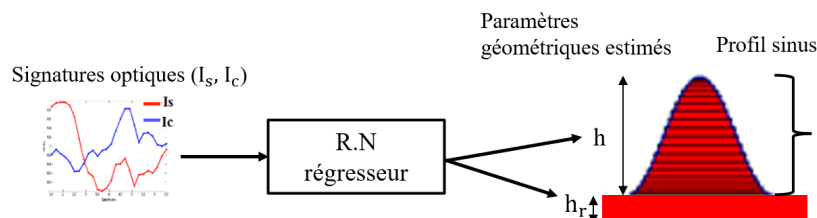


FIGURE 2.16 – Exemple d'utilisation d'un R.N en mode régresseur dans le cadre de la caractérisation scattérométrique permettant d'estimer un profil sinusoïdal défini par deux paramètres géométriques (h et h_r) à partir d'une signature optique fournie en entrée.

Nous avons mis en œuvre un PMC composé d'une couche cachée de 30 neurones en mode régresseur pour déterminer les paramètres géométriques de cet échantillon à partir de sa signa-

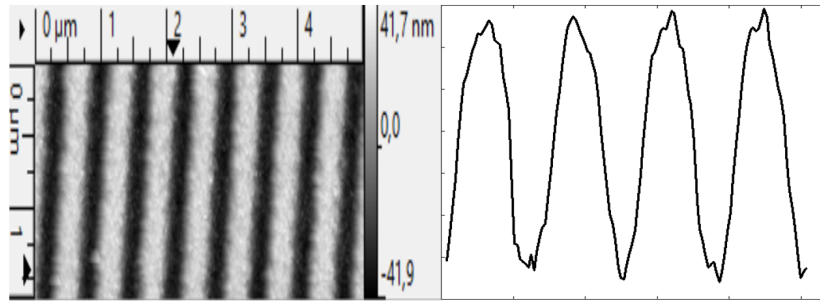


FIGURE 2.17 – Image de l'échantillon obtenue par une caractérisation AFM.

ture optique (figure 2.16). Cette dernière est composée des mesures I_s et I_c réalisées avec un angle d'incidence de 70° sur un spectre allant de 0.7 à 4.5 eV. L'appareil de mesure est un ellipsomètre de marque Jobin Yvon à modulation de phase. L'apprentissage du PMC est effectué sur 3000 échantillons théoriques pour lesquels on a simulé la signature ellipsométrique. Un bruit gaussien ($3\sigma = 0.01$) correspondant à la perturbation d'une mesure réelle est pris en compte. Ces échantillons théoriques sont générés par tirage aléatoire dans les domaines suivants :

$$\begin{aligned} 30 < h < 120 & \text{ nm} \\ 500 < h_r < 600 & \text{ nm} \end{aligned} \quad (2.71)$$

La performance optimale calculée (équation 2.69) sur les données test sur chaque sortie ainsi que les performances globales (équation 2.70) du PMC sont présentées dans la table 2.1. Les valeurs obtenues sont faibles ; ce qui met en valeur la qualité de l'apprentissage. L'évolution des différentes erreurs quadratiques moyennes à l'issue de la phase d'apprentissage est également présentée (figure 2.18). On note ainsi 20 itérations au bout de laquelle la convergence est atteinte.

TABLE 2.1 – Performance du PMC (une couche cachée de 30 neurones) après entraînement évaluée sur les données test (450 échantillons).

Performance du PMC		
$E_{test} = 0.05 \text{ nm}$	$E_{test}^h = 0.06 \text{ nm}$	$E_{test}^{h_r} = 0.03 \text{ nm}$

La mesure de la signature optique de l'échantillon est ensuite fournie à l'entrée du PMC régresseur. La qualité de la caractérisation est évaluée par l'erreur quadratique moyenne sur les signatures optiques (E_{I_s, I_c}) (équation 1.45) calculée entre la signature optique mesurée et celle reconstruite à partir des sorties du PMC (table 2.2).

Le PMC estime des paramètres géométriques inclus dans la plage d'entraînement (équation 2.71) ; ce qui permet de valider le domaine d'apprentissage choisi. L'erreur E_{I_s, I_c} calculée est faible en regard de la précision de mesure de l'appareil ; ce qui montre que le profil supposé est proche de la forme réelle de l'échantillon.

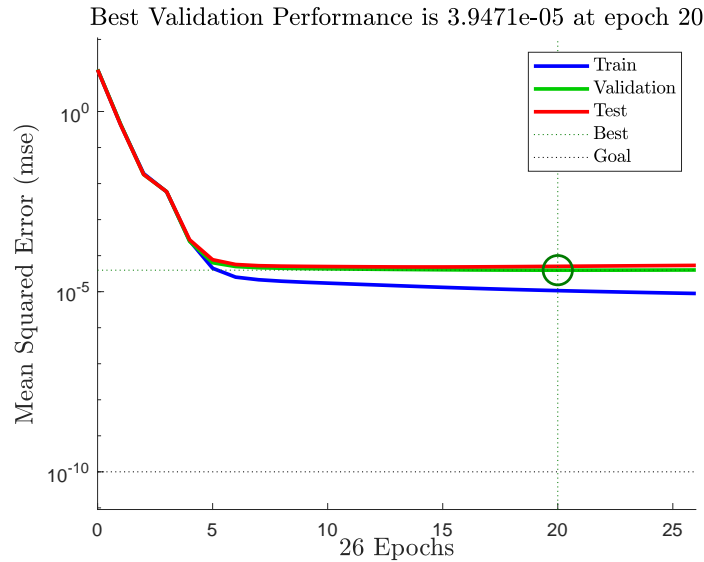


FIGURE 2.18 – Evolution des erreurs quadratiques moyennes calculées sur le corpus de données d'entraînement, de validation et de test dans le cadre de méthode d'arrêt prématuré.

TABLE 2.2 – Résultat de caractérisation scattérométrique du PMC régresseur (une couche cachée de 30 neurones) sur l'échantillon

Paramètres géométriques estimés	Qualité de la caractérisation
$h = 53.75 \pm 0.06 \text{ nm}$	$E_{I_s, I_c} = 0.0479$
$h_r = 575.65 \pm 0.03 \text{ nm}$	

Dans la figure 2.19, on compare les signatures optiques mesurées à celles reconstruites à partir des paramètres géométriques issus de la caractérisation du PMC régresseur.

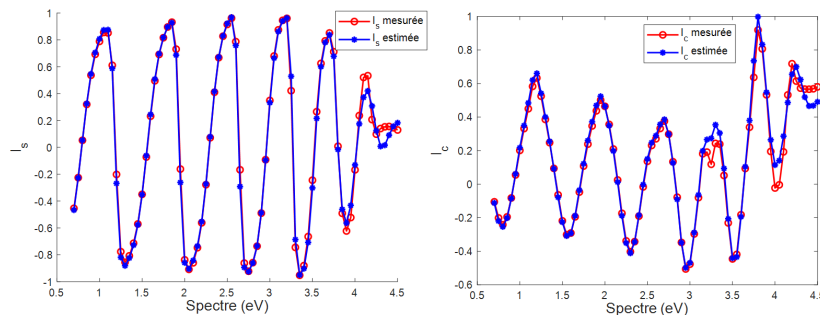


FIGURE 2.19 – Représentation des signatures optiques mesurées de l'échantillon réel et celles reconstruites à partir de la caractérisation par le PMC régresseur fonctionnant sur un modèle géométrique de type sinusoïdal ($h = 53.75 \text{ nm}$ et $h_r = 575.65 \text{ nm}$).

On observe alors que la signature optique calculée est très proche de celle mesurée ; ce qu'on retrouve dans la valeur de E_{I_s, I_c} . Cependant, on peut apercevoir un léger décalage dans la partie UV. Ceci peut provenir de différentes causes. En effet, le profil supposé correspond à une modélisation plus ou moins fidèle à la forme réelle. On peut également s'interroger sur la présence de rugosité non prise en compte dans le modèle.

Nous avons également confronté le résultat de caractérisation par R.N avec celui obtenu par mesure AFM (figure 2.20). Il apparait clairement dans la figure 2.20 que les deux profils estimés sont relativement proches malgré une petite différence sur les hauteurs. Par ailleurs, les résultats de l'AFM ne sont pas réguliers sur plusieurs périodes car la mesure est locale ; ce qui n'est pas le cas de la caractérisation par PMC qui prend en compte plusieurs profils suivant la taille de spot lumineux (1 mm dans ce cas). Enfin, pour finir cette comparaison entre les deux méthodes, notons que l'AFM ne donne aucune information sur l'épaisseur h_r de la couche résiduelle de résine.

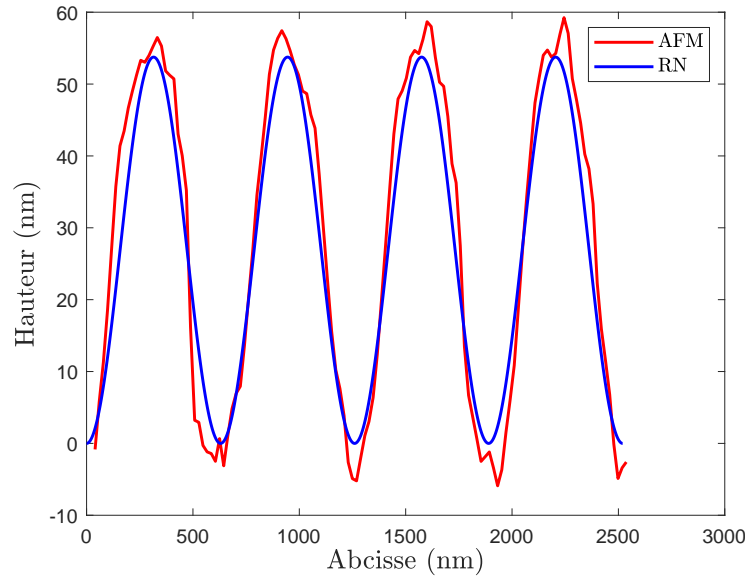


FIGURE 2.20 – Profils géométriques de l'échantillon obtenus par AFM (rouge) et PMC (bleu).

Cette étude nous a permis d'illustrer l'utilisation la plus courante des PMCs en scattérométrie à travers le cas d'un échantillon de forme simple lorsque toutes les conditions préalables étaient satisfaisantes. En pratique, ce n'est malheureusement pas toujours le cas et le travail de cette thèse consiste à apporter des solutions quant à la vérification de ces dernières en amont du processus de caractérisation proprement dit.

2.4 Conclusion

Ce chapitre était consacré à la présentation des méthodes de reconnaissance de formes en particulier la classification automatique. Nous avons ainsi présenté les techniques les plus utilisées dans la littérature : les méthodes indirectes de classification dont les méthodes paramétriques et non paramétriques ainsi que les méthodes directes de classification en s'attardant sur l'utilisation des SVM. Les principaux classifieurs à base de RNs ont ensuite été détaillés. Nous avons notamment décrit le PMC, outil incontournable dans le cadre de cette thèse. Ses potentialités en tant que régresseur ont été mises en valeur mais également en tant que classifieur. Enfin, nous avons effectué des études théoriques et expérimentales dans le cadre de la caractérisation scattérométrique afin de mieux appréhender le fonctionnement classique du PMC en

mode régresseur. Le chapitre suivant est consacré à la caractérisation entièrement neuronale dans le processus de scattérométrie.

Prise en compte des défauts de fabrication dans le processus de caractérisation optique

Comme nous l'avons précisé dans les chapitres précédents, pour garantir l'unicité de la solution, la caractérisation scattérométrique se base sur un certain nombre d'hypothèses comme la connaissance préalable du modèle géométrique de la structure sous test qui est fixé tout au long du processus. Ainsi, l'idée de l'utilisation d'un classifieur est de certifier la forme géométrique et d'écarter tout défaut éventuel de modélisation avant la caractérisation.

Nous allons étudier le comportement du R.N dans le cadre d'une caractérisation classique lorsque toutes les hypothèses de formes sont vérifiées ainsi que sur des formes déviées conduisant à des résultats erronés. Nous proposerons alors une méthode par pondération de profil pour caractériser les formes présentant une déviation minimale par rapport à la cible.

Ensuite, nous nous intéresserons à l'apport du R.N en tant que classifieur pour la détection de formes géométriques dans le processus d'une caractérisation optique. Comme évoquée plus haut, cette application est placée en amont de la phase de caractérisation. Nous démontrerons ainsi la faisabilité de cette méthode en évoquant différents cas d'études. Nous discuterons également d'autres critères pour qualifier la qualité du résultat. De nombreux tests théoriques et expérimentaux ont été réalisés pour valider ces études. Pour finir, nous proposerons une méthode originale basée sur le spectre fréquentiel du profil géométrique spécifiquement adaptée à l'étude de réseaux de forme sinusoidale.

3.1 Cadre de l'étude

3.1.1 Contexte de fabrication

Le domaine visé concerne celui des procédés de fabrication. Dans ce cadre, l'étape de lithographie est cruciale car elle conditionne les dimensions de la structure. La maîtrise parfaite du processus de lithographie ou de gravure est donc importante afin d'écarter certaines dérives ou défauts qui peuvent survenir en cours de fabrication. La déviation de la forme géométrique obtenue par rapport à celle désirée est l'un des principaux défauts qui peut apparaître. En

effet, la forme obtenue est plus ou moins fidèle à celle attendue. Plusieurs raisons peuvent expliquer cela comme une erreur de calibration, un défaut de matériel ou un phénomène lié aux contraintes technologiques.

En scattérométrie, la résolution du problème inverse nécessite la connaissance des certaines conditions expérimentales (plage de longueurs d'onde utilisées, angle d'incidence, etc.) et hypothèses préalables sur la structure à tester (forme géométrique, période, connaissance des indices de réfraction, etc.). La plupart des études dans la littérature n'évoque pas le problème d'une méconnaissance de ces dernières ; ce qui peut avoir des conséquences non négligeables sur les performances de la caractérisation. Il arrive parfois que la structure obtenue soit fautive et ne corresponde pas physiquement à celle attendue. On parle de problème mal posé. Il faut donc se doter d'un moyen de récolter une information qualitative afin de confirmer les hypothèses utilisées lors de la phase de traitement de la signature mesurée.

Le paragraphe suivant est consacré à la présentation des différents types d'échantillons étudiés dans ce chapitre. Nous évoquerons les techniques de fabrications desquelles ils sont issus ainsi que les caractérisations microscopiques réalisées.

3.1.2 Échantillons testés

Plusieurs échantillons de formes, de structures et de dimensions variées issus de différents procédés ont été utilisés pour réaliser notre étude. Le nombre et la diversité de ces échantillons nous a permis de confirmer les résultats dans différents cas de figure.

Le premier échantillon noté E_1 est fabriqué par la société STMicroelectronics par lithographie optique par immersion 193 nm . Il est composé d'un réseau en résine (IM5010) reposant sur une couche du BARC (pour Bottom Anti Reflective Coating), le tout déposé sur un substrat de silicium. Le BARC est une couche intermédiaire qui permet de réduire la réflexion de la lumière lors de l'étape de lithographie. La période du réseau est de $T = 140\text{ nm}$ et l'épaisseur de la couche du BARC est supposée connue (82 nm). La forme cible de l'échantillon est un créneau. L'image obtenue par l'AFM (figure 3.1) montre une forme présentant des arrondis et quelques rugosités sur les flancs.

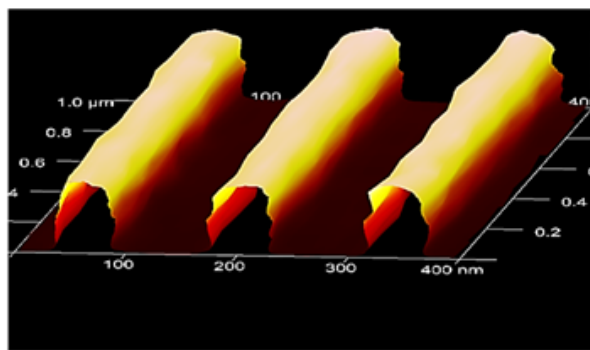


FIGURE 3.1 – Caractérisation AFM de l'échantillon E_1 (période de 140 nm) de réseau en résine (IM5010) reposant sur du BARC ; le tout sur un substrat de silicium.

Le second type d'échantillons concerne ceux fabriqués au Laboratoire des Technologies de la Microélectronique (UMR 5129, Grenoble) par lithographie optique par immersion 248 nm. Les réseaux E_2 , E_3 et E_4 de période $T = 500$ nm sont en résine (M78Y) réalisés sur une couche du BARC, reposant sur une couche de silice (S_iO_2), le tout sur un substrat de silicium. L'épaisseur de la couche du BARC et du S_iO_2 sont connues, respectivement 49 nm et 50 nm. Les résultats de caractérisation de ces échantillons obtenus par le MEB sont présentés dans la figure 3.2. La forme géométrique ciblée est également un créneau.

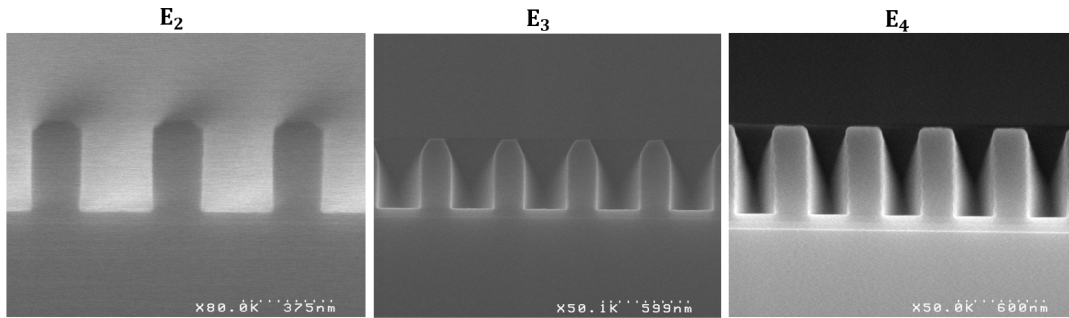


FIGURE 3.2 – Caractérisation MEB des échantillons E_2 , E_3 et E_4 (période de 500 nm) de réseau en résine (M78Y) sur du BARC et une couche du S_iO_2 ; le tout sur un substrat de silicium.

Enfin, la dernière série (de E_5 à E_{10}) concerne les réseaux fabriqués au Laboratoire Hubert Curien (UMR CNRS 5516, Saint-Étienne) par photolithographie optique par interférence de deux ondes lumineuses. Ils sont composés de réseaux en résine (S1805) ne débouchant pas sur le substrat de silicium. La période du motif est de $T = 627$ nm. La figure 3.3 montre différents résultats de caractérisation locale obtenus par AFM pour chaque type d'échantillon. La forme géométrique sinusoïdale est, cette fois ci, attendue. L'obtention des différentes formes plus ou moins éloignées du profil cible ont volontairement été obtenues en modifiant les conditions expérimentales.

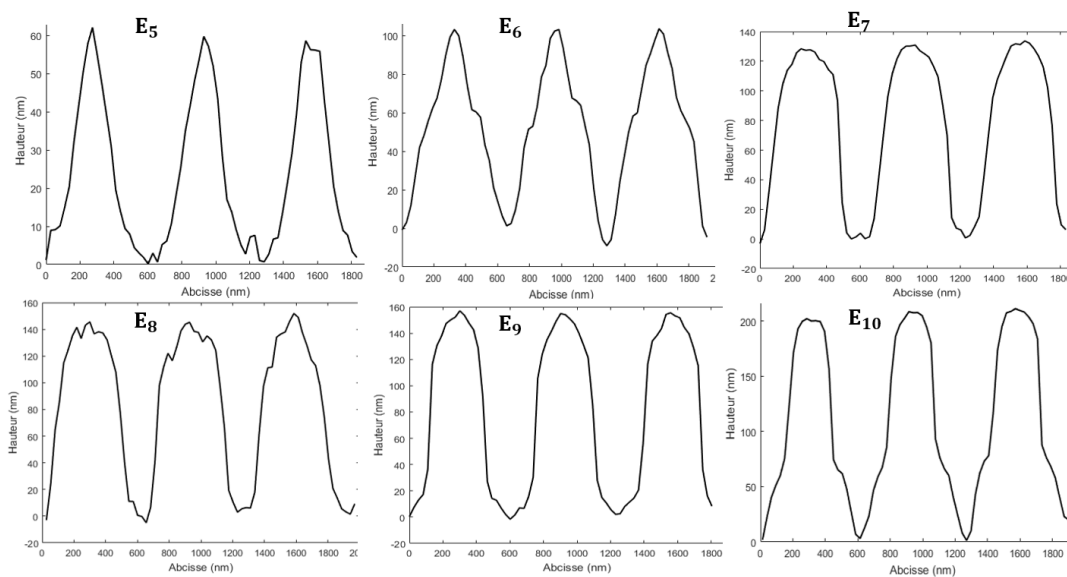


FIGURE 3.3 – Caractérisation AFM des échantillons E_5 , E_6 , E_7 , E_8 , E_9 et E_{10} (période de 627 nm) de réseau en résine (S1805) ne débouchant pas sur le substrat de silicium.

3.1.3 Caractérisation par scattérométrie «classique»

Comme mentionné précédemment (chapitre 1), la connaissance au préalable de la forme géométrique de l'échantillon à caractériser est une des hypothèses les plus importantes de la méthode scattérométrique. L'objectif de cette partie est d'étudier l'influence du choix d'un modèle géométrique sur les performances du R.N. Généralement, on suppose que toutes les hypothèses sont vérifiées mais ce n'est pas toujours le cas (avec des degrés moindres). Par ailleurs, le motif peut aussi évoluer en cours de fabrication et s'éloigner significativement de la forme cible. Il est donc impératif de concevoir un système qui soit capable de détecter cette erreur ou dérive.

Fixons par exemple, un modèle géométrique théorique de type créneau considéré comme profil cible pour décrire une structure périodique composée d'un réseau en ligne de période $T = 140 \text{ nm}$. Le réseau est en résine (IM5010) réalisé sur du BARC (épaisseur de 82 nm), le tout reposant sur un substrat de silicium (modèle type de l'échantillon E_1). Les signatures optiques sont simulées par MMMFE et définies sur un spectre allant de 1.5 à 6.5 eV (soit 52 valeurs) pour un angle d'incidence de 62° . Un bruit gaussien ($3\sigma = 0.01$) correspondant à la perturbation d'une mesure réelle est systématiquement pris en compte. Le modèle du PMC choisi pour effectuer la caractérisation par scattérométrie classique est noté RegPc. Il est composé de 25 neurones cachés. Les 7000 profils composant la base d'entraînement sont générés par tirage aléatoire dans les domaines suivants :

$$\begin{aligned} 40 < CD < 90 \quad nm \\ 40 < h < 130 \quad nm \end{aligned} \tag{3.1}$$

La performance théorique optimale calculée (équation 2.69) sur les données test (1050 échantillons) pour chaque sortie du RegPc est donnée dans la table 3.1. On constate que les erreurs obtenues sont relativement faibles ; ce qui met en valeur la qualité de l'apprentissage.

TABLE 3.1 – Performance de RegPc (une couche cachée composée de 25 neurones) après entraînement évaluée sur les données test (1050 échantillons) fonctionnant exclusivement sur un profil créneau.

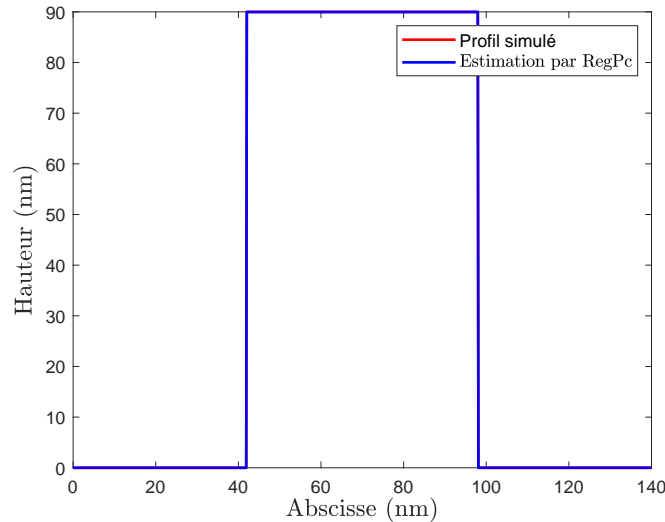
Performance de RegPc	
$E_{test}^{CD} = 0.1 \text{ nm}$	$E_{test}^h = 0.1 \text{ nm}$

Nous avons ensuite défini un échantillon de forme simple, de type rectangulaire puis simulé sa signature optique. Cette dernière est introduite à l'entrée du RegPc. Le résultat de caractérisation est donné dans la table 3.2.

TABLE 3.2 – Estimation par RegPc d'un profil créneau parfait et qualité du résultat obtenu

	Profil simulé	RegPc
Paramètres géométriques	$CD = 56 \text{ nm}$ $h = 90 \text{ nm}$	$CD_{est} = 56.02 \pm 0.1 \text{ nm}$ $h_{est} = 89.97 \pm 0.1 \text{ nm}$
Qualité de la caractérisation		$E_{I_s, I_c} = 4.95 \times 10^{-4}$ $E_{geo} = 1.68 \text{ nm}^2$ $E_{geonorm} = 0.03\%$

Le RegPc estime des paramètres géométriques inclus dans la plage définie (équation 3.1). L'erreur E_{I_s, I_c} et l'écart géométrique E_{geo} sont très faibles confirmant ainsi une bonne qualité de caractérisation. La figure 3.4 représente le profil simulé et celui estimé par le RegPc. Comme on pouvait s'y attendre, les 2 profils sont confondus.

FIGURE 3.4 – Caractérisation d'un profil créneau théorique ($CD = 56 \text{ nm}$ et $h = 90 \text{ nm}$) en rouge par RegPc (en bleu).

Ces excellentes performances doivent bien entendu être nuancées. En effet, le R.N est capable de donner cette précision sur les paramètres géométriques dans le cas où toutes les hypothèses sont vérifiées parfaitement. Ceci inclut les conditions de mesures (angle d'incidence, longueurs d'onde), les conditions liées à la nature du matériau (indice de réfraction, absence de rugosité) mais également les hypothèses structurelles (forme du réseau, épaisseurs des couches résiduelles) ; ce qui, en pratique, est plus ou moins bien respectées.

3.1.4 Hypothèse de forme géométrique inhérente à toute méthode de scattérométrie

Dans les mêmes conditions que l'étude précédente (paragraphe 3.1.3), considérons à présent une forme rectangulaire présentant de faibles arrondis en haut et bas considérés comme un

défaut de forme ($r_h = 35.1 \text{ nm}$, $r_b = 33 \text{ nm}$, $CD = 78 \text{ nm}$ et $h = 119 \text{ nm}$). Les signatures optiques simulées sont introduites à l'entrée du RegPc. Le résultat de caractérisation est donné dans la table 3.3.

TABLE 3.3 – Estimation par RegPc d'un profil rectangulaire arrondis en haut et bas et qualité du résultat obtenu.

	Profil simulé	RegPc
Paramètres géométriques	$r_h = 35.1 \text{ nm}$ $r_b = 33 \text{ nm}$ $CD_h = 78 \text{ nm}$ $h = 119 \text{ nm}$	$CD_{est} = 73.72 \pm 0.1 \text{ nm}$ $h_{est} = 118.85 \pm 0.1 \text{ nm}$
Qualité de la caractérisation		$E_{I_s, I_c} = 0.0947$ $E_{geo} = 1.23 \times 10^3 \text{ nm}^2$ $E_{geonorm} = 14.13\%$

Les paramètres géométriques estimés sont bien inclus dans les domaines de variations définis (équation 3.1), donc tout porte à croire que les hypothèses sont vérifiées. L'erreur E_{I_s, I_c} est plus élevé que précédemment mais reste assez faible. Par contre, l'écart géométrique E_{geo} est largement plus élevé ($\times 1000$). Cela peut correspondre à une déformation du profil géométrique. La figure 3.5 représente les profils correspondants.

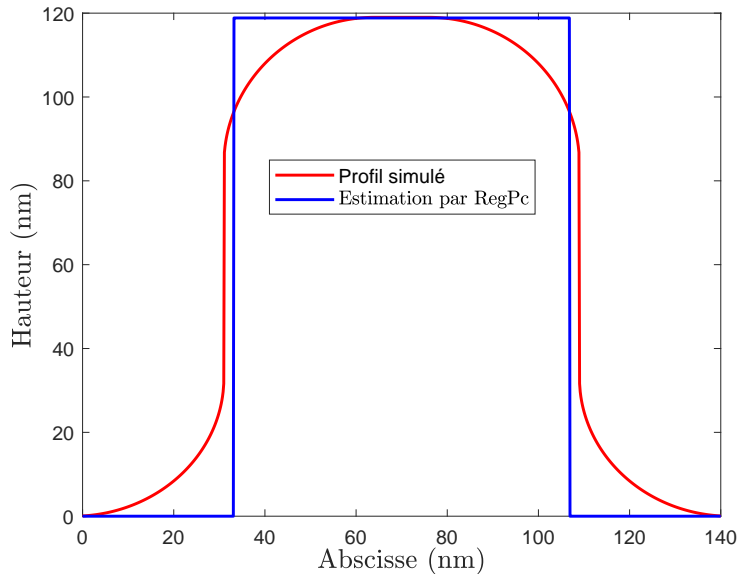


FIGURE 3.5 – Caractérisation d'un profil rectangulaire théorique arrondis en haut et bas ($r_h = 35.1 \text{ nm}$, $r_b = 33 \text{ nm}$, $CD = 78 \text{ nm}$ et $h = 119 \text{ nm}$) en rouge par RegPc (en bleu).

Contrairement au premier exemple (paragraphe 3.1.3), le profil estimé est éloigné de la forme exacte. Le profil présente des arrondis en haut et bas qui ne peuvent être pris en compte par le modèle fixé pour la caractérisation.

La figure 3.6 représente les signatures optiques calculées à partir du résultat du RegPc comparées à celle du profil réel.

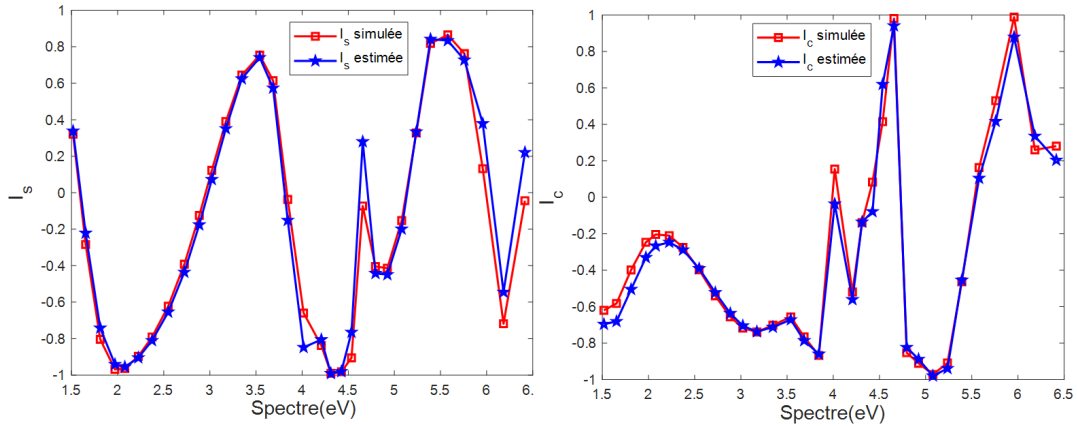


FIGURE 3.6 – Comparaison des signatures optiques simulées à partir du profil réel de forme créneau arrondi en haut et bas (en rouge) et calculées à partir de celui estimé par RegPc (en bleu).

En moyenne, comme le confirme la valeur de E_{I_s, I_c} , les signatures affichées sont relativement proches. Il apparaît cependant localement quelques décalages qui peuvent être représentatifs d'un écart géométrique. Ces zones d'intérêt doivent être explorées.

Le RegPc donne donc comme résultat le profil géométrique créneau produisant une signature optique la plus proche de celle simulée indépendamment de tout écart géométrique induit sur la forme réelle. Le problème est « mal posé ». Ce résultat montre une fois de plus que la seule information sur l'erreur entre les signatures ne nous permet pas de détecter un problème mal posé.

Cette étude sur un cas simple nous a permis de comprendre l'influence des hypothèses en particulier structurelles sur le résultat de la caractérisation optique. Dans certains cas où l'hypothèse est largement fautive, la scattérométrie « classique » peut être suffisante pour alerter sur la suspicion d'un problème mal posé en donnant par exemple un résultat correspondant à un profil non inclus dans la base d'apprentissage. Cependant, la plupart du temps ce n'est pas le cas et cela conduit à des résultats qui peuvent être fortement erronés. Dans sa forme actuelle, la scattérométrie classique est alors incapable de fournir une indication qualitative sur la forme supposée de l'échantillon.

3.2 Méthode de pondération de profils

3.2.1 Principe de la méthode

L'étude par pondération de profil consiste à effectuer une combinaison de plusieurs résultats de caractérisation classique de profils simples. L'idée est d'utiliser le seul critère d'erreur généralement employé E_{I_s, I_c} afin de pondérer les résultats de caractérisation d'un échantillon issu de différents modèles simples pour reconstruire la forme géométrique de l'échantillon. Pour

cela, nous nous appuyons sur le principe suivant : plus une erreur E_{I_s, I_c} calculée est faible plus le résultat correspondant est supposé fiable et inversement. On lui affecte donc un coefficient de pondération noté α_t élevé. Ainsi la forme $F(x)$ reconstruite par pondération est donnée par (équation 3.2) :

$$F(x) = \sum_t \alpha_t f_t(x) \quad (3.2)$$

où $f_t(x)$ est la fonction définissant le résultat de la caractérisation sur le modèle géométrique simple t et α_t le coefficient de pondération qui lui est affecté et défini par :

$$\alpha_t = \frac{\beta_t}{\sum \beta_t}$$

avec

$$\beta_t = \frac{\sum_t E_{I_s, I_c}}{E_{I_s, I_c}}$$

3.2.2 Résultats par scattérométrie classique

Afin de mettre en place la méthode par pondération, il faut définir un ensemble de modèles simples supposés proches de l'échantillon à tester (modèle type de l'échantillon E_1). Prenons, par exemple, 4 modèles de formes géométriques dédiées à la caractérisation par scattérométrie classique : le profil trapézoïdal symétrique noté P_{tr} , le profil rectangulaire arrondi en haut noté P_{ah} , le profil rectangulaire arrondi en bas noté P_{ab} et le profil sinus tronqué noté P_{st} . Les domaines de variation de chacun des paramètres géométriques utilisés pour l'apprentissage sont les suivants :

- Modèle P_{tr}

$$\begin{aligned} 5 < b_1 < 20 & \quad nm \\ 30 < b_2 < 50 & \quad nm \\ 40 < h < 130 & \quad nm \end{aligned} \quad (3.3)$$

- Modèle P_{ah}

$$\begin{aligned} 0.005 < \delta < 1 \\ (0.1 < r_h < 45 & \quad nm) \\ 40 < CD_h < 90 & \quad nm \\ 40 < h < 130 & \quad nm \end{aligned} \quad (3.4)$$

- Modèle P_{ab}

$$\begin{aligned} 5 < r_b < 30 & \quad nm \\ 40 < CD_b < 90 & \quad nm \\ 40 < h < 130 & \quad nm \end{aligned} \quad (3.5)$$

- Modèle P_{st}

$$\begin{aligned} 40 < b < 90 & \quad nm \\ 40 < h < 130 & \quad nm \end{aligned} \quad (3.6)$$

Chaque caractérisation est indépendante et fonctionne sur un seul type de modèle géométrique pour la résolution du problème inverse. Dans ce qui suit, le modèle du PMC noté RegPtr (25 neurones cachés) est associé au profil trapézoïdal symétrique, le modèle RegPah (30 neurones cachés) au profil rectangulaire arrondi en haut, le modèle RegPab (40 neurones cachés) au profil créneau arrondi en bas et enfin le modèle RegPst (25 neurones cachés) au profil sinus tronqué. Chaque modèle du PMC régresseur est entraîné sur 7000 échantillons théoriques.

Les performances optimales de chaque sortie calculées sur le corpus test sont données dans la table 3.4. Ce dernier affiche des faibles valeurs ; ce qui confirme la qualité de chaque entraînement.

TABLE 3.4 – Performances des PMC fonctionnant sur les différents modèles géométriques simples P_{tr} , P_{ah} , P_{ab} et P_{st}

RegPtr	RegPah	RegPab	RegPst
$E_{test}^{b_1} = 0.1 \text{ nm}$	$E_{test}^{\delta} = 0.1$	$E_{test}^{r_b} = 0.3 \text{ nm}$	
$E_{test}^{b_2} = 0.2 \text{ nm}$	$E_{test}^{CD_h} = 0.2 \text{ nm}$	$E_{test}^{CD_b} = 0.2 \text{ nm}$	$E_{test}^b = 0.06 \text{ nm}$
$E_{test}^h = 0.1 \text{ nm}$	$E_{test}^h = 0.3 \text{ nm}$	$E_{test}^h = 0.1 \text{ nm}$	$E_{test}^h = 0.04 \text{ nm}$

Afin de tester la méthode, nous allons, dans un premier temps, simuler une forme géométrique résultant d'une dérive d'un procédé de fabrication d'un profil cible rectangulaire. Cette forme prend en compte l'ensemble des principaux défauts occasionnés lors des étapes de gravure : flancs non verticaux, arrondi en haut et en bas (figure 3.7 en rouge). Ce profil n'est pas représenté par les différents modèles simples des PMC présentés. Les signatures optiques correspondantes sont simulées puis introduites à l'entrée des différents modèles des PMC (RegPtr, RegPah, RegPab et RegPst).

Le résultat de chacune des caractérisations est donné dans la table 3.5. Les différents PMC estiment des valeurs de paramètres géométriques inclus dans les plages initialement définies lors

de l'étape d'apprentissage. L'erreur E_{I_s, I_c} calculée est relativement faible (en particulier pour RegPtr et RegPab). Ceci est confirmé par la valeur de l'écart géométrique E_{geo} .

TABLE 3.5 – Résultats de caractérisation des différents PMC fonctionnant indépendamment sur une forme simple concernant un échantillon présentant une forme complexe simulant la déviation d'un processus de fabrication.

PMC	Paramètres géométriques	Qualité de caractérisation
<i>RegPtr</i>	$b_1 = 8.7 \pm 0.1 \text{ nm}$ $b_2 = 41.9 \pm 0.2 \text{ nm}$ $h = 56.7 \pm 0.1 \text{ nm}$	$E_{I_s, I_c} = 0.008$ $E_{geo} = 145.7 \text{ nm}^2$
<i>RegPah</i>	$\delta = 0.92 \pm 0.1$ $CD_h = 57.5 \pm 0.2$ $h = 59.4 \pm 0.3 \text{ nm}$	$E_{I_s, I_c} = 0.074$ $E_{geo} = 382.9 \text{ nm}^2$
<i>RegPab</i>	$r_b = 20 \pm 0.3 \text{ nm}$ $CD_b = 46.9 \pm 0.2 \text{ nm}$ $h = 55.8 \pm 0.1 \text{ nm}$	$E_{I_s, I_c} = 0.01$ $E_{geo} = 165.8 \text{ nm}^2$
<i>RegPst</i>	$b = 30 \pm 0.06 \text{ nm}$ $h = 65.5 \pm 0.04 \text{ nm}$	$E_{I_s, I_c} = 0.035$ $E_{geo} = 526.8 \text{ nm}^2$

La figure 3.7 illustre les différents profils estimés par rapport à la forme supposée réelle.

Tous les PMC sont limités par la complexité de leur forme fixe pendant l'étape d'apprentissage et ne conviennent pas pour caractériser une forme plus complexe. En effet, le R.N ne peut donner que la forme fixe qu'il décrit produisant une signature optique proche de celle mesurée sur l'échantillon réel. Comme cela a été évoqué précédemment, cela peut conduire à une déformation significative du profil géométrique. C'est le cas pour RegPst qui conduit à une erreur E_{I_s, I_c} faible mais avec un E_{geo} élevé. L'alternative proposée ici consiste à introduire une méthode par pondération de profils en utilisant les résultats de ces PMC minimisant ainsi l'impact de mauvaises caractérisations.

