



HAL
open science

Nouvelle méthodologie hybride pour la mesure de rugosités sub-nanométriques

Jérôme Reche

► **To cite this version:**

Jérôme Reche. Nouvelle méthodologie hybride pour la mesure de rugosités sub-nanométriques. Micro et nanotechnologies/Microélectronique. Université Grenoble Alpes, 2019. Français. NNT : 2019GREAT050 . tel-02520554v1

HAL Id: tel-02520554

<https://theses.hal.science/tel-02520554v1>

Submitted on 26 Mar 2020 (v1), last revised 15 May 2020 (v2)

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 LA COMMUNAUTÉ UNIVERSITÉ GRENOBLE ALPES

Spécialité : **Nano Electronique et Nano Technologie**

Arrêté ministériel : 25 mai 2016

Présentée par

Jérôme RECHE

Thèse dirigée par **Cécile GOURGON**
et codirigée par **Patrice GERGAUD**

préparée au sein du **Laboratoire d'Electronique et des
Technologies de l'Information (CEA Grenoble)** et du
Laboratoire des Technologies de la Microélectronique (CNRS)
dans l'**École Doctorale Electronique, Electrotechnique,
Automatique et Traitement du Signal (EEATS)**

Nouvelle méthodologie hybride pour la mesure de rugosité sub-nanométrique

Thèse soutenue publiquement le 15 octobre 2019,
devant le jury composé de :

Pr, Yves, JOURLIN

Professeur, Université Jean Monnet, Président du jury (et rapporteur)

Dr, David, BABONNEAU

Directeur de recherche, Université de Poitiers, Rapporteur

Dr, Tatiana, NOVIKOVA

Ingénieur de recherche, Polytechnique, Membre

Dr, Raluca, TIRON

Ingénieur de recherche, CEA - LETI, Membre

Dr, Cécile, GOURGON

DR, CNRS, Directrice de thèse

Dr, Patrice, GERGAUD

Ingénieur de recherche, CEA - LETI, Co-directeur de thèse

Dr, Maxime, BESACIER

Maître de conférences, Université Grenoble Alpes, Invité

Dr, Yoann, BLANCQUAERT

Ingénieur de recherche, CEA - LETI, Invité



*On ne va jamais aussi loin
que lorsqu'on ne sait pas où l'on va.*

Christophe Colomb

Remerciements

Trois ans de vie, c'est long et court à la fois. On entre dans ce périple avec deux ou trois affaires et on en ressort grandit avec de grosses valises. J'ai eu l'occasion de rencontrer de nombreuses personnes et beaucoup m'ont aidé par leur travail, leurs actions ou même simplement leur présence.

Mes premiers remerciements vont aux membres du jury pour leur bienveillance, et plus particulièrement aux rapporteurs pour leurs commentaires éclairés sur mon travail et sur ce manuscrit.

Les « super-viseurs »

Ensuite, j'aimerais remercier ceux qui m'ont encadré.

Cécile je te remercie car plus qu'un nom et une signature, tu as su être présente aux bons moments et avec de précieux conseils. Patrice, c'est avec beaucoup d'humilité que je te remercie de la confiance que tu as placée en moi, toujours avec ce recul sur mon travail, mais aussi en me donnant la sensation d'être sur le même pied d'égalité que toi. Tu as apporté ton expertise à mes idées parfois étranges, mais tu m'as aussi poussé vers la lumière (les lignes de lumière). Maxime, je te remercie de ton aide, ta disponibilité, tes conseils, mais aussi d'être venu me chercher pour cette thèse (sachant ce qui t'attendait) et d'avoir pris du temps pour discuter du travail et autres. Yoann (« chef »), je voudrais te remercier d'avoir suivi tout ce travail au quotidien, et aussi pour ton aide dans cet environnement en mouvance perpétuelle.

En somme, un grand merci à tous les quatre !

Les collègues

Je me dois aussi de remercier mes collègues qui ont su, enfin dû, me supporter pendant trois ans. Je pense forcément à mes « co-bureau » Sana et Yann, mais aussi à l'équipe DSA (et DNA) passée et actuelle, dont les divers bureaux ont été des oasis. Je remercie plus particulièrement Aurélie, pour les relectures de ces travaux. De manière générale, je souhaite remercier les différentes personnes du 41.23 que j'ai pu côtoyer, des stagiaires aux chefs et experts, ceux que j'ai pu croiser au café comme en salle blanche (ne mélangeons pas les deux).

A cette liste de remerciements s'ajoute les personnes du LTM pour leur accueil chaleureux et les multiples événements, techniques ou autres, toujours plus intéressants les uns que les autres. De même, je remercie les personnes du SCMC (SMCP) qui m'ont apporté tant de savoir et avec qui j'ai pu avoir de nombreuses discussions, toujours très enrichissantes.

Mes amis

Partager des moments de vie quand le temps est dur à trouver, c'est souvent plus compliqué que de partager le travail, je voudrais donc dire un grand merci au Dr. Alexandre PIAGET, et au Dr. Jérôme GRIFFON (oui j'étais obligé), pour un nombre tellement grand de raisons qu'un manuscrit ne suffirait pas. Je souhaite remercier ici aussi : Anne, Geoffrey et Kevin ; tout comme Aurore et Florian, lui m'attend toujours pour une sortie moto, ainsi que Maxime. Enfin, je remercie Audrey d'être restée présente, même à plus de 6 000 km, et qui a la patience d'attendre encore que je passe la voir, après plus de 3 ans.

Ma famille

On a beau créer des liens, les faire et les défaire au-delà des pertes et des émois, la famille est là. Je voudrais donc remercier tous les membres de ma famille pour leur présence, plus particulièrement un grand merci à mes sœurs, Lauriane, Coraline et Amélie mais aussi mes nièces Cloé, Andréa, Théa et mon fils Mathis, pour tous ces moments d'insouciance pleins de câlins, de bisous mais aussi de bêtises auxquelles il ne faut pas rire. Ces moments qui vous sortent la tête du monde scientifique et vous font respirer pour mieux repartir, je ne les oublie pas. Bien sûr, je remercie aussi mon père de sa présence sans faille et aussi ma mère. Cette dernière, ô combien importante, qui en plus de sa présence a pris le temps (et s'est « pris la tête ») à lire, relire et relire ce manuscrit (presque sans fin). Merci à vous et merci à Célia qui a dû supporter, finalement plus que moi, un travail sans fin.

Les inattendus

Il y a des gens qui m'ont conduit là et qui sans même le savoir m'ont aidé à avancer, je tenais donc à les remercier. Premièrement, je remercie Christophe C. qui un jour de juin 2012 a débarqué dans un amphi, sandales aux pieds et sac de randonnée sur l'épaule, pour nous parler d'un projet révolutionnaire. Son énergie était telle qu'il m'emporta dans un monde qu'on appelle « la science ». 7 ans plus tard, je sens encore cette énergie qui m'anime à présent comme une transmission, un partage, encore merci ! Dans un tout autre registre, je remercie la boulangère en bas de chez moi qui râlait quand je lui demandais 30 pains au chocolat pour amener au travail « il fallait prévenir monsieur ! ». Je remercie aussi tous ceux qui m'ont dit les phrases « mais pourquoi tu fais une thèse ? » et « tu en es où de la rédaction ? », ou encore toute autre phrase de ce type, qui m'ont à chaque fois donné l'envie d'aller « plus vite, plus haut, plus fort », sans pour autant être un grand sportif. Mes derniers remerciements vont à tous les lecteurs de ce manuscrit, bien sûr aux rapporteurs déjà cité, mais aussi toutes les futures personnes (j'espère qu'il y en aura) qui liront en entier (oui je pousse un peu là) ou au moins en partie ce manuscrit.

Table des matières

Introduction	5
Chapitre I - La rugosité de ligne dans le contexte de la microélectronique	9
I.1 La microélectronique	9
I.1.1 Les évolutions	10
I.1.2 Les problématiques actuelles	11
I.2 Le patterning	14
I.2.1 La lithographie	15
I.2.2 La gravure	20
I.2.3 La métrologie	21
I.3 La rugosité de ligne	22
I.4 Mise en place d'une réponse à la problématique de métrologie de rugosité de ligne	24
I.5 Résumé du chapitre	25
Chapitre II - Métrologie de rugosité de ligne	27
II.1 Concept de la métrologie	27
II.1.1 Justesse de mesure	27
II.1.2 Fidélité de mesure	28
II.2 Application à la rugosité de ligne	29
II.2.1 Types de rugosités	29
II.2.2 Métrologie standard	30
II.2.3 La corrélation de bord	30
II.3 Vers une description plus complète de la rugosité.....	32
II.3.1 Notion de corrélation et de fractale	32
II.3.1 Le choix de la PSD	36
II.3.2 Performance de la PSD	38
II.4 Résumé du chapitre	45
Chapitre III - Les techniques de caractérisation	47
III.1 Microscopie électronique à balayage	47
III.1.1 Interaction électrons-matière.....	47
III.1.2 Le principe de fonctionnement	49
III.1.3 Les équipements de microscopie électronique.....	50
III.1.4 Extraction de la rugosité en CD-SEM	51
III.1.5 Limitations	53
III.2 Microscopie à force atomique (AFM)	56

Table des matières

III.2.1 Le principe de fonctionnement	56
III.2.2 Extraction de la rugosité en CD-AFM.....	57
III.2.3 Limitations	58
III.3 La scatterométrie	60
III.3.1 Le principe de fonctionnement	60
III.3.2 Extraction de la rugosité par OCD	62
III.3.3 Limitations	64
III.4 La diffusion des rayons-X aux petits angles	65
III.4.1 Interactions rayons-X, matière	65
III.4.1 Principe et équipements.....	67
III.4.1 Extraction de la rugosité de ligne par CD-SAXS	69
III.4.2 Limitations	77
III.5 Comparaison expérimentale des techniques.....	79
III.5.1 Gamme de travail en fréquences	79
III.5.2 Capacité d'extraction de la rugosité	81
III.5.3 Maturité de développement pour une métrologie en ligne	83
III.6 Résumé du chapitre	84
Chapitre IV - Développement d'étalons de rugosité.....	85
IV.1 Rugosité périodique.....	85
IV.1.1 Les différents types	85
IV.1.2 Contraintes multi-équipements	87
IV.1.3 Mesure de validation.....	88
IV.2 Programmation de rugosité suivant une PSD.....	91
IV.2.1 Génération du signal de bord.....	91
IV.2.2 Paramétrage du dessin de lignes.....	91
IV.2.3 De la ligne au réseau	92
IV.3 La fabrication des étalons et leurs biais.....	93
IV.3.1 Du modèle à la réalité	94
IV.3.2 Les biais du dessin	98
IV.3.3 Les biais métrologiques	98
IV.3.4 Extraction de rugosité sur images simulées	100
IV.4 Améliorations et limites de la programmation.....	103
IV.4.1 Amélioration du signal généré	103
IV.4.2 Les outils de lithographie et les procédés	105

