



# Développement de la technique de scattérométrie neuronale dynamique

Ismail El Kalioubi

► **To cite this version:**

Ismail El Kalioubi. Développement de la technique de scattérométrie neuronale dynamique. Micro et nanotechnologies/Microélectronique. Université Grenoble Alpes, 2015. Français. <NNT : 2015GREAT033>. <tel-01215219>

**HAL Id: tel-01215219**

**<https://tel.archives-ouvertes.fr/tel-01215219>**

Submitted on 13 Oct 2015

**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.

## THÈSE

Pour obtenir le grade de

**DOCTEUR DE L'UNIVERSITÉ GRENOBLE ALPES**

Spécialité : Nano Electronique et Nano Technologies

Arrêté ministériel : 7 août 2006

Présentée par

« **Ismail El Kalyoubi** »

Thèse dirigée par « **Cécile Gourgon** » et

codirigée par « **Maxime Besacier** », « **Stéphane Robert** »

préparée au sein du **Laboratoire des Technologies de la Microélectronique et du Laboratoire Hubert Curien**

dans l'**École Doctorale l'École Doctorale Electronique, Electrotechnique, Automatique & Traitement du Signal**

## Développement de la technique de scattérométrie neuronale dynamique

Thèse soutenue publiquement le « **3 Juin 2015** », devant le jury composé de :

**M. Gérard, GRANET**

Professeur à l'Université Blaise Pascale (IP), Rapporteur

**M. François, FLORY**

Professeur à l'École Centrale Marseille (IM2NP), Rapporteur

**M. Ludovic, ESCOUBAS**

Professeur à l'Université d'Aix-Marseille III (IM2NP), Examineur

**Mme Cécile, GOURGON**

Directeur de Recherche à Grenoble (CNRS-LTM), Directeur de thèse

**M. Maxime, BESACIER**

Maître de conférence à Polytech'Grenoble (CNRS-LTM), Co-Encadrant

**M. Stéphane, ROBERT**

Maître de conférence à Télécom St-Etienne (CNRS-LHC), Co-Encadrant

**M. Patrick, SCHIAVONE**

Ingénieur à Aselta, Invité





عش كل لحظة وكأنها آخر لحظة في حياتك، عش بالإيمان، عش بالحب  
عش بالأمل، عش بالكفاح، عش بالصبر، وقدّر قيمة الحياة..

ابراهيم الفقي

*« Vis chaque instant de ta vie comme si c'était le dernier, vis avec la foi, vis avec l'espoir, vis avec l'amour, vis avec la lutte, vis avec la patience, et estime la valeur de la vie.. »*

Ibrahim El Fikky



*A mes parents, Leila et Essam..*



---

## Remerciements

J'aimerais tout d'abord remercier M. Gérard GRANET et M. François FLORY d'avoir pris le temps de rapporter mon travail. J'exprime également mes remerciements à M. Ludovic ESCOUBAS pour avoir examiné mes travaux et à M. Patrick SCHIAVONE pour m'avoir fait l'honneur de faire partie de mon jury de thèse.

J'exprime mes remerciements à Olivier JOUBERT et à Jean-Jacques ROUSSEAU, directeur respectivement du LTM et de l'ex-LT2C, pour avoir accepté de m'accueillir dans leurs laboratoires respectifs et de l'aide qu'ils m'ont apporté au cours de cette thèse.

Ensuite, je remercie avec toute ma gratitude ma directrice de thèse Cécile GOURGON et mes co-encadrants Maxime BESACIER et Stéphane ROBERT pour leur encadrement. De simples mots ne seraient pas suffisants pour décrire tous ce que vous avez fait pour moi. Merci infiniment. Grâce à vos compétences et votre disponibilité, j'ai appris beaucoup de chose au cours de cette thèse, tant sur le plan scientifique que que personnel. Merci pour votre patience, vos encouragements et vos conseils avisés.

Je souhaite ensuite remercier toutes les personnes au LTM et à l'ex-LT2C qui ont contribué à ce que ces travaux aboutissent: Je remercie tout d'abord Jumana BOUSSEY, directrice adjointe du LTM, pour avoir contribué à ce que ma thèse se déroule dans de bonnes conditions. Merci à Camille PETIT-ETIENNE et à Kevin MENGUILTI pour les gravures plasmas. Merci à Sandra BOS et Mickael MARTIN pour les mesures CD-AFM. Merci à David FUARD, Mohamed SAIB et Nicolas TROSCOMPT pour leur aide précieuse qu'ils m'ont apporté au cours de ma thèse.

Merci à Marielle CLOT, Sylvaine CETRA, Stéphanie BERGER, Céline COUDERT et Nathalie PERITOIRE, pour avoir facilité toute les démarches administratives.

J'adresse particulièrement mes remerciements à mes co-bureaux avec qui j'ai partagé de bons moments: Badr, Faouzi, Seb, Meshkhan, Céline, Ardo, Nicole, Redwan et Schicheng.

Merci à tous les Permanents, Techniciens, Post-doc et thésards qui ont rendu mon séjour agréable, en particulier, : Karim, Marie, Sandrine, Nathalie, Paulin, Alexandru, Seb (AVERTIN), Bilal, Laurent, Nader, Romu, Romain, Michel, Julien, Jérôme, Farah, Aurélien, Mokrane.. et à St Etienne: Bilal, Salim, Faris, Désiré, Amir, Yaya, Khamis, Hanan, Hedi, Elie, Antoine, Meriam, Zana..

Merci à tous les amis sur Grenoble et Saint Etienne en particulier Bilel, Ahmed, Houcine, Salim (MIMOUNI) et Hossam pour tout ce qu'ils ont fait pour moi. Merci à Allal et sa famille pour leur soutien.

Enfin, je remercie avec toute ma gratitude ma mère et mon père, Leila et Essam, pour leur soutien infaillible. Merci de m'avoir motivé et aidé surtout durant les moments difficiles. Aussi, merci infiniment à mes sœurs, Mennen, Heba et Aliaa, pour leur soutien, pour m'avoir aidé et pour leurs conseils qui m'ont permis de positiver à tous les moments. Je remercie aussi de tout mon coeur Heba (MOURAD), mon épouse, pour son soutien et pour avoir toujours été à mes côtés. Cette thèse, c'est grâce à vous.



---

---

---

# Table des matières

---

<b>Introduction .....</b>	<b>23</b>
<b>Chapitre I. Techniques de métrologie pour la microélectronique .....</b>	<b>31</b>
<b>I.1 Les techniques morphologiques .....</b>	<b>32</b>
I.1.1 Microscopie Electronique à Balayage.....	32
I.1.1.1 Le CD-SEM .....	34
I.1.1.2 Le XSEM .....	35
I.1.2 L'AFM et le CD-AFM.....	36
I.1.2.1 L'AFM .....	36
I.1.2.2 Le CD-AFM.....	39
<b>I.2 La scattérométrie .....</b>	<b>41</b>
I.2.1 Principe de la scattérométrie .....	41
I.2.1.1 Définition .....	41
I.2.1.2 Le réseau de diffraction .....	43
I.2.1.3 Principe de la diffraction.....	46
I.2.2 Les différents types de signatures optiques .....	50
I.2.2.1 Scattérométrie ellipsométrique .....	50
<b>I.3 Méthodes de résolution du problème inverse.....</b>	<b>58</b>
I.3.1 Méthodes d'optimisation .....	59
I.3.1.1 L'algorithme de descente de gradient [CUR44] .....	62
I.3.1.2 L'algorithme de Gauss-Newton.....	63
I.3.2 Méthode des bibliothèques.....	68
I.3.2.1 Principe .....	68
I.3.2.2 Mesure spectroscopique et cinétique .....	70
I.3.2.3 Les GPU (Graphics Processing Unit) et autres améliorations .....	70
I.3.3 Une méthode de régression particulière : les réseaux de neurones artificiels.....	71
<b>I.4 Conclusion .....</b>	<b>75</b>
<b>I.5 Bibliographie.....</b>	<b>77</b>

---

## **Chapitre II. Scattérométrie par réseaux de neurones ..... 87**

<b>II.1</b>	<b>Présentation des réseaux de neurones</b> .....	88
II.1.1	Historique des réseaux de neurones .....	88
II.1.2	Applications des réseaux de neurones.....	89
II.1.3	La métaphore biologique.....	91
II.1.4	Formulation mathématique d'un neurone artificiel.....	92
II.1.5	Architecture d'un réseau de neurones .....	95
II.1.6	Les réseaux de neurones à couches .....	97
II.1.6.1	Présentation de l'architecture .....	97
II.1.6.2	Propriétés d'un PMC à une couche cachée .....	98
II.1.6.3	Modélisation d'un processus expérimental par PMC.....	100
<b>II.2</b>	<b>Fonctionnement des réseaux de neurones</b> .....	101
II.2.1	Phase d'apprentissage .....	102
II.2.2	Estimation de l'intervalle de confiance .....	107
<b>II.3</b>	<b>Un réseau de neurones pour la scattérométrie</b> .....	111
II.3.1	Choix de l'architecture optimale .....	112
II.3.1.1	Influence du nombre d'échantillons et du nombre de neurones par couche cachée	113
II.3.1.2	Influence de la plage de variation.....	115
II.3.2	Etude expérimentale .....	117
<b>II.4</b>	<b>Comparaison avec les techniques morphologiques classiques</b> .....	121
II.4.1	Etude comparative avec le X-SEM .....	122
II.4.2	Etude comparative avec le CD-AFM .....	129
<b>II.5</b>	<b>Conclusion</b> .....	142
<b>II.6</b>	<b>BLIOGRAPHIE</b> .....	144

## **Chapitre III. Evaluation des réseaux de neurones pour la scattérométrie en temps réel ..... 149**

<b>III.1</b>	<b>Comparaison de la méthode des bibliothèques et des réseaux de neurones.....</b>	150
--------------	--	-----

---

III.1.1	Conditions de l'étude .....	151
III.1.2	Considérations sur le critère d'erreur calculé sur les signatures reconstruites.	154
III.1.3	Etude théorique .....	156
III.1.3.1	Utilisation d'une signature théorique .....	156
III.1.3.2	Utilisation d'une signature théorique bruitée .....	158
III.1.3.3	Performances suivant la plage de variation des paramètres .....	160
III.1.3.4	Performances suivant le nombre de couche cachées.....	162
III.1.3.5	Résistance de la caractérisation au bruit de mesure .....	163
III.1.4	Etude sur un échantillon expérimental .....	166
III.1.5	Conclusion de l'étude.....	168
<b>III.2</b>	<b>Suivi en temps réel de procédé de gravure plasma par voie neuronale.....</b>	<b>170</b>
III.2.1	La gravure plasma .....	170
III.2.2	Contraintes d'une application en temps réel .....	172
III.2.3	Etude théorique de contrôle de procédés par voie neuronale .....	174
III.2.4	Etude expérimentale de suivi en temps réel de resist trimming .....	178
III.2.4.1	Dispositif expérimental .....	178
III.2.4.2	Résultats scattérométriques .....	181
<b>III.3</b>	<b>Conclusion .....</b>	<b>189</b>
<b>III.4</b>	<b>Bibliographie.....</b>	<b>190</b>
<b>Conclusion.....</b>	<b>.....</b>	<b>193</b>
<b>Annexe I. Les différentes techniques de lithographie.....</b>	<b>.....</b>	<b>199</b>
<b>Annexe 2. La scattérométrie réflectométrique .....</b>	<b>.....</b>	<b>205</b>

---

---

# Tables des Figures

---

<i>Figure I-1. Principe de fonctionnement du SEM [REY] .....</i>	33
<i>Figure I-2. Image de ligne de résine sur Silicium, de période <math>d=189</math> nm, obtenue avec le CD-SEM (tension d'accélération=<math>300V</math>, grossissement=<math>199583</math>, taille du champ d'observation=<math>0.676 \mu m</math>, <math>0.676 \mu m</math>) .....</i>	34
<i>Figure I-3. Image MEB d'un échantillon de Neb22 sur Si, de période <math>d=750nm</math>, en cross section (à gauche), en vue inclinée (à droite) .....</i>	35
<i>Figure I-4. Schéma de fonctionnement du microscope AFM [POI08] .....</i>	37
<i>Figure I-5. Photographie MEB de pointes AFM : PT/Ir (à gauche), diamant (à droite) [POI08] .....</i>	37
<i>Figure I-6. Schéma de la pointe de l'AFM scrutant la surface en mode tapping [POI08] .....</i>	38
<i>Figure I-7. Déformation du profil par une pointe AFM. On observe que le profil reconstitué par AFM (trait en pointillé) n'est pas le même que le profil réel (trait en continue) .....</i>	38
<i>Figure I-8. Exemple d'image brute avant déconvolution obtenue avec un AFM classique .....</i>	39
<i>Figure I-9. Profil représentant la forme de la pointe utilisée par un CD-AFM de type CDR120 nm (Critical Dimension re-entrant profile) [ELK10] .....</i>	39
<i>Figure I-10. Schématisation d'une reconstruction du profil avec un AFM classique (figure du haut) et avec un CD AFM (figure bas). [AZA12]. .....</i>	40
<i>Figure I-11. Image traitée en 3D obtenue par CD AFM pour un profil BARC sur Si .....</i>	40
<i>Figure I-12. Schéma illustratif du principe de la résolution du problème inverse par scattérométrie .....</i>	42
<i>Figure I-13. Modélisation paramétrique de réseau avec profil créneau sur couche résiduelle défini par trois paramètres (CD, h et hr) .....</i>	43
<i>Figure I-14. Modélisation paramétrique de réseau avec profil trapézoïdal sur couche résiduelle défini par 4 paramètres (CD,b1,h,hr) .....</i>	44
<i>Figure I-15. Modélisation paramétrique de réseau avec profil arrondi sur couche résiduelle défini par 4 paramètres (CD,r,h,hr) .....</i>	44
<i>Figure I-16. Profil mesuré par CD-AFM concernant une ligne de résine IM5010 (STMicroelectronics) surplombant une couche de Barc sur du Si. ....</i>	45
<i>Figure I-17. Images XSEM (a et b) obtenues pour des motifs composées de résine Neb22 sur substrat de Si. ....</i>	45
<i>Figure I-18 . Diffraction par un réseau périodique en configuration classique [GER09] .....</i>	47
<i>Figure I-19. Configuration de diffraction conique définie par un angle azimutal <math>\varphi</math> [GER09] .....</i>	47
<i>Figure I-20. Exemple de profil trapézoïdal à bords arrondis ( ) découpé en 10 tranches pour la méthode MMFE .....</i>	49
<i>Figure I-21. Les deux états de polarisation de la lumière a) TE et b) TM [ROB03]. .....</i>	51
<i>Figure I-22. Réflexion d'une onde polarisée [SOU08] .....</i>	51
<i>Figure I-23a. Schéma de principe de l'ellipsomètre à élément tournant [BAO03] .....</i>	53
<i>Figure I-24. Schéma de principe de l'optimisation paramétrique dans le cadre de la scattérométrie ellipsométrique .....</i>	59

Figure I-25. Représentation schématique des minimas d'une fonction $Z$ dépendant d'un seul paramètre $x$ .....	61
Figure I-26: Illustration de la méthode de descente de gradient. ....	63
Figure I-27.Schéma de fonctionnement de la méthode de Gauss Newton .....	65
Figure I-28. . Schéma de principe de la résolution du problème inverse avec la méthode des bibliothèques .....	69
Figure I-29. Diagramme de résolution du problème inverse avec la méthode des réseaux de neurone.....	73
Figure II-1. Structure schématique d'un neurone biologique .....	91
Figure II-2. Représentation mathématique d'un neurone formel. Les $n$ entrées sont désignées par les variables $x_i$ , la sortie est notée $s_j$ et les poids synaptiques pondérant chaque connexion sont désignés par les variables $w_{j,i}$ .....	92
Figure II-3. Exemples de fonctions d'activations communément utilisées pour calculer la sortie d'un neurone formel.....	94
Figure II-4. Exemple d'architecture d'un réseau de neurones non bouclé, complètement connecté comportant $(N_s-1)$ neurones cachées, $N$ entrées et un neurone de sortie. On distingue les poids des connexions reliant les entrées à tous les neurones par $w$ de ceux relatifs aux connexions entre neurones notés $z$ .....	96
Figure II-5. Exemple d'architecture d'un réseau de neurone non bouclé dit à couches présentant une seule couche cachée de $N_c$ neurones comprenant $N$ entrées et $N_s$ sorties . Soit $w_{j,i}$ le poids synaptique reliant le $j^{\text{ième}}$ neurone de la couche cachée à l'entrée $x_i$ et $z_{1,j}$ celui reliant le premier neurone de la couche de sortie avec la sortie du $j^{\text{ième}}$ neurone de la couche cachée.....	98
Figure II-6. Exemple d'utilisation d'un RN dans le cadre de la scattérométrie permettant de caractériser en sortie un profil défini par 3 paramètres ( $CD$ , $h$ et $hr$ ) Le nombre d'entrée est directement relié au nombre de mesures constituant la signature ellipsométrie (typiquement 52 entrées).....	101
Figure II-7. Schéma illustratif représentant les résultats d'un RN suite à 2 apprentissages différents. Les données composant le corpus d'entraînement sont schématisées par les points noirs. Les points verts constituent des données qui permettent de tester les performances du RN. La courbe bleue représente le modèle d'un RN présentant une généralisation optimale et la courbe rouge un RN dans le cas d'un surapprentissage.....	104
Figure II-8.Schéma illustratif de la validation croisée nécessitant $D$ apprentissages successifs. Les parties bleues sont utilisées pour la validation des performances après chaque apprentissage.....	105
Figure II-9. Evolution schématique des erreurs sur les corpus de données lors d'un apprentissage par « early stopping ». $E^{\text{val}}$ (trait en pointillé) représente les erreurs sur les données de validation et $E^{\text{app}}$ (trait plein) sur les données d'apprentissage L'apprentissage est arrêté lorsque $E^{\text{val}}$ atteint un minimum. ....	107
Figure II-10. Architecture à 2 PMC permettant de donner à la fois les résultats de la caractérisation (RN1) mais également une estimation de l'erreur réalisée par celui-ci (RN2).....	109
Figure II-11. Fonction à modéliser $f(x)$ sans bruit (bleu) et présentant différents niveaux de bruit (en rouge): un bruit gaussien d'écart type 0.5 pour $1 < x < 5$ , un bruit gaussien d'écart type 1 pour $5 < x < 10$ , sans bruit pour $10 < x < 15$ .....	110
Figure II-12. Résultats du RN1 avec barres d'erreur à $\sigma$ obtenus par le RN2.....	111
Figure II-13. Résultat du RN2 : Sorties désirées (en rouge) et estimées (en bleu) .....	111
Figure II-14. Profil géométrique trapézoïdal avec couche résiduelle défini par 4 paramètres ( $CD$ , $b_1$ , $h$ , $h_r$ )..	112

Figure II-15. Variation des performances du PMC à une couche cachée en fonction du nombre d'échantillons pour différentes neurones cachée.....	114
Figure II-16. Temps d'entraînement du RN en fonction du nombre d'échantillons et pour différents nombres de neurones .....	115
Figure II-17. Modèle créneau utilisé pour l'échantillon issu de lithographie optique par immersion défini par trois paramètres géométriques ( $CD, h, h_r$ ).....	117
Figure II-18. Performances sur les valeurs de test en fonction du nombre d'échantillons pour différentes architectures testées ( $N_c=10, 15, 20, 25$ ). L'échantillon est composé d'un empilement de résine IM5010 sur du BARC sur Si avec une période $d=140nm$ .....	118
Figure II-19. Schéma récapitulatif expliquant le calcul de l'erreur sur les signatures reconstruites $E_{Is,lc}$ par problème direct à partir des données fournies par l'outil scatterométrique (RN dans la figure). .....	119
Figure II-20. Profils géométriques obtenus avec les algorithmes de Krigeage (rouge), Région de confiance (bleu) et RN (vert) dans le cas de la caractérisation scattérométrique de l'échantillon issu de lithographie optique par immersion .....	120
Figure II-21. Comparaison des signatures ellipsométriques mesurées et calculées à partir des résultats obtenus respectivement par Krigeage, par l'algorithme à région de confiance et par RN dans le cas de la caractérisation scattérométrique de l'échantillon issu de lithographie optique par immersion. ....	121
Figure II-22. Caractérisation par XSEM de l'échantillon Neb22 sur Si de période 750 nm obtenu par nano-impresion en cross section (à gauche) et vue de dessus (à droite) .....	122
Figure II-23. Modèle trapézoïdal symétrique avec couche résiduelle défini par 4 paramètres géométriques ( $CD, b_r, h, et h_r$ ).....	123
Figure II-24. Comparaison de profils géométriques obtenus pour la caractérisation XSEM et scattérométrique supposant une modélisation trapézoïdale de l'échantillon Neb22 sur Si de période 750 nm obtenu par nano-impresion .....	124
Figure II-25. Comparaison de la signature expérimentale avec les signatures reconstruites à partir des résultats de différentes méthodes de caractérisation (RN, MEB) dans le cas d'une modélisation trapézoïdale pour l'échantillon Neb22 sur Si de période 750 nm obtenu par nano-impresion .....	126
Figure II-26. Modèle trapézoïdal avec bords arrondis défini par 5 paramètres ( $CD_{top}, CD_{bottom}, h, h_r, r$ ) incluant les 4 profils couramment utilisés en scattérométrie (créneau, trapézoïdal, créneau arrondi, trapézoïdal à bord arrondi).....	127
Figure II-27. Comparaison de profils géométriques obtenus pour la caractérisation XSEM et scattérométrique supposant une modélisation trapézoïdale avec bord arrondis de l'échantillon Neb22 sur Si de période 750 nm obtenu par nano-impresion .....	128
Figure II-28. Modèle créneau défini par 3 paramètres ( $CD, h et h_r$ ) utilisé pour la caractérisation de l'échantillon Im5010 sur Si.....	129
Figure II-29. Modèle trapézoïdal avec bords arrondis défini par 5 paramètres ( $CD_{top}, CD_{bottom}, r, h, hr$ ) utilisé dans le cadre de la caractérisation scattérométrique de l'échantillon Im5010 sur Si .....	130
Figure II-30. Courbes de régression pour les différents paramètres du profil générique ( $CD_{top}, CD_{bottom}, r, h, hr$ ), calculées sur les valeurs du corpus de test en fin d'entraînement en considérant un PMC d'une seule cachée de 15 neurones entraîné sur 6000 échantillons .....	132



Figure II-31. Représentation géométrique des 3 échantillons ayant en commun les paramètres $CD_{bottom}$ , $h$ et $hr$ pour différentes valeurs de $CD_{top}$ ( $CD_{top}=0$ nm, $CD_{top}=10$ nm, $CD_{top}=20$ nm) et $r$ ( $r=20$ nm, $r=10$ nm, $r=5$ nm) pour $h=90$ nm, $CD_{bottom}=40$ nm et $hr=70$ nm .....	134
Figure II-32. Représentation des signatures ( $I_s, I_c$ ) pour les trois profils étudiés ( $CD_{top}=0$ nm / $r=20$ nm, $CD_{top}=10$ nm / $r=10$ nm, $CD_{top}=20$ nm / $r=5$ nm pour $h=90$ nm, $CD_{bottom}=40$ nm et $hr=70$ nm.....	135
Figure II-33. Comparaison de profils géométriques obtenus par caractérisation AFM et scattérométrie (RN) supposant une modélisation générique dans le cas de l'échantillon Im5010 sur Barc sur Si. ....	136
Figure II-34. Modèle de profil elliptique (gauche) et arrondi (droite) défini chacun par 4 paramètres ( $CD_{top}$ , $CD_{bottom}$ , $h$ , $h_r$ ) utilisé dans le cadre de la caractérisation scattérométrie de l'échantillon Im5010 sur BARC sur Si .....	137
Figure II-35. Comparaison des résultats scattérométriques correspondant à différents profils géométriques supposés dans le cas de l'échantillon Im5010 sur BARC sur Si.....	139
Figure II-36. Juxtaposition des résultats de caractérisations obtenues par CD-AFM et par scattérométrie à partir de différents profils géométriques supposés dans le cas de l'échantillon Im5010 sur BARC sur Si.....	139
Figure II-37. Comparaison de différents profils arrondis (résultats du RN, modélisation optimale de la mesure AFM) avec le profil obtenu par CDAFM dans le cas de l'échantillon Im5010 sur BARC sur Si.....	140
Figure III-1. Profil géométrique défini par trois paramètres ( $CD$ , $h$ , $hr$ ) considéré pour la comparaison bibliothèque/RN. ....	151
Figure III-2. Schéma représentant la méthodologie utilisée pour étudier l'influence sur l'erreur rms d'une variation des paramètres géométriques .....	154
Figure III-3. Exemple d'arborescence pour déterminer les 27 combinaisons correspondant à une variation delta de +/- 1 nm.....	155
Figure III-4. Erreur sur les paramètres en fonction de la variation d'erreur rms dans le cas de la caractérisation de l'échantillon défini par $CD=49$ nm, $h=108$ nm, $hr=80$ nm . ....	156
Figure III-5. Résultats de la caractérisation ( $CD$ , $h$ , $hr$ et erreur rms sur les signatures reconstruites) par bibliothèque et par RN en fonction du nombre d'échantillons utilisés dans le cas d'une signature théorique fonctionnant sur la plage P2. Le RN possède une seule couche cachée. ....	157
Figure III-6. Résultats de la caractérisation ( $CD$ , $h$ , $hr$ et erreur rms) par bibliothèque et par RN en fonction du nombre d'échantillons pour une signature simulée avec un bruit $\sigma_s=0,0033$ fonctionnant sur la plage P2. Le RN est entraîné avec un bruit $\sigma_e=0,0033$ . ....	159
Figure III-7. Résultats de la caractérisation (erreur rms sur signatures reconstruites) par BB et par RN en fonction du nombre d'échantillons pour une signature simulée avec un bruit $\sigma_s=0,0033$ effectué sur les trois plages de variation. Le RN présente une seule couche cachée et est entraîné avec un bruit $\sigma_e=0,0033$ . ....	161
Figure III-8. Résultats de la caractérisation (erreur rms sur signatures reconstruites) par bibliothèque et par RN en fonction du nombre d'échantillons pour une signature simulée avec un bruit $\sigma_s=0,0033$ fonctionnant sur la plage 3. Le RN présente 2 couches cachées et est entraîné avec un bruit $\sigma_e=0,0033$ . 3 cas de figures sont considérés pour $N_{c1}(10,20,30)$ et respectivement 4 cas de figures sont considéré pour $N_{c2}(10,20,30,40)$ . ....	163

Figure III-9. Résultats de la caractérisation (erreur rms sur signatures reconstruites) par BB et par RN en fonction du nombre d'échantillons pour une signature simulée avec différents bruits $\sigma_s$ , fonctionnant sur la plage 2. Le RN présente une couche cachée et est entraîné avec différents bruits $\sigma_e$ .....	165
Figure III-10. Modèle créneau défini par 3 paramètres (CD, h et $h_r$ ) utilisé pour la caractérisation de l'échantillon Im5010 sur Si.....	166
Figure III-11. Résultats de la caractérisation (CD, h, $h_r$ et erreur rms sur les signatures reconstruites) par bibliothèque et par RN en fonction du nombre d'échantillons pour une signature réelle fonctionnant sur la plage P2. Le RN est composé d'une couche cachée de 25 neurones et est entraîné avec un bruit $\sigma_e=0,0033$ .....	168
Figure III-12. Schéma des différentes étapes nécessaires pour la fabrication d'un réseau de ligne périodique	170
Figure III-13. Schéma représentant la réduction de côte de résine sur un profil créneau .....	171
Figure III-14. Résultats de caractérisation d'un profil de résine sur BARC sur Si par CDAFM avant l'étape de resist trimming .....	172
Figure III-15. Profil géométrique créneau défini par 3 paramètres (CD, h, $h_r$ ) .....	175
Figure III-16. Evolution au cours du temps des paramètres géométriques du profil créneau (CD, h, $h_r$ ) lors de la simulation d'un procédé de resist trim.....	175
Figure III-17. Caractérisation dynamique obtenue par RN pour un profil théorique défini les trois paramètres (CD, h, $h_r$ ), en considérant un spectre de 32 longueurs d'ondes et un PMC entraîné respectivement avec les plages définies dans la table III.4 .....	178
Figure III-18. Présentation de la plateforme de gravure Centura 5300.....	179
Figure III-19. Schéma représentatif d'un réacteur de gravure DPS.....	180
Figure III-20. Protocole expérimental de validation de la scattérométrie dynamique neuronale comprenant des mesures CD AFM systématiques.....	181
Figure III-21. Evolution du profil des lignes de la résine IM5010 lors de l'étape de gravure par plasma Hbr/O <sub>2</sub> obtenue par CD-AFM .....	182
Figure III-22. (A gauche) Profil géométrique créneau à bord arrondi défini par 4 paramètres géométriques ( $CD_{top}$ , $CD_{bottom}$ , h, $h_r$ ) correspondant à la structure avant gravure et (à droite) profil géométrique défini par 5 paramètres utilisés pour la modélisation de l'échantillon en cours de gravure ( $CD_{top}$ , $CD_{bottom}$ , h, $h_{r_{sup}}$ , $h_{r_{inf}}$ ). .....	183
Figure III-23. Evolution des paramètres géométriques du profil de ligne calculé par le RN ( $CD_{bottom}$ , hauteur de résine h et hauteur de BARC non gravée $h_{r_{inf}}$ ).....	185
Figure III-25. Evolution des paramètres géométriques reconstruit du profil de ligne (hauteur du pied de barc $h_{r_{sup}}$ , hauteur barc totale $h_{r_{totale}}$ , hauteur totale du motif $h_{totale}$ ).....	186
Figure III-25. Evolution de l'erreur rms sur les signatures reconstruites pour le profil de ligne IM5010 sur BARC.....	188

---

---

## Liste des tables

---

<i>Table II-1. Plages de variation utilisées sur les paramètres géométriques pour l'apprentissage du RN .....</i>	<i>113</i>
<i>Table II-2. Résultats obtenus sur les 4 paramètres géométriques (<math>b_1</math>, <math>CD</math>, <math>h</math>, <math>h_r</math>) en fonction de la plage de variation par un PMC d'une couche cachée avec 15 neurones entraîné sur 5000 échantillons pour un profil trapézoïdal symétrique avec couche résiduelle.....</i>	<i>116</i>
<i>Table II-3. Résultats de la caractérisation scattérométrique pour l'échantillon issu de lithographie optique par immersion obtenu sur les trois paramètres géométriques (<math>CD</math> <math>h</math> <math>hr</math>) par 3 méthodes différentes (Krigage, Région de confiance, Réseau de Neurones) .....</i>	<i>119</i>
<i>Table II-4. Performance du PMC (1 couche cachée de 15 neurones) après apprentissage (300 échantillons) évaluée sur les valeurs de <math>E^{test}</math> .....</i>	<i>124</i>
<i>Table II-5. Résultats de caractérisation scattérométrique (RN) et morphologique (XSEM) de l'échantillon Neb22 sur Si de période 750 nm obtenu par nano-impression en considérant un profil trapézoïdal symétrique avec couche résiduelle.....</i>	<i>125</i>
<i>Table II-6. Résultats de caractérisation scattérométriques (RN) obtenues pour le profil trapézoïdal à bord arrondi avec couche résiduelle dans le cas de l'échantillon Neb22 sur Si de période 750 nm obtenu par nano-impression .....</i>	<i>128</i>
<i>Table II-7. Performances et résultats de caractérisation par RN obtenus pour le profil trapézoïdal à bords arrondis pour la caractérisation de l'échantillon Im5010 sur Barc sur Si.....</i>	<i>131</i>
<i>Table II-8. Simulation des 3 échantillons ayant en commun les paramètres <math>CD_{bottom}</math>, <math>h</math> et <math>hr</math> pour différentes valeurs de <math>CD_{top}</math> et <math>r</math> utilisés pour étudier la sensibilité de la scattérométrie au paramètre <math>CD_{top}</math>.....</i>	<i>133</i>
<i>Table II-9. Performances et résultats de caractérisation par RN obtenus pour le profil trapézoïdal à bords arrondis (<math>CD_{top}</math> fixé à 0) pour la caractérisation de l'échantillon Im5010 sur Barc sur Si.....</i>	<i>135</i>
<i>Table II-10. Résultats de caractérisation par RN obtenus avec 2 profils différents (arrondi et elliptique) pour la caractérisation de l'échantillon Im5010 sur Barc sur Si .....</i>	<i>138</i>
<i>Table II-11. Valeur paramétriques des différents profils arrondis (résultats du RN, modélisation optimale de la mesure AFM) discutés dans le cas de la caractérisation de l'échantillon Im5010 sur BARC sur Si.....</i>	<i>141</i>
<i>Table III-1. Plages de variation des 3 paramètres géométriques (<math>CD</math>, <math>h</math>, <math>hr</math>) utilisés pour l'étude comparative bibliothèque/RN. ....</i>	<i>152</i>
<i>Table III-2. Nombre d'échantillons composant le corpus d'apprentissage pour les réseaux de neurones et les bibliothèques en fonction du pas et de la plage de variation choisis. Des cas ont été exclusivement utilisés pour la bibliothèque (vert) et pour les RN (jaune) .....</i>	<i>153</i>
<i>Table III-3. Temps caractéristiques des deux méthodes scattérométriques (RN et BB) pour la résolution du problème inverse pour une caractérisation d'un motif de un profil elliptique.....</i>	<i>173</i>
<i>Table III-4. Caractéristiques des deux RN effectuées pour caractériser l'évolution temporelle d'un procédé théorique de resist trimming .....</i>	<i>176</i>

---

*Table III-5. Résultats des performances des différents RN utilisés sur les valeurs de tests obtenues pour les paramètres géométriques (CD, h, hr) ..... 176*

*Table III-6. Comparaison des résultats obtenue par RN et par mesures CD AFM pour le CD et la hauteur totale obtenue pour un procédé de réduction de côte de résine IM5010 par plasma Hbr/O2 ..... 187*

---

---

---

# Introduction

---

Les dernières décennies ont été marquées par des avancées importantes dans les secteurs de la microélectronique et des nanotechnologies. Présente dans de nombreux aspects de notre vie quotidienne (médecine, téléphone portable, réseaux Internet, transport..), la microélectronique a investi notre société au point d'en devenir un acteur incontournable. Des industriels aux laboratoires académiques en passant par les centres de Recherche et Développement, tout le monde dans ce domaine s'efforce de répondre aux exigences du futur et contribue à la fabrication et à la production de circuits intégrés toujours plus performants, moins coûteux, plus autonomes en énergie et à haute densité d'intégration .

Pour répondre à ces besoins, les acteurs du monde de la microélectronique ont l'habitude de suivre une feuille de route dictée par l'ITRS ( International Technology Roadmap for Semiconductors) dans laquelle est indiquée la prévision des dimensions des transistors au cours des années à venir. Chaque réduction de cette dimension est une étape importante qualifiée de nœud technologique. Les procédés de fabrication (dépôt, lithographie, gravure de matériaux) se trouvent ainsi régulièrement confrontés à de nouveaux défis. Les échelles nanométriques des transistors actuels ne permettent pas de contrôler « à l'œil nu » les dimensions obtenues. Il est même très délicat de délivrer une mesure de façon rapide, précise et robuste. Pourtant, dans le domaine industriel et particulièrement dans celui de la production, où la notion de rendement est fondamentale, ces 3 contraintes sont des éléments clés pour s'assurer du bon fonctionnement des procédés et atteindre les objectifs visés. Par conséquent la métrologie dimensionnelle est devenue une science à part entière dans le domaine de la microélectronique et des nanostructures en général. Le développement d'équipements performants devient ainsi un enjeu crucial pour le bon fonctionnement des futurs composants.

Plusieurs techniques ont été développées, avec les équipements associés, chacune d'entre elle se basant sur des notions physiques différentes. Deux méthodes de métrologie sont particulièrement utilisées dans le domaine industriel, aussi bien en production qu'en R&D : le Microscope à Force Atomique (AFM) et le Microscope Electronique à Balayage (SEM). Ces deux techniques morphologiques permettent de mesurer, à l'échelle du



---

nanomètre, la grandeur caractéristique appelée CD pour Critical Dimension correspondant, par définition, à la plus petite dimension présente dans un circuit. Dans la plupart des cas, il s'agit, pour les circuits intégrés en microélectronique, de la largeur de la grille des transistors.

Le contrôle de procédé en microélectronique [LIM98, EMA04, SU07] consiste à vérifier que les étapes successives s'enchaînent correctement et que les structures obtenues correspondent aux prévisions. La collection de données issues de capteurs présents dans les machines de procédés permet de mettre en évidence une dérive potentielle des équipements. Pour certaines étapes spécifiques, comme la gravure, un suivi de la dimension CD est également effectué. Les techniques AFM et CDSEM, bien que parfois utilisées pour ce suivi ne sont pas optimales en raison de leur lenteur d'acquisition (AFM) ou de la nécessité de transfert de la plaque vers un équipement spécifique, ce qui impacte le rendement. Ces limitations ont encouragé l'émergence d'une autre technique de métrologie dimensionnelle: la scattérométrie qui est basée sur une mesure optique. Cette technique permet de connaître la forme la plus complète possible d'un motif à condition que celui-ci soit périodique. Son principe repose sur l'exposition du motif à un faisceau lumineux, et sur la récupération du signal de diffraction réfléchi par le réseau. Le signal collecté, appelé signature optique, est directement dépendant de la forme du motif diffractant ainsi que du type de matériaux qui le compose. Cette technique, qui sera détaillée dans ce document, possède toutes les qualités requises pour effectuer un suivi de procédé puisque son extraction de résultat est rapide. Elle est de plus non destructive et non invasive.

Aujourd'hui le suivi de procédé souffre d'une contrainte majeure puisque l'information utile n'est acquise qu'en toute fin de procédé. Cela signifie que l'on sait si un procédé a dérivé seulement lorsque le processus est terminé. La correction réalisée permettra alors de rectifier la dérive pour le passage des plaques suivantes, mais à l'heure d'une optimisation toujours plus poussée de la productivité, tout « gaspillage » de plaque de Silicium n'est pas négligeable. C'est pourquoi, depuis quelques années, le LTM a développé une technique fondée sur l'utilisation de la scattérométrie pour effectuer un suivi en temps réel de procédés et notamment les procédés de gravure. Cette technique n'est pas encore assez mature pour être transférable dans l'industrie, mais les travaux de recherche se poursuivent dans ce sens. L'objectif à long terme est de mettre en place des boucles de rétroaction sur l'équipement durant le procédé pour corriger la « recette » en temps réel.

---

Pour retrouver les dimensions géométriques à partir des signatures optiques récoltées, la scattérométrie se base sur un modèle paramétrique définissant par hypothèse la forme de l'échantillon sous test. Les paramètres du modèle sont alors ajustés pour faire correspondre les signatures expérimentales et simulées. Il s'agit de réaliser ici une « résolution du problème inverse » par une optimisation paramétrique. C'est le cœur de cette méthode et, selon les algorithmes utilisés, la scattérométrie peut s'avérer particulièrement performante.

Ces travaux de thèse se positionnent clairement dans une application de la scattérométrie en temps réel. Très peu de travaux existent sur cette thématique dans la littérature [SOU07, ELK09]. Les seules méthodes employées utilisent la méthode des bibliothèques couplée à des processeurs graphiques (GPU) [SOU07] pour la résolution du problème inverse. Les travaux présentés dans cette thèse, proposent de confronter, dans l'objectif du développement de la scattérométrie en temps réel, la méthode des bibliothèques à une technique originale de résolution de problème inverse, les réseaux de neurones artificiels. Cette technique a déjà été appliquée à la scattérométrie [KAL99,ROB02\*] pour la caractérisation de structures diffractantes en régime statique. Cependant, elle n'a jamais été appliquée dans le cadre de la scattérométrie dynamique.

Ce travail de thèse résulte du fruit de la collaboration entre deux laboratoires: le Laboratoire des Technologies de la Microélectronique (LTM) et le Laboratoire Hubert Curien. Le LTM, notamment l'équipe lithographie avancée, travaille sur la scattérométrie avec des moyens expérimentaux en environnement salle blanche, mais également sur la recherche d'algorithmes performants de résolution de problème inverse [SOU08, ELK10]. Le Laboratoire Hubert Curien, notamment l'équipe « Surfaces complexes et composants HF/RF », s'intéresse également à développer de nouveaux algorithmes pour la résolution du problème inverse pour la scattérométrie, notamment par l'utilisation des réseaux de neurones artificiels [ROB03, GER09]. Ce travail de thèse combine le savoir-faire des deux laboratoires pour répondre aux défis de cette technique de métrologie prometteuse.

La première partie de ce document de thèse rappelle les différentes techniques de métrologie couramment utilisées en microélectronique. Les limites des techniques morphologiques sont mises en avant dans le cadre d'une application en temps réel. Le principe de la scattérométrie ainsi que les équipements utilisés, sont ensuite détaillés. Enfin trois méthodes de résolution de problème inverse sont évaluées avec leur positionnement vis-à-vis d'une application en temps réel.

---

Le second chapitre présente les réseaux de neurones (RN) artificiels. Leur principe de fonctionnement est expliqué dans un premier temps. Ensuite une étude préliminaire vise à appréhender le fonctionnement du RN et à en déterminer l'architecture optimale. Enfin, une comparaison est réalisée entre la scattérométrie neuronale et une technique morphologique sur un cas expérimental.

Enfin, la dernière partie a pour objectif d'évaluer les réseaux de neurones pour une caractérisation dynamique. Tout d'abord les deux méthodes de résolution de problèmes inverses que sont les bibliothèques et les Réseaux de Neurones sont étudiées et confrontés dans ce cadre précis afin d'en extraire les spécificités de fonctionnement. Pour conclure, une évaluation des potentialités des RN en dynamique est menée sur un cas expérimental dans le cadre d'un suivi de procédé de resist trimming.

---

## Bibliographie

- [**ELK09**] El Kodadi M., Soulan S., Besacier M., and Schiavone P. , “Real time scatterometry for profile control during resist trimming process”. *Journal of Vacuum Science & Technology B*, 27(6), 3232-3237, 2009.
- [**ELK10**] El Kodadi M., “Développement des techniques de scattérométrie en temps réel pour le suivi des procédés de gravure plasmas”, PhD thesis, Université de Grenoble, 2010.
- [**EMA04**] Emami-Naeini A., Ebert J., Kosut R., De Roover D., and Ghosal S., “Model-based control for semiconductor and advanced materials processing: An overview”, in *Proc. Amer. Contr. Conf.*, Boston, MA, vol. 5, pp.3902-3909 (Jun. 2004).
- [**GER07**] Gereige I., Robert S., Granet G., Jamon D., et Rousseau J. J., « Rapid Control of submicrometer periodic structures by a neural inversion from ellipsometric measurement ». *Optics Communications* 278, 2:270-273, 2007
- [**GER09**] Gereige I., “ Contribution des réseaux de neurons dans le domaine de l’ellipsométrie. Application à la scattérométrie », PhD thesis, Université de Saint Etienne, 2009
- [**KAL99**] Kallioniemi I., Saarinen J., Oja E., «Optical scatterometry with neural network model for nondestructive measurement of submicron features », *Proc. SPIE*, vol. 3743 , 1999.
- [**LIM98**] Limanond S., Si J., and Tsakalis K., “Monitoring and control of semiconductor manufacturing processes”, *IEEE Control Syst. Mag.*, vol. 18, no. 6, pp.46-58, 1998
- [**ROB02**] Robert S., Mure Ravaud A., Reynaud S., Fourment S., Carcenac F., Arguel P., « Experimental characterization of subwavelength diffraction gratings by an inverse-scattering neural method ». *Journal of the Optical Society of America A* 19, 12: 2394-2402.,2002
- [**ROB03**] Robert S., « Caractérisation optique de réseaux de diffraction submicroniques par des techniques génétique et neuronale ». PhD thesis, Université Jean Monnet de Saint Etienne, 2003.

---

**[SOU07]** Soulan S., Besacier M, Schiavone P, Leveder T, “In-line etching process control using dynamic scatterometry“, Proc. SPIE 6617, Modeling Aspects in Optical Metrology, 661713 (June 18, 2007).

**[SOU08]** Soulan S., “Développement de la scattérométrie dynamique pour le suivi en temps réel de procédés. Application à la microélectronique. ». PhD thesis, Université de Grenoble, 2008.

**[SU07]** Su A.J., Jeng J.C., Huang H.P., Yu C. C., Hung S.Y., Chao C.K. , « Control relevant issues in semiconductor manufacturing: Overview with some new results”, Contr. Eng. Practice, vol. 15, no. 10, pp. 1268-1279 (2007).

---



---

# **Chapitre I. Techniques de métrologie pour la microélectronique**

---



Avec la miniaturisation des composants électroniques, il devient indispensable d'avoir recours à des outils de métrologie performants capables de déterminer rapidement et efficacement les propriétés géométriques d'un motif. Plusieurs techniques de mesure sont utilisées en microélectronique pour atteindre cet objectif. Chacune d'entre elles possède des spécificités qui la rendent plus ou moins adaptée selon le cas de figure. Nous allons, à travers ce chapitre, présenter ces méthodes, que l'on classera en différents groupes. Nous détaillerons particulièrement la technique de scattérométrie qui permet une métrologie rapide, non destructive, précise et à faible coût. Cette technique fondée sur une mesure de réponse optique est une technique indirecte. Cela signifie qu'elle ne délivre pas « directement » les dimensions à connaître mais nécessite l'utilisation d'un algorithme de problème inverse. Nous définirons dans la suite en quoi cela consiste ainsi que les différentes techniques de résolution qui lui sont associées. Nous détaillerons leur principe de fonctionnement afin d'estimer leurs points forts et leurs limites vis-à-vis de notre problématique, la scattérométrie in-situ en temps réel.

## **I.1 Les techniques morphologiques**

Les techniques morphologiques de caractérisation sont largement utilisées aussi bien dans le domaine de la recherche qu'en industrie afin de pouvoir déterminer les dimensions paramétriques d'un motif à une échelle nanométrique. Nous allons présenter deux techniques : la microscopie électronique à balayage (MEB) et la microscopie à force atomique (AFM). Nous évoquerons les spécificités de chacune ainsi que leurs avantages et inconvénients.

### **I.1.1 Microscopie Electronique à Balayage**

Le Microscope Electronique à Balayage [LEE93, LEG] ou le SEM (d'après l'acronyme anglo-saxon Scanning Electron Microscopy) repose sur l'interaction matériau-électrons. Cette technique consiste à envoyer un faisceau d'électrons sur la surface à analyser. Cette dernière émet des électrons « rétro diffusés » de différentes natures qu'il est possible de détecter et

ainsi connaître la topographie de la surface insolée. Au niveau du faisceau d'émission, les électrons, chargés négativement, sont extraits d'un filament chauffé (effet thermoïonique) ou d'une pointe portée à un très haut potentiel négatif (émission de champ). Les électrons sont ensuite accélérés à une forte énergie et sont focalisés par des bobines magnétiques (Figure I-1).

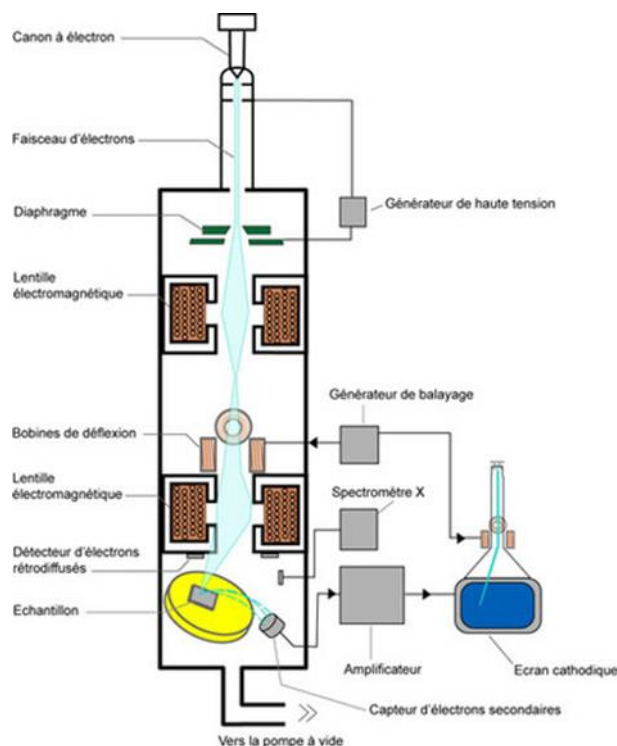


Figure I-1. Principe de fonctionnement du SEM [REY]

La collision entre les électrons et la surface du matériau, lors de chocs inélastiques, crée l'émission d'électrons secondaires ainsi que des électrons rétrodiffusés. Suivant les paramètres de balayage, les électrons secondaires vont être détectés et permettront ainsi de constituer une image restituant un contraste chimique ou topographique.

Lors de l'exposition du matériau aux électrons, on peut voir apparaître des effets de charge: les charges électriques s'accumulent et leur énergie peut chauffer le matériau, entraînant la modification des propriétés géométriques du motif. Le plus souvent le motif va subir une réduction de taille suite à l'échauffement. Pour pallier à cela, on peut diminuer le nombre de frame ou métalliser l'échantillon pour éviter que les charges s'accumulent. Ceci est particulièrement vrai pour l'observation de résines qui sont constituées de polymères et sont donc de nature isolante.

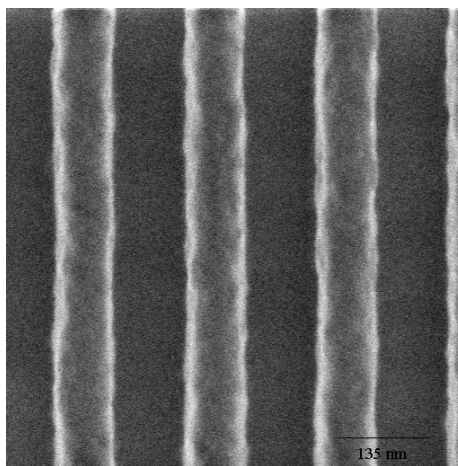
On distingue deux catégories de SEM : le CD-SEM (CD pour désigner « critical dimension ») et le XSEM (X pour désigner «cross section»).

### ***1.1.1.1 Le CD-SEM***

Le CD-SEM permet d'observer les motifs en vue de dessus. Les électrons sont accélérés avec une énergie de 300 à 800 eV. Ceci implique des détecteurs très performants, mais permet de s'affranchir de l'étape de métallisation et de clivage, quelque soit le type de matériaux. C'est donc une technique non destructive. L'échantillon peut en conséquent être réutilisé et c'est pour cette raison que le CD-SEM est une technique largement utilisée en industrie pour le contrôle en ligne sur des sites de production.

Cette technique permet d'extraire le CD de manière très précise grâce à l'emploi de techniques de traitements d'images adaptés. (Figure I-2). La taille de la fenêtre de mesure peut être variable (ici  $0.676 \mu\text{m} \times 0.676 \mu\text{m}$ ).

Avec le CD-SEM, les mesures étant effectuées de façon automatisée, il est possible de réaliser des statistiques sur un grand nombre de mesures d'un même motif. Le CD-SEM présente également l'avantage de pouvoir obtenir de l'information sur la rugosité des motifs.



*Figure I-2. Image de ligne de résine sur Silicium, de période  $d=189 \text{ nm}$ , obtenue avec le CD-SEM (tension d'accélération= $300\text{V}$ , grossissement= $199583$ , taille du champ d'observation= $0.676 \mu\text{m}$ ,  $0.676 \mu\text{m}$ )*

Le modèle de CD-SEM disponible au LTM (Hitachi CG4000 du Leti) ne permet cependant pas la mesure des hauteurs des couches sous-jacentes. Actuellement, peu d'équipements peuvent effectuer cette tâche et ils ne permettent pas encore de bien visualiser le contraste. Nous n'avons donc pas utilisé cette technique dans le cadre de nos travaux. La rapidité de cette technique dépend du nombre d'images souhaité. Non destructive, et précise,

elle fait partie des métrologies présentes sur les sites de production pour le contrôle en ligne des procédés. Cette technique consomme malgré tout du temps car elle utilise des algorithmes pour traiter l'image. Le résultat de la mesure ne peut pas être obtenu immédiatement et cette technique ne peut donc pas être utilisée pour une caractérisation en temps réel.

### 1.1.1.2 Le XSEM

Le XSEM consiste à observer la section d'un échantillon (Figure I-3). Il s'agit d'une méthode de métrologie dite destructive car cette observation nécessite de sectionner la plaque à l'endroit de la mesure. Dans le XSEM, les électrons sont accélérés à une tension élevée (3000 à 30000 eV), et les détecteurs utilisés sont moins sensibles que ceux du CD-SEM. Il est donc nécessaire de passer par une étape de métallisation. Celle-ci rend l'échantillon par la suite inutilisable y compris pour une caractérisation optique. Le XSEM utilisé dans le cadre de cette thèse est conçu par Hitachi. Il permet de caractériser des profils verticalement, et donc de mesurer des motifs de surface, mais aussi des couches sous-jacentes. C'est pour cette raison qu'on l'a utilisé dans le cadre de nos travaux. La Figure I-3 montre un exemple d'images obtenues dans le cas d'un échantillon métallisé de résine NEB22 sur substrat de Silicium. La NEB22 est une résine aussi bien utilisée pour la lithographie électronique que dans la technique de lithographie par nanoimpression.

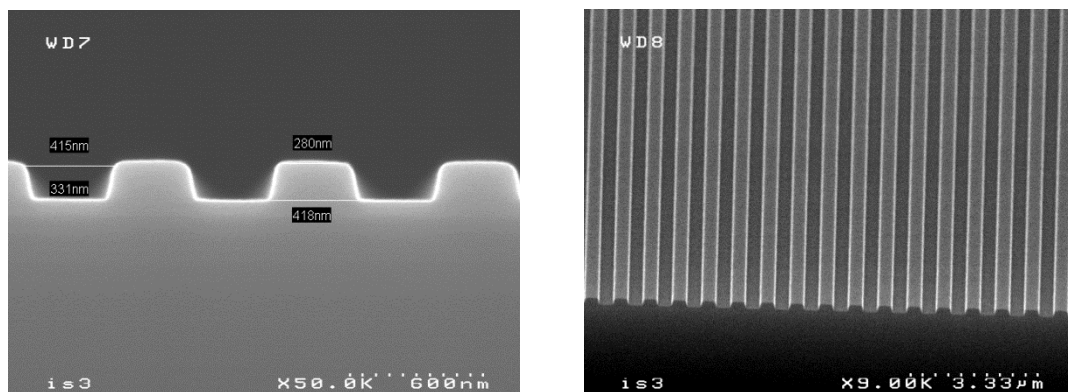


Figure I-3. Image MEB d'un échantillon de Neb22 sur Si, de période  $d=750\text{nm}$ , en cross section (à gauche), en vue inclinée (à droite)

L'utilisation de cette technique de XSEM nécessite un temps important pour la préparation de l'échantillon. Elle est ainsi bien adaptée pour le milieu de la Recherche et Développement. Ces spécificités la rendent difficile à mettre en œuvre dans le milieu de la production industrielle.

Le temps nécessaire pour prendre une image XSEM dépend de plusieurs paramètres. Bien que le temps pour que le faisceau d'électrons balaye la surface est de quelques secondes, il faut considérer le temps de manipulation de l'appareil (temps de pompage, temps pour mettre l'échantillon). Pour donner un ordre de grandeur, une mesure peut prendre de quelques minutes à une dizaine de minutes. De plus, la précision de cet appareil est donnée à 10% près, mais cela reste un ordre de grandeur.

La mesure du profil à partir de l'image obtenue par XSEM se fait manuellement à l'aide d'un éditeur d'image dédié. Dans certaines images, on voit apparaître un filet blanc (Figure I-3) qui entoure le profil. Ceci provient du réglage de l'équipement (paramètres à régler: focus, contraste) mais peut révéler des éléments sur le profil (rugosité de surface).

De manière générale, l'appareil SEM permet d'avoir une information locale, avec des informations sur la rugosité, et l'homogénéité de l'échantillon. Ceci n'est pas le cas des méthodes de caractérisation optique, où le résultat donné est une moyenne faite sur l'ensemble de la surface éclairée par le faisceau optique (taille du spot millimétrique).

## **I.1.2 L'AFM et le CD-AFM**

### ***I.1.2.1 L'AFM***

Le microscope à force atomique [RIV05] a été conçu en premier lieu par G. Binnig, C. Quate et C. Gerber en 1986 pour succéder au microscope à effet tunnel (STM-Scanning Tunneling Microscopy). Il permet de caractériser différents types d'échantillons (conducteur, semi-conducteur, isolant ou biologique). Il présente l'avantage d'être utilisable dans une multitude d'environnements (air, ultravide, liquide). Le principe repose sur la loi d'attraction et de répulsion entre une pointe placée à l'extrémité d'une poutre et la surface du matériau à caractériser (Figure I-4). Un laser est concentré sur la face arrière de la pointe et la réflexion du faisceau est captée par un système de photodiode. Le mouvement de la pointe est ainsi transcrit sur un plan à deux dimensions et permet de reconstituer un profil.

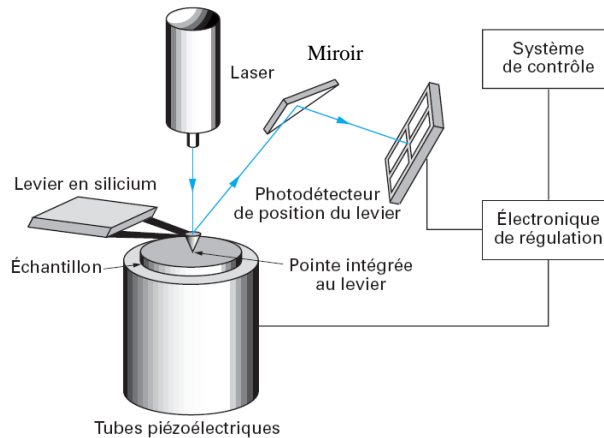


Figure I-4. Schéma de fonctionnement du microscope AFM [POI08]

Il faut noter que la pointe représente l'élément essentiel qui détermine la qualité d'une mesure. Une pointe AFM est composée de silicium et elle est souvent recouverte d'un matériau selon le type d'utilisation. Il existe différents types de pointes : on peut par exemple citer la pointe Pt/Ir, recouverte d'un alliage de Platine/Iridium (Figure I-5-à gauche) et la pointe dite diamant, recouverte d'une couche de diamant (Figure I-5-à droite)

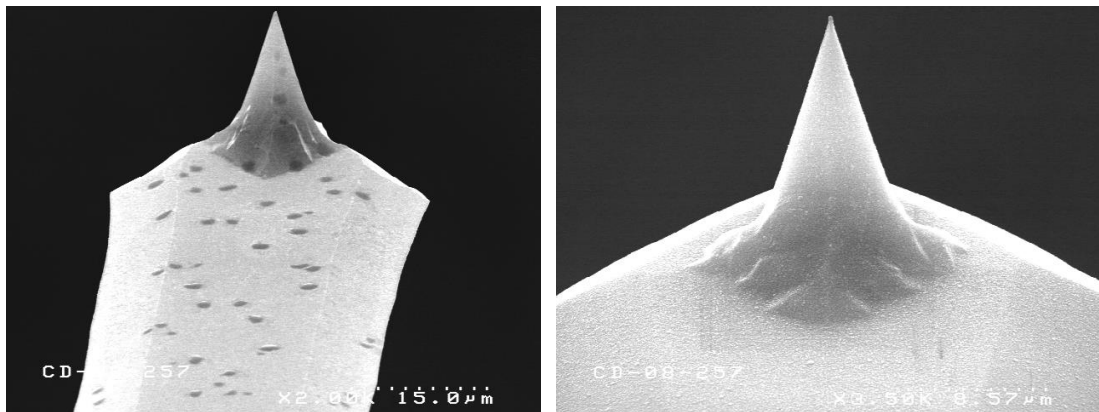


Figure I-5. Photographie MEB de pointes AFM : PT/Ir (à gauche), diamant (à droite) [POI08]

Plusieurs modes permettent de faire une mesure AFM. Dans le mode contact, la pointe est maintenue en contact répulsif permanent avec l'échantillon. Celui-ci est balayé ligne par ligne. La force avec laquelle la pointe appuie sur la surface est fixée à 20 nN. L'asservissement consiste ensuite à maintenir cette force constante en faisant varier la hauteur de la pointe. Ce mode est le plus simple à mettre en œuvre.