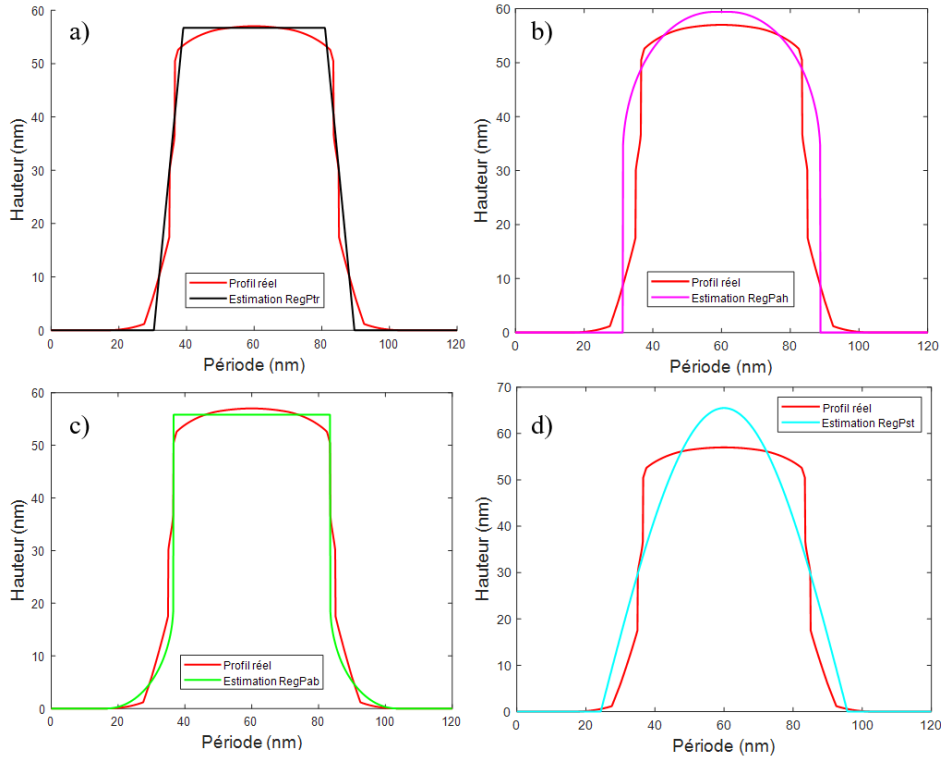


FIGURE 3.7 – Représentations des profils géométriques obtenus par différentes caractérisations d'une forme complexe simulant la déviation d'un processus de fabrication (en rouge) : (a) RegPtr (en noir), (b) RegPah (en magenta), (c) RegPab (en vert) et (d) RegPst (en cyan).

3.2.3 Résultats par pondération de profils

3.2.3.1 Résultats théoriques

Les coefficients de pondération α_t sont calculés conformément à la méthode décrite précédemment (équation 3.2). Le profil reconstruit à l'aide de la méthode par pondération de profils est comparé avec la forme réelle sur la figure 3.8. L'erreur E_{I_s, I_c} et l'écart E_{geo} de la forme reconstruite sont affichés dans la table 3.6.

TABLE 3.6 – Résultat par la méthode de pondération de profils d'une forme complexe simulant la déviation d'un processus de fabrication.

Coefficients de pondération	$\alpha_{Ptr} = 0.49$ $\alpha_{Pah} = 0.05$ $\alpha_{Pab} = 0.35$ $\alpha_{Pst} = 0.11$
Qualité de caractérisation	$E_{I_s, I_c} = 0.006$ $E_{geo} = 88.8 \text{ nm}^2$

L'erreur E_{I_s, I_c} et l'écart géométrique E_{geo} sont réduits par rapport à toutes les caractérisations simples précédentes. Notons, en particulier la valeur de E_{geo} , qui présente une diminution de 61% par rapport au meilleur résultat de la table 3.5 (RegPtr).

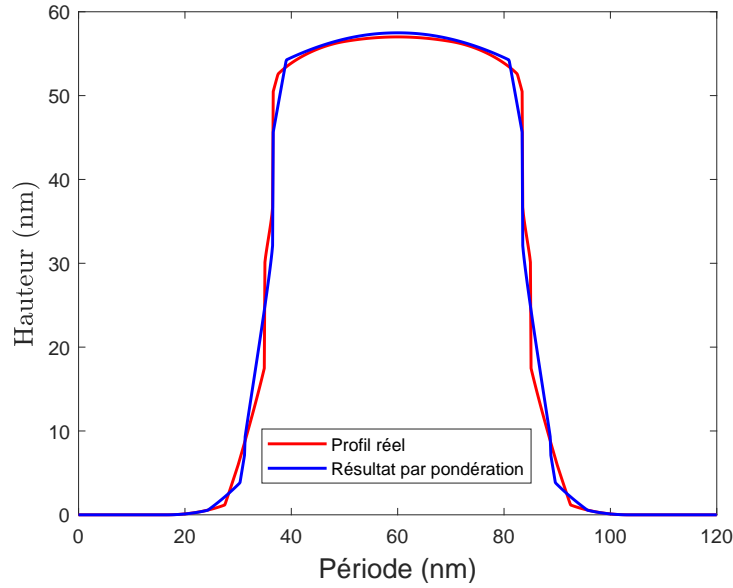


FIGURE 3.8 – Représentation de la forme obtenue par pondération de profils (en bleu) dans le cadre d'une caractérisation d'une forme complexe simulant la déviation d'un processus de fabrication (en rouge).

La méthode par pondération de profils a permis d'améliorer significativement les résultats obtenus par une caractérisation fonctionnant sur différents modèles géométriques simples.

Afin de démontrer la pertinence de l'utilisation d'une telle méthode, une autre approche basée sur la scattérométrie classique est réalisée mais cette fois ci sur un modèle plus complexe dite générique [63]. Le profil est défini par 4 paramètres : CD_{top} et CD_{bottom} (représentant les dimensions critiques, respectivement, sur le haut et le bas du motif), la hauteur h et un paramètre γ ($0 \leq \gamma \leq 1$) quantifiant la forme arrondie au sommet. Mathématiquement, le profil est défini pour : $CD_{top} \leq CD_{bottom}$, $CD_{bottom} < T$ et $h > 0$. Ce modèle inclut plusieurs types de profils tels que le profil rectangulaire, le profil trapézoïdal, le profil rectangulaire et trapézoïdal arrondis au sommet (figure 3.9). Les échantillons théoriques de la base d'entraînement sont générés par tirage aléatoire tout en tenant compte des conditions d'existence physique de chaque profil parmi les domaines suivant incluant les modèles géométriques du paragraphe 3.2.2.

$$\begin{aligned}
 0 &\leq \gamma \leq 1 \\
 35 &< CD_{top} < 45 \quad nm \\
 45 &< CD_{bottom} < 65 \quad nm \\
 45 &< h < 80 \quad nm
 \end{aligned} \tag{3.7}$$

A l'issue de la phase d'apprentissage, la performance calculée sur les données test (1050 échantillons) sur chaque sortie du PMC régresseur générique est donnée dans la table 3.7. Étant donné l'augmentation du nombre de paramètres décrivant la forme, les résultats bruts sont légèrement dégradés par rapport aux modèles simples (table 3.4).

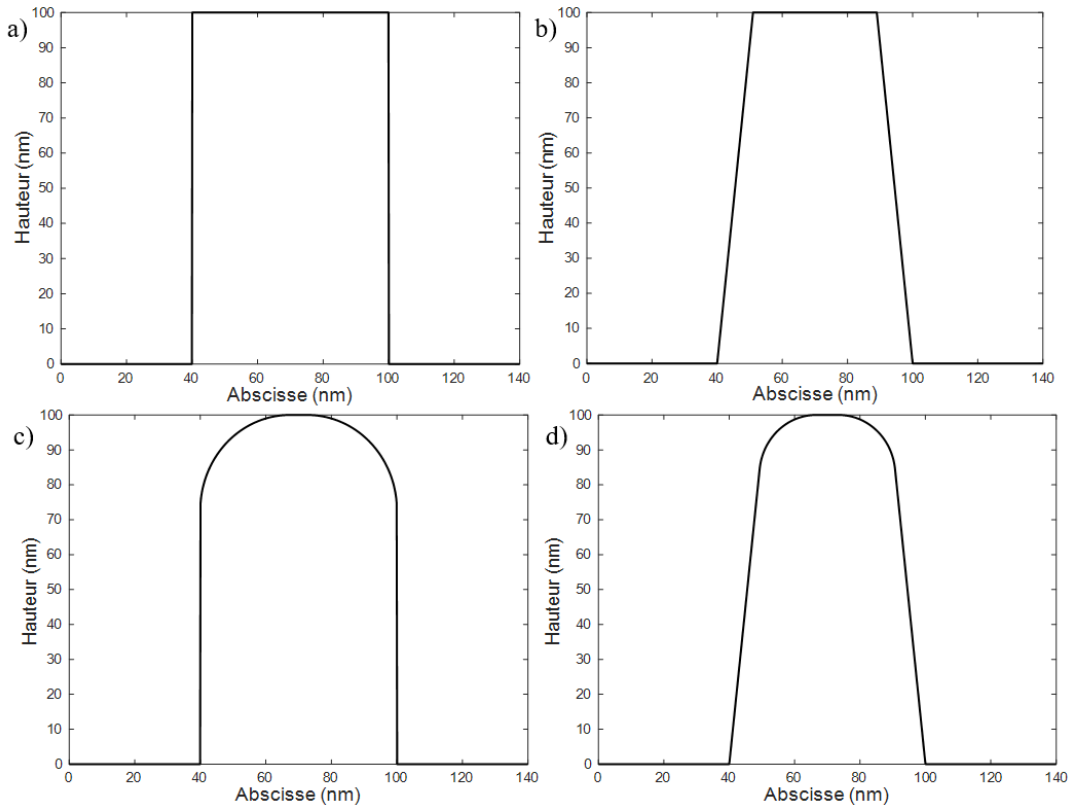


FIGURE 3.9 – Modélisation paramétrique des profils géométriques largement utilisés en scattérométrie : profil rectangulaire (a), profil trapézoïdal (b), profil rectangulaire arrondi au sommet (c) et profil trapézoïdal arrondi au sommet (d).

TABLE 3.7 – Performance du PMC régresseur générique évaluée sur les données test (1050 échantillons) fonctionnant sur un profil générique.

Performance du PMC régresseur générique			
$E_{test}^\gamma = 0.02$	$E_{test}^{CD_{top}} = 0.99 \text{ nm}$	$E_{test}^{CD_{bottom}} = 0.97 \text{ nm}$	$E_{test}^h = 0.13 \text{ nm}$

Les signatures optiques simulées correspondantes à la forme géométrique complexe définie au paragraphe 3.2.3 sont fournies à l'entrée du PMC régresseur générique. Le résultat de la caractérisation est donné dans la table 3.8.

TABLE 3.8 – Résultat de caractérisation du PMC régresseur générique d'une forme complexe simulant la déviation d'un processus de fabrication.

Paramètres géométriques	$\gamma = 0.03 \pm 0.02$ $CD_{top} = 41.42 \pm 0.99 \text{ nm}$ $CD_{bottom} = 59.22 \pm 0.94 \text{ nm}$ $h = 56.44 \pm 0.13 \text{ nm}$
Qualité de caractérisation	$E_{I_s, I_c} = 0.014$ $E_{geo} = 141.42 \text{ nm}^2$

La figure 3.10 compare le profil estimé par rapport au profil réel. Le PMC régresseur générique semble présenter un profil proche de la forme obtenue par le RegPtr (figure 3.7 (a)). Les valeurs des paramètres géométriques estimées sont bien incluses dans la plage (équation 3.7). L'erreur E_{I_s, I_c} calculée est faible et du même ordre de grandeur que E_{ref} . L'écart géométrique E_{geo} est du même ordre de grandeur que le meilleur résultat de caractérisation impliquant une forme simple (table 3.5). Par contre, il reste supérieur au résultat de la méthode par pondération de profils (table 3.6).

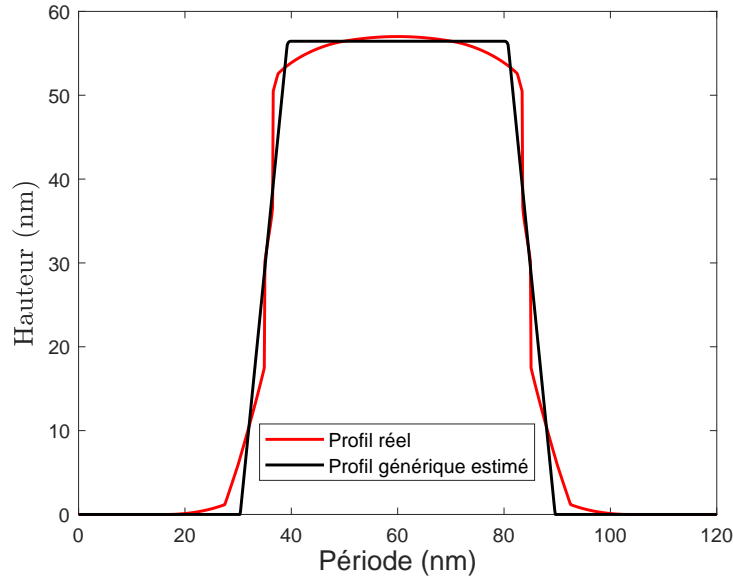


FIGURE 3.10 – Représentation de la forme obtenue par PMC régresseur générique (en noir) d'un profil complexe simulant la déviation d'un processus de fabrication (en rouge).

Ainsi, l'utilisation d'un profil générique peut fournir des résultats fiables. Il permet de prendre en compte plusieurs profils. Cependant, il est souvent limité par la complexité des contraintes d'existence physique imposées sur ses paramètres en particulier lorsque les plages de variation augmentent. Ainsi, le résultat fourni conduit de manière assez fréquente à une forme géométrique non définie mathématiquement. Par ailleurs, dans le cas où la forme existe, (cas de cette étude), nous avons démontré que l'utilisation de la méthode par pondération donne de meilleurs résultats. En effet la forme générique, fonctionnant sur un nombre de paramètres plus élevé, nécessite des architectures de R.N plus complexes et, de ce fait, moins favorables à de bonnes performances en termes de généralisation.

3.2.3.2 Résultats expérimentaux

La méthode par pondération de profils est à présent testée sur un cas expérimental. Considérons l'échantillon noté E_1 (figure 3.1) présenté au paragraphe 3.1.2. Les conditions d'étude sont légèrement différentes de l'étude théorique. En effet, la signature optique de E_1 est mesurée dans une gamme spectrale allant de 1.5 à 6.5 eV (soit 64 valeurs). L'angle d'incidence est de 62° . Cette dernière est fournie aux différents modèles de PMC précédemment entraînés (RegPc, RegPtr, RegPah, RegPab et RegPst). Le résultat de chacune des caractérisations est présenté

dans la table 3.9. L'écart géométrique réel n'étant pas disponible, nous avons calculé l'écart entre les profils estimés par les différents PMC et la mesure AFM tout en gardant à l'esprit que la mesure AFM ne constitue pas une mesure exacte du profil.

TABLE 3.9 – Résultats de caractérisation des différents PMC fonctionnant sur une forme simple concernant la caractérisation de l'échantillon E_1 .

PMC	Paramètres géométriques	Qualité de caractérisation
<i>RegPc</i>	$CD = 44.15 \pm 0.1 \text{ nm}$ $h = 109.44 \pm 0.1 \text{ nm}$	$E_{I_s, I_c} = 0.1288$ $E_{geo} = 1090.8 \text{ nm}^2$
<i>RegPtr</i>	$b_1 = 13.19 \pm 0.1 \text{ nm}$ $b_2 = 33.05 \pm 0.2 \text{ nm}$ $h = 111.49 \pm 0.1 \text{ nm}$	$E_{I_s, I_c} = 0.1003$ $E_{geo} = 798.15 \text{ nm}^2$
<i>RegPah</i>	$\delta = 3.44 \pm 0.1$ $CD_h = 50.53 \pm 0.2 \text{ nm}$ $h = 118.66 \pm 0.3 \text{ nm}$	Profil inexistant
<i>RegPab</i>	$r_b = 30.81 \pm 0.3 \text{ nm}$ $CD_b = 42.32 \pm 0.2 \text{ nm}$ $h = 108.44 \pm 0.1 \text{ nm}$	$E_{I_s, I_c} = 0.1221$ $E_{geo} = 953.65 \text{ nm}^2$
<i>RegPst</i>	$b = 72.6 \pm 0.06 \text{ nm}$ $h = 119.33 \pm 0.04 \text{ nm}$	$E_{I_s, I_c} = 0.1317$ $E_{geo} = 1120.9 \text{ nm}^2$

Les différents modèles de PMC estiment des paramètres géométriques inclus dans les plages d'apprentissage excepté le RegPah qui retourne un profil non défini. Dans ce cas, il est clair que le problème est mal posé pour ce PMC et on peut facilement écarter cette forme supposée. L'information fournie par ce PMC ne sera donc pas utilisée dans la méthode de pondération. Concernant les autres résultats, les erreurs E_{I_s, I_c} calculées entre les signatures optiques sont faibles (mais largement plus grandes que E_{ref}). La forme supposée ne semble donc pas être décrite par les différents modèles sans forcément en être trop éloignée. Ceci est confirmé par la figure 3.11 qui compare les différentes signatures optiques calculées à partir du résultat des différents PMC avec la signature optique mesurée de l'échantillon E_1 . En effet, les signatures estimées par les PMC sont relativement proches de la signature mesurée mais présentent un léger décalage dans la partie UV.

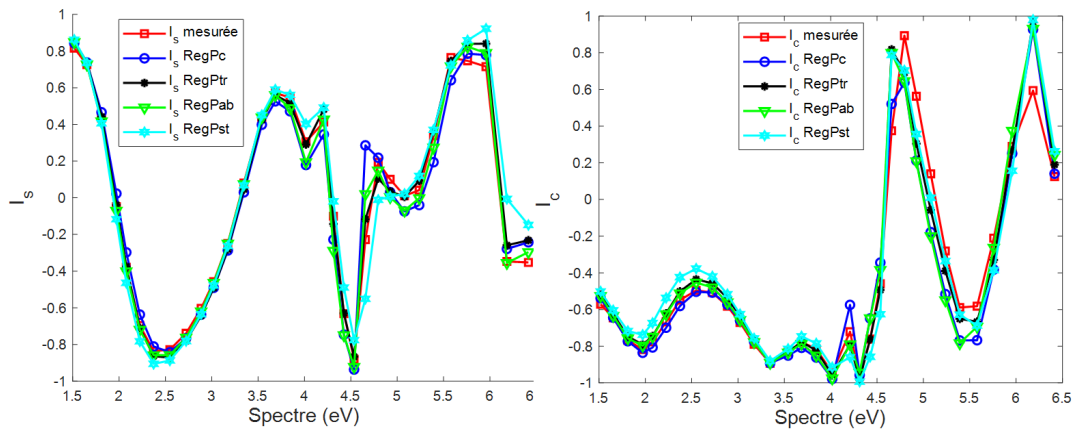


FIGURE 3.11 – Comparaison des signatures optiques calculées à partir des résultats des PMC : RegPc (bleu), RegPtr (noir), RegPab (vert), RegPst (cyan) comparées à celle mesurée (rouge) de l'échantillon E_1

Les profils estimés par les différents PMC sont tracés sur la figure 3.12. La caractérisation par AFM est également ajoutée. L'AFM montre un profil avec des arrondis haut et bas et présentant quelques rugosités de surface. Il faut noter que les profils estimés par les PMC ne tiennent pas compte de la rugosité éventuelle dans la modélisation mathématique.

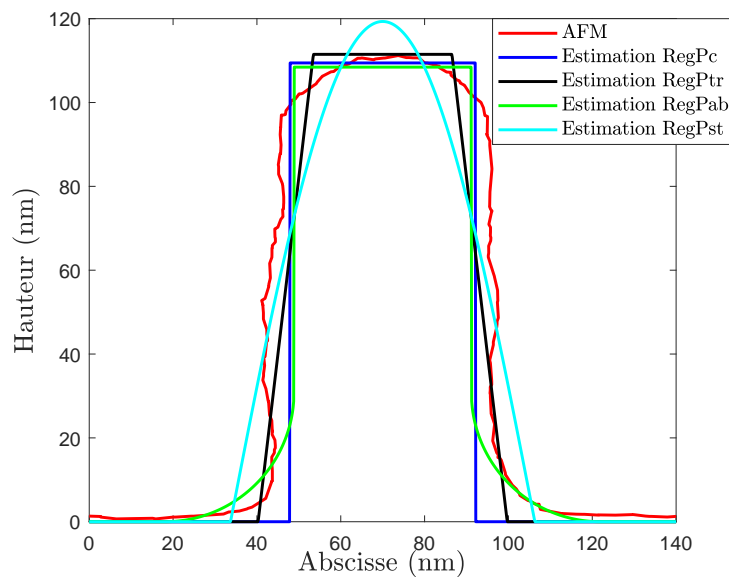


FIGURE 3.12 – Représentation des profils obtenus par caractérisation de l'échantillon E_1 à partir de RegPc (en bleu), RegPtr (en noir), RegPab (en vert), RegPst (en cyan) et par l'AFM (en rouge).

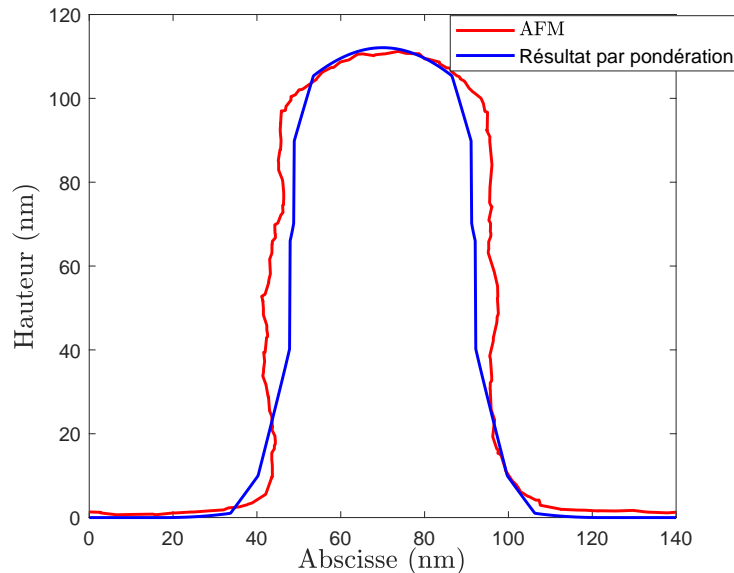
Compte tenu de la complexité de la forme de l'échantillon E_1 , nous avons effectué une étude par pondération de profils à partir des résultats de caractérisation des différents PMC (table 3.9). Le résultat numérique ainsi que sa représentation géométrique sont donnés respectivement dans la table 3.10 et la figure 3.13.

TABLE 3.10 – Caractérisation de l'échantillon E_1 par la méthode de pondération de profils.

Coefficients de pondération	$\alpha_{P_c} = 0.23$ $\alpha_{P_{tr}} = 0.3$ $\alpha_{P_{ah}} = 0$ $\alpha_{P_{ab}} = 0.24$ $\alpha_{P_{st}} = 0.23$
Qualité de caractérisation	$E_{I_s, I_c} = 0.102$ $E_{geo} = 761.03 \text{ nm}^2$

On note que l'erreur E_{I_s, I_c} calculée à partir du profil reconstruit par pondération de profils est faible. Elle est du même ordre de grandeur que la meilleure erreur obtenue par une caractérisation de forme simple (résultat de RegPtr). Cependant, le profil estimé présente une erreur géométrique plus faible que celle obtenue par les autres caractérisations.

La forme reconstruite par pondération de profils est tracée dans la figure 3.13 et comparée à celle obtenue par la caractérisation AFM. On constate que les 2 profils sont proches. Il faut noter que l'AFM ne donne pas la description exacte du profil et présente également un certain nombre d'artefacts (paragraphe 1.2.2) notamment une tendance à augmenter la dimension critique de la forme. Cela pourrait expliquer une partie de cette différence entre les 2 profils.

FIGURE 3.13 – Représentation des profils obtenus par caractérisation de l'échantillon E_1 par la méthode de pondération de profils (en bleu) et par l'AFM (en rouge).

Afin de compléter cette étude, nous avons représenté la signature optique du profil reconstruit par rapport à celle mesurée (figure 3.14). Les 2 signatures optiques semblent relativement proches à l'exception de certaines plages dans l'UV évoqués précédemment.

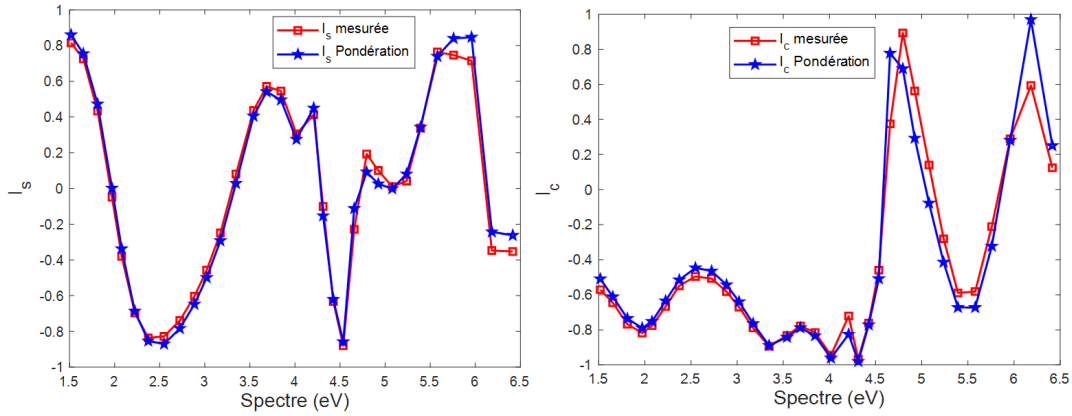


FIGURE 3.14 – Comparaison des signatures optiques calculées à partir des résultats obtenus par la méthode de pondération de profils comparées à celle mesurée (en rouge) de l'échantillon E_1 .

La forme reconstruite par pondération de profils présente des pentes et quelques arrondis en haut et bas. Au delà de la valeur du critère E_{I_s, I_c} , elle semble se rapprocher de la forme décrite par l'AFM. La méthode permet donc d'utiliser chaque caractéristique des profils simples pour proposer une structure présentant une forme plus conforme à l'échantillon.

3.2.4 Limite de la méthode

La méthode par pondération de profils décrite dans le paragraphe précédent n'a été testée jusqu'à présent que sur des profils qui ne s'éloignent pas trop de toutes les formes initiales simples. Dans ce paragraphe, nous allons tester cette méthode sur une forme plus éloignée des profils cibles. La forme géométrique simulée pour l'occasion (figure 3.15 en rouge) présente plusieurs défauts (présence de pentes, arrondis, ...) qui ont été choisis de telle sorte que les modèles de profils simples soient inadaptés. La signature optique est simulée dans les mêmes conditions que l'échantillon E_1 . Cette dernière est ensuite fournie aux différents modèles des 5 PMC de formes simples (RegPc, RegPtr, RegPah, RegPab et RegPst). Le résultat de la caractérisation est donné dans la table 3.11.

Les modèles de PMC définis estiment des paramètres géométriques inclus dans les domaines d'apprentissage excepté pour le RegPah (qui retourne un profil inexistant) et le RegPst (la valeur de la base b est hors plage). En effet, l'erreur E_{I_s, I_c} pour le résultat de RegPst est largement trop élevée; ce qui est confirmé par l'écart géométrique E_{geo} . Par conséquent, les informations fournies par le RegPah et le RegPst ne seront pas utilisées pour la méthode de pondération.

La figure 3.15 illustre les profils estimés par les différents PMC par rapport à la forme réelle. Ainsi, conformément à la valeur de E_{geo} de la table 3.11, les formes estimées sont assez

TABLE 3.11 – Résultat de caractérisation des différents PMC fonctionnant sur une forme simple concernant un échantillon de forme qualifiée d'éloignée des profils cibles.

PMC	Paramètres géométriques	Qualité de caractérisation
<i>RegPc</i>	$CD = 61.67 \pm 0.1 \text{ nm}$ $h = 118.67 \pm 0.1 \text{ nm}$	$E_{I_s, I_c} = 0.1163$ $E_{geo} = 666.95 \text{ nm}^2$
<i>RegPtr</i>	$b_1 = 17.29 \pm 0.1 \text{ nm}$ $b_2 = 48.21 \pm 0.2 \text{ nm}$ $h = 118.36 \pm 0.1 \text{ nm}$	$E_{I_s, I_c} = 0.0537$ $E_{geo} = 883.79 \text{ nm}^2$
<i>RegPah</i>	$\delta = 1.99 \pm 0.1$ $CD_h = 70.57 \pm 0.2 \text{ nm}$ $h = 131.88 \pm 0.3 \text{ nm}$	Profil inexistant
<i>RegPab</i>	$r_b = 45.38 \pm 0.3 \text{ nm}$ $CD_b = 55.48 \pm 0.2 \text{ nm}$ $h = 115 \pm 0.1 \text{ nm}$	$E_{I_s, I_c} = 0.0423$ $E_{geo} = 961.71 \text{ nm}^2$
<i>RegPst</i>	$b = 31.27 \pm 0.06 \text{ nm}$ $h = 107.06 \pm 0.04 \text{ nm}$	$E_{I_s, I_c} = 0.7766$ $E_{geo} = 4.87 \times 10^3 \text{ nm}^2$

éloignées.

La méthode par pondération de profils est alors testée sur cette forme supposée. Le résultat obtenu est donné dans la table 3.12.

TABLE 3.12 – Résultat de la méthode de pondération de profils d'une forme qualifiée d'éloignée des profils cibles.

Coefficients de pondération	$\alpha_{Pc} = 0.19$ $\alpha_{Ptr} = 0.36$ $\alpha_{Pah} = 0$ $\alpha_{Pab} = 0.45$ $\alpha_{Pst} = 0$
Qualité de caractérisation	$E_{I_s, I_c} = 0.1208$ $E_{geo} = 748.76 \text{ nm}^2$

Les résultats de l'erreur E_{I_s, I_c} et de l'écart géométrique E_{geo} ne présentent pas de valeurs significativement meilleures que certains PMCs de formes simples. Dans la figure 3.16, on trace le profil reconstruit à partir du résultat obtenu par pondération de profils par rapport à la forme réelle supposée.

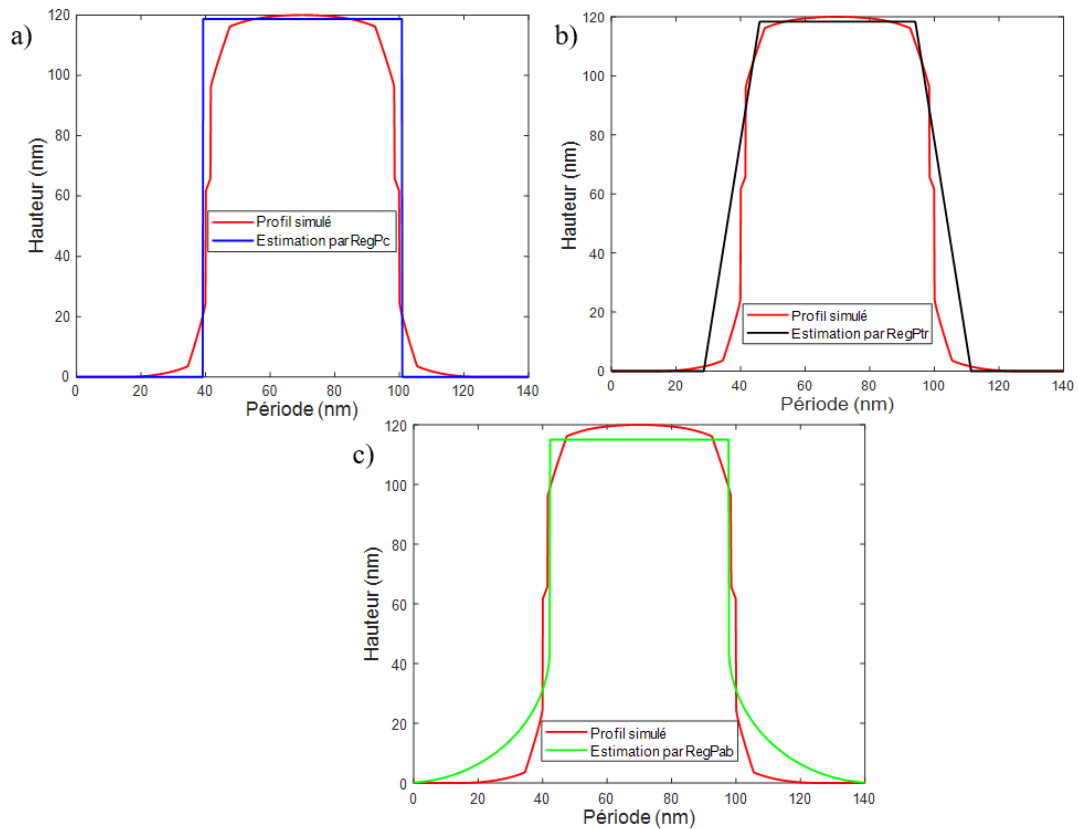


FIGURE 3.15 – Représentations des profils géométriques obtenus par différentes caractérisations d'une forme qualifiée d'éloignée des profils cibles (en rouge) : (a) RegPc (en bleu), (b) RegPtr (en noir) et (c) RegPab (en vert).

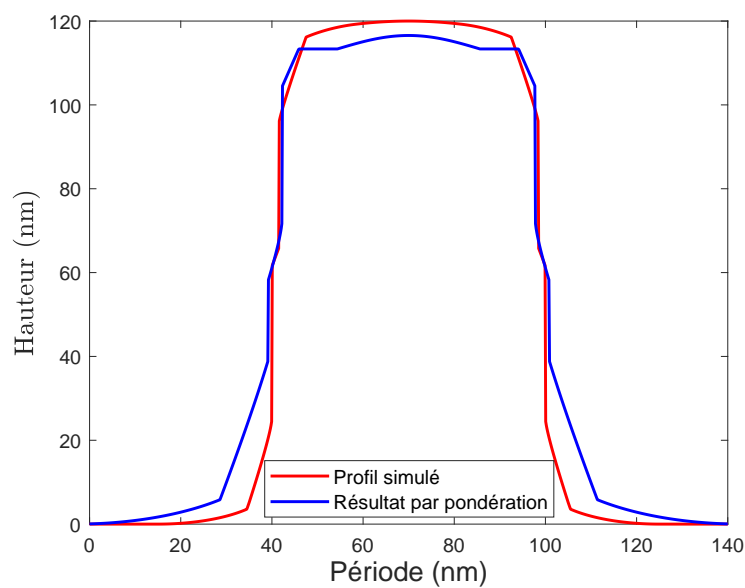


FIGURE 3.16 – Représentation de la forme obtenue par la méthode de pondération de profils (en bleu) d'un profil qualifié d'éloigné des profils cibles (en rouge).

La méthode par pondération de profils semble approcher la modélisation des défauts (pentes et arrondis) présents sur la forme réelle. Elle n'a pas permis d'améliorer significativement les critères de performances obtenus par une caractérisation sur une forme simple. Cependant, elle semble plus robuste quant à la présence d'un problème mal posé. En effet, dans l'étude présentée, le RegPab présente clairement une bonne performance sur le critère E_{I_s, I_c} mais un résultat qui ne correspond pas géométriquement à celui attendu (figure 3.15 (c) et valeur de E_{geo}).

Ce résultat est confirmé par la variation de la signature optique simulée à partir des résultats obtenus par pondération de profils par rapport à celle issue de la forme réelle supposée (figure 3.17). On observe clairement un écart entre les 2 signatures.

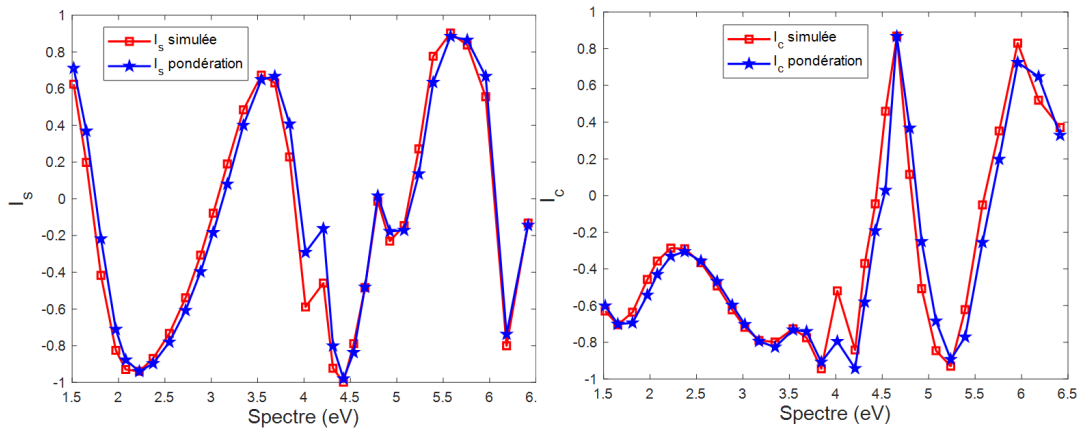


FIGURE 3.17 – Comparaison des signatures optiques simulées (en bleu) à partir des résultats par pondération de profils comparées à celles simulées (en rouge) d'un profil qualifié d'éloigné des profils cibles.

La méthode par pondération de profils n'est composée que de formes simples qui ne nécessitent pas la mise en œuvre d'un R.N supplémentaire. C'est un avantage pour cette méthode qui peut systématiquement être utilisée en complément dans tout processus de scattérométrie. Lorsque les conditions sont réunies, elle conduit à l'amélioration des résultats. Nous avons également démontré qu'elle était plus robuste que les méthodes classiques.

La méthode par pondération de profils est intrinsèquement conditionnée par les résultats des caractérisations simples. Elle donne également une importance prépondérante au critère d'erreur E_{I_s, I_c} qui peut conduire, comme on l'a dit précédemment, à une mauvaise interprétation en particulier dans le cas d'un problème mal posé.

Ainsi, la méthode par pondération de profils est, certes, plus flexible qu'une caractérisation classique sur une forme simple, mais reste tout de même limitée à de petites variations autour des profils géométriques simples composants sa base de départ. Cette approche devient alors inadaptée. Il faut avoir recours à une autre méthode : la classification.

3.3 Classification de profils

Dans cette partie, on s'intéresse à l'identification de la forme géométrique de l'échantillon pour optimiser les performances de tous processus scattérométriques. En effet, l'étude consiste à mettre en place un classifieur neuronal en amont de l'étape de caractérisation proprement dite. Ainsi, le but du classifieur est de certifier le modèle géométrique sur lequel va fonctionner la résolution du problème inverse.

Dans le paragraphe 3.2.2 (table 3.5), nous avons vu que les différents PMC régresseurs donnent des résultats de caractérisation avec une erreur E_{I_s, I_c} relativement faible (table 3.11). Par contre, cette faible valeur d'erreur E_{I_s, I_c} ne conduit pas forcément à un écart géométrique E_{geo} faible. En effet, les différents PMC affichent des profils plus ou moins éloignés de la forme cible ; ce qui conduit à une caractérisation erronée. Nous avons ainsi mis en évidence le fait que la valeur d'erreur E_{I_s, I_c} n'est clairement pas suffisante pour valider un résultat. Par ailleurs, elle ne donne pas d'information sur la pertinence de la forme choisie. Ainsi, il est donc nécessaire d'obtenir d'autres informations permettant d'éviter de se placer dans des conditions de problèmes mal posés où l'interprétation est difficile. Dans ce cadre, la signature optique utilisée classiquement présente, comme nous l'avons dit dans le paragraphe précédent, certaines zones d'intérêt qui peuvent être plus efficacement exploitées (figure 3.14).

Une solution impliquant l'implémentation d'un classifieur est proposée afin de fournir des informations supplémentaires pour lever cette ambiguïté. Ainsi, un nouveau PMC fonctionnant indépendamment du précédent est envisagé pour identifier le profil géométrique.

3.3.1 Détection de la forme cible

Dans ce paragraphe, nous implémentons un classifieur neuronal pour détecter la forme géométrique de la structure à tester. Dans ce cas, chaque forme géométrique correspond à une classe. Ainsi, le système classifieur créé apprend lors de l'entraînement la relation liant les signatures optiques (entrées) aux différentes classes (sorties). Dans toute cette étude, nous avons volontairement choisi la même signature optique que celle utilisée lors de la phase de caractérisation proprement dite. Ceci dans le but de ne pas accroître le temps de caractérisation global. Cependant, une étude spécifique peut être envisagée pour sélectionner parmi les mesures celles présentant une information plus pertinente. Par ailleurs, le PMC classifieur n'attribue pas directement la classe de l'échantillon mais estime une probabilité d'appartenance à une classe. C'est donc à l'expérimentateur de fixer un seuil d'acceptabilité pour attribuer la classe de ce dernier. Deux approches peuvent ainsi être envisagées suivant l'application visée : le classifieur tout ou rien ou le détecteur de profil.