Table des matières

IV.4.3	Prise en compte en amont de la rugosité naturelle	106
IV.4.4	La métrologie de référence	106
IV.5	Résumé du chapitre	107
Chapitre V - Hybridation et simulations.....		109
V.1	Principe de base de l'hybridation	109
V.1.1	Le lien entre les données.....	109
V.1.2	Type de données	111
V.1.3	Niveau d'abstraction	111
V.1.4	L'architecture d'hybridation.....	112
V.2	Les techniques d'hybridation.....	113
V.2.1	Réseau Bayésien	113
V.2.2	Les réseaux de neurones	116
V.3	Application du PMC au cas d'étude	121
V.3.1	Les données d'entrées et de sorties.....	121
V.3.2	Sélection des hyper-paramètres	122
V.4	Vers des données plus proches des métrologies.....	125
V.4.1	Génération SEM.....	126
V.4.2	Génération AFM	127
V.4.3	Génération OCD.....	129
V.4.4	Génération SAXS.....	130
V.5	Résumé du chapitre	131
Conclusions et perspectives		133
1.	Principaux résultats obtenus	133
2.	Perspectives.....	134
2.1	Amélioration des techniques individuellement	134
2.2	Test fonctionnel du premier réseau	135
2.3	Vitesse de traitement globale	135
2.4	Changement du niveau d'hybridation.....	135
Glossaire		137
Bibliographie.....		139

Introduction

Les objets issus de la microélectronique font aujourd’hui partie de notre quotidien, qu’il s’agisse de notre poste informatique, notre téléphone portable ou encore notre tablette, mais aussi des montres connectées ou même de l’automobile, et autres systèmes aux fonctions diverses. Ces objets sont tous constitués d’un nombre important de composants électroniques qui se doivent d’être performants. Les trois paramètres principaux à considérer pour évaluer la performance d’un composant électronique sont : le temps de réponse ou de traitement pour une sollicitation donnée, la fiabilité, et enfin le volume ou la quantité de données qu’il est capable de traiter et/ou stocker. Les applications actuelles des composants électroniques peuvent être réparties en fonction de ces trois facteurs, comme illustré par la Figure 1. Le calcul intensif nécessite de traiter beaucoup de données, avec une grande fiabilité. Dans ce cas, le temps de calcul n’est pas primordial, le nombre de serveurs pouvant être démultiplié si nécessaire. L’IoT (Internet of Things), « internet des objets » en français, ou encore le traitement embarqué, aura besoin de fiabilité sur les résultats et d’un temps de réponse rapide, tandis que le volume d’informations sera restreint. Enfin, les applications streaming (pouvant être traduites par « flux direct » en français) seront, quant à elles, plus assujetties aux performances de temps et de volume d’informations, avec une plus grande latitude sur les performances de fiabilité. Ces exemples, non exhaustifs, sont la représentation directe de ce que visent les développements actuels en microélectronique. D’autre part, des applications encore plus consommatrices de données et de flux sont prévues dans les années à venir, telles que les voitures autonomes ou encore les assistances robotiques. Ces nouvelles applications vont nécessiter de pouvoir répondre aux trois axes de performance à la fois, c’est là un défi pour les futurs développements.

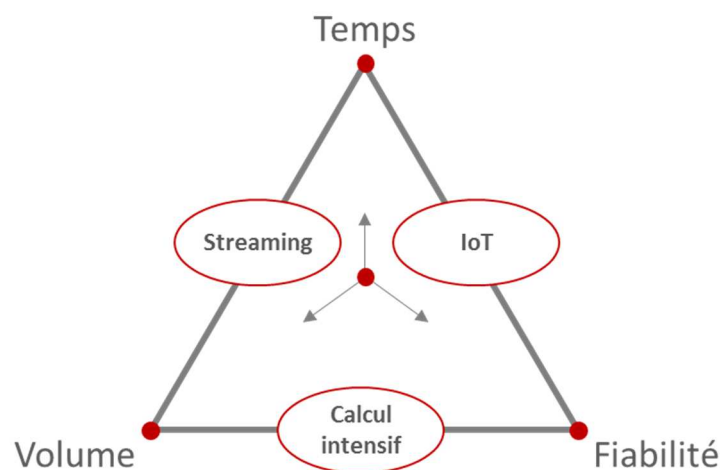


Figure 1 : Exemple d’applications placées sur les trois axes de performance (temps, volume, fiabilité).

Les objets de la microélectronique sont développés pour répondre à ces critères, avec pour conséquence une taille qui ne cesse de diminuer. Ceci nécessite des améliorations constantes des procédés de fabrication ainsi que des ruptures sur les architectures de conception. Un des points clef visé pour ces améliorations est la réduction de la consommation des composants. Pour cela, il a été prouvé qu’il est nécessaire de réduire les défauts de fabrication d’une puce électronique, et notamment la rugosité de ligne présente sur les flancs des motifs constituant les composants élémentaires et qui peut être décrite comme une variation dimensionnelle le long des dessins générés physiquement. Cette réduction nécessite un contrôle métrologique toujours plus poussé et précis.

Cependant, les équipements actuels de métrologie ne répondent pas toujours au cahier des charges, de par leurs propriétés physiques de fonctionnement ou encore leur application première, parfois très différente (inspection, mesures diverses...). Ainsi de nouvelles méthodes, visant à coupler les résultats issus de diverses techniques de métrologie, sont envisagées. C'est l'objet des travaux menés durant cette thèse, avec pour fil conducteur la détermination de la rugosité de ligne.

Le document est organisé de la manière suivante :

Dans le chapitre I, une description des enjeux de la microélectronique permettra de présenter les problématiques actuelles de défauts de fabrication et de la métrologie associée. Une description des principales étapes de procédés permettant la fabrication d'objets de taille caractéristique nanométrique, sera donnée. Nous décrirons alors les différents défauts pouvant être générés lors de ces procédés, et nous nous attacherons à expliquer l'impact de la rugosité de ligne sur les performances des dispositifs. Par la suite, la nécessité d'une métrologie dimensionnelle adaptée sera démontrée, et la réponse proposée pour la mesure de ce défaut, au travers de la combinaison de diverses mesures, sera exposée.

Le chapitre II portera sur un état de l'art sur la métrologie de rugosité de ligne avec l'explication des concepts de la métrologie, puis des métriques historiques de la rugosité de ligne. L'avancée vers une description plus complète de cette rugosité, mise en place depuis une dizaine d'années, sera explicitée grâce à la notion de fractale. On montrera alors que plusieurs modèles peuvent être utilisés pour exprimer la rugosité de ligne dans l'espace physique ou fréquentiel. Différents tests seront effectués afin de déterminer le modèle le plus adéquat par rapport à notre problématique. Ainsi, les performances du modèle choisi seront mises en évidence de manière statistique, sur la base de la variation de ses paramètres.

Le chapitre III donnera le principe de fonctionnement des quatre techniques de métrologie utilisées pour ces travaux. Pour chacune d'entre elles, les principes physiques mis en jeu seront rappelés, ainsi que la manière de réaliser une mesure et d'extraire les paramètres physiques considérés. Enfin, un état de l'art d'application de ces techniques à la mesure de rugosité sera effectué. Sur chaque équipement, les limites d'extraction de la mesure de rugosité seront explicitées, avec les différents points d'amélioration ou de contournement possibles. Un développement particulier sera fait autour de la technique SAXS (Small Angle X-rays Scattering) qui reste actuellement peu utilisée pour ce type d'application. Ensuite, une comparaison expérimentale des techniques sera effectuée grâce à un jeu d'échantillons unique, fabriqué en respectant les contraintes de chacune d'entre elles. Ceci a permis de mettre en évidence, à la fois les résultats de performance de mesure et les capacités des différentes techniques à répondre à notre problématique métrologique, mais aussi les performances de vitesse et de maturité des équipements, afin de définir leur adéquation aux contraintes de l'industrie de la microélectronique.

Dans le chapitre IV, la création d'étalons de rugosité de ligne au travers de la création de dessins non rectilignes sera présentée. Tout d'abord sur la base d'un échantillon de rugosité de ligne périodique, déjà utilisé dans la littérature, avec pour but de déterminer sa faisabilité sur les procédés actuels et de mettre en place la méthodologie de génération et d'extraction. Par la suite, ce sont des rugosités de ligne, sur la base d'un modèle PSD (Power Spectral Density), qui seront générées dans le dessin et réalisées physiquement. Les résultats obtenus seront comparés, à la fois au dessin nominal, et à une chaîne de simulation de l'outil de métrologie utilisé. Nous verrons ainsi les contributions des

Introduction

différentes étapes (arrondi sur le dessin, transfert physique, mesure, etc.) sur la rugosité de ligne visée. Grâce à cela, des améliorations de programmation de la rugosité de ligne, pour l'obtention d'étalons, seront exposées et testées par le biais de simulations.

Finalement le chapitre V abordera l'hybridation de données, au travers de ses principes de base ainsi que des différentes méthodologies pouvant être mises en place. Un point plus particulier sera fait sur la méthodologie de réseau de neurones, utilisée pour ces travaux, et ensuite sur la manière dont cette méthode prend place dans le contexte de mesure de rugosité de ligne. Sur la base des équipements précédemment explicités, des simulateurs seront créés afin de permettre l'apprentissage rapide du réseau de neurones, point crucial pour une intégration en ligne d'une nouvelle métrologie. Le but final est, en effet, de rester dans des temps et une complexité de mise en place acceptables dans l'industrie de la microélectronique, tout en ayant une extraction de la rugosité de ligne avec une résolution sub-nanométrique.

Chapitre I - La rugosité de ligne dans le contexte de la microélectronique

Ce chapitre décrit brièvement les évolutions de la microélectronique depuis sa création, amenant les problématiques actuelles auxquelles cette industrie doit faire face. Les étapes clés de la microélectronique, permettant la réalisation de motifs sur un substrat, sont explicitées. L'attention est portée sur la nécessité d'effectuer une caractérisation physico-chimique ou dimensionnelle à chacune de ces étapes.

Une description des défauts, potentiellement induits par les procédés sur les motifs, est alors abordée, permettant d'introduire le concept de rugosité de ligne. L'impact de ce type de défauts sur le fonctionnement final est exposé, afin de mettre en lumière la nécessité d'une métrologie de rugosité de ligne adaptée.

Enfin, ce chapitre donne une vision globale des travaux menés dans le cadre de cette thèse permettant d'appréhender de ce document telle que la réponse à la problématique de métrologie de rugosité de ligne.