Dans le mode tapping (Figure I-6), le contact permanent est remplacé par un contact intermittent. La pointe touche la surface lors du passage à plus faible amplitude et elle oscille à une fréquence proche de sa fréquence de résonance.

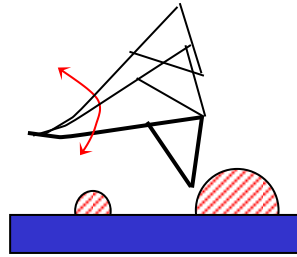


Figure I-6. Schéma de la pointe de l'AFM scrutant la surface en mode tapping [POI08]

Enfin, le troisième mode est intitulé non contact ou résonant. Il est caractérisé par une pointe qui est à distance moyenne d'au moins 10 nm de la surface et elle oscille à une fréquence proche de la fréquence de résonance.

La forme de la pointe présente une limitation pour la mesure de certains profils comme le montre l'illustration de la Figure I-7 avec une pointe conique. En effet, la forme du profil donné par AFM est différente du profil réel, avec l'ajout de pente sur les flancs du motif. Il est parfois nécessaire de faire une étape de traitement qui consiste à déconvoluer la forme de la pointe, pour retrouver la forme du profil original. Cela consiste à prendre en compte la forme de la pointe dans le résultat obtenu.



Figure I-7. Déformation du profil par une pointe AFM. On observe que le profil reconstitué par AFM (trait en pointillé) n'est pas le même que le profil réel (trait en continue)

L'AFM utilisé dans le cadre de cette thèse est l'enviroscope de Veeco. Il nécessite d'avoir à disposition des échantillons de 1 cm<sup>2</sup>. Ce type d'appareil est donc intrinsèquement destructeur. Cette surface à explorer est beaucoup plus importante que celle du XSEM. Cependant, il existe d'autres AFM (cf. I.2.2) où il est possible de faire des mesures sur des plaques entières (la surface mesurée reste petite (1µm) dans ce cas de figure). La Figure I-8 donne un aperçu d'une image obtenue avec un AFM classique pour la mesure de lignes denses.

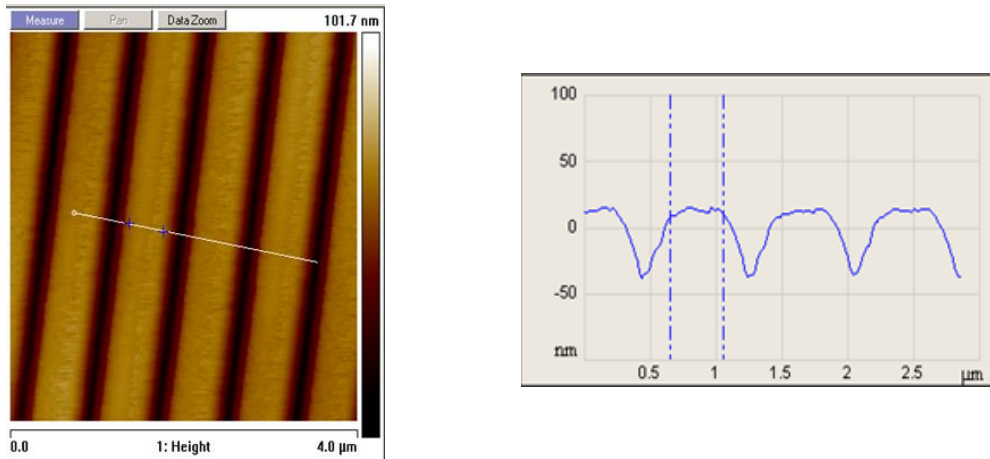


Figure I-8. Exemple d'image brute avant déconvolution obtenue avec un AFM classique

### I.1.2.2 Le CD-AFM

Le principe de fonctionnement du CD-AFM [THI06, MAR94] est identique à celui de l'AFM classique. On se base sur l'interaction entre une pointe fixée sur un levier et une surface à analyser. Le CD-AFM se distingue de l'AFM classique par le fait que la pointe utilisée pour examiner le profil est une pointe évasée, dite « en patte d'éléphant » (Figure I-9).

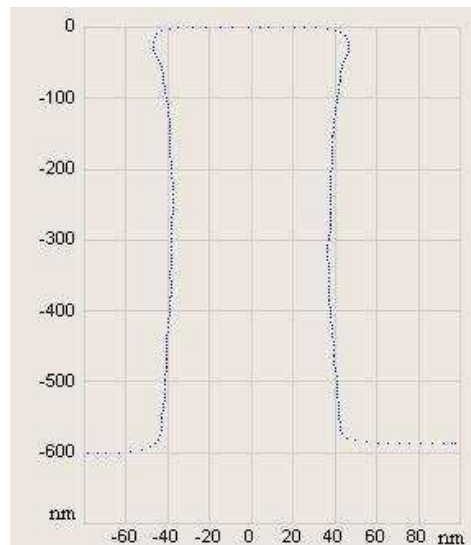


Figure I-9. Profil représentant la forme de la pointe utilisée par un CD-AFM de type CDR120 nm (Critical Dimension re-entrant profile) [ELK10]

Cette forme particulière permet de mesurer la topographie des flancs d'un motif de manière beaucoup plus précise que l'AFM classique. Une information est obtenue sur la rugosité et les ondulations du motif (Figure I-10).



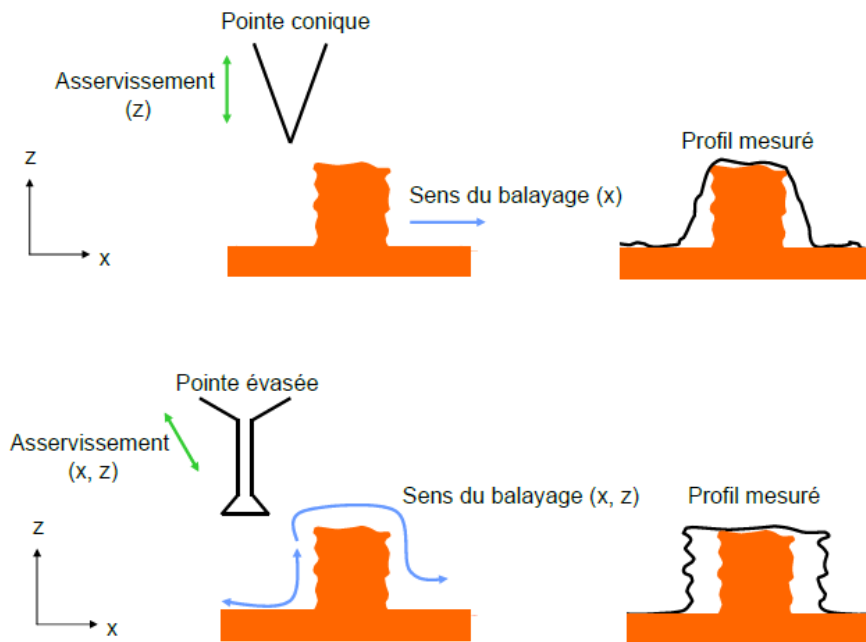


Figure I-10. Schématisation d'une reconstruction du profil avec un AFM classique (figure du haut) et avec un CD AFM (figure bas). [AZA12].

La verticalité est rétablie beaucoup plus précisément. De la même manière que l'AFM classique, il est également nécessaire de déconvoluer la forme de la pointe sur le résultat obtenu pour obtenir l'allure exacte du profil. La Figure I-11 représente un profil 3D constitué d'une couche de résine sur du BARC (Bottom Anti Reflective Coating) et du Silicium. Le BARC est une couche intermédiaire qui permet de réduire la réflexion de la lumière lors de l'étape de lithographie.

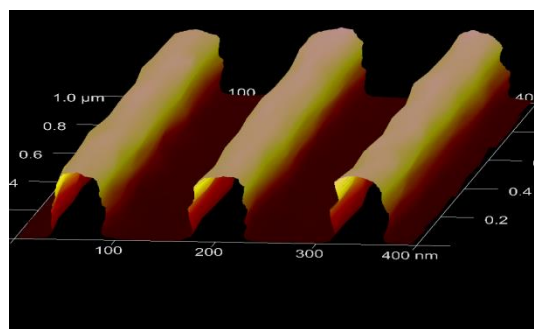


Figure I-11. Image traitée en 3D obtenue par CD AFM pour un profil BARC sur Si

L'utilisation du CD-AFM ou de l'AFM nécessite un temps de mesure assez conséquent (chaque mesure prend plusieurs minutes), de par le déplacement de la pointe et du traitement post mesure. Le CD-AFM possède aussi un caractère complexe d'utilisation. Ce n'est donc pas un outil utilisable sur des lignes de production mais on le retrouve plus fréquemment dans

les environnements de Recherche et Développement car sa capacité à mesurer la pente des flancs d'un profil ainsi que la rugosité de ligne est particulièrement bonne. En contrepartie, l'information sur les couches sous-jacentes au profil est indisponible.

Les méthodes microscopiques présentées permettent de connaître la forme du profil (CD, hauteur, pente, etc...) ainsi qu'éventuellement pour certaines la hauteur des couches sous-jacentes. Il s'agit de méthodes directes ne nécessitant pas de modèle paramétrique et permettant donc de déterminer les grandeurs géométriques sans connaissance à priori du profil. Dans le cadre de ces travaux de thèse, ces techniques feront office de méthodes de mesures comparatives. Cependant, elles sont parfois destructives et nécessitent un temps de mesure important. Ceci rend ces méthodes non adaptées pour l'étude in situ en temps réel. La métrologie optique, notamment avec la scattérométrie, permet quant à elle, de répondre à ces contraintes,

## **I.2 La scattérométrie**

La scattérométrie est une technique de métrologie par voie optique utilisant l'information contenue dans la diffraction d'un rayon lumineux sur un motif périodique pour retrouver notamment les caractéristiques géométriques de ce dernier. Cette technique possède plusieurs avantages qui la rendent adaptée pour une application de mesure en temps réel. Cette technique se base, à partir de l'élaboration d'un modèle paramétrique, sur la résolution d'un problème inverse. Dans ce chapitre, nous allons présenter plusieurs outils qui permettent d'effectuer la mesure à traiter. Puis, nous exposerons les différentes méthodes permettant de résoudre le problème inverse pour extraire, à partir de cette signature mesurée, les paramètres géométriques de la structure.

### **I.2.1 Principe de la scattérométrie**

#### ***I.2.1.1 Définition***

Le mot « scatterométrie » vient de l'association entre deux termes: scatter (diffusion de la lumière), et metro (mesure). Il signifie mesure de la lumière diffusée ou diffractée. La scattérométrie est une technique de métrologie optique qui a eu un grand succès ces dernières années, tant dans le milieu de la recherche que dans le domaine industriel, notamment grâce à sa rapidité de mesure, sa précision et son caractère non destructif. Dans le domaine de la

microélectronique, la scattérométrie est aujourd'hui utilisée dans le contrôle de la qualité de fabrication après l'étape de développement en lithographie et après l'étape de gravure.

Le principe de la scattérométrie est présenté dans la Figure I-12.

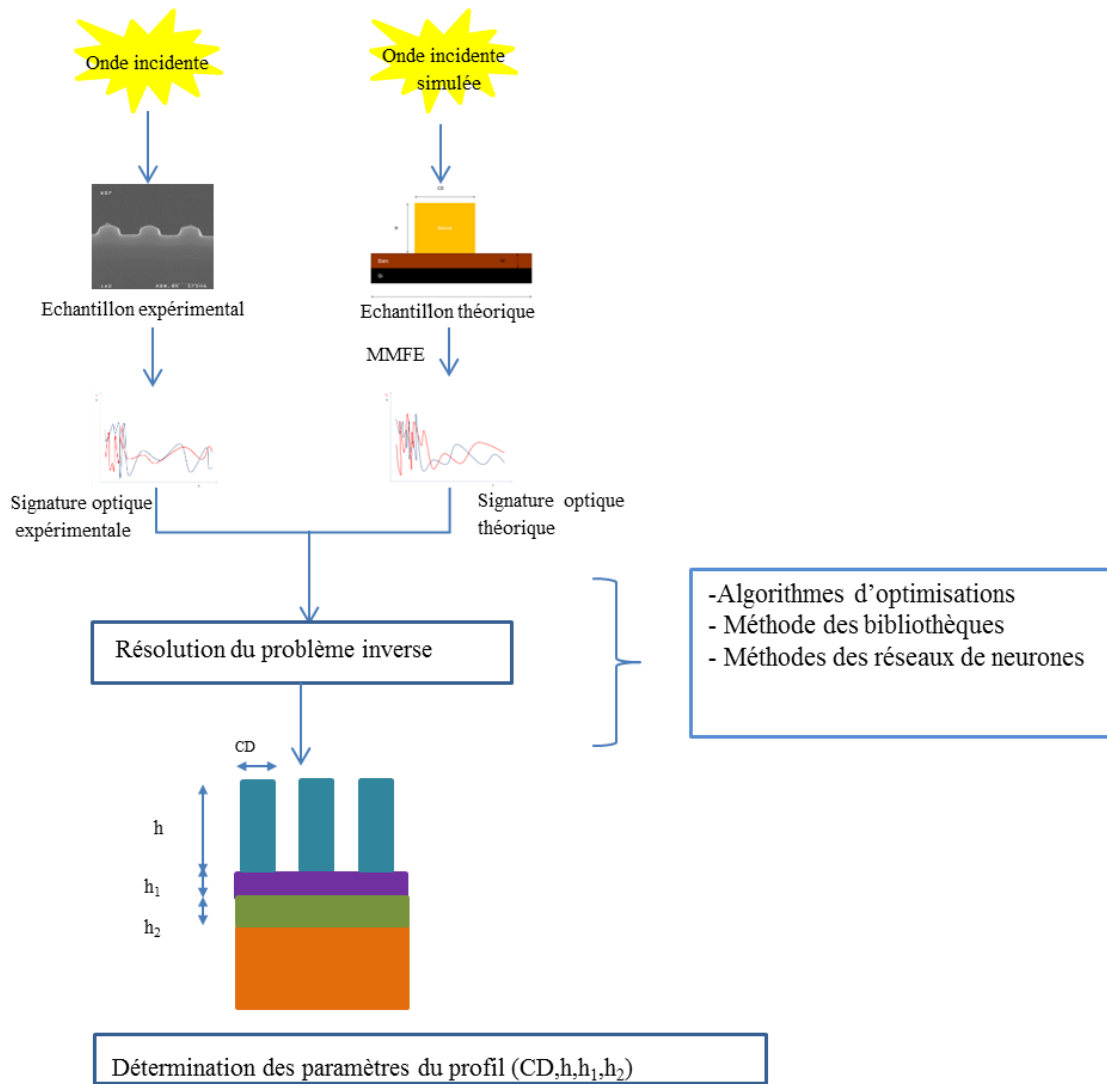


Figure I-12. Schéma illustratif du principe de la résolution du problème inverse par scattérométrie

Dans un premier temps, on mesure une signature optique de l'échantillon. En parallèle, on simule par une méthode numérique exacte et rigoureuse (par exemple la MMFE (cf. §I.2.1.3.)) cette mesure optique obtenue à partir d'une modélisation supposée assez proche du profil de ce même échantillon. On parle de calcul direct. Celui-ci n'est possible que dans le cas où l'échantillon est composé de réseaux périodiques. Dans un second temps on résout le problème inverse par confrontation des signatures théoriques et expérimentales afin d'en déduire les paramètres géométriques recherchés caractérisant la forme du profil. Cette

dernière étape peut être réalisée par différentes méthodes: les algorithmes d'optimisation classiques, la méthode des bibliothèques et les méthodes de régression auxquelles peuvent s'apparenter les réseaux de neurones. La détermination des paramètres géométriques par scattérométrie consiste donc en la résolution du problème inverse, ou problème indirect. Un exemple de structure rencontré dans le cadre de ce travail sera examiné par la suite.

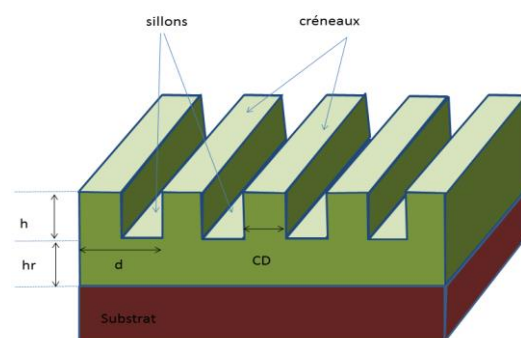
### ***1.2.1.2 Le réseau de diffraction***

La scattérométrie nécessite la présence d'une structure diffractante périodique sur laquelle va se réfléchir l'onde incidente. C'est une condition nécessaire pour pouvoir appliquer la méthode de calcul numérique exacte et rigoureuse (MMFE).

Il est nécessaire de disposer d'un échantillon de référence afin d'estimer et de comparer les performances d'une méthode scattérométrique. Cet échantillon, un réseau de diffraction périodique selon une ou deux dimensions de l'espace, doit être connu d'un point de vue topographique afin de pouvoir établir un modèle paramétrique aussi proche que possible de la réalité physique.

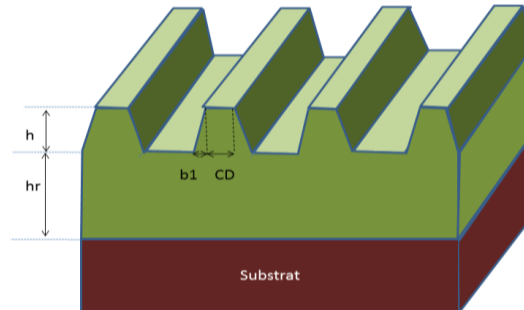
La mise en place de ce modèle paramétrique peut, par exemple, résulter de l'observation de résultats obtenus par les techniques de métrologie telles que le XSEM ou l'AFM (cf § II.1).

La représentation la plus élémentaire est un réseau périodique selon une dimension avec un profil créneau défini avec deux paramètres ( $CD$ ,  $h$ ). Pour s'approcher des structures utilisées dans le cadre de nos travaux, un troisième paramètre  $hr$  a été rajouté caractérisant l'épaisseur de la couche résiduelle présente sous les lignes. (Figure I-13)



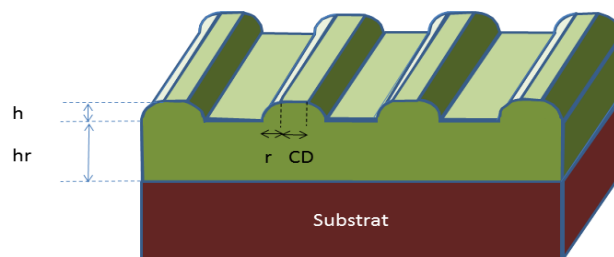
***Figure I-13. Modélisation paramétrique de réseau avec profil créneau sur couche résiduelle défini par trois paramètres ( $CD$ ,  $h$  et  $hr$ )***

On peut apporter une évolution à ce profil en considérant des créneaux avec des flancs non verticaux; ce qui conduit à un profil trapézoïdal défini par 4 paramètres ( $CD, b_1, h, h_r$ )(Figure I-14)



*Figure I-14. Modélisation paramétrique de réseau avec profil trapézoïdal sur couche résiduelle défini par 4 paramètres ( $CD, b_1, h, h_r$ )*

Enfin, un autre cas de figure étudié est un profil arrondi avec 4 paramètres ( $CD, r, h$  et  $h_r$ ) (Figure I-15)



*Figure I-15. Modélisation paramétrique de réseau avec profil arrondi sur couche résiduelle défini par 4 paramètres ( $CD, r, h, h_r$ )*

Ce modèle est également illustré avec une mesure de CD-AFM (Figure I-16). L'échantillon est fabriqué chez STMicroelectronics par lithographie optique par immersion à 193nm. Il est composé d'une couche de résine sur du BARC (Bottom Anti Reflective Coating) sur un substrat de silicium.

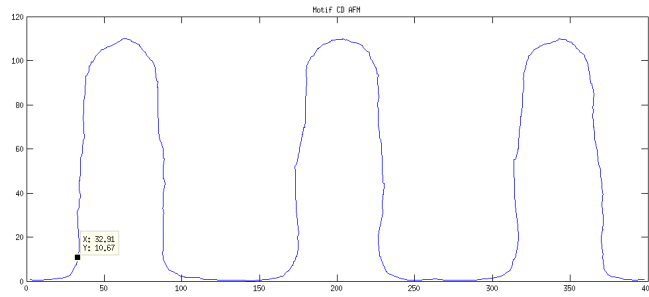


Figure I-16. Profil mesuré par CD-AFM concernant une ligne de résine IM5010 (STMicroelectronics) surplombant une couche de Barc sur du Si.

Différents exemples de profil ont été ainsi présentés. Des formes plus complexes seront également rencontrées par la suite (cf. Chapitre II), prenant en compte un nombre de paramètres plus élevé.

La Figure I-17 représente des images de différents profils de lignes fabriquées par nanoimpression (cf-Annexe 1) obtenues par XSEM. A partir de ces observations et pour ces cas de figure, on constate que ces motifs peuvent être modélisés par des formes trapézoïdales avec des bords arrondis plus ou moins prononcés sur le haut du motif. Ce type de profil est certainement le plus général et le plus simple pour traiter des réseaux de lignes de résine. C'est par conséquent le type de profil le plus adapté pour l'étude qui sera menée dans la suite du problème inverse en scattérométrie. Il faut noter que ce profil inclut le modèle de type créneau, lorsque la pente des flancs est très raide et que l'arrondi est très faible.

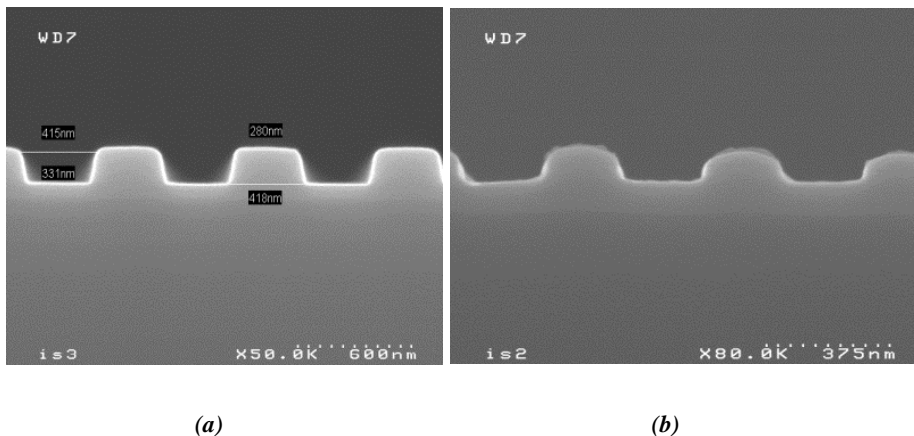


Figure I-17. Images XSEM (a et b) obtenues pour des motifs composés de résine Neb22 sur substrat de Si.

### I.2.1.3 Principe de la diffraction

Pour étudier la diffraction d'une onde sur un réseau, considérons une onde électromagnétique plane de vecteur d'onde  $\vec{k}_i (k_i = N_r k_0, k_0 = \frac{2\pi}{\lambda_0})$  avec  $N_r$ , l'indice de réfraction du milieu dans lequel se propage l'onde et  $\lambda_0$  la longueur d'onde de propagation dans le vide. L'équation de propagation de l'onde dans le vide est donnée par l'équation I.1:

$$\Delta \vec{E} - \frac{1}{c^2} \frac{\partial^2 \vec{E}}{\partial t^2} = \vec{0} \quad (I.1)$$

$c$  étant la célérité de la lumière dans le vide. Cette équation est valable aussi bien pour le champ électrique  $\vec{E}$  que pour le champ magnétique  $\vec{H}$ .

On considère que l'onde est incidente sur une surface périodique avec un angle  $\theta_i$  (Figure I-18). Ce dernier est l'angle que fait la direction de propagation  $\vec{k}_i$  de l'onde incidente avec la normale au réseau. L'incidence classique est considérée ici, ce qui signifie que le plan d'incidence formé par  $\vec{k}_i$  et la normale est perpendiculaire aux lignes. Le réseau contient des motifs de période  $d$  et sépare deux milieux d'indices respectifs  $N_1$  et  $N_2$ .  $N_h$  représentera l'indice du milieu dans lequel on observe la diffraction ( $N_h = N_1$  pour les faisceaux réfléchis et  $N_h = N_2$  pour les faisceaux transmis). L'équation fondamentale des réseaux se note ainsi:

$$N_h \sin \theta_{dm} = N_1 \sin \theta_i + m \frac{K_r}{k_0} \quad (I.2)$$

$$= N_1 \sin \theta_i + m \frac{\lambda_0}{d}, \quad m = 0, \pm 1, \pm 2 \dots \quad (I.3)$$

avec le vecteur réseau  $\vec{K}_r$  tel que  $K_r = 2\pi/d$ , défini comme le vecteur appartenant à la fois au plan d'incidence et au plan du réseau. Le vecteur  $\vec{K}_r$  est donc perpendiculaire à la ligne. La variable  $m$  représente l'ordre de l'onde diffractée. L'angle  $\theta_{hm}$  donne l'angle de diffraction de l'onde dans le milieu  $h$  ( $h=1$  ou  $h=2$ ) à l'ordre  $m$ .

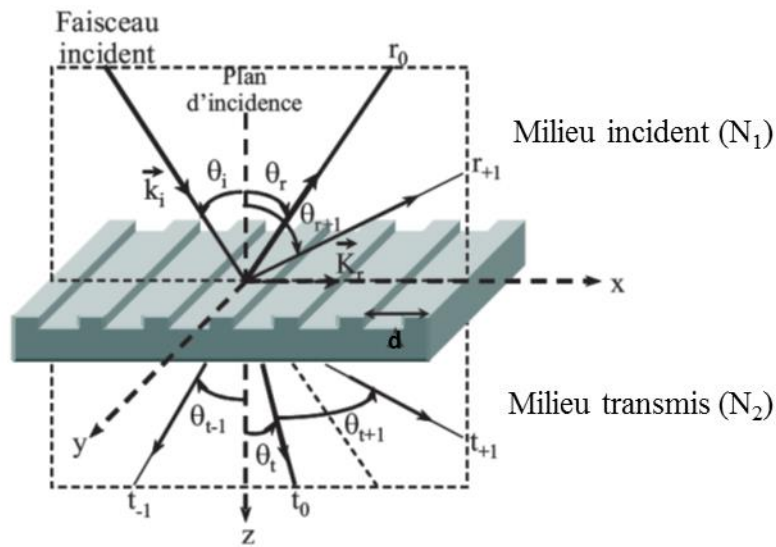


Figure I-18 . Diffraction par un réseau périodique en configuration classique [GER09]

On peut également considérer le cas où le plan d'incidence fait un angle  $\varphi$  avec la normale au réseau. On le nomme angle azimutal. Dans ce cas précis, on parle de diffraction conique par opposition à la configuration classique (Figure I-19).

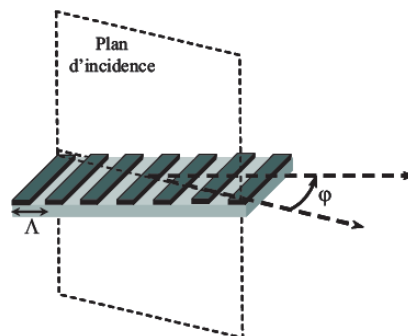


Figure I-19. Configuration de diffraction conique définie par un angle azimutal  $\varphi$  [GER09]

Résoudre le problème de diffraction d'une onde électromagnétique à travers un réseau revient à répondre aux questions suivantes : quelle est la distribution du champ électrique dans les différents ordres et dans les différents milieux ? Quel est l'influence de la géométrie du profil sur le champ électrique diffracté ?

Pour cela, il existe deux façons de résoudre ce problème : soit on utilise des méthodes approximatives (avec des hypothèses contraignantes), soit on fait appel à des méthodes numériques rigoureuses. A noter que les hypothèses mentionnées ici ne concernent pas la géométrie du profil du motif étudié mais bien la façon de représenter le champ électromagnétique.



L'intérêt des méthodes approximatives est principalement la rapidité des calculs permettant leur mise en œuvre dans des délais raisonnables sur des machines classiques. En contrepartie, elles donnent des résultats « grossiers » du comportement physique du phénomène. A titre d'exemple, la théorie scalaire de Fresnel-Kirchoff [GOO05] est une méthode approximative qui consiste à considérer la propagation de la lumière comme un phénomène scalaire, sans prendre en compte la dimension vectorielle de l'onde électromagnétique. Une autre approche approximative se base sur l'hypothèse de Rayleigh en considérant le problème de la diffraction par la voie électromagnétique. Le système se ramène à résoudre un système d'équation linéaire [VAN81][MIL73].

Les méthodes rigoureuses, aussi qualifiées de « méthodes exactes », consistent quant à elles, à résoudre les équations de Maxwell sans aucune approximation sur le plan théorique. On peut citer tout d'abord la méthode intégrale [MAY72, PET65, PET66] qui se base sur le théorème de Green. Le champ électrique est représenté sur chaque point de l'espace par une intégrale sur la surface du réseau. Un courant de surface est généré et le champ diffracté est obtenu par rayonnement de ce courant sur la surface. Dans la méthode des éléments finis [DEL93], on représente la surface par un maillage. Le champ électrique est la somme des fonctions élémentaires sur chaque cellule du maillage. Le problème est ramené à un système d'équations linéaires algébriques. Dans la méthode des sources fictives [TAY94], on considère qu'un champ électrique est généré par des sources fictives. Celles-ci sont choisies de manière à vérifier les équations de Maxwell.

La méthode différentielle [NEV71, PET80], se base sur le fait de considérer les équations de Maxwell sous leur forme différentielle. Le champ électrique et la permittivité sont ainsi représentés en série de Fourier. Dans les méthodes différentielles, on peut également citer la méthode C [CHA80, GRA97] où on effectue un changement de coordonnées simple séparant l'espace en deux milieux homogènes. Enfin la méthode que nous allons utiliser dans la suite de nos travaux est la méthode RCWA (Rigorous coupled Wave Analysis) également connue sous l'acronyme MMFE (Modal Method by Fourier Expansion) [MOH81, LI93, GRA96]. C'est une technique qui consiste à exprimer les solutions de l'équation de Maxwell sous la forme de mode de propagation dans un réseau de forme rectangulaire. Elle se base sur un découpage du motif en tranches rectangulaires dans lesquelles, les modes sont évaluées indépendamment (Figure I-20). Ensuite, les matrices de passage entre chaque couche sont calculées et permettent d'obtenir les coefficients de réflexion nécessaires à la détermination des grandeurs scattérométriques (cf. chap II.2.2.2).

Dans le cadre de nos travaux, nous avons fait le choix de la technique MMMFE car c'est la méthode utilisée dans le laboratoire LTM et le LT2C.

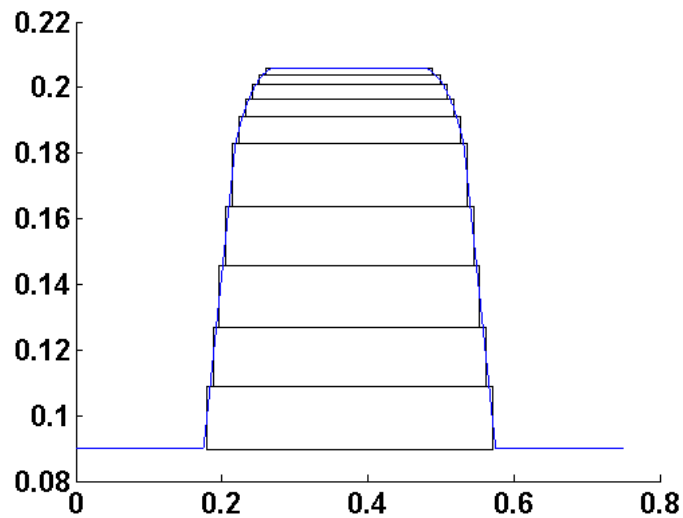


Figure I-20. Exemple de profil trapézoïdal à bords arrondis ( ) découpé en 10 tranches pour la méthode MMFE

En scattérométrie, l'étape de simulation du champ diffracté par le modèle paramétrique est une étape primordiale puisque c'est sur la base de ce modèle là que doit être retrouvé le motif réel. Ce problème direct consiste donc à simuler une signature optique à partir d'une modélisation de l'échantillon testé, ceci à l'aide du champ électromagnétique calculé par la MMFE présentée précédemment.

D'un point de vue plus mathématique, si on représente les paramètres géométriques définissant un profil donné par le vecteur  $\vec{x}$  et l'ensemble de  $b$  mesures (définis par le nombre de longueur d'ondes retenu) composant la signature optique par le vecteur  $\vec{y}$ , ce problème direct peut être représenté par la fonction  $F(\vec{a}, \vec{x})$  définie de la manière suivante :

$$F: \vec{x} \rightarrow \vec{y} = F_{\vec{a}}(\vec{x}) = F(\vec{a}, \vec{x}) \quad (I.4)$$

avec  $\vec{a}$  le vecteur définissant à la fois une configuration expérimentale donnée (angle d'incidence, longueurs d'ondes utilisées) et les propriétés optiques des matériaux supposées constantes (indices  $n$  et  $k$ ).

On va ensuite s'intéresser maintenant aux différents types de signatures optiques couramment utilisées en scattérométrie ainsi que les différents appareils de mesures correspondants.

## I.2.2 Les différents types de signatures optiques

La scattérométrie se base non pas sur la mesure dimensionnelle du motif lui-même comme les méthodes de caractérisation microscopique mais sur le résultat de son interaction avec une onde électromagnétique. Cette mesure étant donc directement liée aux paramètres géométriques de la structure, elle constitue ce qu'on appelle la signature optique.

Pour obtenir une signature optique, on effectue soit une mesure d'intensité (scattérométrie réflectométrique [KAL99, ROB02\*]) (cf. Annexe 2) soit une mesure du changement de polarisation de l'onde (scattérométrie ellipsométrique [HAZ03, GER07]). L'ellipsométrie de Mueller [GAR04] est un cas particulier d'ellipsométrie où on utilise un autre formalisme (le formalisme de Stockes[SCH69]) que l'ellipsométrie classique (formalisme de Jones [FOW12]) pour représenter l'onde polarisée.

Différentes conditions expérimentales existent: la mesure peut être faite en fonction de la longueur d'onde  $\lambda$  (scattérométrie spectroscopique [HAZ03, HOL02]) ou à angle  $\theta$  variable (scattérométrie goniométrique [KAL98, KAL99, ROB02\*, ROB04\*]). Dans ce dernier cas, en général, un seul ordre diffracté, l'ordre 0, est mesuré à une polarisation fixe en fonction de l'angle  $\theta_i$  (2- $\theta$  scatterometry) [BIS97, BIS91]. On peut inclure dans cette même catégorie la photo-goniométrie par transformée de Fourier [PET05]. D'autres techniques, comme la  $\phi$ -ellipsométrie [BEN01] permettent de mesurer la signature optique en fonction de l'angle azimutale.

Nous allons par la suite introduire la scattérométrie ellipsométrique, utilisée dans le cadre de nos travaux.

### I.2.2.1 Scattérométrie ellipsométrique

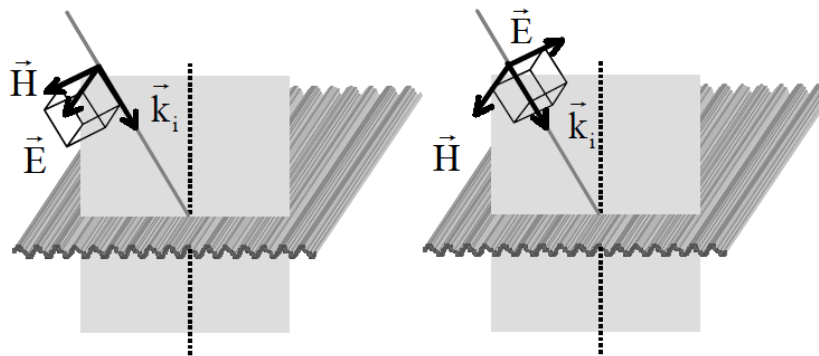
#### I.2.2.1.1 Principe de l'ellipsométrie classique

L'ellipsométrie [AZZ87, TOM05] se base sur le changement de l'état de polarisation de l'onde suite à l'interaction avec la surface d'un matériau (structuré ou non). Dans la scattérométrie ellipsométrique, la signature optique est constituée de mesures ellipsométriques mono ou multi longueurs d'onde.

Soit une onde plane caractérisée par le vecteur d'onde  $\vec{k}_i$ , et les champs  $\vec{E}_i$  et  $\vec{H}_i$  associés tels que  $(\vec{k}_i, \vec{E}_i, \vec{H}_i)$  forment un trièdre direct.

Dans le cas où cette onde est linéairement polarisée, on distingue deux cas particuliers:

- l'onde est Transverse Electrique (polarisation TE ou s): le champ  $\vec{E}$  est perpendiculaire au plan d'incidence où se trouvent le champ  $\vec{H}$  et le vecteur d'onde  $\vec{k}$ . En d'autres termes, il est parallèle aux lignes de réseaux (Figure I-21-a).
- l'onde est Transverse Magnétique (polarisation TM ou p): le champ  $\vec{H}$  est perpendiculaire au plan d'incidence où se trouvent le champ  $\vec{E}$  et le vecteur d'onde  $\vec{k}$ . Cette fois-ci, le champ  $\vec{H}$  est aussi parallèle aux lignes du réseau. (Figure I-21-b).



a) Polarisation TE

b) Polarisation TM

Figure I-21. Les deux états de polarisation de la lumière a) TE et b) TM [ROB03].

Une onde électromagnétique plane de polarisation quelconque peut donc être décomposée sur la base des 2 ondes polarisées linéairement TE et TM (Figure I-22)

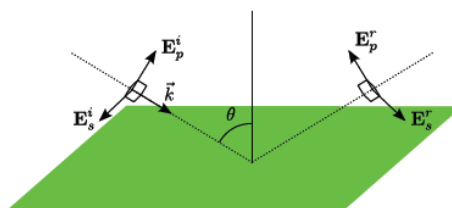


Figure I-22. Réflexion d'une onde polarisée [SOU08]

Le champ  $\vec{E}_i$  incident peut ainsi s'écrire :

$$\vec{E}_i = \vec{E}_{ip} + \vec{E}_{is} \quad (I.5)$$

$$\vec{E}_r = r_p \vec{E}_{ip} + r_s \vec{E}_{is} \quad (I.6)$$

Avec  $r_p$  et  $r_s$  les coefficients de réflexion en amplitude dans le cas d'une polarisation  $p$  et  $s$  respectivement tel que :

$$r_p = \frac{E_{rp}}{E_{ip}} = |r_p| e^{j\delta_p} \quad (I.7)$$

$$r_s = \frac{E_{rs}}{E_{is}} = |r_s| e^{j\delta_s} \quad (I.8)$$

où:

- $|r_p|$  et  $|r_s|$  sont les modules des coefficients de réflexion. Ils indiquent la modification apportée en amplitude aux composantes  $p$  et  $s$  du champ incident.
- $\delta_p$  et  $\delta_s$  sont les retards de phase induits par la réflexion de l'onde dans le cas d'une polarisation  $p$  et  $s$ .

Dans le cadre de notre étude, nous considérons que l'onde incidente est une onde polarisée linéairement. Cette onde devient une onde polarisée elliptiquement après réflexion sur la surface diffractante.

Ainsi, on peut écrire l'équation fondamentale de l'ellipsométrie telle que:

$$\rho = \frac{r_p}{r_s} = \tan\psi \exp(j\Delta) \quad (I.9)$$

dans laquelle les paramètres ellipsométriques  $\psi$  et  $\Delta$  sont tels que:

$$\psi = \arctan \left| \frac{r_p}{r_s} \right| \quad (I.10)$$

$$\Delta = \delta_p - \delta_s \quad (I.11)$$

On précise que les angles  $\psi$  et  $\Delta$  varient dans des domaines bien particuliers :

$$0 < \psi < 90 \quad \text{et} \quad 0 < \Delta < 180$$

$\psi$  et  $\Delta$  sont entièrement dépendants du matériau et caractérisent le changement de polarisation de l'onde incidente. Ils constituent une signature de la structure. Il est possible de remonter aux dimensions paramétriques du motif (CD, hauteur, hauteur des couches, etc ...) à partir de la mesure de ces signaux. Précisons dans ce cas que seul l'ordre 0 est mesuré par l'ellipsométrie.

Le schéma de principe général de l'ellipsomètre est représenté dans la Figure I-23a: une onde incidente polarisée linéairement à l'aide d'un polariseur devient, après réflexion sur le motif, polarisée elliptiquement et est ensuite recueillie par le bras de détection composé d'un analyseur, éventuellement d'un compensateur et d'un spectromètre.

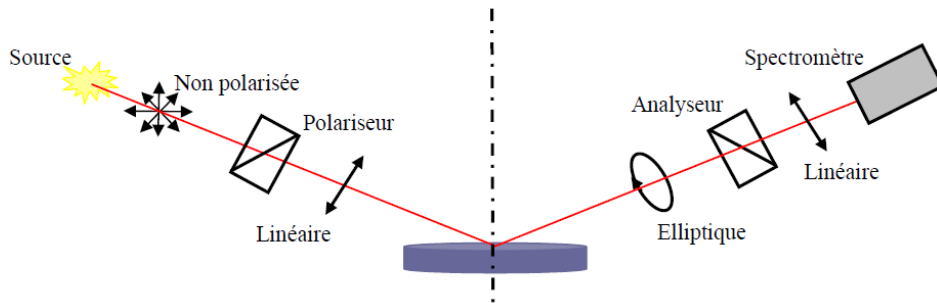


Figure I-23a. Schéma de principe de l'ellipsomètre à élément tournant [BA003]

Après avoir évoqué le principe de l'ellipsométrie, il convient maintenant de présenter les différents types d'ellipsomètres couramment utilisés.

#### I.2.2.1.2 Ellipsomètre à élément tournant

L'ellipsomètre à élément tournant [BER03] se caractérise par le fait que le faisceau peut être modulé en polarisation par la rotation du polariseur, de l'analyseur ou d'un autre élément appelé généralement compensateur.

Dans un premier temps, il est nécessaire de définir des repères afin de pouvoir introduire les notions de rotation. Soit  $Oxyz$  un repère fixe avec  $Oz$  la direction de propagation du faisceau lumineux.

Prenons l'exemple d'un ellipsomètre à polariseur tournant sans compensateur. Le polariseur tourne de manière uniforme au cours du temps, tandis que l'analyseur reste fixe. Cette rotation, de la forme  $\cos(\omega t)$  avec  $\omega$  la vitesse angulaire de rotation constante du polariseur, module la polarisation du faisceau lumineux. On peut décrire l'état de polarisation de la lumière incidente à l'aide du formalisme de Jones. On décompose l'amplitude du champ électromagnétique suivant les axes  $Ox$  et  $Oy$ . On représente l'action de chaque élément (Polariseur, Analyseur) sur la polarisation de l'onde par une matrice de transfert  $2 \times 2$ .

L'action de l'échantillon est définie par :

$$E = \begin{pmatrix} r_p & 0 \\ 0 & r_s \end{pmatrix} \quad (I.12)$$

Le polariseur horizontal (incliné suivant Ox) sera représenté par la matrice :

$$P = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \quad (I.13)$$

On peut définir aussi la matrice rotation par :

$$R(\theta) = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix} \quad (I.14)$$

On désigne par  $\theta_A$  (et  $\theta_P$ ) l'angle que fait l'axe de transmission de l'analyseur (respectivement le polariseur) avec l'axe du repère fixe Ox.

On peut ainsi définir les composantes du champ électrique reçu par le spectromètre dans le repère des axes propres de l'analyseur à l'aide du produit matriciel suivant:

$$\begin{pmatrix} E_p \\ E_s \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \cos \theta_A & -\sin \theta_A \\ \sin \theta_A & \cos \theta_A \end{pmatrix} \begin{pmatrix} r_p & 0 \\ 0 & r_s \end{pmatrix} \begin{pmatrix} \cos \theta_P & -\sin \theta_P \\ \sin \theta_P & \cos \theta_P \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} E_x \\ E_y \end{pmatrix} \quad (I.15)$$

Avec  $\theta_P = \omega t$  et  $\theta_A$  fixé

On peut ainsi déterminer, à partir de l'équation I.15 et en utilisant la définition I.9, les composantes du champ et par extension l'intensité I de la lumière réfléchiée en fonction de  $\psi$  et  $\Delta$ .

$$I = |E_p|^2 + |E_s|^2 = I_0(1 + \alpha \cos 2\theta_P + \beta \sin 2\theta_P) \quad (I.16)$$

avec

$$\alpha = \frac{\tan^2 \psi - \tan^2 \theta_A}{\tan^2 \psi + \tan^2 \theta_A}, \quad (I.17)$$

$$\beta = \frac{2 \cos \Delta \tan \psi \tan \theta_A}{\tan^2 \psi + \tan^2 \theta_A} \quad (I.18)$$

et

$$I_0 = \frac{1}{2} |E_0 r_s|^2 \cos^2 \theta_A (\tan^2 \psi + \tan^2 \theta_A) \quad (I.19)$$

Les paramètres  $\psi$  et  $\Delta$  sont donc déterminés à partir de la mesure des coefficients  $\alpha$  et  $\beta$  en résolvant les équations I.17 et I.18:

$$\tan(\psi) = \sqrt{\frac{1+\alpha}{1-\alpha}} \cdot \tan(\theta_A) \quad \text{et} \quad \cos(\Delta) = \frac{\beta}{\sqrt{1-\alpha^2}} \quad (\text{I.20})$$

A noter que  $\alpha$  et  $\beta$ , donc  $\psi$  et  $\Delta$  ne dépendent pas de l'intensité du faisceau incident.

Ainsi, dans le cas de l'ellipsométrie par polariseur tournant, les signatures sont composées des paramètres  $\alpha$  et  $\beta$  mesurés dans différentes conditions.

### I.2.2.1.3 L'ellipsomètre à modulation de phase

Dans le cas de l'ellipsomètre à modulation de phase, on se place dans les mêmes conditions que dans le montage précédent, avec l'ajout d'un modulateur (photoélectrique par exemple) immédiatement après le polariseur. Cette fois-ci, le polariseur et l'analyseur sont fixes. Ajoutons que l'angle  $\theta_P - \theta_M$  est généralement fixé par le fabricant.

On peut décrire la matrice de transfert  $M$  du modulateur par :

$$M = \begin{pmatrix} \exp j\delta(t) & 0 \\ 0 & 1 \end{pmatrix} \quad (\text{I.21})$$

$$\text{avec } \delta(t) = a \sin(\omega t) \quad (\text{I.22})$$

où  $a$  est l'amplitude de modulation et  $\omega$  la pulsation angulaire du modulateur. La fréquence de modulation est typiquement de l'ordre de 50Khz.

On pourra ainsi exprimer les composantes du champ électrique sur le détecteur par un calcul similaire :

$$\begin{pmatrix} E_p \\ E_s \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \cos(\theta_A) & -\sin(\theta_A) \\ \sin(\theta_A) & \cos(\theta_A) \end{pmatrix} \begin{pmatrix} r_p & 0 \\ 0 & r_s \end{pmatrix} \begin{pmatrix} \cos(\theta_M) & -\sin(\theta_M) \\ \sin(\theta_M) & \cos(\theta_M) \end{pmatrix} \begin{pmatrix} \exp j\delta(t) & 0 \\ 0 & 1 \end{pmatrix}$$

$$\begin{pmatrix} \cos(\theta_P - \theta_M) & -\sin(\theta_P - \theta_M) \\ \sin(\theta_P - \theta_M) & \cos(\theta_P - \theta_M) \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} E_x \\ E_y \end{pmatrix} \quad (\text{I.23})$$

On obtient donc en intensité :

$$I = |E_p|^2 + |E_s|^2 = I_0 + I_s \sin(\delta t) + I_c \cos(\delta t) \quad (\text{I.24})$$

avec



$$I_0 = B(1 - \cos 2\psi \cos 2\theta_A + \cos 2(\theta_P - \theta_M) \cos 2M (\cos 2\theta_A - \cos 2\psi) + \sin 2\theta_A \cos \Delta \cos 2(\theta_P - \theta_M) \sin 2\psi \sin 2M) \quad (I. 25)$$

$$I_s = B[\sin 2(\theta_P - \theta_M) \sin 2\theta_A \sin 2\psi \sin \Delta] \quad (I. 26)$$

$$I_c = B \sin 2(\theta_P - \theta_M) [(\cos 2\psi - \cos 2\theta_A) \sin 2\theta_M + \sin 2\theta_A \cos 2\theta_M \sin 2\psi \cos \Delta] \quad (I. 27)$$

dans laquelle :

$$B = \frac{E_0^2}{4|r_p^2 + r_s^2|} \quad (I. 28)$$

La connaissance des intensités  $I_s$  et  $I_c$  permet de calculer avec précision  $\psi$  et  $\Delta$ . En fonction de l'orientation des angles  $\theta_P$ ,  $\theta_M$  et  $\theta_A$ , on distingue deux configurations particulières :

Configuration I:  $(\theta_P - \theta_M) = \pi/4$ ,  $\theta_M = 0$ ,  $\theta_A = \pi/4$  pour laquelle :

$$I_s = \sin 2\psi \sin \Delta \quad (I. 29)$$

$$I_c = \sin 2\psi \cos \Delta \quad (I. 30)$$

Et la configuration II:  $(\theta_P - \theta_M) = \pi/4$ ,  $\theta_M = \pi/4$ ,  $\theta_A = \pi/4$ , pour laquelle :

$$I_s = \sin 2\psi \sin \Delta \quad (I. 31)$$

$$I_c = \cos 2\psi \quad (I. 32)$$

$I_s$  et  $I_c$  sont directement liées aux paramètres géométriques recherchés par l'intermédiaire de  $\psi$  et  $\Delta$ . Ces intensités constituent donc des signatures de ce motif. Cela signifie qu'il est possible de retrouver les informations sur la géométrie du motif périodique à partir de leur mesure.

On peut noter que dans les cas où la détermination de  $\psi$  et  $\Delta$  est nécessaire, alors l'utilisation stricte de l'une ou l'autre de ces configurations n'est pas suffisante et il réside des ambiguïtés sur la cette détermination . En effet dans le cas de la configuration I, la valeur de  $\psi$  est indéterminée entre  $\psi$  et  $90-\psi$  par la seule connaissance de  $\sin 2\psi$ . Il réside donc une incertitude sur  $\psi$  alors que  $\Delta$  est parfaitement déterminé. A l'inverse, dans le cadre de la configuration II,  $\Delta$  est indéterminé entre deux valeurs,  $\Delta$  et  $180-\Delta$ . On a donc une incertitude

sur  $\Delta$  alors que  $\psi$  est cette fois-ci parfaitement déterminé. Pour éviter d'avoir une indétermination, on mesure généralement  $\Delta$  et  $\psi$  dans les deux configurations.

Dans le cadre de notre travail, on ne s'intéresse qu'à la mesure des paramètres  $I_s$  et  $I_c$  et il n'est pas nécessaire de connaître les paramètres  $\psi$  et  $\Delta$  ; ce qui permet de s'affranchir de ce genre de considération.

Il existe également un troisième type de montage (non détaillé ici) : l'ellipsomètre à extinction [AZZ08, TOM05, BER03]. Cette technique est basée sur la recherche de l'extinction du signal par une orientation particulière des angles du polariseur, de l'analyseur et du compensateur. La mesure de ces angles permet de déterminer les paramètres  $\psi$  et  $\Delta$  de l'échantillon. L'inconvénient majeur de cette technique est sa lenteur par rapport aux autres techniques d'ellipsométrie; ce qui la rend impropre à l'étude menée dans ce manuscrit.

Les systèmes ellipsométriques présentés ci-dessus possèdent l'avantage d'être non destructifs et transposables in-situ. De plus, ils permettent un temps d'acquisition court de la signature ellipsométrique (cf. Chap III).

Dans le cadre de ce travail de thèse, nous utiliserons la gamme d'ellipsomètre de l'équipementier Horiba Jobin Yvon à modulation de phase (Figure I-23b). Nous en avons utilisé deux: un ellipsomètre in situ positionné à l'intérieur de la chambre de gravure Centura300 (de l'équipementier Applied Materials) et un ellipsomètre disponible en salle blanche lié à aucun autre équipement. Les deux ellipsomètres peuvent effectuer des mesures en mode spectroscopique et en mode dynamique.



*Figure I-23b. Ellipsomètre à modulation de phase Horiba Jobin Yvon en standalone*

Le mode spectroscopique consiste à balayer successivement toutes les longueurs d'onde du spectre en imposant un incrément fixe entre deux acquisitions. Il faut noter que l'échantillonnage peut être très important dans ce mode, mais l'acquisition prend beaucoup plus de temps (2 minutes environ)

Le mode dynamique, ou cinétique, consiste à effectuer plusieurs mesures simultanées d'un nombre limité de longueurs d'onde en fonction du nombre de Photorécepteurs présents en sortie du système. Dans ce cas le spectre est faiblement résolu mais obtenu rapidement. Il résulte donc que le mode cinétique est le mieux adapté pour une application en temps réel. L'ellipsomètre disponible dans la chambre de gravure possède 32 PhotoMultiplieurs (PM).

Dans tous les cas, la durée de la mesure dépend du temps d'intégration. Elle est généralement comprise entre 100 et 500ms. Il est important de comprendre que plus le temps d'intégration est faible et plus la mesure sera bruitée. Dans nos travaux, et compte tenu de notre application visée, nous allons prendre un temps d'acquisition de 500ms. Ceci nous semble un bon compromis entre un bon suivi de procédé, un bon échantillonnage et un bruit de mesure faible.

Après avoir présenté les différents types de signatures optiques, intéressons-nous maintenant à la modélisation du problème inverse et aux techniques les techniques utilisées pour le résoudre.

### I.3 Méthodes de résolution du problème inverse

La scattérométrie est une technique de métrologie indirecte dont le but est de retrouver la forme d'une structure à partir de l'analyse d'une signature ellipsométrique mesurée expérimentalement. Pour cela, ce relevé expérimental est confronté, par l'intermédiaire d'une technique ad hoc, aux signatures simulées obtenues théoriquement à partir d'un modèle paramétrique prédéfini (cf. paragraphe II.2.1.3). Cette étape se nomme la résolution du problème inverse. L'objectif est donc d'obtenir, à partir de la signature mesurée, les valeurs des paramètres du modèle géométrique supposé.

Mathématiquement, on exprime le vecteur  $\vec{x}$  des paramètres géométriques caractérisant la forme du profil à partir de l'observation  $\vec{y}$ , représentant la signature scattérométrique correspondante, sous les conditions  $\vec{a}$  définies au I.2.1.3 par  $F^{-1}$

$$F^{-1} : \vec{y} \rightarrow \vec{x} = F^{-1}(\vec{y}, \vec{a}) \quad (\text{I.33})$$

Il existe dans la littérature trois voies pour résoudre le problème inverse : une approche par optimisation, une par l'utilisation de bibliothèques et une dernière par régression.

Chacune de ces différentes techniques est passée en revue par la suite.

### I.3.1 Méthodes d'optimisation

Avant de développer les techniques d'optimisation d'un point de vue mathématique, il convient de présenter le problème d'optimisation dans le cadre particulier de la scattérométrie. Comme le montre la Figure I-24, on compare à chaque itération une réponse optique obtenue par simulation (méthode MMFE) avec la réponse optique expérimentale. Pour cela, le problème direct nous permet de calculer la signature optique correspondant à chaque nouveau jeu de paramètres  $\vec{x}$ .

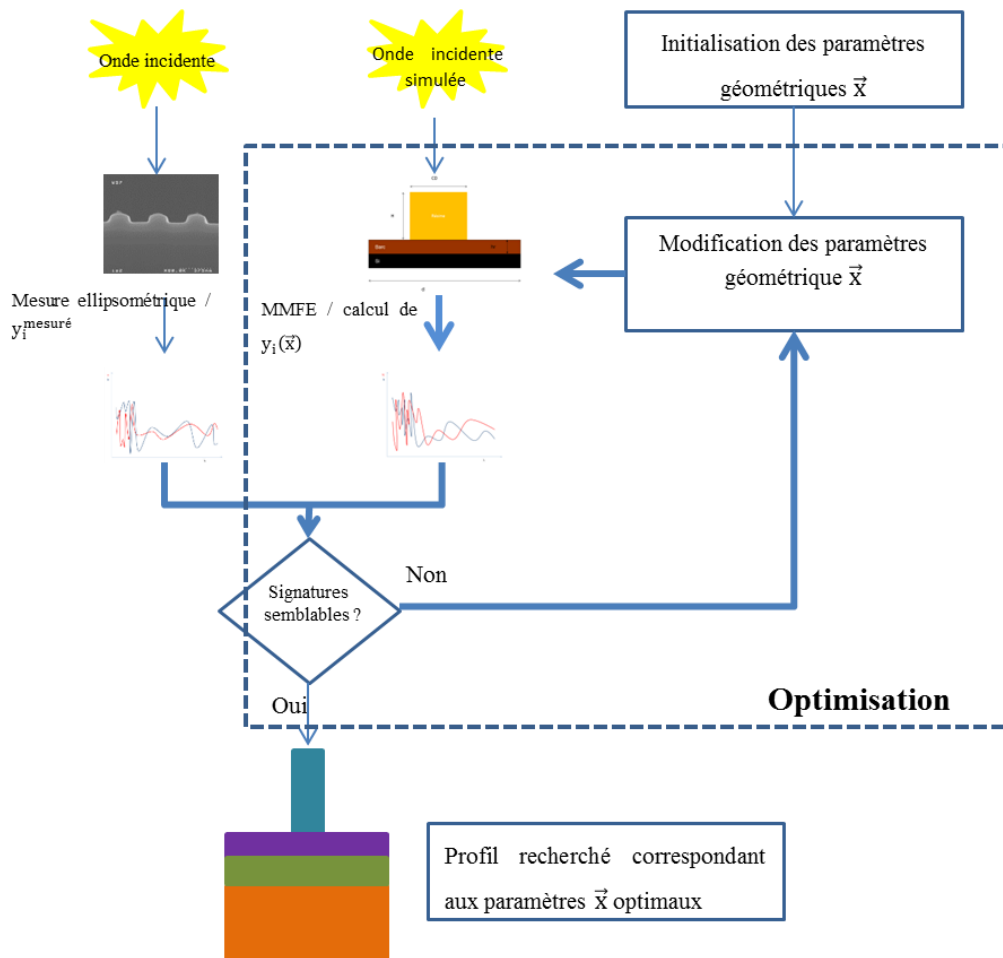


Figure I-24. Schéma de principe de l'optimisation paramétrique dans le cadre de la scattérométrie ellipsométrique

Le critère de comparaison entre les 2 signatures mesurée  $y_i^{\text{mesuré}}$  et simulée  $y_i(\vec{x})$  correspond généralement à l'erreur quadratique moyenne formulée de la façon suivante :

$$Z(\vec{x}) = \sum_{i=1}^b \frac{(y_i^{\text{mesuré}} - y_i(\vec{x}))^2}{\sigma_i^2} \quad (\text{I. 34})$$

où  $b$  représente le nombre de mesures total de la signature ,  $y_i$  la valeur de l'intensité (Is puis Ic) à la longueur d'onde  $\lambda_i$ , composant la signature ,  $\sigma_i^2$  une estimation de la variance de la  $i^{\text{ème}}$  mesure et  $\vec{x}$  l'ensemble des paramètres géométriques variables de la structure tel que  $\vec{x}=(x_1, x_2, \dots, x_N)$ .