3.3.1.1 R.N classifieur tout ou rien

Le PMC classifieur tout ou rien fonctionne sur seulement deux classes C_1 et C_2 (figure 3.18). Il permet de détecter un profil géométrique particulier correspondant par exemple à la classe C_1 à partir d'une signature optique mesurée.

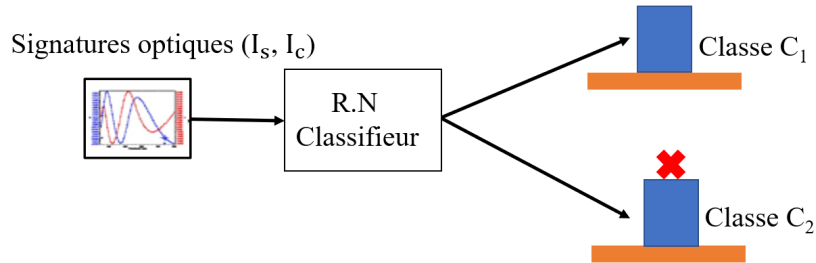


FIGURE 3.18 – Schéma de principe d'un système de vérification d'une forme de profil géométrique donnée par l'utilisation d'un PMC en mode classifieur fonctionnant sur deux classes.

Pour un échantillon quelconque donné, 4 cas peuvent se produire :

- Vrai positif (TP pour True Positive) : l'échantillon appartient à la classe C_1 et l'estimation du classifieur est correcte.
- Faux négatif (FN pour False Negative) : l'échantillon appartient à la classe C_1 et le classifieur identifie la classe C_2 .
- Faux positif (FP pour False Positive) : l'échantillon appartient à la classe C_2 et l'estimation du classifieur est erronée en assignant l'échantillon à la classe C_1 .
- Vrai négatif (TN pour True Negative) : l'échantillon appartient à la classe C_2 et le classifieur ne fait pas d'erreur.

Ces résultats sont généralement rassemblés dans une matrice regroupant le nombre d'échantillons répondant à chacun de ces cas de figure sur une ensemble de tests donné. On parle de matrice de confusion (table 3.13). Plus la matrice est diagonale, plus le classifieur est performant.

TABLE 3.13 – Matrice de confusion résumant les résultats d'un classifieur.

		Classe réelle	
		C_1	C_2
Classe estimée	C_1	<i>TP</i>	<i>FP</i>
	C_2	<i>FN</i>	<i>TN</i>

Trois critères sont utilisés pour mesurer la performance :

- La capacité du classifieur à identifier correctement la classe C_1 :

$$\text{Précision} = \frac{TP}{TP + FP} \quad (3.8)$$

- La capacité à effectuer une classification correcte pour tous les échantillons de la classe C_1 :

$$Sensibilité = \frac{TP}{TP + FN} \quad (3.9)$$

- Le taux global de classification correcte :

$$Exactitude = \frac{TP + TN}{TP + TN + FP + FN} \quad (3.10)$$

Le classifieur neuronal peut ainsi être facilement appliqué à condition de définir clairement en amont lors de la phase d'apprentissage 2 classes distinctes : l'une correspondant à la forme cible et l'autre regroupant les formes à exclure.

3.3.1.2 Cas du réseau de forme rectangulaire

Considérons l'ensemble des modèles géométriques définis précédemment (modèle type des échantillons E_2 à E_4). La classe C_1 est identifiée comme représentant les profils rectangulaires P_c à détecter. La classe C_2 contient alors toutes les autres formes (trapézoïdales P_{tr} , rectangulaires arrondis au sommet P_{ah} et rectangulaires arrondis en bas P_{ab}).

Pour constituer le corpus d'apprentissage du PMC classifieur, un tirage aléatoire est ainsi réalisé dans les domaines de variations des paramètres définis ci dessous :

- Modèle P_c

$$\begin{aligned} 150 < CD < 250 & \quad nm \\ 250 < h < 420 & \quad nm \end{aligned} \quad (3.11)$$

- Modèle P_{tr}

$$\begin{aligned} 5 < b_1 < 30 & \quad nm \\ 130 < b_2 < 220 & \quad nm \\ 250 < h < 420 & \quad nm \end{aligned} \quad (3.12)$$

- Modèle P_{ah}

$$\begin{aligned} 0.1 < \delta < 1 \\ 120 < CD_h < 250 & \quad nm \\ 250 < h < 420 & \quad nm \end{aligned} \quad (3.13)$$

- Modèle P_{ab}

$$\begin{aligned} 10 < r_b < 30 & \text{ nm} \\ 150 < CD_b < 250 & \text{ nm} \\ 250 < h < 420 & \text{ nm} \end{aligned} \quad (3.14)$$

Les signatures optiques (I_s et I_c) sont définies sur une gamme spectrale allant de 0.7 à 4.4 eV. L'angle d'incidence utilisé pour la mesure est de $\theta = 70^\circ$.

3.3.1.2.1 Performance après apprentissage

Un PMC classifieur est implémenté pour la détection du profil rectangulaire et nommé par la suite TOR. Il est entraîné sur 6000 échantillons et contient 20 neurones dans la couche cachée. La performance de ce PMC classifieur pour un seuil d'acceptabilité de 50% est donnée dans la table 3.14.

TABLE 3.14 – Matrice de confusion du TOR (20 neurones cachés) calculée sur le corpus test pour la détection de profils géométriques rectangulaires.

		Classe réelle		<i>Précision</i>
		C_1	C_2	
Classe estimée	C_1	450	23	95.1%
	C_2	0	427	100%
<i>Sensibilité</i>		100%	94.9%	<i>Exactitude</i> : 97.4%

Le TOR montre une excellente performance globale (97.4%). La précision pour la classe C_2 est de 100%; ce qui signifie que lorsque le classifieur rejette le profil, la probabilité que cela soit vrai est maximale. De la même manière, la sensibilité pour la classe C_1 est maximale assurant ainsi un classement parfait pour les échantillons correspondant à la forme cible.

Par contre, 23 profils non-rectangulaires sont mal classés (faux positifs ou FP); ce qui représente 2.5% de l'ensemble du corpus de test. Il est intéressant de s'attarder sur ce lot d'échantillons théoriques. Notons dans un premier temps que tous ces profils sont exclusivement de la même forme. Ce résultat met en évidence une sensibilité significative de la classe C_1 aux profils dont le sommet est faiblement arrondi. Ensuite, la figure 3.19 montre la valeur de chaque sortie brute non seuillée s_1 en fonction de l'écart géométrique (E_{geo}) au modèle rectangulaire le plus proche. Les écarts géométriques correspondant sont relativement faibles voire très faibles. La plus grande déviation géométrique calculée est de 504.92 nm^2 soit 0.55% de la surface totale du réseau. Par conséquent, tous les cas considérés dans la figure 3.19 sont très proches d'un profil rectangulaire. Par ailleurs, on note une corrélation significative entre la sortie s_1 et l'écart géométrique : plus la sortie est élevée plus on se rapproche du profil cible.

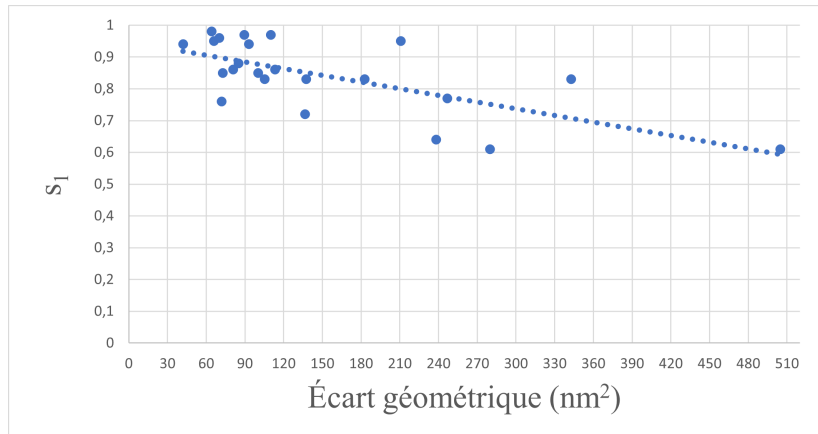


FIGURE 3.19 – Sortie brute s_1 du TOR pour les échantillons mal classés (FP) en fonction de l'écart géométrique au profil rectangulaire le plus proche.

La définition du seuil d'acceptabilité est également primordiale quant à la décision finale. Pour s'en persuader, il suffit d'observer les échantillons faussement acceptés (table 3.14 et figure 3.19) par la règle automatique fixée en amont. Il est donc important, avant de construire un tel système, de se poser la question du seuil d'acceptabilité d'une forme donnée. Ainsi, pour compléter cette étude, une nouvelle classe est créée dans la matrice de confusion. L'idée est d'écartier un échantillon dont aucune des sorties du PMC classifieur n'atteint la valeur seuil initialement fixée. Cette classe est nommée « indéterminée (C.I) ».

Par exemple, un seuillage à 85% permettrait de réduire de moitié le nombre de profils mal classés. La nouvelle matrice de confusion serait alors (table 3.15) :

TABLE 3.15 – Matrice de confusion du TOR évaluée sur les données du corpus test pour un seuil d'acceptabilité fixé à 85%.

		Classe réelle		Précision
		C_1	C_2	
Classe estimée	C_1	421	13	97%
	C_2	0	411	100%
	$C.I$	29	26	
Sensibilité		93.5%	91.3%	Exactitude : 92.4%

Le résultat montre que la performance globale du TOR a diminué, lorsque le seuil d'acceptabilité augmente. On note notamment une dégradation de la sensibilité sur les 2 sorties provoquée par un nombre élevé des échantillons affectés à la classe C.I. Cependant la réduction du nombre de FP a ainsi permis d'augmenter la précision sur la sortie C_1 du classifieur donnant ainsi davantage de certitude à l'expérimentateur quant à la détection automatique de la forme cible. La plus grande déviation géométrique calculée pour un profil faux positif chute alors à 0.22% de la surface totale du réseau.

3.3.1.2.2 Validation expérimentale

Les signatures optiques mesurées des échantillons E_2 et E_3 (figure 3.2) sont introduites à l'entrée du TOR décrit précédemment pour la détection d'une forme rectangulaire. Les résultats bruts avant seuillage sont affichés dans la table 3.16. On observe que les deux échantillons sont clairement identifiés comme appartenant à la classe C_2 donc non conforme à la cible.

TABLE 3.16 – Détection par le TOR de profils rectangulaires sur les échantillons E_2 et E_3 . La sortie s_1 correspond à la classe C_1 (profils cibles rectangulaires) et la sortie s_2 à la classe C_2 (autres profils).

	s_1	s_2
E_2	0.00	1.00
E_3	0.00	1.00

Ce résultat sans équivoque est confirmé par les images MEB de la figure 3.2.

3.3.1.3 Cas du réseau de forme sinusoïdale

Considérons à présent les réseaux E_5 à E_{10} présentés dans le paragraphe 3.1.2. Nous avons mis en œuvre un PMC classifieur tout ou rien pour détecter cette fois ci un profil géométrique sinusoïdal P_s . L'épaisseur de résine h_r sous la structure périodique n'étant pas connue, il est impératif de considérer plusieurs épaisseurs dans le processus ; ce qui constitue une difficulté supplémentaire. Les signatures optiques utilisées sont définies sur une gamme spectrale allant de 0.7 à 4.4 eV. L'angle d'incidence est de 70° .

Pour cette étude, nous avons considéré 3 autres profils supposés correspondre à des défauts de fabrication connus ou à des dérives potentielles de la forme cible : le profil gaussien noté P_g , le profil sinus déformé P_{sd} et P_{st} pour sinus tronqué. Nous définissons un domaine de variation pour chaque paramètre géométrique de chaque modèle en prenant soin de considérer des structures qui englobent les principales déviations de chaque profil sinusoïdal.

- Modèle P_s

$$\begin{aligned} 100 < h < 200 & \quad nm \\ 300 < h_r < 400 & \quad nm \end{aligned} \tag{3.15}$$

- Modèle P_g

$$\begin{aligned} 40 < \sigma < 400 \\ 100 < h < 200 & \quad nm \\ 300 < h_r < 400 & \quad nm \end{aligned} \tag{3.16}$$

- Modèle P_{sd}

$$\begin{aligned} -0.25 < \rho < 2 \\ 100 < h < 200 \quad nm \\ 300 < h_r < 400 \quad nm \end{aligned} \quad (3.17)$$

- Modèle P_{st}

$$\begin{aligned} 0 < \alpha < 1 \\ 100 < b < 500 \quad nm \\ 100 < h < 200 \quad nm \\ 300 < h_r < 400 \quad nm \end{aligned} \quad (3.18)$$

Pour constituer différents corpus d'apprentissage du PMC classifieur, nous avons réalisé un tirage aléatoire dans les domaines de variations des paramètres définis ci dessus tout en calculant systématiquement pour chaque profil l'écart géométrique au profil sinus parfait le plus proche. Dans la suite, nous avons donc construit 3 corpus composés de profils présentant systématiquement une valeur de $E_{geonorm}$ telle que : $E_{geonorm} \geq 1\%$, $E_{geonorm} \geq 5\%$ et $E_{geonorm} \geq 15\%$ afin de créer artificiellement une frontière plus ou moins marquée entre les classes.

3.3.1.3.1 Performance après apprentissage

Les PMC classifieurs sont entraînés sur les 3 types de corpus et sont notés respectivement TOR_a (20 neurones cachés), TOR_b (25 neurones cachés) et TOR_c (30 neurones cachés). Chacun d'eux a nécessité 6000 signatures optiques pour la phase d'apprentissage dont 3000 pour le profil sinusoïdal (classe P_s) et 3000 pour les faux profils sinus (classe P_{fs}) composés de 1000 échantillons de chacun des 3 profils équitablement répartis entre P_g , P_{sd} et P_{st} . Les résultats des 3 TORs sont affichés sous la forme de matrices de confusion dans les tables 3.17 à 3.19.

TABLE 3.17 – Matrice de confusion du TOR_a (corpus assurant $E_{geonorm} \geq 1\%$) axée sur la détection d'un profil sinusoïdal.

		Classe réelle		Précision
		P_s	P_{fs}	
Classe estimée	P_s	450	19	95.9%
	P_{fs}	0	431	100%
Sensibilité		100%	95.8%	Exactitude : 97.9%

Comme on pouvait s'y attendre les performances brutes sont d'autant meilleures que les frontières entre classes sont bien marquées dans la base d'apprentissage. En effet, le TOR_c affiche des performances optimales pour la détection de tout échantillon présentant un écart

TABLE 3.18 – Matrice de confusion du TOR_b (corpus assurant $E_{geonorm} \geq 5\%$) axée sur la détection d'un profil sinusoïdal.

		Classe réelle		Précision
		P_s	P_{fs}	
Classe estimée	P_s	450	8	98.3%
	P_{fs}	0	442	100%
Sensibilité		100%	98.2%	Exactitude : 99.1%

TABLE 3.19 – Matrice de confusion du TOR_c (corpus assurant $E_{geonorm} \geq 15\%$) axée sur la détection d'un profil sinusoïdal.

		Classe réelle		Précision
		P_s	P_{fs}	
Classe estimée	P_s	450	0	100%
	P_{fs}	0	450	100%
Sensibilité		100%	100%	Exactitude : 100%

géométrique minimum à la cible de 15%. Dans les 2 autres cas de figure, si l'exactitude n'est pas de 100%, on peut tout de même assurer que tous les profils sinus exacts seront détectés (sensibilité P_s de 100%) et qu'un profil rejeté par le système présente réellement une déformation par rapport à la cible (précision de 100% pour la sortie P_{fs}). Bien entendu, ces performances sont valables dans les conditions imposées lors de la phase d'apprentissage ; à savoir $E_{geonorm} \geq 1\%$, $E_{geonorm} \geq 5\%$ et $E_{geonorm} \geq 15\%$. Il est donc difficile de les comparer entre eux.

Ainsi, nous avons testé les performances de ces 3 TORs sur le même corpus test ($E_{geonorm} \geq 1\%$). Les résultats obtenus sont donnés dans les différentes matrices de confusions affichées dans les tables 3.17, 3.20 et 3.21 :

TABLE 3.20 – Matrice de confusion du TOR_b évaluée sur les données test du corpus assurant $E_{geonorm} \geq 1\%$.

		Classe réelle		Précision
		P_s	P_{fs}	
Classe estimée	P_s	450	23	95.1%
	P_{fs}	0	427	100%
Sensibilité		100%	94.8%	Exactitude : 97.4%

TABLE 3.21 – Matrice de confusion du TOR_c évaluée sur les données test du corpus assurant $E_{geomorm} \geq 1\%$.

		Classe réelle		Précision
		P_s	P_{fs}	
Classe estimée	P_s	450	33	93.1%
	P_{fs}	0	417	100%
Sensibilité		100%	92.6%	Exactitude : 96.3%

On retrouve le fait que tous les profils sinus sont correctement classés avec une sensibilité maximale pour le profil P_s . La précision sur la seconde sortie reste également maximale quelque soit le TOR. Les PMC classifieurs TOR_b et TOR_c affichent respectivement des performances globales de 97.4% et 96.3% contre 97.9% pour le TOR_a sur le même corpus test. Le nombre de FP passe de 19 pour le TOR_a à 33 pour le TOR_c . Ainsi, il semble primordial d'apporter d'avantage d'information sur les frontières entre les classes lors de la phase d'apprentissage si on veut une détection optimale.

Prenons un seuil d'acceptabilité respectivement de 70% et 90% pour affecter un échantillon dans une classe. Les tables 3.22 et 3.23 présentent les résultats obtenus dans le cadre de l'étude précédente pour le TOR_a . Les performances globales du TOR_a diminuent légèrement, lorsque le seuil d'acceptabilité augmente. En effet, une dégradation de la sensibilité aux faux profils sinusoidaux est provoquée par un nombre élevé des échantillons affectés à la classe C.I. Cependant le nombre de FP diminue conduisant ainsi à une augmentation de la précision sur la première sortie du classifieur. La confiance en cette dernière est primordiale en pratique dans le cas d'un contrôle en cours de processus de fabrication. Dans ce cadre, nous avons étudié dans le paragraphe suivant l'évolution de la sortie P_s en simulant une déformation continue sur un échantillon de forme sinusoidale.

TABLE 3.22 – Matrice de confusion du TOR_a évaluée sur les données test du corpus assurant $E_{geomorm} \geq 1\%$ pour un seuil d'acceptabilité fixé à 70%.

		Classe réelle		Précision
		P_s	P_{fs}	
Classe estimée	P_s	450	15	96.7%
	P_{fs}	0	426	100%
	C.I	0	9	
Sensibilité		100%	94.6%	Exactitude : 97.3%

TABLE 3.23 – Matrice de confusion du TOR_a évaluée sur les données test du corpus assurant $E_{geomorm} \geq 1\%$ pour un seuil d'acceptabilité fixé à 90%.

		Classe réelle		Précision
		P_s	P_{fs}	
Classe estimée	P_s	426	10	97.7%
	P_{fs}	0	420	100%
C.I		24	20	
Sensibilité		94.6%	93.3%	Exactitude : 94.0%

3.3.1.3.2 Performance sur la déformation des faux profils sinus

Dans cette partie, on étudie spécifiquement la déformation continue d'une forme géométrique donnée. Nous avons ainsi pris différents faux profils sinusoidaux, que nous avons déformés en faisant varier le ou les paramètres géométriques correspondants occasionnant ainsi différentes formes plus ou moins proches d'un profil sinus parfait. L'étude est faite sur une forme de profil présentée comme défaut du profil cible.

Prenons par exemple le profil gaussien, on fixe la hauteur $h = 130 \text{ nm}$ et l'épaisseur de la couche résiduelle à $h_r = 250 \text{ nm}$. Plusieurs structures sont simulées en faisant varier uniquement le paramètre σ . Les signatures optiques correspondantes sont fournies aux différents TORs construits dans le paragraphe précédent.

L'évolution de la sortie P_s de chaque TOR en fonction de σ est illustrée dans la figure 3.20. Cette dernière est affichée sous sa forme brute avant seuillage ; à savoir sous la forme d'une probabilité d'appartenance à la classe P_s . L'écart surfacique normalisé ($E_{geomorm}$) calculé entre le profil sinus parfait le plus proche et le profil gaussien défini par σ est également représenté sur le même graphique.

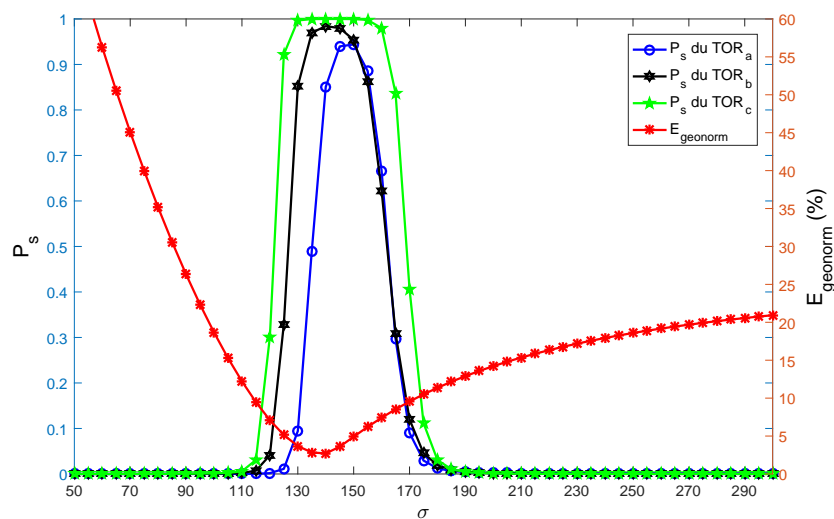


FIGURE 3.20 – Évolution de la sortie brute P_s des différents TOR et de l'écart géométrique normalisé ($E_{geomorm}$) en fonction du paramètre σ du profil gaussien.

Notons dans un premier temps que les déformations occasionnées n'engendrent jamais un profil sinusoidal parfait. Le profil le plus proche est obtenu pour $\sigma = 140$ et qui correspond à un écart géométrique normalisé de 2.7%. Les courbes de chaque sortie présentent un pic de forme non symétrique autour de cette valeur de σ . En effet, l'écart géométrique n'a pas la même incidence suivant le fait que le profil estimé soit globalement réduit ou amplifié par rapport à la forme sinusoidale (courbe en rouge sur la figure).

Pour le TOR_a et TOR_b , la sortie P_s n'atteint pas le maximum, mais en fixant le seuil d'acceptabilité jusqu'à 90%, ils le détectent comme un profil sinusoidal parfait. En l'absence de précision sur les frontières, le TOR_c confirme la détection d'une forme sinusoidale correspondant à la définition qu'on lui a donnée ($E_{geonorm} < 15\%$).

La figure 3.21 représente 2 profils gaussiens présentant le même écart géométrique au sinus parfait ($E_{geonorm} = 16\%$) tout en étant différemment éloignés de la valeur optimale de σ ($\sigma = 110$ et $\sigma = 220$). Le caractère non symétrique de la courbe d'erreur (en rouge) se reproduit sur les réponses des différents TORs.

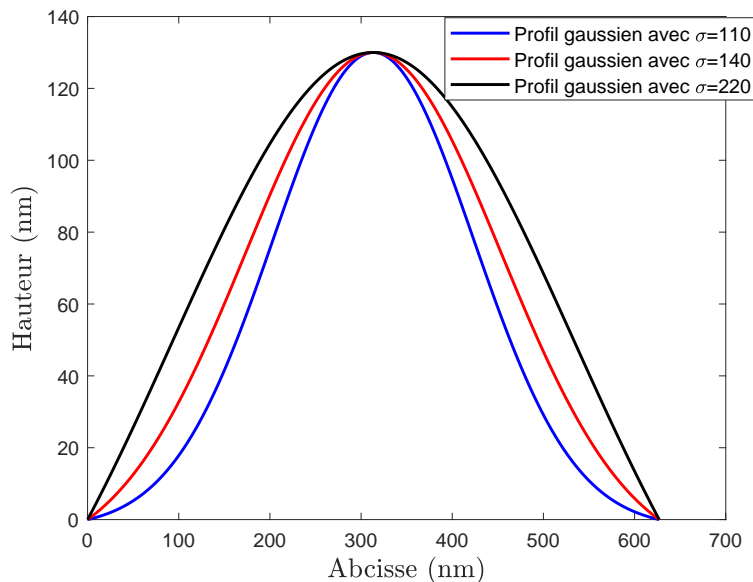


FIGURE 3.21 – Représentation des profils gaussiens ayant une hauteur $h = 130$ nm pour différentes valeurs de $\sigma = 110$ (en bleu) et $\sigma = 220$ (en noir) présentant le même écart géométrique à la forme cible définie $\sigma = 140$ (en rouge).

La table 3.24 regroupe pour chaque seuil d'acceptabilité les 2 valeurs limites de σ permettant une détermination automatique de la forme cible au seuil fixé. On constate que la valeur de $\Delta\sigma$ diminue lorsque le seuil d'acceptabilité augmente. Par ailleurs, la largeur de la détection $\Delta\sigma$ varie en fonction du type de TOR utilisé. Ainsi les courbes ont une largeur de pic plus ou moins étroite suivant le TOR. En effet, le TOR_c a une tolérance au profil sinusoidal parfait plus importante ; ce qui se caractérise par une détection plus étalée sur les valeurs de σ .

TABLE 3.24 – Valeurs limites de σ caractérisant un profil détecté comme forme cible par chacun des PMC classificateurs (TOR_a , TOR_b et TOR_c) pour chaque seuil d'acceptabilité (50%, 70% et 90%).

	Seuil d'acceptabilité		
	50%	70%	90%
TOR_a	$\sigma = 135 \Rightarrow \Delta\sigma = 27.5$ $\sigma = 162.5$	$\sigma = 137.75 \Rightarrow \Delta\sigma = 21.25$ $\sigma = 159$	$\sigma = 142.5 \Rightarrow \Delta\sigma = 11$ $\sigma = 153.5$
TOR_b	$\sigma = 126 \Rightarrow \Delta\sigma = 36$ $\sigma = 162$	$\sigma = 128.5 \Rightarrow \Delta\sigma = 29.5$ $\sigma = 158$	$\sigma = 132 \Rightarrow \Delta\sigma = 21$ $\sigma = 153$
TOR_c	$\sigma = 121 \Rightarrow \Delta\sigma = 47$ $\sigma = 168$	$\sigma = 123 \Rightarrow \Delta\sigma = 44$ $\sigma = 167$	$\sigma = 125 \Rightarrow \Delta\sigma = 37.75$ $\sigma = 162.75$

3.3.1.3.3 Résultats expérimentaux

L'étude expérimentale qui suit est effectuée sur les échantillons E_5 à E_{10} de type sinusoidal (figure 3.3) définis dans le paragraphe 3.1.2.

La figure 3.3 montre les profils locaux obtenus par AFM. Afin de déterminer si ces structures correspondent aux profils cibles, il est nécessaire de traiter ces mesures brutes. Nous allons donc rechercher par optimisation les paramètres des modèles géométriques les plus proches de chaque échantillon à partir de la mesure AFM. Les paramètres géométriques correspondant à chaque type de profil sont déterminés en recherchant l'écart géométrique le plus faible entre les 2 profils. Ensuite, une fois le profil obtenu, l'épaisseur de la couche résiduelle h_r est calculée en minimisant l'erreur E_{I_s, I_c} entre la signature optique mesurée et celle simulée à partir du résultat de l'optimisation précédente. Le résultat est présenté dans les tables 3.25 et 3.26.

Il faut noter que la majorité des profils géométriques identifiés par optimisation sur l'écart géométrique correspondent également à une erreur E_{I_s, I_c} minimale (en gras). Il ressort que l'optimisation identifie les échantillon E_6 et E_{10} à la forme gaussienne, l'échantillon E_8 au profil sinus déformé et les échantillons E_7 et E_9 au profil sinus tronqué. On constate cependant que l'affectation de l'échantillon E_5 n'est pas clairement identifiée. Le profil présentant le plus faible écart géométrique brut (P_g et P_{sd}) n'est pas celui qui a le plus faible écart géométrique normalisé (P_s). Étant donnée la forte valeur de l'écart géométrique normalisé, on peut donc raisonnablement penser que la forme réelle ne peut être représentée par l'un de ces modèles géométriques. Dans tous les cas la valeur de h_r semble quasiment identique pour toutes les études ; ce qui est conforme aux conditions expérimentales utilisées pour la fabrication de ces échantillons.

L'erreur E_{I_s, I_c} est relativement faible pour les échantillons E_5 , E_6 , E_7 et E_8 alors qu'elle est plus élevée pour les échantillons E_9 et E_{10} . Les résultats concernant l'affectation d'un profil pour ces 2 derniers échantillons sont donc à prendre avec précaution.

TABLE 3.25 – Détermination des types de profils des échantillons E_5 , E_6 et E_7 par traitement de la mesure AFM.

	P_s	P_g	P_{sd}	P_{st}
E_5	$h = 47.97 \text{ nm}$ $h_r = 537.66 \text{ nm}$ $E_{I_s, I_c} = 0.045$ $E_{geo} = 4.99 \times 10^3 \text{ nm}^2$ $\mathbf{E_{geonorm} = 33.22\%}$	$\sigma = 113.32$ $h = 52.67 \text{ nm}$ $h_r = 538.09 \text{ nm}$ $\mathbf{E_{I_s, I_c} = 0.0414}$ $\mathbf{E_{geo} = 4.95 \times 10^3 \text{ nm}^2}$ $E_{geonorm} = 34.23\%$	$\rho = 0.12$ $h = 52.38 \text{ nm}$ $h_r = 538.06 \text{ nm}$ $E_{I_s, I_c} = 0.0415$ $\mathbf{E_{geo} = 4.95 \times 10^3 \text{ nm}^2}$ $E_{geonorm} = 34.52\%$	$\alpha = 0$ $b = 439.77 \text{ nm}$ $h = 49.54 \text{ nm}$ $h_r = 538.78 \text{ nm}$ $E_{I_s, I_c} = 0.0425$ $E_{geo} = 5.28 \times 10^3 \text{ nm}^2$ $E_{geonorm} = 38.12\%$
E_6	$h = 104.5 \text{ nm}$ $h_r = 471.28 \text{ nm}$ $E_{I_s, I_c} = 0.0588$ $E_{geo} = 4.39 \times 10^3 \text{ nm}^2$ $E_{geonorm} = 13.4\%$	$\sigma = 181$ $h = 98.44 \text{ nm}$ $h_r = 470.48 \text{ nm}$ $\mathbf{E_{I_s, I_c} = 0.0478}$ $\mathbf{E_{geo} = 2.66 \times 10^3 \text{ nm}^2}$ $\mathbf{E_{geonorm} = 7.61\%}$	$\rho = -0.12$ $h = 96.49 \text{ nm}$ $h_r = 471.12 \text{ nm}$ $E_{I_s, I_c} = 0.0489$ $E_{geo} = 3.47 \times 10^3 \text{ nm}^2$ $E_{geonorm} = 10.2\%$	$\alpha = 0.99$ $b = 500 \text{ nm}$ $h = 97.98 \text{ nm}$ $h_r = 471.7 \text{ nm}$ $E_{I_s, I_c} = 0.056$ $E_{geo} = 4.51 \times 10^3 \text{ nm}^2$ $E_{geonorm} = 13.82\%$
E_7	$h = 144.86 \text{ nm}$ $h_r = 410.8 \text{ nm}$ $E_{I_s, I_c} = 0.0669$ $E_{geo} = 6.03 \times 10^3 \text{ nm}^2$ $E_{geonorm} = 13.29\%$	$\sigma = 162.82$ $h = 139.2 \text{ nm}$ $h_r = 411.07 \text{ nm}$ $E_{I_s, I_c} = 0.0857$ $E_{geo} = 6.15 \times 10^3 \text{ nm}^2$ $E_{geonorm} = 12.95\%$	$\rho = -0.14$ $h = 132.06 \text{ nm}$ $h_r = 411.51 \text{ nm}$ $E_{I_s, I_c} = 0.0781$ $E_{geo} = 3.75 \times 10^3 \text{ nm}^2$ $E_{geonorm} = 7.91\%$	$\alpha = 0.99$ $b = 500 \text{ nm}$ $h = 136.8 \text{ nm}$ $h_r = 411.48 \text{ nm}$ $\mathbf{E_{I_s, I_c} = 0.0598}$ $\mathbf{E_{geo} = 3.53 \times 10^3 \text{ nm}^2}$ $\mathbf{E_{geonorm} = 7.74\%}$

Maintenant que ces différents échantillons ont été affectés par l'AFM à un profil donné, nous allons pouvoir valider expérimentalement la méthode de détection à partir des signatures optiques mesurées sur ces échantillons et introduites à l'entrée de chaque TOR pour l'identification.

TABLE 3.26 – Détermination des types de profils des échantillons E_8 , E_9 et E_{10} par traitement de la mesure AFM.

	P_s	P_g	P_{sd}	P_{st}
E_8	$h = 157.67 \text{ nm}$ $h_r = 418.56 \text{ nm}$ $E_{I_s, I_c} = 0.0813$ $E_{geo} = 9.3 \times 10^3 \text{ nm}^2$ $E_{geonorm} = 18.82\%$	$\sigma = 223.25$ $h = 143.73 \text{ nm}$ $h_r = 418.45 \text{ nm}$ $E_{I_s, I_c} = 0.0516$ $E_{geo} = 5.31 \times 10^3 \text{ nm}^2$ $E_{geonorm} = 9.83\%$	$\rho = -0.25$ $h = 135.25 \text{ nm}$ $h_r = 418.73 \text{ nm}$ $\mathbf{E}_{I_s, I_c} = \mathbf{0.0461}$ $\mathbf{E}_{geo} = \mathbf{4.35 \times 10^3 \text{ nm}^2}$ $\mathbf{E}_{geonorm} = \mathbf{8.22\%}$	$\alpha = 0.99$ $b = 500 \text{ nm}$ $h = 149.03 \text{ nm}$ $h_r = 419.02 \text{ nm}$ $E_{I_s, I_c} = 0.0665$ $E_{geo} = 7.36 \times 10^3 \text{ nm}^2$ $E_{geonorm} = 14.85\%$
E_9	$h = 165.01 \text{ nm}$ $h_r = 392.47 \text{ nm}$ $E_{I_s, I_c} = 0.1265$ $E_{geo} = 7.66 \times 10^3 \text{ nm}^2$ $E_{geonorm} = 14.82\%$	$\sigma = 136.97$ $h = 166.78 \text{ nm}$ $h_r = 392.36 \text{ nm}$ $E_{I_s, I_c} = 0.1364$ $E_{geo} = 8.76 \times 10^3 \text{ nm}^2$ $E_{geonorm} = 16.79\%$	$\rho = -0.02$ $h = 162.1 \text{ nm}$ $h_r = 393.14 \text{ nm}$ $E_{I_s, I_c} = 0.1329$ $E_{geo} = 7.4 \times 10^3 \text{ nm}^2$ $E_{geonorm} = 14.18\%$	$\alpha = 0.99$ $b = 462.08 \text{ nm}$ $h = 161.8 \text{ nm}$ $h_r = 393.41 \text{ nm}$ $\mathbf{E}_{I_s, I_c} = \mathbf{0.1047}$ $\mathbf{E}_{geo} = \mathbf{6.19 \times 10^3 \text{ nm}^2}$ $\mathbf{E}_{geonorm} = \mathbf{12.42\%}$
E_{10}	$h = 219.44 \text{ nm}$ $h_r = 318.81 \text{ nm}$ $E_{I_s, I_c} = 0.195$ $E_{geo} = 8.62 \times 10^3 \text{ nm}^2$ $E_{geonorm} = 12.53\%$	$\sigma = 141.72$ $h = 219.98 \text{ nm}$ $h_r = 318.03 \text{ nm}$ $\mathbf{E}_{I_s, I_c} = \mathbf{0.1869}$ $\mathbf{E}_{geo} = \mathbf{7.66 \times 10^3 \text{ nm}^2}$ $\mathbf{E}_{geonorm} = \mathbf{10.92\%}$	$\rho = -0.01$ $h = 217.84 \text{ nm}$ $h_r = 319.02 \text{ nm}$ $E_{I_s, I_c} = 0.1974$ $E_{geo} = 8.5 \times 10^3 \text{ nm}^2$ $E_{geonorm} = 12.32\%$	$\alpha = 0$ $b = 500 \text{ nm}$ $h = 211.86 \text{ nm}$ $h_r = 321.54 \text{ nm}$ $E_{I_s, I_c} = 0.2269$ $E_{geo} = 1.03 \times 10^4 \text{ nm}^2$ $E_{geonorm} = 15.39\%$

La table 3.27 montre les résultats bruts obtenus avant seuillage. Les valeurs en gras correspondent au profil déterminé par les mesures AFM précédentes. A l'instar des mesures AFM, les TOR_a , TOR_b et TOR_c confirment le caractère non sinusoïdal des profils des échantillons E_6 à E_{10} . On remarque que l'échantillon E_5 est classé par tous les TOR à la sortie P_{f_s} avec une incertitude plus ou moins marquée. Cela confirme les conclusions de la table 3.25 et l'ambiguïté d'affectation à partir de la mesure AFM.

TABLE 3.27 – Résultats bruts de classification des échantillons réels (E_5 , E_6 , E_7 , E_8 , E_9 et E_{10}).

	TOR_a		TOR_b		TOR_c	
	P_s	P_{fs}	P_s	P_{fs}	P_s	P_{fs}
E_5	0.07	0.93	0.17	0.83	0.06	0.94
E_6	0.02	0.98	0.00	1.00	0.03	0.97
E_7	0.00	1.00	0.00	1.00	0.00	1.00
E_8	0.00	1.00	0.00	1.00	0.00	1.00
E_9	0.00	1.00	0.00	1.00	0.00	1.00
E_{10}	0.00	1.00	0.00	1.00	0.00	1.00

Le TOR mis en place a permis de confirmer les résultats de l'AFM sur les 6 échantillons disponibles. Etant donné la faible valeur de la probabilité d'appartenance au modèle cible, le seuil de décision n'a pas été discuté.

Avant de mettre en place un autre type de système classifieur, nous allons d'abord nous attarder sur le comportement du réseau en particulier sur la discrimination de l'espace pour construire les différentes classes.

3.3.1.4 Étude de la frontière

L'étude sur la frontière entre les différents modèles permet de définir le critère d'écart géométrique acceptable entre les profils. Considérons par exemple le cas de profils sinusoidaux. Les faux profils peuvent présenter des paramètres géométriques qui peuvent conduire à la formation de structures proches, voire très proches, d'un profil sinusoidal parfait.

On définit 2 critères pour mesurer l'écart entre 2 profils : un critère indirect, la déviation des signatures optiques générées et un autre plus direct, l'écart géométrique entre les structures physiques. L'écart géométrique normalisé $E_{geomorm}$ peut alors être calculé entre le profil sinus parfait et un profil quelconque (équation 1.44). Ainsi, pour chaque modèle de profil défini, on peut rechercher la forme sinusoidal la plus proche en minimisant $E_{geomorm}$. L'approche par optimisation (méthode de LM) est alors utilisée. La déviation des signatures optiques est caractérisée, quant à elle, par l'erreur E_{I_s, I_c} calculée entre les signatures reconstruites des 2 structures (équation 1.45). Rappelons que la précision de mesure de I_s et I_c est de l'ordre de $E_{ref} = 0.01$, on peut alors déterminer la valeur minimale de $E_{geomorm}$ en dessous de laquelle on ne pourra pas distinguer deux profils différents par la simple mesure de la signature optique.

Nous avons effectué cette étude sur des profils de différentes hauteurs h afin de traiter plusieurs ordres de grandeur en même temps. Ils sont composés de réseaux en résine (S1805) ne débouchant pas sur le substrat de silicium (modèles des échantillons E_5 à E_{10}). La période du réseau est de 627 nm et l'épaisseur h_r de la couche résiduelle est fixée à 500 nm . Les signatures optiques utilisées sont définies sur une gamme spectrale allant de 0.7 à 4.4 eV .

L'angle d'incidence est de 70° .

La catégorie des faux profils sinusoidaux est composée des profils gaussiens, des sinus déformés et des sinus tronqués. Les paramètres géométriques ajustables correspondants sont σ pour le profil gaussien, ρ pour le profil sinus déformé, α et la base b pour le profil sinus tronqué. Les études concernant ces 3 faux profils sinus sont menées conjointement en calculant l'écart géométrique normalisé $E_{geonorm}$ et l'erreur E_{I_s, I_c} pour chacun des 200 échantillons ainsi simulés.