I.1 La microélectronique

Il y a plus de 70 ans que le premier transistor a été créé au sein du laboratoire Bell aux Etats-Unis. Par la suite, en 1958, ce sont plusieurs transistors qui ont été interconnectés sur une même puce pour donner ce qui a été appelé « circuit intégré ». Désormais, les circuits intégrés comportent une complexité importante, comme le confère la Figure I.1 avec de nombreux empilements, de plus on compte par milliards le nombre de transistors interconnectés dans un microprocesseur. Tout cela a été rendu possible grâce à la miniaturisation. Ce monde de la microélectronique, de plus en plus petit, s'est auto-entretenu en permettant, par la création de nouveaux outils plus performants, la fabrication d'autres produits encore plus petits. Ce cercle vertueux a permis de créer des objets indiscernables à l'œil, mais aussi de les mesurer [1–3].

La microélectronique correspond à ce que certains appellent la troisième révolution industrielle [4]. Ces soixante dernières années en sont la preuve avec une évolution importante, à la fois de notre manière de travailler, mais aussi dans notre quotidien avec la technologie qui en est issue. Aujourd'hui, la microélectronique évolue encore. Certaines problématiques, telles que le coût par puce ou encore la densité, sont les mêmes qu'au moment de l'émergence de ce secteur d'activité. Cependant, d'autres défis apparaissent, comme les erreurs dites « aléatoires » qui mènent à de nombreux défauts touchant, parfois même, la fonctionnalité des circuits.

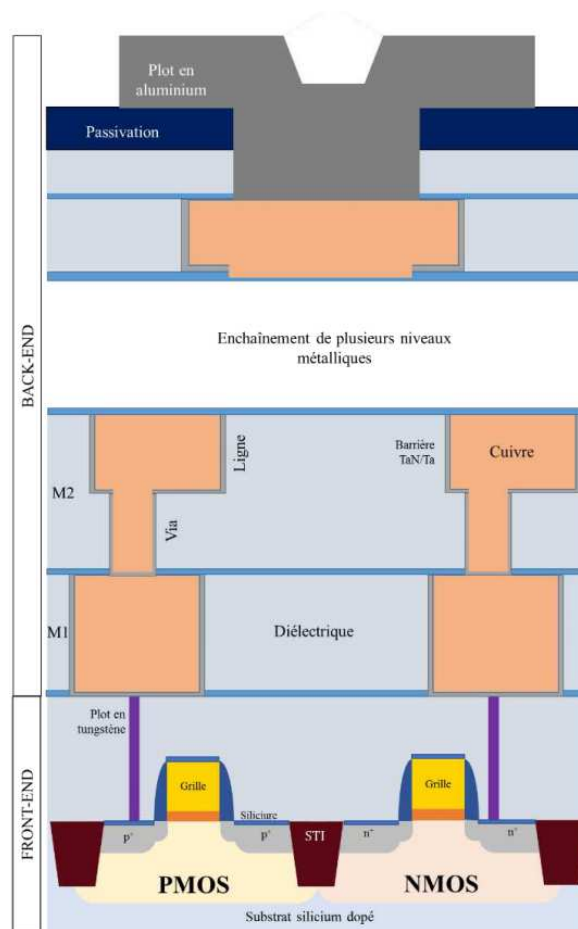


Figure I.1 : Coupe schématique de l'architecture d'un circuit intégré [5].

I.1.1 Les évolutions

Au milieu des années 60, Gordon E. Moore fit le constat que la complexité des puces doublait tous les ans, pour un coût identique. Il insuffla l'idée que cette croissance allait continuer, ainsi fut nommée la « Loi de Moore » [6]. Une dizaine d'années plus tard ce postulat se raffina, en termes de chiffres, avec pour métrique le nombre de transistors par puce qui se devait de doubler tous les deux ans. Cette prédiction utilisée comme feuille de route de la microélectronique a été validée, comme le montre la Figure I.2. Sur cette figure, nous pouvons voir que l'évolution des nœuds technologiques, nom des technologies de fabrication correspondant initialement à la demi-période résolue (distance entre deux motifs élémentaires), suit la courbe théorique de la loi de Moore. En effet, pour doubler le nombre de transistors sur une puce, leur taille a dû diminuer. Ceci a été rendu possible grâce aux nouvelles technologies de fabrication, et notamment la réduction de la longueur d'onde de la lumière utilisée en photolithographie.

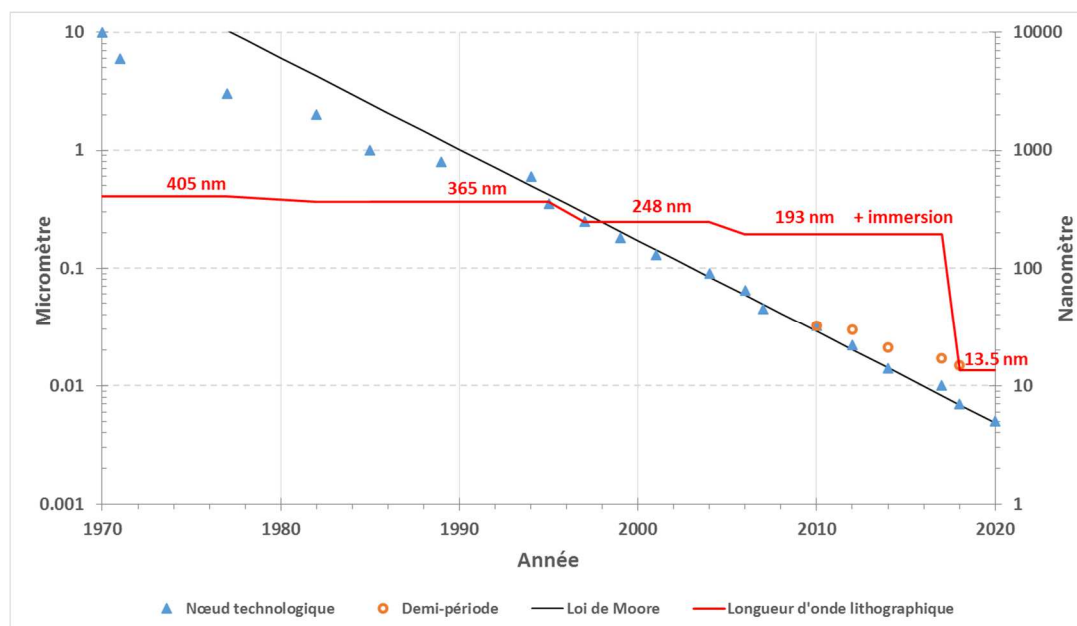


Figure 1.2 : Evolution du nœud technologique au fil des années (triangle bleu) ainsi que de la demi-période résolue en industrie, quand celui-ci ne correspond plus au nœud technologique (rond orange). Représentation de la loi de Moore extrapolée en 1975 (ligne noire), et la longueur d'onde utilisée sur les machines de lithographie qui ont permis d'atteindre ces dimensions (ligne rouge). [3, 7, 8]

Cependant, comme le confère aussi la Figure 1.2, dans les années 2010 un ralentissement de la miniaturisation a eu lieu. Le nombre, donné pour nom au nœud technologique, n'a cessé de diminuer contrairement à la physique réelle, soit au final une différenciation du nœud technologique par rapport à la demi-période. Cette atteinte de limites a donné lieu, parallèlement à la loi de Moore, à l'émergence du « more than Moore ». Le principe est de réaliser, pour chaque besoin, un circuit plus spécifique sans forcément augmenter le nombre de transistors. Cela passe par le développement de nouvelles architectures. Une des avancées les plus perceptibles est le passage à l'empilement 3D, i.e. en superposant un transistor sur un autre.

I.1.2 Les problématiques actuelles

La densité

Cœur du développement de la microélectronique pendant 50 ans, la densité de transistors sur un substrat semble s'être stabilisée, comme le montre la valeur du demi-pitch sur la Figure 1.2. La limite de 10 nanomètres ne semble pas pouvoir être abaissée [9]. En effet, au vu du principe industriel de fabrication par photolithographie, basé sur une source de lumière, une multitude de phénomènes, tels que la diffusion et la diffraction, entrent en jeu et à cela s'ajoute la limite de la matière puisque 10 nm correspondent à un peu moins de 19 mailles de silicium. Pourtant, depuis quelques années, afin de pallier les limites physiques de certains équipements de fabrication et de réduire les aberrations optiques, les efforts se concentrent sur la conception même des circuits. Cela s'est traduit par des règles de dessin de puces électroniques qui ont drastiquement changé, jusqu'à limiter les motifs à des impressions unidirectionnelles et continues sur certaines zones, comme visible sur la Figure 1.3 avec la technique « line-cut » introduite pour le nœud 20 nm [10]. Ainsi, la course à la densité n'est pas encore totalement finalisée et le demi-pas visé dans le dernier rapport de l'International Roadmap for Devices and Systems (IRDS) de 2017 est de 7.7 nm pour 2033 [11].

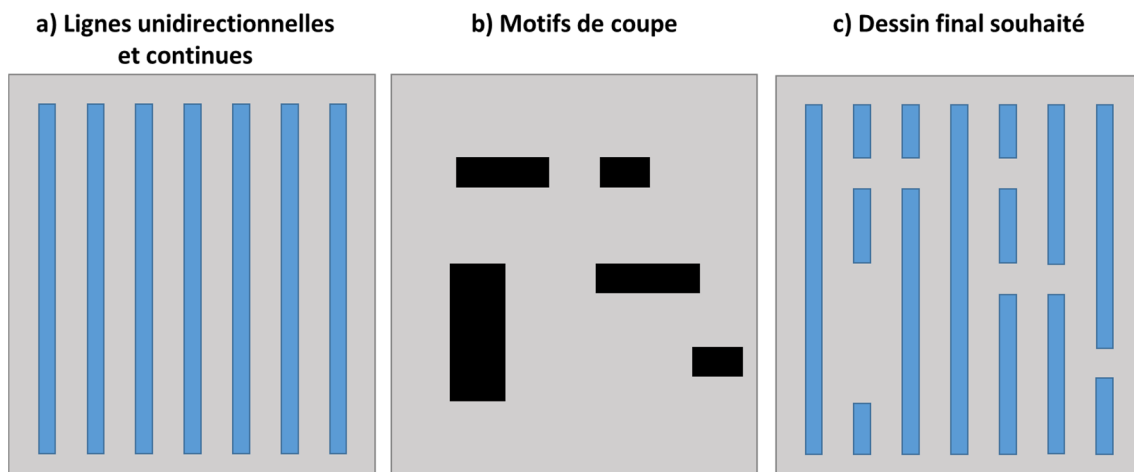


Figure I.3: Schéma de la technique en deux temps avec a) un masque de lignes unidirectionnelles et continues (de dimensions critiques), b) un second masque de coupe de lignes (de dimension relâchée) pour obtenir c) le dessin souhaité aux dimensions critiques.