A noter que le critère à minimiser défini en I.34 est souvent simplifié en omettant l'influence du bruit de mesure :

$$Z(\vec{x}) = \sum_{i=1}^n (y_i^{\text{mesuré}} - y_i(x_1, \dots, x_N))^2 \quad (\text{I. 35})$$

Au début de l'algorithme, les valeurs initiales des paramètres géométriques du modèle sont choisies en fonction des connaissances préalables de la structure (valeurs cibles du cahier des charges par exemple). Au fur et à mesure de la progression de l'algorithme progresse, ces valeurs sont modifiées suivant des règles précises impliquant les informations (locales ou non) récoltées au cours de l'optimisation conduisant à la diminution de la fonction coût. A la fin de l'algorithme, lorsque les deux signatures coïncident, on relève les paramètres géométriques correspondant à la structure optimale dont la signature optique est la plus proche (dans les hypothèses choisies) de celle mesurée. L'algorithme est stoppé lorsque le critère à minimiser (la fonction  $Z(\vec{x})$ ) devient inférieur à une valeur prédéfinie choisi au préalable par l'utilisateur. (par exemple à une valeur  $\epsilon$  ou l'on considère que l'on n'est plus sensible à variation de signature optique)

Intéressons-nous maintenant à la définition de l'optimisation classique d'un point de vue mathématique. Afin de simplifier les notations, nous considérons dans la suite le problème d'optimisation unidimensionnel qui consiste à trouver le minimum de la fonction  $Z$  qui dépend d'une seule variable notée  $x$ . Soit  $M \subset \mathbb{R}$  l'ensemble des solutions admissibles. Trouver le minimum d'une fonction revient à chercher  $x^* \in M$  tel que :

$$Z(x^*) = \min_{x \in M} Z(x) \quad (\text{I. 36})$$

Un point  $x^*$  sera considéré comme un minimum local (Figure I-25) si

$$\forall x \in M \text{ et } d(x, x^*) < \varepsilon \text{ alors } Z(x^*) < Z(x) \quad (\text{I. 37})$$

Où  $d(x, x^*)$  désigne la distance entre le point  $x$  et  $x^*$ . L'équation I.36 signifie que dans le voisinage de  $x^*$  (défini par la valeur de  $\varepsilon$ ), il n'existe aucun point pour lequel la fonction coût  $Z(x)$  est strictement inférieure à  $Z(x^*)$ .

De plus, on considèrera que  $x^*$  est un minimum global (Figure I-25) si :

$$\forall x \in M \text{ alors } Z(x^*) < Z(x) \quad (\text{I. 38})$$

L'objectif de toute méthode d'optimisation est bien entendu de trouver le minimum global de la fonction  $Z$  et de ne pas converger prématurément vers un minimum local qui conduirait à une solution non optimale.

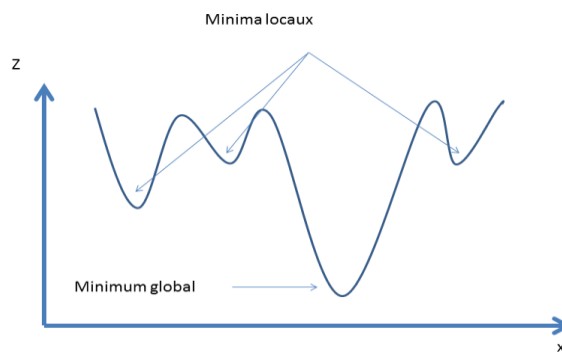


Figure I-25. Représentation schématique des minima d'une fonction  $Z$  dépendant d'un seul paramètre  $x$

Mathématiquement il existe deux conditions que doivent vérifier l'optimum, locale ou globale, d'une fonction [BON00]. La première condition d'optimalité est l'annulation du gradient, soit ici rechercher  $x$  tel que  $\frac{\partial Z}{\partial x} = 0$ . La seconde condition impose que la matrice Hessienne  $H$  (matrice des dérivées secondes) évaluée en  $x^*$  soit définie positive ( $\frac{\partial^2 Z}{\partial x^2} > 0$ ).

La résolution analytique directe est souvent délicate, voire impossible. Ainsi, il existe différentes méthodes itératives qui permettent de se rapprocher avec une précision arbitraire de l'optimum recherché. Par exemple, l'algorithme de descente de gradient se base principalement sur la première condition alors que l'algorithme de Gauss Newton exploite les

deux conditions. Le troisième algorithme qu'on va présenter est l'algorithme de Levenberg Marquardt. Celui-ci se sert des deux algorithmes évoqués précédemment.

### ***1.3.1.1 L'algorithme de descente de gradient [CUR44]***

Dans cette méthode, à chaque itération  $k$ , on calcule la fonction coût  $Z(x^k)$  correspondante à un jeu de paramètres  $x^k$ . L'algorithme d'optimisation consiste donc à trouver la valeur  $x^{k+1}$  pour l'itération  $k+1$  tel que :

$$Z(x^{k+1}) < Z(x^k) \quad (I.39)$$

Le pas  $s^k = \alpha^k d^k$  séparant  $x^{k+1}$  de  $x^k$  est défini par :

$$s^k = x^{k+1} - x^k \quad (I.40)$$

avec  $d^k$  la direction de descente donnée par  $d_k = -\nabla Z(x^k)$  et  $\alpha$  un scalaire défini de manière à respecter la relation I.39

Dans la méthode de la descente du gradient l'équation I.40 devient alors :

$$x^{k+1} = x^k - \alpha^k \nabla Z(x^k) \quad (I.41)$$

La direction de recherche de la solution optimale est donc donnée par l'opposé du gradient. La recherche du scalaire  $\alpha$  à chaque génération peut être déterminée suivant une optimisation unidimensionnelle mais la méthode la plus connue consiste à trouver une valeur qui satisfasse les conditions de Wolfe [NOC99].

La méthode de descente de gradient est illustrée sur la Figure I-26. On représente par le point rouge (position 1), de coordonnées  $(x_1, Z_1)$  le point de départ de l'optimisation. A chaque itération, on calcule le pas donné par l'équation I.42. On obtient ainsi un nouveau point  $(x_2, Z_2)$  qui présente une fonction coût inférieure au point en position 1. On déroule l'algorithme jusqu'à atteindre un minimum local, soit le point en position 6  $(x_6, Z_6)$ . On peut noter que si la valeur du pas est importante, on risque de s'écarter trop rapidement de la solution optimale. Cependant, si le pas est très petit et donc le déplacement relativement faible, la convergence vers la solution optimale nécessitera un temps de calcul important. L'inconvénient majeur de cette technique est sa lenteur de convergence, et le fait qu'elle converge vers le premier minimum trouvé. Ce minimum peut être local ou global. Pour

s'affranchir de ce problème, la solution consiste à relancer l'algorithme plusieurs fois en partant de conditions initiales différentes.

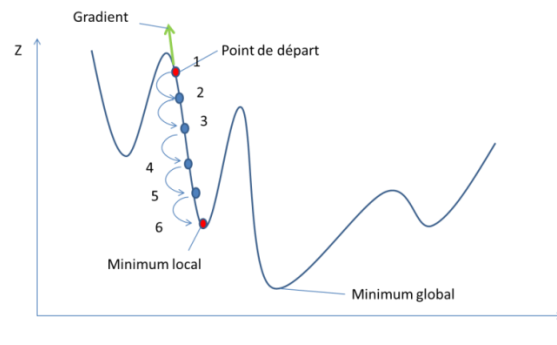


Figure I-26: Illustration de la méthode de descente de gradient.

### I.3.1.2 L'algorithme de Gauss-Newton.

Dans les fonctions présentées précédemment, la recherche de la direction  $d^k$  a été obtenue à l'aide du calcul du gradient de la fonction  $Z(x)$ . La méthode de Gauss-Newton fait appel au calcul de la matrice Hessienne  $H$  de la fonction  $Z$ . On parle de méthode du second ordre. On ne travaille plus directement sur la fonction  $Z$  mais sur un modèle quadratique défini à partir de la valeur de  $Z(x^k)$ , donnée par la formule de Taylor :

$$Z(x^{k+1}) \approx Z(x^k) + \nabla Z(x^k)^T s^k + \frac{1}{2} s^{kT} H \cdot s^k \quad (I.42)$$

Cette approximation est d'autant plus vraie que la fonction  $Z$  est de forme quadratique autour du point  $x^k$ . (Figure I-27). On cherche alors le pas qui nous permettra d'atteindre le minimum de ce modèle soit  $\nabla Z(x^{k+1}) = 0$ . La solution de ce problème est donnée pour :

$$H \cdot s^k = -\nabla Z(x^k) \quad (I.43)$$

soit :

$$s^k = -H^{-1} \nabla Z(x^k) \quad (I.44)$$

On peut représenter graphiquement de manière schématique l'évolution de l'algorithme d'optimisation sur un exemple simple (Figure I-27). Soit la fonction coût représentée par la courbe  $Z(x)$ . L'algorithme est initialisé avec un point choisi en position 1 ( $x_1, Z_1$ ). On



construit ensuite le modèle quadratique en ce point (courbe  $C_1$ ) à partir de la connaissance de la tangente au point  $(x_1, Z_1)$ . L'algorithme de Gauss Newton permet de calculer le pas (équation I.47) afin d'atteindre le minimum de  $C_1$ . Ceci nous donne le nouveau point en position 2  $(x_2, Z_2)$ . On effectue la même démarche en ce point, et on obtient la solution suivante en position 3  $(x_3, Z_3)$ , et ainsi de suite.

L'inconvénient de cette méthode réside dans le fait que la matrice  $H$ , indispensable pour calculer le pas à chaque itération, est très longue, voire délicate à calculer analytiquement (condition d'existence des dérivées secondes). Il existe par conséquent d'autres méthodes, dites de quasi newton, qui en donnent une approximation à chaque itération en fonction des itérations précédentes. La plus connue de ces méthodes est la méthode BFGS ou Broyden Fletcher Goldfarb Shanno [BRO65, FLE63, GOL68, SHA70] :

$$H^{-1k+1} = H^{-1k} - \frac{H^{-1k} s^k (H^{-1k} s^k)^T}{s^{kT} H^{-1k} s^k} + \frac{q^k q^{kT}}{q^{kT} s^k} + \phi_k \left[ s^{kT} H^{-1T} s^k \right] v^k v^{kT} \quad (I.45)$$

où

$$\phi_k \in [0,1] \text{ et } v^k = \left[ \frac{q^k}{q^{kT} s^k} - \frac{H^{-1k} s^k}{s^{kT} H^{-1k} s^k} \right]$$

$$\text{Et } s^k = x^{k+1} - x^k = \alpha^k d^k$$

$$\text{Et } q^k = \nabla Z(x^{k+1}) - Z(x^k)$$

L'avantage de cette variante réside dans le fait qu'il n'est pas nécessaire de connaître avec précision la matrice  $H$  à chaque itération. Une approximation même grossière assure une réduction à chaque pas de la fonction coût et permet de ce fait une convergence plus rapide en terme de temps mais pas en terme d'itérations.

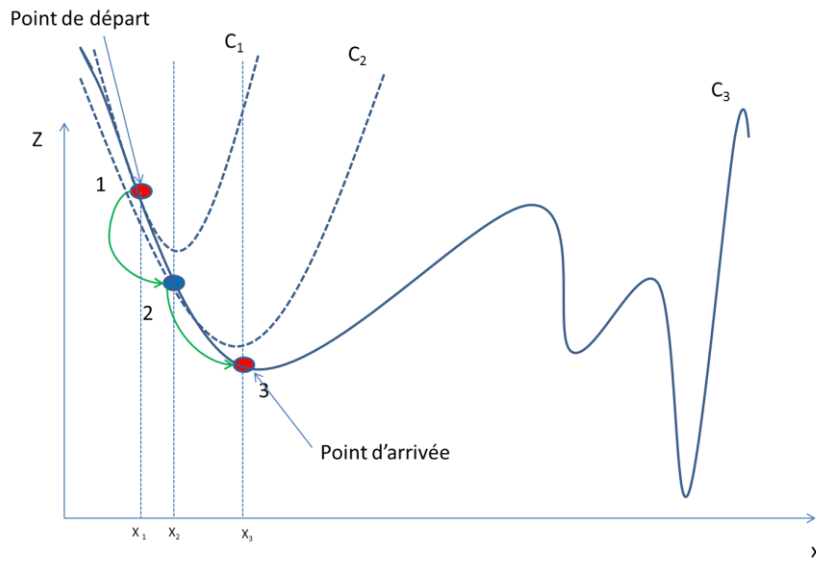


Figure I-27. Schéma de fonctionnement de la méthode de Gauss Newton

La méthode de Levenberg Marquardt (LM) [LEV44, MAR63, MOR77] combine les deux méthodes d'optimisation présentées précédemment.

Avant de détailler le formalisme mathématique de cette méthode, on peut simplifier les notations en posant :

$$\beta_j = -\frac{1}{2} \frac{\partial Z}{\partial x_j} \quad \text{et} \quad C_{jl} = \frac{1}{2} \frac{\partial^2 Z}{\partial x_j \partial x_l} \quad (\text{I. 46})$$

soit vectoriellement :

$$\vec{\beta} = -\frac{1}{2} \vec{\nabla} Z \quad \text{et} \quad [C] = \frac{1}{2} H \quad (\text{I. 47})$$

où C est communément appelée la matrice de courbure.

Reprenons la formule I.34 de Z. En dérivant cette expression par rapport à  $x_j$  ( $j=1, \dots, N$ ) on a les composantes du gradient  $\nabla Z$  :

$$\frac{\partial Z}{\partial x_j} = -2 \sum_{i=1}^n \frac{[y_i^{\text{mesuré}} - y_i(\vec{x})]}{\sigma_i^2} \frac{\partial y_i(\vec{x})}{\partial x_j} \quad (\text{I. 48})$$

On obtient ainsi les différents éléments de la matrice Hessienne en effectuant une dérivation supplémentaire par rapport à  $x_l$ :

$$\frac{\partial^2 Z}{\partial x_j \partial x_l} = 2 \sum_{i=1}^n \frac{1}{\sigma_i^2} \left[ \frac{\partial y_i(\vec{x})}{\partial x_j} \frac{\partial y_i(\vec{x})}{\partial x_l} - [y_i^{mesuré} - y_i(\vec{x})] \frac{\partial^2 y_i(\vec{x})}{\partial x_j \partial x_l} \right] \quad (I.49)$$

En négligeant les termes de second ordre, l'équation précédente devient :

$$\frac{\partial^2 Z}{\partial x_j \partial x_l} = 2 \sum_{i=1}^n \frac{1}{\sigma_i^2} \left[ \frac{\partial y_i(\vec{x})}{\partial x_j} \frac{\partial y_i(\vec{x})}{\partial x_l} \right] \quad (I.50)$$

A la  $k^{\text{ème}}$  itération deux cas de figures sont à considérer :

- Si la fonction à optimiser est supposée de forme quadratique alors le pas d'incrément  $\vec{s}$  (I.40) est donné par la formule de Gauss Newton suivante:

$$\vec{s} = -H^{-1} \nabla Z(\vec{x}^k) \quad (I.51)$$

- Si la fonction à optimiser est « éloignée » d'une forme quadratique alors la méthode des plus fortes pentes sera utilisée par le calcul de  $\vec{s}$  par:

$$\vec{s} = -\alpha \vec{\nabla} Z(\vec{x}^k) \quad (I.52)$$

Où  $\alpha$  est une constante choisie positive.

Le principe de la méthode de Levenberg-Marquardt repose sur un choix différent du pas d'incrément suivant que la fonction se rapproche ou non d'une forme quadratique

On peut résumer la démarche par le formalisme suivant en posant :

$$[C]' = [C] + \lambda I_d \quad (I.53)$$

où  $I_d$  est la matrice identité de dimension  $N \times N$

Ainsi, on peut remplacer les équations I.51 et I.52 par l'équation suivante :

$$[C]' \vec{s} = \vec{\beta} \quad (I.54)$$

En effet, lorsque  $\lambda$  est très grand,  $[C]'$  est à diagonale dominante ( soit le module de chaque terme diagonal est supérieur ou égal à la somme des modules des autres termes de sa ligne) et l'équation I.54 devient :

$$s_j = \frac{1}{\lambda C_{jj}} \vec{\beta} = - \frac{1}{2 \lambda C_{jj}} \vec{v} Z \quad (I.55)$$

L'équation I.55 devient similaire à l'équation I.52 des plus fortes pentes. A l'inverse, lorsque  $\lambda$  tend vers 0, l'équation I.56 équivaut à l'équation( I.51) de Gauss Newton.

L'inconvénient de cet algorithme, comme toutes les autres techniques d'optimisation classique, est le risque de convergence prématurée. Ceci signifie que l'algorithme va converger vers le premier minimum trouvé satisfaisant les conditions d'optimalité, qu'il soit local ou global. Ceci est d'autant plus vrai dans le cas précis de la scattérométrie où les algorithmes se trouvent confrontés à une multitude de minima locaux. En effet, ceux-ci sont induits par le caractère multidimensionnel du problème ainsi que la présence de bruit de mesure. Il est donc primordial de lancer plusieurs fois l'algorithme pour contourner cette difficulté. Il est aussi important de bien choisir le point de départ de l'algorithme pour faciliter l'obtention de la solution optimale. L'autre inconvénient majeur est le temps de calcul nécessaire à l'obtention d'un résultat. En effet, à chaque itération, il est nécessaire de faire appel au problème direct, ce qui présente une contrainte pour une application en temps réel, où la résolution doit être quasi instantanée.

Les techniques d'optimisation classiques ont été largement utilisées afin de résoudre le problème inverse en scattérométrie. Parmi ceux-ci, on peut citer principalement l'algorithme de Levenberg Marquardt [HOL02, HAZ03, HUA04]

Il existe ainsi d'autres algorithmes appartenant à la famille des optimisations stochastiques. La recherche de la solution optimale est donc régie par des règles probabilistes. Parmi ceux-ci, on peut citer la méthode des algorithmes génétiques [HOL75, GOL89] ou le recuit simulé [KIR83]. Ces techniques ont été utilisées pour la résolution du problème inverse en ellipsométrie [COR00] et en scattérométrie [ROB03]. Comme les techniques d'optimisation classique, la plupart de ces méthodes sont itératives et nécessitent plusieurs fois le recours à la fonction coût à chaque itération.

Bien que ces algorithmes puissent résoudre efficacement le problème inverse en scattérométrie, ils présentent l'inconvénient majeur du temps de calcul important, critère crucial dans notre problématique d'application en temps réel. Il est donc indispensable de faire appel à d'autres techniques de résolution, plus adaptées en termes de temps de calcul que l'ensemble des méthodes d'optimisations évoquées ci-dessus. Dans ce cadre-là, nous allons nous intéresser à deux techniques: la méthode des bibliothèques et la méthode des réseaux de neurones. La première citée a déjà fait l'objet de travaux au sein du LTM et a montré sa capacité à répondre à la problématique de temps réel.

### **I.3.2 Méthode des bibliothèques**

#### ***I.3.2.1 Principe***

Niu et al. ont été les précurseurs de la méthode des bibliothèques [NIU99] dans le domaine de la scattérométrie. On retrouve, dans la littérature, différentes appellations: library technique [HAZ03], fitting error interpolation based library search [CHE13], Model Based Library Approach [YAS13], look up table [SOH92]. Dans la suite du manuscrit, nous garderons l'appellation « Méthode des bibliothèques ».

Le principe de la méthode des bibliothèques pour la scattérométrie est représenté dans la Figure I-28. Préalablement à la mesure ellipsométrique réalisée sur un échantillon, on construit une bibliothèque de signatures optiques obtenues de manière théorique à l'aide de la MMFE, dans une plage de variation des paramètres géométriques définie et bornée. L'objectif de cette méthode consiste à retrouver, au sein de la « librairie de signatures », celle qui sera la plus proche de celle obtenue par la mesure.

La recherche dans la bibliothèque est effectuée de manière énumérative. En effet, la signature expérimentale est comparée à chacune des signatures de la bibliothèque correspondant chacune à un motif bien déterminé. Le minimum de la fonction coût définie par le critère I.34 est donc recherché pour déterminer les paramètres géométriques optimaux. L'utilisation de librairie constitue donc un cas particulier de l'optimisation.

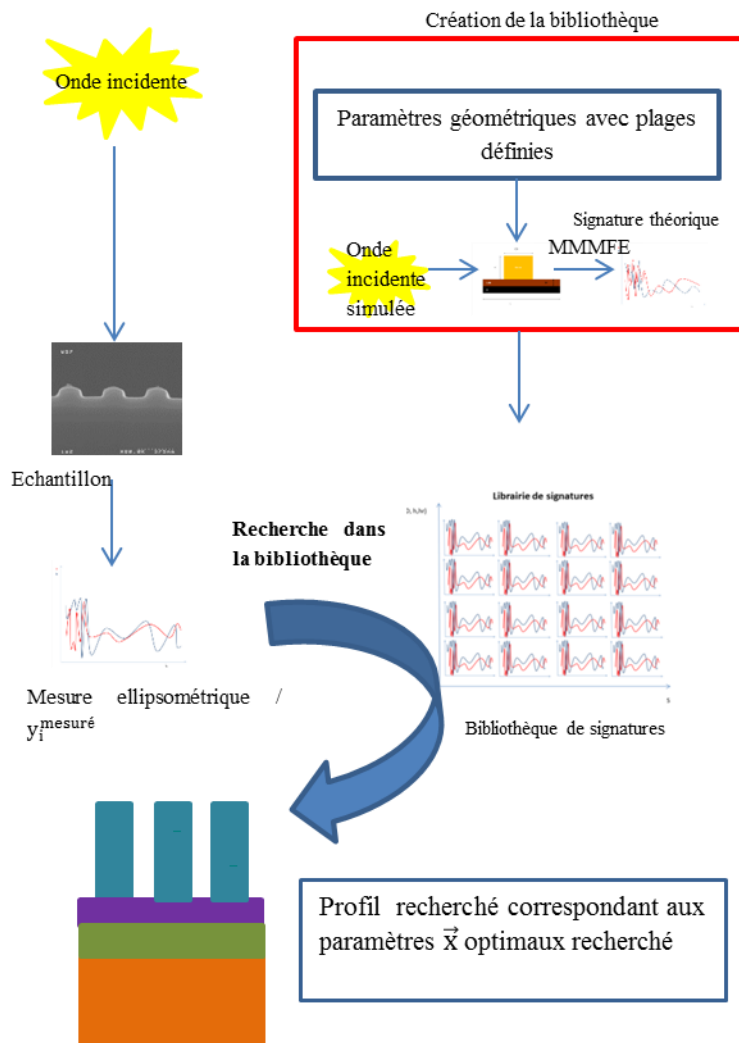


Figure I-28. . Schéma de principe de la résolution du problème inverse avec la méthode des bibliothèques

La différence fondamentale avec une méthode d'optimisation réside dans le choix du pas d'itération. Dans le cas de l'optimisation classique, ce choix est régi par des règles déterministes précises alors que dans le cas des bibliothèques, il est prédéfini lors de la construction de la base de données. Une autre différence réside dans le fait que les méthodes d'optimisation nécessitent le calcul du problème direct (MMFE) à chaque itération alors que pour la bibliothèque ce calcul est effectué pour chaque élément de la bibliothèque.

### ***1.3.2.2 Mesure spectroscopique et cinétique***

La méthode des bibliothèques peut être utilisée pour les deux modes d'acquisition de l'ellipsomètre: en statique ou en dynamique (cinétique). C'est dans cette seconde configuration qu'elle sera développée dans le cadre de cette thèse. En effet, contrairement aux autres méthodes d'optimisation, la méthode des bibliothèques se distingue dans le sens où elle peut être applicable pour une caractérisation en temps réel. Ceci est dû au fait que cette méthode comporte deux phases distinctes: d'une part, la construction de la bibliothèque et d'autre part, la recherche dans la bibliothèque de la signature la plus proche. La construction de la bibliothèque est une étape qui peut être couteuse en temps de calcul mais qui peut se faire préalablement sans contrainte restrictive de temps. L'étape de recherche dans la bibliothèque peut s'effectuer de manière quasi instantanée en utilisant des outils adaptés que nous détaillerons par la suite. Ajoutons que la technique des bibliothèques est utilisé sur des lignes de production dans les milieux industriels.

Certaines améliorations ont été apportées à cette méthode afin de réduire considérablement le temps de recherche dans la bibliothèque. Soulan [SOU08] a proposé dans sa thèse une solution associant une recherche des  $k$  solutions les plus proches de la mesure effectuée au sein de la bibliothèque, appelée principe des  $k$ -ppv ( $k$  plus proches voisins) avec une utilisation des processeurs graphiques (GPU) pour scruter de façon vectorielle la bibliothèque. Dans la méthode des  $k$ -ppv, le choix de  $k$  voisin au lieu d'un seul est essentiel dans une application temps réel. Comme nous allons le voir dans le chapitre III, le spectre dynamique est peu résolu en longueur d'onde et la signature peut par conséquent être fortement bruitée. Augmenter le nombre de voisin  $k$  dans l'algorithme des  $k$ -ppv permet de pallier à cette caractéristique du temps réel et d'obtenir un résultat avec une bonne précision [SOU08].

### ***1.3.2.3 Les GPU (Graphics Processing Unit) et autres améliorations***

L'aspect le plus important à prendre en compte lorsqu'on évoque les GPU sont leur capacité à faire du calcul vectoriel par opposition au calcul scalaire. En effet, le calcul vectoriel consiste à effectuer les tâches en une seule étape au lieu de les faire en plusieurs successivement. Pour essayer d'imager ce propos, prenons un exemple simpliste : des billes sont posées au sol et un joueur souhaite les poser dans une boîte. Deux approches sont

possibles : soit il les prend une à une (approche scalaire), soit il les prend en une seule poignée (approche vectorielle).

Nous pouvons noter que le fait de détourner les GPU de leur utilisation initiale est une opération délicate nécessitant de mettre en place des codes de calculs élaborés [SOU08]. La méthode des GPU permet d'accélérer de manière considérable la recherche dans la bibliothèque. Elle est ainsi parfaitement bien adaptée pour l'étude en temps réel. La méthode des bibliothèques avec GPU dans le cadre de la scattérométrie dynamique a été validée avec succès de manière expérimentale par deux thèses réalisées au LTM [SOU08, ELK10]. Cette méthode a été appliquée pour le suivi de procédés dynamiques, tels que le fluage de résine [SOU07, SOU07\*] ou de réduction de côte de résine dans une application de gravure [ELK09, ELK09\*].

Cependant, une limitation de la méthode des bibliothèques réside dans la quantité de mémoire maximale disponible pour stocker de grandes bibliothèques (1 Go). Ceci signifie que pour une étude sur des cas pratiques avec 7 ou 8 paramètres, l'utilisation des bibliothèques avec GPU sera vite compliquée à mettre en œuvre. On peut penser que le développement des plateformes de programmation parallèles de GPU permettra d'apporter une solution à ce sujet. De plus, le couplage de la méthode des bibliothèques avec la technique du krigeage [KLE08, SAC89], permettra de construire des bibliothèques plus compactes. Cette technique est basée sur l'interpolation statistique de la fonction coût. Elle ne sera pas détaillée dans ce mémoire car elle ne fait pas partie de notre travail de recherche.

Nous allons présenter par la suite une autre méthode de résolution de problème inverse, les réseaux de neurones artificiels. Celle-ci possède également plusieurs atouts qui la rendent compatible avec une utilisation en temps réel.

### **I.3.3 Une méthode de régression particulière : les réseaux de neurones artificiels**

Plusieurs méthodes de régression ont été utilisées dans la littérature pour résoudre le problème inverse de la diffraction. Les plus simples d'entre elles se basent sur des modèles statistiques linéaires [KRU93, NAQ94, KRU93\*, GIA91]. Il existe aussi de nombreuses méthodes de régression non linéaires mais leurs performances dépendent fortement du modèle non linéaire supposé. Le plus connu est le modèle polynomial [SRI].



Notre attention se portera tout particulièrement dans le cadre de cette thèse sur les réseaux de neurones artificiels, qui peuvent s'apparenter à une méthode de régression basée sur un modèle non linéaire. Cette technique sera développée en détail dans le chapitre suivant (chapitre II), mais nous pouvons déjà donner ici le principe de leur utilisation dans le cadre de la scattérométrie .

La première étape (figure I.32) consiste, comme pour la méthode des librairies, à créer un corpus dit d'apprentissage, composé de couples signature ellipsométrique / paramètres géométriques correspondants définis en début de paragraphe par  $\vec{y}/\vec{x}$  . Cela consiste à générer un grand nombre de signatures théoriques à partir de plusieurs simulations d'échantillons obtenues par tirage aléatoire de paramètres géométriques dans une plage de variation fixée. Ensuite, au cours de la phase d'entraînement, le réseau de neurones « apprend » la relation liant ses entrées  $\vec{y}$  à ses sorties  $\vec{x}$ , c'est-à-dire la relation liant les signatures optiques aux paramètres géométriques.

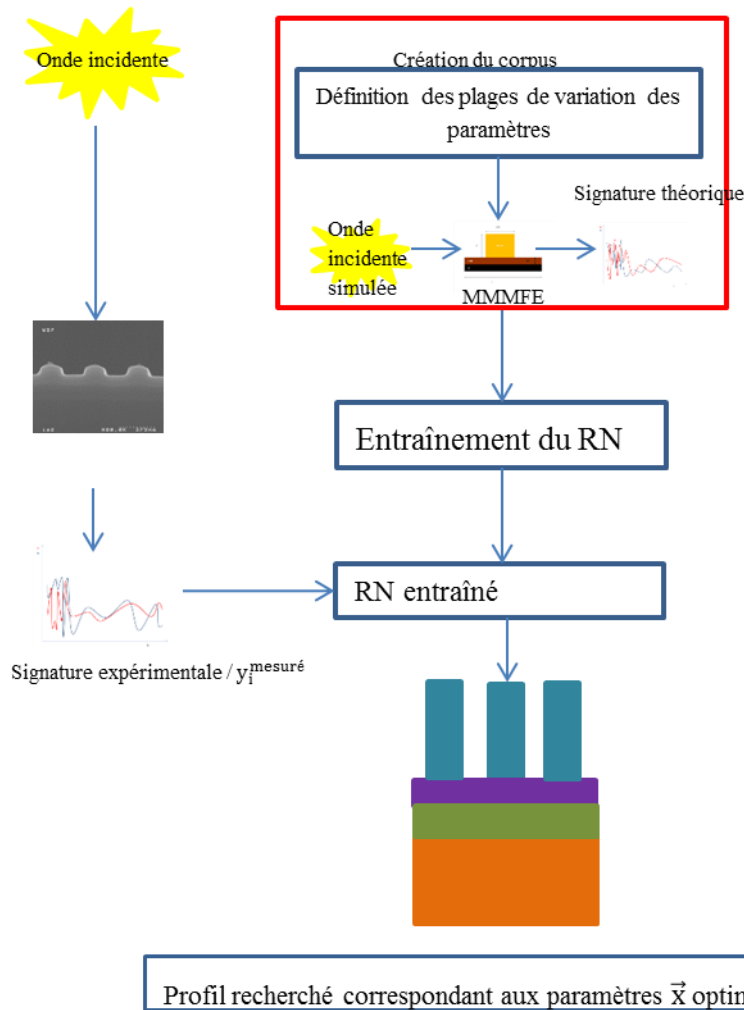


Figure I-29. Diagramme de résolution du problème inverse avec la méthode des réseaux de neurone

Ensuite, il est donc capable de généraliser l'information apprise sur le corpus d'entraînement, à savoir associer à une signature inconnue, un jeu de paramètres donnés. Dans notre cas, la signature proposée sera la signature expérimentale issue d'une mesure ellipsométrique sur l'échantillon testé. Cette dernière étape est intitulée phase de traitement. Elle est instantanée, contrairement à la phase d'apprentissage qui peut prendre plus de temps. Nous verrons plus loin dans le manuscrit que, selon l'architecture du réseau de neurones, cette phase peut durer plusieurs minutes.

A noter que l'utilisation de signatures simulées à la place de signatures expérimentales est justifiée par les mêmes raisons que pour la construction de la bibliothèque. En effet, il n'est pas réaliste de faire une multitude de mesures ellipsométriques pour constituer le corpus car les contraintes de temps et de coûts seraient bien trop importantes. Ainsi, nous utilisons la

méthode MMFE pour simuler les mesures nécessaires à la construction du corpus à partir d'une modélisation paramétrique des différents échantillons.

Plusieurs travaux ont mis en évidence le fait que les RN sont efficaces pour la résolution de problème inverse en scattérométrie [KAL98, KAL99]. Les RN ont montré une meilleure robustesse au bruit par rapport à la technique des Moindres Carré Partiels (PLS pour l'acronyme anglais Partial Least Squares) [BIS98]. Robert et al. [ROB02] ont montré qu'une architecture simple de RN (cf. Chapitre II) permettait d'atteindre une précision raisonnable sur les paramètres géométriques. Les RN ont également été mis en pratique sur un cas expérimental en confrontant les résultats obtenus avec des techniques morphologiques (AFM, XSEM) [ROB02\*]. Une méthode basée sur l'utilisation des caractéristiques propres au réseau de neurones (méthode HVS pour Heuristic Variable Selection) a été mise en place pour choisir la signature optique optimale pour une configuration donnée [ROB04\*]. Ceci permet de réduire le nombre de mesures à effectuer et, par conséquent, de réduire à la fois le temps de la mesure et le temps de traitement. Le contrôle de l'homogénéité d'un réseau optique en silicium a ainsi pu être réalisé [ROB05]. Cette étape est très attendue dans le contrôle in-situ pendant le procédé de fabrication. Enfin, une estimation de l'erreur sur chaque paramètre géométrique a été mise en œuvre grâce à l'implémentation d'un second RN [ROB05\*] entièrement dédié à cette tâche. L'utilisation d'une signature ellipsométrique associée à la méthode des Réseaux de Neurones a également été testée par Gereige et al. [GER07, GER07\*].

Ces derniers ont également étudié expérimentalement la capacité du RN à sélectionner un modèle de forme paramétrique parmi plusieurs à partir de la signature expérimentale utilisée pour l'étape de caractérisation proprement dite. Cette étude est en cours au sein du laboratoire LT2C.

D'autres travaux ont mis en valeur la possibilité de tester la solution neuronale sur des motifs complexes (avec des bords ondulés) [WEI08]. Un cas expérimental a ainsi été traité et les résultats comparés à une technique morphologique (SEM). Enfin, Babin et al. [BAB09] ont montré que les RN donnent des performances similaires à d'autres méthodes de résolutions de problème inverse.

L'utilisation des RN dans le cadre de la scattérométrie statique a ainsi largement été validée au LT2C et a déjà fait l'objet de deux thèses [ROB03, GER09] au sein de ce laboratoire. Au même titre que la méthode des bibliothèques, cette technique prometteuse possède

des avantages indéniables pour une caractérisation en dynamique qui seront exploités dans le cadre de ces travaux de thèse.

Nous avons présenté et dressé un état de l'art des différentes méthodes utilisées pour la résolution du problème inverse en scattérométrie statique. Les méthodes d'optimisation peuvent être applicables efficacement sous certaines conditions mais leur inconvénient réside dans le temps de calcul qui peut être important (utilisation régulière du problème direct). Les deux autres méthodes présentées, les bibliothèques et les réseaux de neurones reposent sur deux principes très proches. En effet, ils nécessitent tous les deux une première phase de construction d'un ensemble de couples paramètres géométriques/signatures optiques et une phase d'extraction des paramètres géométriques à partir d'une signature expérimentale.

Les travaux réalisés sur les RN pour la scattérométrie au cours de cette dernière décennie ont montré que cet outil est efficace et bien adapté pour une caractérisation de réseaux en statique. Il possède un potentiel pour la caractérisation in situ de procédé en temps réel. Bien que cette dernière ait déjà été traitée avec la méthode des bibliothèques avec GPU [SOU07, ELK09, ELK09\*], elle n'a jamais encore été couplée à une solution neuronale dans la littérature.

#### **I.4 Conclusion**

Ce chapitre nous a permis de passer en revue les différentes techniques de métrologie utilisées en microélectronique pour la caractérisation de matériaux. Nous avons à la fois exposé les techniques morphologiques (AFM, MEB) et les techniques optiques par scattérométrie. La première catégorie fait partie des techniques directes permettant un aperçu visuel et ne nécessitant pas d'information à priori sur la forme du profil. Leur inconvénient est qu'elles sont parfois destructives et nécessitent un temps de préparation important de l'échantillon. Elles ne sont donc pas adaptées pour des applications en temps réel.

La scattérométrie se base sur l'analyse de la lumière diffractée par une structure périodique pour en déterminer les valeurs paramétriques caractérisant sa forme. Nous avons utilisé dans le cadre de nos travaux plusieurs types de modèles géométriques de réseaux. Nous avons également détaillé les différents types de réponses optiques utilisées dans la littérature en insistant sur la mesure ellipsométrique utilisée dans le cadre de cette thèse qui permet une caractérisation rapide, non invasive et applicable in-situ.

Nous avons également dressé un aperçu des différentes techniques de résolution de problème inverse en développant les limites et les avantages de chacune. L'optimisation est rapidement écartée pour une application en temps réel car coûteuse en temps de calcul ; ce qui n'est pas le cas pour les deux autres techniques que sont la méthode des librairies et l'utilisation de RN.

Dans le prochain chapitre, nous allons nous intéresser plus particulièrement à la technique des réseaux de neurones pour mieux comprendre leur principe de fonctionnement et leur utilisation dans le cadre bien précis de la scattérométrie

## I.5 Bibliographie

- [AZA12] Azarnouche L., “ Défis liés à la rugosité des motifs de résine photosensible”, Ph.D. thesis, Université de Grenoble, 2012
- [AZZ87] Azzam R.M.A, Bashara N.M, “Ellipsometry and polarized light”, 2<sup>nd</sup> edition, 1987
- [BAB09] Babin, S., Doskolovich L., Kadomina E., Kadomin I., Volotovskiy S., « Restoring pattern CD and cross-section using scatterometry: various approaches », Proc. SPIE Vol. 7272, 2009.
- [BIS91] Bishop K.P., Gaspar S.M., Milner M., Naqvi S.S.H., and MacNeil J.R., “Grating line shape characterization using scatterometry.” Proc. SPIE 1545, 64-73 1991.
- [BIS97] Bischoff J, Truckenbrodt H., and Bauer J, “ Diffraction analysis based characterization of very fine gratings,” Proc SPIE3099, 212-222(1997).
- [BIS98] Bischoff J., Bauer J., Haak U., Hutschenreuther L., Truckenbrodt H., « Optical scatterometry of quarter-micron patterns using neural regression » Proc. SPIE , vol 3332, 1998.
- [BEN01] Benesch N., Hettwer A., Schneider C., Pfitzner, L., Ryssel, Heiner, "Phi-scatterometry for integrated linewidth control in DRAM manufacturing," Semiconductor Manufacturing Symposium, 2001 IEEE International , vol., no., pp.129,132, 2001
- [BER03] Bernoux F., “Ellipsométrie – théorie”, techniques de l’ingénieur, 2003
- [BON00] J. F. Bonnans, A. Shapiro (2000). *Perturbation Analysis of Optimization Problems*. Springer Verlag, New York.
- [BRO65] Broyden G.C., « A class of method for solving non linear simultaneous equations », Math. Comput., pp577-593, 1965.
- [CHA80] Chandezon J., Maystre D., Raoult G., “A new theoretical method for diffraction gratings and its numerical application”. J. Optics (Paris), 11: pp 235-241, 1980.
- [CHE13] Chen X., Liu S., Zhang C., Zhu J., « Improved measurement accuracy in optical scatterometry using fitting error interpolation based library search ». Measurement 46, (octobre 2013): 2638-2646
- [COR00] Cormier G., Boudreau R. “Genetic algorithm for ellipsometric data inversion of absorbing layers”, J. Opt. Soc. Am. A, 17:129-134, 2000.
- [CUR44] Curry H., “The method of steepest descent for nonlinear minimization problems”, Quart Appl. Math, 2: 258-261, 1944.

- [DEL50]** Delort T., Maystre D., “Finite element method for gratings”. *J.Opt. Soc. Am. A*, 10:2592- 2601,1993
- [ELK09]** El Kodadi M., Soulan S., Besacier M., and Schiavone P. , “Real time scatterometry for profile control during resist trimming process”. *Journal of Vacuum Science & Technology B*, 27(6), 3232-3237, 2009.
- [ELK09\*]** El Kodadi M., Soulan, S., Besacier M., and Schiavone P. “Resist trimming etch process control using dynamic scatterometry”. *Microelectronic Engineering*, 86(4), 1040-1042., 2009.
- [ELK10]** El Kodadi M., “Développement des techniques de scattérométrie en temps réel pour le suivi des procédés de gravure plasmas”, PhD thesis, Université de Grenoble, 2010.
- [FLE63]** Fletcher R., “Generalized inverses for non linear equations and optimization”, Numerical method for non linear algebraic equations, London, Gordon & Breach,1963.
- [FLE64]** Fletcher R., C.M. Reeves, “Function minimization by conjugate gradients”, *Computer J*, 7: 149-154, 1964.
- [FOW12]** Fowles, G. R., “Introduction to modern optics”, Courier Corporation, 2012
- [FUA13]** Fuard D.,Troscompt N ,El Kalyoubi I., Soulan S., Besacier B., “S-Genius, a universal software platform with versatile inverse problem resolution for scatterometry”, *Proc. SPIE 8789, Modeling Aspects in Optical Metrology IV*, 878919 (May 13, 2013)
- [GAR04]** Garcia-Caurel E., De Martino A., & Drevillon, B. « Spectroscopic Mueller polarimeter based on liquid crystal devices », *Thin Solid Films*, 455, 120-123, 2004
- [GER07]** Gereige I., Robert S., Granet G., Jamon D., et Rousseau J. J., « Rapid Control of submicrometer periodic structures by a neural inversion from ellipsometric measurement ». *Optics Communications* 278, 2:270-273, 2007
- [GER07\*]** Gereige I., Robert S., Granet G., « Optimal architecture of a neural network for a high precision in ellipsometric scatterometry », *Proc. SPIE*, vol 6648, 2007.
- [GER08]** Gereige I., Robert S., Thiria S., Badran F. , Granet G., Rousseau J.J. « Recognition of diffraction-grating profile using a neural network classifier in optical scatterometry », *J. Opt. Soc. Am A*, Vol 25., 7: 2008, 1661:1667.
- [GER09]** Gereige I., “ Contribution des réseaux de neurons dans le domaine de l’ellipsométrie. Application à la scattérométrie », PhD thesis, Université de Saint Etienne, 2009
- [GER12]** Gereige I., Robert S., Eid J. «Automatic detection of photoresist residual layer in

- lithography using a neural classification approach ». *Microelectronic Engineering*, 97:29-32, 2012
- [GIA91]** Giapas K.P., Gottscho R.A., Clark L.A., Kruskal J., Lambert D., Rietman E.A., and Sinatore D., “Use of light scattering in characterizing reactively ion etched profiles”, *J. Vac. Sci. Technol. A* 9,664-668 (1991).
- [GOL68]** Goldfarb D., Lapidus B., “Conjugate gradient method for non linear programming problems with linear constraints”, *I & EC Fundam.*, pp142-151,1968.
- [GOL89]** Goldberg D.E., “Optimal initial population size for binary coded genetic algorithms”, TCGA Report No 85001, Tuscaloosa: University of Alabama, 1985.
- [GOO05]** Goodman J.W., “Introduction to Fourier Optics”, Third edition, Stanford University, 2005.
- [GRA96]** Granet G., Guizal G., “Efficient implementation of the coupled-wave method for metallic lamellar gratings in TM polarization”, *J. Opt. Soc. Am. A*, Vol. 13, No. 5, May 96.
- [GRA97]** Granet G., Chandezon J., “The method of curvilinear coordinates applied to the problem of scattering from surface-relief gratings defined by parametric equations: application to scattering from a cycloidal grating”. *Pure Appl. Opt.*, 6: 727-740, 1997.
- [HAZ03]** Hazart J., Grand G., Thony P., Herisson D., Garcia S., Lartigue O., « Spectroscopic ellipsometric scatterometry: sources of errors in critical dimension control », *Proceedings of SPIE*, vol 5041, 2003
- [HOL02]** Holden J.M., Gubiotti T., McGahan W.A., Dusa M.V., and Kiers T., “Normal incidence spectroscopic ellipsometry and polarized reflectometry for measurement and control of photoresist critical dimension”, In Daniel J.C. Herr, editor, *Metrology, Inspection, and Process Control for Microlithography XVI*, volume 4689, page 1110-1121. SPIE, 2002
- [HOL75]** Holland J.H., “Adaptation in natural and artificial systems”, Ann Arbor: The University of Michigan Press, 1975.
- [HUA04]** Huang H.T., Terry F.L. Jr., Erratum to “Spectroscopic ellipsometry and reflectometry from gratings (Scatterometry) for critical dimension measurement and in



- situ real time process monitoring” [ Thin Solid Films 455-456(2004) 828-836]. Thin Solid Films, 468(1), 339-346.
- [KAL98]** Kallioniemi I., Saarinen J., Oja E., «Optical Scatterometry of Subwavelength Diffraction Gratings: Neural-Network Approach ». Applied Optics 37, 25:5830- 5835, 1998.
- [KAL99]** Kallioniemi I., Saarinen J., Oja E., «Optical scatterometry with neural network model for nondestructive measurement of submicron features », Proc. SPIE, vol. 3743 , 1999.
- [KAL99\*]** Kallioniemi I., Saarinen J., Oja E., «Characterization of Diffraction Gratings in a Rigorous Domain with Optical Scatterometry: Hierarchical Neural-Network Model ». Applied Optics 38, 28: 5920-5930 , 1999.
- [KIR83]** Kirpatrick S., Gelatt C.D., and Vecchi M.P., “Optimization by simulated annealing”, science, vol 220, pp. 671-680, 1983.
- [KLE08]** Kleijnen J., “Design and analysis of simulation experiments”, Springer, 2008
- [KRU93]** Krukar R., Prins S. , Krukar D. M., Peterson G., Gaspar S., McNeil J, Sohail H.N, Hush D.R . « Using scattered-light modeling for semiconductor critical dimension metrology and calibration », Proc. SPIE, vol 1926, 1993
- [KRU93\*]** Krukar R., Cornblit A., Clark L.A., Kruskal J., Lambert D., Rietman E.A., and Gottscho R.A., “Reactive ion etching profile and depth characterization using statistical and neural analysis of light scattering data”, J. Appl. Phys. 74, 3698-3706 (1993).
- [LEE93]** Lee R. E., “Scanning electron microscopy and x-ray microanalysis”, PTR Prentice Hall Englewood cliffs, N.J. (1993)
- [LEG]** Le Gressus C., “Microscopie Electronique à Balayage”, Technique de l’ingénieur PE 865.
- [LEV44]** Levenberg K., “A method for the solution of certain problems in least squares”, Quart. Appl. Math., 2:164-168, 1944.
- [LI93]** Li L., Haggans C.W., “Convergence of the coupled-wave method for metallic lamellar diffraction gratings”, J. Opt. Soc. Am. A, Vol. 10, pp.1184-1189, 1993.

- [MAI01] Maiken E., Johnson K., Likhachev D., Norton A., Sezginer A., Stanke F., Yonenaga D., « Integrated Critical Dimension Metrology for lithography Process Control », AEC/APC, symposium XIII, Banff, Canada, October 6-11, 2001.
- [MAR94] Martin Y., Wickralasinghe H., “Method for imaging sidewalls by atomic force microscopy”, *Appl. phys. lett.* 64, 2498, (1994).
- [MAR63] Marquardt D., “An algorithm for least squares estimation of nonlinear parameters”, *SIAM. J. Appl. Math.*, 11: 431-441, 1963.
- [MAY72] Maystre D., “Sur la diffraction d’une onde plane par un réseau métallique de conductivité finie”, *Opt. Comm.*, 6 :pp 50-54 1972.
- [MIL73] Millar R.F., “The Rayleigh hypothesis and a related least squares solution to scattering problems for periodic surfaces and other scatterers” *Radio Sci.*, Vol. 8, pp. 785-796, 1973
- [MOH81] Moharam M.G., Gaylord T.K., “Rigorous coupled-wave analysis of planar-grating diffraction. *J. Opt. Soc. Am. A*, 71:pp 811-818, 1981.
- [MOR77] More J.J., “The Levenberg Marquardt algorithm: implementation and theory, *Numerical Analysis*”, Springer Verlag, G. A. Watson, 105-116, 1977.
- [NAQ94] Naqvi S.S.H., Krukar R., McNeil J.R., Franke J.E., Niemczyk T. M., Haaland D.M., Gottsho R.A., and Kornblit, “ Etch depth estimation of large period silicon gratings with multivariate calibration of rigorously simulated diffraction profiles” *J. Opt. Soc. Am. A* 11, 2485-2493 (1994).
- [NEV71] Nevière M., Cerutti-Maori G., Cadilhac M., « Sur une nouvelle méthode de résolution du problème de la diffraction d’une onde plane par un réseau infiniment conducteur. » *Opt. Comm.*, 3 :pp 48-52, 1971
- [NIU99] Niu X., Jakatdar N., Bao J., Spanos C. and Yedur S., « Specular spectroscopic scatterometry in DUV lithography », *SPIE*, vol 3677, 1999.
- [NOC99] Nocedal J. and Wright S. J., “Numerical optimization”, Springer Verlag, NY, 1999.
- [OWE75] Owens John D., Luebke D., Govindaraju N., Harris M., Kruger J., Lefohn A. E., Purcell T. J. “A survey of general-purpose computation on graphics hardware”. *Computer Graphics Forum*, 26(1):80-113, 2007.

- [PET65] Petit R., « Etude théorique de la diffraction d'une onde plane et monochromatique par un réseau métallique infiniment conducteur ». *App. Opt.*, 4: pp 1551-1554, 1965
- [PET66] Petit R., « Contribution à l'étude de la diffraction par un réseau métallique ». *Rev. Opt.*, 45 :pp 249-276, 1966.
- [PET80] Petit R., « Electromagnetic theory of gratings ». Springer-Verlag, Berlin, 1980.
- [PET05] Petit J., Boher P., Leroux T., Barritault P., Hazart J., & Chaton, P. “ Improved CD and overlay metrology using an optical Fourier transform instrument”. In *Microolithography 2005* (pp. 420-428). International Society for Optics and Photonics, May 2005.
- [POI08] Potié A., Stage de fin d'étude, Polytech'Grenoble , 2008
- [RAY95] Raymond C.J., Murnane M.R., Prins S.L., Naqvi S.S., McNeil J.R., « Multi-parameter Process metrology using scatterometry », SPIE Vol.2638, Optical Characterization Techniques for High Performance Microelectronic Device Manufacturing II, 84-93.
- [REY] Reyraud C., “le microscope électronique à balayage”, [www.cea.fr/technologies/les-microscopes/le-microscope-electronique-a-balayage-meb](http://www.cea.fr/technologies/les-microscopes/le-microscope-electronique-a-balayage-meb) , consulté le 20/10/13
- [RIV05] Rivoal J.C., Fretigny C., “Microscopie à force atomique”, *Technique de l'ingénieur*, r1394.
- [ROB02] Robert S., Mure-Ravaud A., Lacour D., « Characterization of optical diffraction gratings by use of a neural method ». *Journal of the Optical Society of America A* 19, 1: 24-32 ,2002
- [ROB02\*] Robert S., Mure Ravaud A., Reynaud S., Fourment S., Carcenac F., Arguel P., « Experimental characterization of subwavelength diffraction gratings by an inverse-scattering neural method ». *Journal of the Optical Society of America A* 19, 12: 2394-2402.,2002
- [ROB03] Robert S., « Caractérisation optique de réseaux de diffraction submicroniques par des techniques génétique et neuronale ». PhD thesis, Université Jean Monnet de Saint Etienne, 2003.
- [ROB04] Robert S., et Mure-Ravaud A., « Resolution of the inverse problem of optical grating testing by means of a neural network », *Proc. SPIE*, vol 5252, 2004.
- [ROB04\*] Robert S., Mure-Ravaud A., Thiria S., Yacoub M., Badran F. « Neural selection of the optimal optical signature for a rapid characterization of a submicrometer period grating ». *Optics Communications* 238, 4-6 : 215-228, 2004.

- [ROB05]** Robert S., Mure-Ravaud A, « Control of the homogeneity of an optical grating by a neural characterization ». *Optical Engineering* 44, 3: 033601-033601, 2005.
- [ROB05\*]** Robert S., Mure-Rauvaud A., Thiria S., Badran F. « Estimation of local error by a neural model in an inverse scattering problem ». *The European Physical Journal Applied Physics* 31, 1: 71-76, 2005.
- [SAC89]** Sacks J., “ Design and analysis of computer Experiments”, *Statistical Science*, Vol. 4., No. 4, 1989
- [SCH69]** Schmieder, R. W, “Stokes-algebra formalism”, *JOSA*, 59(3), 297-302, (1969).
- [SHA70]** Shanno D.F., “ An accelerated gradient projection method for linearly constrained non linear estimation”, *SIAM J. Appl. Math.*, pp. 322-334,1970.
- [SOH92]** Sohail S., Naqvi H., Gaspar S., Hickman K., Bishop K., McNeil John R., “Linewidth measurement of gratings on photomasks: a simple technique”, *Applied Optics*, Vol 31, Issuer 10, pp. 1377-1384 (1992)
- [SOU07]** Soulan S., Besacier M, Schiavone P, Leveder T, “In-line etching process control using dynamic scatterometry“, *Proc. SPIE 6617, Modeling Aspects in Optical Metrology*, 661713 (June 18, 2007).
- [SOU07\*]** Soulan S., Besacier M. , Leveder T. , Schiavone P, “Real-time profile shape reconstruction using dynamic scatterometry”. *Proc. SPIE 6518, Metrology, Inspection, and Process Control for Microlithography XXI*, 65180W (April 05, 2007);
- [SOU08]** Soulan S., “Développement de la scattérométrie dynamique pour le suivi en temps réel de procédés. Application à la microélectronique. » . PhD thesis, Université de Grenoble, 2008.
- [SRI]** Srivastava R., Polynomial regression, consulté le 21/01 2015, <http://www.iasri.res.in/iasriwebsite/designofexpapplication/electronic-book/module5/11PRegres.pdf>
- [TAY94]** Tayeb G., “The method of fictitious sources applied to diffraction gratings”, *Appl. Comput. Electromag, Soc.*, 9:90-100,1994.
- [TIK77]** Tikhonov A. N., Arsenin V.A. *Solution of III-posed problems*. 1977
- [THI06]** Thiault J., « Etude 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”, Ph.D. thesis, Université Joseph Fourier de Grenoble (2006)
- [TOM05]** Tompkins H.G., Irene E.A., “Handbook of Ellipsometry”, William Andrew Publishing, 2005

- [TRA93] Trancoso P., Chralambous M., “Exploring graphics processor performance for general purpose applications”. In DSD '05: Proceedings of the 8<sup>th</sup> Euromicro Conference on Digital System Design, page 306-313, Washington, DC, USA, 2005, IEEE Computer Society
- [YAS13] Yasui N., Isawa M., Ishimoto T., Sekiguchi K., Tanaka M., Osaki M., Shishido C., Hasegawa N., Cheng S., « Application of model-based library approach to photoresist pattern shape measurement in advanced lithography », Proceedings of SPIE, Vol 7638, 2010.
- [VAN81] Van den Berg P.M., “Refelection by a grating: Rayleigh methods”, Journal Society of America, Vol. 71, pp. 1224-1229, 1981
- [WEI08] Wei S., Lifeng Li. « Measurement of photoresist grating profiles based on multiwavelength scatterometry and artificial neural network ». Applied Optics 47, 13: 2524-2532, 2008.





---

## **Chapitre II. Scattérométrie par réseaux de neurones**

---



Dans le chapitre précédent, les différentes techniques utilisées en scattérométrie pour résoudre le problème inverse ont été examinées. Le deuxième chapitre porte sur les réseaux de neurones, une technique prometteuse pour la résolution du problème inverse en scattérométrie. Un réseau de neurones artificiels (RN) est un outil mathématique qui permet de modéliser un système physique à condition de disposer de suffisamment d'exemples décrivant ce phénomène. Dans ce chapitre, le principe général des réseaux de neurones sera parcouru en présentant rapidement leur origine biologique et leur formulation mathématique. La phase d'apprentissage, essentielle au bon fonctionnement des RN, sera ensuite détaillée avec les propriétés sous-jacentes qu'elle exige. Une étude théorique sera ensuite conduite pour définir l'architecture optimale d'un RN dans le cadre précis de la caractérisation qui nous préoccupe. Enfin, une étude expérimentale nous servira de trame de fond pour comparer les performances des RN avec les autres techniques de métrologie morphologique présentées dans le chapitre précédent.

## **II.1 Présentation des réseaux de neurones**

Le fonctionnement d'un RN artificiel est inspiré de celui d'un cerveau humain. Il se décompose en deux phases distinctes: une phase d'apprentissage et une phase de traitement. Prenons la métaphore d'un enfant en cycle primaire. Pour connaître ses tables de multiplication, une première phase d'apprentissage est nécessaire puis une phase de test a lieu pour évaluer s'il a bien appris ou non. Un bref historique des RN sera présenté dans la suite avant de présenter les définitions essentielles permettant de bien comprendre leur fonctionnement et leur utilisation.