Étude du profil gaussien : Le paramètre σ varie aléatoirement entre 50 et 300. L'évolution de E_{I_s, I_c} en fonction de $E_{geonorm}$ est donnée dans la figure 3.22.

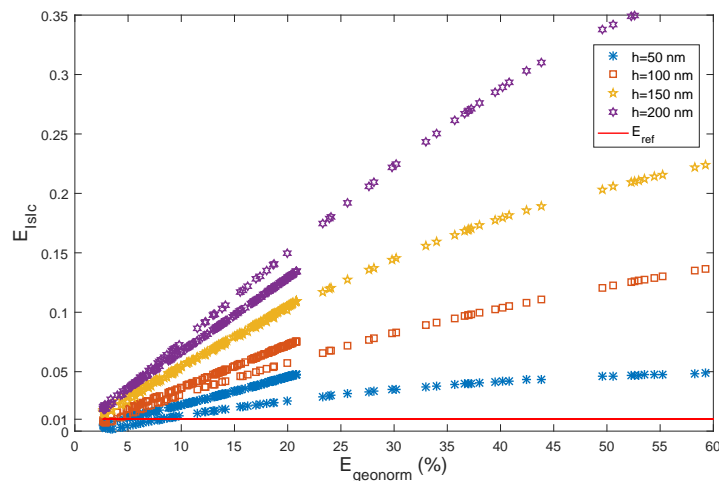


FIGURE 3.22 – Représentation de l'erreur (E_{I_s, I_c}) en fonction de l'écart géométrique normalisé ($E_{geonorm}$) calculé entre un profil gaussien quelconque et le profil sinus le plus proche pour 4 amplitudes différentes.

Étude du profil sinus déformé : Le paramètre ρ varie aléatoirement dans un intervalle $[-0.25; 1]$. La figure 3.23 montre l'évolution de l'erreur E_{I_s, I_c} en fonction de $E_{geonorm}$.

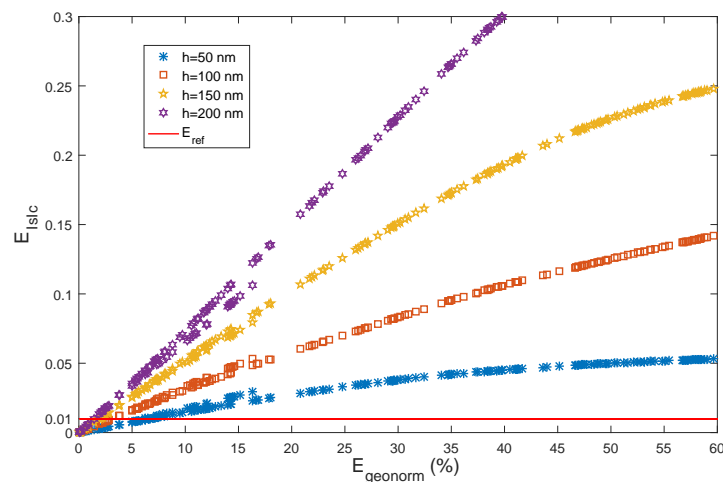


FIGURE 3.23 – Représentation de l'erreur (E_{I_s, I_c}) en fonction de l'écart géométrique normalisé ($E_{geonorm}$) calculé entre un profil sinus déformé quelconque et le profil sinus le plus proche pour 4 amplitudes différentes.

Étude du profil sinus tronqué : Les paramètres α et la base b varient aléatoirement et

respectivement dans les intervalles $[0; 1]$ et $[100; 400]$. L'évolution de E_{I_s, I_c} en fonction de $E_{geonorm}$ est tracée dans la figure 3.24.

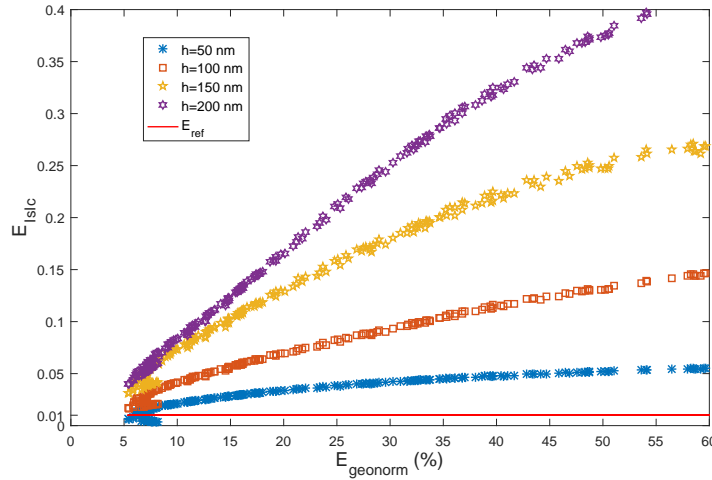


FIGURE 3.24 – Représentation de l'erreur (E_{I_s, I_c}) en fonction de l'écart géométrique normalisé ($E_{geonorm}$) calculé entre un profil sinus tronqué quelconque et le profil sinus le plus proche pour 4 amplitudes différentes.

L'erreur E_{I_s, I_c} varie quasi linéairement en fonction de l'écart géométrique $E_{geonorm}$ pour les 3 cas, en particulier pour les faibles valeurs de $E_{geonorm}$. Cela permet de mettre en évidence la possibilité de discerner les profils qui s'éloignent d'une forme sinus par la simple mesure de la signature optique. Cette constatation est d'autant plus marquée que le profil présente une grande hauteur donc un facteur de forme élevé. Seul le profil sinus déformé permet de déboucher sur la forme sinus parfaite ($E_{geonorm} = 0$) par un jeu de paramètres adéquates ($\rho = 0$). La frontière correspondante entre les 2 profils est donc dans ce cas continue. Il faut noter également la présence d'hystérésis plus ou moins marquée sur chaque courbe suivant le fait que le profil soit globalement réduit ou amplifié par rapport à la forme sinus (écart géométrique positif ou négatif si on prend comme référence la surface de la forme sinusoïdale).

La valeur géométrique minimale $E_{geonormmin}$ en dessous de laquelle on ne pourra pas distinguer 2 profils géométriques différents par l'utilisation de cette signature optique particulière peut ainsi être estimée (table 3.28). Pour certaines amplitudes ($h > 100$ nm) du profil gaussien, les écarts de signatures sont toutes supérieures à E_{ref} , y compris pour le profil gaussien présentant un écart géométrique minimal ($\sigma = 140$). Ainsi dans ce cas de figure la détection d'un faux profil sinus devrait être possible. Pour le profil sinus tronqué, les écarts de signatures sont également tous supérieurs à E_{ref} à partir de $h > 50$ nm.

La table 3.28 montre que l'écart géométrique minimal détectable dépend d'une forme géométrique à une autre mais reste du même ordre de grandeur pour une épaisseur donnée. Par ailleurs, plus le facteur de forme est faible et plus l'écart géométrique minimal détectable augmente. Il faut noter également que, pour certaines formes, il n'y a pas de profil correspondant à cette condition (à partir de $h = 150$ nm pour le profil gaussien et à partir de $h = 100$ nm pour le profil sinus tronqué). Cela correspond au cas idéal où les frontières entre classes sont suffisamment marquées pour pouvoir être détectées.

TABLE 3.28 – Valeurs de l'écart géométrique minimal $E_{geonormmin}$ détectable pour chaque type de profil par la signature optique utilisée.

Différentes hauteurs	Profil gaussien	Profil sinus déformé	Profil sinus tronqué
$h = 50 \text{ nm}$	$E_{geonormmin} = 3.96\%$	$E_{geonormmin} = 6.25\%$	$E_{geonormmin} = 6.76\%$
$h = 100 \text{ nm}$	$E_{geonormmin} = 2.72\%$	$E_{geonormmin} = 2.85\%$	Pas de profil correspondant
$h = 150 \text{ nm}$	Pas de profil correspondant	$E_{geonormmin} = 2.05\%$	Pas de profil correspondant
$h = 200 \text{ nm}$	Pas de profil correspondant	$E_{geonormmin} = 1.41\%$	Pas de profil correspondant

Cette étude nous a permis de mettre en évidence un écart géométrique minimal détectable entre 2 profils pouvant expliquer en partie le classement en faux positifs de certains échantillons de la table 3.17.

3.3.2 Détermination du type de profil géométrique

Le TOR mis en place précédemment ne permet que de détecter la présence d'une forme attendue. Il trouve son utilité sur une chaîne de fabrication afin d'éviter toute dérive du procédé de la forme cible. Le PMC utilisé dans ce paragraphe est destiné, cette fois ci, à apporter une information plus qualitative nécessaire à toute caractérisation de type scattérométrique. En effet, comme nous l'avons vu dans le chapitre 1, la résolution du problème inverse nécessite un certain nombre de connaissances a priori pour assurer des résultats fiables. Dans ce cadre un PMC peut être utilisé afin de déterminer en amont, de manière automatique, une des hypothèses inhérentes à la caractérisation. Dans cette partie nous nous sommes concentrés sur l'hypothèse de forme géométrique de la structure. Nous utilisons un PMC en mode classifieur permettant l'identification du modèle géométrique le plus probable correspondant à l'échantillon présenté.

3.3.2.1 R.N classifieur à 3 classes

Dans ce paragraphe, un PMC classifieur à 3 classes est étudié dans le cas de la discrimination de formes. Ce dernier peut identifier un profil géométrique particulier parmi un nombre limité de possibilités fixées en amont. Les modèles choisis dépendent bien entendu des types d'échantillons à tester et des connaissances a priori sur les déformations et dérives possibles. Nous avons effectué cette étude sur des réseaux de formes rectangulaires et sinusoïdaux.

3.3.2.1.1 Classification de réseaux de type rectangulaire

La classe P_c est définie pour le profil rectangulaire tandis que les classes P_{tr} et P_{ah} correspondent respectivement au profil trapézoïdal et au profil rectangulaire arrondi au sommet. Un PMC nommé ClassP par la suite est mis en place pour caractériser les échantillons E_2 à E_4 (figure 3.2). Chaque classe est composée de 3000 échantillons théoriques générés aléatoirement à partir des domaines de variations définis au paragraphe 3.3.1.2 pour constituer le corpus d'apprentissage. Les signatures optiques (I_s et I_c) sont également simulées dans les mêmes conditions.

3.3.2.1.1.1 Performance après apprentissage

Le PMC ClassP est entraîné sur 9000 échantillons et composé de 25 neurones cachés. Les performances sont données dans la table 3.29 dans le cas d'un seuil d'acceptabilité fixé à 50%.

TABLE 3.29 – Matrice de confusion de ClassP dédié à la classification de 3 formes géométriques. La classe P_c correspond aux profils rectangulaires, les classes P_{tr} et P_{ah} respectivement aux profils trapézoïdaux et aux profils rectangulaires arrondis au sommet.

		Classe réelle			Précision
		P_c	P_{tr}	P_{ah}	
Classe estimée	P_c	450	0	19	95.9%
	P_{tr}	0	450	0	100%
	P_{ah}	0	0	431	100%
Sensibilité		100%	100%	95.8%	Exactitude : 98.6%

Le classifieur ClassP fournit une excellente performance globale (exactitude de 98.6%). Il garantit une précision maximale pour les classes P_{tr} et P_{ah} et une sensibilité maximale pour les classes P_c et P_{tr} . Tous les profils rectangulaires et trapézoïdaux réels sont correctement classés. Seul 19 profils rectangulaires arrondis au sommet sont mal classés dans P_c .

Comme dans le cas du TOR, il est intéressant de s'attarder sur ces profils faussement acceptés dans P_c . La figure 3.25 illustre la sortie brute P_c en fonction du paramètre r_h (équation 1.10) définissant l'arrondi en haut du profil correspondant. On note comme dans le cas de la figure 3.19 une certaine corrélation entre la valeur de sortie du PMC et l'écart géométrique au profil créneau parfait. Les valeurs r_h sont relativement faibles. La valeur la plus élevée est de 23.17 nm ; ce qui correspond, pour ce profil, à un écart géométrique de 1.63% avec le créneau parfait le plus proche. L'erreur E_{I_s, I_c} entre ces 2 profils est de l'ordre de 0.02 (soit $2 \times E_{ref}$). Ainsi, on retrouve le fait que la classe P_c a une sensibilité aux profils rectangulaires avec un faible arrondi au sommet (paragraphe 3.3.1.2).

Afin de diminuer le nombre de FP, on applique les 2 seuils d'acceptabilité (70% et 90%) définis précédemment ainsi que l'introduction d'une classe indéterminée (C.I.). Les tables 3.30 et 3.31 affichent respectivement le résultat pour un seuil fixé à 70% et 90%.

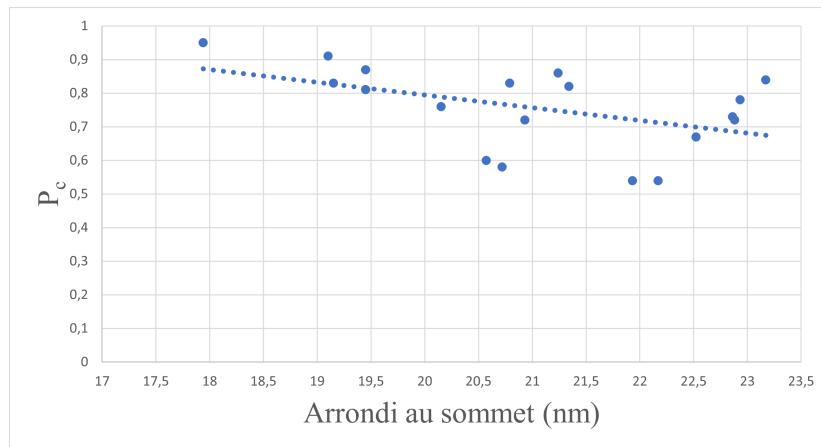


FIGURE 3.25 – Sortie non seuillée correspondant à la classe P_c du ClassP pour les 19 échantillons mal classés (FP) en fonction de la valeur de l'arrondi en haut r_h .

TABLE 3.30 – Matrice de confusion de ClassP dédié à la classification de 3 formes géométriques (P_c , P_{tr} et P_{ah}) pour un seuil d'affectation fixé à 70%.

		Classe réelle			Précision
		P_c	P_{tr}	P_{ah}	
Classe estimée	P_c	449	0	14	96.9%
	P_{tr}	0	450	0	100%
	P_{ah}	0	0	420	100%
	C.I	1	0	16	
Sensibilité		99.7%	100%	93.3%	Exactitude : 97.7%

TABLE 3.31 – Matrice de confusion de ClassP dédié à la classification de 3 formes géométriques (P_c , P_{tr} et P_{ah}) pour un seuil d'affectation fixé à 90%.

		Classe réelle			Précision
		P_c	P_{tr}	P_{ah}	
Classe estimée	P_c	419	0	2	99.5%
	P_{tr}	0	450	0	100%
	P_{ah}	0	0	417	100%
	C.I	31	0	31	
Sensibilité		93.1%	100%	92.6%	Exactitude : 95.2%

Les tables 3.30 et 3.31 montrent que la performance globale de ClassP diminue lorsque le seuil d'acceptabilité augmente ; ce qui entraîne une dégradation de la sensibilité pour certaines classes. En revanche, on augmente significativement la précision de la classe P_c . Le degré de confiance pour chaque sortie s'en trouve amélioré au détriment d'une augmentation des cas indéterminés.

3.3.2.1.1.2 Résultats expérimentaux

Les signatures optiques mesurées pour les échantillons E_2 , E_3 et E_4 (figure 3.2) sont fournies comme entrées du classifieur ClassP pour assurer la validation expérimentale. Les résultats des sorties sont affichés dans la table 3.32. On fait également correspondre pour chaque échantillon la classe automatiquement identifiée par application de différents seuillages.

TABLE 3.32 – Résultats de classification des échantillons E_2 , E_3 et E_4 par ClassP.

	P_c	P_{tr}	P_{ah}		50%	70%	90%
E_2	0.00	0.00	1.00	\Rightarrow	E_2	P_{ah}	P_{ah}
E_3	0.00	0.00	1.00		E_3	P_{ah}	P_{ah}
E_4	0.56	0.00	0.44		E_4	P_c	X

Le ClassP identifie sans ambiguïté les échantillons E_2 et E_3 à la classe P_{ah} . Cette troisième sortie de ClassP bénéficiant d'une précision de 100%, le résultat est donc considéré comme fiable. On retrouve les résultats estimés par le TOR pour ces 2 échantillons (table 3.16).

En revanche, le résultat de la classification de l'échantillon E_4 est plus ambivalent. En effet, le classifieur indique une légère préférence pour le profil rectangulaire au détriment du profil arrondi en haut.

Concernant l'identification automatique, les échantillons E_2 et E_3 seront systématiquement bien classés indépendamment du seuil d'acceptabilité souhaité. Par contre, pour l'échantillon E_4 , si on veut une sensibilité forte sur la première sortie alors l'identification automatique ne pourra décider catégoriquement et classera l'échantillon dans la classe C.I au delà d'un seuil fixé à 50%.

Ces résultats sont totalement conformes aux images MEB des échantillons (figures 3.2) notamment celle de E_4 qui montre une forme rectangulaire légèrement déformée au sommet.

3.3.2.1.1.3 Exemple de caractérisation complète

Le classifieur ClassP peut alors être utilisé pour effectuer la caractérisation en réduisant l'impact d'une forme de profil fixe erronée. Par exemple, l'échantillon E_3 a été clairement identifié par le ClassP comme un profil rectangulaire arrondi au sommet. L'échantillon E_4 semble être à la limite entre le profil rectangulaire et le profil rectangulaire arrondi au sommet. Ainsi, les PMC fonctionnant sur ces formes fixes sont implémentés. Ils sont nommés RegPc et RegPah. L'apprentissage est fait sur 3000 échantillons théoriques sur chaque modèle fixe défini. Chaque PMC contient un nombre optimal de neurones dans la couche cachée déterminé par des résultats préliminaires. Ainsi, le RegPc contient 20 neurones cachés et le RegPah 15 neurones cachés. La performance optimale calculée sur les données test sur chacune des sorties de chaque PMC est donnée dans la table 3.33 et nous permet de vérifier la qualité de l'apprentissage.

TABLE 3.33 – Performances des PMC nommés *RegPc* et *RegPah* fonctionnant respectivement sur les modèles P_c et P_{ah} dans les plages définies précédemment (équations 3.11 et 3.13).

RegPc	RegPah
$E_{test}^{CD} = 0.03 \text{ nm}$	$E_{test}^{\delta} = 0.01$
$E_{test}^h = 0.03 \text{ nm}$	$E_{test}^{CD_h} = 0.66 \text{ nm}$
	$E_{test}^h = 0.4 \text{ nm}$

Les signatures optiques mesurées des échantillons E_3 et E_4 sont fournies aux entrées de ces PMC pour la caractérisation. Les résultats sont donnés dans la table 3.34. L'erreur E_{I_s, I_c} calculée entre la signature optique mesurée et celle simulée est ajoutée pour qualifier la qualité des résultats.

TABLE 3.34 – Caractérisations par *RegPc* et *RegPah* des échantillons E_3 et E_4 pour une forme de profil fixée par le classifieur *ClassP*.

PMC		E_3	E_4
RegPc	Paramètres géométriques		$CD = 246.44 \pm 0.03 \text{ nm}$ $h = 559.48 \pm 0.03 \text{ nm}$
	Qualité de la caractérisation		$E_{I_s, I_c} = 0.16$
RegPah	Paramètres géométriques	$\delta = 0.32 \pm 0.01$ $CD_h = 209.61 \pm 0.66 \text{ nm}$ $h = 383.05 \pm 0.4 \text{ nm}$	$\delta = 0.09 \pm 0.01$ $CD_h = 245.76 \pm 0.66 \text{ nm}$ $h = 556.76 \pm 0.4 \text{ nm}$
	Qualité de la caractérisation	$E_{I_s, I_c} = 0.12$	$E_{I_s, I_c} = 0.16$

Le *RegPah* estime les paramètres géométriques de l'échantillon E_3 inclus dans la plage d'apprentissage. Le rayon estimé au sommet est significatif ($r_h = 33.53 \text{ nm}$). Ceci est cohérent avec l'image MEB.

Concernant l'échantillon E_4 , les paramètres géométriques estimés par *RegPc* et *RegPah* sont du même ordre de grandeur. L'arrondi au sommet estimé par *RegPah* est très faible ($r_h = 11.05 \text{ nm}$) et le profil semble proche d'un modèle rectangulaire. En effet, l'écart géométrique E_{geo} calculé entre le profil estimé par *RegPah* et le modèle rectangulaire le plus proche donné par *RegPc* est de 101.56 nm^2 . Cela représente seulement 0.07% de la surface totale du réseau. La figure 3.26 représente les résultats de profils estimés par *RegPc* et *RegPah*. On constate que les 2 profils sont relativement proches.

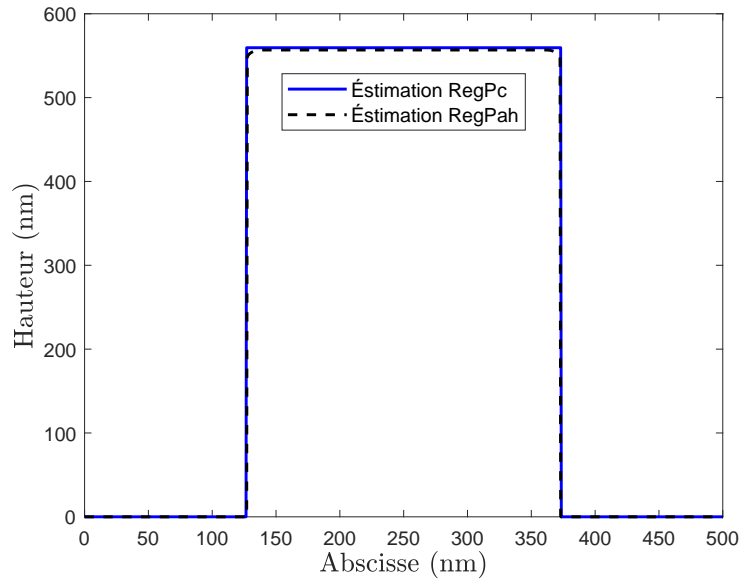


FIGURE 3.26 – Représentations des profils géométriques obtenus par différentes caractérisations de l'échantillon E_4 : RegPc (en bleu) et RegPah (en noir pointillée).

Ainsi le ClassP fournit à l'expérimentateur la forme de profil supposé de l'échantillon à tester afin d'assurer une caractérisation optimale. Dans le cas de l'échantillon E_3 , l'identification est primordial et permet de confirmer l'hypothèse de forme évitant ainsi les erreurs liées à la résolution d'un problème mal posé. Dans le cas particulier de l'échantillon E_4 , la forme peut être décrite aussi bien par un créneau que par un modèle avec un faible arrondi en haut. Ainsi quelle que soit la structure choisie en fonction de la sortie de ClassP, le résultat de la caractérisation finale ne sera que peu affecté par la décision prise quant à la forme estimée.

3.3.2.1.2 Classification de réseau du type sinusoïdal

Cette fois-ci, nous étudions le cas des réseaux de formes sinusoïdales (E_5 à E_{10}) en utilisant un classifieur composé de 3 sorties nommé Class3 par la suite. La classe P_s est définie pour le profil sinus, la classe P_g pour le profil gaussien et la classe P_{sd} pour le profil sinus déformé. Il est entraîné sur 3 types de corpus correspondant respectivement à différentes définitions de frontières entre classes. Ils sont dénommés respectivement $Class3_a$, $Class3_b$ et $Class3_c$. Les échantillons sont tirés aléatoirement dans les domaines de variation définis précédemment (équations 3.15, 3.16 et 3.17) tout en calculant pour chaque profil l'écart géométrique aux 2 autres profils les plus proches et en s'assurant que celles ci soient telles que $E_{geonorm} \geq 1\%$, $E_{geonorm} \geq 5\%$ et $E_{geonorm} \geq 15\%$. Le PMC classifieur $Class3_a$ contient 20 neurones cachés, 25 pour $Class3_b$ et 20 pour $Class3_c$.

3.3.2.1.2.1 Performance après apprentissage

Nous avons entraîné chaque PMC classifieur sur 9000 échantillons théoriques dont 3000 pour chaque classe. Les résultats sont donnés sous la forme de matrices de confusion à la suite de chaque apprentissage en prenant un seuil d'acceptabilité de 50% :

TABLE 3.35 – Matrice de confusion du $Class3_a$ dédié à la classification parmi 3 profils géométriques (P_s , P_g et P_{sd}) pour un seuil d'acceptabilité fixé à 50%.

		Classe réelle			Précision
		P_s	P_g	P_{sd}	
Classe estimée	P_s	450	10	4	97.0%
	P_g	0	433	14	96.9%
	P_{sd}	0	7	432	98.4%
Sensibilité		100%	96.2%	96.0%	Exactitude : 97.4%

TABLE 3.36 – Matrice de confusion du $Class3_b$ dédié à la classification parmi 3 profils géométriques (P_s , P_g et P_{sd}) pour un seuil d'acceptabilité fixé à 50%.

		Classe réelle			Précision
		P_s	P_g	P_{sd}	
Classe estimée	P_s	450	4	2	98.7%
	P_g	0	437	10	97.8%
	P_{sd}	0	9	438	98.0%
Sensibilité		100%	97.1%	97.3%	Exactitude : 98.1%

TABLE 3.37 – Matrice de confusion du $Class3_c$ dédié à la classification parmi 3 profils géométriques (P_s , P_g et P_{sd}) pour un seuil d'acceptabilité fixé à 50%.

		Classe réelle			Précision
		P_s	P_g	P_{sd}	
Classe estimée	P_s	450	0	0	100%
	P_g	0	445	8	98.2%
	P_{sd}	0	5	442	98.9%
Sensibilité		100%	98.9%	98.2%	Exactitude : 99.0%

Comme dans le cas du TOR, les performances sont meilleures lorsque les frontières entre classes sont bien marquées y compris dans les échantillons utilisés pour les tests. Le $Class3_c$ estime ainsi des bonnes performances globales (exactitude de 99%) pour des échantillons aux formes clairement distinctes ($E_{geonorm} \geq 15\%$). Il assure également une précision de 100% sur la sortie P_s . Tous les classifieurs assurent un classement correcte des profils sinus (sensibilité P_s de 100%). Toutefois, ces performances ne sont pas évaluées sur le même corpus. On ne peut donc pas les comparer entre elles.

Nous avons alors testé les performances de ces 3 classifieurs sur le même corpus test ($E_{geonorm} \geq 1\%$). Les résultats obtenus sont donnés dans les différentes matrices de confusions (tables 3.35, 3.38 et 3.39) :

TABLE 3.38 – Matrice de confusion du $Class3_b$ évaluée sur les données test du corpus assurant $E_{geonorm} \geq 1\%$ pour un seuil d'acceptabilité fixé à 50%.

		Classe réelle			Précision
		P_s	P_g	P_{sd}	
Classe estimée	P_s	450	22	13	92.7%
	P_g	0	421	12	97.2%
	P_{sd}	0	7	425	98.3%
Sensibilité		100%	93.5%	94.4%	Exactitude : 96.0%

TABLE 3.39 – Matrice de confusion du $Class3_c$ évaluée sur les données test du corpus assurant $E_{geonorm} \geq 1\%$ pour un seuil d'acceptabilité fixé à 50%.

		Classe réelle			Précision
		P_s	P_g	P_{sd}	
Classe estimée	P_s	450	60	36	82.4%
	P_g	0	375	14	96.4%
	P_{sd}	0	15	400	96.3%
Sensibilité		100%	83.3%	88.8%	Exactitude : 90.7%

Les résultats des tables 3.38 et 3.39 montrent que tous les profils sinus sont correctement classés en gardant une sensibilité maximale pour le profil P_s . Cependant, le $Class3_b$ et $Class3_c$ affichent respectivement des performances globales de 96.0% et 90.7% contre 97.4% pour le $Class3_a$ sur le même corpus. Ce qui fait passer le nombre des échantillons mal classés de 35 pour le $Class3_a$ à 125 pour le $Class3_c$. On note également une dégradation de la sensibilité et de la précision.

Comme dans le cas de l'étude du TOR, il est donc intéressant de s'attarder sur le seul résultat obtenu par le $Class3_a$. On note plusieurs profils faussement classés dans les résultats. Les figures 3.27 et 3.28 représentent le résultat de sortie du $Class3_a$ pour les échantillons mal classés respectivement gaussien et sinus déformé en fonction de leur paramètre σ et ρ .

On note que les échantillons faussement classés P_s (en rouge) affichent des profils avec des paramètres géométriques σ et ρ autour de leurs valeurs optimales ($\sigma = 140$ cf figure 3.20 et $\rho = 0$). Ces valeurs représente le plus faible écart géométrique au profil sinusoïdal parfait. Ainsi l'écart surfacique maximal au sinus parfait le plus proche correspondrait à 2.7% pour le profil gaussien et 0% pour le profil sinus déformé. Les mêmes constatations peuvent être faites sur les échantillons faussement classés par P_g (figure 3.27) et P_{sd} (figure 3.28).

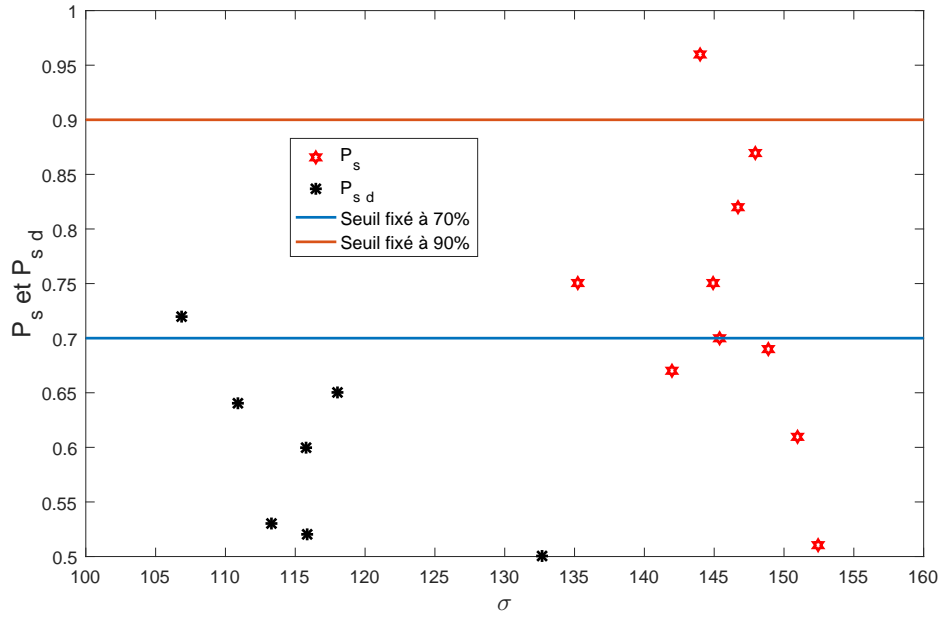


FIGURE 3.27 – Sorties brutes P_s et P_{sd} supérieures à 0.5 du PMC classifieur $Class3_a$ pour les échantillons au profil gaussien mal classés en fonction du paramètre σ .

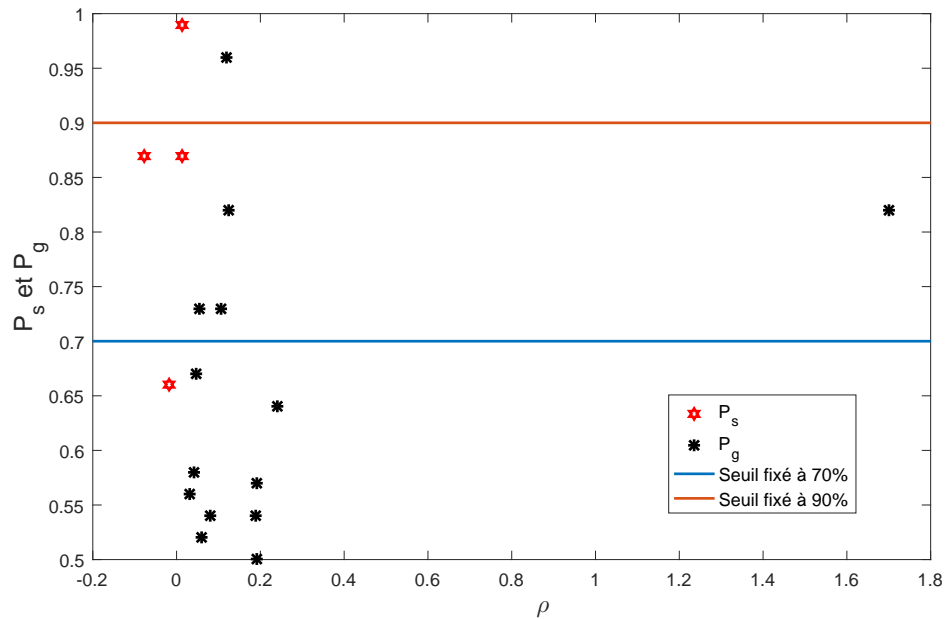


FIGURE 3.28 – Sorties brutes P_s et P_g supérieures à 0.5 du PMC classifieur $Class3_a$ pour les échantillons mal classés au profil sinus déformé en fonction du paramètre ρ .

Par ailleurs, en modifiant le seuil d'acceptabilité (traits horizontaux sur les figures 3.27 et 3.28), le nombre d'échantillons mal classés diminue. Cela permet d'augmenter la précision sur toutes les sorties (tables 3.40 et 3.41) au détriment de la sensibilité sur les classes par l'ajout de la classe C.I.

TABLE 3.40 – Matrice de confusion du $Class3_a$ évaluée sur les données test du corpus assurant $E_{geonorm} \geq 1\%$ pour un seuil d'acceptabilité fixé à 70%.

		Classe réelle			Précision
		P_s	P_g	P_{sd}	
Classe estimée	P_s	448	6	3	98.0%
	P_g	0	416	5	98.8%
	P_{sd}	0	1	412	99.7%
	C.I	2	27	30	
Sensibilité		99.5%	92.4%	91.5%	Exactitude : 94.5%

TABLE 3.41 – Matrice de confusion du $Class3_a$ évaluée sur les données test du corpus assurant $E_{geonorm} \geq 1\%$ pour un seuil d'acceptabilité fixé à 90%.

		Classe réelle			Précision
		P_s	P_g	P_{sd}	
Classe estimée	P_s	438	1	1	99.5%
	P_g	0	376	1	99.7%
	P_{sd}	0	0	384	100%
	C.I	12	74	64	
Sensibilité		97.3%	83.5%	85.3%	Exactitude : 88.7%

Il est clair que le $Class3_a$ présente les meilleures performances en terme de généralisation sur des données inconnues en particulier lorsqu'on veut discriminer deux formes proches. Ainsi, dans la suite, nous retenons exclusivement ce classifieur.

3.3.2.1.2.2 Estimation des performances

Dans cette section, nous allons comme dans le cas du TOR, étudier le comportement du classifieur $Class3_a$ lors de la déformation théorique d'un profil donné. L'évolution des formes décrites au paragraphe 3.3.1.3.2 ont été prises en compte. On utilise ainsi les différents seuils d'acceptabilité fixés (50%, 70% et 90%).

Nous étudions le cas du profil gaussien. Le classifieur $Class3_a$ estime une forme pour chaque échantillon défini par une valeur particulière de σ . La figure 3.29 représente l'évolution des profils identifiés par le $Class3_a$ suivant le seuil d'acceptabilité fixé. La valeur 1 correspond

à un profil identifié, la valeur 0.5 à une classe indéterminée et la valeur 0 à un profil rejeté. L'écart surfacique normalisé ($E_{geonorm}$) calculé entre le profil sinus parfait le plus proche et le profil gaussien défini par σ est également représenté sur le même graphique.

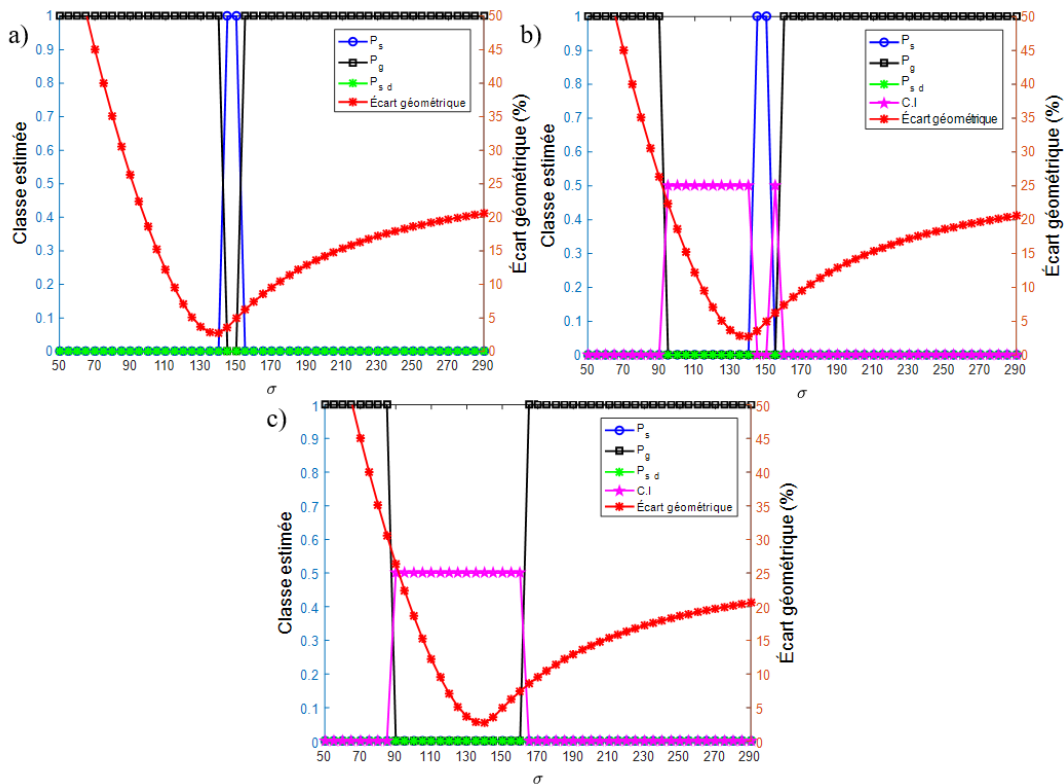


FIGURE 3.29 – Évolution des profils identifiés P_s , P_g et P_{sd} par le PMC classifieur $Class3_a$ en fonction du paramètre σ du profil gaussien pour un seuil d'acceptabilité fixé respectivement à (a) 50%, (b) 70% et (c) 90%.

Il faut noter dans un premier temps que la classe P_{sd} n'est jamais détectée. Le profil P_g est logiquement identifié suivant la majorité des valeurs de σ ; ce qui est conforme avec l'étude. Quant au profil P_s , il est identifié automatiquement autour de la valeur de $\sigma = 140$ pour les seuils de 50% et 70% à différent degré de sensibilité. Rappelons que cette valeur particulière de σ correspond bien à l'écart géométrique le plus faible entre le profil gaussien et le profil sinus parfait le plus proche (courbe en rouge sur la figure 3.29). Le nombre d'échantillons de la classe indéterminée augmente avec le seuil d'acceptabilité. Ainsi comme aucune valeur de σ ne permet d'approcher la forme sinusoidale parfaite à moins de 2.7%, le $Class3_a$ ne peut assurer avec une précision de 99.5% que l'échantillon $\sigma = 140$ est de forme sinusoidale. Dans le cas d'une classification automatique, il est donc important de définir en amont le niveau d'exigence voulue pour la définition d'un profil donné.

En complément de cette étude, la figure 3.30 compare l'évolution de la sortie P_s du $Class3_a$ avec celle du TOR_a en fonction de σ utilisé dans le cadre de la détection de forme sinusoidale. L'écart géométrique normalisé entre le profil gaussien et le profil sinus parfait le plus proche est également ajouté dans le même graphe.