Le coût

Cette problématique de coût est l'axe majeur de la prédiction de Moore en 1965. En effet, bien que son extrapolation ait défini l'évolution de la taille des transistors pour la microélectronique, le sens premier portait sur le coût de production, et donc celui de vente, qui devait rester identique par puce. L'augmentation du nombre de puces fabriquées devait alors permettre plus de profit. Ainsi, au fil des ans, de multiples changements pour maintenir cette diminution des coûts par puce ont eu lieu. Par exemple, ces quinze dernières années, l'augmentation de la taille des substrats, de 200 mm de diamètre à 300 mm, a permis une réduction des coûts par puce de 28 % [12]. Cependant, la transition actuelle vers des substrats de 450 mm, censée permettre un gain similaire, représente un investissement colossal. En effet, les nouvelles chaînes de production atteignent un coût de mise en place de 10 milliards de dollars, soit 30% de plus que pour les substrats 300 mm [13]. Au global, le coût des environnements propres, le prix des dernières générations d'outils (notamment de lithographie qui atteint 120 millions de dollars [14] contre 30 précédemment [15]), ainsi que la complexité des étapes de fabrication, de l'accroissement de leur nombre, et l'augmentation des temps de conception des nouveaux circuits mènent désormais à la fin du modèle proposé par Moore. Par ailleurs, l'alternative évoquée du « more than Moore » nécessite des volumes plus faibles par type de puces et de nouveaux développements, ces derniers points posent aussi la question de la rentabilité [16].

Les défauts

L'IRDS [11] définit comme point clef d'amélioration la maîtrise de divers types de défauts impactant les performances des puces. Il est possible de regrouper ces défauts en plusieurs catégories:

- **La défektivité** qui correspond aux changements de la forme fonctionnelle du motif souhaité. Il peut s'agir d'ajout ou de manque de matière, comme l'ajout de pont entre deux motifs ou la rupture d'une ligne de connexion, par exemple.
- **Les erreurs de positionnement**, comme visible en haut à droite de la Figure I.4. ces erreurs sont imputables, soit à des enchaînements de procédés non alignés ou calibrés différemment (le terme « overlay » est utilisé couramment en anglais [17]), soit à la dérive

au cours du temps de l'écriture des motifs (exprimée par le terme « stitching » signifiant défaut de raccord entre champs).

- **La rugosité et les effets stochastiques**, visibles sur la partie en bas à gauche de la Figure I.4. Ce type de défaut est relié à des phénomènes physico-chimiques, qu'il s'agisse des matériaux utilisés, du principe même du procédé considéré, ou encore du temps de réalisation de ce dernier [18].

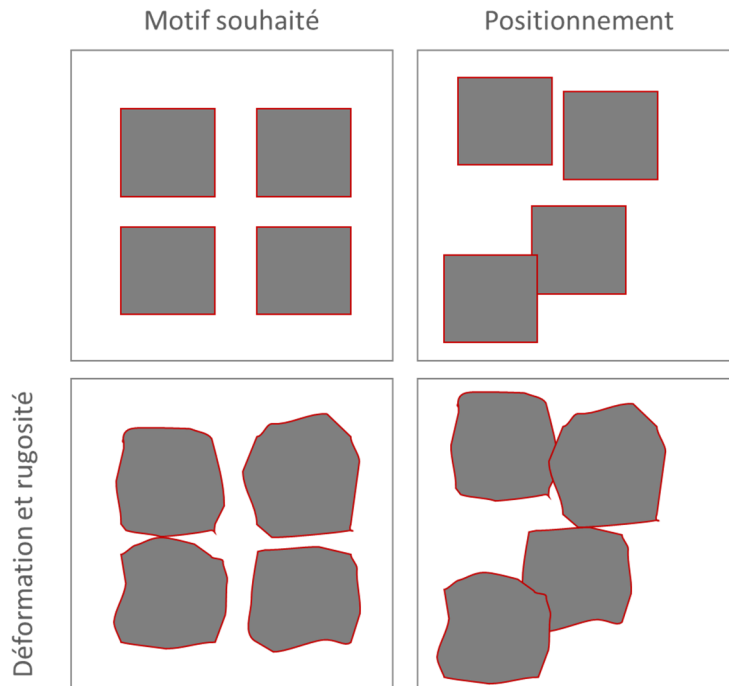


Figure I.4 : Schématisation des problématiques de positionnement et de déformation ou rugosité, ainsi que la combinaison des deux, en comparaison du motif souhaité.

Outre la défektivité qui impacte directement le fonctionnement, ces défauts sous la forme de fluctuations impactent les performances et la consommation des composants. On note par exemple, la création de capacités parasites ainsi que l'augmentation des courants de fuites (voir Chapitre II pour les enjeux liés à la rugosité de ligne). Dans le cas de la combinaison des défauts de positionnement et de ceux de rugosité (comme visible en bas de la Figure I.4), de la même manière que la défektivité, des contacts peuvent être créés entre les motifs sans que cela ne soit prévu dans le dessin initial, induisant des courts-circuits ou un manque de contact entre deux niveaux, ce qui empêche tout autant le fonctionnement du composant final. C'est donc la fiabilité des composants finis, mais surtout celle de la chaîne de production complète, qui n'est pas assurée si la totalité de ces trois catégories de défauts n'est pas contrôlée.

Les travaux développés ici portent donc sur la mesure de la rugosité de ligne. Afin de comprendre l'impact de celle-ci ainsi que sa caractérisation, il est important de cerner comment ce défaut apparaît et évolue lors des différentes étapes de la fabrication des composants. La section suivante est donc dédiée à l'explication de la réalisation des motifs pour la fabrication de composants.

I.2 Le patterning

Le patterning correspond à la structuration des formes dans la matière et donc des défauts qui y sont liées ; c'est aussi la clef du passage de l'électronique à la microélectronique. Au lieu d'être créés séparément, les composants sont directement fabriqués les uns à côté des autres, sur un même substrat, et connectés entre eux. Le terme patterning, pouvant être traduit littéralement en français par « façonnage », correspond à la création contrôlée de formes dans un substrat, support de base de la microélectronique. Cela inclut donc, comme représenté en Figure I.5, les différentes étapes de fabrication telles que la lithographie et la gravure, explicitées ci-après, ainsi que la vérification des formes réalisées grâce à la métrologie. Le point de départ étant une multitude de dessins 2D de ce que l'on souhaite réaliser en 3D, telle que des couches superposées, après un patterning, d'autres s'ensuivent jusqu'à la réalisation finale d'une puce (ici la préparation des matériaux à structurer n'est pas mentionnée).

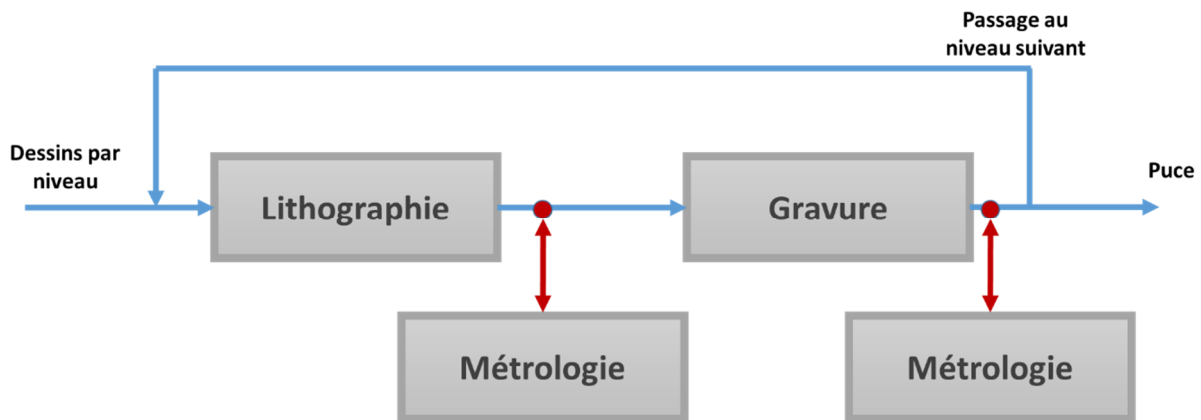


Figure I.5 : Schématisation du patterning au travers de ces étapes clefs.

La variété et la diversité des motifs pouvant être réalisés sont quasiment infinies. Dans le cadre de ces travaux de thèse, l'attention est portée sur les réseaux de lignes. Ce type de motifs, très répandu dans la microélectronique, est représenté en Figure I.6. Les paramètres primordiaux sont notés, certains pouvant être physiquement liés. On utilisera donc les termes : CD (Critical Dimension) pour désigner la largeur de ligne, Pitch pour le pas de répétition d'une ligne, CD_{Espace} l'espace entre deux lignes et enfin SWA (Side-Wall Angle) pour exprimer l'angle formé entre le bord de la ligne et le support. Comme explicité sur la Figure I.6, tous les paramètres ne sont pas visibles sur la « vue du dessus » (a), contrairement à la « vue en coupe » (b). Cependant, obtenir une vue en coupe est plus complexe et nécessite parfois des métrologies dites destructives (altérant le motif ou, à minima, son support). Or, dans ces travaux, utiliser des caractérisations non destructives est une des contraintes permettant la mesure sur divers équipements, tour à tour. Ainsi, suivant les techniques, l'accès à la mesure des différents paramètres peut être possible ou non (voir Chapitre IV). On notera cependant que la Figure I.6 n'illustre pas les paramètres de rugosité de ligne, ceux-ci seront explicités en détail dans le Chapitre II.

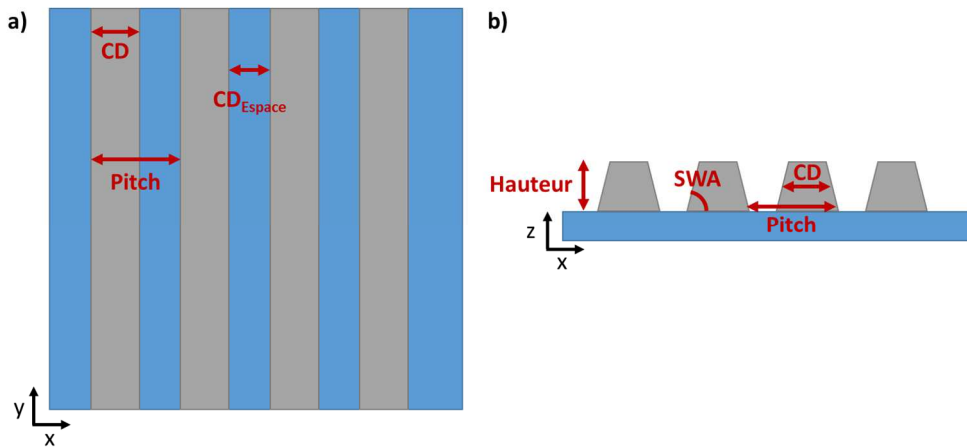


Figure I.6 : Représentation du motif "réseau de lignes" avec ses paramètres géométriques
a) vue du dessus, b) vue en coupe.