### **II.1.1 Historique des réseaux de neurones**

Les premiers travaux sur les réseaux de neurones remontent à 1890, date à laquelle W. James, un psychologue américain, a présenté le concept de mémoire associative. Un demi-

siècle plus tard, les deux neurologues Mc Culloch et W. Pitts [MCC43] mettent en place un modèle simplifié du modèle biologique qu'ils appellent le neurone formel. Celui-ci est simplement un neurone binaire, c'est-à-dire qu'il donne des sorties 0 ou 1. Ils démontrent que ce modèle est capable d'effectuer des calculs logiques, arithmétiques et symboliques simples. D. Hebb, un physiologiste canadien, indique en 1949 une règle d'apprentissage, qui porte le nom de son auteur, dans son ouvrage «The organization of Behaviour » [HEB49]. Quelques années plus tard, F. Rosenblatt propose le premier modèle de RN, le perceptron [ROS57], suite à de premières études sur la compréhension du cerveau humain ainsi que les processus cognitifs. En 1960, B. Widrow et Hoff mettent en place le modèle de l'ADALINE (ADaptive LINear Element) qui est aujourd'hui reconnu comme le premier réseau de neurones, bien que composé d'un seul neurone. Mais ce succès sera suivi d'une période de déclin, suite à la publication en 1969 par M. Minsky et S. Papert [MIN69] d'une critique des propriétés du perceptron mettant en doute la possibilité de traiter des problèmes non linéaires. Plusieurs scientifiques ont néanmoins poursuivi les recherches et ont réussi à proposer de nouveaux modèles. Entre temps, certains scientifiques comme T. Kohonen, proposent des réseaux pouvant servir de mémoires associatives [KOH82] en développant le concept de carte auto-organisée. J. Hopfield [HOP82] apporte une forte contribution dans le domaine en publiant une étude d'un réseau dynamique qui évolue jusqu'à atteindre un état de stabilité par un comportement autonome. C'est dans le milieu des années 80 que des avancées majeures ont été réalisées par Y. Le Cun et Rumelhart [LEC85, RUM86] avec la mise au point de l'algorithme de rétropropagation du gradient de l'erreur permettant l'apprentissage des Perceptrons Multi Couches (PMC). Ces avancées ont permis de pallier les difficultés mentionnées ci-dessus par M. Minsky et S. Papert. C'est le début de l'essor des RN et surtout des PMC.

### **II.1.2 Applications des réseaux de neurones**

Les RN possèdent un grand nombre d'applications dans de nombreux domaines aussi divers que variés. On les retrouve aussi bien en traitement d'image où la reconnaissance de motifs est importante (reconnaissance faciale [AGA10]) que dans des domaines plus singuliers comme les jeux vidéo [STA06]. On peut également citer l'application des RN dans la bio-ingénierie pour la prédiction des propriétés pharmacologiques ainsi que l'activité thérapeutique des molécules pour la découverte de nouveaux médicaments [HAN95]. En robotique, on a utilisé les RN pour le pilotage automatique d'un véhicule autonome [RIV94, 95]. Les RN interviennent alors au niveau du pilotage pour déterminer les actions souhaitables

en fonction des consignes de cap et de vitesse. Les RN sont employés aussi en aéronautique [WU13] pour améliorer la trajectoire d'urgence d'un avion dans le cas d'une panne totale de moteur. Dans le domaine du marketing, on assiste de plus en plus à l'augmentation du volume d'informations accessibles grâce à l'apport du numérique. Il est indispensable d'avoir à disposition des outils efficaces, comme les RN, permettant de traiter efficacement et rapidement les informations pertinentes [ENK05]. Ce domaine connu sous le terme de data-mining ou fouille de données.

Les RN trouvent aussi des applications dans des secteurs de la vie quotidienne. Citons, par exemple dans le domaine de l'environnement, le contrôle de systèmes de collectes des eaux pluviales [ROU01] ou l'étude de la contamination des ressources hydriques souterraines [FOD11]. On les trouve également dans la reconnaissance automatique des codes postaux [LEC89].

Certaines problématiques d'ingénierie font également appel à une résolution par RN, comme par exemple, la prédiction de l'évolution d'une espèce chimique afin de simuler et optimiser la conduite de certains réacteurs chimiques [AMM07]. Dans ses travaux, Boueri [BOU10] a également employé les RN pour l'identification d'une famille de polymères dans le but de mettre en place une identification en temps réel de matériel plastique sur une chaîne de tri.

Cette énumération non exhaustive de plusieurs applications sur un large spectre de domaines montre que cet outil a largement fait ses preuves notamment en termes de performance et de robustesse. Cela fait également partie des arguments qui nous ont poussés à développer l'utilisation de cet outil dans notre domaine de recherche : la métrologie en micro et nanoélectronique.

Un RN artificiel peut être utilisé dans différentes catégories, comme la classification, le data-mining ou la prédiction. C'est cette dernière fonction qui sera exploitée dans ce mémoire. Le comportement d'un système est ainsi prédit en connaissant une entrée donnée. On parle alors de généralisation.

### II.1.3 La métaphore biologique

Les neurones artificiels se basent sur le principe de fonctionnement des neurones biologiques composant le cerveau humain. Ce dernier est un système très complexe et très puissant. Aujourd'hui encore, les scientifiques sont incapables de réaliser un système qui posséderait les mêmes capacités.

Le système nerveux est composé de milliards de cellules spécialisées nommées neurones. Ensemble, ces neurones commandent l'action de l'être humain (alimentation, action, repos) et forment ce qu'on appelle un réseau de neurones. Chaque neurone possède différentes portes d'entrée: les dendrites (Figure II-1). Ces derniers reçoivent un signal, provenant des différents organes du corps humain ou d'autres neurones, qui est ensuite traité dans le noyau, nommé corps cellulaire. La réponse est alors émise à travers les axones représentant les lignes de transmission de l'information du système nerveux. La zone de contact avec un autre neurone est la synapse. Celle-ci peut être excitatrice ou inhibitrice, en fonction de l'information envoyée par le corps cellulaire. Elle est connectée, soit à un autre neurone, soit à un muscle afin de réaliser une action. L'information est véhiculée sous forme de signal électrique par le biais d'un potentiel d'action.

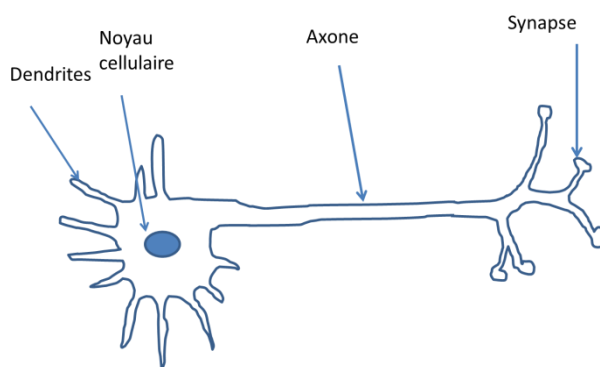


Figure II-1. Structure schématique d'un neurone biologique

Pour appréhender le processus de fonctionnement du cerveau, prenons un exemple simple, celui d'une personne qui tient un objet très chaud. Le cerveau va donner l'ordre au muscle de la main de relâcher l'objet immédiatement. Dans cet exemple, le signal d'entrée est la température de l'objet transmise au cerveau par les récepteurs sensoriels de la main. Le cerveau va analyser la situation et envoyer un signal via les neurones aux muscles de la main. Le signal de sortie est la réaction de l'individu qui consiste à relâcher l'objet chaud. Cette expérience va permettre au cerveau d'apprendre et de se souvenir de cette situation afin de

guider les réactions futures. Cette expérimentation permet à l'être humain de s'entraîner pour s'adapter à son environnement.

La richesse de l'architecture du cerveau humain a poussé certains scientifiques à vouloir s'en inspirer pour réaliser des modèles mathématiques utilisés dans de nombreuses disciplines scientifiques.

### II.1.4 Formulation mathématique d'un neurone artificiel

Suite à l'observation du neurone biologique, Mc Culloch et W. Pitts ont mis en place une structure mathématique permettant de réaliser des actions du même ordre : le neurone formel. La Figure II-2 présente la structure d'un neurone  $j$  connecté à plusieurs entrées et la formulation mathématique utilisée.

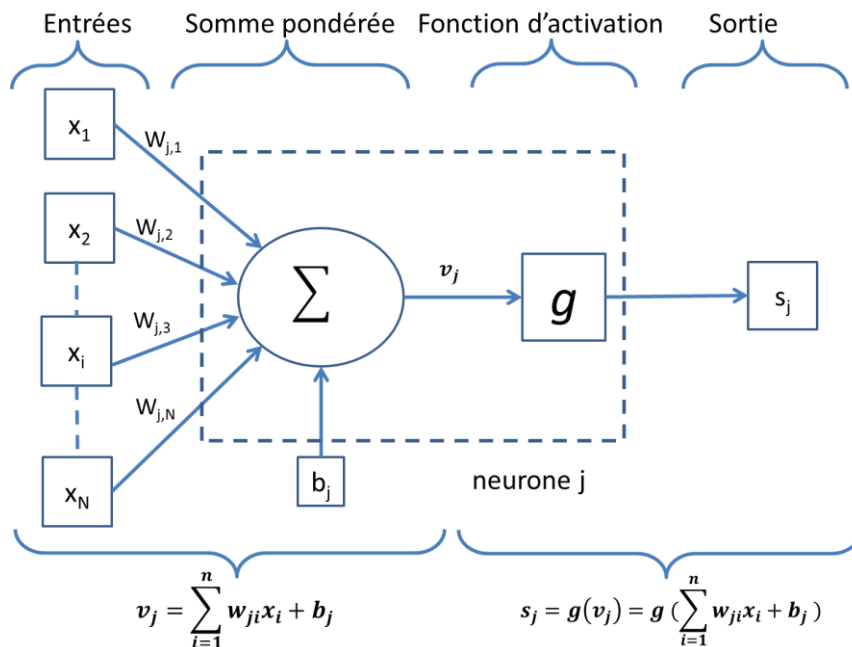


Figure II-2. Représentation mathématique d'un neurone formel. Les  $n$  entrées sont désignées par les variables  $x_i$ , la sortie est notée  $s_j$  et les poids synaptiques pondérant chaque connexion sont désignés par les variables  $w_{j,i}$

Le fonctionnement du neurone se déroule en deux étapes :

Dans un premier temps, on effectue un prétraitement de l'information reçue en calculant le potentiel  $v_j$  du neurone en fonction des entrées  $x_i$  ( $i=1,2,\dots,n$ ) qui lui sont affectées, par l'intermédiaire de la fonction d'entrée totale suivante:

$$v_j = \sum_{i=1}^n w_{ji}x_i + b_j \quad (\text{II. 1})$$

dans laquelle  $w_{j,i}$  est le poids synaptique de la connexion qui lie l'entrée  $i$  au neurone et  $b_j$  représente un paramètre propre au neurone nommé biais.

L'information est ainsi transmise, des entrées  $x_i$  ( $i=1,2,\dots,n$ ) jusqu'au neurone  $j$ , à travers les différentes connexions. Chacune de ces liaisons est associée à un poids synaptique, pour pondérer l'information reçue par le corps cellulaire à travers les dendrites. Le poids représente donc la puissance d'une synapse en amplifiant ou en diminuant l'influence d'un message transmis par un neurone.

Il est généralement commode de considérer le biais comme le poids d'une entrée constante  $x_0$  égale à 1. L'équation II.1 devient alors :

$$v_j = \sum_{i=0}^n w_{ji} x_i \quad (\text{II. 2})$$

**Remarque:** L'équation II.2 présente un exemple particulier de fonction d'entrée totale, généralement appliquée à un neurone qualifié de linéaire. C'est celui qui sera utilisé par la suite. Cependant, il existe dans la littérature d'autres fonctions d'entrée totale (fonction polynomiale, fonction distance, fonction de type noyau) utilisées pour d'autres applications, en particulier dans le domaine de la classification [MIS 06].

La deuxième étape concerne le calcul de la sortie neurone  $s_j$  définissant ainsi l'état interne du neurone en fonction du potentiel  $v_j$  :

$$s_j = g(v_j) = g\left(\sum_{i=0}^n w_{ji} x_i\right) \quad (\text{II. 3})$$

La fonction  $g$  est nommée fonction de transfert, fonction d'activation, ou même fonction d'état suivant les ouvrages du domaine.

Il existe plusieurs possibilités pour la fonction d'activation du neurone comme par exemple la fonction seuil, la fonction identité ou la fonction gaussienne (Figure II-3).

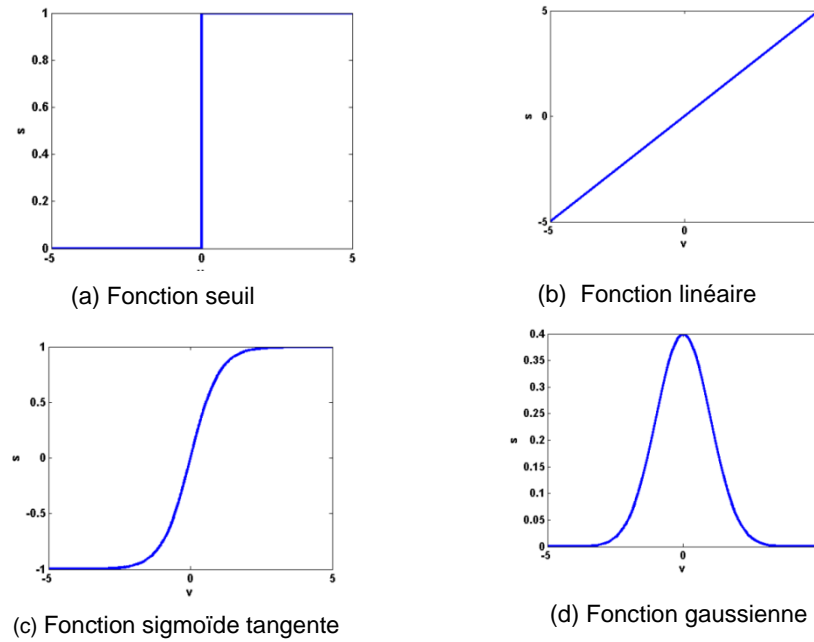


Figure II-3. Exemples de fonctions d'activations communément utilisées pour calculer la sortie d'un neurone formel

Chaque fonction est choisie suivant le type de sorties à modéliser (binaires, décimales, positives) mais également suivant la propriété qu'elle confère au RN ainsi construit. Par exemple, la fonction sigmoïde (en forme de « s »), continue et différentiable, (Figure II-3-c) est utilisée dans certaines catégories des RN car elle introduit une non linéarité. On la retrouve sous deux formes qui sont respectivement la fonction logistique (équation II.4) et la fonction hyperbolique (équation II.5).

$$g(v) = \frac{1}{e^{-Kv} + 1} \quad (\text{II. 4})$$

$$g(v) = \frac{e^{Kv} - 1}{e^{Kv} + 1} = \tanh\left(\frac{K}{2}v\right) \quad (\text{II. 5})$$

La valeur du paramètre K est généralement égale à 1.

La fonction seuil (Figure II-3-a) est la première fonction à avoir été utilisée dans le tout premier RN. Elle est utilisée dans le cas où l'on souhaite avoir une sortie binaire (0 ou 1). Elle est définie de la façon suivante :

$$g(x) = 0 \quad \text{si} \quad x < 0 \quad (\text{II. 6})$$

$$g(x) = 1 \quad \text{si } x > 0 \quad (\text{II. 7})$$

La fonction identité est également nommée linéaire (Figure II-3-b) et s'exprime simplement par :

$$g(x) = x \quad \forall x \quad (\text{II. 8})$$

Enfin, la fonction gaussienne (Figure II-3-d) est définie par:

$$g(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (\text{II. 9})$$

avec les paramètres  $\mu$  et  $\sigma$  représentant respectivement l'espérance mathématique et l'écart type de la modélisation.

En résumé, un neurone  $j$  donné peut ainsi être défini comme une fonction algébrique globale  $f$  non linéaire, paramétrée par  $w_{j,i}$  (avec  $i=0,1, 2,\dots,n$ ), à valeurs bornées, de variables réelles  $x_i$  (avec  $i=1,2,\dots,n$ ) nommée entrées:

$$s_{j,n} = f(x_1, x_2, \dots, x_n, w_{j,0}, w_{j,1}, w_{j,2}, \dots, w_{j,n}) \quad (\text{II. 10})$$

### II.1.5 Architecture d'un réseau de neurones

Les neurones présentés dans le paragraphe précédent offrent des propriétés remarquables lorsqu'ils sont combinés les uns aux autres : on parle tout naturellement de Réseau de Neurones. Il s'agit donc d'un ensemble de neurones interconnectés entre eux. Une première manière d'aborder un RN artificiel est de considérer un graphe, possédant des entrées et des sorties et où chaque nœud représente un neurone artificiel. Au sein de ce réseau, les connexions effectuent la liaison entre les neurones. Les entrées désignent l'information issue de l'environnement extérieur. La sortie du RN est constituée par les neurones qui effectuent les derniers calculs. Enfin, les neurones qui effectuent des calculs intermédiaires sont qualifiés de neurones cachés. Ils sont par conséquent situés entre les entrées et les neurones de sorties.

On peut distinguer deux grandes classes: les réseaux bouclés (appelés aussi dynamiques ou encore récurrents) et les réseaux non bouclés (ou statique).



La première famille permet de modéliser des processus dynamiques en introduisant une hiérarchisation temporelle des activations de certains neurones appartenant à un même cycle. Par exemple, les sorties du RN à l'instant  $t$  seront utilisées comme entrée du RN à l'instant  $t+1$ . Notons que cette notion de « RN dynamique » concerne ici un mode de fonctionnement particulier des RN. Ceci est à distinguer d'un « processus dynamique » physique tel qu'un procédé de gravure comme nous le verrons plus loin dans le manuscrit (cf. Chapitre III). Dans l'étude effectuée dans cette thèse, ce type de RN dynamique n'a pas été utilisé pour le suivi d'un procédé dynamique tel que la gravure plasma.

Dans les travaux développés ici, nous avons utilisé les réseaux de neurones non bouclés où l'information circule des entrées vers les sorties sans qu'il y ait un retour en arrière (à la différence des réseaux dynamiques). Le temps n'a aucune influence dans ce type de réseaux. C'est pour cette raison qu'on les qualifie de réseaux statiques.

On retrouve dans la littérature plusieurs possibilités d'arrangement des neurones pour constituer un réseau. On parle généralement d'architecture du RN. Citons en guise d'exemple le réseau complètement connecté (Figure II-4) où chaque neurone est connecté à toutes les entrées  $x_i$  du réseau ainsi qu'aux sorties des neurones de numéro inférieur. Ce type de RN correspond au modèle le plus complet car il possède le plus grand nombre de connexions (donc de poids), pour un nombre de neurones fixés.

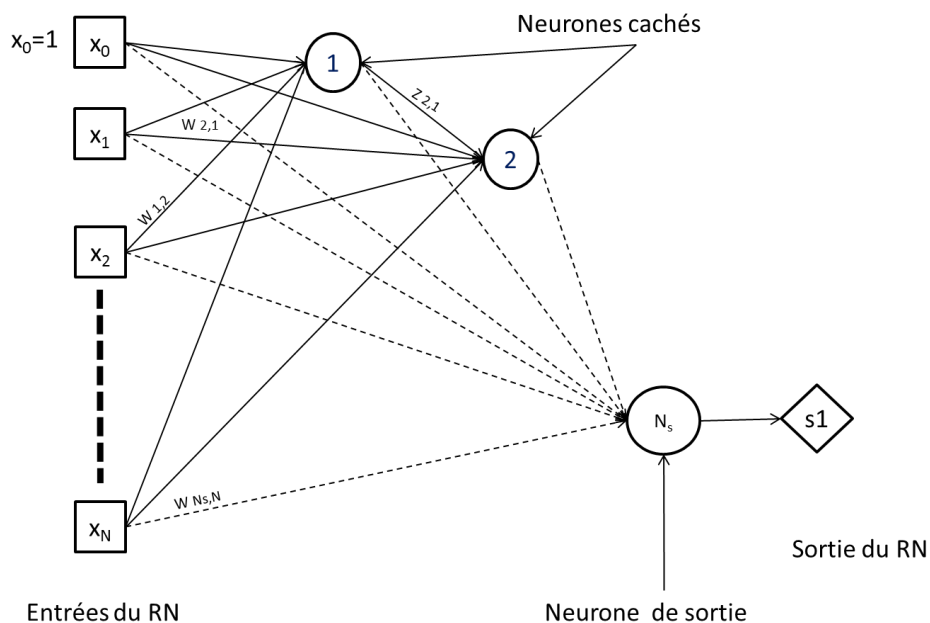


Figure II-4. Exemple d'architecture d'un réseau de neurones non bouclé, complètement connecté comportant  $(N_s-1)$  neurones cachés,  $N$  entrées et un neurone de sortie. On distingue les poids des connexions reliant les entrées à tous les neurones par  $w$  de ceux relatifs aux connexions entre neurones notés  $z$ .

A partir de ce modèle, il est possible de créer des architectures de moindre complexité en supprimant certaines connexions. Nous ne détaillerons pas toutes les possibilités mais seulement le modèle le plus répandu: le modèle dit à couches. Ce modèle utilisé par la suite possède des propriétés intéressantes et est présenté en détail dans le paragraphe suivant.

## II.1.6 Les réseaux de neurones à couches

### II.1.6.1 Présentation de l'architecture

Dans une organisation de réseaux en couche (Figure II-5), les neurones sont disposés en couches verticales successives dans lesquelles toute connexion entre les neurones d'une même couche est interdite. Les neurones d'une même couche se trouvent donc tous connectés aux neurones de la couche précédente et à ceux de la couche suivante. Il faut également noter que les neurones d'une même couche possèdent tous les mêmes propriétés (notamment la même fonction d'activation). Le nombre de neurones constituant la couche dite de sortie est conditionné par le nombre de sorties souhaitées. Toutes les autres couches de neurones réalisant des calculs intermédiaires sont qualifiées de couches cachées.

**Remarque:** Il est commun de trouver dans la littérature l'appellation couche d'entrée, voire même neurones d'entrée pour qualifier les entrées du RN. Cette expression est trompeuse car les entrées (représentées par des carrés sur la Figure II-5) ne sont pas des neurones: elles ne réalisent aucun traitement de l'information puisqu'elles ne font que transmettre les valeurs des variables d'entrées.

Ce type de structure est également dénommé perceptron multicouches (PMC). Le lecteur pourra trouver un état de l'art détaillé des perceptrons multicouches dans [DRE97], [DAV90], [JOD94].

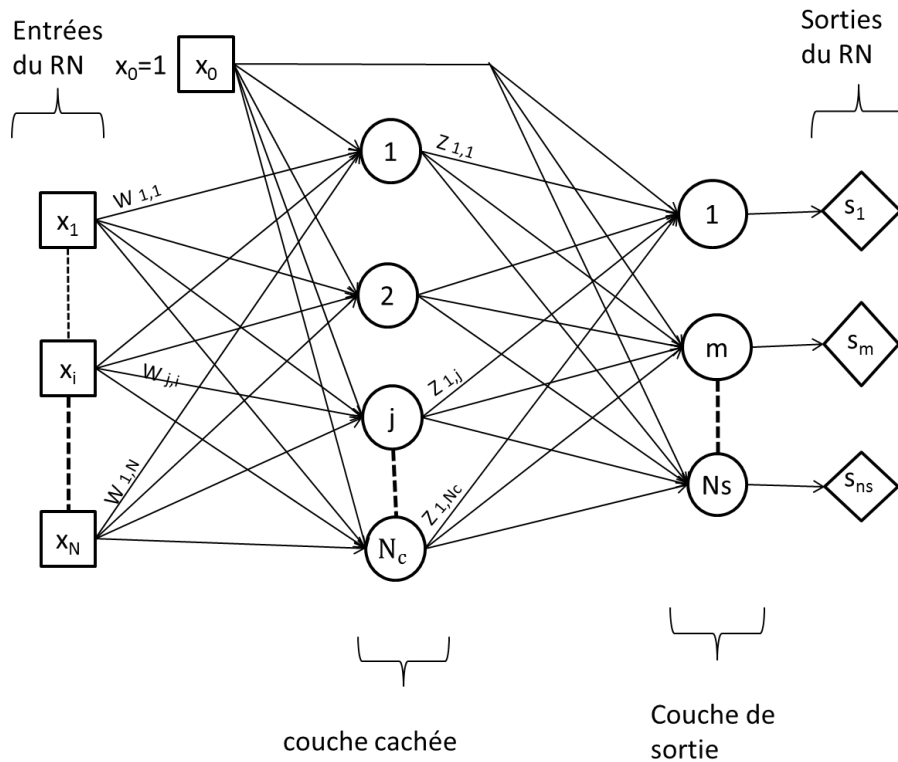


Figure II-5. Exemple d'architecture d'un réseau de neurone non bouclé dit à couches présentant une seule couche cachée de  $N_c$  neurones comprenant  $N$  entrées et  $N_s$  sorties. Soit  $w_{j,i}$  le poids synaptique reliant le  $j^{\text{ième}}$  neurone de la couche cachée à l'entrée  $x_i$  et  $z_{l,j}$  celui reliant le premier neurone de la couche de sortie avec la sortie du  $j^{\text{ième}}$  neurone de la couche cachée.

Chaque PMC utilisé par la suite sera entièrement caractérisé par le nombre d'entrées, le nombre de couches du réseau, le nombre de neurones par couche, la nature des différentes connexions entre les neurones et le type de neurones (fonction de prétraitement, fonction d'activation) composant chaque couche. Cet ensemble définit l'architecture du RN.

### II.1.6.2 Propriétés d'un PMC à une couche cachée

Nous allons décrire dans ce paragraphe un type particulier de PMC, comprenant une seule couche cachée, qui présente des propriétés remarquables.

Considérons tout d'abord un PMC constitué d'une couche cachée de  $N_c$  neurones linéaires présentant une fonction d'activation a priori non linéaire notée  $g$ . Les  $N_s$  neurones de sorties, linéaires également, utilisent, quant à eux, une fonction d'activation identité. Le vecteur noté  $\vec{x}=[x_1, x_2, \dots, x_i, \dots, x_N]^T$  est introduit en entrée du PMC. Le calcul de la sortie  $s_m$  (soit la sortie du  $m^{\text{ième}}$  neurone de la dernière couche), peut donc se formuler de la façon suivante:

$$s_m(\vec{x}, W, Z) = \sum_{j=0}^{N_c} z_{m,j} \left[ g \left( \sum_{i=0}^N w_{j,i} x_i \right) \right] \quad (\text{II. 11})$$

où  $W$  est la matrice regroupant les poids des connexions entre la couche cachée et l'entrée du PMC et  $Z$  celle des connexions entre la couche de sortie et la couche cachée définies respectivement par:

$$W = \begin{pmatrix} w_{1,1} & \cdots & w_{1,i} & \cdots & w_{1,N} \\ \vdots & \ddots & & & \vdots \\ w_{j,1} & & w_{j,i} & & w_{j,N} \\ \vdots & & & \ddots & \vdots \\ w_{N_c,1} & \cdots & w_{N_c,i} & \cdots & w_{N_c,N} \end{pmatrix} \quad Z = \begin{pmatrix} z_{1,1} & \cdots & z_{1,j} & \cdots & z_{1,N_c} \\ \vdots & \ddots & & & \vdots \\ z_{m,1} & & z_{m,j} & & z_{m,N_c} \\ \vdots & & & \ddots & \vdots \\ z_{N_s,1} & \cdots & z_{N_s,j} & \cdots & z_{N_s,N_c} \end{pmatrix}$$

L'équation II.11 indique que  $s_m$  est une fonction linéaire des paramètres  $Z$  mais elle est aussi fonction non linéaire des paramètres  $W$ . On peut ainsi énoncer qu'un RN est un véritable outil permettant une modélisation non linéaire. C'est sur la base de cette constatation que sont énoncées les 2 principales propriétés remarquables des PMC à une couche cachée :

La première, la propriété d'approximation universelle, a été démontrée par Cybenko et Funuhashi [CYB89][FUN89]. Celle-ci peut être énoncée comme suit :

*« Toute fonction bornée suffisamment régulière peut être approchée uniformément, avec une précision arbitraire, dans un domaine fini de l'espace de ses variables, par un réseau de neurones comportant une couche de neurones cachés en nombre fini, possédant tous la même fonction d'activation, et un neurone de sortie linéaire. »*

La seconde propriété importante à citer est celle de la parcimonie. En effet, il est important de pouvoir modéliser un processus physique avec un nombre minimal de variables ajustables. Ainsi, plusieurs études [BAR93], [HOR94] ont montré que :

*Si le résultat de l'approximation (c'est-à-dire la sortie du RN) est une fonction non linéaire des paramètres ajustables, elle est plus parcimonieuse que si elle est une fonction linéaire de ces paramètres (comme par exemple une approximation polynomiale). De plus, pour des réseaux de neurones à fonction d'activation sigmoïdale, l'erreur commise dans l'approximation varie comme l'inverse du nombre de neurones cachés, et elle est indépendante du nombre de variables de la fonction à approcher. Par conséquent, pour une*

précision donnée, donc pour un nombre donné de neurones cachés, le nombre de paramètres du réseau est proportionnel au nombre de variables de la fonction à approcher. La parcimonie est d'autant plus importante que le nombre d'entrées du modèle est grand.

Ainsi, les architectures des PMC utilisés par la suite respecteront ces contraintes. Nous montrerons plus tard dans le manuscrit (cf. II.3.1) que le choix du nombre de neurones cachés sera déterminé en fonction de la complexité du problème étudié.

### II.1.6.3 Modélisation d'un processus expérimental par PMC

Supposons que l'on dispose de données expérimentales issues d'un processus physique que l'on cherche à modéliser. Ces informations sont accessibles sous la forme de  $n_t$  observations de  $N_s$  mesures distinctes  $\vec{y}^q = [y_1^q, y_2^q, \dots, y_m^q, \dots, y_{N_s}^q]^T$  ( $q=1, \dots, n_t$ ) supposées liées à  $N$  paramètres physiques  $\vec{x}^q = [x_1^q, x_2^q, \dots, x_i^q, \dots, x_N^q]^T$ . La modélisation consiste à approcher la relation univoque liant  $\vec{y}$  à  $\vec{x}$  sur la base des  $n_t$  observations. Une fois le modèle construit, il pourra être efficacement utilisé afin de prédire la réponse  $\vec{x}^{test}$  correspondant à toute nouvelle observation  $\vec{y}^{test}$ . Il faut noter qu'il s'agit ici de résoudre un problème inverse comme cela a été évoqué dans le chapitre précédent par une méthode de régression.

Il existe dans la littérature plusieurs modèles de régression pour effectuer cette tâche. Cependant, ils sont généralement limités aux cas simples de fonctions linéaires comme dans le cas de la PLS (Partial Least Square method) [HAA88].

L'avantage d'une régression par PMC réside dans le fait qu'ils utilisent, comme nous l'avons montré dans le paragraphe précédent, une modélisation non linéaire par rapport aux paramètres ajustables, ce qui constitue la base de leur propriété d'approximateur universel. De ce fait, les RN ont déjà montré qu'ils étaient indiqués pour modéliser efficacement tout processus physique non linéaire [BAD01] à condition de disposer d'un nombre  $n_t$  suffisamment grand de couples d'observations  $\vec{y}^q / \vec{x}^q$ .

Dans le cadre de la scattérométrie, les couples  $\vec{y}^q / \vec{x}^q$  à disposition sont formés de l'ensemble des  $n_t$  simulations {signatures ellipsométriques / valeurs de paramètres géométriques}. Cet ensemble est nommé le corpus d'apprentissage. Le RN sera utilisé pour modéliser la relation liant une signature ellipsométrique fournie en entrée aux paramètres géométriques de la structure définissant ainsi ses propres sorties. (Figure II-6).

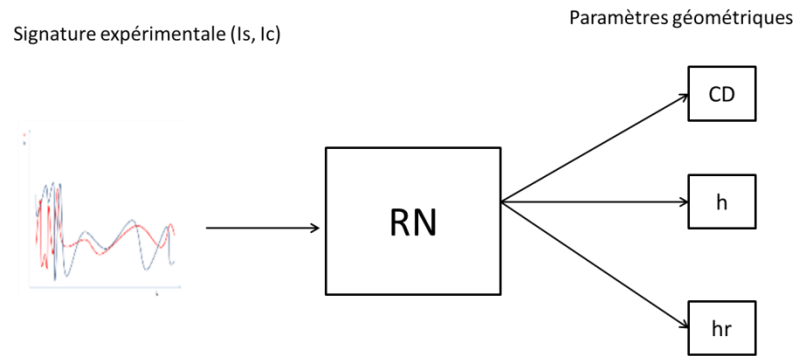


Figure II-6. Exemple d'utilisation d'un RN dans le cadre de la scattérométrie permettant de caractériser en sortie un profil défini par 3 paramètres (CD, h et hr) Le nombre d'entrée est directement relié au nombre de mesures constituant la signature ellipsométrique (typiquement 52 entrées)

D'un point de vue mathématique, les sorties  $\vec{s} = [s_1, s_2, \dots, s_m, \dots, s_{N_s}]^T$  du RN définies en II.11 représentent après apprentissage une bonne approximation de l'espérance mathématique de  $\vec{x}$  connaissant  $\vec{y}$ .

Après avoir défini ce que sont les RN, il convient maintenant de s'intéresser à leur principe de fonctionnement.

## II.2 Fonctionnement des réseaux de neurones

Comme nous l'avons vu précédemment, le RN nécessite une phase d'apprentissage où il apprend la relation liant des entrées à des sorties connues. Durant cette phase, les paramètres W et Z du RN sont ajustés de façon à permettre une généralisation optimale. Cette phase fait appel à des algorithmes d'optimisation et permet d'estimer, à la fin du processus, les performances globales du RN.

Cependant, avant de détailler la phase d'apprentissage, il faut préciser que chacun de nos entraînements a été précédé d'une phase de pré-traitement qui consiste à normaliser les données du corpus d'apprentissage. En effet, lors de la construction du modèle, on peut se trouver face à une situation où les entrées peuvent être de nature et d'ordre de grandeur différents. Il est alors essentiel de réaliser un traitement pour qu'elles aient toutes la même influence sur la construction du modèle. En pratique, il est généralement recommandé de pré-traiter systématiquement l'ensemble de ces données.

Le traitement appliqué par la suite consiste à centrer et réduire toutes les variables d'entrée et de sortie du RN. Par exemple, les nouvelles entrées du RN seront représentées par le vecteur  $\vec{x}'^q = [(x'_1)^q, (x'_2)^q, \dots, (x'_i)^q, \dots, (x'_N)^q]$  pour lequel :

$$(x'_i)^q = \frac{x_i^q - \mu_i}{\sigma_i} \quad (\text{II. 12})$$

avec  $\mu_i$  et  $\sigma_i$  respectivement la moyenne et l'écart type calculés sur l'ensemble des données du corpus d'apprentissage pour chaque entrée  $i$ .

Les sorties sont également normalisées selon le même principe. Lors de la phase de traitement, tous les résultats provenant de la sortie du RN devront donc être dénormalisés avant toute interprétation physique.

### II.2.1 Phase d'apprentissage

Il existe deux types d'apprentissage pour les RN: supervisé et non supervisé. Dans l'apprentissage supervisé, l'utilisateur fournit les valeurs désirées à la sortie du RN pour une série d'entrées connues. Pour cela, on doit disposer d'un comportement de référence précis (corpus d'apprentissage) dans lequel des sorties sont associées aux entrées connues. C'est ce type d'apprentissage qui sera adopté dans la suite de ce travail de thèse. Dans l'apprentissage non supervisé, on ne dispose pas d'information à priori sur les sorties du RN. Ce type d'apprentissage ne sera pas détaillé par la suite car il concerne des applications particulières comme par exemple le data-mining.

Dans le cas de la régression, l'objectif de la phase d'apprentissage est d'apprendre la relation liant le vecteur  $\vec{y}$  au vecteur  $\vec{x}$  du RN. Pour cela, on dispose d'un corpus d'apprentissage composé de  $n_t$  couples d'exemples  $\{\vec{y}^q, \vec{x}^q\}$  ( $q = 1, 2, \dots, n_t$ ). L'apprentissage revient à trouver les poids  $W$  et  $Z$  qui vont permettre au RN de donner en sortie une bonne approximation du vecteur  $\vec{x}$  lorsqu'on lui présente en entrée l'observation  $\vec{y}$ . Pour cela, on définit la fonction coût notée  $C(W, Z)$  représentant l'erreur quadratique entre les sorties  $\vec{s}$  calculées et les valeurs désirées  $\vec{y}$  sur l'ensemble des  $n_t$  données utilisées pour l'apprentissage :

$$C(W, Z) = \sum_{q=1}^{n_t} \left[ \sum_{m=1}^{Ns} (y_m^q - s_m^q(\vec{x}, W, Z))^2 \right] \quad (\text{II. 13})$$

Il s'agit donc d'un problème de minimisation pouvant être résolu par les différentes méthodes d'optimisation présentées dans le chapitre I. Rappelons qu'au cours de tout algorithme d'optimisation, les solutions, ici les poids, vont évoluer de manière itérative de

façon à réduire l'erreur de prédiction jusqu' à l'obtention d'une valeur optimale de  $C(W, Z)$ . Ainsi, le poids  $w_{j,i}$  liant le neurone  $j$  à l'entrée  $i$  sera modifié à l'itération  $k+1$  comme suit:

$$w_{j,i}^{k+1} = w_{j,i}^k + \Delta w_{j,i} \quad (\text{II. 14})$$

tel que  $\Delta w_{j,i}$  constitue la modification apportée qui dépend bien entendu de la méthode d'optimisation retenue.

On distingue ainsi deux types de fonctionnement: un fonctionnement dit adaptatif dans lequel les poids sont modifiés après le passage de chaque couple du corpus d'entraînement et un fonctionnement dit non adaptatif (ou batch training en anglais) dans lequel la mise à jour des poids est effective après le passage de l'ensemble des  $n_t$  couples d'entraînement. C'est ce dernier type d'entraînement qui sera utilisé par la suite car il est non seulement plus rapide que l'entraînement adaptatif mais surtout plus adapté à une grande quantité de données d'entraînement comme cela est le cas dans ce travail de thèse.

Dans le cas où on utilise les méthodes d'optimisation classiques, il est nécessaire de connaître à chaque itération les dérivées premières et secondes de la fonction coût par rapport aux paramètres ajustables, les poids du RN. Or, si cela ne pose aucune difficulté apparente pour les poids  $Z$  relatifs à la couche de sortie, cela n'est pas le cas pour les poids  $W$  car la sortie des neurones cachés n'est pas connue pour un couple d'observations donné. L'algorithme de rétropropagation du gradient [RUM86] permet de répondre à cette problématique en calculant de manière récursive le gradient de la fonction coût relatif aux paramètres des couches cachées, c'est-à-dire  $\partial C / \partial w_{j,i}$ . Il est ainsi possible d'utiliser n'importe quelle méthode d'optimisation classique développée dans le chapitre I pour cette phase d'apprentissage. Dans notre cas, l'algorithme Levenberg-Marquardt a été privilégié pour les avantages présentés dans le chapitre précédent.

Un des problèmes les plus fréquemment rencontré lors de cette phase est le surapprentissage, nommé aussi surajustement (Figure II-7). En effet, ce terme désigne un entraînement conduisant à un RN incapable de généraliser sur de nouvelles entrées car il a appris « par cœur » les couples d'apprentissage. La Figure II-7 montre par exemple les résultats d'un RN (courbe rouge) conduisant à un apprentissage parfait des points d'entraînement (en noir) présentant une erreur de prédiction extrêmement faible. Ce RN est pourtant incapable de fournir un résultat correct pour de nouvelles entrées fournies lors d'une utilisation ultérieure (en vert). Pour éviter cela, plusieurs solutions existent: On compte parmi



celles-ci, celle qui supprime les modèles dont la matrice jacobienne ( $\partial C/\partial w_{j,i}$ ) n'est pas de rang plein [SAA93] (c'est à dire une matrice carré dont toutes les colonnes et les lignes sont linéairement indépendantes) et la méthode locale LOCL (Local Overfitting Control via Laverages) [MON99] qui permet d'estimer l'influence de chaque exemple sur le modèle.

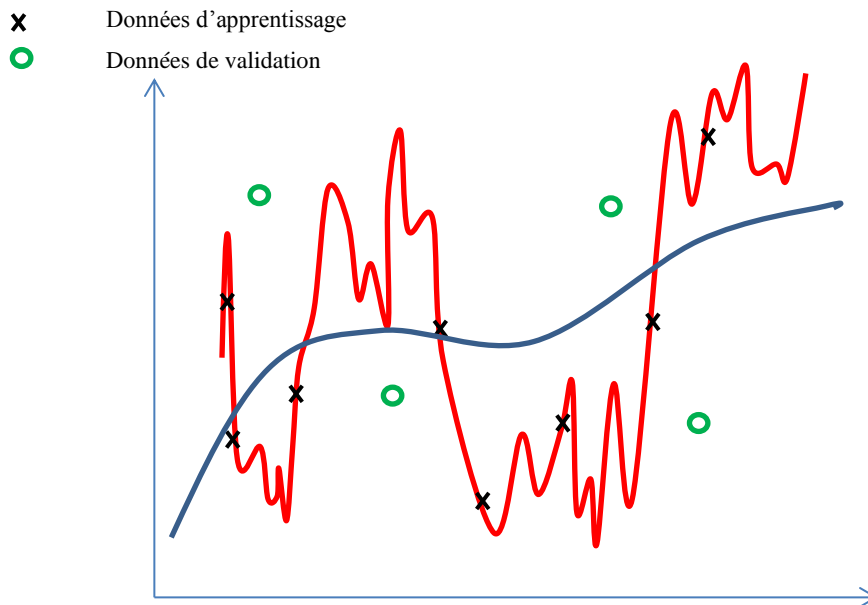


Figure II-7. Schéma illustratif représentant les résultats d'un RN suite à 2 apprentissages différents. Les données composant le corpus d'entraînement sont schématisées par les points noirs. Les points verts constituent des données qui permettent de tester les performances du RN. La courbe bleue représente le modèle d'un RN présentant une généralisation optimale et la courbe rouge un RN dans le cas d'un surapprentissage.

D'autres méthodes comme la validation croisée [STO74] consistent à estimer les performances du RN à partir d'exemples n'ayant pas servi pendant l'apprentissage (Figure II-8). Dans ce cas, le corpus d'entraînement est réparti en  $D$  parties de taille égale ( $P_1, P_2, \dots, P_D$ ), dont une partie est réservée systématiquement aux données de validation (case en bleu sur la figure).  $D$  apprentissages successifs sont alors réalisés par le même RN, en initialisant à chaque fois les poids du RN à des valeurs différentes. L'erreur commise par chacun des  $D$  RN entraînés est alors calculée sur la fraction de validation. La performance globale du modèle étudié, intitulée score de validation croisée, est calculée en réalisant la moyenne des  $D$  erreurs. La validation croisée permet donc de déterminer parmi différentes architectures celle qui possède des performances optimales en généralisation en choisissant systématiquement celle présentant le plus faible score de validation.

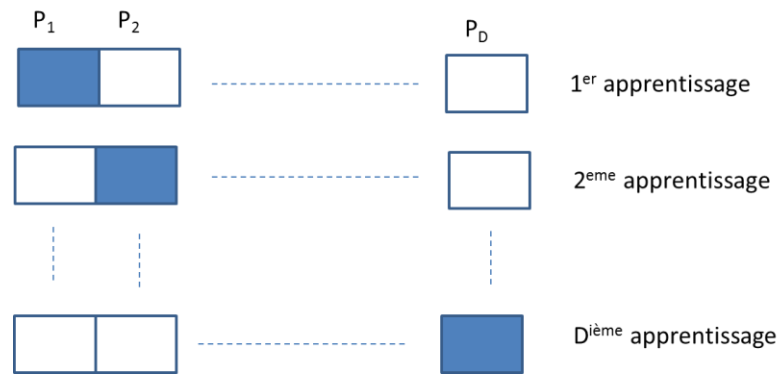


Figure II-8. Schéma illustratif de la validation croisée nécessitant  $D$  apprentissages successifs. Les parties bleues sont utilisées pour la validation des performances après chaque apprentissage

Il existe également une autre méthode, assez populaire: la méthode de l'arrêt prématuré, ou early stopping [POG85]. Le principe est simple. On arrête l'apprentissage lorsque les performances en généralisation sont estimées optimales en évitant ainsi le surapprentissage.

Pour cela, le corpus d'entraînement est réparti en trois sous-groupes: le corpus d'apprentissage (points noirs sur la courbe II.7), le corpus de validation (points verts sur la courbe II.7) et le corpus de test. Le premier permet d'effectuer l'apprentissage proprement dit en utilisant la procédure décrite précédemment. Le second corpus est complètement indépendant du premier et permet le calcul de l'erreur réalisée par le RN à chaque itération sur des données « nouvelles ». L'entraînement sera stoppé lorsque cette estimation de l'erreur cessera de diminuer. Ainsi, la capacité du RN en généralisation sera supposée optimale pour l'architecture en question (courbe bleue sur la Figure II-7). Le troisième corpus permet d'estimer des performances globales, en utilisant des données qui n'ont pas été utilisées, ni pour la phase apprentissage proprement dite, ni pour la validation, évitant ainsi tout biais de mesure.

Toute estimation d'erreur réalisée par la  $m^{\text{ième}}$  sortie du RN est basée sur le calcul suivant :

$$E_m^{\text{corpus}} = \sqrt{\frac{1}{n_{\text{corpus}}} \sum_{q=1}^{n_{\text{corpus}}} (y_m^q - s_m^q)^2} \quad (\text{II. 15})$$

ou  $y_m^q$  correspond à la  $m^{\text{ième}}$  sortie désirée,  $s_m^q$  la  $m^{\text{ième}}$  sortie calculée par le RN et  $n_{\text{corpus}}$  représente suivant les cas le nombre de données d'apprentissage (napp), de validation

(val), ou de test (test). (Par exemple, l'erreur sur les données de test pour la  $m^{\text{ième}}$  sortie sera nommé  $E_m^{\text{test}}$ )

Enfin, on peut également estimer les performances globales sur un corpus donné en calculant l'erreur moyenne sur l'ensemble des  $N_s$  sorties :

$$E^{\text{corpus}} = \sqrt{\frac{1}{n_{\text{corpus}}} \sum_{q=1}^{n_{\text{corpus}}} \left[ \frac{1}{N_s} \sum_{m=1}^{N_s} (y_m^q - s_m^q)^2 \right]} \quad (\text{II. 16})$$

L'erreur quadratique  $E^{\text{test}}$  est généralement employée pour comparer plusieurs RN entre eux. Elle estime la performance globale du modèle en termes de généralisation sur des données non biaisées.

Pour comprendre le fonctionnement de la méthode de l'arrêt prématuré, reprenons l'exemple présenté dans la Figure II-7 qui illustre de manière schématique deux apprentissages réalisés par RN. Le modèle représenté par la courbe bleue représente une régression correspondant à une généralisation maximale. Les points d'entraînement et de validation sont à des distances comparables de la courbe bleue. Dans ce cas  $E^{\text{app}}$  et  $E^{\text{val}}$  sont proches. La courbe rouge représente le cas d'un surapprentissage où la régression cherche à passer par tous les points d'apprentissage et s'écarte par conséquent des points de validation. Dans ce cas la valeur de  $E^{\text{app}}$  est faible mais la valeur de  $E^{\text{val}}$  est élevée.

Résumons donc le processus d'arrêt prématuré sur la Figure II-9 qui représente les erreurs quadratiques au cours de l'entraînement. On peut remarquer que lors des premières itérations, les erreurs  $E^{\text{app}}$  et  $E^{\text{val}}$  tendent à diminuer, jusqu'à ce que  $E^{\text{val}}$  atteigne un minimum. C'est à ce moment que l'apprentissage est arrêté et que la généralisation du RN est supposée optimale.

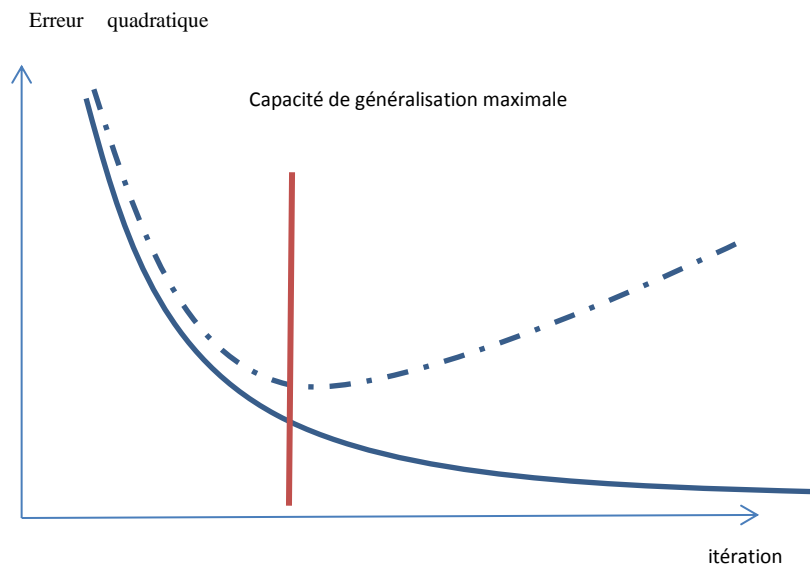


Figure II-9. Evolution schématique des erreurs sur les corpus de données lors d'un apprentissage par « early stopping ».  $E^{val}$  (trait en pointillé) représente les erreurs sur les données de validation et  $E^{app}$  (trait plein) sur les données d'apprentissage. L'apprentissage est arrêté lorsque  $E^{val}$  atteint un minimum.

Si l'apprentissage se poursuit, le RN entre dans une phase de surapprentissage et cherche à approcher parfaitement tous les points d'apprentissage. Dans ce cas l'erreur  $E^{val}$  augmente et  $E^{app}$  diminue fortement.

Cette technique a été retenue dans ce travail car elle est simple à mettre en œuvre. Cependant, elle a tendance à se satisfaire de RN surdimensionnés inutilement. Par la suite, la répartition des données suivante a été retenue pour le early stopping: 70% du corpus d'entraînement pour l'apprentissage, 15% pour la validation et 15% pour les test.

## II.2.2 Estimation de l'intervalle de confiance

Un moyen pour estimer les performances du RN par le calcul de l'erreur quadratique moyenne sur l'ensemble des valeurs de test a été défini. Cette technique permet de mesurer l'erreur de façon globale sur l'ensemble du corpus. Pour avoir une information plus locale, il est nécessaire de définir une barre d'erreur ou un intervalle de confiance propre à chaque sortie.

Il existe une technique permettant d'estimer l'intervalle de confiance à l'aide des RN. La méthode consiste à effectuer un apprentissage de la variance des sorties du RN sur les données d'entraînement [BAD01].

Pour cela, reprenons le processus défini au paragraphe II.1.6.3 que l'on cherche à modéliser à l'aide du PMC. Ce processus consiste à déterminer  $x$  (paramètres géométriques) en connaissant  $y$  (signature expérimentale). On dispose donc de  $n_t$  couples d'observations  $\{\overline{y^q}|\overline{x^q}\}$  ( $q=1,2,\dots,n_t$ ). Chaque vecteur est composé respectivement de  $N_s$  et  $N$  composantes.

La première étape consiste à effectuer un entraînement classique des données par un premier PMC pour apprendre la relation entre  $\overline{y^q}$  et  $\overline{x^q}$  tel que cela a été décrit précédemment. On a énoncé précédemment que chaque sortie du réseau  $s_m$  ( $m=1, 2, \dots, N_s$ ) représente une bonne approximation de l'espérance mathématique de  $\vec{x}$  sachant  $\vec{y}$ , notée  $E(\vec{x}|\vec{y})$ .

La seconde étape consiste à calculer les erreurs quadratiques  $e_m$  réalisées par le PMC sur chacune des sorties  $m$ , et, ce pour tous les échantillons de la base d'apprentissage.

$$e_m^q = (y_m^q - E(y_m^q/\overline{x^q}))^2 = (y_m^q - s_m^q)^2 \quad (\text{II. 17})$$

Un second RN (Figure II-10) sera ainsi utilisé pour apprendre la relation liant les erreurs quadratiques  $\overline{e^q}=[e_1^q, e_2^q, \dots, e_m^q, \dots, e_{N_s}^q]^T$  réalisées par le premier RN aux entrées  $\overline{x^q}=[x_1^q, x_2^q, \dots, x_m^q, \dots, x_N^q]^T$ .

Les sorties  $s'_m$  de ce RN seront assimilées à l'espérance mathématique de  $(y_m^q/\overline{x^q} - E(y_m^q/\overline{x^q}))^2$  sachant  $\overline{y^q}$  donc à une estimation de la variance de  $y_m^q/\overline{x^q}$

$$E(e_m^q/\overline{x^q}) = E((y_m^q/\overline{x^q} - E(y_m^q/\overline{x^q}))^2/\overline{x^q}) = V(y_m^q/\overline{x^q}) = s'_m \quad (\text{II. 18})$$

La fonction d'activation des neurones de sortie de ce second RN doit donc être telle qu'elle ne donne que des valeurs strictement positives. L'apprentissage se fera donc à l'aide des  $n_t$  couples  $\{\overline{y^q}|\overline{e^q}\}$  pour ( $q=1,2,\dots,n_t$ ) fournis par le premier PMC.

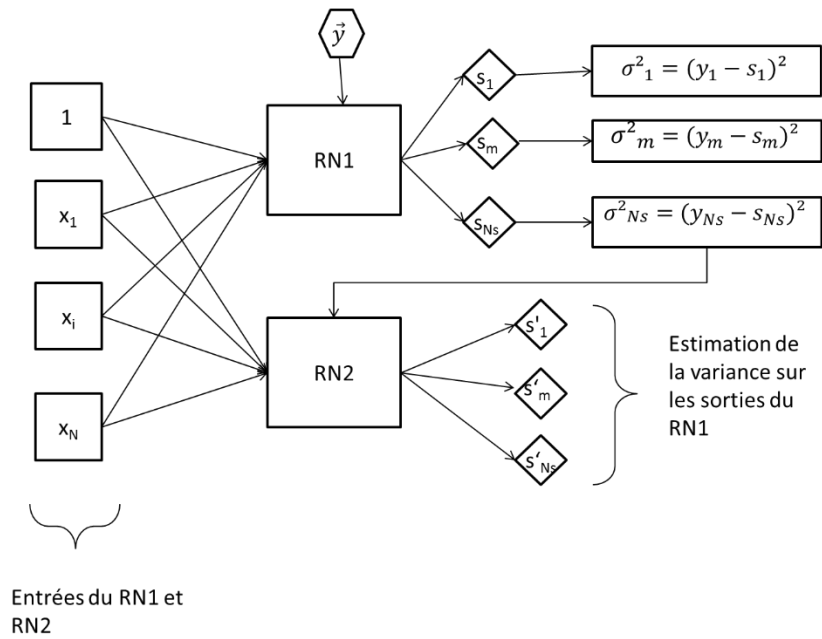


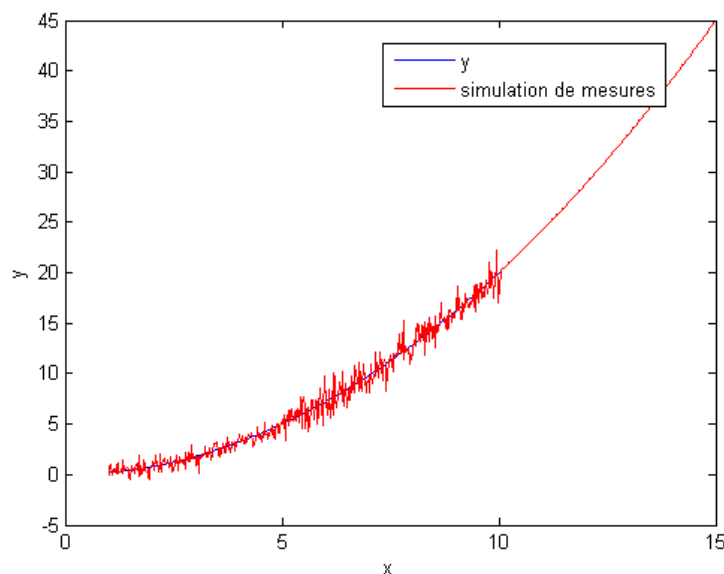
Figure II-10. Architecture à 2 PMC permettant de donner à la fois les résultats de la caractérisation (RN1) mais également une estimation de l'erreur réalisée par celui-ci (RN2)

A partir de la connaissance de la variance  $V(\vec{y}/\vec{x})$ , on est capable de déterminer les barres d'erreurs pour chaque sortie du premier RN à condition de se placer dans le cadre de l'hypothèse d'une distribution Gaussienne. Nous représenterons dans la suite des résultats à  $1\sigma$  (correspondant à une probabilité d'appartenance de 68%).

Pour illustrer cette technique, nous allons considérer un problème théorique simple mais avec une approche similaire au problème inverse rencontré par la suite en scattérométrie. Ainsi, des mesures sont simulées avec différents niveaux de bruit et l'impact occasionné sur la précision des résultats sera analysé.

Soit la fonction simple  $y(x) = \frac{x^2}{5}$  qui représente le processus physique à modéliser dans le domaine  $1 < x < 15$ . (Figure II-11). Différents relevés sont réalisés dans le domaine de variation avec des niveaux de bruit variables pour construire un corpus d'apprentissage :

- un bruit gaussien d'écart type 0.5 pour  $1 < x < 5$
- un bruit gaussien d'écart type 1 pour  $5 < x < 10$
- sans bruit pour  $10 < x < 15$



**Figure II-11.** Fonction à modéliser  $f(x)$  sans bruit (bleu) et présentant différents niveaux de bruit (en rouge): un bruit gaussien d'écart type 0.5 pour  $1 < x < 5$ , un bruit gaussien d'écart type 1 pour  $5 < x < 10$ , sans bruit pour  $10 < x < 15$

Un premier RN (RN1) est entraîné afin d'estimer la valeur  $y$  à partir du corpus d'apprentissage ainsi formé. Puis un second RN (RN2) est entraîné avec le mode opératoire explicité plus haut pour l'estimation des barres d'erreur associées à chaque point  $x$  de mesure entrée dans le RN1.

Les RN1 et RN2 ont été entraînés sur 701 échantillons avec respectivement 15 et 20 neurones dans la couche cachée. Les performances  $E^{\text{test}}$  obtenues sont meilleures pour le RN1 ( $E^{\text{test}}=0.53$ ) que pour le RN2 ( $E^{\text{test}}=0,76$ ). La différence peut être expliquée par le fait que la fonction à approcher par le RN2 est beaucoup plus complexe que pour le RN1. C'est d'ailleurs pour cela que l'architecture de ce dernier a été augmentée.

Les résultats obtenus par le RN1 agrémentés des barres d'erreurs estimées par le RN2 sont affichés sur la Figure II-12. On peut noter dans un premier temps que le RN1 donne des valeurs qui sont assez proches de la fonction initiale. Ceci montre que le RN résiste assez bien au bruit introduit. Ensuite on peut remarquer que les barres d'erreurs évaluées par RN2 englobent les valeurs désirées dans 60% des cas (correspondant à l'écart type à  $1\sigma$  de la gaussienne). Ces dernières sont, bien entendu, plus ou moins importantes suivant le niveau de bruit présent sur les données  $x$  fournies au RN1.