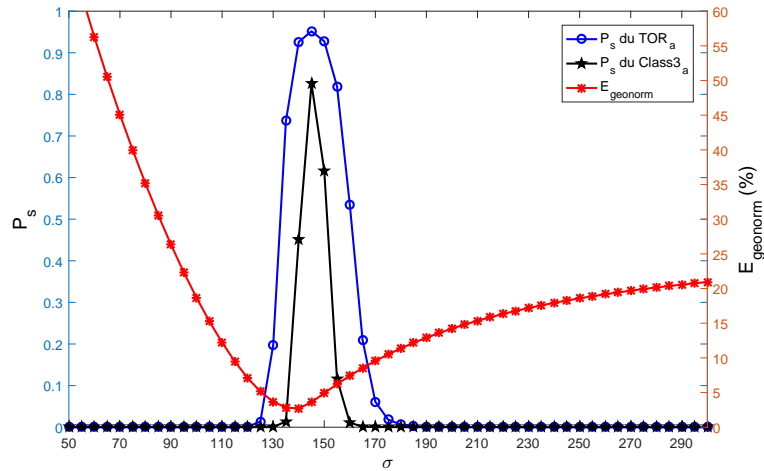


FIGURE 3.30 – Évolution de la sortie P_s des PMC classifieurs TOR_a et $Class3_a$ en fonction du paramètre σ du profil gaussien. L'écart géométrique normalisé ($E_{geonorm}$) est également tracé (en rouge)

Dans les 2 cas de figure, les sorties P_s n'atteignent pas le maximum. Les profils présentés ne correspondent pas à la définition d'un profil sinusoïdal parfait tel que cela a été défini lors de l'apprentissage. Cependant il s'en rapproche pour une valeur de sigma autour de 140 et c'est ce qu'indiquent les 2 classifieurs d'une manière plus ou moins bien marquée. Par ailleurs, il semble que, contrairement à ce qu'on pourrait attendre, l'identification d'un profil sinusoïdal est plus précise dans le cas du classifieur $Class3_a$ qui présente une largeur de détection plus faible que le TOR_a (table 3.24) et ce, quel que soit le seuil d'acceptabilité choisi par exemple, pour un seuil acceptabilité de 50%, $\Delta\sigma = 10.5$ pour le $Class3_a$ contre $\Delta\sigma = 27.5$ pour le TOR_a .

3.3.2.1.2.3 Résultats expérimentaux

La validation expérimentale de cette étude est réalisée sur les échantillons (E_5 , E_6 , E_8 et E_{10}) présentés dans la partie 3.1.2 (figure 3.3). Les signatures optiques mesurées de ces derniers sont introduites à l'entrée de $Class3_a$. La table 3.42 affiche le résultat brut de classification de ces échantillons ainsi que la classe identifiée automatiquement dans le cas de différents seuils d'acceptabilité. Les valeurs en gras correspondent comme précédemment au profil déterminé par les mesures AFM.

TABLE 3.42 – Résultats de classification des échantillons réels (E_5 , E_6 , E_8 et E_{10}).

	P_s	P_g	P_{sd}		50%	70%	90%	
E_5	0.00	0.69	0.31	⇒	E_5	P_g	X	X
E_6	0.00	0.90	0.10		E_6	P_g	P_g	P_g
E_8	0.00	0.22	0.78		E_8	P_{sd}	P_{sd}	X
E_{10}	0.00	0.98	0.02		E_{10}	P_g	P_g	P_g

Les résultats d'identification des échantillons E_6 , E_8 et E_{10} correspondent parfaitement à la forme identifiée par l'AFM (tables 3.25 et 3.26) ; et ce quel que soit le seuil d'acceptabilité choisi (exception pour E_8 au delà de 70%). On remarque également que la forme de l'échantillon E_5 n'est pas clairement identifiée conformément à ce qui a été dit précédemment (paragraphe 3.3.1.3.3). Par contre, les résultats sont conformes à ceux du TOR_a (table 3.27) en rejetant le profil sinusoidal parfait. L'ambiguïté repose sur les profils P_g et P_{sd} confirmée par les résultats de l'AFM (table 3.25).

3.3.2.1.2.4 Exemple d'une caractérisation complète

Pour finaliser l'étude, nous allons présenter un exemple de caractérisation complète du réseau E_8 identifié comme P_{sd} par $Class3_a$. Ainsi, le PMC régresseur de type RegPsd est implémenté avec une forme fixe correspondant au profil sinus déformé. L'apprentissage est fait sur 3000 échantillons théoriques pour chaque profil. Le nombre de neurones cachés est de 15. La performance optimale calculée sur les données test sur chacune de sortie de RegPsd est donnée dans la table 3.43.

TABLE 3.43 – Performance de RegPsd (une couche cachée composée de 15 neurones) après entraînement évaluée sur les données test (450 échantillons) fonctionnant exclusivement sur un profil sinus déformé.

Performance de RegPsd		
$E_{test}^p = 0.004$	$E_{test}^h = 0.3 \text{ nm}$	$E_{test}^{h_r} = 0.21 \text{ nm}$

La signature optique mesurée de l'échantillon E_8 est alors fournie à l'entrée de RegPsd. Le résultat est donné dans la table 3.44.

TABLE 3.44 – Résultat de caractérisation par RegPsd de l'échantillon E_8 .

	RegPsd
Paramètres géométriques	$\rho = -0.13 \pm 0.004$ $h = 140.78 \pm 0.3 \text{ nm}$ $h_r = 420.3 \pm 0.21 \text{ nm}$
Qualité de la caractérisation	$E_{I_s, I_c} = 0.0564$ $E_{geo} = 5.48 \times 10^3 \text{ nm}^2$ $E_{geonorm} = 10.99\%$

Le RegPsd estime des paramètres géométriques inclus dans la plage définie. L'épaisseur h_r estimée est conforme à celle obtenue par optimisation à partir des mesures AFM (table 3.27).

Nous avons représenté les signatures optiques mesurées et celles reconstruites à la fois à partir des résultats de RegPsd et à partir des mesures AFM (figure 3.31). Les signatures sont confondues dans la partie infrarouge mais on observe un léger écart entre les signatures mesurées et estimées dans la partie UV.

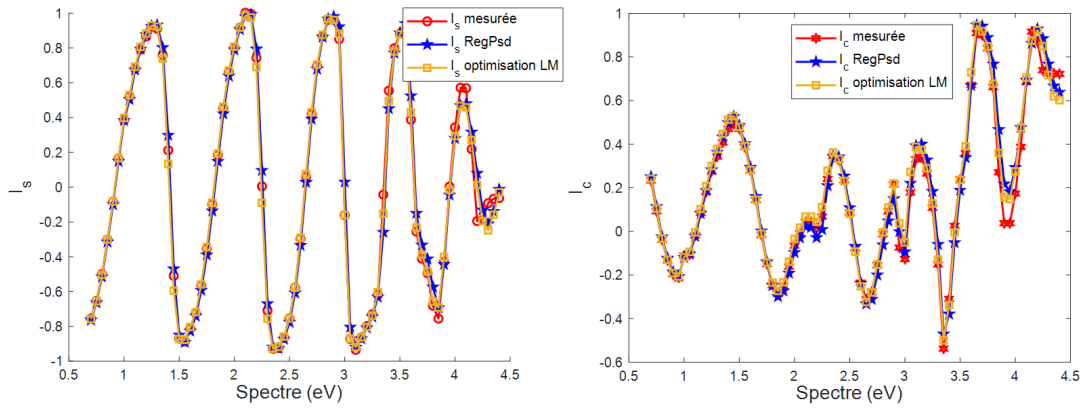


FIGURE 3.31 – Comparaison des signatures optiques simulées à partir des résultats de *RegPsd* et à partir des résultats de l'optimisation des mesures AFM avec celles mesurées sur l'échantillon E_8 .

La figure 3.32 représente les profils estimés par le *RegPsd*, les mesures AFM brutes et celui obtenu après traitement sur 4 périodes. On constate que la forme de profil estimée par le *RegPsd* est relativement proche de celle obtenue par l'AFM. Il faut rappeler, comme cela a déjà été évoqué plus haut, que la référence choisie dans le cadre de cette étude est la mesure AFM. Toutefois, comme on le note ici cette mesure est très locale et peut présenter des artefacts liés à la forme de la pointe, à un défaut localisé, etc. Dans cet exemple, on note clairement une dérive du profil sur seulement 4 périodes.

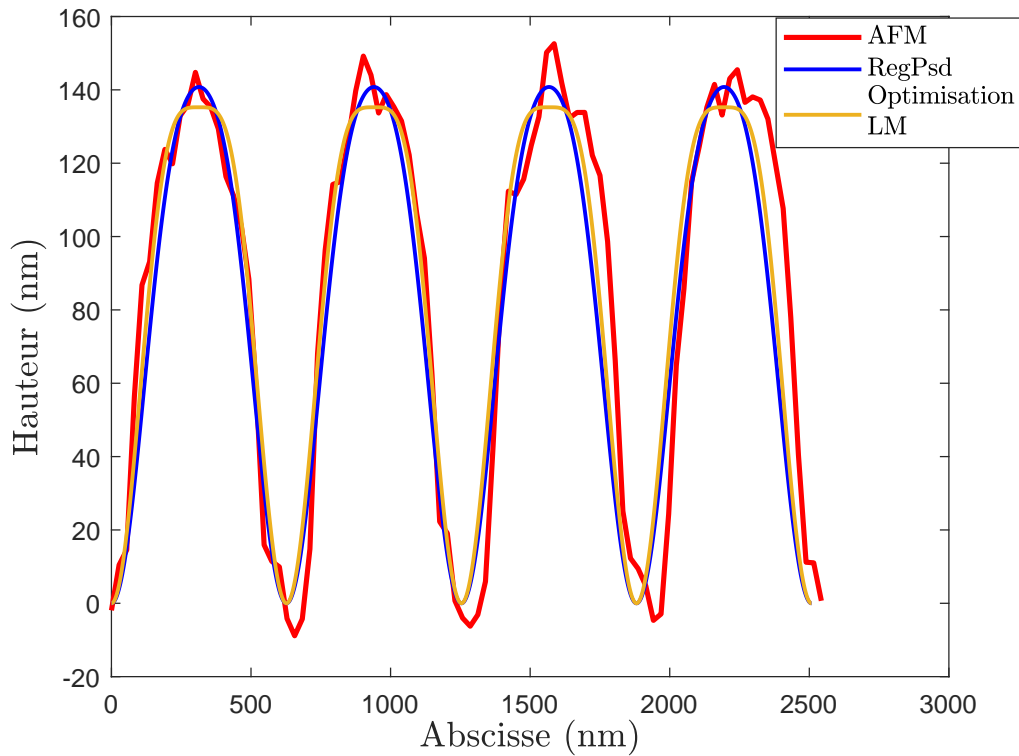


FIGURE 3.32 – Représentation des profils obtenus par caractérisation de l'échantillon E_8 par *RegPsd* (en bleu), par AFM (en rouge) et par optimisation (en or).

3.3.2.2 R.N classifieur à 4 classes

Pour un fonctionnement optimal, il est primordial de connaître à l'avance le nombre et le type de défauts possibles que peut occasionner le procédé de fabrication. Un classifieur peut alors être créé prenant en compte l'ensemble des classes ainsi modélisées. Ce nombre n'est pas limité mais conditionne nécessairement les performances du classifieur. Reprenons l'étude précédente en introduisant par exemple une classe supplémentaire correspondant à la forme sinus tronquée notée précédemment P_{st} . Le PMC mis en place sera nommé $Class4_a$ par la suite. On effectue alors un tirage aléatoire dans les domaines de variation définis (équations 3.15, 3.16, 3.17 et 3.18) tout en calculant systématiquement pour chaque profil (P_s , P_g , P_{sd} et P_{st}) l'écart géométrique aux 3 autres profils les plus proches pour définir une frontière. Cependant, suivant les conclusions des paragraphes 3.3.1.3 et 3.3.2.1, seul le corpus avec des données $E_{geomorm} \geq 1\%$ sera exploité.

3.3.2.2.1 Performance après apprentissage

Le classifieur a nécessité 12000 échantillons théoriques dont 3000 pour chaque classe. On voit clairement que le passage de 3 à 4 classes nécessite une architecture plus complexe avec un corpus d'apprentissage plus grand. Il contient également 30 neurones cachés. La performance du PMC classifieur est donnée sous la forme d'une matrice de confusion dans la table 3.45.

TABLE 3.45 – Matrice de confusion de $Class4_a$ axée sur l'identification d'un profil géométrique particulier.

		Classe réelle				Précision
		P_s	P_g	P_{sd}	P_{st}	
Classe estimée	P_s	450	13	9	0	95.3%
	P_g	0	414	15	1	96.3%
	P_{sd}	0	19	406	39	87.5%
	P_{st}	0	4	4	410	94.5%
Sensibilité		100%	92%	90.2%	91.1%	Exactitude : 93.3%

On remarque les mêmes constatations que les études précédentes sur les classifieurs TOR et Class3 : le $Class4_a$ assure un classement correct de tous les profils sinus (sensibilité P_s de 100%). La performance globale est légèrement plus faible que le $Class3_a$. Le nombre d'échantillons mal classés passe également à 5.8% du corpus test contre 2.6%. Ce qui confirme l'importance de la connaissance préalable des défauts possibles afin de limiter le nombre de classes possibles.

Le $Class4_a$ est évalué dans le paragraphe suivant dans le cadre de la déformation continue d'une forme géométrique donnée dans les mêmes conditions d'études utilisées dans le cas précédent.

3.3.2.2.2 Estimation des performances sur la déformation d'un profil géométrique

Considérons le cas du profil gaussien, où plusieurs structures sont simulées en faisant varier uniquement le paramètre σ (paragraphe 3.3.2.1.2). Les signatures optiques sont alors fournies au $Class4_a$. En considérant les différents seuils d'acceptabilités fixés (50% 70% et 90%), la figure 3.33 présente l'évolution des profils identifiés (P_s , P_g , P_{sd} et P_{st}) par le $Class4_a$ en fonction de σ . L'écart géométrique entre le profil gaussien et le profil sinus parfait le plus proche est également ajouté dans le même graphe.

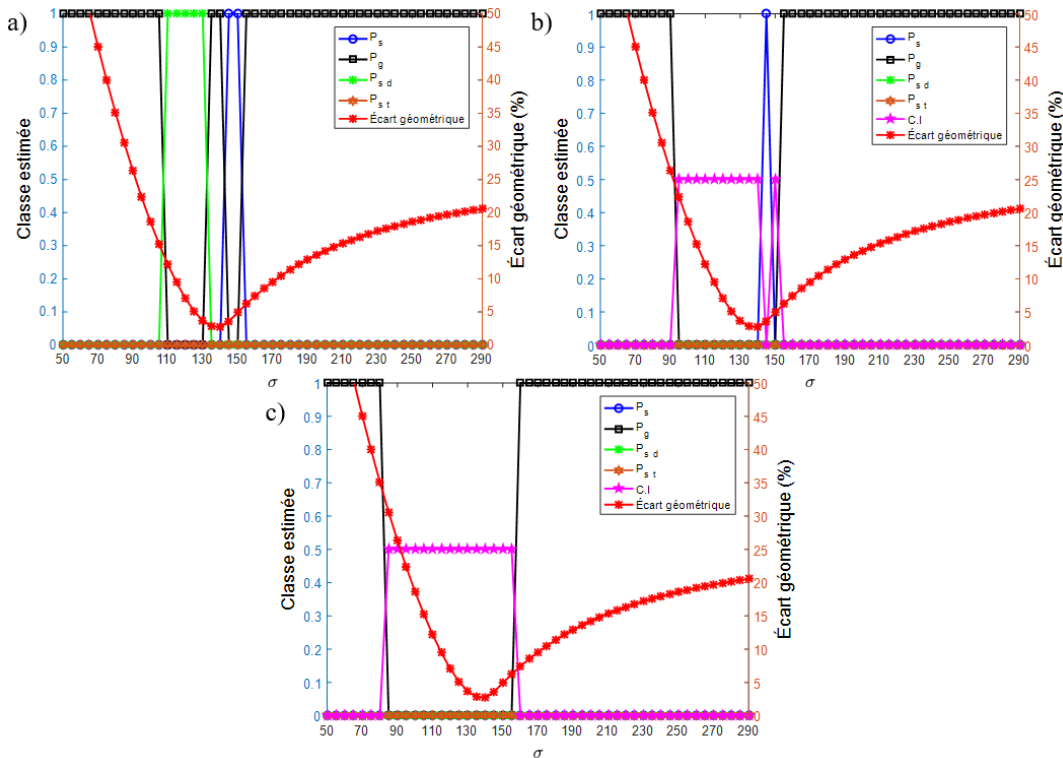


FIGURE 3.33 – Évolution des profils identifiés P_s , P_g , P_{sd} et P_{st} par le PMC classifieur $Class4_a$ en fonction du paramètre σ du profil gaussien pour un seuil d'acceptabilité fixé respectivement à (a) 50%, (b) 70% et (c) 90%.

Comme attendu, la sortie P_g est identifiée sur plusieurs valeurs de σ ; et ce indépendamment du seuil d'acceptabilité. Pour un seuil d'acceptabilité de 50%, l'identification d'un profil donné autour de $\sigma = 140$ varie. En effet, le profil estimé passe de P_{sd} , P_g à P_s avec le degré de confiance choisi. Par contre, pour les 2 autres seuils d'acceptabilités, le classifieur affiche clairement qu'il ne peut identifier avec certitude dans une zone plus ou moins large, une forme donnée; ce qui peut être interprété par le fait que la forme en question ne peut être décrite avec précision par les classes en présence.

La figure 3.34 illustre l'évolution de la sortie P_s du $Class4_a$ avec celles du TOR_a et $Class3_a$ en fonction de σ comme dans le cas de l'étude précédente (figure 3.30). On constate les mêmes conclusions que précédemment, la courbe P_s du $Class4_a$ est sensiblement proche de celle du $Class3_a$. On note une largeur de détection plus faible pour ces 2 cas de figure. Par exemple, pour un seuil d'acceptabilité choisi à 50%, la valeur de $\Delta\sigma$ pour le $Class4_a$ est du même ordre

de grandeur que pour le $Class3_a$ ($\Delta\sigma = 10.5$) contre $\Delta\sigma = 27.5$ pour le TOR_a .

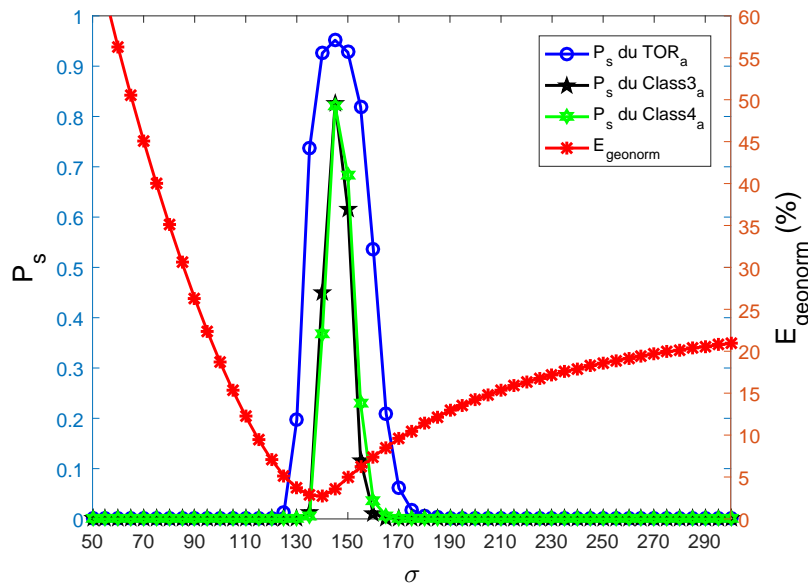


FIGURE 3.34 – Évolution de la sortie P_s des PMC classifieurs TOR_a , $Class3_a$ et $Class4_a$ en fonction du paramètre σ du profil gaussien. L'écart géométrique normalisé ($E_{geonorm}$) est également tracé (en rouge)

3.3.2.2.3 Résultats expérimentaux

L'étude expérimentale est effectuée sur les échantillons E_5 à E_{10} (figure 3.3). Les signatures optiques de ces derniers sont fournies en entrée de ce PMC. Le résultat de la classification est résumé dans la table 3.46. Nous avons également fait correspondre pour chaque résultat la classe identifiée suivant le seuil d'acceptabilité choisi.

TABLE 3.46 – Résultat de la classification des échantillons réels (E_5 , E_6 , E_7 , E_8 , E_9 et E_{10}) par la $Class4_a$.

	P_s	P_g	P_{sd}	P_{st}		50%	70%	90%	
E_5	0.01	0.95	0.04	0.00	⇒	E_5	P_g	P_g	P_g
E_6	0.00	0.87	0.12	0.01		E_6	P_g	P_g	×
E_7	0.00	0.00	0.00	1.00		E_7	P_{st}	P_{st}	P_{st}
E_8	0.00	0.01	0.60	0.39		E_8	P_{sd}	×	×
E_9	0.00	0.00	0.00	1.00		E_9	P_{st}	P_{st}	P_{st}
E_{10}	0.00	0.99	0.01	0.00		E_{10}	P_g	P_g	P_g

Le $Class4_a$ confirme les résultats obtenus par l'AFM (tables 3.25 et 3.26). Par contre, il donne un résultat assez net concernant l'échantillon E_5 qui est identifié à la classe P_g alors que l'AFM était plus ambiguë. Cependant, cette tendance avait déjà été notée dans les résultats du $Class3_a$ (table 3.42). D'une manière générale, l'identification des profils, bien que correcte, est

moins marquée que le $Class3_a$. En effet, si on prend l'exemple de l'échantillon E_8 , il ne sera détecté comme P_{sd} que dans le cas d'un seuil de 50% pour le $Class4_a$.

3.3.2.2.4 Exemple de caractérisation complète

L'échantillon E_7 est classé par le $Class4_a$ à la sortie P_{st} , nous allons alors procéder à la caractérisation complète de ce dernier. Le PMC régresseur de type RegPst correspondant au profil sinus tronqué est implémenté. Il contient 30 neurones cachés. L'apprentissage est effectué sur 3000 échantillons théoriques de forme sinus tronquée. La performance optimale calculée sur les données test sur chacune des sorties du PMC est donnée dans le tableau 3.47.

TABLE 3.47 – Performance de RegPst (une couche cachée composée de 30 neurones) après entraînement évaluée sur les données test (450 échantillons) fonctionnant exclusivement sur un profil sinus tronqué.

Performance de RegPst			
$E_{test}^\alpha = 0.17$	$E_{test}^b = 6.13 \text{ nm}$	$E_{test}^h = 1.24 \text{ nm}$	$E_{test}^{h_r} = 1.1 \text{ nm}$

La signature optique mesurée de l'échantillon E_7 est fournie en entrée de RegPst. Le résultat est donné dans tableau 3.48.

TABLE 3.48 – Résultat de caractérisation par RegPst de l'échantillon E_7 .

	RegPst
Paramètres géométriques	$\alpha = 0.61 \pm 0.17$ $b = 502.13 \pm 6.13 \text{ nm}$ $h = 139.79 \pm 1.24 \text{ nm}$ $h_r = 409.45 \pm 1.1 \text{ nm}$
Qualité de la caractérisation	$E_{I_s, I_c} = 0.0586$ $E_{geo} = 3.88 \times 10^3 \text{ nm}^2$ $E_{geonorm} = 8.42\%$

On remarque que les paramètres géométriques estimés sont inclus dans le domaine de variation définie (équation 3.18) sauf b (à la limite de la plage définie). L'épaisseur h_r de la couche résiduelle estimée est proche de celle obtenue par optimisation. L'erreur E_{I_s, I_c} calculée entre les signatures optiques est faible. Ces résultats sont conformes à ceux donnés par l'AFM.

Dans la figure 3.35, on représente les signatures optiques mesurées par rapport à celles reconstruites à partir des résultats du RegPst et à partir des mesures AFM. On observe que la signature optique mesurée et celle calculée à partir du résultat du RegPst sont relativement proches ; ce qui est confirmé par la valeur de E_{I_s, I_c} faible.

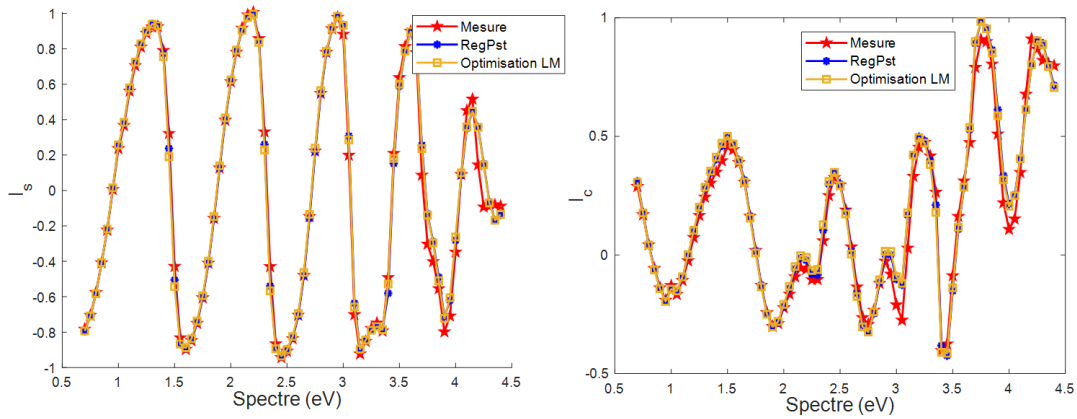


FIGURE 3.35 – Comparaison des signatures optiques simulées à partir des résultats de RegPst et à partir des résultats de l’optimisation des mesures AFM avec celles mesurées sur l’échantillon E_7 .

La figure 3.36 représente les profils estimés par le RegPst, les mesures AFM brutes et après traitement sur 4 périodes. Le RegPst montre un profil relativement proche de ceux obtenus par caractérisation AFM et par optimisation. Par ailleurs, on retrouve les mêmes constatations évoquées dans le paragraphe 3.3.2.1.2.4 sur les artefacts des mesures AFM brutes.

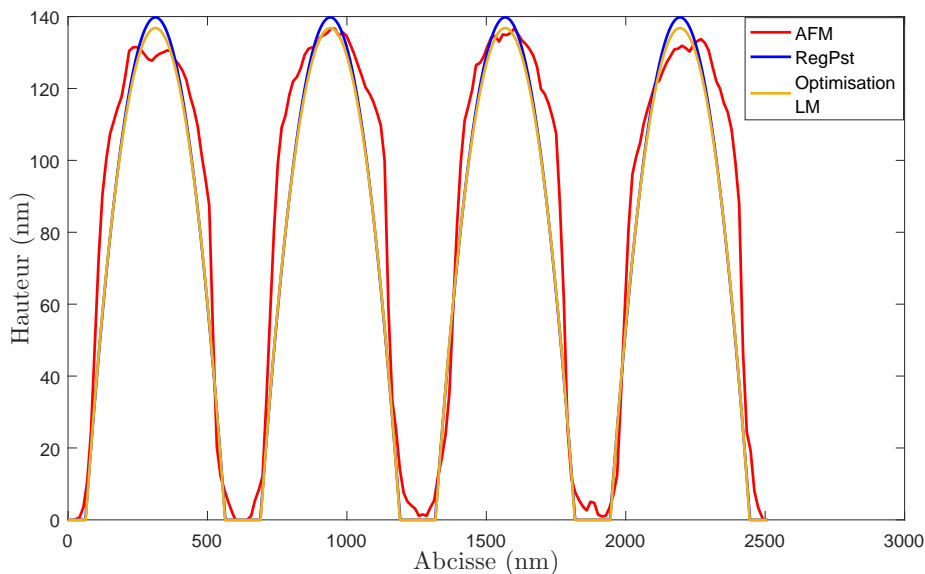


FIGURE 3.36 – Représentation des profils obtenus par : AFM (en rouge), RegPst (en bleu) et par optimisation de LM (en or) pour la caractérisation de l’échantillon E_7 .

3.3.2.3 Bilan sur la détermination du type de profil géométrique

Dans cette partie, nous avons mis en œuvre un réseau de neurones classifieur fonctionnant en amont de l’étape de caractérisation. Cette étude a permis de présenter la potentialité des PMCs à fournir une information complémentaire en exploitant la même signature optique utilisée en scattérométrie classique. Ainsi, la première application a montré la capacité d’un PMC à détecter une forme géométrique précise par exemple en ligne de production. La seconde application a permis de concevoir un classifieur permettant d’identifier une forme structurale parmi

un nombre prédéfini de profils. Elle permet ainsi de déterminer en amont, une des hypothèses inhérentes (hypothèse de forme) nécessaire à la caractérisation scattérométrique.

3.4 Méthode complémentaire dédiée aux réseaux de forme sinusoïdale

Dans cette partie, nous nous intéressons à une nouvelle approche complémentaire basée sur le calcul des harmoniques d'un profil géométrique pour la détection exclusive d'un réseau de forme sinusoïdale. En effet, nous allons mettre à profit la représentation fréquentielle de ce profil particulier pour qualifier le modèle.

3.4.1 Généralités sur les signaux périodiques

Un signal est une représentation physique d'une information de diverse nature : électronique, optique, acoustique, etc. Il permet d'analyser, de concevoir et de caractériser un système de traitement d'information.

Un signal périodique de période T peut être décomposé en une série de fonctions trigonométriques sinusoïdales appelée également décomposition en série de Fourier. Elle peut s'écrire sous la forme trigonométrique réelle suivante :

$$\begin{aligned} x(t) &= a_0 + \sum_{n=1}^{+\infty} a_n \cos(n\omega t) + b_n \sin(n\omega t) \\ &= a_0 + \sum_{n=1}^{+\infty} A_n \cos(n\omega t - \phi_n) \end{aligned} \quad (3.19)$$

avec $A_n = \sqrt{a_n^2 + b_n^2}$, $\phi_n = \arctan \frac{b_n}{a_n}$ et $\omega = 2\pi f$

où f est la fréquence du signal $f = \frac{1}{T}$

Le coefficient a_0 est la valeur moyenne du signal ou composante continue, le terme $A_n \cos(n\omega t - \phi_n)$ est appelé harmonique de rang n où A_n est son amplitude et ϕ_n sa phase. Le signal $x(t)$ peut être pair ($b_n = 0 \quad \forall n$) ou impair ($a_n = 0 \quad \forall n$).

En utilisant les formules d'Euler :

$$\cos(n\omega t) = \frac{e^{jn\omega t} + e^{-jn\omega t}}{2}$$

et

$$\sin(n\omega t) = \frac{e^{jn\omega t} - e^{-jn\omega t}}{2}$$

Cette décomposition de $x(t)$ peut s'écrire sous la forme exponentielle complexe :

$$x(t) = \sum_{n=-\infty}^{+\infty} c_n e^{jn\omega t} \quad (3.20)$$

Les c_n sont appelés coefficients de Fourier et se calculent à partir de la fonction $x(t)$:

$$c_n = \frac{1}{T} \int_0^T x(t) e^{-jn\omega t} dt \quad (3.21)$$

Ils peuvent s'écrire aussi sous la forme complexe :

$$c_n = |c_n| e^{j \arg(c_n)} \quad (3.22)$$

L'harmonique de rang n s'écrit alors :

$$\begin{aligned} u_n(t) &= c_{-n} e^{-jn\omega t} + c_n e^{jn\omega t} \\ &= 2|c_n| \cos(n\omega t + \arg(c_n)) \end{aligned} \quad (3.23)$$

L'harmonique de rang n possède donc une amplitude $2|c_n|$ et un déphasage $\arg(c_n)$.

Il apparait dans l'expression de $x(t)$ des termes pour les fréquences s'étendant de $-\infty$ à $+\infty$, d'où le nom de spectres bilatéraux.

En faisant l'analogie entre la forme trigonométrique réelle et la forme exponentielle complexe, on obtient la relation suivante :

$$|c_0| = a_0$$

et

$$|c_n| = \frac{A_n}{2}$$

Soit un signal $x(t)$ qui est intégrable et continue. Il admet alors une Transformée de Fourier (TF) définie par :

$$X(f) = \int_{-\infty}^{+\infty} x(t) e^{-j2\pi f t} dt \quad (3.24)$$

On peut également passer du domaine fréquentiel au domaine temporel, cela permet de calculer le signal $x(t)$ à partir $X(f)$.

$$x(t) = \int_{-\infty}^{+\infty} X(f) e^{j2\pi f t} df \quad (3.25)$$

C'est la Transformée de Fourier inverse.

On peut déterminer ces différents paramètres par la Transformée de Fourier Discrète (TFD). C'est une transformation qui permet de calculer le spectre d'un signal discret, obtenu

par échantillonnage d'un signal continu. En considérant un échantillonnage de la fonction $x(t)$ sur $[0, T]$, comportant N points, la période d'échantillonnage et la fréquence d'échantillonnage sont respectivement définies par :

$$T_e = \frac{T}{N} \quad (3.26)$$

$$f_e = \frac{1}{T_e} \quad (3.27)$$

Ainsi, la TFD d'un signal échantillonné $x(n)$ est donnée par :

$$X(k) = \sum_{n=0}^{N-1} x(n) e^{-j2\pi \frac{nk}{N}} \quad (3.28)$$

La TFD inverse correspond à :

$$x(n) = \frac{1}{N} \sum_{k=0}^{N-1} X(k) e^{j2\pi \frac{nk}{N}} \quad (3.29)$$

Dans le cadre de notre étude, la transformée de Fourier discrète est calculée numériquement par l'algorithme dit de Transformée de Fourier Rapide désigné par FFT (Fast Fourier Transform).

Si la condition de Shannon est vérifiée, c'est à dire la fréquence d'échantillonnage f_e est supérieure à 2 fois la fréquence maximale du signal continu alors les premières valeurs de la TFD sont exactement les coefficients de Fourier. En d'autres termes, le premier terme correspond à la valeur moyenne, le second au fondamental, le troisième à l'harmonique du rang 2 et ainsi de suite jusqu'au rang n .

Dans notre étude, le signal représente la fonction géométrique définissant le profil d'une forme géométrique donnée. En guise d'illustration, considérons la forme sinusoïdale parfaite de hauteur $h = 80 \text{ nm}$. Elle est définie sur une période $T = 627 \text{ nm}$. En appliquant la FFT sur ce profil, nous obtenons une décomposition constituée d'une valeur moyenne et d'un harmonique fondamental (figure 3.37).

La valeur moyenne est $c_0 = 39.96 \text{ nm}$. L'harmonique fondamental à $f_1 = 0.0016 \text{ (nm)}^{-1}$ possède une amplitude $c_1 = 20.01 \text{ nm}$ et une phase $\phi_1 = -3.14 \text{ rad}$.

Il apparaît dans la figure 3.37 que le spectre est représenté pour des fréquences s'étendant de $-\infty$ à $+\infty$. Dans le cadre de notre étude, on tiendra compte que du caractère symétrique du spectre dans la phase de reconstruction.

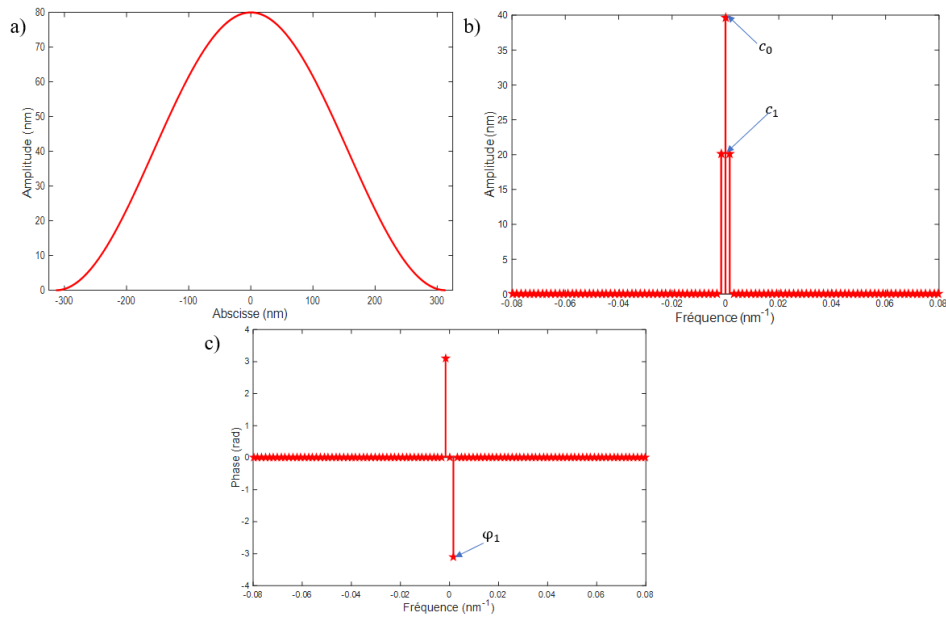


FIGURE 3.37 – Représentation d'un profil sinusoïdal dans le domaine temporel (a) et dans le domaine fréquentiel (amplitude (b) et phase (c)).

3.4.2 Décomposition d'un profil géométrique quelconque

Prenons le cas des formes dérivées du profil sinusoïdal (profil gaussien, profil sinusoïdal déformé et profil sinusoïdal tronqué) de même période $T = 627 \text{ nm}$ et présentant une hauteur de 80 nm :

- Le profil gaussien : La figure 3.38 et la table 3.49 donnent le spectre fréquentiel du profil défini par $\sigma = 110$ et $h = 80 \text{ nm}$.

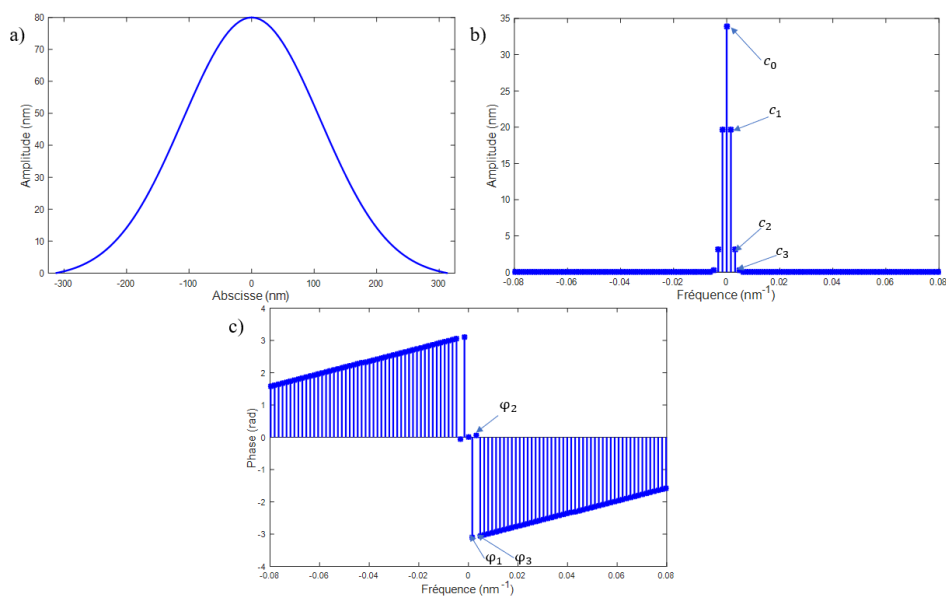


FIGURE 3.38 – Représentation d'un profil gaussien dans le domaine temporel (a) et dans le domaine fréquentiel (amplitude (b) et phase (c)).

TABLE 3.49 – Résultat de la décomposition fréquentielle du profil gaussien défini par $\sigma = 110$ et $h = 80$ nm.

Valeur moyenne	Harmoniques	
	Amplitudes	Phases
$c_0 = 33.89$ nm	$c_1 = 19.66$ nm	$\phi_1 = -3.14$ rad
	$c_2 = 3.17$ nm	$\phi_2 = 0.006$ rad
	$c_3 = 0.23$ nm	$\phi_3 = -3.13$ rad

— Le profil sinus déformé : La figure 3.39 et la table 3.50 montrent le spectre fréquentiel du profil défini par $\rho = 0.46$ et $h = 80$ nm.

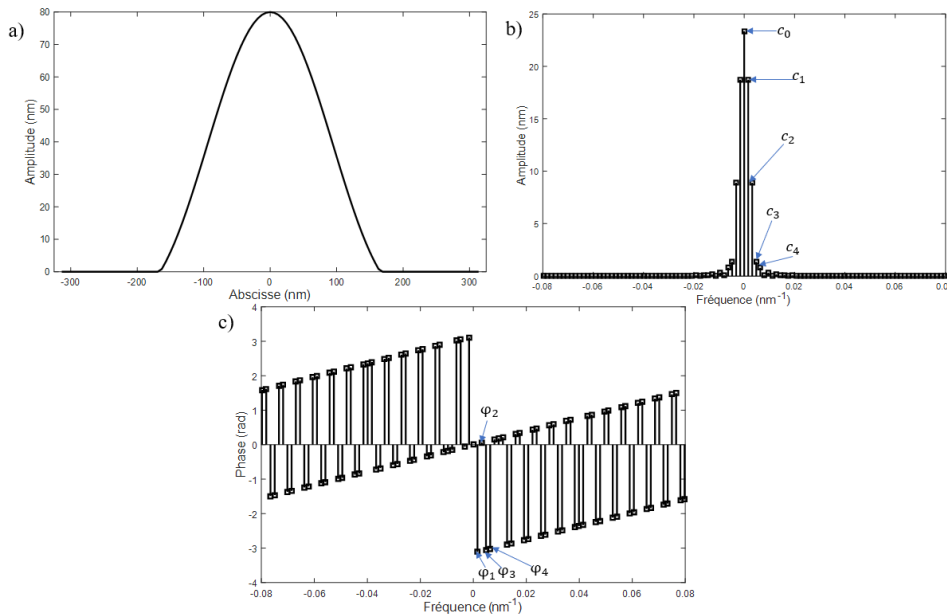


FIGURE 3.39 – Représentation d'un profil sinus déformé dans le domaine temporels (a) et dans le domaine fréquentiel (amplitude (b) et phase (c)).