I.2.1 La lithographie

Il s'agit là de la première étape physique du patterning pour la création d'un transistor. Le but est de passer d'un dessin à une réalisation sur un substrat. La lithographie est vue comme la première étape critique de la réalisation, car même si le motif peut encore évoluer par la suite, cette étape fixe une partie des dimensions finales à plusieurs échelles. Pour exemple, cette étape fixe à l'échelle macroscopique la taille de la puce, et à l'échelle du micromètre et du nanomètre, le pitch. C'est l'évolution des techniques de lithographie qui a permis les progrès de la microélectronique, à savoir la diminution de la taille des transistors et donc des composants. Suivant la technique utilisée ainsi que les matériaux, la rugosité créée lors de cette étape sera différente en amplitude mais aussi en forme. Il existe aujourd'hui une multitude de techniques, les plus répandues sont décrites ci-après.

I.2.1-A La photolithographie

La plus industrielle des techniques pour la réalisation sur substrat est la photolithographie [19], de par sa vitesse et sa qualité d'exécution. Cette technique, représentée en Figure I.7, utilise une source de lumière monochromatique mise en forme par un jeu de lentilles. Cette lumière traverse ensuite un masque, support des motifs à réaliser. Ce masque, composé de parties transparentes et d'autres opaques, permet de laisser passer sélectivement la lumière. Un second système optique permet la réduction de la taille de ce dessin (actuellement d'un facteur 4) et sa focalisation sur le substrat, on parle alors d'écriture ou d'exposition de motifs. Ce même substrat est recouvert d'une résine photosensible qui, sous l'effet de la lumière, change de propriété chimique et ainsi est amenée à rester ou à être dissoute lors du développement, dernière sous-étape de la lithographie.

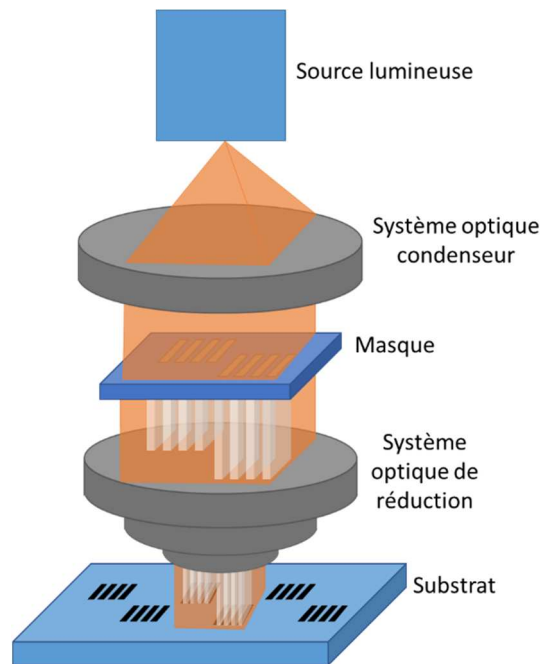


Figure I.7 : Schéma simplifié du principe de fonctionnement de la photolithographie.

Sur ce principe de réduction de taille par un système optique, on peut penser que plus la taille du motif sur le masque sera réduite, plus le motif sera petit lors de sa projection sur le substrat, cependant la photolithographie est limitée par les phénomènes de diffraction. La taille du plus petit élément pouvant être écrit est régie par la résolution atteignable. Celle-ci correspond à la capacité du système à garder deux éléments qui sont séparés sur le masque, encore séparés lors de la projection. A cela s'ajoute la capacité résolvente de la résine utilisée. En somme, la résolution limite d'un procédé de photolithographie, associée à un équipement, est donnée par le critère de Rayleigh :

$$R = k_1 \times \frac{\lambda_0}{NA} \quad (I.1)$$

avec k_1 la constante de procédé (dépendant du type de masque, de résine ou encore la polarisation de l'illumination), λ_0 la longueur d'onde de la lumière utilisée et NA (de l'anglais Numerical Aperture) l'ouverture numérique du système de projection, représentée Figure I.8, suivant :

$$NA = n \times \sin(\theta_{Max}) \quad (I.2)$$

où n est l'indice optique du milieu en sortie de lentille, et θ_{Max} le demi-angle maximal du cône de lumière, également après la lentille. Pour améliorer la résolution, plusieurs voies peuvent ainsi être envisagées [20] :

- **La réduction du facteur k_1** : ce paramètre, qui était près de 0.9 dans les années 80, est désormais aux alentours de 0.2. Cette réduction a été possible grâce à l'amélioration du contraste des résines photosensibles, ainsi que le contrôle des effets de diffraction avec des sources de lumière à plusieurs pôles.
- **La diminution de la longueur d'onde** de la lumière utilisée. Comme cité précédemment, cette longueur d'onde, qui au fil des années donne son nom aux technologies, est passée de 405 nm à 13.5 nm. Cette dernière en date, nommée EUV (pour Extrême Ultra-Violet), est à peine émergente et a demandé de nombreuses années de développement. Par

exemple, la problématique d'absorption rapide de l'énergie liée à cette longueur d'onde de la gamme UV a nécessité de revoir entièrement l'architecture des machines en utilisant le masque, ainsi que tous les systèmes optiques, en réflexion, et non en transmission comme précédemment.

- **L'augmentation de l'ouverture numérique** par le changement de matériaux de fabrication des optiques et, depuis une quinzaine d'années par le remplacement de l'air (milieu dit « sec ») entre la lentille et le substrat, par un liquide tel que l'eau ayant un coefficient $n=1.44$ pour une lumière à 193 nm, donnant lieu à la technologie dite « immersion ».

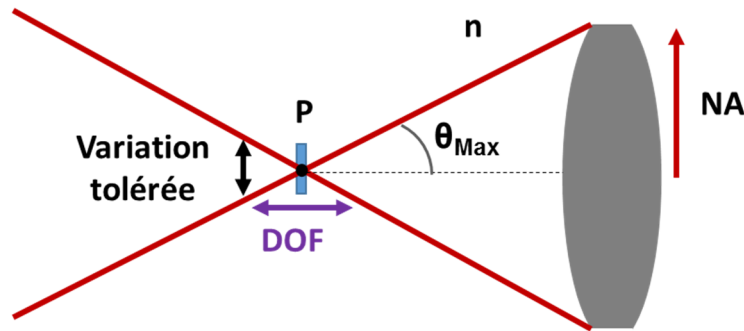


Figure I.8 : Représentation schématique de l'ouverture numérique et de la profondeur de champ.

En regard de la résolution, il est important de voir que le changement de ces différents paramètres influe sur la profondeur de champ (Depth of Focus, DoF), à savoir la distance verticale sur laquelle la variation d'un motif est acceptée (Figure I.8) qui s'exprime selon :

$$\text{DoF} \propto \frac{\lambda_0}{\text{NA}^2} \quad (\text{I.3})$$

Pour garder un DoF assez important et permettre l'écriture sur un substrat qui n'est pas parfaitement plan, et dans des résines de grande épaisseur sans induire une rugosité trop importante, il est donc préférable de diminuer la longueur d'onde plutôt que d'augmenter NA dont l'influence suit une loi en puissance de deux. Par ailleurs, on note que l'équation du DoF doit être adaptée au cas de la lithographie immersion [21], puisqu'il faut prendre en considération les divers changements d'indices en présence.

I.2.1-B La lithographie par faisceau d'électrons

Aussi appelée lithographie électronique ou « e-beam », cette technique a été très largement utilisée durant cette thèse, notamment grâce à la flexibilité qu'elle autorise au niveau des motifs à exposer. En effet, comme le suggère sa structure schématique (Figure I.9), et contrairement à la photolithographie, la lithographie par faisceau d'électrons n'a pas besoin de masque. Le faisceau d'électrons est défléchi par le biais de lentilles électroniques et vient réagir avec une résine électrosensible. Ainsi, les motifs sont exposés un par un. Ce dernier point induit un temps d'écriture beaucoup plus long, mais permet de changer aisément le type ou les dimensions des motifs voulus. De plus, avec cette technologie de lithographie, la résolution dépend essentiellement de la taille du faisceau d'électrons utilisé et de son énergie [22]. Par exemple, la lithographie électronique est la technique qui permet, aujourd'hui, la réalisation des réticules utilisés ensuite par la photolithographie. Pour tenter de pallier les limites de vitesse de la lithographie électronique classique, qui utilise un seul faisceau d'électrons de forme gaussienne, il existe de multiples dérivés. L'un d'entre eux nommé

« shape beam », *a contrario* de la technique standard, utilise une forme spécifique du faisceau (tel un masque). Cette technique peut par exemple, en un seul impact, exposer un rectangle pouvant être équivalent à une ligne complète.

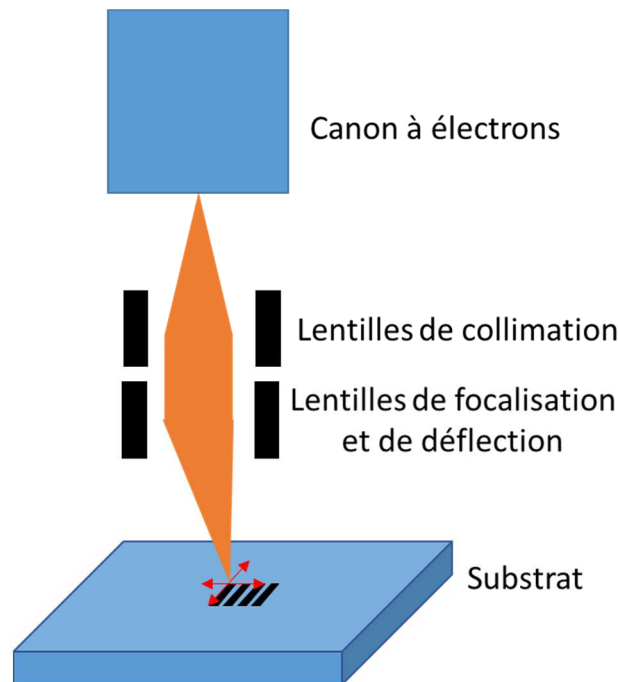


Figure I.9 : Schéma de principe de fonctionnement d'un outil de lithographie électronique à faisceau gaussien.

En parallèle de la photolithographie et de la lithographie électronique, d'autres techniques dites de lithographie avancée sont développées. Ces techniques ont pour but la reproduction de motifs déjà créés, sans réduction de taille ou, au contraire, la sous-multiplication du motif.

I.2.1-C Les techniques avancées de reproduction et de multiplication

Parmi les techniques de reproduction, l'une des plus répandues est la Nano-Impression. Cette dernière permet de répéter, à la même échelle et sur un nouveau substrat, les motifs déjà réalisés sur un moule nommé « master d'impression ». Ce master peut-être directement appuyé, tel un tampon, sur le nouveau substrat comportant un matériau semi-rigide, créant ainsi des motifs inversés (les lignes deviennent des tranchées et inversement) [23]. Il existe, aujourd'hui, des outils de nano-impimpression qui répètent plusieurs fois un master de la taille d'une puce, pour couvrir finalement un substrat entier [24], tandis que d'autres permettent la réplique d'un substrat complet (plusieurs puces) en une seule fois [25]. Certains travaux de métrologie, présentés dans ce manuscrit, se basent sur des lithographies réalisées grâce à cette dernière technologie.