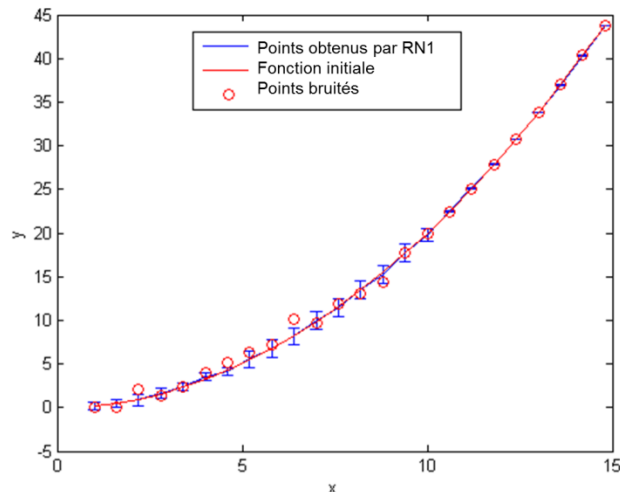


Figure II-12. Résultats du RN1 avec barres d'erreur à  $\sigma$  obtenus par le RN2

Pour visualiser la fonction à approcher par le RN2, on peut également tracer sur une autre figure les sorties désirées (correspondant aux erreurs réalisées par le RN1) (équation II.18) avec les sorties estimées par le second RN (Figure II-13).

Le RN2 est ainsi capable de détecter les trois niveaux de bruit initialement introduit sur les données. Il représente donc une bonne estimation de l'erreur réalisée par le premier RN et pourra ainsi être utilisé efficacement dans ce but.

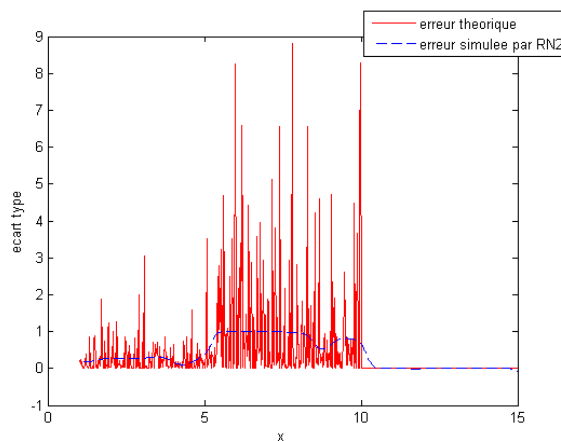


Figure II-13. Résultat du RN2 : Sorties désirées (en rouge) et estimées (en bleu)

### II.3 Un réseau de neurones pour la scattérométrie

Afin de mieux saisir le fonctionnement général d'un RN, nous avons procédé à deux études distinctes. La première est une étude théorique qui vise à identifier l'architecture optimale d'un RN dans le cadre précis de la scattérométrie. La seconde traite un cas



expérimental de caractérisation par RN en comparant les résultats obtenus avec d'autres techniques microscopiques couramment utilisées dans le domaine.

### II.3.1 Choix de l'architecture optimale

Pour utiliser au mieux un RN, il est nécessaire de mettre en place l'architecture la plus adaptée au problème traité. Nous allons pour cela évaluer l'influence de chacun de ses paramètres. Les études menées par la suite ont pour seul but d'appréhender le fonctionnement du RN. Nous nous positionnerons dans un cadre proche des problématiques visées dans cette thèse.

L'étude scattérométrique qui nous servira de support d'illustration tout au long de ce travail correspond à la caractérisation théorique d'une structure diffractante en résine présentant une couche résiduelle déposée sur un substrat de Si parfaitement connu. La résine utilisée dénommée NEB22 est une résine commercialisée par la société Sumitomo Chemicals pour la lithographie électronique, mais elle est également utilisée pour des procédés de lithographie par exposition à des UV profonds (Deep UV) ou de nanoimpression à chaud. L'évolution de son indice optique en fonction de l'énergie (ou de la longueur d'onde) est également connue. Le motif périodique (de période 500 nm) est supposé de forme trapézoïdale, défini par 4 paramètres géométriques ( $CD$ ,  $b_1$ ,  $h$ ,  $h_r$ ) (Figure II-14).

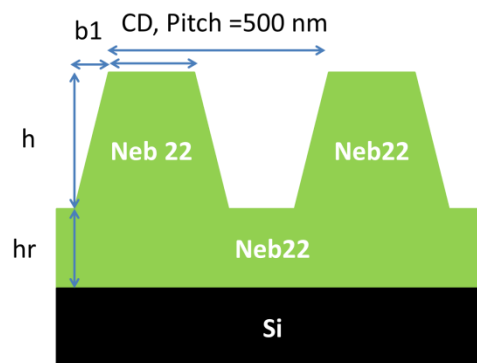


Figure II-14. Profil géométrique trapézoïdal avec couche résiduelle défini par 4 paramètres ( $CD$ ,  $b_1$ ,  $h$ ,  $h_r$ )

Nous considérons un spectre ellipsométrique obtenu sur la gamme [1.5 eV- 4 eV] avec un pas de 0.1 eV. L'angle d'incidence utilisé pour les simulations de mesures est de  $60^\circ$ . Ces valeurs ont été choisies de manière à s'approcher le plus possible du cas expérimental.

Deux plages de variations P1 et P2 (Table II-1), considérant l'ensemble des paramètres, ont été utilisées au cours de l'étude de telle façon que la plage P1 soit plus réduite que la plage P2 tout en englobant la même valeur centrale pour chacun des 4 paramètres.

P1 (nm)	P2 (nm)
$0 < b_1 < 80$	$0 < b_1 < 100$
$200 < b_2 < 300$	$180 < b_2 < 320$
$150 < h < 220$	$140 < h < 240$
$50 < h_r < 100$	$30 < h_r < 120$

Table II-1. Plages de variation utilisées sur les paramètres géométriques pour l'apprentissage du RN

Ces deux plages seront utilisées notamment pour l'étude de l'influence de la largeur du domaine d'étude sur les performances du RN.

### II.3.1.1 Influence du nombre d'échantillons et du nombre de neurones par couche cachée

L'influence du nombre d'échantillons et du nombre de neurones cachés sur les performances globales du RN a été étudiée sur la plage P2. Cette plage est la plus large et permet de mieux observer l'influence du nombre d'échantillons car la fonction à modéliser est plus complexe.

La Figure II-15 représente les erreurs  $E^{\text{test}}$  (II.16) en fonction du nombre d'échantillons  $n_t$  utilisés lors de l'apprentissage, et en fonction du nombre de neurones  $N_c$  composant la couche cachée du PMC.

On remarque que, plus le nombre d'échantillons augmente, plus l'erreur  $E^{\text{test}}$  diminue jusqu'à atteindre une valeur limite qui dépend du nombre de neurones composant la couche cachée. Quand l'architecture devient plus complexe, la convergence est atteinte plus tardivement. Ceci s'explique par le fait que, lorsqu'on augmente le nombre de neurones dans la couche cachée, l'architecture de la fonction générée par le PMC est plus complexe et nécessite plus d'échantillons pour une bonne représentation des données.

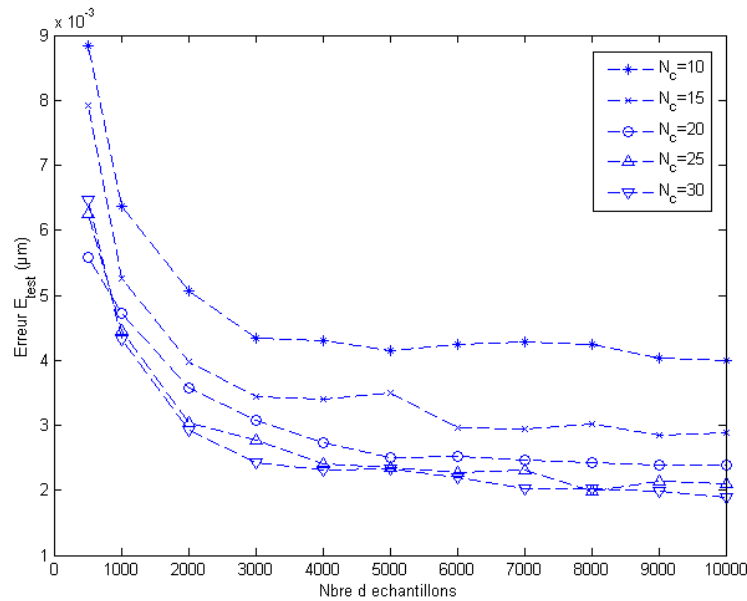


Figure II-15. Variation des performances du PMC à une couche cachée en fonction du nombre d'échantillons pour différentes neurones cachés.

Par ailleurs, on constate également que cette valeur limite est de plus en plus faible suivant le nombre de neurones cachés et converge vers une performance optimale atteinte pour une architecture de  $N_c=25$  neurones. On en déduit alors que le degré de complexité de la fonction générée par le PMC est suffisant pour représenter efficacement la fonction de régression recherchée dans le domaine P2.

L'architecture optimale sera dans ce cas un RN avec 25 neurones dans la couche cachée, entraîné sur 6000 échantillons. Ces valeurs ne sont pas critiques. Pour illustrer notre propos, prenons un exemple sur la courbe de la Figure II-15. Plaçons nous à  $N_e=8000$  échantillons, et comparons les erreurs des deux architectures  $N_c=15$  et  $N_c=20$ . On observe que la différence entre les deux est de  $\Delta E_{\text{test}}=0,6$  nm. On pourra donc dans ce cas considérer une architecture plus simple au prix d'un résultat moins précis, voire « acceptable ». Il est donc nécessaire de réaliser un compromis entre précision et temps de calcul. Dans cet exemple la précision du RN est de 2nm mais on ne peut pas généraliser cette valeur pour le RN. Dans le chapitre III, on évaluera de manière plus précise la précision du RN.

Ainsi, il est important de bien choisir le nombre de neurones et le nombre d'échantillons pour ne pas augmenter la complexité du réseau inutilement. Pour fixer les idées, une étude portant sur le temps nécessaire à la phase d'apprentissage est réalisée. Les résultats sont présentés sur la Figure II-16 pour différents nombres de neurones cachés, en fonction du nombre d'échantillons.

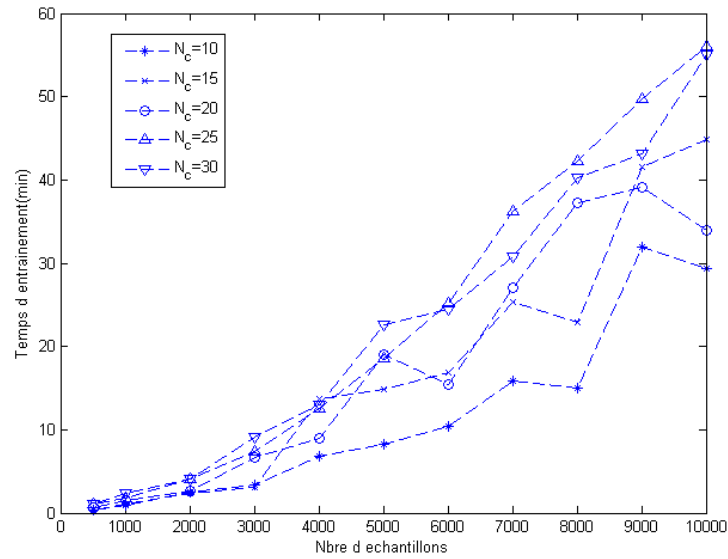


Figure II-16. Temps d'entraînement du RN en fonction du nombre d'échantillons et pour différents nombres de neurones .

Il est clairement visible sur la figure que le temps d'entraînement augmente en fonction du nombre d'échantillons et, dans un degré moindre, en fonction de la complexité de l'architecture. Pour être complet, il faut ajouter à ce temps d'entraînement, le temps de construction du corpus d'apprentissage, dépendant aussi du nombre d'échantillons. Tout cela doit guider l'utilisateur dans le choix d'une architecture assurant de bonnes performances avec des ressources en temps acceptables.

### II.3.1.2 Influence de la plage de variation

Considérons un PMC à une couche cachée entraîné sur 5000 échantillons pour 15 neurones cachés. L'étude précédente, réalisée sur la plage P2, a montré que cette architecture de PMC n'est pas optimale, mais elle permet d'obtenir des résultats satisfaisants avec des marges de temps acceptables. Cette architecture a donc été volontairement choisie pour étudier l'influence de la plage de variation. Les résultats obtenus respectivement sur les plages P1 et P2 sont regroupés dans la Table II-2 et représentent l'erreur globale sur les valeurs de test  $E^{test}$  (équation II.16) et l'erreur sur chaque paramètre sur les valeurs de tests  $E_m^{test}$  (équation II.15)

Plage de variation	P1	P2
$E^{\text{test}}$ (nm)	0,98	3,12
$E_{b1}^{\text{test}}$ (nm)	1,17	3,14
$E_{CD}^{\text{test}}$ (nm)	1,35	4,1
$E_h^{\text{test}}$ (nm)	0,63	2,68
$E_{hr}^{\text{test}}$ (nm)	0,46	2,25

Table II-2. Résultats obtenus sur les 4 paramètres géométriques ( $b$ ,  $CD$ ,  $h$ ,  $h_r$ ) en fonction de la plage de variation par un PMC d'une couche cachée avec 15 neurones entraîné sur 5000 échantillons pour un profil trapézoïdal symétrique avec couche résiduelle

Les performances obtenues sur la plage la plus étroite (P1) sont presque 3 fois meilleures en termes d'erreur quadratique globale pour une même architecture de PMC. Ce résultat est dû au fait que la complexité de la modélisation augmente avec la largeur de la plage et, qu'à nombre d'échantillons égal, on a une meilleure représentativité des données sur une petite plage.

Cependant, notons qu'en termes de valeur absolue, les résultats obtenus pour les deux plages sont satisfaisants avec une estimation de l'erreur sur les paramètres de l'ordre du nanomètre pour la plage 1 et de l'ordre de 3 nm pour la plage 2. On note également que l'erreur augmente d'un facteur 3 alors que la variation de la plage n'est pas très importante. Ceci s'explique par le fait que l'augmentation de la complexité de la fonction à approximer n'est pas linéaire, contrairement à l'augmentation de la plage qui est linéaire.

Il sera donc important dans le traitement d'un cas concret de réduire au mieux la plage de variation en fonction des informations connues sur l'échantillon. Néanmoins, la plage sélectionnée ne doit pas être trop réduite si l'on veut exploiter le même RN pour la caractérisation de plusieurs motifs de tailles différentes inclus dans la plage en question. Il peut ainsi être parfois plus judicieux de considérer une plage plus large, au risque d'avoir un coût plus élevé (architecture, temps de calcul) pour atteindre de bonnes performances.

On peut noter que ces résultats concernent un RN d'architecture constante. Il sera montré dans le chapitre III (cf. III.1.3.4) qu'en ajustant l'architecture du RN, il est possible

d'atteindre pour une large plage des performances semblables à celles d'une plage plus réduite.

L'étude menée dans ces deux derniers paragraphes a permis de mettre en évidence l'influence des paramètres du RN sur ses performances. Cette étude, bien que non exhaustive, servira de base pour l'optimisation des RN utilisés dans le chapitre III.

Intéressons-nous maintenant à la comparaison des performances du RN obtenues sur un cas réel avec d'autres algorithmes d'optimisation couramment employés en scattérométrie.

### II.3.2 Etude expérimentale

Considérons un échantillon composé d'une couche de résine IM5010 et d'une couche de BARC (Bottom Anti Reflective Coating) AR19G sur un substrat Si (Figure II-17). Le BARC est une couche intermédiaire qui permet de réduire la réflexion de la lumière lors de l'étape de lithographie. La période ( $d$ ) du motif est considérée comme fixe et égale à 140 nm. Le modèle considéré est un modèle créneau, défini par trois paramètres géométriques ( $CD$ ,  $h$  et  $h_r$ ). Cet échantillon a été fabriqué par la société ST Microelectronics par lithographie optique par immersion à 193 nm.

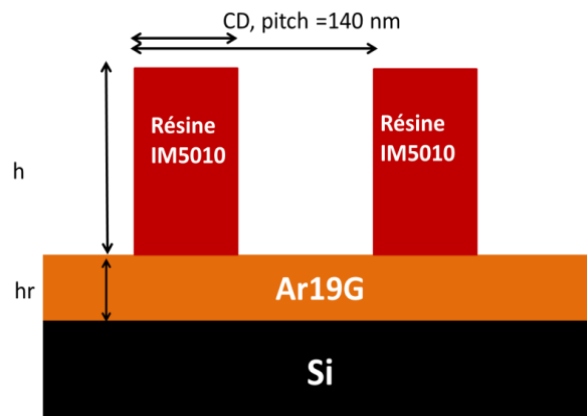


Figure II-17. Modèle créneau utilisé pour l'échantillon issu de lithographie optique par immersion défini par trois paramètres géométriques ( $CD$ ,  $h$ ,  $h_r$ )

La gamme spectrale utilisée pour cette étude est de 1,41 eV à 6,51 eV par pas de 0,2 eV correspondant à 52 entrées pour le RN. L'angle d'incidence est de  $62^\circ$ . L'ellipsomètre utilisé est un ellipsomètre Jobin Yvon à modulation de phase. L'apprentissage du RN est réalisé sur la plage de variation suivante :

$$20 \text{ nm} < CD < 90 \text{ nm}$$

$$80 \text{ nm} < h < 170 \text{ nm}$$

$$40 \text{ nm} < hr < 120 \text{ nm}$$

Pour le choix de la meilleure architecture, une étude préalable a été réalisée en se basant sur les considérations présentées précédemment dans le paragraphe II.3.1.

La Figure II-18 représente les performances du RN évaluées sur les valeurs du corpus test en fonction du nombre d'échantillons pour différents nombres de neurones. On remarque que les erreurs sont beaucoup plus faibles que dans le cas de la Figure II-15 et que la convergence des courbes, bien que similaire à la Figure II-15, est atteinte très rapidement et semble pas être affectée par l'augmentation du nombre de neurones. Ceci s'explique par le fait que les plages des paramètres sont beaucoup plus restreintes et que la fonction à modéliser devient plus simple et plus facilement modélisable avec une simple architecture. De plus le nombre de paramètres est moins important et conduit également à atteindre de bonnes performances avec peu de neurones.

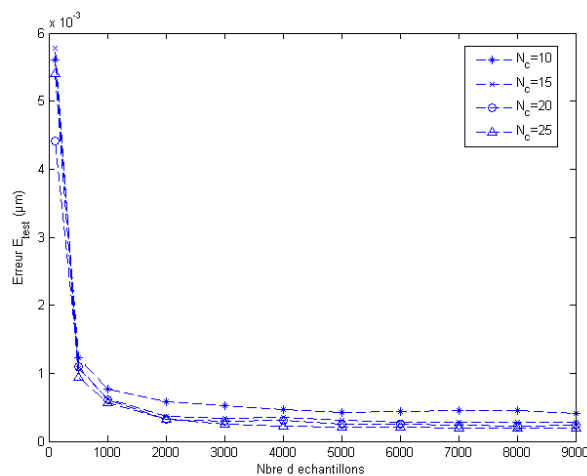


Figure II-18. Performances sur les valeurs de test en fonction du nombre d'échantillons pour différentes architectures testées ( $N_c=10, 15, 20, 25$ ). L'échantillon est composé d'un empilement de résine IM5010 sur du BARC sur Si avec une période  $d=140\text{nm}$ .

Avant de commencer l'étude du cas réel, précisons que la qualité des résultats fournis sera évaluée par le critère d'erreur quadratique moyenne sur les signatures reconstruites  $E_{Is, Ic}$ . Ce critère est obtenu à partir des signatures  $I_s, I_c$  reconstruites par problème direct à partir des paramètres géométriques comme le montre la Figure II-19 :

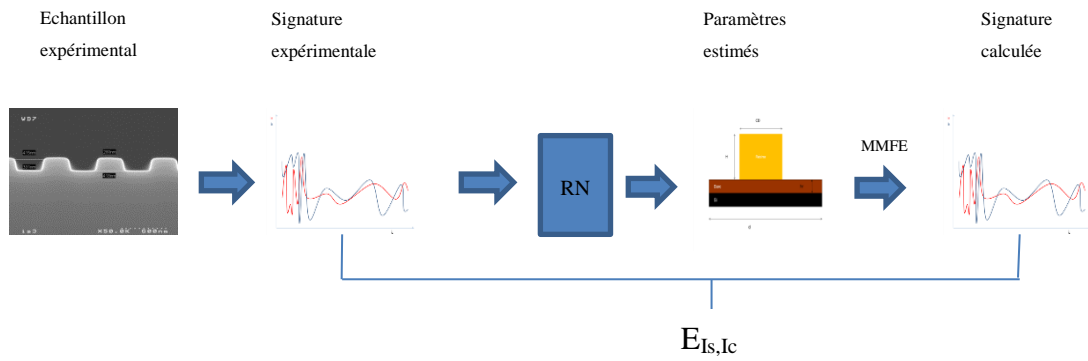


Figure II-19. Schéma récapitulatif expliquant le calcul de l'erreur sur les signatures reconstruites  $E_{Is,Ic}$  par problème direct à partir des données fournies par l'outil scattérométrique (RN dans la figure).

$$E_{Is,Ic} = \sqrt{\frac{1}{2N} \sum_{\lambda=1}^N ((I_{s_{exp}}(\lambda) - I_{s_{cal}}(\lambda))^2 + (I_{c_{exp}}(\lambda) - I_{c_{cal}}(\lambda))^2)} \quad (II. 19)$$

N est le nombre de longueurs d'ondes composant la signature.

Ce critère n'est pas propre au RN et il est généralement utilisé en scattérométrie pour apprécier les performances obtenues, voire même pour construire la fonction coût à minimiser pour les techniques utilisant des méthodes d'optimisation.

Les résultats obtenus par le RN sur l'échantillon présenté sont donnés dans la Table II.3 et la Figure II-20 et comparés avec d'autres méthodes scattérométriques basées sur des méthodes d'optimisation (cf Chapitre I.3): le krigeage et l'algorithme à région de confiance

	Krigeage	Région de confiance	RN
CD (nm)	53,5	53,1	57,4
H(nm)	118	116	116,6
hr(nm)	83,7	84,7	84,3
$E_{Is,Ic}$	0,106	0,105	0,14

Table II-3. Résultats de la caractérisation scattérométrique pour l'échantillon issu de lithographie optique par immersion obtenu sur les trois paramètres géométriques (CD h hr) par 3 méthodes différentes (Krigeage, Région de confiance, Réseau de Neurones)



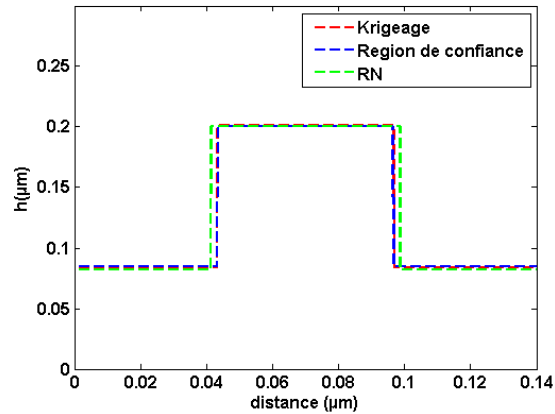
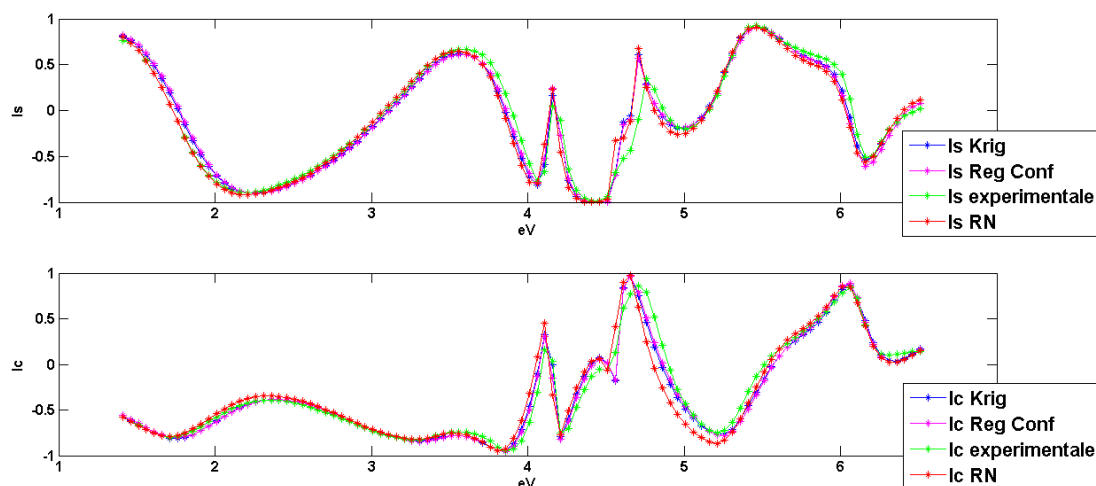


Figure II-20. Profils géométriques obtenus avec les algorithmes de Krigeage (rouge), Région de confiance (bleu) et RN (vert) dans le cas de la caractérisation scattérométrique de l'échantillon issu de lithographie optique par immersion

Les valeurs des paramètres estimées par les trois méthodes sont semblables sauf pour le paramètre CD, qui diffère légèrement pour le RN. Par ailleurs, les erreurs sur les signatures reconstruites sont du même ordre de grandeur quel que soit l'algorithme utilisé.

On remarque néanmoins une valeur plus importante d'erreur  $E_{I_s, I_c}$  pour les RN que pour les autres méthodes due à la valeur de CD qui est plus importante. Rappelons que les méthodes d'optimisation utilisées dans le cadre de la scattérométrie tendent justement à minimiser directement le critère  $E_{I_s, I_c}$  au détriment de toute autre considération, ce qui n'est pas le cas du RN. La valeur minimale de ce critère est donc forcément obtenue par une méthode d'optimisation. Dans tous les cas, nous verrons plus loin dans le manuscrit (cf III.1.2) que cet écart d'erreur rms ( $\Delta RMS = 0,04$ ) entre dans la marge d'erreur relative à un bruit de mesure.

Pour compléter cette étude, les signatures expérimentales et simulées ont été tracées à partir des résultats des 3 méthodes sur le spectre considéré (Figure II-21). On peut constater que les trois méthodes ont des signatures qui sont relativement proches de la signature expérimentale. Cependant, on peut observer que dans la région proche de 5eV, les signatures du RN (notamment  $I_c$ ) s'éloignent légèrement de la signature expérimentale expliquant l'écart relevé dans le tableau.



*Figure II-21. Comparaison des signatures ellipsométriques mesurées et calculées à partir des résultats obtenus respectivement par Krigeage, par l'algorithme à région de confiance et par RN dans le cas de la caractérisation scattérométrique de l'échantillon issu de lithographie optique par immersion.*

Ce résultat peut être expliqué par le fait qu'une petite variation dans les paramètres génère une grande variation au niveau des signatures dans la région de l'UV [GER09]. On a donc une meilleure sensibilité aux paramètres dans cette zone.

On en conclut néanmoins que les trois méthodes scattérométriques donnent des résultats sensiblement identiques et relativement corrects quant à la reconstruction de la signature mesurée.

#### **II.4 Comparaison avec les techniques morphologiques classiques**

Les deux techniques couramment utilisées en microélectronique (XSEM et CDAFM cf I.1) sont comparées aux performances de la scattérométrie neuronale dans ce paragraphe. Cette comparaison permettra de confronter des résultats provenant de méthodes fonctionnant sur des principes radicalement distincts. Les échantillons traités dans chaque cas sont obtenus par deux techniques différentes: nanoimpression pour l'un et lithographie par immersion à 193nm. Utiliser différents types d'échantillons, avec des techniques de caractérisation variées, tend à rendre plus robuste notre analyse

Avant de présenter les résultats, il convient de rappeler brièvement la démarche retenue pour chacune des méthodes. D'une part les méthodes morphologiques sont des techniques directes qui ne nécessitent pas d'informations a priori sur l'échantillon. La mesure est faite sur un nombre limité de lignes sur lesquelles une moyenne est effectuée. Le résultat obtenu est

donc une image très localisée de la morphologie. A l'inverse, la scattérométrie est une méthode indirecte dont le résultat découle d'un modèle paramétrique. Elle nécessite la connaissance d'une information à priori sur le modèle (forme géométrique, indice des matériaux). Par ailleurs, la mesure est effectuée sur une région importante, fixée par la taille du spot du faisceau utilisé (de l'ordre millimétrique pour les outils utilisés dans le cadre de ces travaux). L'information obtenue est donc une moyenne globale des différents motifs

Ces 2 techniques abordant la caractérisation par des fonctionnements relativement différents, il est difficile de choisir un critère de comparaison des résultats qui ne facilite pas plus l'une que l'autre. Nous avons opté pour un affichage en 2 temps: une démarche qualitative, reposant sur la superposition du motif scattérométrique avec l'image XSEM ou CDAFM. L'autre approche est quantitative et repose sur l'étude du critère rms.

#### II.4.1 Etude comparative avec le X-SEM

L'échantillon étudié dans ce paragraphe est fabriqué par nano-impression au LTM. Il est composé d'une couche de NEB22 sur Si. D'après l'image XSEM obtenue dans la Figure II-22, on peut en déduire que le profil est proche d'une forme trapézoïdale avec des bords arrondis sur le haut des lignes. La période du motif est de  $d=750$  nm.

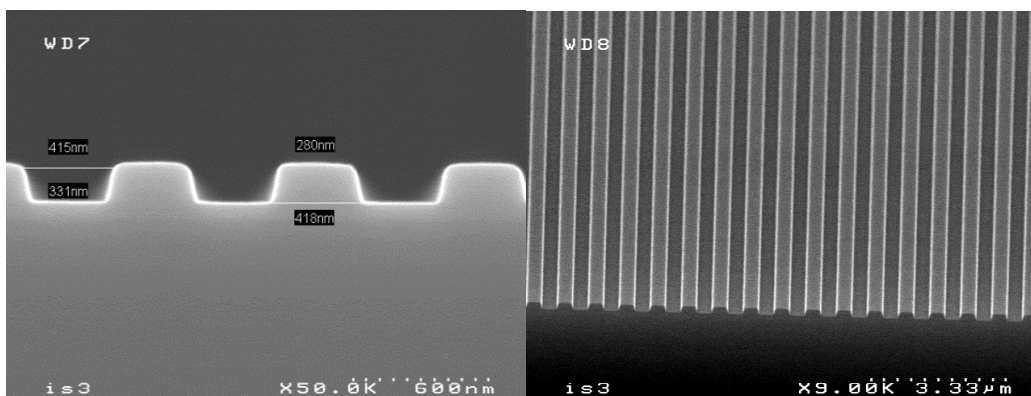


Figure II-22. Caractérisation par XSEM de l'échantillon Neb22 sur Si de période 750 nm obtenu par nano-impression en cross section (à gauche) et vue de dessus (à droite)

Pour résoudre le problème inverse par scattérométrie, un profil trapézoïdal avec 4 paramètres a donc été sélectionné ( $CD$ ,  $b_1$ ,  $h$  et  $h_p$ ) (Figure II-23). Nous avons volontairement opté pour un modèle simple pour commencer l'étude. Ce profil correspond d'ailleurs au modèle le plus couramment utilisé dans la littérature [KAL99, HAZ03, CHE13]. De plus, il a été montré dans [ELK10] qu'on est peu sensible aux arrondis en haut du motif en

scattérométrie. Une modélisation plus complexe pourra être envisagée si les résultats ne sont pas satisfaisants.

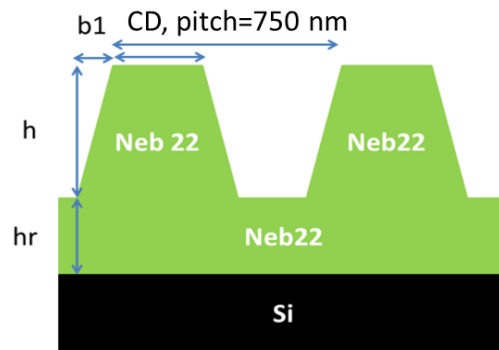


Figure II-23. Modèle trapézoïdal symétrique avec couche résiduelle défini par 4 paramètres géométriques (CD,  $b_1$ ,  $h$ , et  $h_r$ )

Le spectre variant de 1.5 eV à 6.5 eV et composé de 52 mesures a été de nouveau utilisé pour la mesure de la signature ellipsométrique. L'angle d'incidence est de  $70^\circ$ . L'ellipsomètre utilisé est l'ellipsomètre Jobin Yvon à modulation de phase.

L'entraînement du PMC (composé d'une seule couche cachée composée de 15 neurones) a été effectué avec 3000 échantillons. Les domaines de variation de chacun des paramètres utilisés pour l'apprentissage sont les suivants :

$$250 \text{ nm} < \text{CD} < 370 \text{ nm}$$

$$10 \text{ nm} < b_1 < 90 \text{ nm}$$

$$140 \text{ nm} < h < 200 \text{ nm}$$

$$40 \text{ nm} < h_r < 110 \text{ nm}$$

Les performances sur les valeurs des  $E^{\text{test}}$  sont données dans la Table II-4 et sont d'un ordre de grandeur acceptable par rapport au cas étudié ce qui met en valeur la qualité de l'entraînement.

Paramètre géométrique	$E^{\text{test}}$ pour chaque paramètre (nm)
CD	1,7
b1	1,3
h	0,67
hr	0,49
$E^{\text{test}}$	1,14

Table II-4. Performance du PMC (1 couche cachée de 15 neurones) après apprentissage (300 échantillons) évaluée sur les valeurs de  $E^{\text{test}}$

Comme énoncé plus haut, nous avons confronté les deux méthodes en redessinant sur l'image XSEM le profil correspondant aux paramètres obtenus par le RN (Figure II-24). On peut constater sur la figure que le profil obtenu par RN s'approche bien du motif du XSEM: les pentes ainsi que la largeur et la hauteur du motif sont bien conformes à la caractérisation microscopique. Cependant, le profil réel présente des arrondis qui n'ont pas été représentés par RN dû au choix d'un modèle simplifié sans arrondi. Par ailleurs, cette figure montre aussi certains inconvénients de la comparaison de profil par superposition de motifs. En effet, comme il a déjà été dit, l'analyse XSEM est forcément entachée d'erreurs. Ceci est particulièrement visible sur ce type d'image où la zone blanche sur les flancs des motifs est liée, soit à un focus imparfait, soit à des rugosités de bords de lignes. Il est donc assez difficile de définir clairement un profil à partir de cette image.

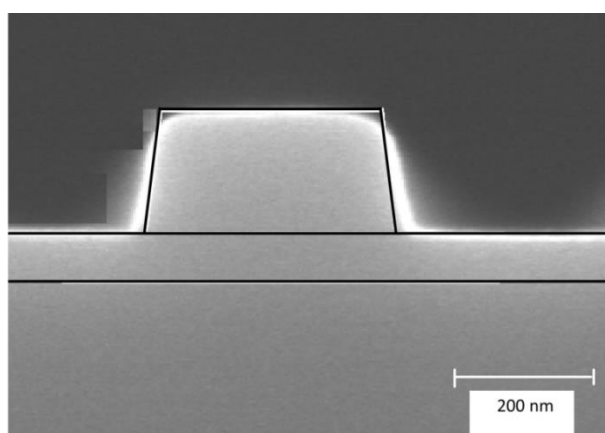


Figure II-24. Comparaison de profils géométriques obtenus pour la caractérisation XSEM et scattérométrique supposant une modélisation trapézoïdale de l'échantillon Neb22 sur Si de période 750 nm obtenu par nano-impression

L'autre méthode consiste à évaluer le critère de l'erreur rms sur la signature reconstruite [CHE13, HAZ03, QUI05] défini dans l'équation II.19. On analyse ainsi les résultats en reconstituant par le problème direct, en utilisant la MMFE, des signatures calculées à partir des résultats géométriques issus du RN ou de l'image XSEM et en les comparant à la signature expérimentale. Pour cela, la première phase consiste à extraire les paramètres de l'image XSEM selon le modèle défini par la Figure II-23. La Table II-5 représente les résultats des 2 méthodes sous la forme des valeurs des paramètres définissant la forme géométrique supposée. Les valeurs XSEM sont obtenues à partir de l'exploitation de l'image (cf. Chapitre I.I). On observe que les hauteurs  $h$  et  $h_r$  sont très proches, mais que les paramètres  $CD$  et  $b_1$  obtenus divergent de façon nette.

Paramètre géométrique	RN (nm)	XSEM (nm)
CD	332	284
$b_1$	26	50
$h$	178	174
$h_r$	75	75

Table II-5. Résultats de caractérisation scattérométrique (RN) et morphologique (XSEM) de l'échantillon Neb22 sur Si de période 750 nm obtenu par nano-impession en considérant un profil trapézoïdal symétrique avec couche résiduelle.

Dans le cadre de cette comparaison paramétrique du XSEM avec les RN, une extraction des paramètres du profil supposé a été effectuée à partir de l'image XSEM. Ces paramètres choisis deviennent les seules informations retenues concernant le profil XSEM alors que celui-ci en contient beaucoup plus. Cela entraîne donc une perte d'information. Ceci peut expliquer l'écart noté entre les paramètres des deux méthodes. Il est ainsi difficile d'extraire des paramètres de l'image XSEM comparables aux résultats scattérométriques.

L'analyse des signatures reconstruites à partir des valeurs paramétriques de la table II.4 est donnée dans la Figure II-25.

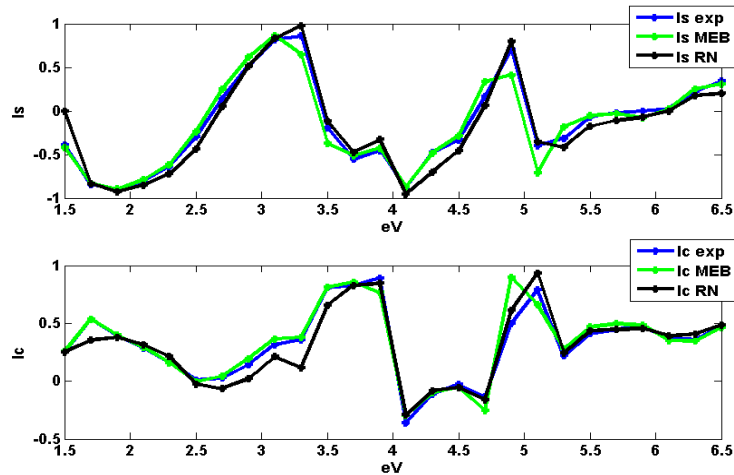


Figure II-25. Comparaison de la signature expérimentale avec les signatures reconstruites à partir des résultats de différentes méthodes de caractérisation (RN, MEB) dans le cas d'une modélisation trapézoïdale pour l'échantillon Neb22 sur Si de période 750 nm obtenu par nano-impression

Les erreurs sur les signatures reconstruites des deux méthodes sont semblables et sont respectivement de  $E_{Is,Ic}=0,107$  pour le RN et de  $E_{Is,Ic}=0,106$  pour le XSEM. Les signatures obtenues par les deux techniques s'approchent plus ou moins bien de la signature expérimentale. Pour le XSEM, la perte d'information constatée lors de l'extraction des paramètres est reflétée sur les signatures et peut expliquer en partie l'écart constaté à certaines énergies. Pour le RN, la divergence sur les paramètres  $CD$  et  $b_1$  introduite par le choix d'un profil incomplet est retranscrite également au niveau des signatures. L'inconvénient de cette méthode de comparaison réside dans le fait que les hypothèses de départ (forme du profil, indice des matériaux) ne sont pas toujours vérifiées et entraînent des erreurs par rapport à la signature expérimentale. Néanmoins, c'est une méthode quantitative qui permet d'apprécier de manière précise si le modèle choisi correspond au profil réel.

Pour évaluer l'importance de la modélisation de départ, on va désormais considérer un profil trapézoïdal qui prend en compte l'arrondi au sommet comme le suggère l'image XSEM (Figure II-22). Le modèle utilisé (Figure II-26) est un profil complet, ou générique, puisqu'il permet de modéliser les profils les plus classiquement rencontrés en scattérométrie à savoir : le profil créneau, le profil trapézoïdal, le profil créneau arrondi et le profil trapézoïdal à bords arrondis. Il est défini par 5 paramètres: aux paramètres  $h$  et  $h_r$  définis plus haut, on ajoute le paramètre  $CD_{top}$  qui désigne le plateau situé en haut du motif, le paramètre  $CD_{bottom}$  pour le bas du motif et le rayon  $r$  pour la courbure de l'arrondi.

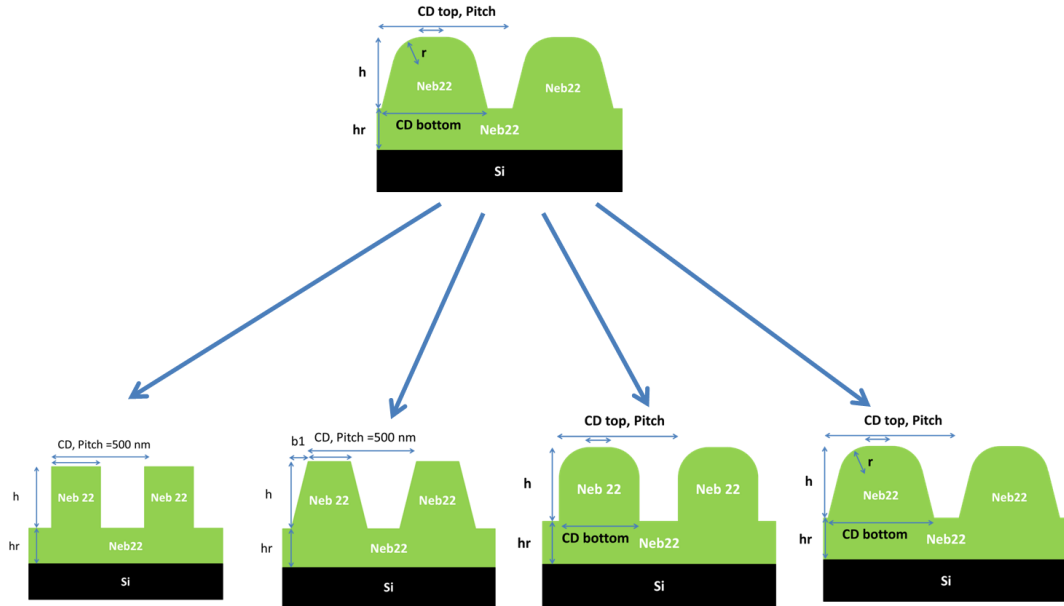


Figure II-26. Modèle trapézoïdal avec bords arrondis défini par 5 paramètres ( $CD_{top}$ ,  $CD_{bottom}$ ,  $h$ ,  $h_r$ ,  $r$ ) incluant les 4 profils couramment utilisés en scattérométrie (créneau, trapézoïdal, créneau arrondi, trapézoïdal à bord arrondi).

Ce modèle implique cependant une contrainte sur le paramètre  $r$  :

$$r < \frac{CD_{bottom} - CD_{top}}{\sqrt{2}} \quad (II. 20)$$

Ce modèle est employé pour analyser l'échantillon testé dans ce paragraphe. Un nouveau RN est mis en place: les meilleurs résultats sont obtenus avec un RN entraîné sur 6000 échantillons, présentant 20 neurones dans la couche cachée. Les plages de variation des paramètres sont respectivement :

$$0 \text{ nm} < CD_{top} < 270 \text{ nm}$$

$$330 \text{ nm} < CD_{bottom} < 410 \text{ nm}$$

$$150 \text{ nm} < h < 210 \text{ nm}$$

$$50 \text{ nm} < h_r < 90 \text{ nm}$$

Il faut noter que la plage du paramètre  $r$  est définie une fois que les valeurs de  $CD_{bottom}$  et de  $CD_{top}$  sont fixées :

$$0 < r < \frac{CD_{bottom} - CD_{top}}{\sqrt{2}} \quad (II. 21)$$



La Table II-6 représente à la fois les performances globales du RN après apprentissage en considérant la modélisation générique présentée ci-dessus ainsi que les résultats correspondant à l'échantillon.

Paramètre	Valeur obtenue par RN (nm)	$E^{\text{test}}$ pour chaque paramètre (nm)
$CD_{\text{top}}$	253	6,1
$CD_{\text{bottom}}$	385	1,7
r	52	5,0
h	178	0,8
hr	74	0,45

Table II-6. Résultats de caractérisation scattérométriques (RN) obtenus pour le profil trapézoïdal à bord arrondi avec couche résiduelle dans le cas de l'échantillon Neb22 sur Si de période 750 nm obtenu par nano-impession

Pour compléter ce tableau, précisons que les performances globales sur les valeurs de tests sont de  $E^{\text{test}}=3,6$  nm et l'erreur sur les signatures reconstruites est  $E_{\text{Is, Ic}}=0,068$ . Celle-ci est plus faible que dans le cas précédent avec un profil trapézoïdal simple. Cette information est confirmée par la juxtaposition des résultats de scattérométrie sur l'image XSEM (Figure II-27). Le profil générique corrobore davantage les informations fournies par l'image XSEM.

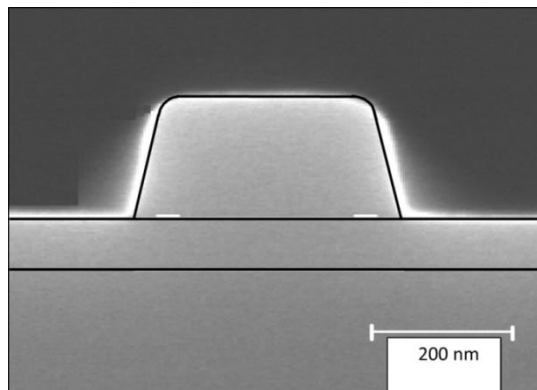


Figure II-27. Comparaison de profils géométriques obtenus pour la caractérisation XSEM et scattérométrie supposant une modélisation trapézoïdale avec bords arrondis de l'échantillon Neb22 sur Si de période 750 nm obtenu par nano-impession

On constate en comparant la table II-4 et la table II-5 que h et hr sont les mêmes et que  $CD_{\text{bottom}} = CD+2*b1$ . Ceci met en évidence le fait la méthode employée est robuste et que l'arrondi introduit dans le cas générique n'est pas un paramètre influant sur la signature (soit

un paramètre avec une faible sensibilité). En effet, le fait de négliger ce paramètre ne modifie pas les valeurs sur les paramètres hauteur et largeur.

A partir de ces résultats, l'intérêt de considérer un profil générique peut être discuté. On note qu'augmenter la complexité du modèle nécessite de mettre en place des ressources plus importantes en scattérométrie (Figure II-16). Celles-ci se traduisent, pour le RN, par une augmentation du nombre d'échantillons, du nombre de neurones cachés (Pour le profil trapèze, 3000 échantillons et 15 neurones et pour le profil générique, 6000 échantillons et 20 neurones). Cependant, dans le cas présent l'amélioration du profil par rapport au profil simple n'est pas flagrante si l'on considère seulement la valeur de l'erreur rms. Il paraît donc nécessaire de trouver un bon compromis entre l'augmentation de la complexité du problème et la précision souhaitée dans l'analyse. Selon les cas de figure, les technologues pourront se contenter de paramètres issus d'un profil plus simple qui ne modélise pas parfaitement la structure, ou auront besoin d'avoir accès à une analyse plus fine du profil.

#### II.4.2 Etude comparative avec le CD-AFM

Dans cette partie, les performances de la scattérométrie neuronale sont comparées avec les résultats obtenus par CDAFM (cf. I.1.2). L'échantillon est composé d'un réseau de lignes réalisées dans la résine IM5010, reposant sur du BARC, le tout sur un substrat de Si. Le composant a été fabriqué chez STMicroelectronics par lithographie optique par immersion à 193nm. D'après les spécificités données par le fabricant, le profil a une forme créneau (Figure II-28) et les valeurs cibles sont les suivantes:  $CD=50\text{nm}$ ,  $h=130\text{nm}$  et  $h_r=82\text{nm}$ . La période du motif est de 140nm. Ce paramètre est le seul fiable dans la mesure où il est défini sur le réticule et ne dépend pas des conditions du procédé lithographique.

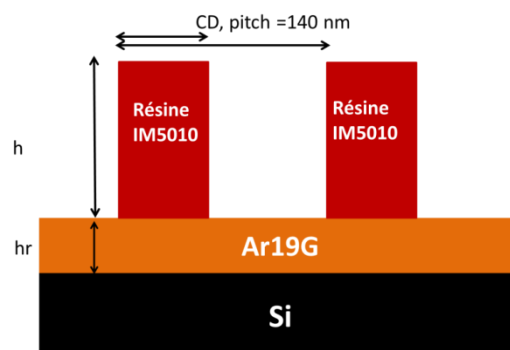


Figure II-28. Modèle créneau défini par 3 paramètres ( $CD$ ,  $h$  et  $h_r$ ) utilisé pour la caractérisation de l'échantillon Im5010 sur Si

Nous considérerons directement le profil générique défini plus haut (Figure II-29).

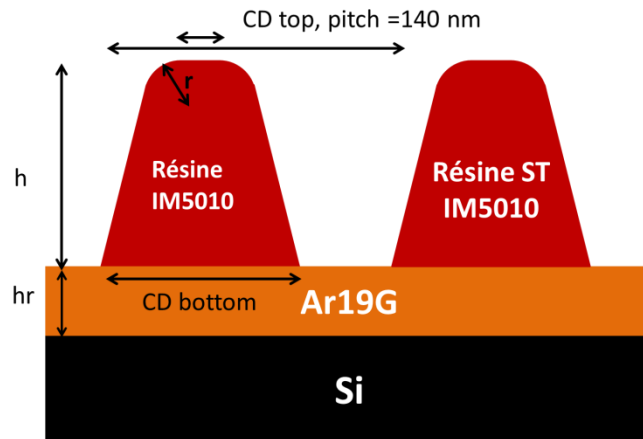


Figure II-29. Modèle trapézoïdal avec bords arrondis défini par 5 paramètres ( $CD_{top}$ ,  $CD_{bottom}$ ,  $r$ ,  $h$ ,  $hr$ ) utilisé dans le cadre de la caractérisation scattérométrique de l'échantillon Im5010 sur Si

Le RN utilise le même type de signature ellipsométrique que les cas précédents (Cf. II.4.1). L'angle d'incidence est de  $62^\circ$ . Le RN composé d'une couche cachée de 15 neurones est entraîné à l'aide de 6000 échantillons. Les plages de variations des paramètres utilisés lors de l'entraînement sont les suivantes:

$$0 \text{ nm} < CD_{top} < 20 \text{ nm}$$

$$30 \text{ nm} < CD_{bottom} < 70 \text{ nm}$$

$$80 \text{ nm} < h < 130 \text{ nm}$$

$$60 \text{ nm} < hr < 100 \text{ nm}$$

La plage du paramètre  $r$  est définie comme cela a été énoncé dans l'étude précédente. La Table II-7 représente à la fois les performances globales du RN après apprentissage ainsi que les résultats correspondant à l'échantillon testé.

Les performances globales sur les valeurs de test  $E^{\text{test}}$  sont de 3,11 nm et l'erreur  $E_{Is, Ic}$  sur la signature reconstruite est de 0,066. Les résultats obtenus par scattérométrie nous donnent une première idée sur l'allure du profil, qui présente un important rayon de courbure au sommet.

Paramètre	Valeur RN (nm)	$E^{\text{test}}$ pour chaque paramètre (nm)
$CD_{\text{top}}$	9,5	5,3
$CD_{\text{bottom}}$	59	1,2
r	107	4
h	23	1,6
hr	82	0,7

Table II-7. Performances et résultats de caractérisation par RN obtenus pour le profil trapézoïdal à bords arrondis pour la caractérisation de l'échantillon Im5010 sur Barc sur Si

Les erreurs pour l'ensemble des paramètres sont acceptables sauf pour le paramètre  $CD_{\text{top}}$  qui possède une erreur relative très grande (environ 50%). Afin de comprendre cette mauvaise performance pour le  $CD_{\text{top}}$ , les courbes de régression à l'issue de l'apprentissage ont été analysées (Figure II-30). Celles-ci permettent d'apprécier les performances de l'entraînement sur les valeurs de tests pour chaque paramètre de sortie du RN. L'abscisse représente les valeurs cibles pour chaque échantillon de test et l'ordonnée représente les valeurs calculées par le RN. Le coefficient de régression mesure la pente de la courbe de régression (courbe en bleu). Plus celui-ci s'approche de 1, plus l'entraînement est correct. Ceci signifie que les valeurs calculées par le RN sont exactes et correspondent bien aux valeurs cibles. A l'inverse, un coefficient de régression faible indique que l'entraînement a été mal effectué. Cela peut également permettre de détecter un défaut de modélisation du problème comme par exemple un choix inopportun d'une sortie peu dépendante des entrées du RN. Dans notre cas, il se trouve que seul le coefficient de régression pour le paramètre  $CD_{\text{top}}$  est faible. La courbe est presque horizontale, ce qui nous amène à penser que ce paramètre influe peu sur la signature retenue en entrée du RN.

Les paramètres  $CD_{\text{bottom}}$ , h et hr donnent quant à eux de bons résultats. Le paramètre r donne également, mais dans un degré moindre, des valeurs acceptables. Ceci peut être

expliqué d'une part par le fait que ce paramètre est lié aux autres par l'équation II.20. Le RN doit ainsi mémoriser cette contrainte, ce qui représente un degré de complexité supplémentaire pour l'apprentissage de ce paramètre. D'autre part, il faut noter qu'il s'agit d'un paramètre géométrique de second ordre intrinsèquement moins influant que d'autres paramètres comme l'épaisseur par exemple.

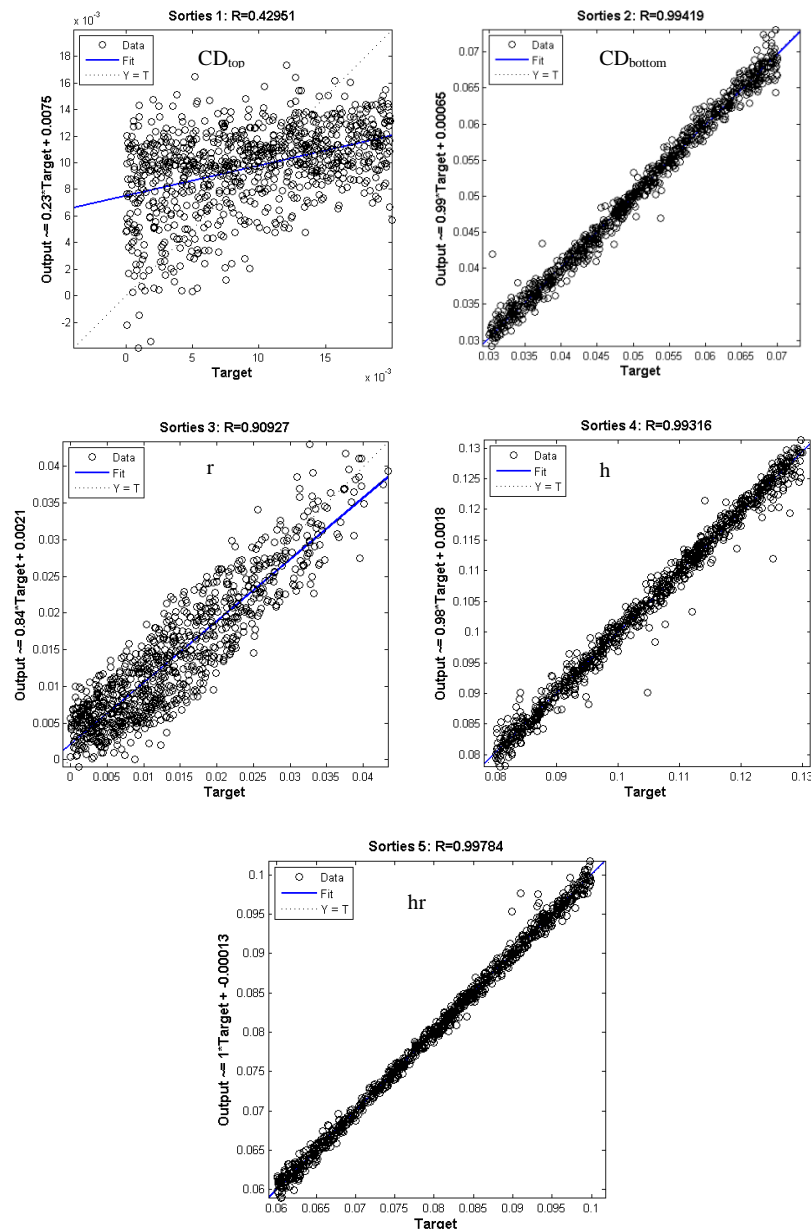


Figure II-30. Courbes de régression pour les différents paramètres du profil générique ( $CD_{top}$ ,  $CD_{bottom}$ ,  $r$ ,  $h$ ,  $hr$ ), calculées sur les valeurs du corpus de test en fin d'entraînement en considérant un PMC d'une seule cachée de 15 neurones entraîné sur 6000 échantillons

Afin de confirmer les remarques énoncées concernant le paramètre  $CD_{top}$ , une étude a été conduite pour évaluer la sensibilité de la signature utilisée au paramètre  $CD_{top}$ .

Pour cela, considérons 3 profils ayant en commun les paramètres  $CD_{bottom}$ ,  $h$  et  $hr$  (Table II-8) pour différentes valeurs de  $CD_{top}$  et  $r$  pris de façon à prendre en compte différentes formes géométriques.

Paramètre	Profil 1	Profil 2	Profil 3
$CD_{top}$ (nm)	0	10	20
$CD_{bottom}$ (nm)	40	40	40
$r$ (nm)	20	10	5
$h$ (nm)	90	90	90
$hr$ (nm)	70	70	70

*Table II-8. Simulation des 3 échantillons ayant en commun les paramètres  $CD_{bottom}$ ,  $h$  et  $hr$  pour différentes valeurs de  $CD_{top}$  et  $r$  utilisés pour étudier la sensibilité de la scattérométrie au paramètre  $CD_{top}$*

Les profils correspondant à ces trois cas de figure sont superposés afin de pouvoir apprécier les différences géométriques occasionnées (Figure II-31). Les cas extrêmes  $CD_{top}=0$  nm et  $CD_{top}=20$  nm montrent une différence notable et cohérente au sommet des motifs, traduisant la différence de valeur du paramètre  $r$ . En revanche on note des pentes des motifs très proches l'une de l'autre ce qui donne à ces deux géométries une bonne similarité. Le cas  $CD_{top}=10$  nm présente un profil au sommet du motif assez proche de celui observé pour  $CD_{top}=0$ nm mais avec une pente significativement différente. L'évaluation à partir de l'aperçu visuel des motifs, bien que présentant une part de subjectivité, permet donc de déceler les différences géométriques qui apparaissent de façons plus ou moins prononcées suivant les géométries.