TABLE 3.50 – Résultat de la décomposition fréquentielle du profil sinus déformé défini par $\rho = 0.46$ et $h = 80$ nm.

Valeur moyenne	Harmoniques	
	Amplitudes	Phases
$c_0 = 23.34$ nm	$c_1 = 18.72$ nm	$\phi_1 = -3.14$ rad
	$c_2 = 8.9$ nm	$\phi_2 = 0.006$ rad
	$c_3 = 1.33$ nm	$\phi_3 = -3.13$ rad
	$c_4 = 0.79$ nm	$\phi_4 = -3.12$ rad

- Le profil sinus tronqué : La figure 3.40 et la table 3.51 donnent le spectre fréquentiel du profil défini par $\alpha = 0.9$, $b = 300 \text{ nm}$ et $h = 80 \text{ nm}$.

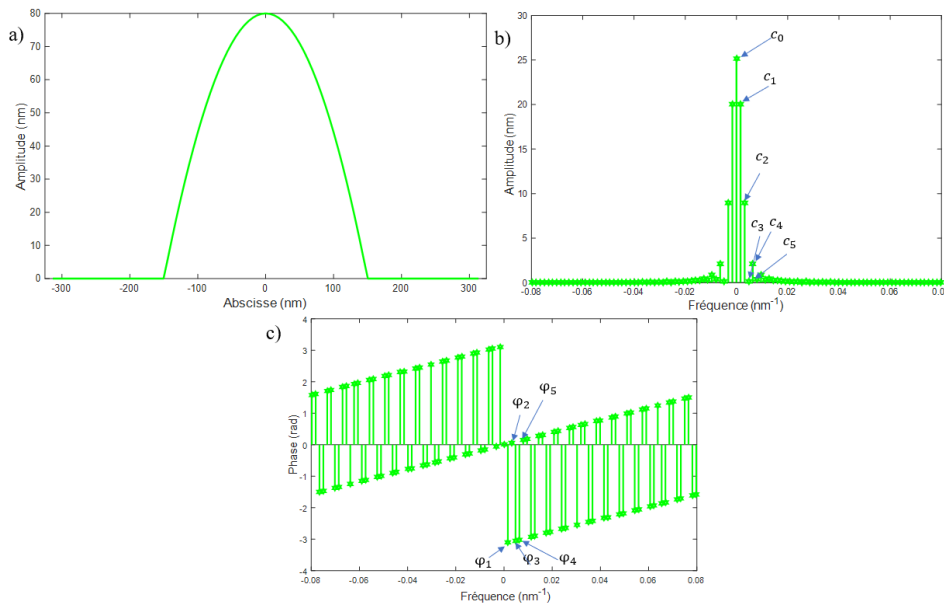


FIGURE 3.40 – Représentation d'un profil sinus tronqué dans le domaine temporel (a) et dans le domaine fréquentiel (amplitude (b) et phase (c)).

TABLE 3.51 – Résultat de la décomposition fréquentielle du profil sinus tronqué défini par $\alpha = 0.9$, $b = 300 \text{ nm}$ et $h = 80 \text{ nm}$.

Valeur moyenne	Harmoniques	
	Amplitudes	Phases
$c_0 = 25.17 \text{ nm}$	$c_1 = 20.04 \text{ nm}$	$\phi_1 = -3.14 \text{ rad}$
	$c_2 = 8.93 \text{ nm}$	$\phi_2 = 0.006 \text{ rad}$
	$c_3 = 0.16 \text{ nm}$	$\phi_3 = -3.13 \text{ rad}$
	$c_4 = 2.11 \text{ nm}$	$\phi_4 = -3.12 \text{ rad}$
	$c_5 = 0.38 \text{ nm}$	$\phi_5 = 0.01 \text{ rad}$

Tous ces profils étant de même période, on retrouve sensiblement la même valeur de l'harmonique fondamental. Par contre, l'apparition d'harmoniques de rang supérieur à 1 est caractéristique d'une forme qui s'éloigne d'un profil sinusoïdal parfait. Il paraît donc opportun d'utiliser cette information pour la détection d'un profil qui présenterait des défauts par rapport à cette forme cible particulière.

3.4.3 Reconstruction d'un profil

La reconstruction d'un profil géométrique à partir des données d'un signal fréquentiel complet ou partiel est alors envisageable par FFT inverse.

Considérons le profil sinus déformé de période $T = 627 \text{ nm}$ défini par $\rho = 0.46$ et $h = 80 \text{ nm}$ pour lequel nous avons précédemment calculé son spectre fréquentiel. Pour reconstruire le profil géométrique, on tronque le spectre respectivement sur les 2, 3 et 4 premières harmoniques. Les différents profils reconstruits par FFT inverse sont représentés sur la figure 3.41 et comparés à la forme initiale.

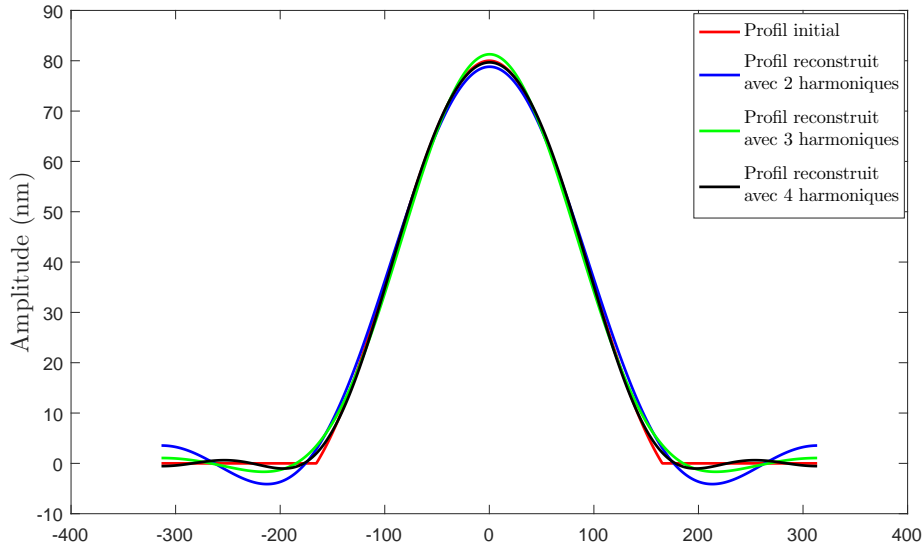


FIGURE 3.41 – Reconstruction d'un profil sinus déformé simulé ($\rho = 0.46$ et $h = 80 \text{ nm}$) (en rouge) à partir des 2 (en bleu), 3 (en vert) et 4 (en noir) premières harmoniques.

Nous constatons qu'en utilisant uniquement les 2 et 3 premiers harmoniques, les profils reconstruits affichent une forme proche de la forme cible avec des défauts plus ou moins marqués suivant la troncature. Ils présentent notamment des ondulations bien connues au pied des formes reconstruites dues à la forme discontinue du profil dans cette zone. Cette étude montre qu'il est possible de reconstruire un profil sans hypothèse de forme. Cependant, cela reste limité à certains type de profils. Dans la suite, le spectre fréquentiel sera utilisé exclusivement pour apporter une information qualitative sur la forme supposée et ne servira pas à la reconstruction du profil.

3.4.4 Critères de détection

Nous avons vu que l'apparition d'harmoniques dans le spectre fréquentiel est liée à la présence d'un profil non purement sinusoidal. Cependant, suivant le nombre et l'amplitude de ces harmoniques le profil peut toujours être considéré comme répondant aux exigences de la forme cible. Il est donc important de quantifier l'incidence de cette déviation sur le spectre afin de mettre en place une méthode permettant à terme une détection automatique répondant aux spécifications de l'expérimentateur. On cherche alors à mettre en place une relation liant la présence de ces différentes harmoniques à l'écart géométrique au profil sinus le plus proche. La catégorie des faux profils sinus est composée comme dans les paragraphes précédents des

profils gaussiens, sinus déformés et sinus tronqués. Dans la suite, on se focalisera sur l'étude des 4 premiers harmoniques du spectre.

3.4.4.1 Rapport entre les deux premiers harmoniques

Définissons β_1 comme le rapport entre l'amplitude de l'harmonique du rang 2 (c_2) et l'amplitude du fondamental (c_1).

$$\beta_1 = \frac{c_2}{c_1} \quad (3.30)$$

Ainsi, plus le profil géométrique tend vers une forme sinusoïdale parfaite plus le rapport tend vers 0.

Les 3 profils considérés comme des défauts de la forme sinusoïdale, sont simulés dans les domaines de variations suivants basés sur le modèle structurel des échantillons E_5 à E_{10} :

— Profil gaussien :

$$\begin{aligned} 30 < \sigma < 400 \\ 0 < h < 300 \quad nm \end{aligned} \quad (3.31)$$

— Profil sinus déformé :

$$\begin{aligned} -0.25 < \rho < 1 \\ 0 < h < 300 \quad nm \end{aligned} \quad (3.32)$$

— Profil sinus tronqué :

$$\begin{aligned} 0 < \alpha < 1 \\ 100 < b < 500 \quad nm \\ 0 < h < 300 \quad nm \end{aligned} \quad (3.33)$$

On représente alors pour chaque échantillon simulé le rapport β_1 en fonction de $E_{geonorm}$ qui représente pour chaque profil l'écart géométrique à la forme sinus la plus proche (figure 3.42).

On constate que le rapport β_1 varie quasi linéairement en fonction de $E_{geonorm}$. Au delà de $E_{geonorm} \geq 30\%$ les courbes de β_1 sont confondues. Par ailleurs, pour une forme donnée, 2 profils présentant le même écart géométrique peuvent conduire à 2 valeurs distinctes de β_1 . Cette constatation est cependant moins marquée pour le profil sinus déformé. Cela peut s'expliquer par les variations géométriques globalement positives ou négatives au profil sinusoïdal et le profil sinus parfait le plus proche (déjà mentionnés dans les paragraphes 3.3.1.3.2). Il ressort que les valeurs de β_1 dépendent du profil géométrique en question et ce, d'autant plus que l'écart surfacique est faible. Dans la table 3.52, nous avons regroupé pour chaque profil les conditions

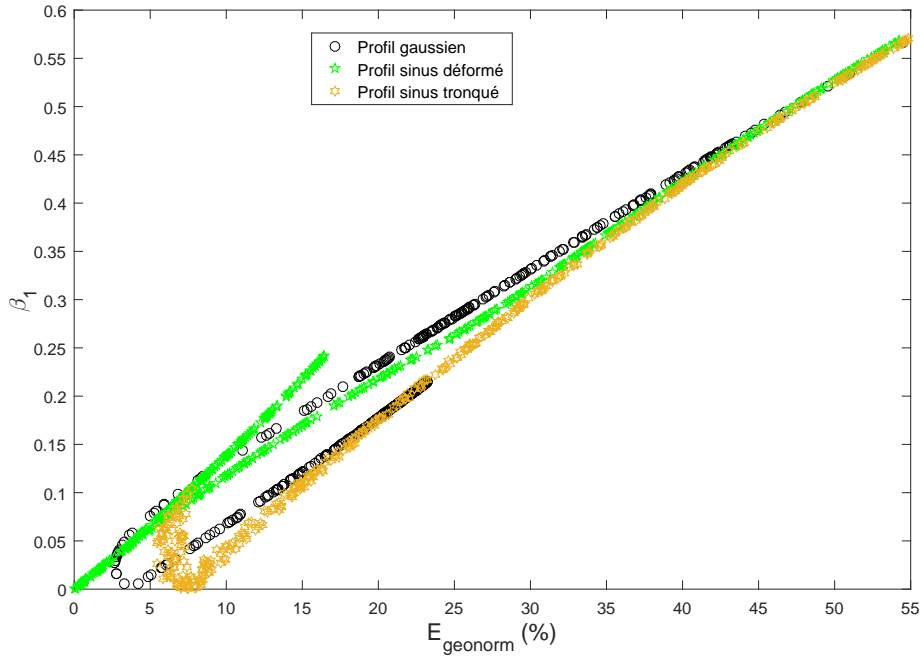


FIGURE 3.42 – Représentation de β_1 en fonction de l'écart surfacique pour différents profils : gaussien (en noir), sinus déformé (en vert) et sinus tronqué (en or)

sur β_1 qui correspondent à une déviation maximale donnée par rapport au profil cible évaluée en termes d'écart surfacique.

TABLE 3.52 – Conditions sur les valeurs de β_1 correspondants à une déviation donnée au profil cible en termes d'ecart géométrique pour chaque type de profil.

Écart géométrique	Profil gaussien	Profil sinus déformé	Profil sinus tronqué
$E_{geonorm} < 1\%$	Pas de profil correspondant	$\beta_1 \leq 0.01$ Condition nécessaire	Pas de profil correspondant
$E_{geonorm} < 5\%$	$\beta_1 \leq 0.02$ Condition suffisante	$\beta_1 \leq 0.07$ Condition nécessaire	Pas de profil correspondant
$E_{geonorm} < 10\%$	$\beta_1 \leq 0.07$ Condition suffisante	$\beta_1 \leq 0.12$ Condition suffisante	$\beta_1 \leq 0.04$ Condition suffisante
$E_{geonorm} < 15\%$	$\beta_1 \leq 0.12$ Condition suffisante	$\beta_1 \leq 0.17$ Condition suffisante	$\beta_1 \leq 0.11$ Condition nécessaire
$E_{geonorm} < 20\%$	$\beta_1 \leq 0.18$ Condition suffisante	$\beta_1 \leq 0.22$ Condition suffisante	$\beta_1 \leq 0.18$ Condition nécessaire
$E_{geonorm} < 25\%$	$\beta_1 \leq 0.29$ Condition suffisante	$\beta_1 \leq 0.27$ Condition suffisante	$\beta_1 \leq 0.24$ Condition nécessaire

Il faut préciser aussi que certaines valeurs de β_1 pour des profils gaussien et sinus tronqué sont quasiment proches de 0 (faibles valeurs de c_2) alors qu'elles correspondent à des écarts géométriques non négligeables.

Pour chaque déviation maximale, les valeurs de β_1 sont significativement différentes d'un type de profil à un autre. Il paraît alors difficile d'utiliser β_1 pour une détection fiable d'une déviation par rapport à une forme sinus indépendamment de la connaissance de la forme réelle.

3.4.4.2 Rapport entre les 4 premiers harmoniques

Le rapport β_2 est défini, cette fois ci, en tenant compte des amplitudes des harmoniques du rang 2 (c_2) au rang 4 (c_4) :

$$\beta_2 = \frac{c_2 + c_3 + c_4}{c_1} \quad (3.34)$$

La figure 3.43 montre les 3 courbes de β_2 en fonction de $E_{geonorm}$ pour les 3 profils précédents.

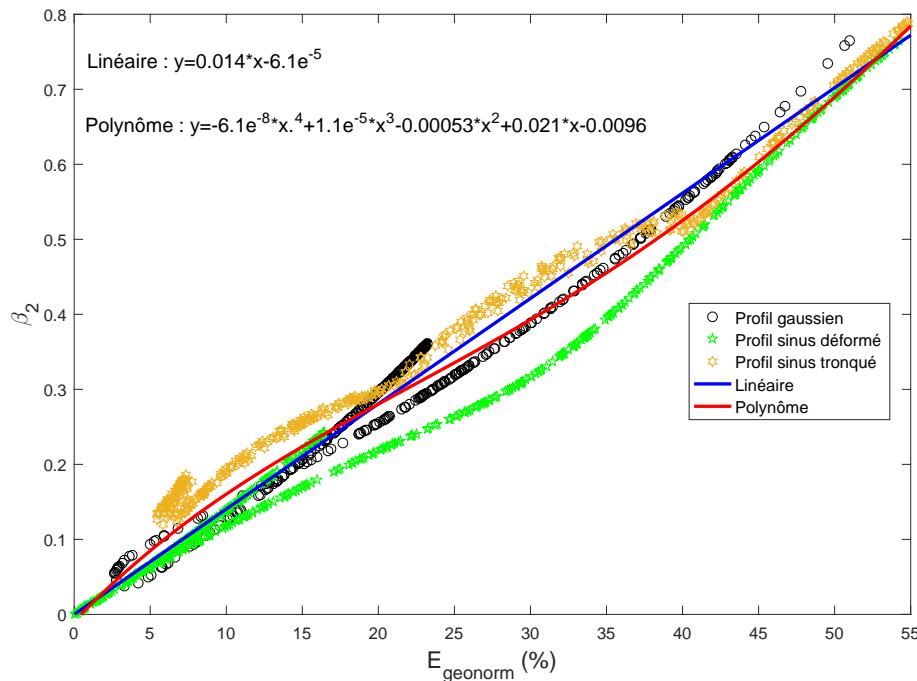


FIGURE 3.43 – Représentation de β_2 en fonction de l'écart surfacique pour différents profils : gaussien (en noir), sinus déformé (en vert), sinus tronqué (en or), un modèle linéaire (en bleu) et un modèle polynômial (en rouge)

Les valeurs de β_2 varie globalement de façon linéaire avec l'écart géométrique. La pente de la courbe est également plus importante en utilisant ce critère plutôt que β_1 donc β_2 est plus sensible aux écarts géométriques. On note également une réduction des effets correspondants aux écarts géométriques positifs et négatifs de part et d'autre du sinus parfait.

L'étude montre que les valeurs de β_2 restent cependant spécifiques d'une forme de profil à une autre mais beaucoup moins marqué que β_1 (table 3.53).

TABLE 3.53 – Conditions sur les valeurs de β_2 correspondants à une déviation donnée au profil cible en termes d'écart géométrique pour chaque type de profil.

Écart géométrique	Profil gaussien	Profil sinus déformé	Profil sinus tronqué
$E_{geonorm} < 1\%$	Pas de profil correspondant	$\beta_2 \leq 0.01$ Condition nécessaire	Pas de profil correspondant
$E_{geonorm} < 5\%$	$\beta_2 \leq 0.05$ Condition suffisante	$\beta_2 \leq 0.07$ Condition nécessaire	Pas de profil correspondant
$E_{geonorm} < 10\%$	$\beta_2 \leq 0.12$ Condition suffisante	$\beta_2 \leq 0.12$ Condition suffisante	$\beta_2 \leq 0.19$ Condition suffisante
$E_{geonorm} < 15\%$	$\beta_2 \leq 0.2$ Condition suffisante	$\beta_2 \leq 0.17$ Condition suffisante	$\beta_2 \leq 0.26$ Condition nécessaire
$E_{geonorm} < 20\%$	$\beta_2 \leq 0.25$ Condition suffisante	$\beta_2 \leq 0.22$ Condition suffisante	$\beta_2 \leq 0.3$ Condition nécessaire
$E_{geonorm} < 25\%$	$\beta_2 \leq 0.32$ Condition suffisante	$\beta_2 \leq 0.27$ Condition suffisante	$\beta_2 \leq 0.37$ Condition nécessaire

Ainsi, il est alors possible d'approcher la relation globale liant β_2 à $E_{geonorm}$. La table 3.54 présente les valeurs limites de β_2 à ne pas dépasser correspondant à une exigence de détection suivant les spécificités imposées par l'expérimentateur et ce, indépendamment du type de profil. Nous avons utiliser 2 modèles pour approcher au mieux cette dépendance : un modèle linéaire simple et une forme polynômiale d'ordre 4 (figure 3.43).

TABLE 3.54 – Conditions sur les valeurs de β_2 correspondants à une déviation donnée au profil cible en termes d'écart géométrique indépendamment du type de profils.

Écart géométrique	Critère de détection	
	Linéaire	Polynômiale
$E_{geonorm} < 1\%$	$\beta_2 \leq 0.01$	$\beta_2 \leq 0.01$
$E_{geonorm} < 5\%$	$\beta_2 \leq 0.07$	$\beta_2 \leq 0.08$
$E_{geonorm} < 10\%$	$\beta_2 \leq 0.14$	$\beta_2 \leq 0.15$
$E_{geonorm} < 15\%$	$\beta_2 \leq 0.21$	$\beta_2 \leq 0.22$
$E_{geonorm} < 20\%$	$\beta_2 \leq 0.28$	$\beta_2 \leq 0.27$
$E_{geonorm} < 25\%$	$\beta_2 \leq 0.35$	$\beta_2 \leq 0.33$

3.4.5 Méthode de détection de forme sinusoïdale

La méthode présentée dans ce paragraphe permet de détecter la présence d'une forme sinusoïdale. Le modèle de structures à caractériser sera celui des échantillons E_5 à E_{10} . Deux techniques sont alors envisagées :

- Un R.N régresseur $\text{Reg}x_h$ qui donne en sortie l'épaisseur h_r de la couche résiduelle et les caractéristiques des x harmoniques de la forme fréquentielle du profil (figure 3.44). Les critères β_1 et/ou β_2 sont ensuite calculés ;

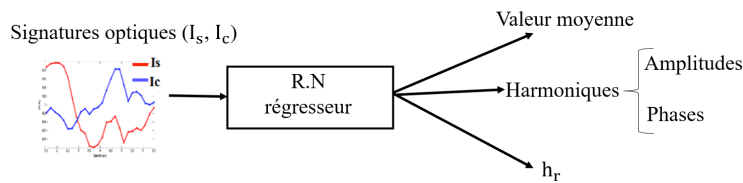


FIGURE 3.44 – Schéma de principe du PMC régresseur utilisé dans le cadre de la détection d'une forme sinusoïdale en utilisant le spectre fréquentiel du profil géométrique permettant d'estimer en sortie deux paramètres (h_r et les caractéristiques des harmoniques) à partir d'une signature optique fournie en entrée.

- Un R.N régresseur $\text{Reg}\beta_2$ qui donne en sortie directement l'épaisseur h_r de la couche résiduelle et directement un des 2 critères β_1 ou β_2 (figure 3.45).

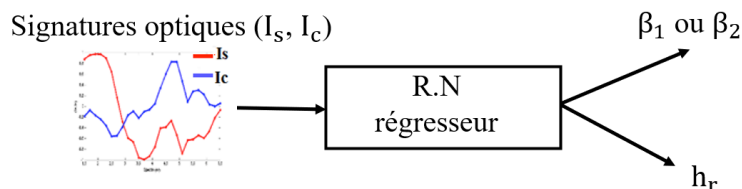


FIGURE 3.45 – Schéma de principe du PMC régresseur utilisé dans le cadre de la détection d'une forme sinusoïdale en utilisant le spectre fréquentiel du profil géométrique permettant d'estimer en sortie deux paramètres (h_r et β_1 ou β_2) à partir d'une signature optique fournie en entrée.

L'expérimentateur doit ensuite fixer, suivant les exigences expérimentales, la précision souhaitée sur la détection d'un profil sinusoïdal parfait. En d'autres termes il faut définir à l'instar de ce qui a été fait dans les paragraphes précédents un seuil d'acceptabilité lié cette fois ci au paramètre β_1 ou β_2 . Les tables 3.52 et 3.54 permettent de préciser les choses. Ainsi si on fixe arbitrairement dans la suite une contrainte correspondant à $E_{geonorm} < 10\%$, β_1 ne doit pas dépasser 0.07 pour un profil gaussien, 0.12 pour un profil sinus déformé et 0.04 pour un profil sinus tronqué. On prendra une valeur moyenne de 0.08. Pour β_2 , par contre, on peut utiliser l'approximation indépendante de la forme du profil. Ainsi il faudrait $\beta_2 \leq 0.14$ pour le modèle linéaire simple et $\beta_2 \leq 0.15$ pour la forme polynômiale d'ordre 4.

Les corpus d'apprentissage sont constitués des profils cibles et de tous les profils géométriques qualifiés de défauts (profil sinus, gaussien, sinus déformé et sinus tronqué) définis dans les paragraphes précédents indépendamment des contraintes particulières sur les frontières entre

les modèles. Nous utilisons les données définies dans le paragraphe 3.3.1.2 pour construire cette base de données.

3.4.5.1 Performances théoriques après apprentissage

Le premier PMC régresseur implémenté nommé Reg2h (figure 3.44) ne tient compte que des 2 premiers harmoniques. L'apprentissage est effectué sur 3000 échantillons équitablement repartis entre les types de profils (dont 750 échantillons pour chaque profil). Il contient 15 neurones cachés. La performance optimale calculée sur les données test est donnée dans la table 3.55. Il est difficile a priori d'interpréter la qualité de l'apprentissage à partir des erreurs estimées sur les harmoniques. Ainsi, une estimation de l'erreur sur β_1 est alors donnée en calculant à posteriori ce rapport avec les sorties du PMC.

TABLE 3.55 – Performance de Reg2h (une couche cachée composée 15 neurones) après entraînement évaluée sur les données test (450 échantillons).

Valeur moyenne	Harmoniques		Couche résiduelle	Critère de détection
	Amplitudes	Phases		
$E_{test}^{c_0} = 0.95 \text{ nm}$	$E_{test}^{c_1} = 0.22 \text{ nm}$ $E_{test}^{c_2} = 0.3 \text{ nm}$	$E_{test}^{\phi_1} = 1.27 \times 10^{-16} \text{ rad}$ $E_{test}^{\phi_2} = 0.3 \text{ rad}$	$E_{test}^{h_r} = 0.91 \text{ nm}$	$E_{test}^{\beta_1} = 0.01$

Le PMC Reg4h, quant à lui, prend en compte les 4 premiers harmoniques. Compte tenu du nombre de sorties plus important, l'entraînement est effectué sur 8000 échantillons équitablement repartis entre les types de profils (dont 2000 échantillons pour chaque profil). L'augmentation du nombre de sorties nécessite également une architecture plus complexe. Le Reg4h contient donc 2 couches cachées respectivement de 10 et 25 neurones. La performance optimale calculée sur les données test sur chaque sortie du modèle Reg4h est fournie dans la table 3.56. Le calcul de β_2 est alors possible en complément de β_1 . Une estimation des erreurs occasionnées sur ces 2 paramètres est alors fournie.

TABLE 3.56 – Performance de Reg4h (deux couches cachées composées respectivement 10 et 25 neurones) après entraînement évaluée sur les données test (1200 échantillons).

Valeur moyenne	Harmoniques		Couche résiduelle	Critère de détection
	Amplitudes	Phases		
$E_{test}^{c_0} = 1.05$	$E_{test}^{c_1} = 0.34 \text{ nm}$ $E_{test}^{c_2} = 0.35 \text{ nm}$ $E_{test}^{c_3} = 0.33 \text{ nm}$ $E_{test}^{c_4} = 0.23 \text{ nm}$	$E_{test}^{\phi_1} = 5.26 \times 10^{-16} \text{ rad}$ $E_{test}^{\phi_2} = 0.25 \text{ rad}$ $E_{test}^{\phi_3} = 0.3 \text{ rad}$ $E_{test}^{\phi_4} = 0.44 \text{ rad}$	$E_{test}^{h_r} = 1.35 \text{ nm}$	$E_{test}^{\beta_1} = 0.02$ $E_{test}^{\beta_2} = 0.04$

Les critères β_1 et β_2 sont estimés correctement par les différents PMC ; ce qui permet d'envisager sereinement la détection automatique de la forme cible.

3.4.5.2 Résultats théoriques

3.4.5.2.1 Présentation des échantillons théoriques

Afin d'évaluer la performance de chacun des PMC entraînés précédemment, nous avons simulé 3 faux profils sinus définis sur une même période $T = 627 \text{ nm}$ tout en calculant l'écart géométrique théorique avec le sinus parfait le plus proche :

- Profil gaussien : $\sigma = 135$, $h = 130 \text{ nm}$ et $h_r = 250 \text{ nm}$;
- Profil sinus déformé : $\rho = 0.1$, $h = 130 \text{ nm}$ et $h_r = 250 \text{ nm}$;
- Profil sinus tronqué : $\alpha = 0.7$, $b = 370 \text{ nm}$, $h = 130 \text{ nm}$ et $h_r = 250 \text{ nm}$.

TABLE 3.57 – Résultat de caractérisation des 3 faux profils sinus (gaussien, sinus déformé et sinus tronqué) par la FFT.

		Profil gaussien	Profil sinus déformé	Profil sinus tronqué
Valeur moyenne (nm)		$c_0 = 64.24$	$c_0 = 58.44$	$c_0 = 50.54$
Harmoniques	Amplitudes (nm)	$c_1 = 31.41$	$c_1 = 32.5$	$c_1 = 35.37$
		$c_2 = 1.15$	$c_2 = 3.27$	$c_2 = 8.13$
		$c_3 = 0.48$	$c_3 = 0.002$	$c_3 = 4.03$
		$c_4 = 0.29$	$c_4 = 0.002$	$c_4 = 0.89$
	Phases (rad)	$\phi_1 = -3.14$	$\phi_1 = -3.14$	$\phi_1 = -3.14$
		$\phi_2 = 0.006$	$\phi_2 = 0.006$	$\phi_2 = 0.006$
		$\phi_3 = -3.13$	$\phi_3 = 0.009$	$\phi_3 = 0.009$
		$\phi_4 = -3.12$	$\phi_4 = 0.01$	$\phi_4 = -3.12$
Critère de détection		$\beta_1 = 0.03$ $\beta_2 = 0.06$	$\beta_1 = 0.1$ $\beta_2 = 0.1$	$\beta_1 = 0.23$ $\beta_2 = 0.36$
Valeur théorique de $E_{geonorm}$		$E_{geonorm} = 2.83\%$	$E_{geonorm} = 8.43\%$	$E_{geonorm} = 24.37\%$
$E_{geonorm}$ estimé par β_1 (valeur moyenne)		$E_{geonorm} < 5\%$	$E_{geonorm} < 15\%$	$E_{geonorm} < 25\%$
$E_{geonorm}$ estimé par le modèle linéaire		$E_{geonorm} = 4.29\%$	$E_{geonorm} = 7.14\%$	$E_{geonorm} = 25.71\%$
$E_{geonorm}$ estimé par le modèle polynômial		$E_{geonorm} = 4.08\%$	$E_{geonorm} = 6.64\%$	$E_{geonorm} = 26.43\%$

La table 3.57 montre le résultat de la FFT obtenu pour les 4 premiers harmoniques. Les critères théoriques β_1 et β_2 sont alors calculés et ajoutés dans le tableau. En se référant aux tables 3.52 et 3.54, nous avons ajouté l'écart $E_{geonorm}$ maximal correspondant à ces valeurs. L'écart $E_{geonorm}$ estimé à partir des valeurs de β_1 des profils gaussien et sinus tronqué est du

même ordre de grandeur que les valeurs exactes de l'écart théorique. Concernant le profil sinus déformé cette erreur est légèrement sur-dimensionnée. Pour le critère β_2 , les résultats de 2 modèles d'approximation (linéaire et polynômial) sont proches et sensiblement conformes aux valeurs exactes de l'écart géométrique sauf pour le profil gaussien ($E_{geonorm}$ de β_2 supérieur à la valeur exacte de l'écart). Ainsi, en tenant compte de la contrainte du début du paragraphe 3.4.5 ($E_{geonorm} < 10\%$), l'utilisation de β_1 montre que seul le profil gaussien est considéré comme sinus parfait et les autres seront rejetés. Par contre, en considérant les valeurs de β_2 , les résultats sont conformes avec les valeurs théoriques ; à savoir ; que seul le profil sinusoidal tronqué est rejeté.

Les valeurs estimées de $E_{geonorm}$ à partir de β_1 et β_2 dans ce cadre correspondent globalement aux valeurs théoriques. Le critère β_2 permet de donner une information plus précise et fiable conformément aux conclusions établies au paragraphe 3.4.4. Les signatures optiques correspondantes de ces 3 profils sont simulées, bruitées et fournies à l'entrée des différents PMC précédemment entraînés (Reg2h, Reg4h et Reg β_2).

3.4.5.2.2 Validation théorique de Reg2h

Les sorties Reg2h regroupées dans la table 3.58 sont relativement proches des valeurs théoriques et conformes aux performances estimées après apprentissage (table 3.55). Il en est de même pour les valeurs du critère β_1 calculées en post traitement qui correspondent à ceux obtenus théoriquement par la FFT. On retrouve donc les valeurs de $E_{geonorm}$ estimées à partir de ce dernier fidèles à ceux de la table 3.57. Ainsi en considérant la contrainte donnée en début de paragraphe 3.4.5, seul le profil gaussien est estimé comme un profil sinusoidal parfait. Le Reg2h permet de donner une très bonne estimation de β_1 mais comme cela a été évoqué, ce critère n'est pas assez précis pour discriminer correctement les profils. Il faut donc avoir recours au calcul de β_2 et, pour cela, estimer un plus grand nombre d'harmoniques

TABLE 3.58 – Résultat de caractérisation des 3 faux profils sinus (gaussien, sinus déformé et sinus tronqué) par Reg2h.

		Profil gaussien	Profil sinus déformé	Profil sinus tronqué
Valeur moyenne (nm)		$c_0 = 63.64 \pm 0.95$	$c_0 = 60.08 \pm 0.95$	$c_0 = 50.33 \pm 0.95$
Harmoniques	Amplitudes (nm)	$c_1 = 31.41 \pm 0.22$ $c_2 = 0.82 \pm 0.3$	$c_1 = 32.41 \pm 0.22$ $c_2 = 3.74 \pm 0.3$	$c_1 = 35.38 \pm 0.22$ $c_2 = 8.1 \pm 0.3$
	Phases (rad)	$\phi_1 = -3.14 \pm 1.27 \times 10^{-16}$ $\phi_2 = 0.82 \pm 0.3$	$\phi_1 = -3.14 \pm 1.27 \times 10^{-16}$ $\phi_2 = 0.008 \pm 0.3$	$\phi_1 = -3.14 \pm 1.27 \times 10^{-16}$ $\phi_2 = 0.01 \pm 0.3$
Couche résiduelle (nm)		$h_r = 250.65 \pm 0.91$	$h_r = 248.55 \pm 0.91$	$h_r = 250.18 \pm 0.91$
Critère de détection		$\beta_1 = 0.02 \pm 0.01$	$\beta_1 = 0.11 \pm 0.01$	$\beta_1 = 0.22 \pm 0.01$
$E_{geonorm}$ estimé		$E_{geonorm} < 5\%$	$E_{geonorm} < 15\%$	$E_{geonorm} < 25\%$

3.4.5.2.3 Validation théorique de Reg4h

Le résultat obtenu par le Reg4h est donné dans la table 3.59. Par souci de clarté, la valeur des caractéristiques de chaque harmonique ne sera pas affichée mais nous présenterons seulement la valeur de β_1 , β_2 et h_r estimées par le PMC.

TABLE 3.59 – Résultat de caractérisation des 3 faux profils sinusoidaux (gaussien, sinus déformé et sinus tronqué) par Reg4h.

	Profil gaussien	Profil sinus déformé	Profil sinus tronqué
Couche résiduelle	$h_r = 252.5 \pm 1.35 \text{ nm}$	$h_r = 250.48 \pm 1.35 \text{ nm}$	$h_r = 250 \pm 1.35 \text{ nm}$
Critère de détection	$\beta_1 = 0.02 \pm 0.02$	$\beta_1 = 0.1 \pm 0.02$	$\beta_1 = 0.22 \pm 0.02$
	$\beta_2 = 0.04 \pm 0.04$	$\beta_2 = 0.11 \pm 0.04$	$\beta_2 = 0.35 \pm 0.04$
$E_{geonorm}$ estimé par β_1	$E_{geonorm} < 5\%$	$E_{geonorm} < 15\%$	$E_{geonorm} < 25\%$
$E_{geonorm}$ estimé par le modèle linéaire	$E_{geonorm} = 2.86\%$	$E_{geonorm} = 7.86\%$	$E_{geonorm} = 25\%$
$E_{geonorm}$ estimé par le modèle polynômial	$E_{geonorm} = 2.9\%$	$E_{geonorm} = 7.31\%$	$E_{geonorm} = 25.64\%$

Les critères β_1 et β_2 calculés en post traitement sont clairement conformes à ceux obtenus par la FFT.

Les conclusions établis au paragraphe 3.4.5.2.1 et reprises dans la table 3.59 peuvent donc s'appliquer à partir des valeurs de β_1 et β_2 estimées par Reg4h. Ainsi en considérant la contrainte donnée en début de paragraphe 3.4.5, seul le profil sinusoidal tronqué est rejeté conformément aux objectifs de l'étude.

3.4.5.2.4 Validation théorique de Reg β_2

Les PMC précédents offrent l'avantage de fournir une partie du spectre qui pourrait être utile dans le cadre d'une reconstruction de profil par FFT inverse. Cependant, dans notre étude, nous n'exploiteront que les harmoniques permettant le calcul des critères β_1 et β_2 . Ainsi il paraît judicieux de mettre en place un PMC dédié exclusivement à la détermination du critère de détection. Dans ce cadre, nous avons implémenté une architecture qui donne directement en sortie la valeur de la couche résiduelle h_r et le seul critère β_2 (figure 3.45). Le nouveau PMC régresseur est entraîné sur 8000 échantillons théoriques équitablement repartis entre les types de profils (dont 2000 échantillons pour chaque profil) et contient seulement 10 neurones cachés. L'architecture est clairement simplifiée. La table 3.60 montre la performance sur les données test de chaque sortie. Les résultats sur l'estimation de h_r et β_2 sont nettement meilleurs que ceux affichés par Reg4h (table 3.56) pour une architecture simplifiée.

TABLE 3.60 – Performance de $\text{Reg}\beta_2$ (une couche cachée composée de 10 neurones) après entraînement évaluée sur les données test (1200 échantillons).

Performance de $\text{Reg}\beta_2$	
$E_{test}^{h_r} = 0.78 \text{ nm}$	$E_{test}^{\beta_2} = 0.02$

Comme dans le cas des PMC précédents, on teste la performance du $\text{Reg}\beta_2$ sur les signatures théoriques de 3 faux profils sinus (gaussien, sinus déformé et sinus tronqué) définis précédemment. Les résultats estimés sont donnés dans la table 3.61. L'écart géométrique estimé à partir de la valeur de β_2 pour chaque modèle d'approximation est également ajouté.

TABLE 3.61 – Résultats de caractérisation des 3 faux profils sinus (gaussien, sinus déformé et sinus tronqué) par le $\text{Reg}\beta_2$.

	Profil gaussien	Profil sinus déformé	Profil sinus tronqué
Couche résiduelle	$h_r = 250.78 \pm 0.78 \text{ nm}$	$h_r = 249.84 \pm 0.78 \text{ nm}$	$h_r = 249.44 \pm 0.78 \text{ nm}$
Critère de détection	$\beta_2 = 0.04 \pm 0.02$	$\beta_2 = 0.1 \pm 0.02$	$\beta_2 = 0.36 \pm 0.02$
$E_{geonorm}$ estimé par le modèle linéaire	$E_{geonorm} = 2.86\%$	$E_{geonorm} = 7.14\%$	$E_{geonorm} = 25.71\%$
$E_{geonorm}$ estimé par le modèle polynômial	$E_{geonorm} = 2.9\%$	$E_{geonorm} = 6.64\%$	$E_{geonorm} = 26.43\%$

Le $\text{Reg}\beta_2$ estime un β_2 et une épaisseur h_r relativement proche du résultat théorique obtenu par la FFT. L'écart $E_{geonorm}$ estimé par β_2 (pour les 2 modèles) est donc comme énoncé précédemment très proche de la valeur exacte.

Ces résultats permettent de valider l'utilisation des PMC pour estimer le critère choisi (β_1 ou β_2) dans le cadre d'une détection automatique de profil sinusoidal. Le paragraphe suivant fait état des premiers résultats expérimentaux.

3.4.5.3 Résultats expérimentaux

L'étude est effectuée sur les échantillons E_5 à E_{10} (figure 3.3) définis dans le paragraphe 3.1.2. Pour comparer le résultat des différents PMCs régresseurs ($\text{Reg}2h$, $\text{Reg}4h$ et $\text{Reg}\beta_2$) avec les mesures AFM, il est nécessaire de traiter les profils bruts obtenus.