Par ailleurs, des techniques adjointes au moins à l'une des techniques précédemment citées, permettent de réduire le pas critique final, par sous-multiplication. Parmi elles, on peut citer le double patterning (Self-Aligned Double Patterning, SADP), représenté schématiquement en Figure I.10, qui utilise, suite à une première formation de motifs, une répétition de dépôts et de gravures. Cette technique est déjà en production [26] pour les nœuds avancés, de même que le quadruple patterning effectuant cette méthode deux fois consécutivement. On remarque dans ce cas Figure I.10.e que la rugosité n'est pas la même de part et d'autre du motif final.

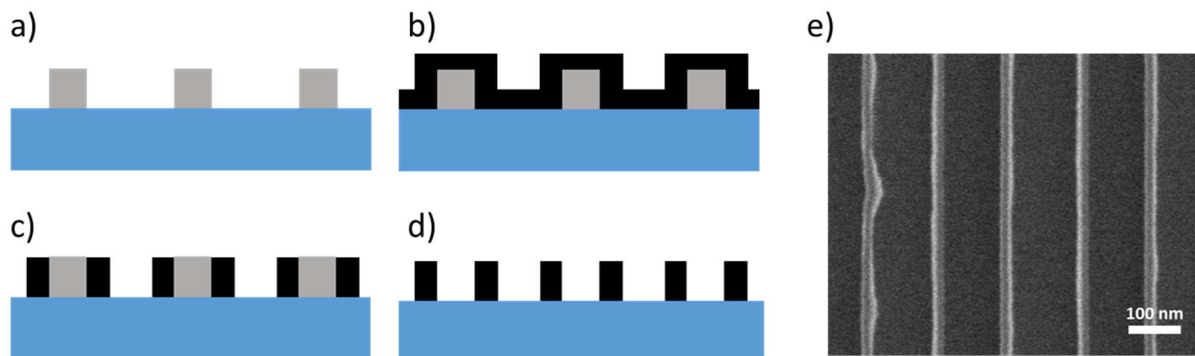


Figure I.10 : Schématisation des étapes principales du SADP, a) création d'un premier motif (gris), b) dépôt (noir), c) gravure du dépôt au dessus du premier motif et en fond de motif, d) retrait du premier motif ; et e) image vue du dessus (la parties blanches correspondent aux lignes finales visible sur le schéma d).

Pour second exemple de technique de sous-multiplication du pas lithographique, on peut citer l'auto-assemblage des copolymères à blocs (DSA) par grapho-épitaxie [27]. Cette technologie met en œuvre des matériaux polymères comportant deux chaînes de monomères différents qui, lors de l'ajout d'énergie, et de par leur faible affinité, s'auto-organisent. Suivant des conditions spécifiques (proportion des monomères, température utilisée, etc.) cette auto-organisation aboutit à la création de lamelles verticales (équivalente à des lignes nanométriques) sur de courtes distances. L'emploi de guides physiques, comme visible sur la Figure I.11, obtenus par une autre technologie de lithographie, permet d'orienter correctement cet auto-assemblage sur de grandes distances. La rugosité visible sur l'image en vue du dessus Figure I.11.e est à la fois du au copolymère et au guide d'alignement. La largeur des lignes obtenues par ce type de technique est actuellement aux alentours d'une dizaine de nanomètres, mais cette technique ne cesse d'être développée afin d'obtenir des facteurs de sous-multiplication de plus en plus importants et aussi de réduire la défektivité des procédés [28].

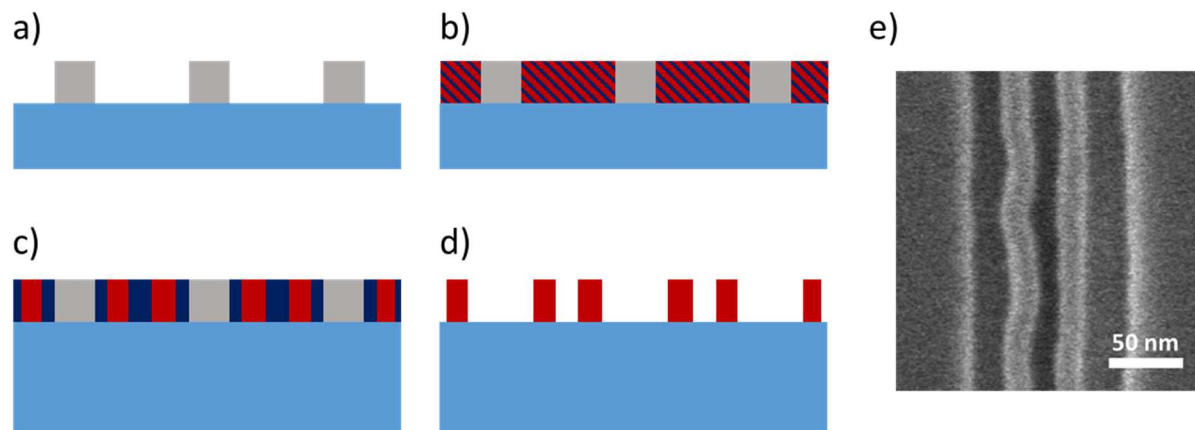


Figure I.11 : Schématisation des étapes principales de grapho-épitaxie en DSA, a) motifs guides, b) mise en couche de copolymère à blocs, c) ajout d'énergie (traitement thermique), d) retrait des matériaux non désirés ; et e) image prise en vue du dessus après retrait d'un des polymères (à gauche et à droite les bords du guide, et au centre de ligne d'un des polymères)

On remarque que ces dernières techniques ont besoin, en plus d'une étape de lithographie, de diverses étapes de gravure. Le principe de gravure nécessaire au patterning est décrit dans la section suivante.

I.2.2 La gravure

Cette partie du patterning suit l'étape de lithographie (Figure I.5). La gravure a pour but principal de transférer les motifs précédemment réalisés, dans le matériau final, par retrait de matière. Il peut s'agir du silicium, matériau nécessaire à la réalisation de transistors, de couches métalliques permettant les interconnexions, ou bien d'autres matériaux fonctionnels. La gravure est aussi utilisée pour la préparation de substrat, ou encore comme nous l'avons vu précédemment la sous-multiplification du pitch, par exemple avec la méthode SADP. Des étapes de nettoyage, afin de retirer entièrement un type de matériaux tel que les polymères, font aussi partie des procédés de gravure.

La principale contrainte de la gravure est de ne pas modifier le motif lors du transfert. Pour ce faire, la sélectivité est un facteur permettant ce contrôle, elle donne le ratio de vitesse de gravure du matériau à graver par rapport à celle du matériau initial. Ainsi, comme dans l'exemple donné en Figure I.12, avec une sélectivité de 1 (vitesse de gravure identique entre le matériau supérieur et celui inférieur), ce facteur permet de donner la hauteur maximale qu'il est possible de graver avec une bonne définition du motif. En effet, en considérant une gravure seulement dans la direction verticale, lorsque le matériau du dessus sera entièrement consommé (Figure I.12c), la distance entre le haut du motif et son fond restera identique.

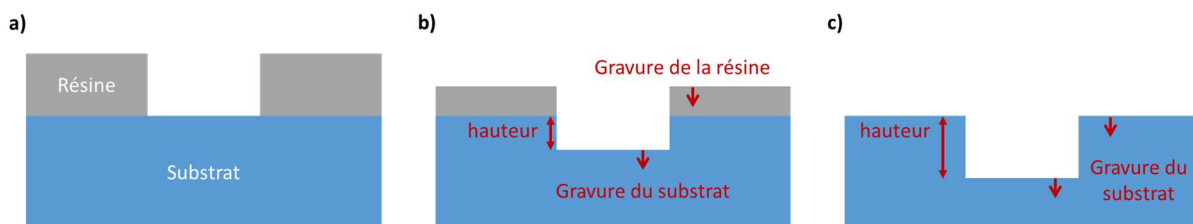


Figure I.12 : Principe de gravure avec une sélectivité de 1 entre la résine et le substrat,
a) l'état initial, b) gravure simultanée du substrat et de la résine
c) il n'y a plus de résine, la hauteur de gravure reste constante dans le substrat.

Le choix d'un matériau, ainsi que celui du procédé de gravure, influe fortement sur le résultat de la rugosité. La sélectivité de gravure entre les polymères (résines) et le silicium est malheureusement limitée. Pour pallier à la problématique de sélectivité, de nombreuses recherches sont en cours sur les matériaux eux-mêmes et les procédés utilisés ; cependant il est rare que ces derniers prennent en compte la rugosité de ligne, il s'agit là d'un phénomène souvent relégué au second plan, la première problématique étant le transfert des formes aux bonnes dimensions moyennes. Par ailleurs, pour augmenter le champ des possibilités, on peut utiliser des épaisseurs plus importantes de résine qui constitue le motif initial, mais cela impacte fortement les capacités lithographiques et dégrade la rugosité. Une solution plus poussée est l'emploi d'empilements plus complexes, avec l'utilisation d'un « masque dur ». Le principe, visible en Figure I.13, est l'ajout d'un matériau entre la résine et le substrat. Dans un premier temps, le masque dur est gravé. La sélectivité à considérer est alors celle entre la résine et le masque dur, pour un procédé donné. Puis, un autre procédé est utilisé pour graver le substrat. Ainsi, c'est le dépôt intermédiaire du masque dur qui doit être adapté, en composition et en épaisseur pour permettre de graver le substrat selon la profondeur souhaitée.

Pour la gravure l'empilement est donc important, tant sur sa morphologie que sur sa composition. On note cependant que ces empilements plus complexes nécessitent, en plus de leur gravure à l'aide de multiples procédés, des étapes de retrait des couches supérieures pour ne laisser, au final, que le motif

dans le matériau voulu. Il s'agit là d'autant d'étapes qui impactent la rugosité, tant vers une amélioration qu'une détérioration.

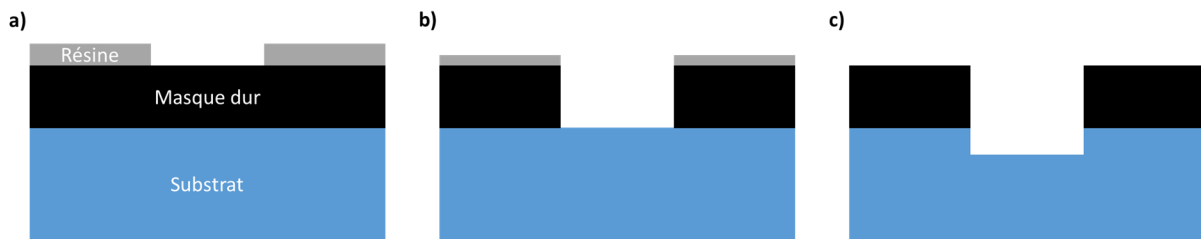


Figure I.13 : Principe d'utilisation d'un masque dur, a) état initial, b) ouverture du masque dur, c) gravure du substrat.