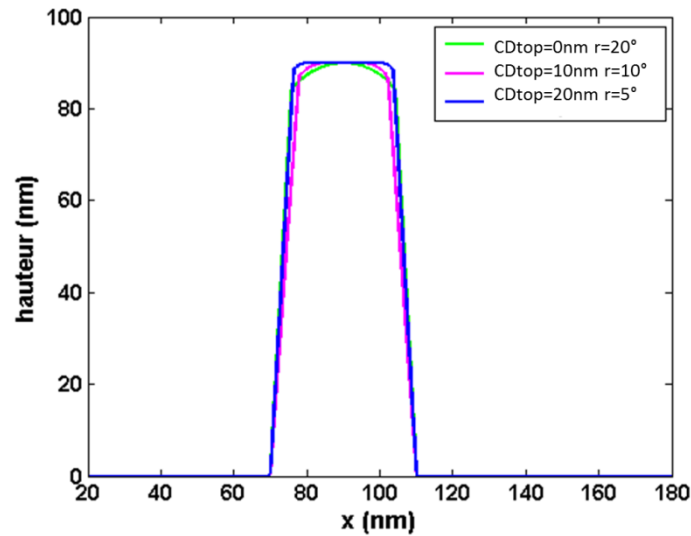


Figure II-31. Représentation géométrique des 3 échantillons ayant en commun les paramètres  $CD_{bottom}$ ,  $h$  et  $hr$  pour différentes valeurs de  $CD_{top}$  ( $CD_{top}=0$  nm,  $CD_{top}=10$  nm,  $CD_{top}=20$  nm) et  $r$  ( $r=20$  nm,  $r=10$  nm,  $r=5$  nm) pour  $h=90$  nm,  $CD_{bottom}=40$  nm et  $hr=70$  nm

Les signatures  $I_s$  et  $I_c$  pour les trois profils ont été calculées et tracées sur la Figure II-32. On observe trois signatures relativement proches alors que toute la plage de variation du paramètre  $CD_{top}$  est utilisée. Un léger écart entre le cas  $CD_{top}=10$  nm et les 2 autres ( $CD_{top}=0$  et 20 nm) est décelable sur  $I_s$  et sur  $I_c$ , notamment dans l'UV. Ce résultat montre que l'influence de la pente est plus importante que la forme au sommet du motif. Cela confirme également le fait que dans l'UV on a une meilleure sensibilité sur les paramètres. Une étude plus fine semble montrer que 2 échantillons qui sont proches en termes de signature optique, ne le sont pas forcément lorsqu'on regarde les représentations physiques des profils correspondants.

Cette étude montre que le RN et donc la scattérométrie est peu sensible au paramètre  $CD_{top}$ . Etant donné que la valeur de ce paramètre n'a pas d'influence notable sur les signatures, ce paramètre a été fixé à 0 dans la suite de l'étude.

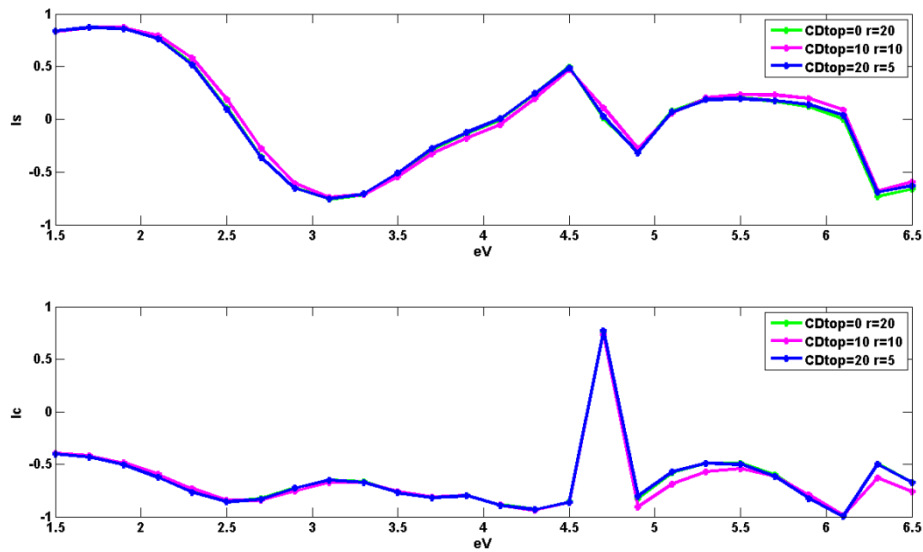


Figure II-32. Représentation des signatures ( $I_s, I_c$ ) pour les trois profils étudiés ( $CD_{top}=0$  nm /  $r=20$  nm,  $CD_{top}=10$  nm /  $r=10$  nm,  $CD_{top}=20$  nm /  $r=5$  nm pour  $h=90$  nm,  $CD_{bottom}=40$  nm et  $hr=70$  nm

La Table II-9 représente les résultats de caractérisation par RN dans le cas d'un réseau PMC entraîné sur 6000 échantillons avec 20 neurones cachés. A noter que pour cette étude le paramètre  $CD_{top}$  a été fixé à 0nm.

paramètre	Valeur RN (nm)	Erreur sur les tests (nm)
$CD_{bottom}$	59	0,4
r	29	0,3
h	109	0,67
hr	82	0,22

Table II-9. Performances et résultats de caractérisation par RN obtenus pour le profil trapézoïdal à bords arrondis ( $CD_{top}$  fixé à 0) pour la caractérisation de l'échantillon Im5010 sur Barc sur Si

La valeur d'erreur globale sur les valeurs de test est  $E^{test}=0,38$  et l'erreur sur les signatures reconstruites est  $E_{I_s, I_c}=0,060$ . Les performances sur les valeurs de test sont beaucoup plus réduites que dans le cas précédent. Ceci montre que l'erreur sur un paramètre influe sur les performances des autres paramètres. Il est donc essentiel de définir clairement les sorties du RN en accord avec la définition du problème à résoudre. L'erreur sur les



signatures avec le cas  $CD_{top}=0$  est sensiblement égale au cas précédent. Ceci confirme une fois de plus que ce paramètre n'a pas d'influence sur les signatures reconstruites.

Le profil correspondant à ces valeurs a été tracé sur la Figure II-33 et comparé à celui obtenu par CD-AFM.

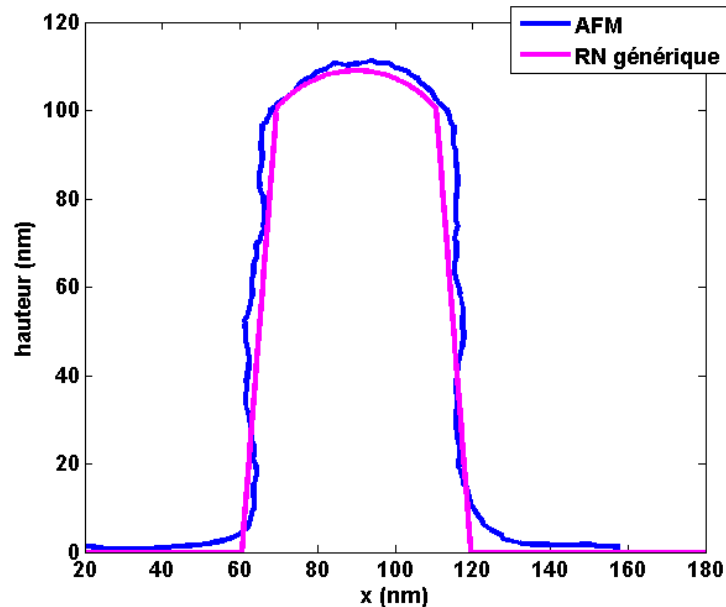


Figure II-33. Comparaison de profils géométriques obtenus par caractérisation AFM et scattérométrie (RN) supposant une modélisation générique dans le cas de l'échantillon Im5010 sur Barc sur Si.

Sur le profil fourni par CDAFM, on voit apparaître des bords arrondis et l'absence de plateau en haut du motif. L'hypothèse donnée par le constructeur, sur la forme du profil créneau, n'est donc a priori pas vérifiée si l'on considère la mesure par CDAFM comme une mesure de référence. Ceci peut être dû aux conditions du procédé lithographique.

Le profil obtenu par CDAFM traduit la présence de rugosité sur les flancs. Comme on l'a précisé en introduction du paragraphe II.4, les techniques morphologiques sont des techniques où l'information est moyennée sur une zone très localisée. L'outil de mesure est donc sensible aux rugosités présentes sur les flancs du motif. La scattérométrie, quant à elle, fait apparaître une légère pente des flancs qui n'est pas présente dans la caractérisation par CDAFM. Elle donne une information plus globale sur une surface plus grande. A l'inverse, l'inclinaison détectée par la scattérométrie peut être présente pour compenser la présence de rugosité sur les flancs. Le profil générique utilisé peut ainsi faire apparaître une pente sans que celle-ci soit aussi prononcée sur le profil réel. Ceci est dû au fait que la scattérométrie est

fondée sur la minimisation du critère rms et elle exploite tous les degrés de liberté disponibles du profil pour obtenir la valeur d'erreur la plus faible.

On peut aussi signaler la présence d'un arrondi au fond des lignes. Il est difficile donc de conclure sur l'exactitude des résultats donnés par AFM en ce qui concerne le fond des lignes car la forme de la pointe génère nécessairement ce type de morphologie en bas des motifs. Dans le cas étudié, cet arrondi n'a pas été pris en compte dans le modèle scattérométrique, mais il peut avoir un effet sur les résultats obtenus.

Néanmoins, on peut conclure que le RN utilisant le profil générique reproduit assez fidèlement les motifs. Par ailleurs, de par ses nombreux degrés de liberté, l'utilisation du profil générique doit être faite avec précaution.

Il est important de savoir si le profil générique est le mieux adapté en étudiant le cas de profils plus simples à disposition, à savoir les profils elliptique et arrondi définis sur la Figure II-34. Pour le profil elliptique, on considère un quart d'ellipse de part et d'autre du  $CD_{top}$  (le demi-petit axe correspondant ainsi à  $b = \frac{CD_{bottom} - CD_{top}}{2}$  et le demi-grand axe à  $h$ ). Pour le profil arrondi, on considère un quart de cercle de rayon  $r$  de part et d'autre du  $CD_{top}$  ( $r = \frac{CD_{bottom} - CD_{top}}{2}$ ).

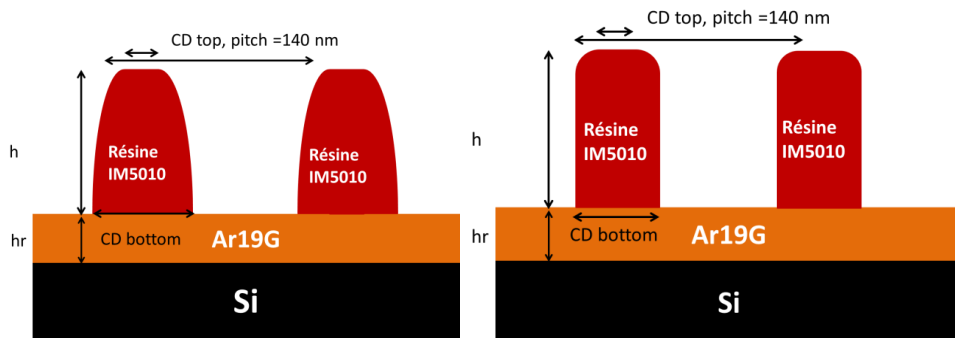


Figure II-34. Modèle de profil elliptique (gauche) et arrondi (droite) défini chacun par 4 paramètres ( $CD_{top}$ ,  $CD_{bottom}$ ,  $h$ ,  $h_r$ ) utilisé dans le cadre de la caractérisation scattérométrique de l'échantillon Im5010 sur BARC sur Si

Pour chacun des deux modèles de profil utilisés, un RN avec 15 neurones dans la couche cachée est entraîné sur 3000 échantillons. Les plages de variation utilisées lors de l'entraînement sont respectivement pour le profil elliptique et le profil arrondi :

$$0 \text{ nm} < CD_{top} < 20 \text{ nm}$$

$$CD_{top} = 0 \text{ nm}$$

30 nm < CD <sub>bottom</sub> < 70 nm	Et	30 nm < CD <sub>bottom</sub> < 70 nm
80 nm < h < 130 nm		80 nm < h < 130 nm
60 nm < hr < 100 nm		60 nm < hr < 100 nm

Pour le profil arrondi, la valeur de CD<sub>top</sub> a été fixée à 0, car l'étude précédente avec le profil générique a montré que le RN était peu sensible à ce paramètre. Pour le profil elliptique, l'étude des courbes de régression (non retranscrite ici pour ne pas surcharger le document) montre qu'on possède une sensibilité à ce paramètre et il sera donc considéré comme paramètre flottant.

Les résultats de caractérisation obtenus avec les deux profils sont présentés dans la Table II-10.

Paramètre	Profil elliptique (nm)	Profil arrondi (nm)
CD <sub>top</sub>	15	0
CD <sub>bottom</sub>	57	49
H	116	107
Hr	80	85

*Table II-10. Résultats de caractérisation par RN obtenus avec 2 profils différents (arrondi et elliptique) pour la caractérisation de l'échantillon Im5010 sur Barc sur Si*

Concernant les paramètres, les valeurs données par les deux profils sont cohérentes à l'exception du paramètre CD<sub>top</sub> qui diffère du fait que le profil considéré n'est pas le même. Les résultats soulignent la présence d'un arrondi, dû au fait que les valeurs de CD<sub>top</sub> et CD<sub>Bottom</sub> sont différentes pour un même profil. Les erreurs sur les signatures reconstruites sont respectivement E<sub>Is,Ic</sub>=0,058 pour le profil elliptique et E<sub>Is,Ic</sub>=0,089 pour le profil arrondi.

L'ensemble des résultats précédents obtenus en termes d'erreur sur les signatures reconstruites a été synthétisé dans la Figure II-35. Nous avons également ajouté ceux obtenus dans le cas d'un profil créneau et générique.

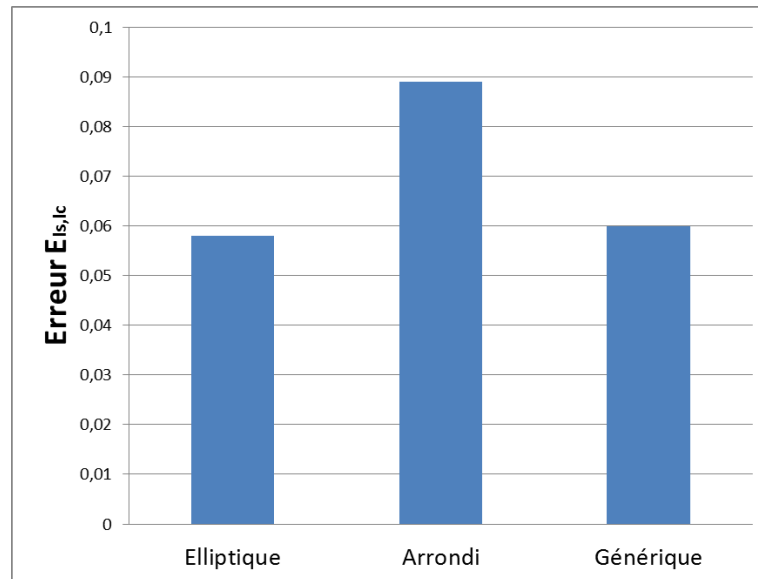


Figure II-35. Comparaison des résultats scattérométriques correspondant à différents profils géométriques supposés dans le cas de l'échantillon Im5010 sur BARC sur Si

Les résultats concernant le profil générique sont satisfaisants et représentent avec le profil elliptique les valeurs optimales.

Nous avons ensuite représenté sur une même figure l'ensemble des profils géométriques reconstruits par les 2 méthodes (Figure II-36).

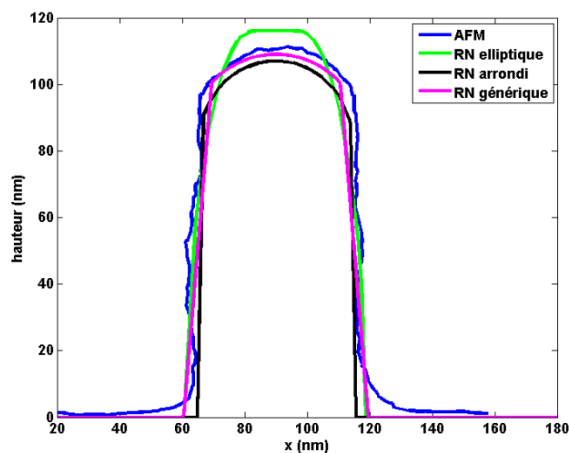


Figure II-36. Juxtaposition des résultats de caractérisations obtenues par CD-AFM et par scattérométrie à partir de différents profils géométriques supposés dans le cas de l'échantillon Im5010 sur BARC sur Si

On observe que le profil elliptique est le plus éloigné du profil obtenu par CD-AFM et que ce sont les profils générique et arrondi qui sont les plus proches. Pourtant, le profil elliptique donne l'erreur rms la plus faible. Ceci s'explique par le fait que la scattérométrie va compenser l'erreur du modèle sur les arrondis au sommet par une hauteur plus importante, qui

conduit à une erreur rms plus petite. Dans ce cas, une diminution de l'erreur rms conduit à s'écarter de la position géométrique optimale !

Il est donc nécessaire de prendre avec précaution le critère de l'erreur estimée sur les signatures reconstruites, car il peut être trompeur dans certains cas. Cette étude montre l'importance du choix initial du modèle : un mauvais choix de modèle peut, dans certains cas, conduire à une erreur rms plus faible, mais qui ne correspond pas au motif réel.

Concernant les autres résultats, il est difficile de trancher entre les deux profils générique et arrondi. En effet, le profil arrondi représente mieux la forme verticale des flancs, mais conduit à une hauteur de motif un peu trop faible comparativement au profil donné par le CDAFM.

En conclusion, il ne faut pas pour autant discréditer le critère rms qui se base sur une information réelle, la signature expérimentale. La mesure AFM possède elle aussi des sources d'erreur et d'incertitudes par rapport au motif qui peuvent contribuer à trouver une erreur rms plus grande pour certains motifs scattérométriques même si leur profil semble mieux correspondre. Chaque critère d'évaluation des performances, erreur rms ou la superposition des profils, possède ses limites. En complément de cette étude comparative avec le CDAFM, une étude sur le critère rms a été réalisée pour mieux cerner son comportement. Un profil se rapprochant au mieux du profil fourni par CDAFM a été choisi manuellement, en considérant un profil arrondi. (courbe noire sur la Figure II-37).

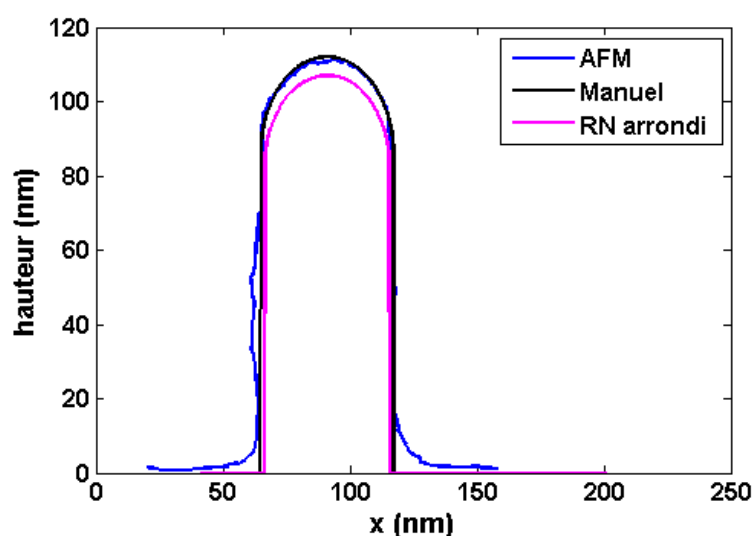


Figure II-37. Comparaison de différents profils arrondis (résultats du RN, modélisation optimale de la mesure AFM) avec le profil obtenu par CDAFM dans le cas de l'échantillon Im5010 sur BARC sur Si

Les valeurs de ce profil correspondant à une extraction paramétrique de la mesure AFM sont relevées dans la Table II-11 et comparées avec les résultats donnés par le RN.

Paramètres	Profil manuel (nm)	RN arrondi (nm)
$CD_{top}$	0	0
$CD_{bottom}$	52	49
H	112	107
Hr	85	85

Table II-11. Valeur paramétriques des différents profils arrondis (résultats du RN, modélisation optimale de la mesure AFM) discutés dans le cas de la caractérisation de l'échantillon Im5010 sur BARC sur Si

Contrairement à ce qu'on aurait pu attendre, l'erreur pour le profil choisi manuellement est plus élevée que pour le profil fourni par RN (table II-12). Le profil s'approchant le plus du profil obtenu par CDAFM possède l'erreur rms la plus élevée. Ceci confirme l'analyse déjà mentionnée ci-dessus concernant les sources d'erreurs imputables à la mesure CDAFM.

	Profil manuel (nm)	RN arrondi (nm)
$E_{Is,Ic}$	0,17	0,10

Tableau II-12. Erreur sur les signatures des différents profils arrondis (résultats du RN, modélisation optimale de la mesure AFM) discutés dans le cas de la caractérisation de l'échantillon Im5010 sur BARC sur Si

Nous pouvons conclure que les résultats obtenus par scattérométrie neuronale et par les techniques morphologiques sont cohérentes et en accord. Dans les 2 cas étudiés, deux méthodes pour évaluer les performances ont été utilisées: l'erreur sur les signatures reconstruites, et l'aperçu visuel lié à la superposition des profils. Chaque méthode possède ses avantages et ses limites et nous avons montré que les résultats doivent être analysés avec précaution. Cependant, les deux méthodes sont à considérer de manière complémentaire pour obtenir la précision optimale.

D'autre part, le choix du profil en scattérométrie présente une importance capitale. En effet, le profil générique utilisé permet d'avoir une sensibilité maximale (bord arrondi, pente). Il représente un atout pour la scattérométrie car il permet d'analyser des structures avec peu d'information a priori sur le motif. Néanmoins, il est nécessaire de prendre des précautions lors de son usage, car, en plus d'être coûteux en ressources (temps de calcul, architecture plus complexe) il peut parfois générer des degrés de liberté et induire des erreurs sur le motif. Un profil plus simple peut dans certains cas être plus approprié. Tout dépendra de la priorité à fixer par l'utilisateur. Celui-ci peut par exemple nécessiter une information simple sur le CD

pouvant être satisfaite par un profil simple. Il peut aussi avoir besoin d'un degré de précision maximal sur la structure et dans ce cas, l'emploi d'un modèle plus complexe sera privilégié. Dans le cas d'un profil simple, il faut quand même comprendre que ce dernier doit être suffisamment cohérent, et tenir compte notamment des paramètres principaux pour obtenir une valeur fiable du résultat. En effet avec la corrélation entre les paramètres, si le modèle simple choisi comporte des erreurs sur les paramètres fixes celles-ci vont se répercuter sur les paramètres flottants avec un risque élevé d'avoir des erreurs sur le résultat souhaité. Les simplifications ne doivent concerner que les paramètres secondaires. (L'étude sur la comparaison des profils trapèze ou générique (cf II.4.1) a clairement montré que le CD<sub>bottom</sub> et la hauteur du motif n'étaient pas impactés par le changement de la forme au sommet du motif).

Ainsi, les méthodes de caractérisation morphologique et scattérométrique apparaissent comme des techniques complémentaires. Une solution de caractérisation qui combine les deux méthodes permettra de tirer profit des avantages de deux techniques. Notons qu'il est cependant nécessaire pour la scattérométrie d'avoir un minimum d'information sur le profil (période, modélisation paramétrique, plage de variation des paramètres à déterminer). L'objectif pour la scattérométrie étant de tendre vers une solution efficace qui permette de s'affranchir des techniques morphologiques qui sont plus longues et destructives.

## **II.5 Conclusion**

Le RN est un outil mathématique permettant de prédire le fonctionnement d'un système physique à condition de disposer d'une base d'apprentissage conséquente décrivant le phénomène. Les différents types de RN ont été exposés rapidement en s'attardant particulièrement sur le PMC pour ses propriétés intéressantes qui sont exploitées en scattérométrie. La phase d'apprentissage fait appel à des algorithmes d'optimisation et nécessite une évaluation des performances.

Des études, théoriques et expérimentales, ont été menées pour appréhender le fonctionnement du RN. Il a été montré que la scattérométrie permet une caractérisation exacte et précise du motif, avec un nombre restreint d'informations sur la structure étudiée. La scattérométrie est donc un outil de métrologie efficace pouvant être déployé de façon presque

autonome pour la caractérisation de motifs complexes. Ces études en statique seront utilisées pour l'initialisation de la caractérisation des échantillons en temps réel par voie neuronale.



## II.6 BLIOGRAPHIE

- [AGA10] Agarwal M., Jain N., Kumar M. M., & Agrawal H “Face recognition using eigen faces and artificial neural network”. International Journal of Computer Theory and Engineering, vol. 2, no 4, p. 1793-8201, 2010.
- [AMM07] Ammar M. Y., « Mise en œuvre de réseaux de neurones pour la modélisation de cinétique réactionnelle en vue de la transposition Batch/continu », PhD thesis, Institut National Polytechnique de Toulouse , juillet 2007.
- [BAD01] Badran F., Thiria S., « les perceptrons multicouches: de la régression non-linéaire aux problèmes inverse ». J. Phys. IV, 92 :157-188,2002.
- [BAR93] Barron A., “Universal approximation bounds for superposition of a sigmoïdal function”, IEE Transactions on Information Theory, Vol 39, pp. 930-945, 1993.
- [BOU10] Boueri M., “Laser-induced plasma on polymeric materials and applications for the discrimination and identification of plastics”, PhD Thesis, Université de Lyon, octobre 2010.
- [CHE13] Chen X., Liu S., Zhang C., Zhu J., « Improved measurement accuracy in optical scatterometry using fitting error interpolation based library search ». Measurement 46, (octobre 2013): 2638-2646
- [CON00] Conn A.R., Gould N. I. M., Toint Ph. L., “*Trust-Region Methods*”.MPS-SIAM Series on Optimization 1. SIAM and MPS, Philadelphia, 2000.
- [CYB89] Cybenko G., “Approximation by superpositions of sigmoidal function”. Mathematics of control, Signals and Systems, 2:303-314,1989.
- [DAV90] Davalo E., Naim P., « Des réseaux de neurones », Editions Eyrolles, Paris, 1990.
- [DRE97] Dreyfus G., Personnaz L., Toulouse G., « Perceptrons, Past and Present », Enciclopedia Italiana, In press, 1997.
- [DRE02] Dreyfus G., Martinez J.M.,Samuelides M.,Gordon M.B.,Badran F., Thiria S. et Heralut L. , “Réseaux de neurones \_ Méthodologie et applications”, Eyrolles, Paris, 2002.
- [ENK05] Enke, D., & Thawornwong, S., “The use of data mining and neural networks for forecasting stock market returns”, Expert Systems with applications, vol 29, no.4, p.

- 927-940, 2005.
- [FOD11]** Foddis M. L., “Application of artificial neural networks in hydrogeology: identification of unknown pollution sources in contaminated aquifers”, PhD thesis, Université de Strasbourg, Mars 2011
- [FUN89]** Funuhashi K. “On the approximate realization of continuous mapping by neural networks”. *Neural Networks*, 2:183-192,1989.
- [GER09]** Gereige I., “ Contribution des réseaux de neurons dans le domaine de l’ellipsométrie. Application à la scattérométrie », PhD thesis, Université de Saint Etienne, 2009
- [GRO88]** Carpenter G., Grossberg S., «The ART of Adaptive Pattern Recognition by self-Organizing Neural Network », *Computer*, v.21 n.3, p. 77-88, March 1988.
- [HAA88]** Haaland D.M. and Thomas E.V., “Partial Least-Square methods for spectral analysis. 1. Relation to other quantitative calibration information”, *Anal. Chem.*, Vol. 60, pp. 1193-1202, 1988.
- [HAZ03]** Hazart J., Grand G., Thony P., Herisson D., Garcia S., Lartigue O., « Spectroscopic ellipsometric scatterometry: sources of errors in critical dimension control », *Proceedings of SPIE*, vol 5041, 2003
- [HEB49]** Hebb D., “The Organization of Behavior”. Wiley : New York, 1949.
- [HOP82]** Hopfield J. J., “Neural networks and physical systems with emergent collective computational abilities”, *Proceedings of the National Academy of Sciences, Etats-Unis*, 79, p. 2554-2558.
- [HOR94]** Hornik K.,Stinchcombe M.,White H. and Auer P., “Degree of approximation results for feedforward networks approximating unknown mappings and their derivatives”, *Neural Computation*, Vol 6, pp. 1262-1275, 1994.
- [JOD94]** Jodouin J.F., “les réseaux neurométriques-modèle et application” Editions Hermes, Paris,1994.
- [KAL99]** Kallioniemi I., Saarinen J., Oja E., «Characterization of Diffraction Gratings in a Rigorous Domain with Optical Scatterometry: Hierarchical Neural-Network Model ». *Applied Optics* 38, 28: 5920-5930 , 1999.
- [KLE08]** Kleijnen J., “Design and analysis of simulation experiments”, Springer, 2008
- [KOH82]** Kohonen T., “Self Organized Formation of Topologically correct Feature Maps”, *Biological Cybernetics*, vol. 46,pp.59-69,1982.

- [LEC86]** LeCun Y., “A learning Scheme for Asymmetric Threshold Networks”, Proceedings of Cognitiva 85, 599-604, Paris, France, 1985
- [LEC89]** LeCun Y., Boser B et al., “Backpropagation applied to handwritten zip code recognition, In Neural computation”, 1, p 541-551, 1989.
- [LET59]** Lettvin J.Y., Maturana H.R, McCulloch W.S., Pitts W.H., “What the frog’s Eyes Tells the Frog’s Brain”, Proceedings of the IRE, Vol. 47, No. 11, pp1940-51, 1959
- [MCC43]** McCulloch, W. S., & Pitts, W. “A logical calculus of the ideas immanent in nervous activity”. The bulletin of mathematical biophysics, 5(4), 115-133, 1943.
- [MIS06]** Misra B. B., Satapathy S. C., Biswal B. N., Dash P. K., & Panda G. “Pattern classification using polynomial neural network”. In Cybernetics and Intelligent Systems, 2006 IEEE Conference on (pp. 1-6). IEEE, June 2006.
- [MIN69]** Minsky M. L. and Papert S.A., “Perceptrons”. Cambridge, MIT press, 1969
- [MON99]** Monari G., « 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, 1999.
- [POG85]** Poggio T, Torre V and Koch C., “Computational vision and regularization theory”, nature, Vol. 317, pp.314-319, 1985.
- [PFI93]** Pfister M., Rojas R., “Backpropagation algorithms: their characteristics and efficiency”, Universität Berlin, Fachbereich Mathematik- Institut für Informatik, B3, 1993.
- [QUI05]** Quintanilha R., “Etude du problème inverse en diffractométrie spectroscopique, application à la métrologie dimensionnelle en microélectronique”, PhD thesis, INPG, 2005
- [RIV94]** Rivals I., Canas D., Personnaz L., Dreyfus G., “Modeling and control of mobile robots and intelligent vehicles by neural networks”, Proceedings of the IEEE conference on Intelligent Vehicles, p. 137-142, 1994
- [ROB03]** Robert S., « Caractérisation optique de réseaux de diffraction submicroniques par des techniques génétique et neuronale ». PhD thesis, Université Jean Monnet de Saint Etienne, 2003.

- [ROS57]** Rosenblatt F., “The perceptron: a perceiving and recognizing automaton”. In Technical report, Cornell Aeronautical Lab, 1957.
- [ROU01]** Roussel P., Moncet F., Barrieu B., Viola A., « 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, G.R.A.I.E.
- [RUM86]** Rumelhart D. E., Hinton G. E., Williams R. J., “Learning internal representations by error back-propagation”. Parallel Distributed Processing : Explorations in the Microstructure of Cognitions, pages 318-362, 1986.
- [SAA93]** Saarinen S., Bramley R., Cybenko G., “Iii-conditioning in neural network training problems”. SIAL J. Sci. Stat. Comp., 14: 693-714, 1993.
- [SAC89]** Sacks J., “Design and analysis of computer Experiments”, Statistical Science, Vol. 4., No. 4, 1989
- [STA06]** Stanley K. O., Bryant B. D., Karpov I., & Miikkulainen R. “Real-time evolution of neural networks in the nero video game”. In AAAI, Vol. 6, pp. 1671-1674, July 2006.
- [STO74]** Stone M., “Cross-validatory choice and assessment of statistical predictions”. Journal of the Royal Statistical Society B, 8 :349-359, 1974.
- [WU13]** Hongying WU, “Contribution to Engine-out Aircraft Trajectory Management and Control”, PHD Thesis, Université de Toulouse, Avril 2013



---

## **Chapitre III. Evaluation des réseaux de neurones pour la scattérométrie en temps réel**

---

Dans les chapitres précédents, deux techniques de résolution de problème inverse utilisées au LHC et au LTM, respectivement les RN et les bibliothèques, ont été présentées. Au cours de précédents travaux, la méthode des Bibliothèques a été validée pour caractériser l'évolution en temps réel de motifs nanostructurés pour un suivi de procédé de gravure [SOU08, ELK10]. Dans le même temps, les RN ont été introduits et validés comme outil performant pour la scattérométrie statique [ROB03, GER09]. Les travaux développés durant cette thèse s'inscrivent dans l'étude de l'apport des RN dans le cadre d'une caractérisation temps réel ou dynamique.

Il convient dans un premier temps de comparer le comportement de ces deux méthodes, sur un cas statique. A notre connaissance aucune étude comparative des deux techniques n'est présente dans la littérature scientifique. Ce travail nous apparait en effet essentiel pour analyser le fonctionnement et la complémentarité de ces 2 techniques adaptées à la caractérisation en temps réel afin d'en tirer le maximum d'efficacité (repousser les limitations de la bibliothèque développées au chapitre I (cf.I.3.2.2.B.b)). Par la suite, les potentialités des RN pour un suivi de procédé en temps réel ont été évaluées sur un cas théorique puis sur un cas expérimental.

### **III.1 Comparaison de la méthode des bibliothèques et des réseaux de neurones**

L'objectif de cette section est d'étudier le comportement des deux méthodes sur un cas statique. Les conditions de l'étude sont définies dans un premier temps puis les résultats sont exposés par la suite.

### III.1.1 Conditions de l'étude

Le profil théorique considéré pour cette étude est composé d'une superposition d'une couche structurée de résine IM5010 sur du BARC, le tout reposant sur un substrat de silicium (Figure III-1). Le profil géométrique est donc défini par trois paramètres à ajuster (CD, h, hr).

La période du motif demeure fixe et vaut  $d=140$  nm. A partir de ce modèle, nous considérons un échantillon théorique dont les dimensions sont respectivement  $CD=49$  nm,  $h=108$  nm,  $hr=80$  nm pour se rapprocher au mieux des valeurs cibles du profil expérimental présenté dans le chapitre II (cf. II.4.2). Ces valeurs correspondent en effet aux valeurs obtenues par caractérisation par RN de cet échantillon.

Concernant la signature utilisée, le spectre considéré varie de 1.5 eV à 6.5 eV avec un pas de 0.2 eV ; ce qui correspond à 52 mesures. L'angle d'incidence est de  $62^\circ$ , choisi afin de se placer dans les mêmes conditions que le cas expérimental. Une signature est simulée à l'aide de la fonction MMFE en introduisant un bruit gaussien sur les intensités  $I_s$  et  $I_c$  d'écart type  $\sigma_s=0,0033$  afin de prendre en compte le bruit de mesure de l'ellipsomètre utilisé [GER09].

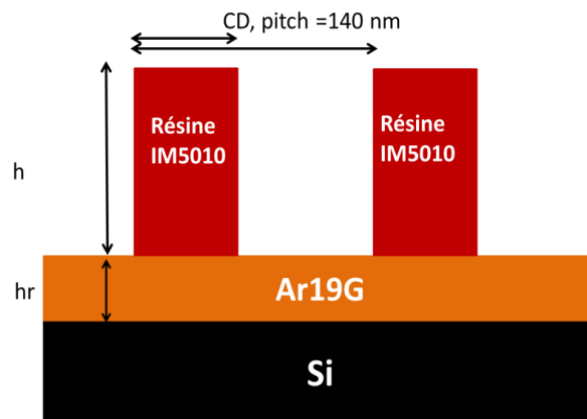


Figure III-1. Profil géométrique défini par trois paramètres (CD, h, hr) considéré pour la comparaison bibliothèque/RN.

En métrologie, les valeurs attendues des paramètres du profil sont souvent données par le fabricant, définies par les contraintes technologiques (réticule utilisé par exemple). Le composant fabriqué peut être légèrement différent de quelques nanomètres, dû au fait que le process utilisé n'est pas toujours parfait. L'objectif de la métrologie est de quantifier précisément le profil fabriqué. Si on fait le lien avec les méthodes de résolution de problème inverse, ceci signifie que les plages de variations des deux méthodes, bibliothèque et RN, doivent être choisies en fonctions des valeurs données par le fabricant. Dans le cas d'un



procédé en production, les plages choisies ne sont pas très larges (quelques nanomètres) car le process étant stable, on s'attend à une forte reproductibilité et à des valeurs proches de la valeur cible. Par contre, dans le cadre d'une application temps réel, la variation des paramètres devient importante, et il devient nécessaire de considérer des plages plus larges. C'est dans ce cadre que s'inscrit l'étude menée dans ce paragraphe, qui s'intéresse à l'influence de la largeur de la plage de variation des paramètres géométriques dans le cas d'une bibliothèque et d'un RN. On souhaite savoir comment se comportent les RN dans des conditions « difficiles », c'est-à-dire où la fonction à modéliser devient plus complexe, car elle est décrite sur un intervalle plus large. Pour la bibliothèque, on s'attend à ce que les résultats ne changent pas quand la plage augmente car la complexité du problème (recherche du minimum de la fonction coût) reste la même ; le seul facteur qui change est l'augmentation du temps de recherche de la solution optimale dans le cas où la bibliothèque est parcourue itérativement, sans l'utilisation de GPU. Mais nous allons vérifier cela dans notre étude.

Pour cela 3 cas de plages de tailles différentes ont été analysés (Table III-1). Les paramètres de la plage 2 et 3 ont été choisis tels que la largeur de la plage soit augmentée respectivement de 100% et 300% par rapport à celle de la plage initiale notée 1.

	Plage de variation 1	Plage de variation 2	Plage de variation 3
CD(nm)	40-65	27-77	10-120
h(nm)	100-136	64-172	28-208
Hr(nm)	68-100	36-132	20-128

Table III-1. Plages de variation des 3 paramètres géométriques (CD, h, hr) utilisés pour l'étude comparative bibliothèque/RN.

Pour la construction de la bibliothèque, les plages de la Table III-1 ont été utilisées. Le pas fixe les valeurs prises par chacun des paramètres géométriques de chaque plage. Le corpus total des échantillons composant la bibliothèque est défini par toutes les combinaisons de celle-ci. Pour chaque jeu de 3 paramètres, soit pour chaque échantillon théorique simulé, une signature est créée à l'aide de la MMFE (cf. I.2.1.3). L'ensemble des signatures constitue la bibliothèque de signatures.

Pour la méthode des RN, le nombre d'échantillons correspond aux couples (signatures/paramètres géométriques) utilisés lors de l'entraînement du RN. Les valeurs des paramètres sont cette fois-ci choisies aléatoirement dans la plage des paramètres selon une loi

uniforme. Nous avons rajouté un bruit gaussien  $\sigma_e$  sur les signatures du corpus d'entraînement afin de se rapprocher d'un apprentissage sur des cas réels.

Pas d'échantillonnage	Plage 1	Plage 2	Plage 3
1nm	31746	539223	2189919
	15000		
	7500		
2nm	4199	70070	280280
		30000	
3nm	1287	20757	83509
		15000	
4nm	630	9100	36064
			26000
5nm	336	4840	18722
6nm		2907	11191
7nm		1792	6656
8nm		1274	4508
9nm		858	3549
10nm		660	2508
			1000
			500

Table III-2. Nombre d'échantillons composant le corpus d'apprentissage pour les réseaux de neurones et les bibliothèques en fonction du pas et de la plage de variation choisis. Des cas ont été exclusivement utilisés pour la bibliothèque (vert) et pour les RN (jaune)

La Table III-2 regroupe le nombre d'échantillons utilisés à la fois pour les réseaux de neurones et pour la bibliothèque pour les trois plages définies dans la table III.1. Concernant la bibliothèque, nous avons inséré le pas correspondant (de 1 à 10 nm). Toutes les bibliothèques de Table III-2 ont été ainsi générées pour la suite de l'étude. Afin de faciliter la comparaison sans biais des résultats, nous avons pris soin d'utiliser, autant que possible, le même nombre d'échantillons pour chaque technique. Quelques variantes résident néanmoins: certains cas (en vert) ont été exclusivement utilisés pour la construction de bibliothèques fortement résolues. Les autres (en jaune), correspondant à un pas non entier, permettront d'étudier les performances du RN avec une meilleure discrétisation de l'espace d'étude.

Dans les travaux réalisés précédemment avec la méthode bibliothèque [SOH92, NU99, HAZ03], y compris au sein du LTM [SOU07, ELK09], les signatures composant la librairie sont non bruitées et permettent d'atteindre de bonnes performances. Nous allons faire de même et construire une librairie de signatures théoriques.

Concernant les caractéristiques structurelles du RN, différentes architectures ont été testées avec, dans un premier temps, une seule couche cachée comprenant respectivement 10, 15, 20, 25 neurones, puis dans un second temps plusieurs couches cachées. (cf. III.1.3.4). Les autres caractéristiques sont celles utilisées dans le chapitre II

### III.1.2 Considérations sur le critère d'erreur calculé sur les signatures reconstruites

Une étude préliminaire a été réalisée afin d'étudier l'influence d'une variation des paramètres géométriques sur l'erreur rms des signatures reconstruites (Equation II.19).

La procédure appliquée par la suite est décrite sur la Figure III-2. Nous considérons un échantillon expérimental, identique à celui de la Figure III-1. Après caractérisation par RN, nous obtenons les valeurs suivantes:  $CD=49$  nm,  $h=108$  nm, et  $hr=80$  nm. Nous avons modifié les paramètres  $CD$ ,  $h$ , et  $hr$  dans une plage variant de 1 à 10 nm ( $\Delta=1$ nm,  $\Delta=2$ nm, ...,  $\Delta=10$ nm). Les paramètres résultants sont notés  $CD'$ ,  $h'$ ,  $hr'$ . Les différentes signatures théoriques sont ainsi systématiquement recalculées puis confrontées à la signature expérimentale par le calcul de  $RMS_1$ . En parallèle, à partir des résultats de la caractérisation ( $CD$ ,  $h$ ,  $hr$ ), nous avons recalculé une signature théorique ( $I_{s_2}$ ,  $I_{c_2}$ ) puis nous l'avons comparé à la signature expérimentale ( $I_{s_{exp}}$ ,  $I_{c_{exp}}$ ) pour obtenir l'erreur  $RMS_2$  qui correspond donc au minimum d'erreur atteignable sur cet échantillon dans les hypothèses structurelles retenues.

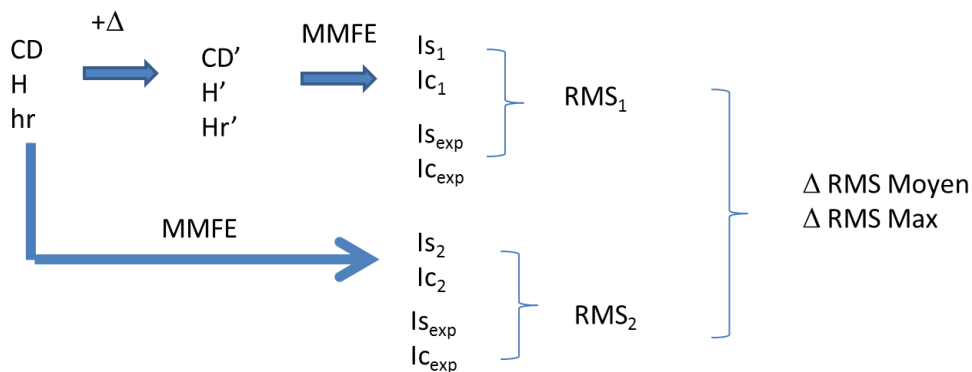


Figure III-2. Schéma représentant la méthodologie utilisée pour étudier l'influence sur l'erreur rms d'une variation des paramètres géométriques

Les paramètres  $CD'$ ,  $h'$ ,  $hr'$  ne peuvent s'écarter de la valeur optimale que dans une limite maximale fixée à l'avance ( $\pm 1$  nm,  $\pm 2$  nm, ...,  $\pm 10$  nm). Prenons l'exemple d'une

variation maximale de  $\pm 1$  nm (Figure III-3). La valeur de CD' peut prendre aléatoirement les valeurs CD-1 nm, CD ou CD+1nm. Ensuite, la valeur de h' peut prendre aléatoirement les valeurs h-1nm, h ou h+1nm et ainsi de suite pour  $h_r$ . Au final on obtient  $3^3=27$  cas de figure et par conséquent 27 profils générés pour chaque variation étudiée( $\Delta$ ). Ainsi, 27 erreurs  $RMS_1$  sont donc calculées à l'issue de la première variation et ainsi de suite. Nous avons alors calculé pour chacune d'elle une différence d'erreur  $\Delta RMS=RMS_2-RMS_1$  maximale.

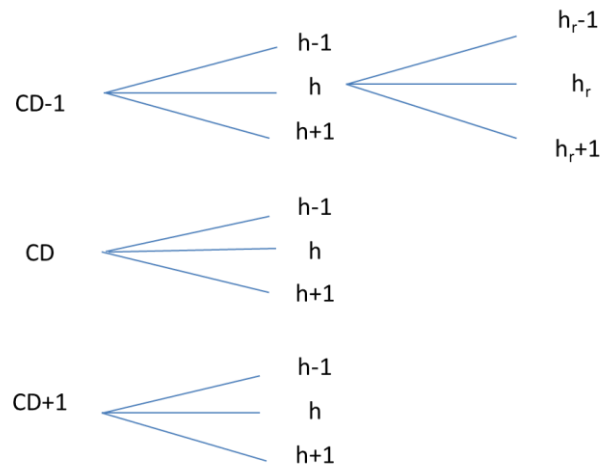


Figure III-3.Exemple d'arborescence pour déterminer les 27 combinaisons correspondant à une variation delta de +/- 1 nm

Les résultats obtenus sont présentés sur la Figure III-4. Ces courbes permettent de confronter la variation d'erreur rms occasionnée par une variation de paramètres. Ainsi, elles peuvent permettre de comparer la qualité de 2 résultats scattérométriques par la seule connaissance des erreurs rms correspondantes reconstruites sur les signatures. Par exemple, une différence maximale  $\Delta RMS_{max}$  de 0.06 correspond à une variation de l'ordre de grandeur de +/- 2 nm sur les paramètres.

Cette courbe sera mise à profit pour l'analyse des performances des méthodes présentées par la suite. Notons que cette courbe est indépendante de la technique de résolution de problème inverse (BB ou RN). On pourrait obtenir une courbe similaire si l'on choisissait de caractériser l'échantillon de la Figure III-1 avec la méthode des bibliothèques. En effet, en dehors de l'étape de caractérisation, les autres calculs consistent à effectuer un calcul direct via MMFE (cf. Figure III-2).

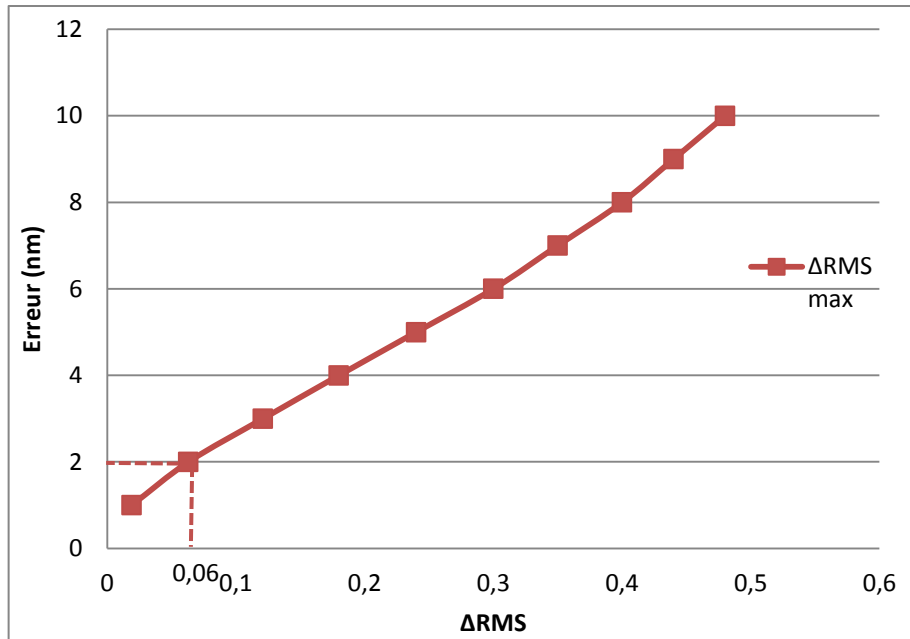


Figure III-4. Erreur sur les paramètres en fonction de la variation d'erreur rms dans le cas de la caractérisation de l'échantillon défini par  $CD=49$  nm,  $h=108$  nm,  $hr=80$  nm.

Cependant, il faut noter que cette courbe n'est exploitable que dans le cas de l'échantillon défini précédemment dans les conditions particulières de l'étude. Elle a pour seul but de donner une information quantitative sous la forme d'écart géométrique sur les paramètres beaucoup plus significative et exploitable que la simple erreur rms sur les signatures reconstruites.

### III.1.3 Etude théorique

Il convient d'abord d'étudier le comportement des deux méthodes sur une signature théorique. L'étude porte sur l'influence du nombre d'échantillons suivant la largeur du domaine de recherche. Une étude est également effectuée pour évaluer la résistance de la caractérisation au bruit de mesure éventuel.

#### III.1.3.1 Utilisation d'une signature théorique

Dans ce paragraphe, une signature théorique sans bruit ( $\sigma_s=0$ ) est modélisée à partir de l'échantillon simulé décrit dans la section précédente ( $CD=49$  nm,  $h=108$  nm et  $hr=80$  nm) (Figure III-1). Elle représente la signature simulée qui sera fournie respectivement au RN et à la bibliothèque par la caractérisation.

Un RN à une couche cachée présentant un nombre de neurones déterminé (10 à 25) est entraîné sur des signatures elles aussi non bruitées ( $\sigma_e=0$ ). L'étude est effectuée sur la page 2.

La Figure III-5 représente les résultats de la caractérisation pour chaque paramètre obtenu par la méthode des bibliothèques et par le RN en fonction du nombre d'échantillons utilisés conformément à la table III.2. Les performances en terme d'erreur rms sur les signatures reconstruites sont également reproduites.

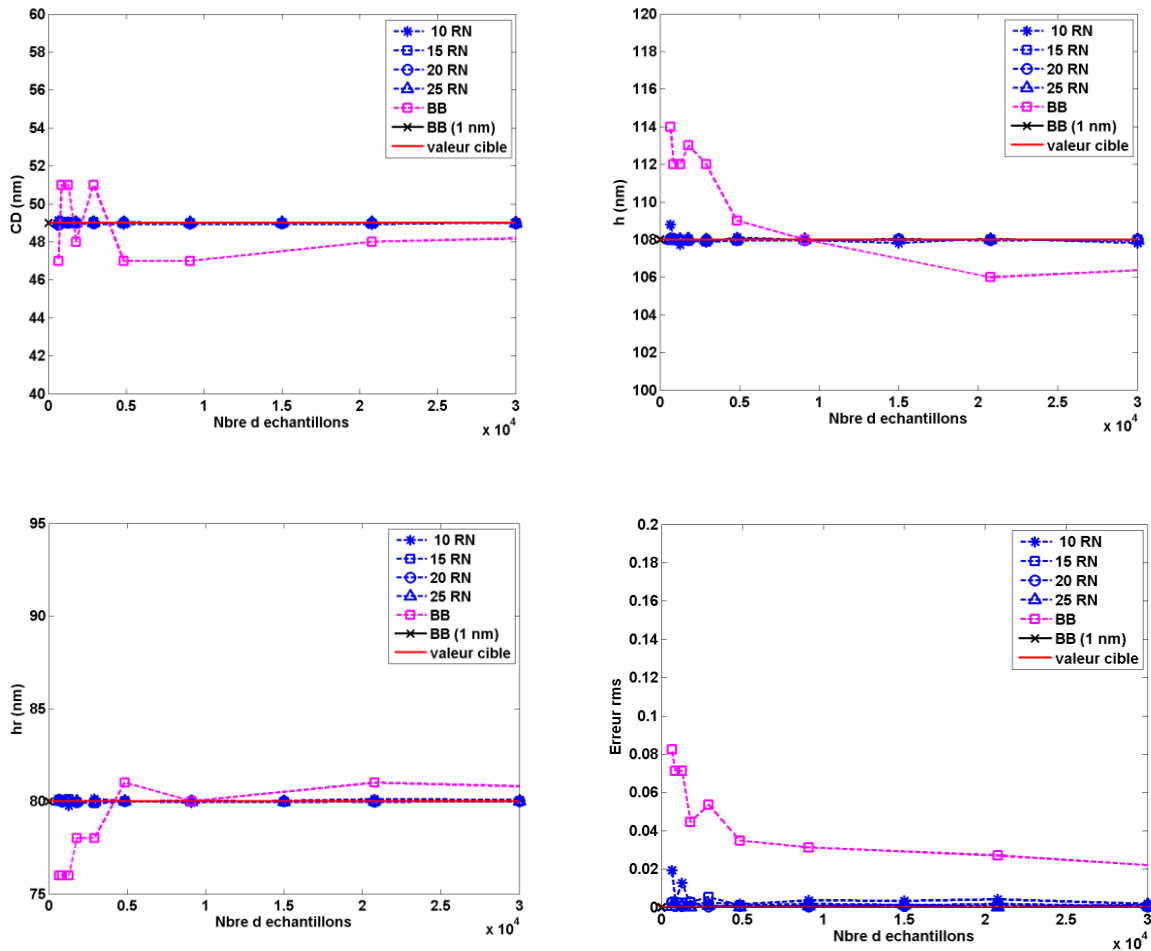


Figure III-5. Résultats de la caractérisation (CD, h, hr et erreur rms sur les signatures reconstruites) par bibliothèque et par RN en fonction du nombre d'échantillons utilisés dans le cas d'une signature théorique fonctionnant sur la plage P2. Le RN possède une seule couche cachée.

Chaque valeur cible correspond à la valeur théorique connue de l'échantillon défini ci-dessus. Les résultats de la bibliothèque correspondant à différentes valeurs du pas sont donnés par la courbe en rose. Rappelons que la bibliothèque construite avec un pas de 1nm est composée de 539223 échantillons. Il est ainsi difficile pour des raisons de clarté évidente de représenter le point correspondant sur la figure. Par conséquent, une droite noire a été tracée pour le pas 1 nm. Cette courbe noire est confondue dans cette étude avec la courbe des valeurs cible (en rouge) confirmant les performances de la bibliothèque fortement résolue.

Analysons d'abord les résultats du RN. Si l'on considère les paramètres géométriques (CD,  $h$  et  $h_r$ ), nous pouvons constater que le RN reproduit bien le motif, et ce indépendamment du nombre de neurones et du nombre d'échantillons. Une architecture simple est donc tout à fait adaptée dans ce cas (par exemple 10 neurones pour 5000 échantillons). Augmenter le nombre de neurones n'est pas souhaitable car cela entraînerait un accroissement de la complexité du modèle inutilement nécessitant plus de ressources (temps de calcul) et pouvant potentiellement conduire à une capacité de généralisation plus faible.

Les résultats obtenus avec la bibliothèque dont le pas est fixé à 1 nm sont tout aussi concluants. Par contre, pour les autres pas, la bibliothèque conduit à des valeurs un peu moins précises, avec une variation maximale de quelques nanomètres, surtout pour un nombre d'échantillons inférieur à 5000. Par conséquent on retrouve logiquement une erreur rms plus importante sur la dernière courbe. Les deux méthodes donnent des résultats excellents mais la méthode des RN nécessite moins d'échantillons pour atteindre les mêmes performances. Notons que si l'on considère que la différence d'erreur rms entre RN et BB est de 0.04 au delà de 5000 échantillons, et que l'on se réfère à la courbe de la Figure III-4, la variation sur les paramètres est de l'ordre de 1.5 nm. Cet écart moyen confirme les résultats obtenus sur chacun des paramètres et visibles sur les 3 courbes de la figure.

Il est important aussi de corréler ces résultats avec les particularités de chaque méthode. Rappelons que les RN s'apparentent à une régression non linéaire ce qui signifie que les valeurs des paramètres obtenues sont interpolées dans le domaine de variation prédéfini. A l'inverse, les résultats de la bibliothèque sont extraits directement dans un domaine discret des paramètres défini par le pas. Ainsi, la solution retenue se rapprochera de la valeur paramétrique la plus proche dans les valeurs discrètes prédéfinies. Le choix du pas détermine ainsi la précision de la mesure. Plus celui-ci est important, plus le résultat obtenu est susceptible de s'écarter de la solution escomptée.

### ***III.1.3.2 Utilisation d'une signature théorique bruitée***

Le même cas a ensuite été traité en prenant en compte cette fois-ci l'incertitude de mesure par un bruit gaussien parfaitement connu ( $\sigma_s=0,0033$ ). Ce bruit correspond à l'estimation du bruit de mesure de l'ellipsomètre. Le RN est cette fois ci entraîné sur des signatures présentant le même type de bruit ( $\sigma_e=0,0033$ ).

Les résultats sont relevés sur la Figure III-6. Le comportement observé dans le cas de la bibliothèque est similaire au cas précédent. Les résultats ne semblent pas ou très peu affectés par le bruit de la signature.

Il n'en est pas de même pour le RN. En effet, les résultats obtenus avec le RN, bien que toujours de meilleure qualité que la bibliothèque, montrent une sensibilité non négligeable au bruit de mesure introduit dans la signature. Seul le paramètre CD semble résister davantage au bruit. En effet, pour extraire correctement les paramètres géométriques, il faut cette fois-ci utiliser 20 ou 25 neurones et un minimum de 5000 échantillons. En pratique, on observe donc que le RN peut répondre aux besoins scattérométriques, mais avec une architecture plus complexe que dans le cas non bruité.