La table 3.62 donne les valeurs de β_1 et β_2 obtenues à partir des profils bruts donnés par l'AFM. Cependant, il faut noter que les profils AFM ne sont pas homogène sur plusieurs périodes (déjà évoqué au paragraphe 3.3.2.1.2.4). Par conséquent, les valeurs de β_1 et β_2 sont légèrement différentes d'une période à une autre. Nous avons donc calculé la valeur moyenne de β_1 et β_2 sur les 4 périodes disponibles. L'écart $E_{geonorm}$ correspondant à chacune des valeurs de β_1 et β_2 calculées est ajouté. Pour certains échantillons (E_6 , E_9 et E_{10}), l'estimation de $E_{geonorm}$

à partir de β_1 reste cependant assez différente de celle obtenue par β_2 . Ceci peut provenir de la limite du critère β_1 à détecter une forme cible sinusoïdale. En comparant avec les tables 3.25 et 3.26 on note tout de même que les valeurs de $E_{geonorm}$ estimées par les critères β_1 et β_2 sont éloignées des valeurs obtenues théoriquement en particulier pour les échantillons E_7 , E_8 et E_9 . Ainsi, il est clair que dans ce cadre ces critères ne sont pas suffisants. Ces échantillons ne seront donc pas testés par les PMCs.

TABLE 3.62 – Résultats des critères de détection β_1 et β_2 des échantillons E_5 à E_{10} calculés à partir des profils obtenus par mesures brutes AFM.

	E_5	E_6	E_7	E_8	E_9	E_{10}
Critère de détection (en moyenne)	$\beta_1 = 0.26$	$\beta_1 = 0.07$	$\beta_1 = 0.19$	$\beta_1 = 0.24$	$\beta_1 = 0.06$	$\beta_1 = 0.05$
	$\beta_2 = 0.37$	$\beta_2 = 0.27$	$\beta_2 = 0.35$	$\beta_2 = 0.38$	$\beta_2 = 0.32$	$\beta_2 = 0.25$
$E_{geonorm}$ correspondant à β_1	$E_{geonorm}$	$E_{geonorm}$	$E_{geonorm}$	$E_{geonorm}$	$E_{geonorm}$	$E_{geonorm}$
	< 20%	< 10%	< 25%	< 25%	< 10%	< 5%
$E_{geonorm}$ correspondant au modèle linéaire (β_2)	$E_{geonorm}$	$E_{geonorm}$	$E_{geonorm}$	$E_{geonorm}$	$E_{geonorm}$	$E_{geonorm}$
	= 26.43%	= 19.29%	= 25%	= 27.14%	= 22.86%	= 17.86%
$E_{geonorm}$ correspondant au modèle polynômial (β_2)	$E_{geonorm}$	$E_{geonorm}$	$E_{geonorm}$	$E_{geonorm}$	$E_{geonorm}$	$E_{geonorm}$
	= 27.22%	= 19.27%	= 25.64%	= 28.02%	= 23.24%	= 17.69%

Les signatures optiques mesurées des échantillons E_5 , E_6 et E_{10} (figure 3.3) sont introduites à l'entrée de chacun des PMCs régresseurs précédemment entraînés (Reg2h, Reg4h et Reg β_2). Les résultats sont donnés dans la table 3.63. Comme précédemment, l'affichage est amputé des valeurs des caractéristiques des harmoniques pour Reg2h et Reg4h ainsi que de l'épaisseur h_r pour ne garder que les critères β_1 et β_2 calculés soit directement par le PMC soit en post traitement.

TABLE 3.63 – Résultats de caractérisation des échantillons E_5 , E_6 et E_{10} par Reg2h, Reg4h et Reg β_2 .

	Reg2h	Reg4h	Reg β_2
E_5	$\beta_1 = 0.28 \pm 0.01$	$\beta_1 = 0.3 \pm 0.02$	$\beta_2 = 0.38 \pm 0.02$
		$\beta_2 = 0.45 \pm 0.04$	
E_6	$\beta_1 = 0.14 \pm 0.01$	$\beta_1 = 0.1 \pm 0.02$	$\beta_2 = 0.15 \pm 0.02$
		$\beta_2 = 0.18 \pm 0.04$	
E_{10}	$\beta_1 = 0.23 \pm 0.01$	$\beta_1 = 0.24 \pm 0.02$	$\beta_2 = 0.27 \pm 0.02$
		$\beta_2 = 0.31 \pm 0.04$	

Les β_1 et β_2 estimés par les différents PMCs sont du même ordre de grandeur. D'après les résultats précédents, on peut estimer que Reg2h est plus précis que Reg4h pour la détermination de β_1 (tables 3.55 et 3.56) alors que Reg β_2 est plus adapté pour β_2 (tables 3.56 et 3.60).

Cependant, les paramètres β_1 et β_2 calculés à partir des profils AFM bien que proches de ceux estimés par les PMCs régresseurs (Reg2h, Reg4h et Reg β_2) présentent des erreurs supérieures à celles évaluées théoriquement (tables 3.25 et 3.26). Gardons à l'esprit que les mesures AFM brutes ne constituent pas une photographie exacte du profil. Il s'agit de la convolution de la pointe par le profil réel. Par ailleurs, la valeur obtenue par les mesures AFM ont été réalisées ponctuellement sur une seule période. Or nous avons montré précédemment que le profil n'est pas homogène sur quelques périodes. Toutes ces raisons peuvent expliquer les écarts entre les valeurs estimées et théoriques des critères β_1 et β_2 .

Nous avons regroupé les résultats suivant le critère de détection choisi en se plaçant dans le cadre de contrainte choisie ($E_{geonorm} < 10\%$) pour une détection du profil sinusoidal. La table 3.64 montre les résultats suivant le critère β_1 estimé par Reg2h et Reg4h sous la forme d'une estimation de l'écart géométrique $E_{geonorm}$ maximal à la forme cible ainsi que de la décision prise dans le cas défini ci dessus. Les décisions affichées en gras correspondantes aux échantillons détectés par optimisation de mesure AFM comme proche d'une forme sinusoidale dans la contrainte choisie ($E_{geonorm} < 10\%$). Tous les échantillons testés sont rejetés conformément au résultats des tables 3.25 et 3.26. Plus précisément, si on compare avec les valeurs de $E_{geonorm}$ obtenues à partir des résultats de l'AFM (table 3.62), l'estimation de l'écart géométrique à partir du paramètre β_1 de ce dernier reste assez grossière. Cela confirme que ce critère n'est pas suffisant pour réaliser une détection fiable d'une forme cible sinusoidale.

TABLE 3.64 – Détection des échantillons réels (E_5 , E_6 et E_{10}) à partir des résultats de β_1 estimés par Reg2h et Reg4h.

		E_5	E_6	E_{10}
Reg2h	$E_{geonorm}$ estimé	$E_{geonorm} < 30\%$	$E_{geonorm} < 20\%$	$E_{geonorm} < 25\%$
	Détection du profil sinus	non	non	non
Reg4h	$E_{geonorm}$ estimé	$E_{geonorm} < 30\%$	$E_{geonorm} < 15\%$	$E_{geonorm} < 25\%$
	Détection du profil sinus	non	non	non

Pour le critère de détection β_2 , les résultats sont donnés sous le même format que précédemment dans les tables 3.65 et 3.66 respectivement pour Reg4h et Reg β_2 . Cette fois ci les valeurs de $E_{geonorm}$ estimées à partir des 2 approximations (linéaire et polynômiale) sont affichées.

Les résultats sont sensiblement les mêmes d'un PMC à l'autre. L'utilisation du PMC Reg β_2 semble apporter une estimation plus précise de l'écart géométrique au profil cible. Comme dans

TABLE 3.65 – Détection des échantillons réels (E_5 , E_6 et E_{10}) à partir du résultat de β_2 estimé par Reg_4h .

		E_5	E_6	E_{10}
Linéaire	$E_{geonorm}$ estimé	$E_{geonorm} = 32.14\%$	$E_{geonorm} = 13.57\%$	$E_{geonorm} = 22.14\%$
	Détection du profil sinus	non	non	non
Polynômial	$E_{geonorm}$ estimé	$E_{geonorm} = 33.48\%$	$E_{geonorm} = 13.06\%$	$E_{geonorm} = 22.45\%$
	Détection du profil sinus	non	non	non

TABLE 3.66 – Détection des échantillons réels (E_5 , E_6 et E_{10}) à partir du résultat de β_2 estimé par Reg_{β_2} .

		E_5	E_6	E_{10}
Linéaire	$E_{geonorm}$ estimé	$E_{geonorm} = 27.14\%$	$E_{geonorm} = 10.71\%$	$E_{geonorm} = 19.29\%$
	Détection du profil sinus	non	non	non
Polynômial	$E_{geonorm}$ estimé	$E_{geonorm} = 28.02\%$	$E_{geonorm} = 10.11\%$	$E_{geonorm} = 19.27\%$
	Détection du profil sinus	non	non	non

le cas précédent, tous les échantillons sont rejetés suivant le critère choisi et conformément aux mesures AFM. Au regard de ce résultat, on note que le critère β_2 fournit une meilleure détection du profil cible sinusoidal. Les valeurs numériques de $E_{geonorm}$ des échantillons sont sensiblement proches de ceux issus des mesures brutes AFM (sauf pour E_6).

3.4.5.4 Bilan sur la méthode de détection de forme sinusoïdale

Dans cette étude, nous avons présenté un travail préliminaire sur une nouvelle approche pour détecter un profil sinus parfait en utilisant le spectre fréquentiel du profil géométrique comme critère. Cette méthode spécifique aux profils sinusoidaux s'avère prometteuse pour détecter une forme sinusoïdale. En effet, il n'est théoriquement pas nécessaire de définir en amont un nombre restreint de défauts de profil. Cependant il est nécessaire de trouver le critère adéquate qui permettrait une discrimination sans ambiguïté. A cet effet, l'emploi du critère β_2 , bien que donnant des résultats satisfaisants, ne semble pas optimal. Il est nécessaire de tester d'autres critères plus riches en information.

3.5 Conclusion

Nous avons présenté dans ce chapitre, la potentialité du PMC à identifier des motifs géométriques dans le cadre de la caractérisation scattérométrique. Nous avons montré la capacité du PMC à caractériser des structures diffractantes de diverses formes géométriques.

Nous avons ainsi présenté une méthode dite de pondération de profils qui permet de caractériser la forme déviée de sa cible lors d'un processus de fabrication. Cette approche est basée sur la combinaison de plusieurs résultats de caractérisation classique obtenus par PMC simple. Cette méthode est efficace pour de faibles déviations.

Ensuite, nous avons testé les potentialités de la classification de formes géométriques en amont du processus de scattérométrie. Nous avons montré à travers cette étude, la possibilité de certifier ou non la forme de l'échantillon choisie en plaçant en amont de l'étape de caractérisation un PMC fonctionnant en mode classifieur. Ainsi, une première application est conçue permettant de détecter une forme donnée. Dans notre cas, nous avons utilisé les formes rectangulaires et sinusoïdales comme formes cibles. La seconde application concerne l'identification d'une forme géométrique parmi un nombre limité de profils. Des tests théoriques et expérimentaux sont réalisés sur des réseaux de diffraction de diverses formes géométriques. Les 2 approches utilisant des PMC ont montré de très bons résultats et peuvent être utilisées par exemple dans le cadre d'un contrôle en ligne de production ou dans le cadre d'une détection de défaut ou dans l'objectif d'identifier une forme cible.

Pour finir, nous avons proposé une nouvelle approche complémentaire pour détecter exclusivement des réseaux de forme sinusoïdale. Cette méthode est basée sur le calcul des harmoniques d'un profil géométrique. En effet, nous avons employé le spectre fréquentiel comme information pour détecter un profil géométrique particulier. Ainsi, nous avons étudié des critères de détection permettant de distinguer un bon d'un faux profil sinusoïdal. Les résultats obtenus sont encourageant et montrent que la méthode présentée est capable de détecter une forme présentant des caractéristiques définies en amont. Cependant, ces travaux doivent être poursuivis afin de valider complètement le processus en recherchant d'autres critères plus sensibles et en exploitant notamment d'autres résultats de caractérisation en plus de l'AFM.

Conclusion générale

Les travaux réalisés dans ce manuscrit s'inscrivent dans le cadre précis de la caractérisation scattérométrique. En effet, la maîtrise parfaite des différentes étapes de fabrication est importante afin d'écartier certaines dérives ou défauts qui peuvent survenir en cours de processus. Les applications visées concernent la détection et l'identification des formes géométriques grâce à l'apport d'informations qualitatives complémentaires fournies par un PMC fonctionnant en mode classifieur en amont du processus de caractérisation. Par exemple l'hypothèse de forme géométrique de la structure, fixée tout au long du processus de scattérométrie, peut ainsi être certifiée ou validée. Le réseau de neurones mis en place fonctionne sur la même signature optique utilisée pour la caractérisation. La solution proposée pour répondre aux besoins de caractérisation scattérométrique est donc entièrement neuronale.

Dans un premier temps, nous avons montré que le choix du modèle géométrique conditionne la qualité des résultats lors de la caractérisation en insistant sur les critères qui peuvent guider l'expérimentateur à accorder davantage de confiance au résultat (erreur calculée entre les signatures, résultats non inclus dans la base d'apprentissage, résultats erronés, ...). Ensuite, nous avons étudié le comportement du PMC sur des formes déviées de leur cible. Dans ce cadre, nous avons mis en place une méthode dite de pondération de profils basée sur la combinaison de plusieurs résultats de caractérisation classique. Cette méthode s'est avérée efficace sur des formes présentant de faibles déviations. Par exemple, elle a permis dans certains cas de réduire de 61% l'écart géométrique entre le profil réel et celui reconstruit par pondération par rapport au meilleur résultat obtenu par caractérisation classique. Elle estime également un profil proche du résultat de caractérisation AFM. Cependant, elle reste tout de même limitée à de petites variations autour des profils géométriques simples composants sa base de départ. Cette approche devient alors inadaptée pour des formes avec une plus grande déviation. L'alternative est d'avoir recours à d'autres approches comme la classification.

Nous avons ensuite montré l'apport de la classification de formes géométriques dans le cadre de la caractérisation. Nous avons ainsi mis en place deux approches suivant le type d'application visée : la détection ou l'identification d'une forme géométrique.

Dans la première application, nous avons montré la capacité du PMC à détecter une forme géométrique donnée. Cette approche, basée sur une classification binaire, peut être utilisée dans le cadre de la détection de défauts de forme ou dans le cadre de l'identification d'une forme cible précise. L'étude est réalisée sur des formes de profils de types rectangulaires et sinusoïdaux.

Le PMC classifieur mis en place offre une détection optimale du profil cible. Il affiche des performances théoriques proches de 98% pour les formes rectangulaires et sinusoïdales. Les validations expérimentales sont effectuées sur ces différents types d'échantillons. Ce classifieur binaire a montré qu'il était capable de détecter efficacement un profil cible. Il trouve donc son utilité dans une chaîne de fabrication et permet ainsi d'éviter toutes dérives d'un procédé de fabrication.

La seconde application a permis de concevoir un PMC classifieur qui permet de déterminer une forme structurelle parmi un nombre limité de profils. Elle est utilisée pour sélectionner automatiquement en amont, la forme géométrique la plus proche de l'échantillon présenté. Cette approche apporte ainsi une information plus qualitative nécessaire à toute caractérisation scattérométrique ultérieure. Le PMC classifieur de 3 classes affiche des performances théoriques proches de 98% pour des formes rectangulaires et sinusoïdales. Il a également montré qu'en utilisant un seuil d'acceptabilité plus élevé, la précision sur les différentes sorties augmente (par exemple elle est de 99.5% sur la sortie P_c et 100% sur les autres sorties pour les formes rectangulaires); ce qui permet de diminuer le nombre d'échantillons mal classés. Nous avons montré que ce type d'application arrive notamment à discriminer efficacement jusqu'à 4 classes de différentes formes à partir des signatures optiques. Ce dernier a affiche une performance théorique de 93.3%. La validation de la méthode est démontrée théoriquement puis sur des échantillons variés obtenus au cours d'un processus de fabrication.

Ainsi, ces 2 applications présentées confirment que le PMC peut être utilisé efficacement dans le cadre du processus de scattérométrie. Il peut par exemple être mis à profit dans un contrôle en ligne de production ou lors du suivi d'un processus de fabrication. L'ensemble de ces méthodes nécessite cependant la connaissance préalable du nombre et du type de défauts possibles que peut engendrer le procédé de fabrication.

Enfin, nous avons proposé une nouvelle méthode complémentaire basée sur le calcul des harmoniques pour détecter exclusivement des réseaux de type sinusoïdal. Des critères de détection permettant de distinguer un bon d'un mauvais profil sinusoïdal ont été mis en place. Les premiers résultats théoriques sont encourageants et montrent que cette approche est capable de détecter une forme ayant un écart géométrique au profil sinusoïdal le plus proche fixé par l'utilisation (par exemple $E_{geonorm} < 10\%$). Cependant, il est nécessaire de poursuivre ces travaux tout en exploitant davantage d'autres critères pouvant permettre de valider le processus.

Ce travail a permis de mettre en place une méthode qui ouvre la voie à d'autres applications. En effet, nous nous sommes volontairement bornés à l'étude de l'hypothèse de forme géométrique. Mais la signature optique contient d'autres informations comme la rugosité, les indices des matériaux, etc. Ainsi l'extraction de ces informations peut être utile lors du suivi d'un processus de fabrication pour modéliser au plus proche la structure de l'échantillon. La grande diversité de ces informations potentielles pourrait conduire à l'utilisation d'outils plus performant comme le réseau de neurones convolutif à apprentissage profond qui permettrait

d'améliorer significativement les performances d'extraction.

Bibliographie

- [1] Gordon E Moore. Gramming more components onto integrated circuits. *Electronics*, 38 :8, 1965.
- [2] Larry F Thompson, C Grant Willson, and Murray J Bowden. Introduction to microlithography. *Crystal research and technology*, 30(4), 1995.
- [3] Roderick R Kunz, Robert D Allen, William D Hinsberg, and Gregory M Wallraff. Acid-catalyzed single-layer resists for ArF lithography. *Optical Engineering*, 32(10) :2363–2367, 1993.
- [4] Arnaud Bazin. *Analyse de procédés de traitement plasma des résines photosensibles à 193 nm pour le développement de technologies CMOS sub-65 nm*. PhD thesis, Institut National Polytechnique de Grenoble-INPG, 2009.
- [5] LF Johnson, GW Kammlott, and KA Ingersoll. Generation of periodic surface corrugations. *Applied Optics*, 17(8) :1165–1181, 1978.
- [6] Valentin Gâté. *Écriture de motifs périodiques submicrométriques sur films TiO₂ sol-gel par lithographie interférométrique dynamique sur de grandes surfaces*. PhD thesis, Université Jean Monnet-Saint-Etienne, 2013.
- [7] Charles W Gwyn, R Stulen, D Sweeney, and D Attwood. Extreme ultraviolet lithography. *Journal of Vacuum Science & Technology B : Microelectronics and Nanometer Structures Processing, Measurement, and Phenomena*, 16(6) :3142–3149, 1998.
- [8] Burn J Lin. Optical lithography—present and future challenges. *Comptes Rendus Physique*, 7(8) :858–874, 2006.
- [9] JP Beasley and DG Squire. An electron beam maskmaker. *IEEE Transactions on Electron Devices*, 22(7) :376–384, 1975.
- [10] Stephen Y Chou, Peter R Krauss, and Preston J Renstrom. Imprint of sub-25 nm vias and trenches in polymers. *Applied physics letters*, 67(21) :3114–3116, 1995.
- [11] Céline Masclaux. *Étude de la structuration spontanée de films polymères en nano-impression thermique*. PhD thesis, Grenoble, 2012.
- [12] Pauline Voisin. *Lithographie de nouvelle génération par nanoimpression assistée par UV : étude et développement de matériaux et procédés pour l'application microélectronique*. PhD thesis, Université Joseph-Fourier-Grenoble I, 2007.
- [13] Ch Elachi and C Yeh. Periodic structures in integrated optics. *Journal of Applied Physics*, 44(7) :3146–3152, 1973.

- [14] Stephen W James and Ralph P Tatam. Optical fibre long-period grating sensors : characteristics and application. *Measurement science and technology*, 14(5) :R49, 2003.
- [15] Ralf Waldhäusl, Bernd Schnabel, Peter Dannberg, Ernst-Bernhard Kley, Andreas Bräuer, and Wolfgang Karthe. Efficient coupling into polymer waveguides by gratings. *Applied optics*, 36(36) :9383–9390, 1997.
- [16] J Vörös, JJ Ramsden, G Csúcs, I Szendrő, SM De Paul, M Textor, and ND Spencer. Optical grating coupler biosensors. *Biomaterials*, 23(17) :3699–3710, 2002.
- [17] Anne-Laure Fehrembach. *Réseaux résonnants à Bande Interdite Photonique, nouveaux filtres pour le DWDM*. PhD thesis, Université de droit, d'économie et des sciences-Aix-Marseille III, 2003.
- [18] M Niggemann, M Glatthaar, A Gombert, A Hinsch, and V Wittwer. Diffraction gratings and buried nano-electrodes—architectures for organic solar cells. *Thin solid films*, 451 :619–623, 2004.
- [19] Stéphane Robert. *Caractérisation optique de réseaux de diffraction submicroniques par des techniques génétique et neuronale*. PhD thesis, Saint-Etienne, 2003.
- [20] Joseph W Goodman. Fourier optics, roberts and company. *Englewood, CO*, 2005.
- [21] RF Millar. The rayleigh hypothesis and a related least-squares solution to scattering problems for periodic surfaces and other scatterers. *Radio Science*, 8(8-9) :785–796, 1973.
- [22] PM Van den Berg. Reflection by a grating : Rayleigh methods. *JOSA*, 71(10) :1224–1229, 1981.
- [23] Erwin G Loewen, Michel Nevière, and Daniel Maystre. On an asymptotic theory of diffraction gratings used in the scalar domain. *JOSA*, 68(4) :496–502, 1978.
- [24] R Petit. Etude théorique de la diffraction d'une onde plane et monochromatique par un réseau métallique infiniment conducteur. *Applied Optics*, 4(12) :1551–1554, 1965.
- [25] D Maystre. Sur la diffraction d'une onde plane par un réseau métallique de conductivité finie. *Optics Communications*, 6(1) :50–54, 1972.
- [26] T Delort and D Maystre. Finite-element method for gratings. *JOSA A*, 10(12) :2592–2601, 1993.
- [27] G Tayeb. The method of fictitious sources applied to diffraction gratings. *The Applied Computational Electromagnetics Society Journal (ACES)*, pages 90–100, 1994.
- [28] M Nevière, G Cerutti-Maori, and M Cadilhac. Sur une nouvelle méthode de résolution du problème de la diffraction d'une onde plane par un réseau infiniment conducteur. *Optics Communications*, 3(1) :48–52, 1971.
- [29] R Petit and M Cadilhac. Electromagnetic theory of gratings : some advances and some comments on the use of the operator formalism. *JOSA A*, 7(9) :1666–1674, 1990.
- [30] J Chandezon, G Raoult, and D Maystre. A new theoretical method for diffraction gratings and its numerical application. *Journal of Optics*, 11(4) :235, 1980.

- [31] Gérard Granet and Jean Chandezon. La méthode des coordonnées curvilignes appliquée à la diffraction par des réseaux dont le profil est donné par des équations paramétriques : application à la diffraction par un réseau cycloïdal. *Pure and Applied Optics : Journal of the European Optical Society Part A*, 6(6) :727, 1997.
- [32] MG Moharam and TK Gaylord. Rigorous coupled-wave analysis of planar-grating diffraction. *JOSA*, 71(7) :811–818, 1981.
- [33] Lifeng Li and Charles W Haggans. Convergence of the coupled-wave method for metallic lamellar diffraction gratings. *JOSA A*, 10(6) :1184–1189, 1993.
- [34] G Granet and B Guizal. Efficient implementation of the coupled-wave method for metallic lamellar gratings in TM polarization. *JOSA A*, 13(5) :1019–1023, 1996.
- [35] Robert Edward Lee. *Scanning electron microscopy and X-ray microanalysis*. 1993.
- [36] Claude Le Gressus. Microscopie électronique à balayage. *Techniques de l'ingénieur. Analyse et caractérisation*, 1(P865) :P865–1, 1995.
- [37] Ismail El Kalioubi. *Développement de la technique de scattérométrie neuronale dynamique*. PhD thesis, Université Grenoble Alpes, 2015.
- [38] Issam Gereige. *Contribution des réseaux de neurones dans le domaine de l'ellipsométrie : application à la scatterométrie*. PhD thesis, Université Jean Monnet-Saint-Etienne, 2008.
- [39] Jean-Claude Rivoal and Christian Frétigny. Microscopie à force atomique (AFM). *Techniques de l'ingénieur*, 1 :394, 2005.
- [40] Jérôme Thiault. *Étude par microscopie à force atomique en trois dimensions de l'évolution de la rugosité de bord de ligne lors de la fabrication d'une grille de transistor MOS*. PhD thesis, Université Joseph-Fourier-Grenoble I, 2007.
- [41] L Azarnouche. *Défis liés à la rugosité des motifs de résine photosensible*. PhD thesis, Ph. D. thesis, Univeristé de Grenoble, 2012.
- [42] Ilkka Kallioniemi, Jyrki Saarinen, and Erkki Oja. Optical scatterometry of subwavelength diffraction gratings : neural-network approach. *Applied optics*, 37(25) :5830–5835, 1998.
- [43] Ilkka JP Kallioniemi and Jyrki Saarinen. Optical scatterometry with neural network model for nondestructive measurement of submicron features. In *In-Line Characterization, Yield Reliability, and Failure Analyses in Microelectronic Manufacturing*, volume 3743, pages 33–40. SPIE, 1999.
- [44] Stéphane Robert, Alain Mure Ravaud, Stéphanie Reynaud, Sabine Fourment, Franck Carcenac, and Philippe Arguel. Experimental characterization of subwavelength diffraction gratings by an inverse-scattering neural method. *JOSA A*, 19(12) :2394–2402, 2002.
- [45] Joerg Bischoff, Horst Truckenbrodt, and Joachim J Bauer. Diffraction-analysis-based characterization of very fine gratings. In *Micro-optical Technologies for Measurement, Sensors, and Microsystems II and Optical Fiber Sensor Technologies and Applications*, volume 3099, pages 212–222. SPIE, 1997.

- [46] Joerg Bischoff, Joachim J Bauer, Ulrich Haak, Lutz Hutschenreuther, and Horst Trukenbrodt. Optical scatterometry of quarter-micron patterns using neural regression. In *Metrology, Inspection, and Process Control for Microlithography XII*, volume 3332, pages 526–537. SPIE, 1998.
- [47] Norbert Benesch, A Hettwer, C Schneider, L Pfitzner, and Heiner Ryssel. Phi-scatterometry for integrated linewidth control in DRAM manufacturing. In *2001 IEEE International Symposium on Semiconductor Manufacturing. ISSM 2001. Conference Proceedings (Cat. No. 01CH37203)*, pages 129–132. IEEE, 2001.
- [48] Brooke S Stutzman, Hsu-Ting Huang, and Fred L Terry Jr. Two-channel spectroscopic reflectometry for in situ monitoring of blanket and patterned structures during reactive ion etching. *Journal of Vacuum Science & Technology B : Microelectronics and Nanometer Structures Processing, Measurement, and Phenomena*, 18(6) :2785–2793, 2000.
- [49] Xinhui Niu, Nickhil H Jakatdar, Junwei Bao, Costas J Spanos, and Sanjay K Yedur. Specular spectroscopic scatterometry in DUV lithography. In *Metrology, Inspection, and Process Control for Microlithography XIII*, volume 3677, pages 159–168. SPIE, 1999.
- [50] Jerome Hazart, Gilles Grand, Philippe Thony, David Herisson, Stephanie Garcia, and Oliver Lartigue. Spectroscopic ellipsometric scatterometry : sources of errors in critical dimension control. In *Process and Materials Characterization and Diagnostics in IC Manufacturing*, volume 5041, pages 9–20. SPIE, 2003.
- [51] Issam Gereige, Stéphane Robert, Gérard Granet, Damien Jamon, and Jean-Jacques Rousseau. Rapid control of submicrometer periodic structures by a neural inversion from ellipsometric measurement. *Optics communications*, 278(2) :270–273, 2007.
- [52] Grant R Fowles. *Introduction to modern optics*. Courier Corporation, 1989.
- [53] Enric Garcia-Caurel, Antonello De Martino, and Bernard Drevillon. Spectroscopic mueller polarimeter based on liquid crystal devices. *Thin Solid Films*, 455 :120–123, 2004.
- [54] Robert W Schmieder. Stokes-algebra formalism. *JOSA*, 59(3) :297–302, 1969.
- [55] S Babin, L Doskolovich, E Kadomina, I Kadomin, and S Volotovskiy. Restoring pattern CD and cross-section using scatterometry : various approaches. In *Metrology, Inspection, and Process Control for Microlithography XXIII*, volume 7272, pages 1192–1200. SPIE, 2009.
- [56] Issam Gereige, Stéphane Robert, Sylvie Thiria, Fouad Badran, Gérard Granet, and Jean Jacques Rousseau. Recognition of diffraction-grating profile using a neural network classifier in optical scatterometry. *JOSA A*, 25(7) :1661–1667, 2008.
- [57] Jinlong Zhu, Shiyuan Liu, Chuanwei Zhang, Xiuguo Chen, and Zhengqiong Dong. Classification and recognition of diffraction structures using support vector machine in optical scatterometry. In *Metrology, Inspection, and Process Control for Microlithography XXVI*, volume 8324, pages 885–895. SPIE, 2012.

- [58] Jinlong Zhu, Shiyuan Liu, Chuanwei Zhang, Xiuguo Chen, and Zhengqiong Dong. Identification and reconstruction of diffraction structures in optical scatterometry using support vector machine method. *Journal of Micro/Nanolithography, MEMS, and MOEMS*, 12(1) :013004, 2013.
- [59] Jinlong Zhu, Shiyuan Liu, Hao Jiang, Chuanwei Zhang, and Xiuguo Chen. Improved deep-etched multilayer grating reconstruction by considering etching anisotropy and abnormal errors in optical scatterometry. *Optics Letters*, 40(4) :471–474, 2015.
- [60] Jinlong Zhu, Hao Jiang, Chuanwei Zhang, Xiuguo Chen, and Shiyuan Liu. Application of support vector machine for the fast and accurate reconstruction of nanostructures in optical scatterometry. *arXiv preprint arXiv :1905.06857*, 2019.
- [61] Issam Gereige, Stéphane Robert, and Jessica Eid. Automatic detection of photoresist residual layer in lithography using a neural classification approach. *Microelectronic engineering*, 97 :29–32, 2012.
- [62] Zaki SF Philippe, Stéphane Robert, and Bernard Bayard. Automatic inspection of a residual resist layer by means of self-organizing map. *Optical Engineering*, 55(5) :054106, 2016.
- [63] Zaki Sabit Fawzi, Stéphane Robert, Ismail El Kalyoubi, and Bernard Bayard. Analysis and detection of an incorrect profile shape in a classical scatterometric process. *The European Physical Journal Applied Physics*, 76(3) :31001, 2016.
- [64] Emil Agocs, Bernd Bodermann, Sven Burger, Gaoliang Dai, Johannes Endres, Poul-Erik Hansen, Lars Nielson, Morten Hannibal Madsen, Sebastian Heidenreich, Michael Krumrey, et al. Scatterometry reference standards to improve tool matching and traceability in lithographical nanomanufacturing. In *Nanoengineering : Fabrication, Properties, Optics, and Devices XII*, volume 9556, pages 153–164. SPIE, 2015.
- [65] Issam Gereige, Stéphane Robert, M Stchakovsky, Damien Jamon, Frederic Celle, S Reynaud, Jean-Claude Pommier, and Yves Jourlin. Ellipsometric characterization of photoresist gratings using artificial neural network. *physica status solidi c*, 5(5) :1133–1136, 2008.
- [66] Haskell B Curry. The method of steepest descent for non-linear minimization problems. *Quarterly of Applied Mathematics*, 2(3) :258–261, 1944.
- [67] Charles G Broyden. A class of methods for solving nonlinear simultaneous equations. *Mathematics of computation*, 19(92) :577–593, 1965.
- [68] Kenneth Levenberg. A method for the solution of certain non-linear problems in least squares. *Quarterly of applied mathematics*, 2(2) :164–168, 1944.
- [69] Donald W Marquardt. An algorithm for least-squares estimation of nonlinear parameters. *Journal of the society for Industrial and Applied Mathematics*, 11(2) :431–441, 1963.
- [70] James Matt Holden, Thomas Gubiotti, William A McGahan, Mircea V Dusa, and Ton Kiers. Normal-incidence spectroscopic ellipsometry and polarized reflectometry for mea-

- surement and control of photoresist critical dimension. In *Metrology, Inspection, and Process Control for Microlithography XVI*, volume 4689, pages 1110–1121. SPIE, 2002.
- [71] Hsu-Ting Huang and Fred L Terry Jr. Spectroscopic ellipsometry and reflectometry from gratings (scatterometry) for critical dimension measurement and in situ, real-time process monitoring. *Thin Solid Films*, 455 :828–836, 2004.
- [72] Holland Jh. Adaptation in natural and artificial systems. *Ann Arbor*, 1975.
- [73] David E Goldberg. Optimal initial population size for binary-coded genetic algorithms. In *Proceeding of the International Joint Conference on Artificial Intelligence*, volume 9, pages 588–592, 1985.
- [74] Gabriel Cormier and Roger Boudreau. Genetic algorithm for ellipsometric data inversion of absorbing layers. *JOSA A*, 17(1) :129–134, 2000.
- [75] Sébastien Soulan. *Développement de la scatterométrie dynamique pour le suivi en temps réel de procédés. Application à la microélectronique*. PhD thesis, Université Joseph-Fourier-Grenoble I, 2008.
- [76] Sébastien Soulan, Maxime Besacier, Tanguy Leveder, and Patrick Schiavone. Real-time profile shape reconstruction using dynamic scatterometry. In *Metrology, Inspection, and Process Control for Microlithography XXI*, volume 6518, pages 315–325. SPIE, 2007.
- [77] M El Kodadi, S Soulan, M Besacier, and P Schiavone. Real time scatterometry for profile control during resist trimming process. *Journal of Vacuum Science & Technology B : Microelectronics and Nanometer Structures Processing, Measurement, and Phenomena*, 27(6) :3232–3237, 2009.
- [78] Xiuguo Chen, Shiyuan Liu, Chuanwei Zhang, and Jinlong Zhu. Improved measurement accuracy in optical scatterometry using fitting error interpolation based library search. *Measurement*, 46(8) :2638–2646, 2013.
- [79] Shiming Wei and Lifeng Li. Measurement of photoresist grating profiles based on multiwavelength scatterometry and artificial neural network. *Applied optics*, 47(13) :2524–2532, 2008.
- [80] Stéphane Robert, Alain Mure-Ravaud, Sylvie Thiria, Méziane Yacoub, and Fouad Badran. Neural selection of the optimal optical signature for a rapid characterization of a submicrometer period grating. *Optics communications*, 238(4-6) :215–228, 2004.
- [81] Sabit Fawzi Philippe Zaki. *Classification par réseaux de neurones dans le cadre de la scatterométrie ellipsométrique*. PhD thesis, Lyon, 2016.
- [82] Rasheed MA Azzam, Nicholas Mitchell Bashara, and Stanley S Ballard. Ellipsometry and polarized light. *Physics Today*, 31(11) :72, 1978.
- [83] Frank Bernoux, Jean-Philippe Piel, Bernard Castellon, Christophe Defranoux, Jean-Hervé Lecat, and Pierre Boher. Ellipsométrie. théorie. *Techniques de l'ingénieur. Mesures et contrôle*, (R6490) :R6490–1, 2003.

- [84] SN Jaspersen and SE Schnatterly. An improved method for high reflectivity ellipsometry based on a new polarization modulation technique. *Review of Scientific Instruments*, 40(6) :761–767, 1969.
- [85] Chun-Hung Ko, Yi-sha Ku, and Nigel Smith. Correlation of scatterometry sensitivities to variation in device parameters. In *Metrology, Inspection, and Process Control for Microlithography XX*, volume 6152, pages 720–728. SPIE, 2006.
- [86] Peter E Hart, David G Stork, and Richard O Duda. *Pattern classification*. Wiley Hoboken, 2000.
- [87] P Macnaughton-Smith, WT Williams, MB Dale, and LG Mockett. Dissimilarity analysis : a new technique of hierarchical sub-division. *Nature*, 202(4936) :1034–1035, 1964.
- [88] Bruce Hendrickson, Robert W Leland, et al. A multi-level algorithm for partitioning graphs. *SC*, 95(28) :1–14, 1995.
- [89] Anil K Jain and Richard C Dubes. *Algorithms for clustering data*. Prentice-Hall, Inc., 1988.
- [90] Richard O Duda, Peter E Hart, and David G Stork. *Pattern classification and scene analysis*, volume 3. Wiley New York, 1973.
- [91] Lynn M Fletcher-Heath, Lawrence O Hall, Dmitry B Goldgof, and F Reed Murtagh. Automatic segmentation of non-enhancing brain tumors in magnetic resonance images. *Artificial intelligence in medicine*, 21(1-3) :43–63, 2001.
- [92] Jinchang Ren. ANN vs. SVM : Which one performs better in classification of MCCs in mammogram imaging. *Knowledge-Based Systems*, 26 :144–153, 2012.
- [93] Ahmed Kharrat, Nacéra Benamrane, Mohamed Ben Messaoud, and Mohamed Abid. Evolutionary support vector machine for parameters optimization applied to medical diagnostic. In *VISAPP*, pages 201–204, 2011.
- [94] Bernard Gosselin. Application de réseaux de neurones artificiels à la reconnaissance automatique de caractères manuscrits. *Faculté polytechnique de Mons*, 1996.
- [95] Khaoula Elagouni, Christophe Garcia, and Pascale Sébillot. A comprehensive neural-based approach for text recognition in videos using natural language processing. In *Proceedings of the 1st ACM International Conference on Multimedia Retrieval*, pages 1–8, 2011.
- [96] Tal Steinherz, Ehud Rivlin, and Nathan Intrator. Offline cursive script word recognition—a survey. *International Journal on Document Analysis and Recognition*, 2(2) :90–110, 1999.
- [97] Fabian Nasse, Christian Thureau, and Gernot A Fink. Face detection using GPU-based convolutional neural networks. In *International Conference on Computer Analysis of Images and Patterns*, pages 83–90. Springer, 2009.
- [98] Monowar H Bhuyan, Dhruva Kumar Bhattacharyya, and Jugal K Kalita. Network anomaly detection : methods, systems and tools. *Ieee communications surveys & tutorials*, 16(1) :303–336, 2013.

- [99] Sepp Hochreiter, Martin Heusel, and Klaus Obermayer. Fast model-based protein homology detection without alignment. *Bioinformatics*, 23(14) :1728–1736, 2007.
- [100] Zahra Hamou Mamar, Pierre Chainais, and Alex Aussem. Combining learning methods and time-scale analysis for defect diagnosis of a tramway guiding system. In *2008 16th Mediterranean Conference on Control and Automation*, pages 1780–1785. IEEE, 2008.
- [101] Chia-ying Lee and James Glass. A nonparametric bayesian approach to acoustic model discovery. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 1 : Long Papers)*, pages 40–49, 2012.
- [102] David Enke and Suraphan Thawornwong. The use of data mining and neural networks for forecasting stock market returns. *Expert Systems with applications*, 29(4) :927–940, 2005.
- [103] Z Guennoun, F Hamza, et al. Stocks portfolio optimization using classification and genetic algorithms. *Applied Mathematical Sciences*, 6(94) :4673–4684, 2012.
- [104] Jean-Paul Benzécri et al. *L'analyse des données*, volume 2. Dunod Paris, 1973.
- [105] Ian T. Jolliffe. Principal components in regression analysis. In *Principal component analysis*, pages 129–155. Springer, 1986.
- [106] P Demartines. Analyse de données par réseaux de neurones autoorganisés [ph. d. dissertation]. *Institut National Polytechnique de Grenoble, Valence, France*, 1995.
- [107] Teuvo Kohonen. Self-organized formation of topologically correct feature maps. *Biological cybernetics*, 43(1) :59–69, 1982.
- [108] Chantal Hajjar. *Cartes auto-organisatrices pour la classification de données symboliques mixtes, de données de type intervalle et de données discrétisées*. PhD thesis, Supélec, 2014.
- [109] Roy De Maesschalck, Delphine Jouan-Rimbaud, and Désiré L Massart. The mahalanobis distance. *Chemometrics and intelligent laboratory systems*, 50(1) :1–18, 2000.
- [110] Paul D McNicholas. Model-based classification using latent gaussian mixture models. *Journal of Statistical Planning and Inference*, 140(5) :1175–1181, 2010.
- [111] Emanuel Parzen. On estimation of a probability density function and mode. *The annals of mathematical statistics*, 33(3) :1065–1076, 1962.
- [112] Christopher M Bishop and Nasser M Nasrabadi. *Pattern recognition and machine learning*, volume 4. Springer, 2006.
- [113] C Delannoy. Un algorithme rapide de recherche de plus proches voisins. 1980.
- [114] Keinosuke Fukunaga and Patrenahalli M. Narendra. A branch and bound algorithm for computing k-nearest neighbors. *IEEE transactions on computers*, 100(7) :750–753, 1975.
- [115] Peter Hart. The condensed nearest neighbor rule (corresp.). *IEEE transactions on information theory*, 14(3) :515–516, 1968.
- [116] Jerome H Friedman, Forest Baskett, and Leonard J Shustek. An algorithm for finding nearest neighbors. *IEEE Transactions on computers*, 100(10) :1000–1006, 1975.