Finalement, il est important de comprendre que la gravure fixe la forme 3D du motif. Même si l'objet a déjà une épaisseur dans les couches supérieures, c'est bien la gravure qui détermine au final : la hauteur, le SWA et la rugosité.

I.2.3 La métrologie

La métrologie correspond littéralement à la science de la mesure, ce qui implique à la fois les connaissances des techniques utilisées pour faire la mesure, mais aussi les méthodes liées ainsi que leurs biais et limites. Dans le cadre du patterning, la métrologie prend une place de plus en plus importante du fait de la précision qu'elle requiert et du temps de mesure qu'elle implique. En effet, la réduction de la taille des transistors, ainsi que la création de nouvelles architectures, ont conduit à l'augmentation du nombre de mesures mais aussi le besoin d'une plus grande justesse, qui correspond à la proximité de la mesure avec la réalité, et d'une meilleure fidélité, qui correspond à la variation lors de la réalisation d'une même mesure, l'IRDS [11] pointe la nécessité de mesures sub-nanométriques. En termes de morphologie on remarque, sur la vue en coupe présentée en Figure I.6 que pas moins de 4 paramètres sont nécessaires pour décrire une structure comme celle-ci, mais ce nombre est encore en augmentation avec la prise en considération des imperfections comme par exemple les arêtes du motif (corner rounding) [29].

En microélectronique on peut classer les techniques de métrologies selon plusieurs types. Tout d'abord la métrologie « in-situ » concerne le fait de mesurer, pendant qu'un procédé opère, cette méthodologie permet notamment de détecter la fin de gravure. Les espèces volatilisées durant ce procédé sont analysées en temps réel par spectroscopie. Il est donc possible de repérer la présence d'un composé chimique créé lors de la gravure et ainsi de déterminer la fin de cette gravure, lorsque cette espèce chimique n'apparaît plus. On peut noter que la métrologie in-situ est aussi très développée pour les mesures de contraintes des matériaux et elle est présente dans le domaine de la lithographie, avec des mesures déportées, par exemple le suivi et la correction des équipements en temps réel (maintien de l'énergie de la source de lumière, correction de positionnements) [30].

Dans le cas où la métrologie est faite *a posteriori* de l'étape, comme visible sur la Figure I.5, on peut distinguer « l'off-line » qui correspond à la mesure d'un objet qui ne sera pas utilisé pour la suite du procédé, *a contrario* de « l'in-line » [31]. Dans le cas de mesure in-line, la condition principale est la non détérioration de l'objet mesuré, mais aussi du substrat de départ qui doit pouvoir rester dans la chaîne de fabrication. La première condition est souvent loin d'être respectée [32], un impact restreint par rapport à la suite du procédé est donc souvent accepté.

Par ailleurs, les buts de la métrologie peuvent être très différents suivant les utilisateurs (industriel, développeur, chercheur) et peuvent donc induire l'utilisation d'une technique ou d'une autre pour l'obtention d'un même paramètre. On catégorisera donc, de la plus restrictive à la plus généraliste, les métrologies ayant pour but :

- **Le contrôle** : lors de production industrielle, la validation de la qualité du produit est le point central. Il n'est pas nécessaire d'obtenir une mesure chiffrée, mais simplement une information concernant le dépassement d'une limite, fixée par rapport à la cible visée ou une référence précédemment qualifiée.
- **La comparaison** : pour les développements de procédés, c'est une comparaison d'une étape à une autre qui peut être utilisée. La justesse (voir Chapitre II) n'est pas recherchée. Les métriques sont alors relatives et signées. Le résultat est souvent représenté en pourcentage. L'outil et la procédure de mesure doivent rester identiques pour ne pas influencer le résultat.
- **La quantification** : cette métrologie, plus générique, a pour but de donner un chiffre qui vient donc apporter une quantification. De plus, à l'aide d'outils statistiques, cette métrologie donne des informations qualitatives sur notre mesure. La notion de fidélité de mesure (voir Chapitre II) est alors utilisée.

Dans le cadre des travaux présentés dans ce manuscrit, la métrologie de quantification est utilisée. Par sa flexibilité, ce type de métrologie permet de comparer des mesures obtenues sur divers équipements, mais aussi de confronter celles réalisées par différents instituts. Il est de plus possible d'évaluer la différence par rapport à une valeur nominale.

I.3 La rugosité de ligne

Comme explicité précédemment, il existe divers types de défauts dont la rugosité. On note, dans cette catégorie, que la « rugosité de ligne » correspond à la rugosité sur la longueur de la ligne et non celle sur la hauteur de celle-ci, nommée SWR pour Side-Wall Roughness en anglais. Il y a plusieurs moyens de caractériser la rugosité de ligne quantitativement (voir Chapitre II). Quelle que soit la méthode employée il est notable que la rugosité a, de manière nominale, diminué avec l'avancement des nœuds technologiques, alors que le ratio rugosité de ligne/CD a fortement augmenté ces dernières années. En effet, celui-ci était aux alentours des 3% dans les années 2000 [33, 34], alors que sur les derniers chiffres de 2018 ce ratio est de plus de 10% [35], avec une valeur nominale de rugosité de 2.5 nm.

Le principe de défaut doit être relié aux problématiques qu'il pose, et non seulement au simple fait que ce n'est pas le résultat attendu. Dans le cas de la rugosité de ligne, la littérature fait mention de trois grandes catégories de problèmes :

- **L'augmentation de la consommation** des transistors de par une élévation du courant de fuite à l'état « bas » [34, 36]. Comme illustré en Figure I.14, on remarque une perte de fiabilité des composants dans les deux cas : (i) avec des petites grilles (représentées en rouge) menant à des fluctuations à l'intérieur même d'un dispositif et (ii) avec des grilles larges induisant des variations entre différents dispositifs [37, 38].

- **Une variation du courant drain**, i.e. des performances qualitatives des transistors, due au rugosité de ligne a été démontrée [39, 40].
- **Une détérioration des interconnexions** dont les performances sont impactées par la rugosité [41].

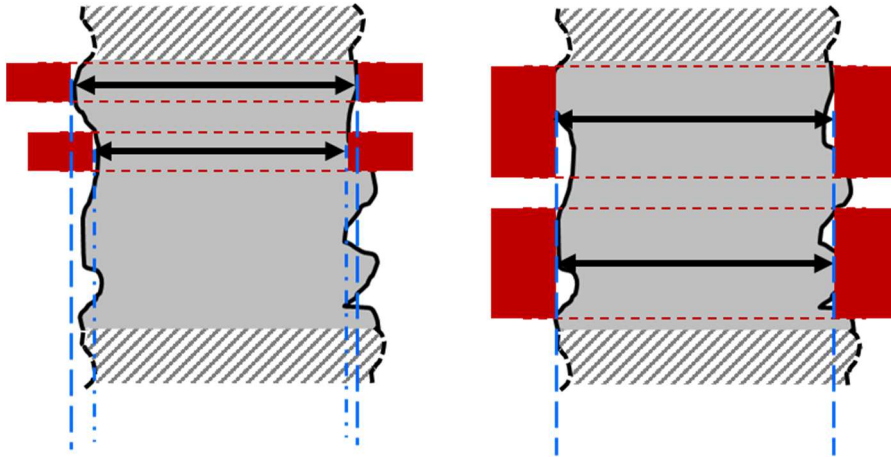


Figure I.14 : Schématisation en vue du dessus de la variation du CD (fleches noires) impactant la grille (en rouge). Dans le cas où la grille est fine, à gauche, le CD change dans une même puce (fluctuation intra-composant). Dans le cas où la grille est large, à droite, le CD ne fluctue pas dans le composant mais entre composants (fluctuation inter-composants).

Par ailleurs, diverses études cherchent à comprendre la provenance de la rugosité, ou du moins à attribuer à chaque étape sa part dans la valeur finale de rugosité. Parmi elles, plusieurs sont focalisées sur la rugosité induite par la lithographie au niveau du masque [42], mais aussi au niveau de la chimie de la résine [43–46]. De même, il existe des recherches relatives à l’impact de la gravure sur la rugosité de ligne [47], notamment celle de T. Wallow [48] qui montre la variation de rugosité avant et après gravure.

Au fil des avancées sur la problématique de rugosité de ligne, le fil conducteur reste la diminution de cette rugosité. En effet, comme vu précédemment, l’impact de celle-ci est multiple et sa réduction entraînerait donc une multitude d’améliorations dans les appareils électroniques actuels, notamment sur leur consommation d’énergie qui serait grandement réduite. Dans cette idée, en plus de voir l’impact des étapes courantes sur la rugosité de ligne comme montré précédemment, des étapes supplémentaires ont été introduites, souvent entre la lithographie et la gravure, et parfois après cette dernière, avec pour seul but de réduire la rugosité sans changer les autres caractéristiques des motifs [49–51].

Dans le cadre de cette réduction de la rugosité de ligne, il semble évident que la mesure de ce paramètre se doit d’être la plus juste et précise possible (voir Chapitre II). Ainsi, le développement des diverses métrologies pour la rugosité de ligne est présenté Chapitre III. Comme pour toute autre métrique, les feuilles de route de la microélectronique formulent des objectifs quant à la valeur de rugosité maximale acceptable pour les nouveaux nœuds technologiques. La dernière recommandation de l’IRDS de 2017 [11] prévoyait la nécessité d’obtenir une rugosité de ligne sub-nanométrique, plus exactement inférieure à 0.6 nm pour le nœud 7 nm, en 2019. Ceci induit une résolution de mesure à l’Angstrom. C’est dans ce contexte que ces travaux prennent place.

I.4 Mise en place d'une réponse à la problématique de métrologie de rugosité de ligne

Diverses techniques sont aujourd'hui utilisées en microélectronique, pour obtenir les informations morphologiques déjà évoquées. Dans le cadre de la rugosité de ligne, le microscope électronique à balayage de contrôle dimensionnel (CD-SEM, voir Chapitre III) est le plus courant. Cependant, des biais lui sont attribués de par la physique qu'il utilise. Ainsi dans ce domaine, il est habituel de prendre pour outil de référence la microscopie à force atomique (AFM, voir Chapitre III). Ces deux techniques, développées depuis plus d'une trentaine d'années, sont désormais limitantes au vu des besoins de justesse et de précision (Chapitre II). De ce fait, d'autres techniques telles que la réflectométrie (OCD, voir Chapitre III), déjà très présente dans l'industrie mais non développée pour la mesure de rugosité de ligne, ou encore la diffusion des rayons-X aux petits angles (SAXS, voir Chapitre III), technique dite « de laboratoire » mais émergente dans la microélectronique pour la mesure morphologique, sont proposées par les experts [52].

Ainsi en réponse globale à cette problématique, le développement de nouvelles méthodologies est nécessaire. Cela passe par l'amélioration des techniques actuelles en les rendant disponibles pour ce type de mesures. De plus, l'utilisation de l'hybridation ou « fusion de données », aussi indiquée par l'IRDS, semble pertinente.