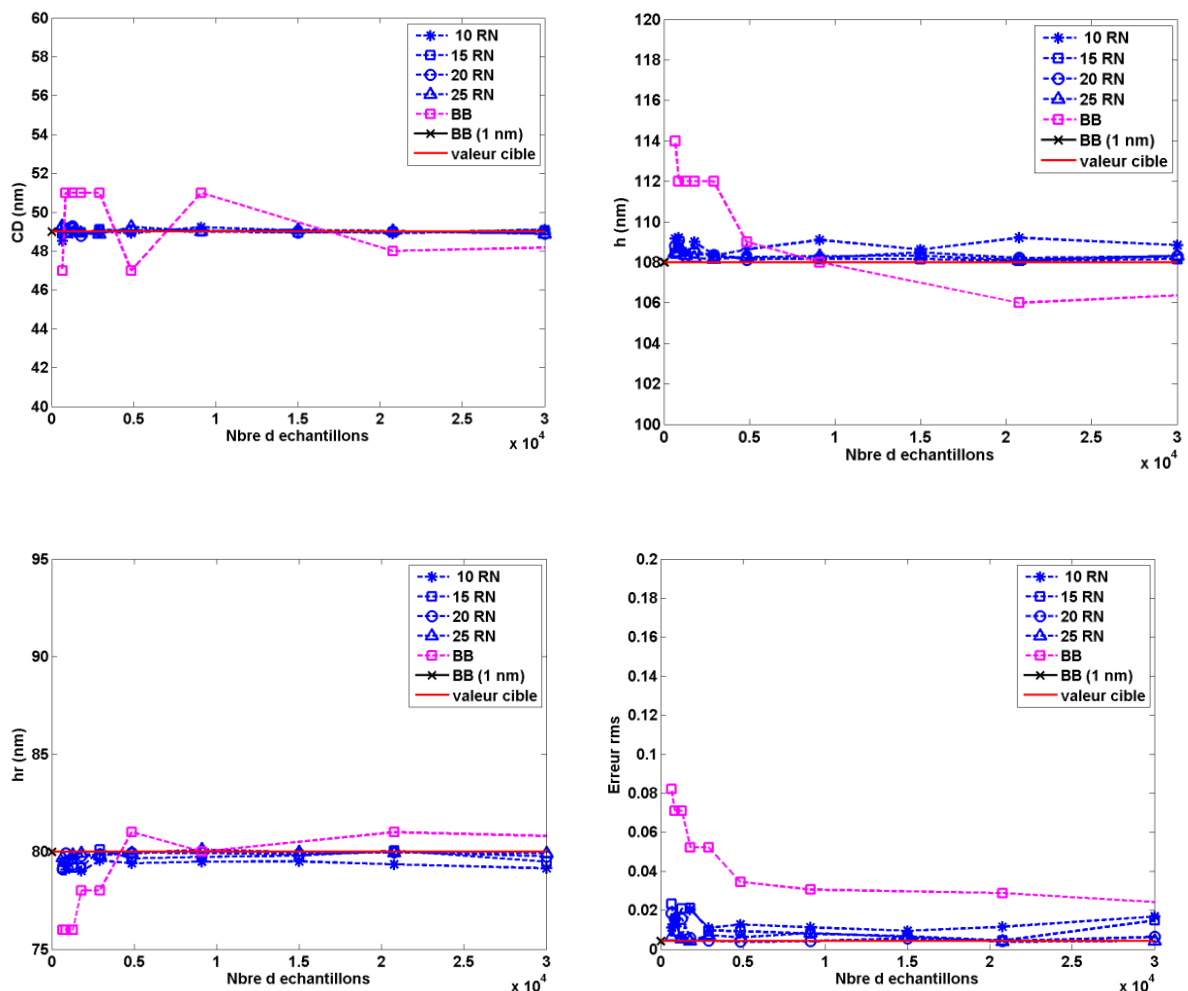


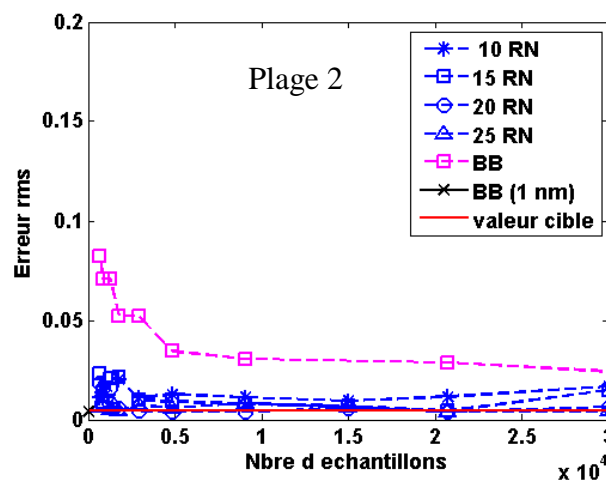
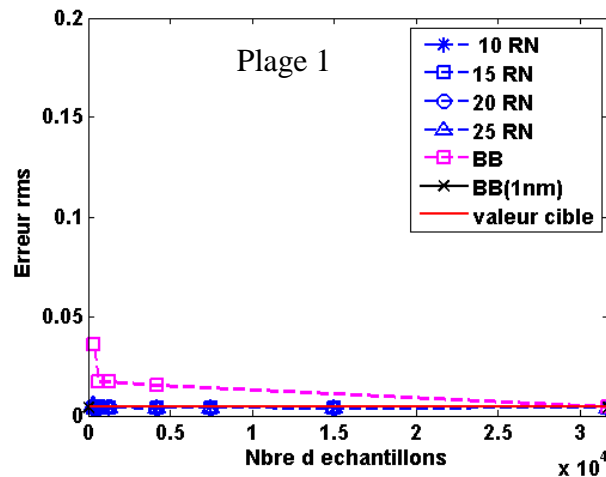
Figure III-6. Résultats de la caractérisation (CD, h, hr et erreur rms) par bibliothèque et par RN en fonction du nombre d'échantillons pour une signature simulée avec un bruit  $\sigma_s=0,0033$  fonctionnant sur la plage P2. Le RN est entraîné avec un bruit  $\sigma_e=0,0033$ .



La valeur cible (courbe en rouge) est obtenue comme dans le paragraphe précédent, en calculant l'erreur rms entre la signature bruitée et la signature théorique sans bruit. Par conséquent on retrouve une valeur de rms égale à 0,0033. Ceci constitue dans notre cas la limite atteignable en terme d'erreur rms par les deux méthodes.

### III.1.3.3 Performances suivant la plage de variation des paramètres

Reprenons les conditions de l'étude précédente ( $\sigma_s=0,0033$ ,  $\sigma_e=0,0033$ ) afin d'examiner les résultats sur chaque plage de variation définie précédemment (Table III-1). Par souci de clarté la Figure III-7 ne représente que les résultats en terme d'erreur rms sur les signatures reconstruites pour les 3 plages de variation.



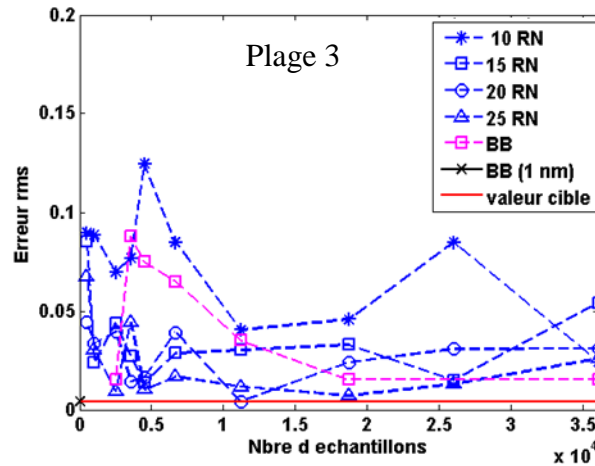


Figure III-7. Résultats de la caractérisation (erreur rms sur signatures reconstruites) par BB et par RN en fonction du nombre d'échantillons pour une signature simulée avec un bruit  $\sigma_s=0,0033$  effectué sur les trois plages de variation. Le RN présente une seule couche cachée et est entraîné avec un bruit  $\sigma_e=0,0033$ .

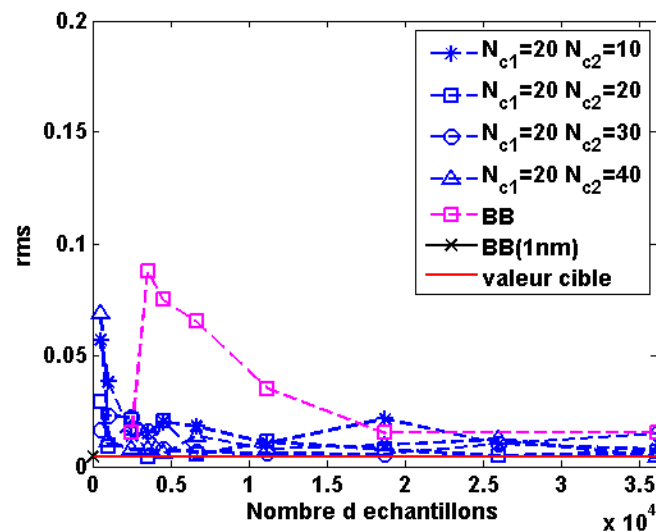
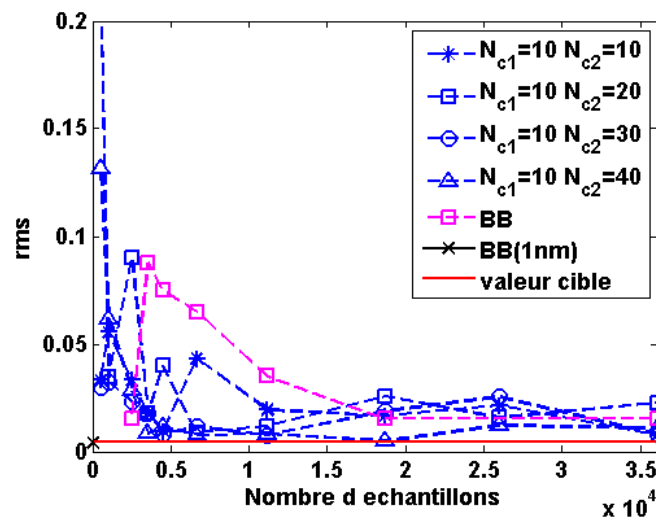
Ce qui est constaté en premier lieu est que le RN donne des valeurs proches des valeurs cibles dans le cas de la plage 1, et ce malgré le bruit gaussien. On observe ensuite une dégradation des performances du RN lorsque la taille de la plage de variation augmente. Pour la plage 3, les résultats dépendent fortement du nombre de neurones composant l'unique couche cachée. Ainsi seul le RN avec  $N_c=25$ , présente des résultats comparables avec ceux obtenus pour la plage 2. A noter également que le nombre d'échantillons influe peu sur les résultats, ceci pour les trois cas de figure.

Enfin, les bibliothèques semblent moins dépendantes de l'augmentation de la plage de variation. On retrouve des erreurs rms quasiment du même ordre de grandeur entre les plages 2 et 3. En effet, dans la méthode des bibliothèques, le principe repose sur la recherche, dans la bibliothèque élaborée, de la signature qui ressemble le plus à la signature expérimentale, ceci indépendamment de la largeur de la plage de variation. Pour les RN, l'augmentation de la plage entraîne l'augmentation de la complexité de la fonction à modéliser et nécessite par conséquent plus de ressources (architecture plus complexe). Pour la méthode des bibliothèques, les performances augmentent avec le nombre d'échantillons composant la bibliothèque.

En conclusion, cette étude a montré que les performances du RN sont beaucoup plus dépendantes de l'espace de variation des paramètres que la bibliothèque. De bonnes performances peuvent cependant être obtenues à condition d'optimiser l'architecture du RN.

### III.1.3.4 Performances suivant le nombre de couche cachées

Afin d'améliorer les performances obtenues avec la page 3 du paragraphe précédent, une architecture présentant deux couches cachées est étudiée. Différentes configurations sont testées :  $N_{c1}=10, 20$  et  $30$  neurones dans la couche 1 et respectivement  $N_{c2}=10, 20, 30, 40$  dans la couche 2. Les résultats sont présentés dans la Figure III-8.



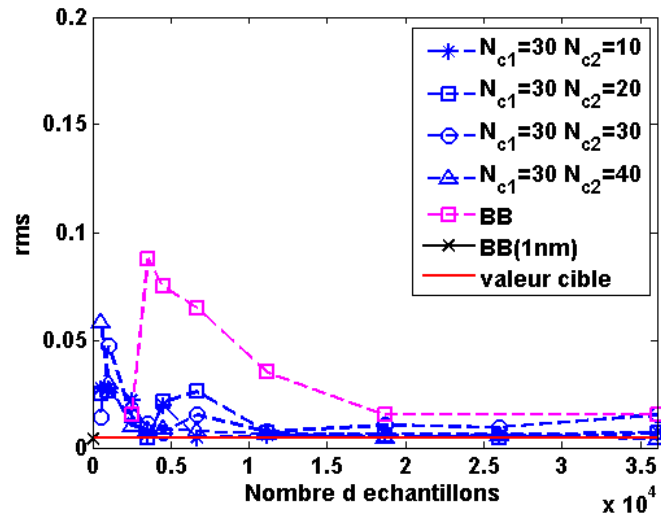


Figure III-8. Résultats de la caractérisation (erreur rms sur signatures reconstruites) par bibliothèque et par RN en fonction du nombre d'échantillons pour une signature simulée avec un bruit  $\sigma_s=0,0033$  fonctionnant sur la plage 3. Le RN présente 2 couches cachées et est entraîné avec un bruit  $\sigma_c=0,0033.3$  cas de figures sont considérés pour  $N_{c1}$  (10,20,30) et respectivement 4 cas de figures sont considéré pour  $N_{c2}$  (10,20,30,40).

Retenons d'après l'analyse de ces courbes que la double couche améliore significativement les résultats. En effet, l'architecture présentant 30 neurones sur la première couche cachée présente des résultats comparables à ceux obtenus sur la plage 1 dans le cas précédent (Figure III-7). Par ailleurs, le nombre de neurones sur la seconde couche ne semble pas affecter les résultats. Une architecture satisfaisante est donc donnée par 2 couches cachées composées respectivement de 30 et 10 neurones.

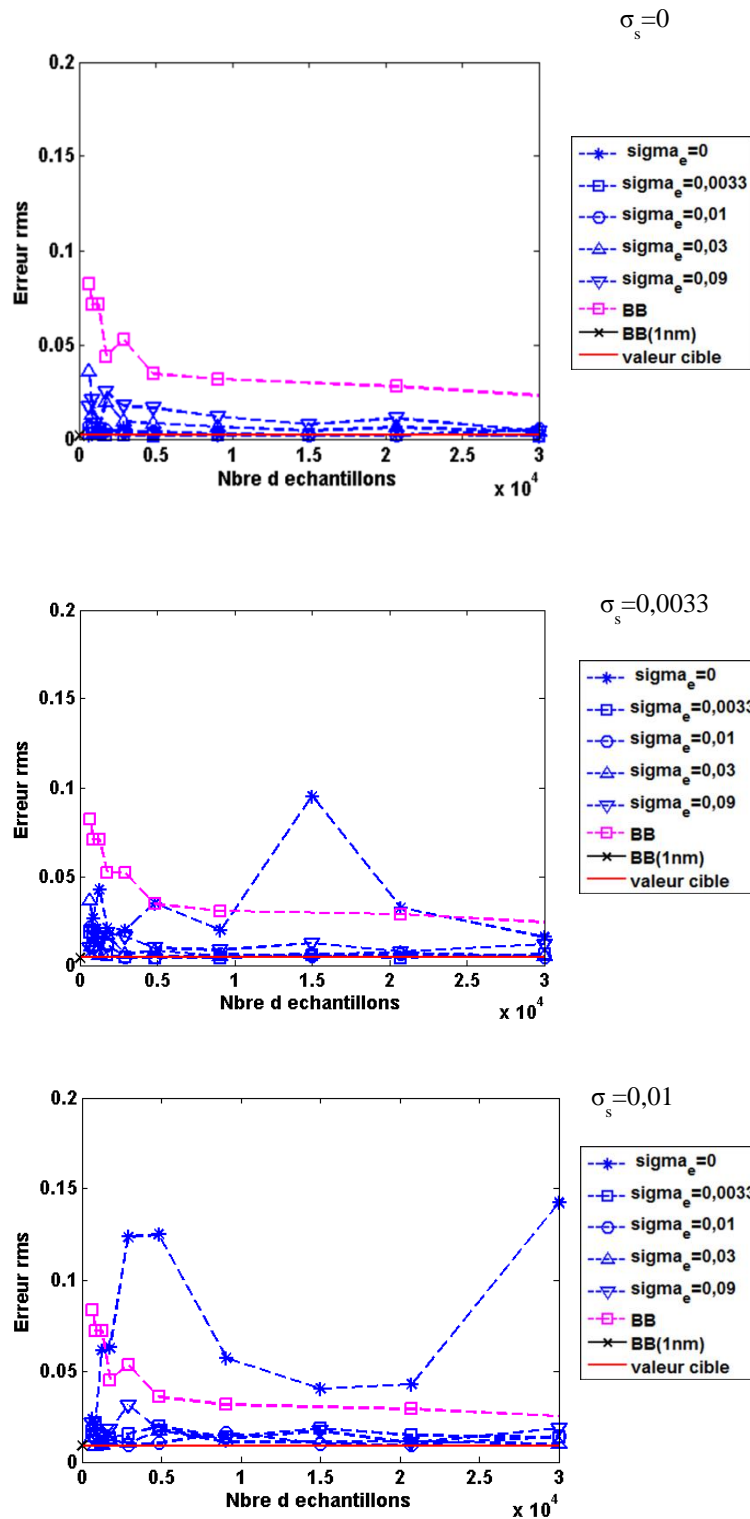
Cette étape supplémentaire, de recherche de l'architecture optimale, bien qu'elle puisse être parfois intuitive et réduite avec l'expérience de l'utilisateur, nécessite du temps et peut présenter dans certains cas, une contrainte de l'utilisation des RN.

Précisons qu'augmenter le nombre de couches implique nécessairement plus de ressources à utiliser, se traduisant généralement par un accroissement du temps d'entraînement (cf. étude effectuée en II.3.1).

### III.1.3.5 Résistance de la caractérisation au bruit de mesure

Il devient intéressant d'évaluer maintenant la résistance des deux méthodes au bruit de mesure. Reprenons les mêmes conditions que celles utilisées dans le paragraphe III.1.3.2 (RN entraîné sur la plage 2 avec une seule couche cachée) et considérons différents bruits  $\sigma_s$  sur la signature simulant la mesure de l'échantillon. 5 cas de figures ont été traités ( $\sigma_s=0$ ,  $\sigma_s=0,0033$ ,  $\sigma_s=0,01$ ,  $\sigma_s=0,03$  et  $\sigma_s=0,09$ ). Pour chacun d'entre eux, l'architecture du RN optimale en

termes de neurones a été retenue. Les performances (Figure III-9) ont été relevées pour la bibliothèque et pour des RN entraînés sur des signatures présentant différents niveaux de bruit ( $\sigma_e=0$ ,  $\sigma_e=0,0033$ ,  $\sigma_e=0,01$ ,  $\sigma_e=0,03$  et  $\sigma_e=0,09$ )



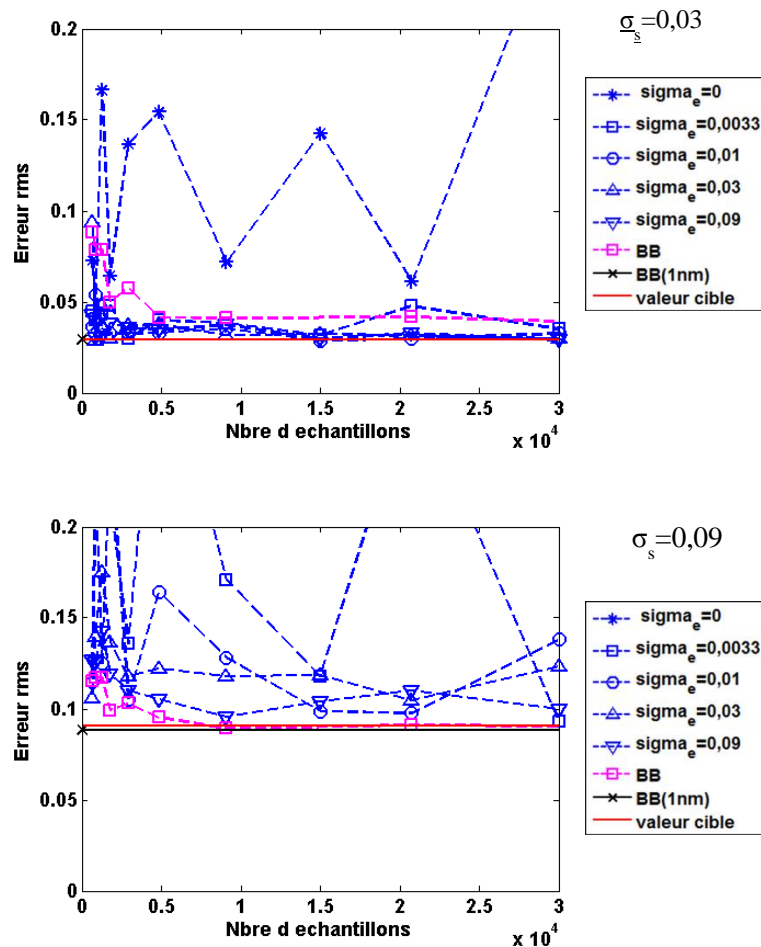


Figure III-9. Résultats de la caractérisation (erreur rms sur signatures reconstruites) par BB et par RN en fonction du nombre d'échantillons pour une signature simulée avec différents bruits  $\sigma_s$ , fonctionnant sur la plage 2. Le RN présente une couche cachée et est entraîné avec différents bruits  $\sigma_e$ .

Les résultats obtenus montrent que la présence de bruit altère les résultats des 2 méthodes mais de différente façon. Au vu des courbes, la valeur du bruit de mesure introduit sur la signature théorique a plus d'influence sur les performances du RN que sur la méthode des bibliothèques.

Concernant le RN, l'introduction de bruit dans les données d'apprentissage est indispensable pour avoir des performances acceptables. En effet, dans le cas où le RN est entraîné sur des signatures non bruitées, les performances sont systématiquement dégradées en présence de bruit sur la signature à analyser. Par contre, la valeur du bruit sur les données d'entraînement ne semble pas critique. Notons également que, plus le bruit de la signature est important, plus il faut augmenter le bruit  $\sigma_e$  pour limiter l'erreur rms.

Concernant la bibliothèque, une dégradation de performances dans un degré moindre que les RN est observée lorsque le bruit  $\sigma_s$  augmente. Rappelons que les signatures composant la bibliothèque sont théoriques alors que pour le RN l'ajout d'un bruit est essentiel dans la

phase d'entraînement. Le RN doit absolument être entraîné dans les mêmes conditions que celles correspondant à la phase de traitement afin d'assurer une généralisation optimale.

Cette étude montre que, lors de l'utilisation des RN en scattérométrie, il faut prendre un soin tout particulier concernant l'optimisation de son architecture. Si ce dernier est mal configuré, ceci peut avoir un impact sur les performances. De son côté l'utilisation des bibliothèques ne semble pas souffrir de telles contraintes.

### III.1.4 Etude sur un échantillon expérimental

Toutes les études précédentes ont été réalisées sur une signature simulée par MMFE. Intéressons-nous à présent à l'étude du comportement des deux méthodes sur un cas réel. L'échantillon est celui présenté dans le chapitre II (Figure III-10)

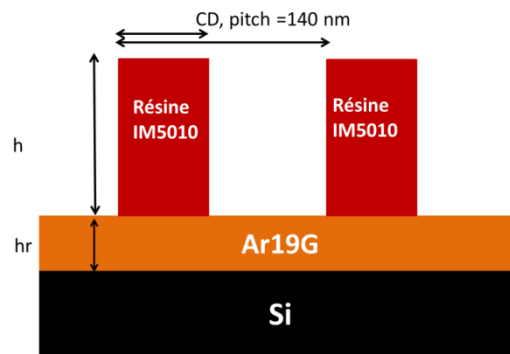


Figure III-10. Modèle crête défini par 3 paramètres ( $CD$ ,  $h$  et  $h_r$ ) utilisé pour la caractérisation de l'échantillon Im5010 sur Si

Les bibliothèques construites précédemment ainsi que les différents RN entraînés dans les conditions détaillées ci-dessus correspondent au cas de l'étude et peuvent être utilisés pour la caractérisation.

La configuration optimale du RN a été sélectionnée en s'inspirant de l'étude précédente, notamment dans le cas où l'on considère une signature théorique avec un bruit gaussien  $\sigma_s=0,0033$ , qui correspond au bruit de mesure mesuré de l'ellipsomètre [GER09]. La page 2 a été retenue pour cette section. L'étude théorique (cf. III.1.3.2) a permis de déterminer l'architecture la plus adaptée, à savoir un RN à une couche cachée de 25 neurones, entraîné avec  $\sigma_e=0,0033$ . Les paramètres géométriques extraits par les deux méthodes ainsi que les

erreurs sur les signatures reconstruites sont présentés sur la Figure III-11 en fonction du nombre d'échantillons utilisé.

Les valeurs issues de la BB avec un pas de 1 nm correspondraient aux valeurs optimales de l'échantillon dans le cas où les hypothèses sous-jacentes à la caractérisation (modèle du profil, indices et conditions de mesures) seraient vérifiées.

Comparons les résultats obtenus avec les résultats de la Figure III-6. On observe que les valeurs de caractérisation expérimentales obtenues sur l'échantillon à 1 nm sont légèrement différentes de l'échantillon théorique considéré. En effet, les résultats sont respectivement de  $CD=48$  nm,  $h=105$  nm,  $hr=83$  nm pour le cas de la bibliothèque à 1nm(Figure III-11) et de  $CD=49$ ,  $h=108$ ,  $hr=80$ (Figure III-6) pour les valeurs cibles de l'étude théorique. Cet écart est relatif au choix effectué dans notre étude théorique, de considérer les résultats issus de la caractérisation par RN de l'échantillon de la Figure II-28 comme étant la valeur cible. Il aurait été plus précis de considérer les valeurs de caractérisation à 1 nm par bibliothèque comme étant la valeur cible. Cependant, cette étude théorique a été effectuée au début de nos travaux de thèse, avant l'étude expérimentale. Les conclusions obtenues pour l'étude théorique devraient s'appliquer à l'étude expérimentale, car on considère que l'écart entre les résultats des caractérisations paramétriques expérimentale et théorique est faible en terme de pourcentage ( $\Delta CD=2\%$ ,  $\Delta h=3\%$ ,  $\Delta hr=4\%$  obtenu en considérant le rapport entre l'écart entre les deux valeurs et la valeur minimale).

Notons une élévation de la valeur limite atteignable en terme d'erreur rms (0.09 au lieu de 0.0033); ce qui nous conduit à penser que ces hypothèses ne sont pas totalement vérifiées. En effet, le profil réel présente des bords arrondis ainsi que de la rugosité sur les flancs (Figure II-33). Le modèle créneau considéré ne reflète donc pas tout à fait la réalité et ceci se retranscrit sur les résultats d'erreur rms obtenus.



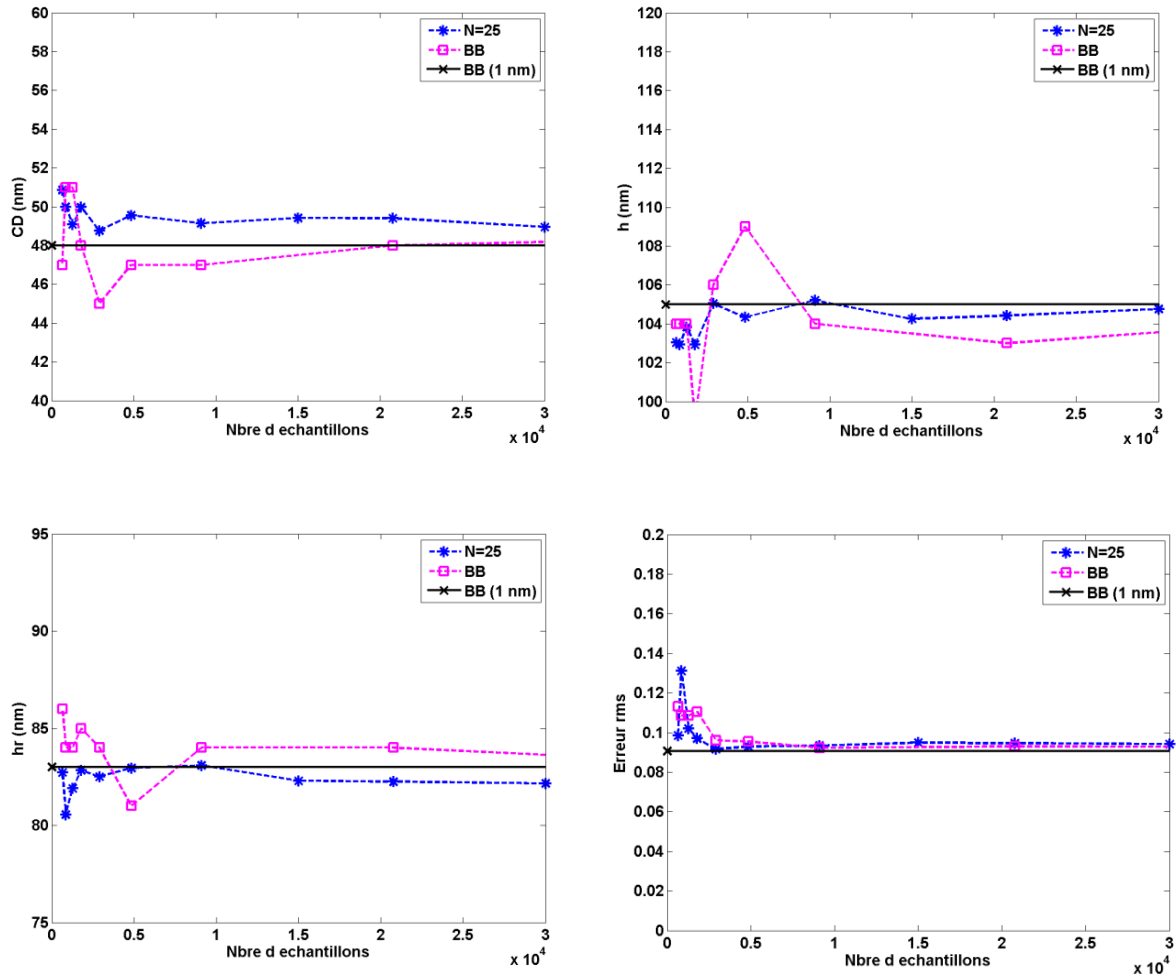


Figure III-11. Résultats de la caractérisation (CD, h, hr et erreur rms sur les signatures reconstruites) par bibliothèque et par RN en fonction du nombre d'échantillons pour une signature réelle fonctionnant sur la plage P2. Le RN est composé d'une couche cachée de 25 neurones et est entraîné avec un bruit  $\sigma_e=0,0033$ .

Le RN permet, à partir d'un certain nombre d'échantillons, de déterminer les paramètres au nanomètre près par rapport aux valeurs optimales obtenues par la bibliothèque. Par exemple, pour le paramètre h, la bibliothèque requière au moins 10 000 échantillons pour atteindre la valeur optimale alors que le RN nécessite au moins 4000 échantillons. Par contre, cette différenciation n'est pas décelable dans le calcul de l'erreur rms sur les signatures reconstruites car la sensibilité n'est pas la même pour les différents paramètres.

### III.1.5 Conclusion de l'étude

Cette étude a permis d'apprécier le comportement de deux méthodes bibliothèque et RN sur des signatures théoriques et expérimentales. Rappelons que cette étude a été réalisée dans un cadre précis (profil, type de matériaux, domaine d'étude, etc ...). Il est ainsi difficile de généraliser les conclusions établies pour tout type d'échantillon mais on peut raisonnablement

penser que le comportement, en terme qualitatif, serait très proche pour d'autres cas expérimentaux sur des structures du même type.

Pour le RN, on a montré qu'une architecture simple en termes de neurones et d'échantillons est suffisante pour bien approcher la fonction à modéliser, mais que l'augmentation de la complexité de la structure, notamment le nombre de couches, est nécessaire pour des plages importantes. La connaissance à priori des valeurs géométriques des paramètres recherchés permettra ainsi de limiter la plage de variation et donc d'utiliser des architectures plus simples. Cependant, il est parfois nécessaire de considérer des plages plus larges pour des applications temps réel, où le profil évolue au cours du temps. Dans ce cas, il sera nécessaire de considérer des architectures plus fortes pour atteindre les mêmes performances. Contrairement à la bibliothèque, la présence de bruit dans l'entraînement est indispensable pour atteindre de bonnes performances. La modélisation précise de ce dernier n'est cependant pas critique.

Pour la méthode des bibliothèques, on a constaté qu'un pas de 1nm permet d'atteindre les performances optimales mais que les résultats obtenus avec un pas plus important peuvent être satisfaisants et permettent d'avoir des bibliothèques plus compactes.

La méthode des RN présente des performances similaires voire dans certains cas meilleures que la bibliothèque pour un nombre d'échantillons égal. Notons qu'il est nécessaire au préalable de prendre soin d'optimiser l'architecture du RN en question, ce qui rajoute une étape supplémentaire et peut présenter une contrainte d'utilisation de cet outil. Cette étude étant effectuée pour appréhender les spécificités d'un RN, l'expérience de l'utilisateur des RN jouera un rôle pour épargner du temps et pour faire le bon choix initial de l'architecture.

Il a été montré également qu'il réside entre les deux méthodes un écart très petit en terme d'erreur rms correspondant à une variation nanométrique sur les paramètres. Il convient de se demander si une telle différence a une importance pour l'utilisateur.

Après cette étude comparative des deux outils largement utilisés en scattérométrie, il convient de se servir de ces résultats pour s'intéresser à l'application des RN pour une application en dynamique de suivi de procédé de gravure plasma.

## III.2 Suivi en temps réel de procédé de gravure plasma par voie neuronale

Le suivi d'une procédure de gravure de résine en régime dynamique a déjà été validé sur un cas simulé [SOU07] et sur une étude expérimentale [ELK10]. Les études effectuées au sein du LTM utilisaient la méthode des bibliothèques (Chapitre I.3) avec l'utilisation des GPU pour une approche vectorielle de la recherche de la solution [SOU08]. Nous allons présenter une alternative à ces études en utilisant la méthode des RN. Nous évaluerons dans ce chapitre les potentialités des RN pour le contrôle in-situ de procédés de gravure plasma. Les contraintes d'une application en temps réel seront alors évoquées puis une validation de la méthode neuronale sera conduite sur un cas théorique puis sur un cas expérimental.

Il convient avant de commencer notre étude de rappeler quelques principes généraux sur la gravure.

### III.2.1 La gravure plasma

La fabrication d'un réseau périodique de dimension nanométrique repose sur deux étapes clé (Figure III-12) : La lithographie (Annexe 1) et la gravure: la première consiste à définir le motif à produire dans une couche de résine photosensible (dans le cas d'une photolithographie) déposée préalablement sur le matériau à graver (silicium par exemple). La seconde est l'étape de gravure (gravure plasma dans notre cas), qui consiste à transférer dans le matériau le motif à graver à partir du masque en résine.

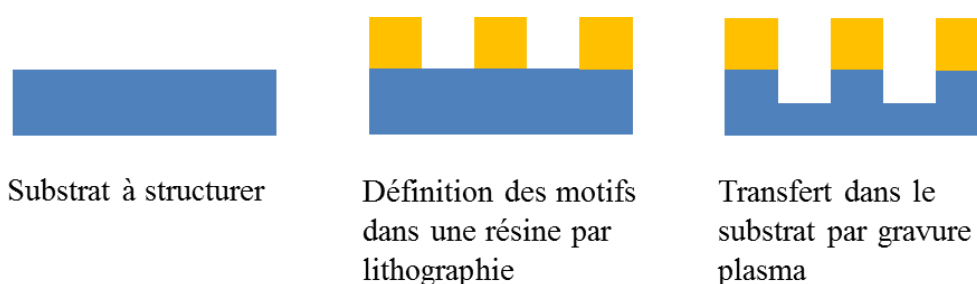


Figure III-12. Schéma des différentes étapes nécessaires pour la fabrication d'un réseau de ligne périodique

Le plasma est un milieu gazeux ionisé neutre qui contient des électrons, des ions négatifs et positifs ainsi que des particules neutres (atomes, molécules, etc.) soumis à des interactions collectives. Les plasmas occupent une place importante dans l'industrie de la microélectronique car ils sont souvent utilisés dans l'étape de gravure.

Plusieurs types de gravures sont utilisés selon la nature des matériaux :

- La gravure IBE (en anglais Ion Beam Etching) [SIG69, HAR89] qui est de nature physique. Le faisceau ne comporte pas d'éléments qui interagissent chimiquement avec le matériau à graver. La surface du matériau est bombardée par les ions du plasma, qui transfèrent aux atomes de surface suffisamment d'énergie pour être libérés de cette surface.
- La gravure RIE (en anglais Reactive Ion Etching) [WIN83]. Cette attaque est de nature chimique et ne dépend que de l'interaction entre le matériau à graver et les espèces réactives du plasma.
- La gravure RIBE (en anglais Reactive Ion Beam Etching) qui est de nature physico-chimique. Le bombardement des ions ainsi que la réaction chimique des particules du plasma avec les matériaux conduisent à l'arrachement de la matière.
- La gravure ICP (« Inductively Coupled Plasma ») qui sera expliquée plus loin dans le manuscrit cf. III.2.4.1). C'est cette méthode qui sera utilisée dans le cadre du travail réalisé

Pour un complément d'information sur les techniques de gravure, le lecteur intéressé pourra se référer à la référence [PAR04].

Le procédé de « réduction de côte de résine » [YEN04] a été introduit afin de faire face aux limites dimensionnelles imposées par la lithographie concernant les dimensions des grilles de transistor. Le principe de cette technique repose sur la possibilité d'éroder latéralement et verticalement par gravure plasma les motifs de résine (Figure III-13), ce qui permet de réduire la largeur du motif. Dans la suite du manuscrit l'appellation anglaise « resist trimming » sera utilisée pour désigner ce procédé.

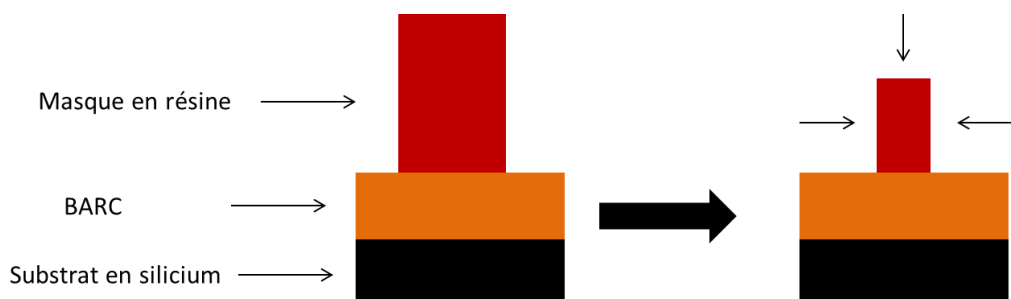


Figure III-13. Schéma représentant la réduction de côte de résine sur un profil créneau

L'inconvénient de cette technique est que l'espace entre les lignes augmente après l'érosion de cote de résine, maintenant ainsi la densité d'intégration constante. Le resist

trimming n'est donc pas intéressant pour accroître les capacités de stockage des DRAMs [PAR04] par exemple. Dans le cas des circuits logiques CMOS [YEN04], cette méthode devient très intéressante car elle augmente la vitesse à laquelle les électrons se déplacent de la source au drain.

Les profils considérés pour modéliser les motifs utilisés dans le resist trimming dans ces travaux de thèse sont différents (cf.II.4.2.). Ils présentent un profil plus ou moins créneau. Dans certains cas, on s'éloigne du profil créneau et on s'approche d'un profil arrondi. La Figure III-14 représente une caractérisation par CD-AFM réalisée avant le début du resist trimming sur un motif résine sur BARC sur Si. C'est ce motif qui sera pris en référence dans la suite. Le profil de ce motif s'approche plus d'une forme arrondie notamment en haut des sillons et présente des rugosités sur le flanc.

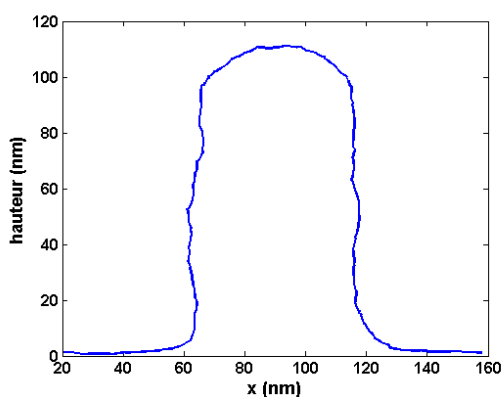


Figure III-14. Résultats de caractérisation d'un profil de résine sur BARC sur Si par CD-AFM avant l'étape de resist trimming

Dans la suite, un outil de suivi en temps réel de procédé de gravure utilisant des RN est étudié. Nous avons choisi de tester la méthode sur un procédé de réduction de côte de résine tel que celui présenté dans le paragraphe précédent dont le nombre de matériaux impliqués est restreint.

### III.2.2 Contraintes d'une application en temps réel

La capacité d'application d'un suivi en temps réel par la scattérométrie est liée à la configuration des paramètres utilisés par l'outil de mesure scattérométrique (ellipsomètre dans le cas étudié): le temps d'intégration et la période d'acquisition de la signature.

La période T correspond au rythme d'acquisition de la signature ellipsométrique. Notons que plus ce temps est long, plus le bruit de mesure est réduit. Ainsi, une nouvelle

signature est mesurée toutes les T secondes. Le temps maximal disponible pour effectuer la résolution du problème inverse correspond donc à T. Pour que l'application traitée soit qualifiée de temps réel, il est nécessaire que le temps de résolution du problème inverse soit inférieur à la période d'acquisition ellipsométrique. Pour nos expérimentations, l'outil ellipsométrique que nous utilisons possède une période d'acquisition de 500ms, qu'il est possible de réduire en contrepartie d'un bruit de mesure plus important.

Il est ainsi nécessaire d'évaluer le temps de résolution de problème inverse par les deux méthodes envisagées. Prenons le cas d'un échantillon précédemment utilisé dans le chapitre II (cf II.4.2). Le profil retenu pour effectuer cette étude est le profil elliptique (Figure II-34 gauche). Le PC utilisé pour cette étude est un Intel Xeon@3,6GHz et 12 G.o. de RAM.

Les temps de résolution de problème inverse sont présentés dans la Table III-3.

	RN	BB
Temps de chargement (ms)	69	7270
Temps de résolution du problème inverse (ms)	76	140
Temps maximal disponible pour la caractérisation temps réelle	500 ms	500 ms

Table III-3. Temps caractéristiques des deux méthodes scattérométriques (RN et BB) pour la résolution du problème inverse pour une caractérisation d'un motif d'un profil elliptique.

En premier lieu, il faut noter que les temps de résolution de problème inverse sont inférieurs à la période d'acquisition (500 ms) pour les 2 méthodes. Elles respectent donc la condition de temps réel.

On remarque que la méthode de RN est deux fois plus rapide que la méthode des bibliothèques. Il faut noter que le temps de résolution du problème inverse pour la méthode des bibliothèques peut être grandement réduit dans le cas de l'utilisation des GPU non disponible sur le PC utilisé pour cette comparaison. Soulan et al [SOU08] ont montré un pouvoir d'accélération de l'ordre de 6 des GPU par rapport à un CPU « standard ». Ces résultats peuvent être exploités dans le cas où le temps d'acquisition des signatures est réduit, permettant un suivi plus échantillonné de l'évolution des structures.

Le temps de chargement des données nécessaires pour la caractérisation d'un motif donné a également été relevé dans la table III.3 pour les 2 méthodes. Pour la méthode des RN, ce fichier contient seulement les matrices des poids  $Z$  et  $W$  qui permettent d'obtenir les sorties du RN (cf. chapitre II.1.6.1). Ce temps est indépendant du nombre d'échantillons constituant la base d'apprentissage. Seuls l'architecture utilisée et le nombre de sorties conditionnent la dimension des matrices en question. Pour la méthode des bibliothèques, il contient l'ensemble des signatures de la bibliothèque. Il est donc clair que la taille du fichier dépend du nombre de signatures, qui dépend lui-même du nombre de paramètres utilisés pour définir le motif. Pour le temps de chargement, la bibliothèque consomme donc beaucoup plus de temps par rapport à la méthode des RN. Ceci n'est pas contraignant pour une application en statique du fait que les fichiers sont chargés avant le début de la mesure. Par contre, dans le cas particulier d'une application nécessitant un changement de profil en cours de procédé, il est nécessaire de charger le fichier possédant les données modélisant le nouveau profil. La contrainte en temps réel n'est donc plus vérifiée pour les bibliothèques compte tenu du temps de chargement. La méthode des RN sera alors plus adaptée à l'évolution du profil car le temps de chargement est quasi instantané.

Ces deux techniques sont donc parfaitement applicables pour un suivi en temps réel (sans changement de forme du motif pour la bibliothèque).

On va à présent s'intéresser à la validation de l'outil scattérométrique neuronal pour le suivi de procédés sur un cas théorique puis expérimental.

Pour notre étude, l'ellipsomètre est configuré en mode dynamique (cf I.2.2.1.3) selon le cas étudié: dans l'étude théorique, le cas avec 16 ou 32 PM (PhotoMultiplificateurs) est considéré, chaque PM étant fixé sur une longueur d'onde particulière. Ces cas ont été inspirés des travaux effectués par S. Soulan, qui a caractérisé des échantillons dans les deux cas de figure [SOU08]. Ainsi, le spectre parcouru est de 1.4eV à 6.1eV avec un pas respectivement de 0,3eV et 0,15eV. L'étude expérimentale (cf III.2.4) est quant à elle réalisée sur un spectre de 1,5 à 6,5 eV en utilisant 32 PM (configuration imposée par l'équipement).

### **III.2.3 Etude théorique de contrôle de procédés par voie neuronale**

L'objectif de cette étude est de simuler un procédé de «resist trimming » afin d'évaluer les potentialités des RN pour le temps réel. Un profil créneau constitué d'une couche de résine

(IM5010) déposée sur du BARC reposant sur un substrat Si a été considéré (Figure III-15). Le choix du profil créneau a été effectué car c'est le motif le plus simple.

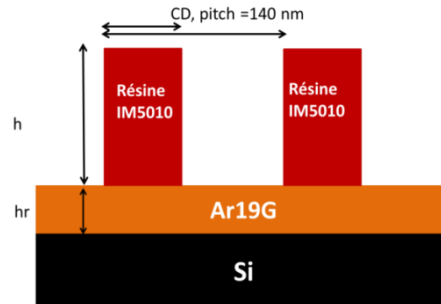


Figure III-15. Profil géométrique créneau défini par 3 paramètres (CD, h, hr)

Le procédé de « resist trimming » est simulé en faisant varier au cours du temps les paramètres géométriques du profil selon la loi définie dans la Figure III-16. Une signature est simulée par problème direct (MMFE) pour chaque jeu de paramètres et pour chaque incrément temporel. On obtient ainsi un jeu de simulation de mesures correspondant au suivi de procédé de resist trimming au cours de temps. Un bruit gaussien ( $\sigma_s=0,0033$ ) est considéré sur les signatures théoriques calculées par problème direct, afin de simuler des signatures expérimentales.

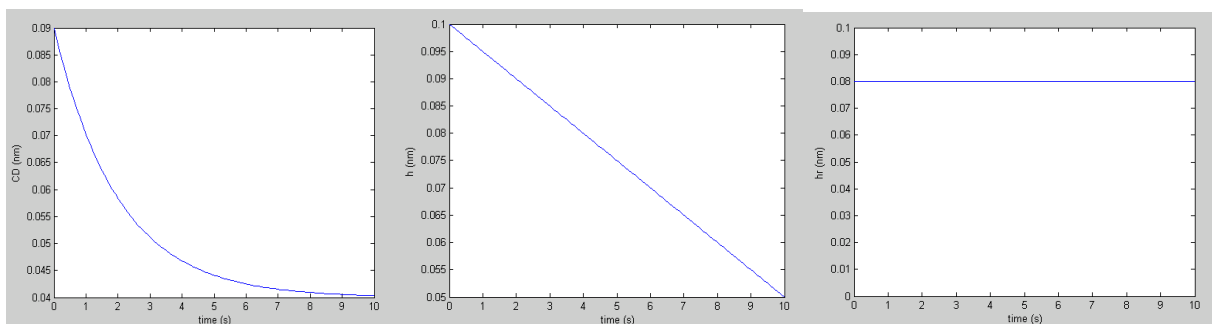


Figure III-16. Evolution au cours du temps des paramètres géométriques du profil créneau (CD, h, hr) lors de la simulation d'un procédé de resist trim

Le procédé théorique dure 10s et permet la génération de 50 signatures théoriques, puisque la période d'acquisition théorique choisie est 200 ms, s'approchant ainsi des conditions expérimentales (500ms).

L'objectif de notre étude est de retrouver cette évolution des paramètres par scattérométrie. Deux RN ont ainsi été entraînés suivant les caractéristiques indiquées dans la



Table III-4. Précisons que le nombre d'échantillons est plus important pour le  $RN_B$  par rapport au  $RN_A$  pour compenser l'augmentation de la plage de variation. Pour chaque étude, les cas avec respectivement 16 et 32 longueurs d'ondes ont été considérés. Les PMC (Perceptrons MultiCouche - cf. II.1.6.2) utilisés pour cette étude sont composés d'une seule couche cachée de 15 neurones. Le PMC est entraîné avec des données bruitées ( $\sigma_e=0,0033$ ).

	$RN_A$	$RN_B$
Nombre d'échantillons	1000	3000
domaine pour CD (en nm)	35-100	10-110
domaine pour h (en nm)	40-110	30-130
domaine pour hr (en nm)	70-90	60-110
Type de bruit Bruit	Gaussien $\sigma_e=0,0033$	Gaussien $\sigma_e=0,0033$
Nombre d'entrées	16 et 32	16 et 32

Table III-4. Caractéristiques des deux RN effectuées pour caractériser l'évolution temporelle d'un procédé théorique de resist trimming

Les performances du RN sur les valeurs de tests (15% du corpus d'apprentissage) ont été évaluées après chacun des 4 entraînements. Les erreurs  $E_{param}^{test}$  et  $E^{test}$  (cf. Equation II.15 et II.16) sont relevées dans la Table III-5.

	$RN_{A16}$	$RN_{A32}$	$RN_{B16}$	$RN_{B32}$
Nombre de longueur d'onde du spectre	16	32	16	32
$E_{CD}^{test}$ (nm)	0,43	0,41	0,76	0,74
$E_h^{test}$ (nm)	0,36	0,37	0,62	0,79
$E_{hr}^{test}$ (nm)	0,19	0,17	0,38	0,36
$E^{test}$ (nm)	0,34	0,33	0,60	0,65

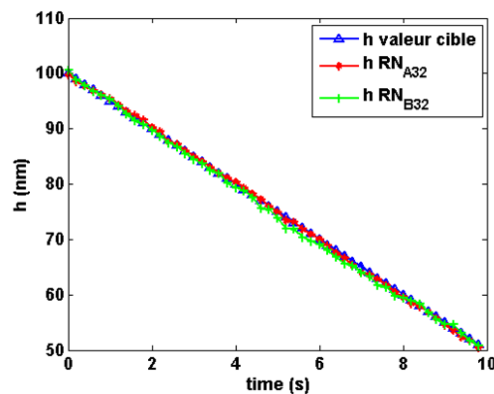
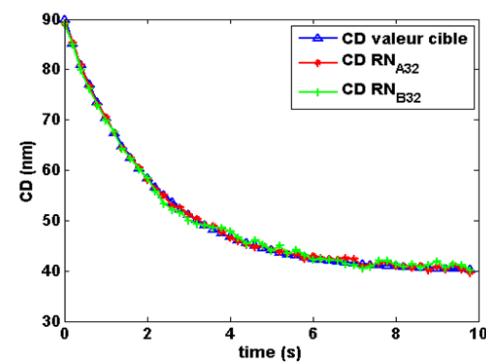
Table III-5. Résultats des performances des différents RN utilisés sur les valeurs de tests obtenues pour les paramètres géométriques (CD, h, hr)

Il faut noter en premier lieu la bonne performance obtenue par les 4 RN. En effet, les résultats sont tous en dessous du nanomètre. Rappelons qu'il s'agit ici de l'erreur minimale atteignable lorsque toutes les hypothèses sont vérifiées.

On décèle ensuite une faible différence de performances entre les RN à 16 et 32 entrées ; ce qui nous amène à penser qu'une signature composée de 16 longueurs d'onde est suffisante pour ce cas.

Puis, comme on pouvait s'y attendre, les performances obtenues dans l'étude avec un plus large domaine de variation des paramètres sont légèrement moins bonnes; la fonction à modéliser étant potentiellement plus complexe.

Il est alors possible de tracer en fonction du temps l'évolution des paramètres calculés par les  $RN_{A32}$  et  $RN_{B32}$  à partir des signatures dynamiques simulées. Le choix du spectre composé de 32 longueurs d'onde est guidé par le souci de se placer dans les mêmes conditions que le cas expérimental. La Figure III-17 représente les résultats obtenus sur chacun des paramètres étudiés.



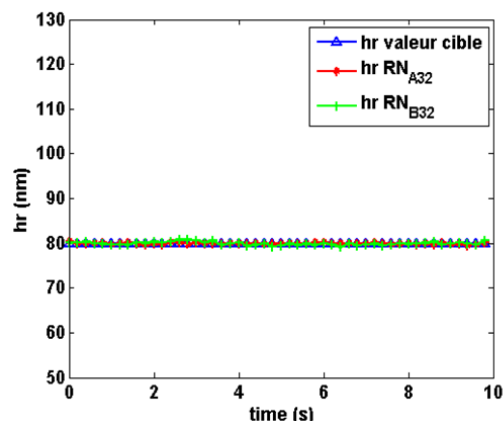


Figure III-17. Caractérisation dynamique obtenue par RN pour un profil théorique défini par les trois paramètres ( $CD$ ,  $h$ ,  $hr$ ), en considérant un spectre de 32 longueurs d'ondes et un PMC entraîné respectivement avec les plages définies dans la table III.4

On note une très bonne cohérence entre les résultats théoriques et ceux provenant des RNs sur chacun des trois paramètres géométriques pour les deux études menées.

On peut conclure que le RN permet de reproduire parfaitement l'évolution en fonction du temps d'un profil théorique simple à trois paramètres à partir des signatures simulées théoriquement. Afin de pouvoir généraliser ces résultats, il convient de considérer des cas plus complexes proches de la réalité, notamment un cas expérimental.

### III.2.4 Etude expérimentale de suivi en temps réel de resist trimming

Le procédé de resist trimming, utilisé dans ce paragraphe a été réalisé au LTM sur la plateforme Centura5300 de la société Applied Materials. Cette plateforme est également équipée d'un ellipsomètre in situ, qui nous a servi pour la caractérisation en temps réel. Le dispositif expérimental sera présenté puis on s'intéressera aux résultats obtenus dans le cadre d'une caractérisation par RN.

#### III.2.4.1 Dispositif expérimental

La plateforme Centura5300 (Figure III-18) est constituée des éléments suivants :

- 2 sas de chargement permettant chacun de charger et de décharger jusqu'à 25 substrats de 300 mm de diamètre.

- une chambre de transfert sous vide munie d'un bras robotisé pour le transfert des plaques d'une chambre à l'autre. Le vide dans la chambre de transfert et dans la chambre de chargement est assuré par une pompe primaire.

- une chambre d'alignement équipée d'un laser afin d'orienter les substrats vis-à-vis de leur encoche (« notch »).

- une chambre d'analyse de surface par XPS (X-ray photoelectron spectroscopy) permettant des caractérisations quasi in-situ sans que le substrat ne soit exposé à une atmosphère oxydante

- 2 réacteurs de gravure: réacteur A et réacteur B

- un ellipsomètre Jobin-Yvon UVISEL MWL-32 pour les mesures in-situ

- un système de détection de fin d'attaque: « l'EyeD » de Verify Instrument. Celui-ci combine les techniques d'interférométrie et d'émission optique.

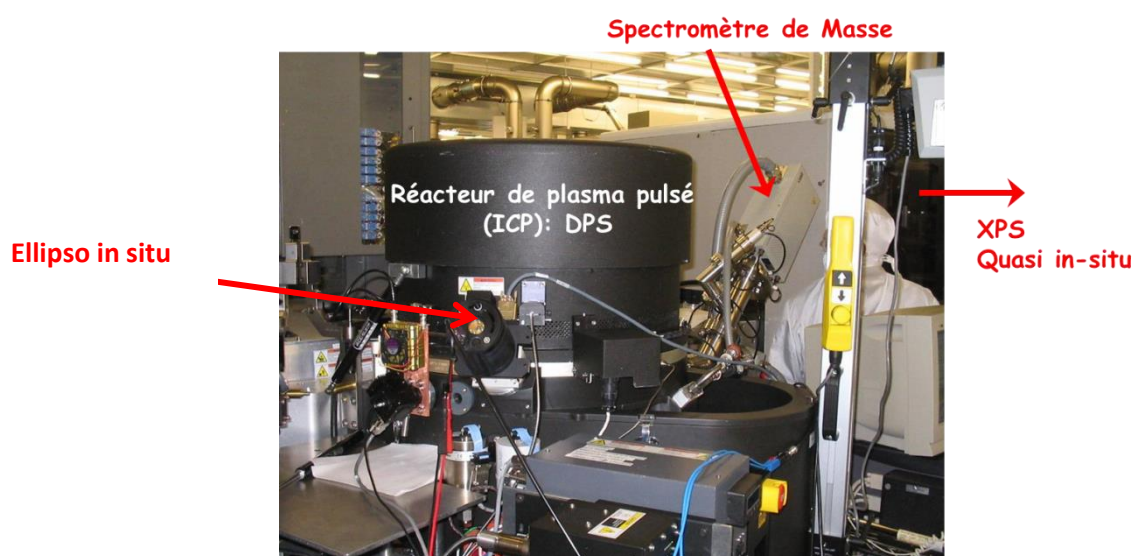


Figure III-18. Présentation de la plateforme de gravure Centura 5300

Le réacteur DPS («Decoupled Plasma Source») est composé d'un réacteur à couplage inductif (ICP pour «inductively Coupled Plasma»). Le principe de fonctionnement de ce réacteur est présenté dans la Figure III-19. Dans le réacteur DPS, une antenne enroulée autour d'un dôme cylindrique est alimentée par un générateur radiofréquence via une boîte d'accord. Celle-ci permet, par le biais d'un système d'asservissement, de maintenir l'impédance vue par les générateurs à  $50\Omega$ . Le générateur permet de délivrer une puissance source comprise entre

200 et 1500W, à une fréquence de 12.56 MHz. La puissance source générée entraîne la création du plasma à l'aide d'un couplage inductif de haute densité (de  $10^{11}$  à  $10^{12}$  ions.cm<sup>-3</sup>). Cette puissance permet également de contrôler la densité d'électrons dans l'enceinte et le flux ionique atteignant le substrat.