- [117] Baek S Kim and Song B Park. A fast k nearest neighbor finding algorithm based on the ordered partition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (6) :761–766, 1986.
- [118] Olivier Ondel. *Diagnostic par reconnaissance des formes : Application à un ensemble convertisseur-machine asynchrone*. PhD thesis, Ecole centrale de Lyon, 2006.
- [119] Dimitri P Bertsekas. Nonlinear programming. *Journal of the Operational Research Society*, 48(3) :334–334, 1997.
- [120] MO Stitson, JAE Weston, A Gammerman, V Vovk, and V Vapnik. Theory of support vector machines. *University of London*, 117(827) :188–191, 1996.
- [121] Slim Essid. *Classification automatique des signaux audio-fréquences : reconnaissance des instruments de musique*. PhD thesis, Université Pierre et Marie Curie-Paris VI, 2005.
- [122] Vladimir N Vapnik. The nature of statistical learning. *Theory*, 1995.
- [123] Richard Courant and David Hilbert. *Methods of mathematical physics : partial differential equations*. John Wiley & Sons, 2008.
- [124] Guobin Zhu and Dan G Blumberg. Classification using ASTER data and SVM algorithms ;: The case study of beer sheva, israel. *Remote sensing of Environment*, 80(2) :233–240, 2002.
- [125] Yoav Goldberg and Michael Elhadad. splitSVM : fast, space-efficient, non-heuristic, polynomial kernel computation for NLP applications. In *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics on Human Language Technologies : Short Papers*, pages 237–240, 2008.
- [126] Wenjian Wang, Zongben Xu, Weizhen Lu, and Xiaoyun Zhang. Determination of the spread parameter in the gaussian kernel for classification and regression. *Neurocomputing*, 55(3-4) :643–663, 2003.
- [127] S Sathiya Keerthi and Chih-Jen Lin. Asymptotic behaviors of support vector machines with gaussian kernel. *Neural computation*, 15(7) :1667–1689, 2003.
- [128] Bernhard E Boser, Isabelle M Guyon, and Vladimir N Vapnik. A training algorithm for optimal margin classifiers. In *Proceedings of the fifth annual workshop on Computational learning theory*, pages 144–152, 1992.
- [129] U Krebel. Pairwise classification and support vector machines, advances in kernel methods : support vector learning. *Cambridge, MA*, pages 255–268, 1999.
- [130] Edgar Osuna, Robert Freund, and Federico Girosi. Training support vector machines : an application to face detection. In *Proceedings of IEEE computer society conference on computer vision and pattern recognition*, pages 130–136. IEEE, 1997.
- [131] François Poulet, Thanh-Nghi Do, and Van Hoa Nguyen. SVM incrémental et parallèle sur gpu. In *EGC*, pages 103–114, 2009.
- [132] Jae Yeon Baek, Philippe Leray, Anne-Laure Charley, and Costas J Spanos. Real-time inspection system utilizing scatterometry pupil data. *Journal of Micro/Nanolithography, MEMS, and MOEMS*, 13(4) :041403, 2014.

- [133] Warren S McCulloch and Walter Pitts. A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5(4) :115–133, 1943.
- [134] Donald Olding Hebb. *The organization of behavior : a neuropsychological theory*. J. Wiley ; Chapman & Hall, 1949.
- [135] Frank Rosenblatt. Principles of neurodynamics : Perceptions and the theory of brain mechanisms. 1962. Publisher : Spartan.
- [136] Bernard Widrow and Marcian E Hoff. Adaptive switching circuits. Technical report, Stanford Univ Ca Stanford Electronics Labs, 1960.
- [137] Marvin Minsky and Seymour Papert. Perceptron : an introduction to computational geometry, 1969.
- [138] John J Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the national academy of sciences*, 79(8) :2554–2558, 1982.
- [139] David E Rumelhart, Geoffrey E Hinton, and Ronald J Williams. Learning internal representations by error propagation. Technical report, California Univ San Diego La Jolla Inst for Cognitive Science, 1985.
- [140] Yann Le Cun and Françoise Fogelman-Soulié. Modèles connexionnistes de l'apprentissage. *Intellectica*, 2(1) :114–143, 1987.
- [141] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet : A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009.
- [142] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6) :84–90, 2017.
- [143] Yann LeCun, Bernhard Boser, John S Denker, Donnie Henderson, Richard E Howard, Wayne Hubbard, and Lawrence D Jackel. Backpropagation applied to handwritten zip code recognition. *Neural computation*, 1(4) :541–551, 1989.
- [144] AL-Mansoori Saeed. Intelligent handwritten digit recognition using artificial neural network. *Int. Journal of Engineering Research and Applications*, 5(5) :46–51, 2015.
- [145] Xiaojun Tong, Zhu Wang, and Haining Yu. A research using hybrid RBF/elman neural networks for intrusion detection system secure model. *Computer physics communications*, 180(10) :1795–1801, 2009.
- [146] Reyadh Shaker Naoum and Zainab Namh Al-Sultani. Learning vector quantization (LVQ) and k-nearest neighbor for intrusion classification. *World of Computer Science and Information Technology Journal (WCSIT)*, 2(3) :105–109, 2012.
- [147] Pierre Roussel, Frédérique Moncet, Bruno Barrieu, and Arnaud 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*, 1 :919–926, 2001.

- [148] Isabelle Rivals, Daniel Canas, Leon Personnaz, and Gerard Dreyfus. Modeling and control of mobile robots and intelligent vehicles by neural networks. In *Proceedings of the Intelligent Vehicles' 94 Symposium*, pages 137–142. IEEE, 1994.
- [149] Hongying Wu. *Contribution to engine-out aircraft trajectory management and control*. PhD thesis, Toulouse, INSA, 2013.
- [150] Humberto González-Díaz, Isis Bonet, Carmen Terán, Erik De Clercq, Rafael Bello, Maria M García, Lourdes Santana, and Eugenio Uriarte. ANN-QSAR model for selection of anticancer leads from structurally heterogeneous series of compounds. *European journal of medicinal chemistry*, 42(5) :580–585, 2007.
- [151] Weiwei Zhang, Abul Hasan, Fadhel M Ghannouchi, Mohamed Helaoui, Yongle Wu, and Yuanan Liu. Novel calibration algorithm of multiport wideband receivers based on real-valued time-delay neural networks. *IEEE Transactions on Microwave Theory and Techniques*, 64(11) :3540–3548, 2016.
- [152] Ajit Kumar Sahoo and Siba Kumar Udgata. A novel ANN-based adaptive ultrasonic measurement system for accurate water level monitoring. *IEEE Transactions on Instrumentation and Measurement*, 69(6) :3359–3369, 2019.
- [153] Jungme Park, Dai Li, Yi L Murphey, Johannes Kristinsson, Ryan McGee, Ming Kuang, and Tony Phillips. Real time vehicle speed prediction using a neural network traffic model. In *The 2011 International Joint Conference on Neural Networks*, pages 2991–2996. IEEE, 2011.
- [154] Xiaolei Ma, Zhuang Dai, Zhengbing He, Jihui Ma, Yong Wang, and Yunpeng Wang. Learning traffic as images : a deep convolutional neural network for large-scale transportation network speed prediction. *Sensors*, 17(4) :818, 2017.
- [155] Caio César Teodoro Mendes, Vincent Frémont, and Denis Fernando Wolf. Exploiting fully convolutional neural networks for fast road detection. In *2016 IEEE International Conference on Robotics and Automation (ICRA)*, pages 3174–3179. IEEE, 2016.
- [156] Peikang Li, Yu Zang, Cheng Wang, Jonathan Li, Ming Cheng, Lun Luo, and Yao Yu. Road network extraction via deep learning and line integral convolution. In *2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pages 1599–1602. IEEE, 2016.
- [157] Yann Battie, Stéphane Robert, Issam Gereige, Damien Jamon, and Michel Stchakovsky. Demonstration of the feasibility of a complete ellipsometric characterization method based on an artificial neural network. *Applied optics*, 48(28) :5318–5323, 2009.
- [158] L Penlap Woguia, J Reche, M Besacier, P Gergaud, and G Rademaker. Data fusion by artificial neural network for hybrid metrology development. In *Metrology, Inspection, and Process Control for Semiconductor Manufacturing XXXV*, volume 11611, pages 647–658. SPIE, 2021.
- [159] Yann Morere. *Les réseaux de neurones récurrents*, 1998.

- [160] Bernard Widrow and Samuel D Stearns. Adaptive signal processing prentice-hall. *Englewood Cliffs, NJ*, 1985.
- [161] S Grossenberg. Studies of the mind and brain, 1982.
- [162] Degang Yang, Guo Chen, Hui Wang, and Xiaofeng Liao. Learning vector quantization neural network method for network intrusion detection. *Wuhan University Journal of Natural Sciences*, 12(1) :147–150, 2007.
- [163] Aleksandra Vuckovic, Vlada Radivojevic, Andrew CN Chen, and Dejan Popovic. Automatic recognition of alertness and drowsiness from EEG by an artificial neural network. *Medical engineering & physics*, 24(5) :349–360, 2002.
- [164] Mohamed Boubaker, Mohamed Akil, Khaled Ben Khalifa, Thierry Grandpierre, and Mohamed Hedi Bedoui. Implementation of an LVQ neural network with a variable size : algorithmic specification, architectural exploration and optimized implementation on fpga devices. *Neural Computing and Applications*, 19(2) :283–297, 2010.
- [165] Teuvo Kohonen. Learning vector quantization. In *Self-organizing maps*, pages 175–189. Springer, 1995.
- [166] Teuvo Kohonen. Learning vector quantization. In *Self-organizing maps*, pages 203–217. Springer, 1997.
- [167] Furkan Elmaz, Özgün Yücel, and Ali Yener Mutlu. Predictive modeling of biomass gasification with machine learning-based regression methods. *Energy*, 191 :116541, 2020.
- [168] Rami Matarneh, Svitlana Maksymova, Zhanna Deineko, and Vyacheslav Lyashenko. Building robot voice control training methodology using artificial neural net. 2017.
- [169] Ivan Lorencin, Nikola Anđelić, Josip Španjol, and Zlatan Car. Using multi-layer perceptron with laplacian edge detector for bladder cancer diagnosis. *Artificial Intelligence in Medicine*, 102 :101746, 2020.
- [170] George Cybenko. Approximation by superpositions of a sigmoidal function. *Mathematics of control, signals and systems*, 2(4) :303–314, 1989.
- [171] Kurt Hornik, Maxwell Stinchcombe, Halbert White, and Peter Auer. Degree of approximation results for feedforward networks approximating unknown mappings and their derivatives. *Neural computation*, 6(6) :1262–1275, 1994.
- [172] Mervyn Stone. Cross-validatory choice and assessment of statistical predictions. *Journal of the royal statistical society : Series B (Methodological)*, 36(2) :111–133, 1974.
- [173] Gaétan Monari. *Sélection de modèles non linéaires par "leave-one-out" : étude théorique et application des réseaux de neurones au procédé de soudage par points*. PhD thesis, Université Pierre et Marie Curie-Paris VI, 1999.
- [174] Sirpa Saarinen, Randall Bramley, and George Cybenko. Ill-conditioning in neural network training problems. *SIAM Journal on Scientific Computing*, 14(3) :693–714, 1993.
- [175] Lutz Prechelt. Early stopping-but when? In *Neural Networks : Tricks of the trade*, pages 55–69. Springer, 1998.

-
- [176] Eric Davalo, Patrick Naïm, and FM Clément. *Des réseaux de neurones*. Eyrolles Paris, 1990.
- [177] Jean-François Jodouin. *Les réseaux neuromimétiques : modèles et applications*. Hermès, 1994.
- [178] Enciclopedia Italiana. Articles de synthese ou de vulgarisation review papers, papers for the general public. *Neural Computation*, 10 :133–164, 1998.

Table des figures

1.1	Schéma de principe des différentes étapes d'un procédé de lithographie.	12
1.2	Principe de la lithographie optique par projection.	13
1.3	Représentation du phénomène de diffraction par réseau.	16
1.4	Représentation des deux états de polarisation de la lumière a)TE et b)TM [19].	18
1.5	Modélisation paramétrique des profils géométriques largement utilisés en scattérométrie : profil rectangulaire (a), profil symétrique trapézoïdal (b), profil rectangulaire arrondi en haut (c) et profil rectangulaire arrondi en bas (d).	19
1.6	Modélisation paramétrique des profil géométriques utilisés en scattérométrie : profil sinusoidal (a), profil gaussien (b), profil sinus déformé (c) et profil sinus tronqué (d).	20
1.7	Exemple de découpe d'un profil géométrique quelconque en couches rectangulaires d'épaisseurs différentes.	21
1.8	Procédure de découpe d'un profil quelconque.	22
1.9	Image CD-SEM d'une ligne de résine sur Silicium de période $T=189$ nm [37]. .	23
1.10	Exemple d'image d'un réseau en silice provenant du X-SEM (Department of Applied Physics and Photonics, Vrije Universiteit Brussel, Bruxelles) [38].	24
1.11	Exemple d'une image obtenue par AFM classique	25
1.12	Influence de la forme de la pointe AFM sur la mesure du profil géométrique. Les traits en pointillés et continus désignent respectivement le profil mesuré et le profil réel [38].	25
1.13	Schématisation d'une reconstruction de profil avec un CD-AFM [41].	25
1.14	Schéma illustrant le principe de la scattérométrie.	26
1.15	Représentation d'une onde plane incidente en réflexion et transmission avec un angle θ_i et d'indice n_0 et n_1	33
1.16	Montage de base d'un ellipsomètre classique.	34
1.17	Représentation schématique de l'ellipsomètre spectroscopique à modulation de phase (UVISEL de Jobin Yvon) [38].	37

1.18	Courbe de convergence selon l'ordre de troncature M pour $N = 40$ pour deux réseaux en résine définis par un profil trapèze ($b_1 = 30 \text{ nm}$, $b_2 = 180 \text{ nm}$ et $h = 390 \text{ nm}$) et un profil rectangulaire arrondi en haut ($\delta = 0.6$, $CD_h = 210 \text{ nm}$ et $h = 390 \text{ nm}$).	38
1.19	Courbe de convergence selon le nombre de couche N pour deux réseaux en résine définis un profil trapèze ($b_1 = 30 \text{ nm}$, $b_2 = 180 \text{ nm}$ et $h = 390 \text{ nm}$ avec $M = 10$) et un profil rectangulaire arrondi en haut ($\delta = 0.6$, $CD_h = 210 \text{ nm}$ et $h = 390 \text{ nm}$ avec $M = 10$).	38
1.20	Représentation des signatures optiques mesurées (rouge) et simulées à partir d'un profil trapézoïdal ($b_1 = 5 \text{ nm}$, $b_2 = 200 \text{ nm}$ et $h = 300 \text{ nm}$) (bleu).	39
1.21	Représentation graphique d'un profil trapézoïdal simulé ($b_1 = 30 \text{ nm}$, $b_2 = 180 \text{ nm}$ et $h = 390 \text{ nm}$) (rouge) et du profil reconstruit (bleu) à partir de la signature optique théorique bruitée.	40
2.1	Schéma des différentes étapes du système de classification d'un objet.	43
2.2	Exemple d'une structure arborescente hiérarchique pour différentes observations (a, b, c, d et e).	43
2.3	Exemple de topologie de classe obtenue suivant le choix de la distance utilisée dans un espace tridimensionnel : a) euclidienne, b) city-block et c) Chebychev [108].	46
2.4	Interprétation géométrique de la règle de décision de Bayes sur le classement d'un vecteur \mathbf{X} en 2 classes distinctes (C_1 et C_2).	48
2.5	Définition de l'hyperplan optimal et de la marge d'un classifieur SVM dans le cas d'une classification en 2 classes linéairement séparables dans un espace à 2 dimensions.	52
2.6	Représentation d'un neurone artificiel.	57
2.7	Exemple de fonctions d'activations g utilisées pour calculer la sortie d'un neurone.	58
2.8	Exemple d'architecture de réseau multicouche à deux couches cachées.	59
2.9	Exemple d'architecture de réseau à connexions locales.	59
2.10	Schéma expliquant le fonctionnement de l'apprentissage supervisé.	60
2.11	Architecture d'une SOM en 2D composée de n entrées et de $5 * 5$ neurones de sortie répartis suivant une topologie rectangulaire.	62
2.12	Représentation d'un neurone j utilisé dans un réseau de Kohonen.	63
2.13	Architecture d'un LVQ composée de n entrées, d'une couche de compétition de N_c neurones et d'une couche de sortie de N_s neurones correspondants chacune à une classe connue.	65

2.14	Structure d'un réseau de neurones de type perceptron multicouche à une couche cachée.	66
2.15	Evolution schématique des erreurs quadratiques calculées sur le corpus de données d'entraînement E_{entr} et de validation E_{val} dans le cadre de méthode d'arrêt prématuré.	71
2.16	Exemple d'utilisation d'un R.N en mode régresseur dans le cadre de la caractérisation scattérométrique permettant d'estimer un profil sinusoidal défini par deux paramètres géométriques (h et h_r) à partir d'une signature optique fournie en entrée.	71
2.17	Image de l'échantillon obtenue par une caractérisation AFM.	72
2.18	Evolution des erreurs quadratiques moyennes calculées sur le corpus de données d'entraînement, de validation et de test dans le cadre de méthode d'arrêt prématuré.	73
2.19	Représentation des signatures optiques mesurées de l'échantillon réel et celles reconstruites à partir de la caractérisation par le PMC régresseur fonctionnant sur un modèle géométrique de type sinusoidal ($h = 53.75 \text{ nm}$ et $h_r = 575.65 \text{ nm}$).	73
2.20	Profils géométriques de l'échantillon obtenus par AFM (rouge) et PMC (bleu).	74
3.1	Caractérisation AFM de l'échantillon E_1 (période de 140 nm) de réseau en résine (IM5010) reposant sur du BARC ; le tout sur un substrat de silicium.	77
3.2	Caractérisation MEB des échantillons E_2 , E_3 et E_4 (période de 500 nm) de réseau en résine (M78Y) sur du BARC et une couche du S_iO_2 ; le tout sur un substrat de silicium.	78
3.3	Caractérisation AFM des échantillons E_5 , E_6 , E_7 , E_8 , E_9 et E_{10} (période de 627 nm) de réseau en résine (S1805) ne débouchant pas sur le substrat de silicium.	78
3.4	Caractérisation d'un profil créneau théorique ($CD = 56 \text{ nm}$ et $h = 90 \text{ nm}$) en rouge par RegPc (en bleu).	80
3.5	Caractérisation d'un profil rectangulaire théorique arrondis en haut et bas ($r_h = 35.1 \text{ nm}$, $r_b = 33 \text{ nm}$, $CD = 78 \text{ nm}$ et $h = 119 \text{ nm}$) en rouge par RegPc (en bleu).	81
3.6	Comparaison des signatures optiques simulées à partir du profil réel de forme créneau arrondi en haut et bas (en rouge) et calculées à partir de celui estimé par RegPc (en bleu).	82
3.7	Représentations des profils géométriques obtenus par différentes caractérisations d'une forme complexe simulant la déviation d'un processus de fabrication (en rouge) : (a) RegPtr (en noir), (b) RegPah (en magenta), (c) RegPab (en vert) et (d) RegPst (en cyan).	86

3.8	Représentation de la forme obtenue par pondération de profils (en bleu) dans le cadre d'une caractérisation d'une forme complexe simulant la déviation d'un processus de fabrication (en rouge).	87
3.9	Modélisation paramétrique des profils géométriques largement utilisés en scattérométrie : profil rectangulaire (a), profil trapézoïdal (b), profil rectangulaire arrondi au sommet (c) et profil trapézoïdal arrondi au sommet (d).	88
3.10	Représentation de la forme obtenue par PMC régresseur générique (en noir) d'un profil complexe simulant la déviation d'un processus de fabrication (en rouge).	89
3.11	Comparaison des signatures optiques calculées à partir des résultats des PMC : RegPc (bleu), RegPtr (noir), RegPab (vert), RegPst (cyan) comparées à celle mesurée (rouge) de l'échantillon E_1	91
3.12	Représentation des profils obtenus par caractérisation de l'échantillon E_1 à partir de RegPc (en bleu), RegPtr (en noir), RegPab (en vert), RegPst (en cyan) et par l'AFM (en rouge).	91
3.13	Représentation des profils obtenus par caractérisation de l'échantillon E_1 par la méthode de pondération de profils (en bleu) et par l'AFM (en rouge).	92
3.14	Comparaison des signatures optiques calculées à partir des résultats obtenus par la méthode de pondération de profils comparées à celle mesurée (en rouge) de l'échantillon E_1	93
3.15	Représentations des profils géométriques obtenus par différentes caractérisations d'une forme qualifiée d'éloignée des profils cibles (en rouge) : (a) RegPc (en bleu), (b) RegPtr (en noir) et (c) RegPab (en vert).	95
3.16	Représentation de la forme obtenue par la méthode de pondération de profils (en bleu) d'un profil qualifié d'éloigné des profils cibles (en rouge).	95
3.17	Comparaison des signatures optiques simulées (en bleu) à partir des résultats par pondération de profils comparées à celles simulées (en rouge) d'un profil qualifié d'éloigné des profils cibles.	96
3.18	Schéma de principe d'un système de vérification d'une forme de profil géométrique donnée par l'utilisation d'un PMC en mode classifieur fonctionnant sur deux classes.	98
3.19	Sortie brute s_1 du TOR pour les échantillons mal classés (FP) en fonction de l'écart géométrique au profil rectangulaire le plus proche.	101
3.20	Évolution de la sortie brute P_s des différents TOR et de l'écart géométrique normalisé ($E_{geonorm}$) en fonction du paramètre σ du profil gaussien.	106
3.21	Représentation des profils gaussiens ayant une hauteur $h = 130 \text{ nm}$ pour différentes valeurs de $\sigma = 110$ (en bleu) et $\sigma = 220$ (en noir) présentant le même écart géométrique à la forme cible définie $\sigma = 140$ (en rouge).	107

3.22	Représentation de l'erreur (E_{I_s, I_c}) en fonction de l'écart géométrique normalisé ($E_{geonorm}$) calculé entre un profil gaussien quelconque et le profil sinus le plus proche pour 4 amplitudes différentes.	112
3.23	Représentation de l'erreur (E_{I_s, I_c}) en fonction de l'écart géométrique normalisé ($E_{geonorm}$) calculé entre un profil sinus déformé quelconque et le profil sinus le plus proche pour 4 amplitudes différentes.	112
3.24	Représentation de l'erreur (E_{I_s, I_c}) en fonction de l'écart géométrique normalisé ($E_{geonorm}$) calculé entre un profil sinus tronqué quelconque et le profil sinus le plus proche pour 4 amplitudes différentes.	113
3.25	Sortie non seuillée correspondant à la classe P_c du ClassP pour les 19 échantillons mal classés (FP) en fonction de la valeur de l'arrondi en haut r_h	116
3.26	Représentations des profils géométriques obtenus par différentes caractérisations de l'échantillon E_4 : <i>RegPc</i> (en bleu) et <i>RegPah</i> (en noir pointillée).	119
3.27	Sorties brutes P_s et P_{sd} supérieures à 0.5 du PMC classifieur $Class3_a$ pour les échantillons au profil gaussien mal classés en fonction du paramètre σ	122
3.28	Sorties brutes P_s et P_g supérieures à 0.5 du PMC classifieur $Class3_a$ pour les échantillons mal classés au profil sinus déformé en fonction du paramètre ρ	122
3.29	Évolution des profils identifiés P_s , P_g et P_{sd} par le PMC classifieur $Class3_a$ en fonction du paramètre σ du profil gaussien pour un seuil d'acceptabilité fixé respectivement à (a) 50%, (b) 70% et (c) 90%.	124
3.30	Évolution de la sortie P_s des PMC classifieurs TOR_a et $Class3_a$ en fonction du paramètre σ du profil gaussien. L'écart géométrique normalisé ($E_{geonorm}$) est également tracé (en rouge)	125
3.31	Comparaison des signatures optiques simulées à partir des résultats de RegPsd et à partir des résultats de l'optimisation des mesures AFM avec celles mesurées sur l'échantillon E_8	127
3.32	Représentation des profils obtenus par caractérisation de l'échantillon E_8 par RegPsd (en bleu), par AFM (en rouge) et par optimisation (en or).	127
3.33	Évolution des profils identifiés P_s , P_g , P_{sd} et P_{st} par le PMC classifieur $Class4_a$ en fonction du paramètre σ du profil gaussien pour un seuil d'acceptabilité fixé respectivement à (a) 50%, (c) 70% et (c) 90%.	129
3.34	Évolution de la sortie P_s des PMC classifieurs TOR_a , $Class3_a$ et $Class4_a$ en fonction du paramètre σ du profil gaussien. L'écart géométrique normalisé ($E_{geonorm}$) est également tracé (en rouge)	130
3.35	Comparaison des signatures optiques simulées à partir des résultats de RegPst et à partir des résultats de l'optimisation des mesures AFM avec celles mesurées sur l'échantillon E_7	132

3.36	Représentation des profils obtenus par : AFM (en rouge), RegPst (en bleu) et par optimisation de LM (en or) pour la caractérisation de l'échantillon E_7 . . .	132
3.37	Représentation d'un profil sinusoïdal dans le domaine temporel (a) et dans le domaine fréquentiel (amplitude (b) et phase (c)).	136
3.38	Représentation d'un profil gaussien dans le domaine temporel (a) et dans le domaine fréquentiel (amplitude (b) et phase (c)).	136
3.39	Représentation d'un profil sinus déformé dans le domaine temporels (a) et dans le domaine fréquentiel (amplitude (b) et phase (c)).	137
3.40	Représentation d'un profil sinus tronqué dans le domaine temporel (a) et dans le domaine fréquentiel (amplitude (b) et phase (c)).	138
3.41	Reconstruction d'un profil sinus déformé simulé ($\rho = 0.46$ et $h = 80 \text{ nm}$) (en rouge) à partir des 2 (en bleu), 3 (en vert) et 4 (en noir) premières harmoniques.	139
3.42	Représentation de β_1 en fonction de l'écart surfacique pour différents profils : gaussien (en noir), sinus déformé (en vert) et sinus tronqué (en or)	141
3.43	Représentation de β_2 en fonction de l'écart surfacique pour différents profils : gaussien (en noir), sinus déformé (en vert), sinus tronqué (en or), un modèle linéaire (en bleu) et un modèle polynômial (en rouge)	142
3.44	Schéma de principe du PMC régresseur utilisé dans le cadre de la détection d'une forme sinusoïdale en utilisant le spectre fréquentiel du profil géométrique permettant d'estimer en sortie deux paramètres (h_r et les caractéristiques des harmoniques) à partir d'une signature optique fournie en entrée.	144
3.45	Schéma de principe du PMC régresseur utilisé dans le cadre de la détection d'une forme sinusoïdale en utilisant le spectre fréquentiel du profil géométrique permettant d'estimer en sortie deux paramètres (h_r et β_1 ou β_2) à partir d'une signature optique fournie en entrée.	144

Liste des tableaux

2.1	Performance du PMC (une couche cachée de 30 neurones) après entraînement évaluée sur les données test (450 échantillons).	72
2.2	Résultat de caractérisation scattérométrique du PMC régresseur (une couche cachée de 30 neurones) sur l'échantillon	73
3.1	Performance de RegPc (une couche cachée composée de 25 neurones) après entraînement évaluée sur les données test (1050 échantillons) fonctionnant exclusivement sur un profil créneau.	79
3.2	Estimation par RegPc d'un profil créneau parfait et qualité du résultat obtenu	80
3.3	Estimation par RegPc d'un profil rectangulaire arrondis en haut et bas et qualité du résultat obtenu.	81
3.4	Performances des PMC fonctionnant sur les différents modèles géométriques simples P_{tr} , P_{ah} , P_{ab} et P_{st}	84
3.5	Résultats de caractérisation des différents PMC fonctionnant indépendamment sur une forme simple concernant un échantillon présentant une forme complexe simulant la déviation d'un processus de fabrication.	85
3.6	Résultat par la méthode de pondération de profils d'une forme complexe simulant la déviation d'un processus de fabrication.	86
3.7	Performance du PMC régresseur générique évaluée sur les données test (1050 échantillons) fonctionnant sur un profil générique.	88
3.8	Résultat de caractérisation du PMC régresseur générique d'une forme complexe simulant la déviation d'un processus de fabrication.	88
3.9	Résultats de caractérisation des différents PMC fonctionnant sur une forme simple concernant la caractérisation de l'échantillon E_1	90
3.10	Caractérisation de l'échantillon E_1 par la méthode de pondération de profils.	92
3.11	Résultat de caractérisation des différents PMC fonctionnant sur une forme simple concernant un échantillon de forme qualifiée d'éloignée des profils cibles.	94
3.12	Résultat de la méthode de pondération de profils d'une forme qualifiée d'éloignée des profils cibles.	94
3.13	Matrice de confusion résumant les résultats d'un classifieur.	98

3.14	Matrice de confusion du TOR (20 neurones cachés) calculée sur le corpus test pour la détection de profils géométriques rectangulaires.	100
3.15	Matrice de confusion du TOR évaluée sur les données du corpus test pour un seuil d'acceptabilité fixé à 85%.	101
3.16	Détection par le TOR de profils rectangulaires sur les échantillons E_2 et E_3 . La sortie s_1 correspond à la classe C_1 (profils cibles rectangulaires) et la sortie s_2 à la classe C_2 (autres profils).	102
3.17	Matrice de confusion du TOR_a (corpus assurant $E_{geonorm} \geq 1\%$) axée sur la détection d'un profil sinusoïdal.	103
3.18	Matrice de confusion du TOR_b (corpus assurant $E_{geonorm} \geq 5\%$) axée sur la détection d'un profil sinusoïdal.	104
3.19	Matrice de confusion du TOR_c (corpus assurant $E_{geonorm} \geq 15\%$) axée sur la détection d'un profil sinusoïdal.	104
3.20	Matrice de confusion du TOR_b évaluée sur les données test du corpus assurant $E_{geonorm} \geq 1\%$	104
3.21	Matrice de confusion du TOR_c évaluée sur les données test du corpus assurant $E_{geonorm} \geq 1\%$	105
3.22	Matrice de confusion du TOR_a évaluée sur les données test du corpus assurant $E_{geonorm} \geq 1\%$ pour un seuil d'acceptabilité fixé à 70%.	105
3.23	Matrice de confusion du TOR_a évaluée sur les données test du corpus assurant $E_{geonorm} \geq 1\%$ pour un seuil d'acceptabilité fixé à 90%.	106
3.24	Valeurs limites de σ caractérisant un profil détecté comme forme cible par chacun des PMC classifieurs (TOR_a , TOR_b et TOR_c) pour chaque seuil d'acceptabilité (50%, 70% et 90%).	108
3.25	Détermination des types de profils des échantillons E_5 , E_6 et E_7 par traitement de la mesure AFM.	109
3.26	Détermination des types de profils des échantillons E_8 , E_9 et E_{10} par traitement de la mesure AFM.	110
3.27	Résultats bruts de classification des échantillons réels (E_5 , E_6 , E_7 , E_8 , E_9 et E_{10}).	111
3.28	Valeurs de l'écart géométrique minimal $E_{geonormmin}$ détectable pour chaque type de profil par la signature optique utilisée.	114
3.29	Matrice de confusion de ClassP dédié à la classification de 3 formes géométriques. La classe P_c correspond aux profils rectangulaires, les classes P_{tr} et P_{ah} respectivement aux profils trapézoïdaux et aux profils rectangulaires arrondis au sommet.	115
3.30	Matrice de confusion de ClassP dédié à la classification de 3 formes géométriques (P_c , P_{tr} et P_{ah}) pour un seuil d'affectation fixé à 70%.	116

3.31	Matrice de confusion de ClassP dédié à la classification de 3 formes géométriques (P_c , P_{tr} et P_{ah}) pour un seuil d'affectation fixé à 90%.	116
3.32	Résultats de classification des échantillons E_2 , E_3 et E_4 par ClassP.	117
3.33	Performances des PMC nommés RegPc et RegPah fonctionnant respectivement sur les modèles P_c et P_{ah} dans les plages définies précédemment (équations 3.11 et 3.13).	118
3.34	Caractérisations par RegPc et RegPah des échantillons E_3 et E_4 pour une forme de profil fixée par le classifieur ClassP.	118
3.35	Matrice de confusion du $Class3_a$ dédié à la classification parmi 3 profils géométriques (P_s , P_g et P_{sd}) pour un seuil d'acceptabilité fixé à 50%.	120
3.36	Matrice de confusion du $Class3_b$ dédié à la classification parmi 3 profils géométriques (P_s , P_g et P_{sd}) pour un seuil d'acceptabilité fixé à 50%.	120
3.37	Matrice de confusion du $Class3_c$ dédié à la classification parmi 3 profils géométriques (P_s , P_g et P_{sd}) pour un seuil d'acceptabilité fixé à 50%.	120
3.38	Matrice de confusion du $Class3_b$ évaluée sur les données test du corpus assurant $E_{geonorm} \geq 1\%$ pour un seuil d'acceptabilité fixé à 50%.	121
3.39	Matrice de confusion du $Class3_c$ évaluée sur les données test du corpus assurant $E_{geonorm} \geq 1\%$ pour un seuil d'acceptabilité fixé à 50%.	121
3.40	Matrice de confusion du $Class3_a$ évaluée sur les données test du corpus assurant $E_{geonorm} \geq 1\%$ pour un seuil d'acceptabilité fixé à 70%.	123
3.41	Matrice de confusion du $Class3_a$ évaluée sur les données test du corpus assurant $E_{geonorm} \geq 1\%$ pour un seuil d'acceptabilité fixé à 90%.	123
3.42	Résultats de classification des échantillons réels (E_5 , E_6 , E_8 et E_{10}).	125
3.43	Performance de RegPsd (une couche cachée composée de 15 neurones) après entraînement évaluée sur les données test (450 échantillons) fonctionnant exclusivement sur un profil sinus déformé.	126
3.44	Résultat de caractérisation par RegPsd de l'échantillon E_8	126
3.45	Matrice de confusion de $Class4_a$ axée sur l'identification d'un profil géométrique particulier.	128
3.46	Résultat de la classification des échantillons réels (E_5 , E_6 , E_7 , E_8 , E_9 et E_{10}) par la $Class4_a$	130
3.47	Performance de RegPst (une couche cachée composée de 30 neurones) après entraînement évaluée sur les données test (450 échantillons) fonctionnant exclusivement sur un profil sinus tronqué.	131
3.48	Résultat de caractérisation par RegPst de l'échantillon E_7	131
3.49	Résultat de la décomposition fréquentielle du profil gaussien défini par $\sigma = 110$ et $h = 80$ nm.	137

3.50	Résultat de la décomposition fréquentielle du profil sinus déformé défini par $\rho = 0.46$ et $h = 80 \text{ nm}$	137
3.51	Résultat de la décomposition fréquentielle du profil sinus tronqué défini par $\alpha = 0.9$, $b = 300 \text{ nm}$ et $h = 80 \text{ nm}$	138
3.52	Conditions sur les valeurs de β_1 correspondants à une déviation donnée au profil cible en termes d'écart géométrique pour chaque type de profil.	141
3.53	Conditions sur les valeurs de β_2 correspondants à une déviation donnée au profil cible en termes d'écart géométrique pour chaque type de profil.	143
3.54	Conditions sur les valeurs de β_2 correspondants à une déviation donnée au profil cible en termes d'écart géométrique indépendamment du type de profils.	143
3.55	Performance de Reg2h (une couche cachée composée 15 neurones) après entraînement évaluée sur les données test (450 échantillons).	145
3.56	Performance de Reg4h (deux couches cachées composées respectivement 10 et 25 neurones) après entraînement évaluée sur les données test (1200 échantillons).	145
3.57	Résultat de caractérisation des 3 faux profils sinus (gaussien, sinus déformé et sinus tronqué) par la FFT.	146
3.58	Résultat de caractérisation des 3 faux profils sinus (gaussien, sinus déformé et sinus tronqué) par Reg2h.	147
3.59	Résultat de caractérisation des 3 faux profils sinusoidaux (gaussien, sinus déformé et sinus tronqué) par Reg4h.	148
3.60	Performance de Reg β_2 (une couche cachée composée de 10 neurones) après entraînement évaluée sur les données test (1200 échantillons).	149
3.61	Résultats de caractérisation des 3 faux profils sinus (gaussien, sinus déformé et sinus tronqué) par le Reg β_2	149
3.62	Résultats des critères de détection β_1 et β_2 des échantillons E_5 à E_{10} calculés à partir des profils obtenus par mesures brutes AFM.	150
3.63	Résultats de caractérisation des échantillons E_5 , E_6 et E_{10} par Reg2h, Reg4h et Reg β_2	150
3.64	Détection des échantillons réels (E_5 , E_6 et E_{10}) à partir des résultats de β_1 estimés par Reg2h et Reg4h.	151
3.65	Détection des échantillons réels (E_5 , E_6 et E_{10}) à partir du résultat de β_2 estimé par Reg4h.	152
3.66	Détection des échantillons réels (E_5 , E_6 et E_{10}) à partir du résultat de β_2 estimé par Reg β_2	152

Résumé :

La miniaturisation des composants électroniques a permis le développement de nombreuses applications dans divers domaines de la science et de la technologie. Les procédés de fabrication comme la lithographie ont énormément évolué et permettent aujourd'hui une réduction importante des dimensions caractéristiques de la structure. Ces progrès technologiques nécessitent des techniques de caractérisation rapides, fiables, précises, et si possible, à moindre coût. La scattérométrie est l'une des méthodes optiques qui se présente comme une alternative répondant à cette problématique de caractérisation. Le processus de caractérisation scattérométrique nécessite un certain nombre d'hypothèses contraignantes comme la connaissance préalable de la forme géométrique de la structure à tester qui est fixée tout le long du processus. Ce manuscrit présente un processus complet de caractérisation en utilisant le réseau de neurones comme outils d'aide à la décision en limitant le nombre de connaissances a priori sur la structure. Nous avons ainsi montré la possibilité de valider ou non la forme d'un échantillon en plaçant en amont de l'étape de caractérisation un réseau de neurones fonctionnant en mode classifieur. La méthode développée permet la caractérisation des différents types de structures aux formes variées issues des procédés de fabrication courants. Les premières applications mises en œuvre concernent la détection de défaut ou l'identification d'une forme géométrique précise sur un mode de fonctionnement binaire. La seconde concerne l'identification d'une forme géométrique parmi un nombre limité de profils. Ces méthodes présentées sont parfaitement bien adaptées à la problématique de classification en scattérométrie en apportant une information qualitative complémentaire. Elles peuvent être utilisées par exemple dans le cadre d'un contrôle en ligne de production ou lors d'un suivi du processus de fabrication.

Mots-clés : scattérométrie, réseaux de neurones, classification, problème inverse, réseau de diffraction.

Abstract:

Miniaturizing electronic component had allowed development of a lot applications in various fields of sciences and technologies. Fabrication processes such as lithography have highly progressed. This technology allowed a significant size reduction of structure characteristic dimensions. This technological progress requires reliable, accurate, fast and low-cost characterization techniques if possible. Scatterometry is one of optical methods that presents an alternative solution to this characterization problem. The scatterometric process requires a number of constraining assumptions such as prior knowledge of structure geometrical shape to be tested which is fixed throughout process. This manuscript presents a complete characterization process using neural network method. This method is used as a decision support tool by limiting amount of priori knowledge about structure. Thus, we shown possibility of validating or not the sample shape by placing a neural network operating in classifier mode of characterization step. Obtained results concern different structures with various shapes resulting from fabrication processes. The first implemented application concern default detection or geometrical target identification of a binary shape in operating mode. The second concerns identification of a geometrical shape among a limited number of profiles. These developed methods are perfectly adapted to classification problem in scatterometry by providing additional qualitative information. These methods can be used, for instance in a control of online production or for monitoring when manufacturing process.

Keywords: scatterometry, neural networks, classification, inverse problem, diffraction grating.