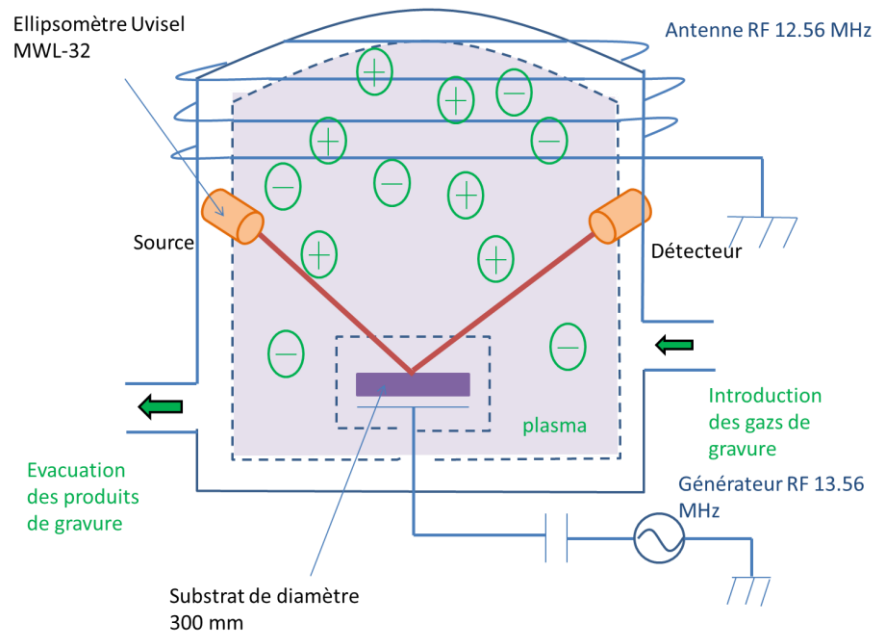


Figure III-19. Schéma représentatif d'un réacteur de gravure DPS

Un second générateur radiofréquence est relié au porte substrat (électrode) via une seconde boîte d'accord. Ce générateur délivre une puissance d'autopolarisation comprise entre 0 et 250W à une fréquence de 13.56MHz. La particularité de ce générateur RF réside dans le fait qu'il est couplé de manière capacitive au substrat permettant ainsi de contrôler l'énergie des ions bombardant le substrat. Par ailleurs, on peut noter que les générateurs de puissance RF qui fournissent la puissance source et la puissance d'autopolarisation sont calibrés à des fréquences différentes afin d'éviter tout phénomène d'interférence.

L'ellipsomètre utilisé permettant le suivi en temps réel du procédé de gravure est un ellipsomètre UVISEL MWL32 du groupe Horiba Jobin Yvon. Cet ellipsomètre est positionné dans la chambre de gravure DPS5300 et permet d'effectuer des mesures in-situ. Il s'agit d'un ellipsomètre à modulation de phase (cf chapitre I.2.2).

Il est notamment composé :

- d'un modulateur photoélastique (50KHz) stabilisé thermiquement.

- d'une source lumineuse à arc xénon de puissance 75W. Elle est maintenue fixe de façon à produire un faisceau incident sur l'échantillon de  $62^\circ$ . La taille du spot est de l'ordre de 2 mm x 3 mm.
- D'un système d'acquisition multi longueur d'onde MWL32 (32 longueurs d'onde en parallèle) dans la gamme 250-826 nm avec une vitesse d'acquisition de 2 signatures par seconde en mode dynamique (soit une période d'acquisition de 500ms). Il est également possible d'effectuer des mesures en statique, pour mesurer des signatures sur la gamme 250-826 nm avec un pas d'échantillonnage plus petit, fixé par l'utilisateur.

### III.2.4.2 Résultats scattérométriques

Afin de pouvoir valider la méthode neuronale en dynamique, il est nécessaire de confronter les résultats obtenus à une technique de caractérisation morphologique. Le protocole expérimental suivi est détaillé dans la Figure III-20.

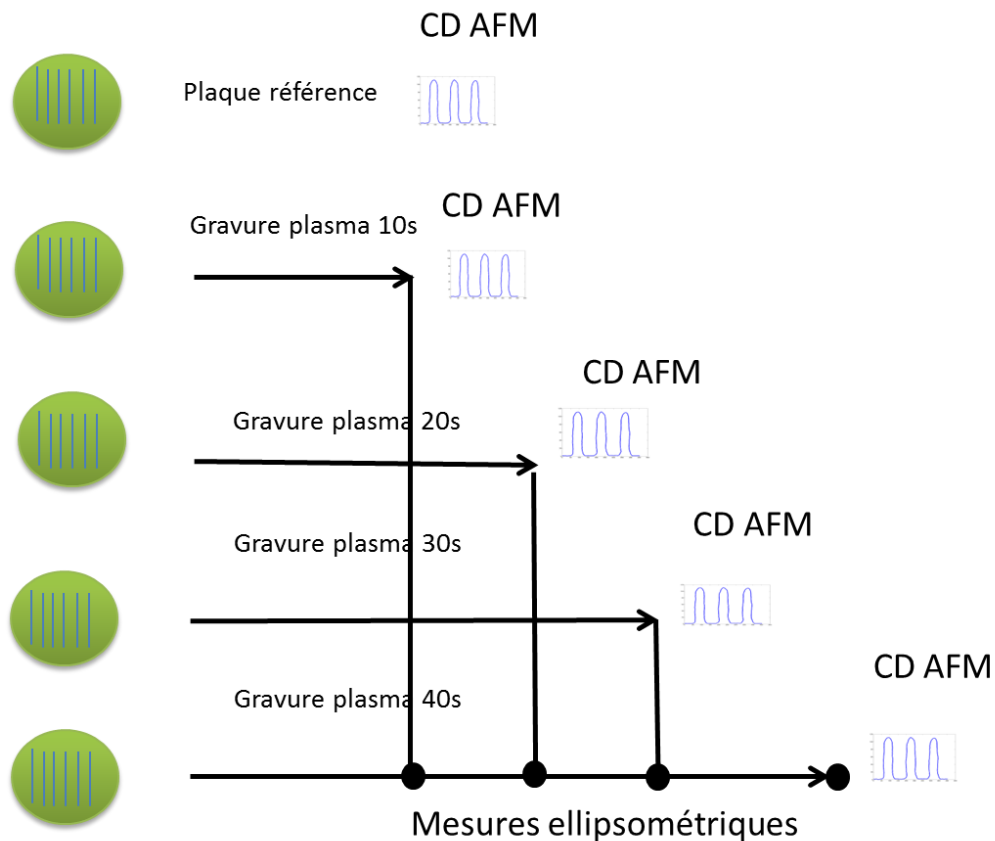


Figure III-20. Protocole expérimental de validation de la scattérométrie dynamique neuronale comprenant des mesures CD AFM systématiques

5 plaques appartenant à un même lot sont gravées successivement dans les mêmes conditions. La première est considérée comme la plaque témoin et les 4 autres ont été gravées à des temps différents (10s, 20s, 30s, 40s). A la fin de chaque étape de gravure, une mesure CDAFM (cf. paragraphe I.1.2.2) est conduite afin d'analyser le profil obtenu. Une mesure scattérométrique en dynamique est réalisée sur la dernière plaque (gravure pendant 40s) et comparée ensuite avec les mesures obtenues par caractérisation CD AFM.

Avant chaque étape de gravure, le réacteur DPS est nettoyé par l'utilisation d'un plasma à base d'un gaz SF<sub>6</sub>/O<sub>2</sub> [ULA02] afin de se mettre toujours dans les mêmes conditions de gravure.

Le plasma utilisé pour la gravure des résines est le plasma HBr/O<sub>2</sub>. Cette chimie est largement utilisée dans les procédés de « resist trimming » [PAR04\*, PAR05]. Les conditions expérimentales utilisées lors de la gravure sont: HBr 100 sccm/O<sub>2</sub> 100sccm, P<sub>s</sub>=300 W, P<sub>bias</sub>=0 W, pression=10 mT. P<sub>s</sub> et P<sub>bias</sub> représentent respectivement la puissance RF injectée dans l'antenne source et la puissance RF du générateur de polarisation du substrat.

La Figure III-21 présente les résultats de caractérisation par CD AFM obtenus aux différents temps de gravure définis dans la Figure III-20. Le profil présente des bords arrondis et des rugosités sur les flancs. Par conséquent, le profil géométrique sélectionné pour l'étude scattérométrique est le profil créneau à bord arrondi (Figure III-21).

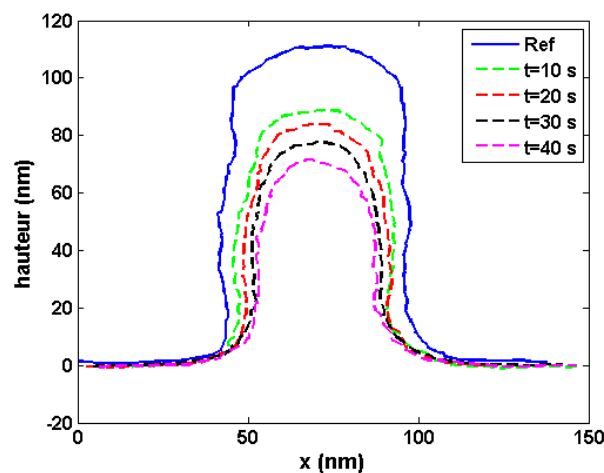


Figure III-21. Evolution du profil des lignes de la résine IM5010 lors de l'étape de gravure par plasma Hbr/O<sub>2</sub> obtenue par CD-AFM

Au cours de la gravure, nous avons considéré que la couche de BARC pouvait partiellement être gravée par le plasma. Ainsi, un paramètre supplémentaire  $hr_{sup}$  est introduit dans le modèle afin de prendre en compte ce phénomène. Le profil géométrique considéré en

cours de gravure (Figure III-22 droit) a donc été modélisé par 5 paramètres définis par  $CD_{top}$ ,  $CD_{bottom}$ ,  $h$ ,  $hr_{inf}$ ,  $hr_{sup}$ .

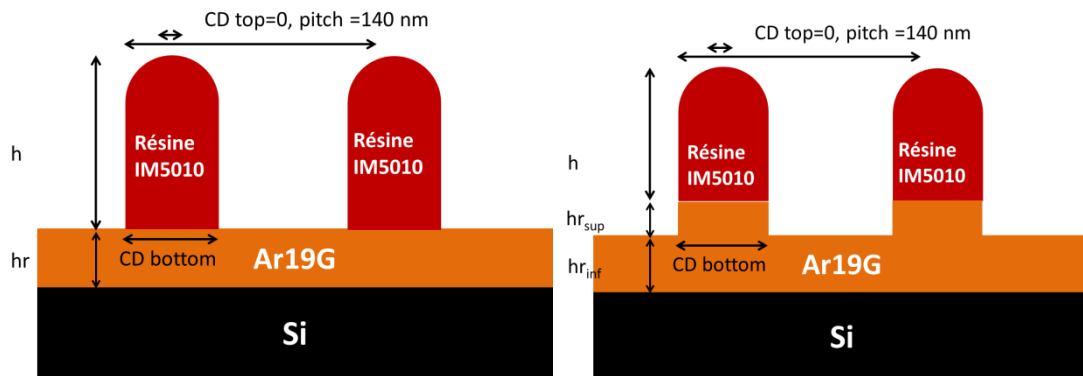


Figure III-22. (A gauche) Profil géométrique crête à bord arrondi défini par 4 paramètres géométriques ( $CD_{top}$ ,  $CD_{bottom}$ ,  $h$ ,  $hr$ ) correspondant à la structure avant gravure et (à droite) profil géométrique défini par 5 paramètres utilisés pour la modélisation de l'échantillon en cours de gravure ( $CD_{top}$ ,  $CD_{bottom}$ ,  $h$ ,  $hr_{sup}$ ,  $hr_{inf}$ ).

Notons cependant que deux paramètres ne font pas partie des sorties du RN:  $CD_{top}$  fixé à 0 et le paramètre  $hr_{sup}$  puisque  $hr_{inf}$  et  $hr_{sup}$  sont reliés par la formule  $hr_{inf} + hr_{sup} = hr_{totale}$ . En effet sous la couche de résine la hauteur de BARC est constante à sa valeur fixée initialement soit 83 nm. Ainsi seule la connaissance de  $hr_{inf}$  (par exemple) est nécessaire pour la définition du profil. Il faut noter que la hauteur déterminée par le CD-AFM correspond à la somme de la hauteur de résine et de BARC gravé soit  $h_{totale} = h + hr_{sup}$ .

Par ailleurs, le profil déterminé par le CD-AFM (Figure III-21) nous incite à penser que la longueur du plateau sur les sillons est extrêmement restreinte et ce, dès la première mesure. A cela notons que la scattérométrie est peu sensible au paramètre  $CD_{top}$  (cf chapitre II.4.2). Ainsi, le  $CD_{top}$  sera fixé à 0 dans la suite de l'étude. Le profil est donc parfaitement défini par la connaissance des 3 paramètres  $CD_{bottom}$ ,  $h$  et  $hr_{inf}$ .

En métrologie, il est difficile d'identifier un appareil de mesure qui puisse être qualifié de référence. Toutefois les outils de mesure à disposition, notamment le CD-AFM, sont fiables et permettent de détecter si les résultats obtenus avec d'autres techniques de métrologie sont proches ou non. On s'attend donc à obtenir une correspondance entre les valeurs CD-AFM et les valeurs calculées par RN, ceci pour les paramètres  $CD_{bottom}$  et  $h_{totale}$ .

Le RN utilisé est composé d'une couche cachée de 15 neurones entraîné avec 9000 échantillons différents. Cette architecture est le fruit d'une optimisation telle qu'exposée dans le chapitre II. Le spectre utilisé est composé de 32 longueurs d'onde (imposé par



l'appareillage de mesure) régulièrement espacées dans l'intervalle 1.5 eV à 6.5 eV. Les plages de variation utilisées lors de l'entraînement sont :

$$20 \text{ nm} < \text{CD}_{\text{bottom}} < 70 \text{ nm}$$

$$40 \text{ nm} < h < 130 \text{ nm}$$

$$60 \text{ nm} < \text{hr}_{\text{inf}} < 83 \text{ nm}$$

La Figure III-23 (en vert) montre l'évolution temporelle des 3 paramètres géométriques ( $\text{CD}_{\text{bottom}}$ ,  $h$ ,  $\text{hr}_{\text{inf}}$ ) calculés par RN au cours du procédé de gravure avec plasma  $\text{Hbr}/\text{O}_2$  et la Figure III-24 celle des paramètres reconstruits ( $\text{hr}_{\text{sup}}$ ,  $\text{hr}_{\text{totale}}$ ,  $h_{\text{totale}}$ ). Les résultats des mesures du CD AFM concernant pour les paramètres  $\text{CD}_{\text{bottom}}$  et  $h_{\text{totale}}$  ont été rajoutés sur la même figure (points bleus). Rappelons que ces deux paramètres sont les seuls accessibles par caractérisation CDAFM et seront donc retenus pour effectuer la comparaison. En effet, le CD-AFM ne permet pas d'obtenir, de manière robuste et précise, la hauteur  $h$  ni  $\text{hr}_{\text{inf}}$ .

Le plasma est déclenché à l'instant  $t=38\text{s}$ . Les différents temps de gravure (10s,20s,30s,40s) indiqués dans le protocole (Figure III-21) sont décomptés à partir de l'instant où le plasma se déclenche. Notons également que nous avons mentionné sur les figures les vitesses de gravure des différents matériaux calculées par intervalle de 10s.

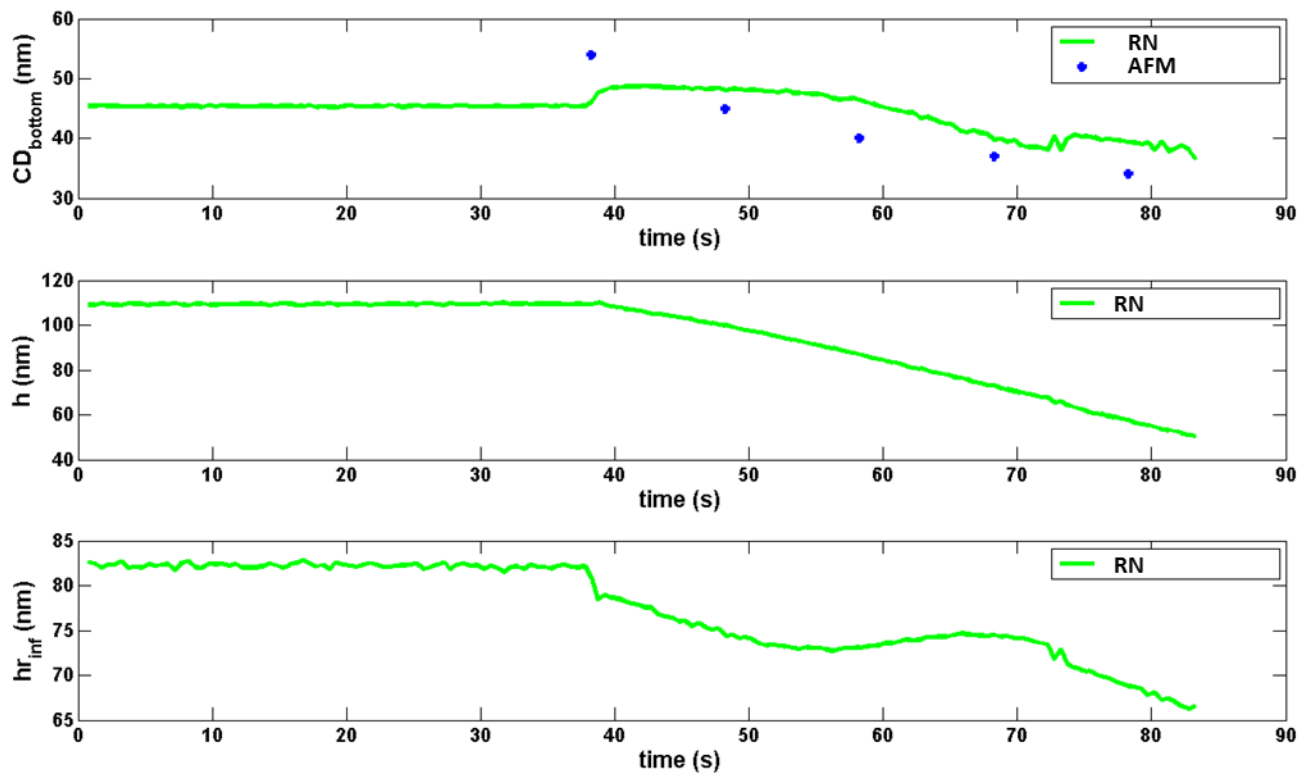


Figure III-23. Evolution des paramètres géométriques du profil de ligne calculé par le RN ( $CD_{bottom}$ , hauteur de résine  $h$  et hauteur de BARC non gravée  $hr_{inf}$ )

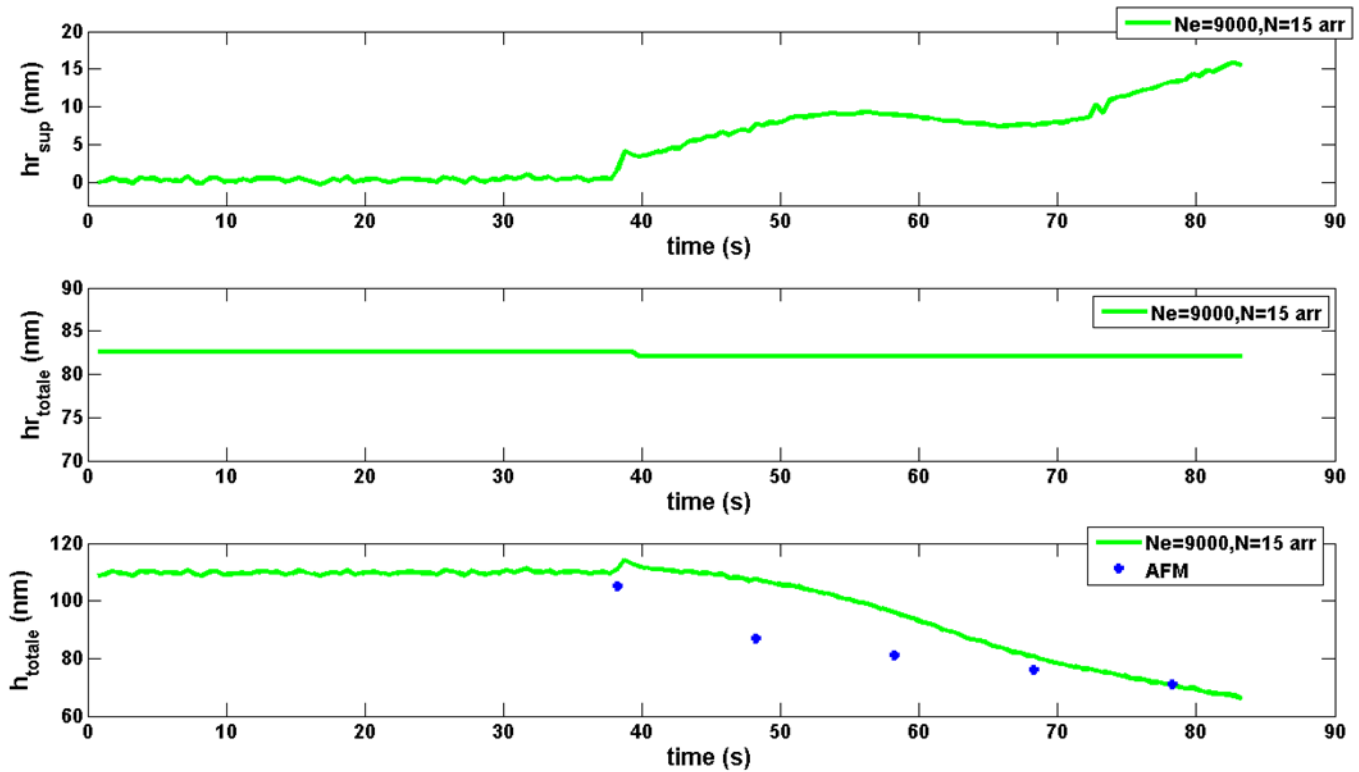


Figure III-24. Evolution des paramètres géométriques reconstruit du profil de ligne (hauteur du pied de barc  $hr_{sup}$ , hauteur barc totale  $hr_{totale}$ , hauteur totale du motif  $h_{totale}$ )

D'un point de vue absolu, on peut constater que les valeurs obtenues par RN sont éloignées des valeurs du CD-AFM. Ceci est confirmé par les valeurs de la Table III-6 où l'on observe des écarts  $\Delta$  importants entre les deux techniques, notamment pour le paramètre  $h_{totale}$  à 10s ( $\Delta=20,6\text{nm}$ ). Les écarts notés sont supérieurs aux erreurs de mesures nanométriques que l'on peut tolérer par chacune des deux méthodes.

D'un point de vue relatif, concernant l'allure globale des paramètres calculés par RN, on observe une évolution cohérente des paramètres pour un procédé de «resist trimming». Les valeurs des paramètres sont stables jusqu'à l'instant  $t=38,25\text{s}$ , instant à partir duquel le plasma est déclenché. La hauteur ( $h$ ) de résine est ensuite réduite ainsi que sa largeur ( $CD_{bottom}$ ) en fonction du temps. Le BARC ( $hr_{inf}$ ) semble également affecté par le procédé et sa hauteur est réduite.

Temps (s)	CD <sub>bottom</sub> (nm)			h <sub>totale</sub> (nm)		
	CD AFM	RN	$\Delta$	CD AFM	RN	$\Delta$
0	54	46,02	7,98	105	111,1	6,1
10	45	48,07	3,07	87	107,6	20,6
20	40	46,28	6,28	81	95,81	14,81
30	37	39,71	2,71	76	80,76	4,76
40	34	39,35	5,35	71	71	0

Table III-6. Comparaison des résultats obtenue par RN et par mesures CD AFM pour le CD et la hauteur totale obtenue pour un procédé de réduction de côte de résine IM5010 par plasma Hbr/O2

Pour les paramètres reconstruits, on observe l'apparition d'un pied de BARC ( $h_{r_{sup}}$ ) tout en gardant une hauteur totale de BARC constante ( $h_{r_{totale}}$ ). La hauteur totale ( $h_{totale}$ ) suit l'évolution des paramètres présentés précédemment et diminue au cours du temps.

Au début de la gravure, le CD<sub>bottom</sub> obtenu par RN reste constant. Le régime stabilisé n'est pas encore atteint au cours de cette période et ceci peut expliquer l'absence de gravure. On note de manière globale que les vitesses de gravures de résine et du BARC sont cohérentes avec les résultats attendus [SIN02, PAR04, PAR05].

Notons aussi que les indices de la résine IM5010, peuvent varier au cours de la gravure, notamment à l'allumage du plasma [ELK10]. Dans notre modèle, l'indice de la résine est supposé le même tout au long de la gravure.

La Figure III-25 présente l'évolution au cours du temps de l'erreur rms calculée systématiquement sur les signatures reconstruites. Le suivi de cette valeur donne une idée sur la précision des résultats calculés par RN. La valeur d'erreur reste raisonnable et stable avant le début de la gravure. Ensuite, l'erreur augmente une première fois entre l'intervalle 40 s et 60 s atteignant une valeur maximale de 0,16 à 48 s. Une deuxième augmentation est observée entre l'instant 60 s et 85 s avec une valeur maximale de 0,18 à 76s. Ce comportement de l'erreur traduit partiellement les écarts notés avec le CD-AFM.

Cette évolution peut aussi être expliquée par plusieurs facteurs. Une hypothèse consiste à envisager un changement de forme du profil au cours du procédé. Cependant, le suivi de l'évolution par CD-AFM (Figure III-21) permet d'écarter cette hypothèse puisque le profil semble conserver la forme arrondie tout au long du procédé. On devrait donc s'attendre à ce que l'erreur rms reste constante tout au long du process. Une autre hypothèse plausible pour expliquer la première augmentation de l'erreur réside dans le changement des indices des matériaux, respectivement de la résine et du BARC .

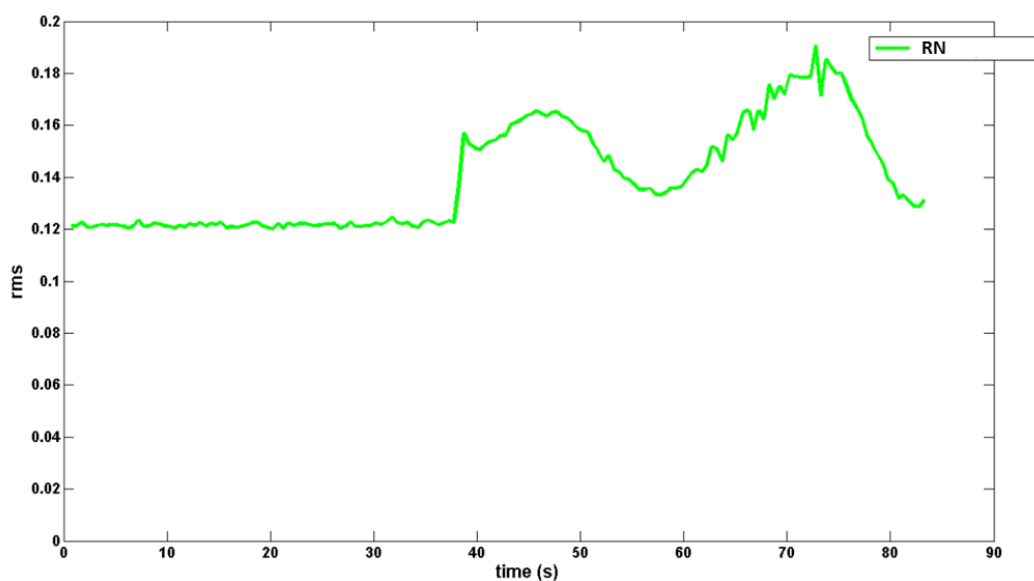


Figure III-25. Evolution de l'erreur rms sur les signatures reconstruites pour le profil de ligne IM5010 sur BARC

Les écarts avec la méthode morphologique CDAFM suggère que les conditions de l'utilisation de la méthode scattérométrique doivent être mises en cause. Il est ainsi nécessaire de tenir compte de la variation d'indice de l'IM5010 et du BARC au cours du temps dans le modèle utilisé ; ce qui n'a pas été fait durant cette thèse faute de temps.

Rappelons au cours de cette analyse que l'outil AFM a été considéré comme outil de référence auquel on compare nos résultats obtenus par scattérométrie. Or nous savons (cf. II.4.2) que cet outil comporte des sources d'erreur et qu'il peut être également mis en cause dans la démarche utilisée dans cette étude. Les écarts obtenus reflètent donc l'influence de l'erreur entraînée par le choix de l'AFM comme outil de référence.

D'autre part, on note un écart important ( $\Delta CD=8$  nm et  $\Delta h_{totale}=6$  nm) entre les deux outils de mesure à l'instant  $t=38s$ , soit avant que le plasma se déclenche. Ceci montre que le changement d'indice des matériaux n'est pas le seul élément responsable de l'écart noté. Des

sources d'erreur peuvent aussi provenir du signal ellipsométrique acquis. En effet, des pertes de signal au-delà de 5,5eV ont été relevées sur l'appareil (fibre qui s'abîme au cours du temps) pouvant avoir des conséquences sur le signal, surtout s'il s'agit des informations dans l'UV. Une erreur sur la signature ellipsométrique induira à son tour des variations sur les résultats scattérométriques. Il aurait été nécessaire de traiter le problème sans considérer cette partie du spectre ellipsométrique. (Considérer un spectre de 1.5 à 5.5eV uniquement).

Pour mieux cerner la source du problème, il aurait été nécessaire de traiter le même cas avec la méthode des bibliothèques, déjà validée en dynamique. Ceci apporterait plus de lumière sur les sources éventuelles d'erreur et permettra de mieux évaluer l'outil scattérométrique neuronal en dynamique. Ce travail n'a pas pu être effectué par manque de temps. De même, il aurait été nécessaire de réaliser d'autres expériences en changeant les conditions du plasma, mais ceci n'a pas pu être réalisé en raison de pannes récurrentes sur l'équipement. Les résultats obtenus peuvent donc encore être améliorés pour obtenir une caractérisation plus précise par RN au cours du temps.

### **III.3 Conclusion**

Ce chapitre a permis l'étude de l'outil scattérométrique neuronal pour le suivi in situ de procédé de gravure. L'utilisation de bibliothèques et de réseaux de neurones, a été comparée sur une étude statique et a montré de très bonnes performances. Ces deux méthodes ont ainsi prouvé qu'elles pouvaient répondre aux contraintes du temps réel (sans changement de forme de profil pour la bibliothèque). La validation de la voie neuronale a ensuite été menée sur une étude théorique puis sur un cas expérimental. Les résultats obtenus semblent montrer que les RN sont capables de suivre l'évolution d'un procédé de gravure. Ils nécessitent cependant une certaine expérience pour définir l'architecture optimale.

Ces travaux ont permis la mise en place des fondements de la caractérisation scattérométrique par voie neuronale pour une application temps réel. Des travaux futurs permettront de prendre en compte l'ensemble des contraintes du temps réel sur les matériaux et d'apprécier ainsi les potentialités de l'outil neuronal pour une caractérisation précise des motifs en temps réel. Par exemple, il serait intéressant de considérer l'évolution de l'indice de résine 193 et du BARC dans le modèle utilisé. De plus, des outils de caractérisation plus précis, comme le FIB-STEM, peuvent être mis en place pour obtenir un degré de précision maximale.

### **III.4 Bibliographie**

- [WIN83] Winters H.F., Coburn J.W., Chuang T.J., “Surface processes in plasma-assisted etching environments”. *J. Vac. Sci. Technol. B*, 1:496, 1983
- [ELK09] El Kodadi M., Soulan S., Besacier M., and Schiavone P. , “Real time scatterometry for profile control during resist trimming process”. *J. Vac. Sci. Technol. B*, 27(6): 3232-3237, 2009.
- [ELK09\*] El Kodadi M., Soulan, S., Besacier M., and Schiavone P. “Resist trimming etch process control using dynamic scatterometry”. *Microelectronic Engineering*, 86(4) : 1040-1042., 2009.
- [ELK10] El Kodadi M., “Développement des techniques de scattérométrie en temps réel pour le suivi des procédés de gravure plasmas”, PhD thesis, Université de Grenoble, 2010.
- [GER12] Gereige I., Robert S., Eid J. «Automatic detection of photoresist residual layer in lithography using a neural classification approach ». *Microelectronic Engineering*, 97:29-32, 2012
- [HAR89] Harper J.M.E., “Ion beam etching, in *Plasma Etching*”, by Manos D.M. and Flamm D.L. (eds.), Academic Press, San Diego,391-423, 1989
- [HAZ03] Hazart J., Grand G., Thony P., Herisson D., Garcia S., Lartigue O., « Spectroscopic ellipsometric scatterometry: sources of errors in critical dimension control », *Proceedings of SPIE*, vol 5041, 2003
- [NIU99] Niu X., Jakatdar N., Bao J., Spanos C. and Yedur S., « Specular spectroscopic scatterometry in DUV lithography», *SPIE* , vol 3677, 1999.
- [PAR04] Pargon E., “Analyse des mécanismes mis en jeu lors de l’élaboration par gravure plasma de structures de dimension déca-nanométriques : Application au transistor CMOS ultime », PhD Thesis, Université Joseph Fourier, 2004.
- [PAR04\*] Pargon E., Joubert O., Xu S. and Lill T., « Characterization of resist trimming processes by quasi in situ x ray photoelectron spectroscopy.», *J. Vac. Sci. Technol. B*, 22:1869-1879, 2004.

- [PAR05]** Pargon E., Joubert O., Chevolleau T., Cunge G., Xu S., and Lill T., “Mass spectrometry studies of resist trimming processes in Hbr/O<sub>2</sub> and Cl<sub>2</sub>/O<sub>2</sub> chemistries.” J. Vac. Sci. Technol. B., 23:103-112, 2005.
- [SIG69]** Sigmund P., “Theory of sputtering. i. sputtering yield of amorphous and polycrystalline targets”, J. Phy. Rev, 184:383-416, 1969.
- [SIN02]** Sin C.-Y. and Chen B.-H., “Resist trimming in high-density CF<sub>4</sub>/O<sub>2</sub> plasmas for sub-0.1 μm device fabrication,” J. Vac. Sci. Technol. B 20, 1974, 2002.
- [SOH92]** Sohail S., Naqvi H., Gaspar S., Hickman K., Bishop K., McNeil John R., “Linewidth measurement of gratings on photomasks: a simple technique”, Applied Optics, Vol 31, Issuer 10, pp. 1377-1384 (1992)
- [SOU07]** Soulan S., Besacier M, Schiavone P, Leveder T, “In-line etching process control using dynamic scatterometry“, Proc. SPIE 6617, Modeling Aspects in Optical Metrology, 661713 ,June 18, 2007.
- [SOU07\*]** Soulan S., Besacier M. , Leveder T. , Schiavone P, “Real-time profile shape reconstruction using dynamic scatterometry”. Proc. SPIE 6518, Metrology, Inspection, and Process Control for Microlithography XXI, 65180W, April 05, 2007;
- [SOU08]** Soulan S., “Développement de la scattérométrie dynamique pour le suivi en temps réel de procédés. Application à la microélectronique. ». PhD thesis, Université de Grenoble, 2008.
- [TRO13]** Troscompt N., Besacier M., Saib M., “Assesment of the scatterometry capability to detect an etch process deviation”, Proceedings of SPIE Vol 8789, 878902,2013 .
- [ULA02]** Ullal S.J., Singh J, Daugherty J., Vahedi V. and Aydil E.S.,” Maintaining reproducible plasma reactor wall conditions: SF<sub>6</sub> plasma cleaning of films deposited on chamber walls during Cl<sub>2</sub>/O<sub>2</sub> plasma etching of Si.” J. Vac. Sci. Technol. A, 20:43, 2002.
- [WIN83]** Winters , Harold F., Coburn,J.W., Chuang T.J., “Surface processes in plasma assited etching environements” J. Vac. Sci. Technol. B, 1:496, 1983
- [YEN04]** Yen T. F., Chiu C.M. and Chiu K.-F., “Photo-resist trimming for sub 0.1 μm generation integrated circuit devices by cl<sub>2</sub>/o<sub>2</sub> transform coupled plasma”, J. Microelectronic Engineering, 75:201–209, 2004.





---

## Conclusion

Ces travaux de thèse s'inscrivent dans le cadre du développement d'un outil de métrologie par voie optique pour le suivi en temps réel de procédés dynamiques. L'application visée ici concerne la microélectronique et les procédés de gravure utilisés lors de la fabrication des structures de taille nanométrique. Des précédents travaux, utilisant la méthode des bibliothèques, avaient montré la compatibilité de cette technique avec les contraintes des mesures en temps réel. En effet, la scattérométrie est reconnue pour être non destructive, non invasive et applicable in-situ. La méthode des bibliothèques possède néanmoins des contraintes en terme de temps de construction, d'espace mémoire et de flexibilité dans l'évolution du profil. Il est donc apparu nécessaire de confronter cette technique avec un outil mathématique prometteur pour la résolution de ce type de problématique : Les réseaux de neurones artificiels.

Les réseaux de neurones artificiels ont déjà été utilisés pour une caractérisation en statique de réseaux périodiques de taille nanométrique et ces travaux de thèse ont permis d'évaluer leur potentiel pour la scattérométrie dynamique. Les résultats obtenus incitent à poursuivre les recherches dans cette voie. Ce manuscrit, composé en trois parties, rappelle le principe de cette méthode neuronale et expose comment elle peut être mise à profit dans le cadre de la scattérométrie dynamique tout en confrontant les résultats avec ceux obtenus par la méthode des bibliothèques.

Dans la première partie, les différentes techniques de métrologie utilisées pour caractériser des composants nanométriques ont été présentées. On peut distinguer 2 grandes familles de techniques. Les techniques morphologiques, effectuant une mesure directe des profils locaux, sont efficaces pour une caractérisation statique mais ne sont pas adaptées pour des mesures à l'intérieur même des équipements spécifiques des procédés visés. La scattérométrie, dont le principe repose sur une mesure indirecte sur des motifs uniquement périodiques, possède quant à elle des atouts pour le suivi temps réel. Différents types de montages expérimentaux (ellipsométrique ou réflectométrique) permettent de mesurer la réponse optique, véritable signature de l'échantillon, à partir de laquelle sont extraites les informations structurales correspondantes. Ainsi, différents algorithmes de résolution de problème inverse existent. Les méthodes d'optimisation classique permettant de comparer

---

itérativement la mesure expérimentale et les résultats issus d'un modèle paramétrique ne sont pas adaptées au temps réel principalement à cause de leur variabilité de convergence et leur lenteur d'utilisation. Les outils fondés sur la méthode des bibliothèques et des réseaux de neurones artificiels, permettent une résolution quasi-instantanée du problème inverse et se présentent donc comme de véritables atouts pour le temps réel. La première partie a ainsi permis d'exposer les différents éléments nécessaires pour comprendre la problématique développée dans cette thèse visant à développer une technique de scattérométrie dynamique neuronale.

La seconde partie a été consacrée aux réseaux de neurones artificiels et plus particulièrement à leur application dans le cadre précis de la scattérométrie statique de réseaux périodiques. Les réseaux de neurones permettent de modéliser le comportement d'un système physique, à condition de disposer de suffisamment d'exemples décrivant ce phénomène. Ils sont composés d'un assemblage de petites entités mathématiques, les neurones, disposés suivant un ordre préétabli, l'architecture, propre à chaque problème. Ils nécessitent deux phases: une étape d'entraînement plus ou moins longue, apprenant itérativement la relation entre les signatures optiques et les paramètres géométriques issus d'un corpus prédéfini et une phase de traitement quasi instantanée, où il associe à une signature inconnue mesurée les paramètres géométriques correspondants. Chaque RN entraîné nécessite l'évaluation des performances pour trouver l'architecture optimale, qui sera choisie en fonction du problème traité. Les résultats obtenus montrent que la scattérométrie neuronale permet d'obtenir des résultats précis et robustes, comparables aux techniques morphologiques et autres algorithmes de résolution de problème inverse. Nous avons également souligné le fait que le choix du modèle neuronal résultant du compromis entre une large gamme d'étude et une bonne précision doit être fixé selon les priorités du technologue.

La dernière partie concernait la validation de la voie neuronale pour le temps réel. Une comparaison du comportement de deux méthodes applicables au temps réel, bibliothèque et RN a été réalisée en mode statique afin de définir leurs spécificités. Les résultats obtenus avec les deux méthodes sont très encourageants. La méthode des bibliothèques résiste bien au bruit de mesure mais nécessite des bibliothèques très grandes pour garantir une bonne précision. Les RN donnent de bonnes performances avec peu d'échantillons mais nécessitent du temps et une certaine expérience pour le choix optimale de l'architecture. Les deux méthodes répondent parfaitement aux exigences du temps réel, mais les réseaux de neurones sont incontournables dans le cas où il faut considérer un changement de type de profil, et donc de

---

représentation des motifs, en cours de procédé. Une étude expérimentale de suivi de procédé de « resist trimming » en temps réel a été menée par voie neuronale. Elle consistait à observer l'évolution d'un motif de résine lors de l'étape de réduction des dimensions. Les résultats ont été comparés avec ceux obtenus par CD-AFM, considéré pour cette étude comme la mesure de référence. Les résultats de comparaison entre les deux méthodes montrent une évolution d'ensemble semblable mais présentent cependant des différences ponctuelles notables, ce qui ne permet pas de tirer encore de conclusions sur la validité de la méthode neuronale pour le suivi en temps réel de procédé. Cependant les allures des évolutions des paramètres mesurés avec la scattérométrie par réseaux de neurones sont physiquement cohérentes et semblent montrer que les RN répondent à toutes les exigences du temps réel et peuvent être appliqués dans ce cadre-là. Cette étude comparative mérite d'être poursuivie en considérant une autre technique de métrologie comme référence. En effet un outil comme le FIB SEM, bien que plus lourd à mettre en place, devrait donner des mesures plus proches du véritable profil et pourrait mettre en avant certaines limitations induites par l'utilisation du CD AFM.

Ces travaux de thèse constituent une première brique dans l'utilisation des RN pour un contrôle in-situ et en temps réel de procédé. Cette méthode a été testée sur un cas proche des contraintes industrielles. Elle est attendue dans le domaine, notamment pour des cas plus complexes, comme le changement de modèle en cours de procédé. Deux perspectives à ce travail peuvent être proposées:

La première consiste à continuer l'évaluation de l'apport des RN en dynamique pour le suivi de procédé en temps réel. En utilisant des outils de caractérisation plus précis comme le FIB SEM, puis en considérant l'évolution de l'indice de résine dans le modèle, des résultats plus précis peuvent être obtenus.

La seconde étude consiste à développer une technique basée sur les RN qui permet de détecter le changement de profil. En effet les RN sont capables de remplir cette tâche à partir de l'analyse de la signature mesurée. Le RN est alors utilisé en mode classifieur, c'est-à-dire qu'il est capable de déterminer la forme géométrique du profil sous test parmi plusieurs modèles utilisés dans la phase d'apprentissage. Une extension de cette étude sera de considérer une caractérisation exclusivement neuronale. Deux RN sont utilisées dans ce cas : un premier est utilisé pour détecter la forme du profil (créneau, trapèze, arrondi..) par classification, le second permettra une caractérisation paramétrique comme exposé dans nos

---

travaux. Ceci fait actuellement l'objet d'une nouvelle thèse qui permettra d'évaluer les apports des RN pour les profils changeant de forme au cours de procédé.

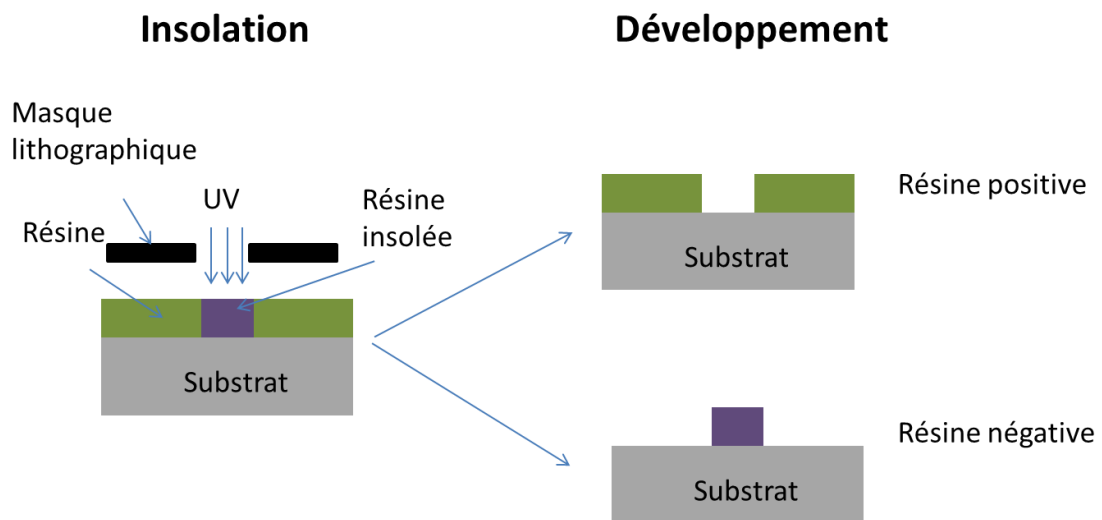




---

## Annexe I. Les différentes techniques de lithographie

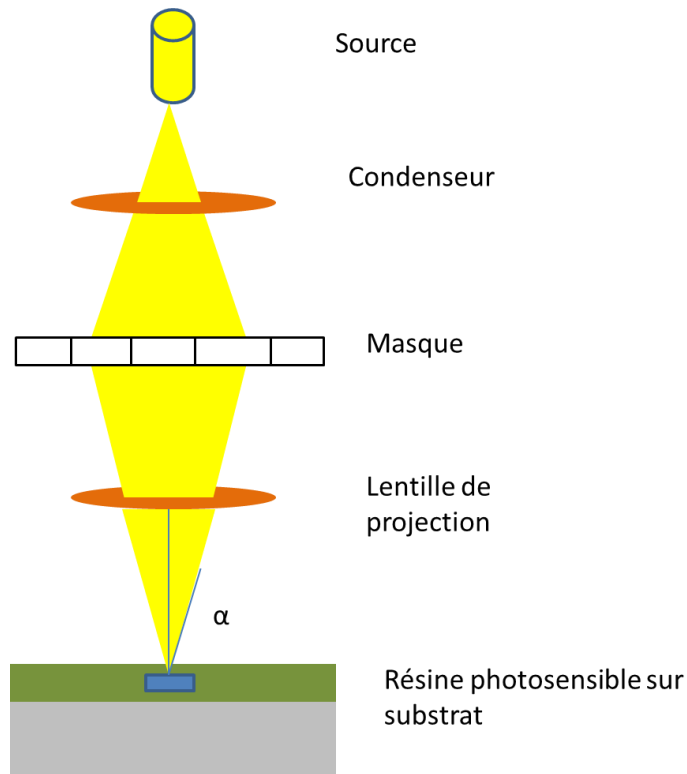
La lithographie optique [THO83] a été un facteur clé pour diminuer la dimension des motifs dans l'industrie microélectronique. Elle consiste à déposer dans un premier temps une couche de polymère photosensible (résine) sur le substrat. Le faisceau lumineux passe à travers un masque pour reproduire le motif sur la résine par insolation, modifiant ainsi les propriétés de solubilité de la résine. Ainsi, après révélation dans un développeur (solution basique), la zone exposée va subsister (résine négative) ou être éliminée (résine positive) (Figure I-1).



*Figure I-1. Insolation et développement d'une résine positive et négative*

La plaque de silicium est alors recouverte des motifs qui sont l'image du masque à travers un système optique permettant de réduire la taille du motif d'un facteur égale à quatre (Figure I-2)





*Figure I-2. Principe de la lithographie optique par projection*

On peut déterminer la résolution maximale de cette technique grâce à l'équation de Rayleigh :

$$R = \frac{k\lambda}{ON}$$

$\lambda$  est la longueur d'onde d'insolation (la longueur d'onde de 193 nm est de nos jours utilisée dans l'industrie de la microélectronique).  $K$  est un facteur dépendant du procédé technologique mis en place et qui varie entre 0,6 et 0,8. Il prend en compte la qualité du masque, des éléments du système optique et de la résine.  $ON$ , l'ouverture numérique de l'optique de projection, est dérivée de l'équation suivante :

$$ON = n \sin(\alpha) \frac{k\lambda}{ON}$$

Où  $\alpha$  est le demi-angle maximum d'ouverture de l'objectif et  $n$  l'indice de réfraction du milieu où se trouve le système optique.

---

L'industrie de la microélectronique a vu se développer des variantes de la lithographie optique : la lithographie en ultraviolet extrême (ou EUV pour le terme Anglais Extreme Ultra Violet) [GWY98], utilisant la longueur d'onde de 13,5 nm.

On peut citer également la lithographie par immersion [LIN06] qui permet d'améliorer la résolution en augmentant l'ouverture numérique par introduction d'un liquide entre le système optique de projection et l'échantillon.

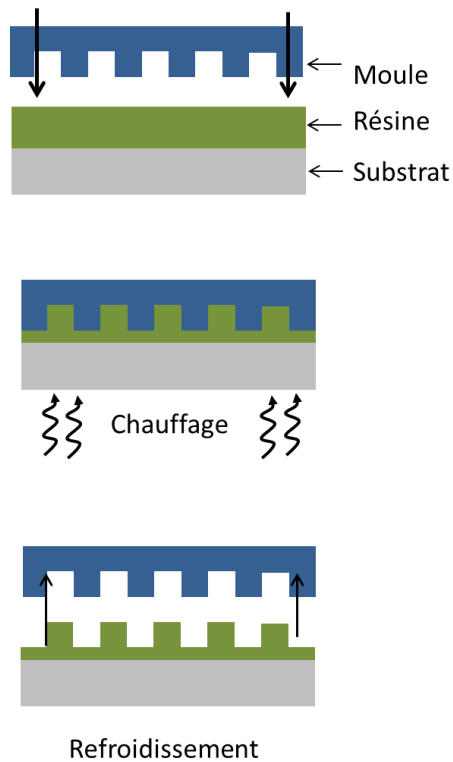
La lithographie par faisceau d'électrons [BEA75] est une technique alternative à la lithographie optique. Elle met en place un faisceau d'électron pour balayer la surface de la résine. Les motifs sont formés sur une résine électrosensible. Cette technique permet d'atteindre une résolution importante car elle n'est pas limitée par la longueur d'onde. Cette technique est adaptée pour de faibles volumes de production en raison de son très faible rendement lié au fait que l'exposition a lieu point par point dans la résine.

Certaines imperfections peuvent apparaître lors de l'étape de lithographie et entraîner quelques défauts sur le motif. Pour l'échantillon traité dans la partie II.4.2, composé d'un empilement de résine IM5010 sur du BARC sur Si, le profil supposé par le fabricant est un profil créneau. Après caractérisation par CDAFM et par scattérométrie, le profil présente des arrondis au sommet (figure II-34). Ceci est dû à l'étape de fabrication technologique par lithographie. En effet, l'interaction photon résine n'est pas parfaite et peut faire apparaître des imperfections comme les arrondis.

La lithographie par nanoimpression [CHO95] est une méthode simple adaptée pour la fabrication de structures nanométrique. Cette méthode consiste à dupliquer des motifs contenus sur un moule dans une couche de résine déposée sur un substrat à graver. (Figure I-3)

Le moule étant fabriqué par les techniques de lithographie standards. Il existe deux grand type de nanoimpression : la nanoimpression thermique et la nanoimpression assistée par ultraviolet (UV-NIL pour l'acronyme anglais Ultra Violet curing Nanoimprint lithography).

Dans la première technique, un moule contenant les motifs à reproduire est pressé dans le polymère chauffé au-delà de sa température de transition visqueuse. Le polymère se déforme bien à cette température car il est visqueux. L'ensemble est refroidi et le moule est retiré, transférant ainsi les motifs dans le polymère. Une couche résiduelle est laissée au fond des motifs pour éviter tout contact entre le moule et le substrat.



*Figure I-3. Principe de la lithographie par nanoimprint thermique*

L'UV-NIL est une évolution de la nanoimpression thermique [COL99]. Un moule transparent et une résine photosensible fluide à température ambiante, sont utilisés. Le moule et la résine sont mis en contact par l'application d'une faible pression. La fluidité de la résine permet à celle-ci de remplir la cavité du moule. La résine est exposée à un faisceau UV à travers le moule, ce qui entraîne sa polymérisation. La couche de résine résiduelle au fond des motifs est éliminée, après démoulage, grâce à une étape de gravure. Le principe de cette technique est détaillé dans la référence [MAS12].

L'avantage de la nanoimpression thermique et qu'elle n'est pas limitée par la diffraction, la diffusion ou l'interférence de l'onde dans le polymère vue qu'elle n'utilise aucun faisceau énergétique.

---

## Bibliographie

- [**BEA75**] Squire D.G., Beasley J.P., “ An electron beam mask maker”, IEEE Trans. on Elec. Dev; 22:409, 1975.
- [**CHO95**] Chou S.Y., Krauss P.R., and Renstrom P.J., « Imprint of sub-25 nm vias and trenches in polymers », Appl. Phy. Lett., 67:3114, 1995.
- [**COL99**] Colburn M., Johnson S., Stewart M., Damle S., Bailey T., Choi B.J., Wedlake M., Michaelson R., Screenivasan S.V., Ekerdt J., and Willson C.G., “Step and flash imprint lithography: a new approach to high-resolution patterning.” Proceedings SPIE, 3676:379-389, 1999.
- [**GWY98**] Gwyn C., Sweeney D., Stulen R. and Attwood D., “Extreme ultraviolet Lithography”, J. Vac. Sci. Technol. B, 16:3142, 1998.
- [**LIN06**] Lin. J. B., “Optical lithography - present and future challenges”, C. R. Physique, 7:856–874, 2006.
- [**MAS12**] Masclaux C., “Etude de la Structuration spontanée de films polymers en nano-impression thermique”, PhD Thesis, Grenoble University, 2012.
- [**THO83**] Thompson L. F., “Introduction to Microlithography”, ACS Symposium , 1983.



---

## Annexe 2. La scattérométrie réflectométrique

### 2.1 La réflectométrie goniométrique

En réflectométrie goniométrique [RAY95], l'intensité de l'onde réfléchie par l'échantillon est mesurée en fonction de l'angle d'incidence  $\theta_i$  dans un ou plusieurs ordres de diffraction.

Le principe repose sur la mesure du coefficient de réflexion en intensité  $R$  qui est le rapport entre l'intensité réfléchie  $I_r$  et l'intensité incidente  $I_0$ :

$$R = \frac{I_r}{I_0} = |r^2|$$

Ce coefficient dépend donc fortement des conditions d'excitation ( $\theta_i, \lambda$ ) et de la nature des matériaux considérés (indices optiques), mais également de la structure (période, forme etc...); Il constitue donc une signature de l'échantillon, de la même façon que  $\psi$  et  $\Delta$  en ellipsométrie. La différence fondamentale entre les 2 mesures concerne la non-prise en compte pour la réflectométrie de la modification de phase induite par la réflexion de l'onde sur l'échantillon.

Cette technique utilise généralement une seule longueur d'onde et un angle d'incidence  $\theta_i$  variable. Un exemple de montage a été utilisé par [ROB02\*]. Une source laser ( $\lambda=670$  nm) traverse différents systèmes optiques dont un polariseur pour le choix de la polarisation TE ou TM. L'onde est ensuite divisée en deux : l'une, servant de référence, sera collectée par un photodétecteur et l'autre constitue le faisceau incident sur l'échantillon. Ce dernier peut tourner autour d'un axe vertical permettant de se positionner dans les conditions d'incidence données. Les efficacités des ondes diffractées (pourcentage d'intensité de l'onde diffractée par rapport à l'onde incidente) sont mesurées par un détecteur pivotant afin de récolter les différents ordres de diffraction.

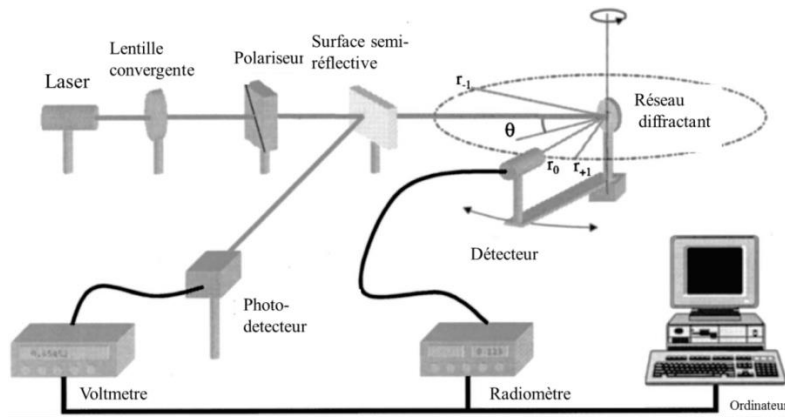
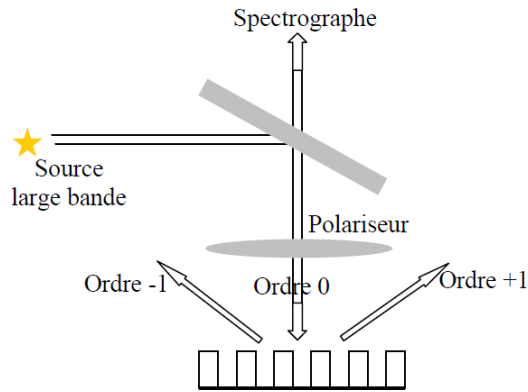


Figure II-1. Dispositif expérimental pour la mesure de signature optique par réflectométrie goniométrique [ROB02\*]

La signature optique est dans ce cas composée des efficacités de diffraction dans les différents ordres mesurés à partir de plusieurs angles d'incident  $\theta_i$  pour une longueur d'onde fixée.

## 2.2 Réflectomètre à angle d'incidence normal

La réflectométrie spectroscopique à angle d'incidence normal [HOL02, MAI01] fonctionne sur le même principe que la réflectométrie goniométrique, mais avec un angle d'incidence  $\theta_i$  nul. Par contre, la signature est cette fois-ci mesurée pour plusieurs longueurs d'onde. Comme c'est le cas de toutes les méthodes réflectométriques, une onde incidente est séparée en deux parties (Figure II-2). Une partie sert de référence et l'autre partie est transmise vers l'échantillon via le polariseur. L'onde diffractée à l'ordre 0 est récupérée et redirigée vers le spectromètre pour la mesure du coefficient  $R$ . Le polariseur permet de sélectionner la polarisation incidente et donc de mesurer les grandeurs  $R_p$  et  $R_s$  ; on parle de signatures réflectométriques.



*Figure II-2. Réflectomètre à angle d'incidence normal [ELK10]*

Le principal avantage de ce système optique réside dans le fait qu'il est facile à mettre en œuvre in-situ. En effet, il existe en mode spectroscopique et en mode cinétique. Les conclusions sont donc les mêmes que pour l'ellipsomètre en terme de temps de mesure. Toutefois, ne possédant pas ce type d'équipement au sein du laboratoire, nous n'avons pu faire des mesures reflectométriques. Nous avons orienté nos travaux expérimentaux sur des mesures ellipsométriques.