L'hybridation, déjà utilisée dans de multiples secteurs, a déjà fait l'objet de recherches en microélectronique, notamment pour l'accès à une mesure de CD ou de profil plus fiable, et avec une meilleure résolution [53, 54]. Le principe de l'hybridation correspond dans notre cas, comme représenté en Figure I.15, à l'utilisation de différentes techniques apportant une ou plusieurs informations morphologiques sur le motif, pouvant être redondantes ou non. Ces informations sont traitées en connaissance des autres afin d'obtenir, au final, un résultat complet comprenant toutes les informations en une unique fois, plus juste et plus fiable.

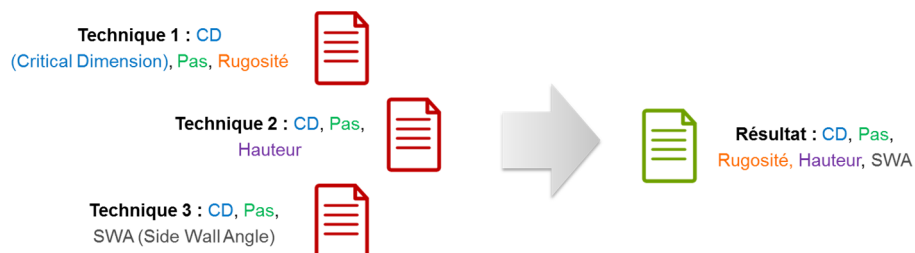


Figure I.15 : Principe général de l'hybridation, explicité dans le cadre des données morphologiques disponibles en microélectronique.

L'utilisation de différentes techniques peut mener à un temps de traitement plus important. Toutefois cela peut être compensé par des mesures plus rapides. En effet, en utilisant chacune des techniques au niveau de sa limite de capacité, une partie de l'information récupérée est inexploitable car trop entachée d'erreurs. *A contrario*, en limitant la mesure aux capacités réelles d'un équipement, et à son domaine d'application, on obtient moins d'informations mais celles-ci sont plus pertinentes nous le verrons dans le Chapitre III et beaucoup plus rapides à obtenir.

I.5 Résumé du chapitre

Ce chapitre a permis d'introduire le domaine de la microélectronique, en montrant son évolution depuis son apparition et jusqu'à aujourd'hui. Ainsi, nous nous sommes intéressés aux grandes problématiques historiques, toujours d'actualité, à savoir l'augmentation de la densité et la diminution du coût par puce, avec l'arrivée de nouvelles architectures. La problématique de défauts a aussi été explicitée au travers des défektivités dues aux erreurs de placement, ainsi qu'à la rugosité et aux effets stochastiques.

Par la suite, nous avons parcouru les étapes liées au patterning (lithographie, gravure et métrologie). Ces étapes demandant toujours plus d'innovation, nous avons présenté succinctement les techniques mises en place aujourd'hui avec, entre autres pour la lithographie, celles utilisées lors de ces travaux de recherche.

La rugosité de ligne a été explicitée en termes d'impact sur les composants finaux, mais aussi comme un défi vis-à-vis de la métrologie. Ceci nous a conduits à la nécessité de mesurer la rugosité de ligne avec une précision sub-nanométrique. Ainsi, la dernière partie a exposé la vision globale que nous avons mise en place pour répondre à cette problématique. Cette réponse se résume dans l'utilisation de diverses techniques de métrologie, chacune développée ou à développer pour la mesure de rugosité, combinées toutes ensembles pour pallier leurs limites individuelles.

Chapitre II - Métrologie de rugosité de ligne

Dans cette partie, les concepts fondamentaux de la métrologie sont décrits afin de comprendre le besoin d'un nombre de mesure suffisamment grand. Cela permet aussi d'appréhender le principe de la calibration et du raccordement des outils.

Par la suite, la problématique de la métrologie pour la rugosité de ligne est introduite. La rugosité est décomposée sous forme de différentes métriques et le principe de corrélation des bords est abordé. Cette partie montre aussi qu'un seul paramètre n'est pas suffisant pour décrire entièrement la rugosité de ligne et qu'il est nécessaire d'utiliser des modèles plus complexes.

L'accent est finalement mis sur l'utilisation de la PSD (Power Spectral Density) pour caractériser la rugosité, notamment sur les capacités de ce modèle, et son comportement suivant les paramètres utilisés. Les limites de ce modèle sont mises en évidence.

II.1 Concept de la métrologie

II.1.1 Justesse de mesure

Dans le cadre de la métrologie, la justesse correspond à l'accord de la mesure avec une référence [55]. La justesse de mesure est souvent confondue avec « l'exactitude de mesure » que l'on peut communément appeler la « valeur vraie », autrement dit la dimension unique et absolue. Il est important de comprendre que cette « valeur vraie » est inaccessible car toute technique de mesure comporte une ou plusieurs erreurs, d'où la notion de justesse. La représentation d'une mesure juste ou non, donnée sur la Figure II.1, laisse entendre que l'on connaît la valeur que l'on souhaite mesurer (centre de la cible). Dans le cas de procédés de fabrication tels que la lithographie, cela peut correspondre à la valeur prévue par le dessin. De ce fait, pour obtenir la justesse de mesure d'un équipement, les métrologues utilisent le principe de raccordement avec étalonnage.

Ce raccordement consiste à mesurer, sur l'outil à calibrer, des étalons de référence. Ces étalons ont une valeur certifiée, obtenue préalablement grâce à des mesures utilisant diverses techniques et effectuées par divers instituts. Il est ainsi possible de définir le biais de l'outil, erreur systématique à la référence, et de le corriger ou au moins de le prendre en compte. On note cependant que ce décalage peut varier suivant le type de motif mesuré, le biais n'étant pas forcément une constante sur la totalité de la gamme de mesures de l'outil. En effet, le biais est très souvent similaire à une droite liée à la sensibilité mais peut, dans des cas plus complexes, être non linéaire. Il est nécessaire, dans ce dernier cas, d'effectuer un nombre de mesures important en termes de diversité des motifs, pour couvrir la gamme de mesures de l'outil.

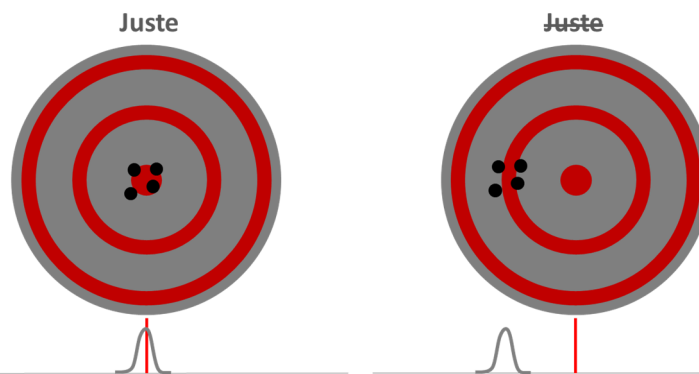


Figure II.1 : Représentation du principe de justesse de mesure au travers de cibles dont le centre rouge est la valeur visée, et les points, les mesures obtenues. Visualisation sous chaque cas de la densité de probabilité associée.

La notion de justesse correspond donc à une cascade de référencements : de la machine à l'étalon, lui-même mesuré par un équipement qui est calibré sur un étalon primaire, etc. La justesse permet de comparer les mesures entre elles, aussi bien à petite échelle (par exemple des équipements différents dans une même entreprise), qu'à grande échelle (d'une entreprise ou d'un institut à un autre).

II.1.2 Fidélité de mesure

Cette notion fait référence aux différences obtenues lors de la répétition d'une même mesure [55]. Le terme de « precision », utilisé en anglais, est souvent évoqué pour parler de fidélité. Il s'agit là d'une valeur numérique souvent exprimée comme un écart-type, moyenne de l'écart à la moyenne, mais d'autres calculs d'expression de la fidélité existent, coefficient de variation par exemple. Ainsi, comme le montre la Figure II.2, la fidélité est représentative de la largeur de la dispersion autour d'une valeur moyenne. On ne parlera pas de mesure fidèle ou non, mais plutôt de faible ou haute fidélité, en donnant un chiffre ou un ordre de grandeur.

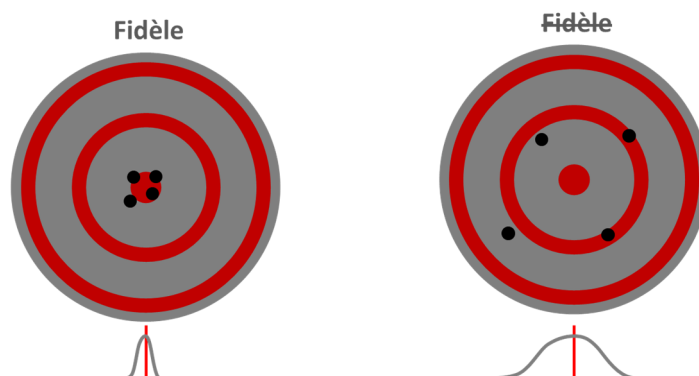


Figure II.2 : Représentation du principe de fidélité de mesure au travers de cibles dont le centre rouge est la valeur visée, et les points, les mesures obtenues. Visualisation sous chaque cas de la densité de probabilité associée.

On voit ici, avec les Figure II.1 et Figure II.2, que les notions de justesse et de fidélité sont indépendantes. La fidélité, contrairement à la justesse, est associée à une erreur aléatoire (non systématique). Ce concept induit donc d'effectuer une multitude de mesures d'un même objet, afin de pouvoir déterminer la fidélité de mesure comme une distribution de valeurs, d'où le terme d'erreur. Ce dernier est attribuable à la fluctuation du procédé de mesure et peut être imputable à cinq facteurs : l'opérateur, la procédure, le système de mesure, l'environnement et le temps entre deux mesures. Ainsi on séparera deux principes qui sont la répétabilité et la reproductibilité.

II.1.2-A La répétabilité

Pour effectuer un test de répétabilité, il faut être dans des conditions qui ne changent pas au niveau opérateur, procédure, système de mesure et environnement [55]. Contenir tous les paramètres d'environnement étant difficile, ce facteur est ramené au temps, laissant imaginer qu'en un temps court l'environnement change peu. Cette notion de temps court, afin de conserver l'état de l'équipement, est très dépendante du type de mesure et des paramètres d'environnement considérés. Cela peut être de l'ordre de la seconde, comme celui de la journée.

Par ailleurs, suivant les cas, on peut distinguer la répétabilité dite « statique » de celle dite « dynamique ». Ces deux types permettent une caractérisation des différents éléments composant un équipement. En effet, dans le cas de répétabilité dynamique on va, par exemple, effectuer à nouveau la totalité de l'alignement, permettant de centrer la mesure sur le motif et donc déplacer l'objet à mesurer, tandis que dans le cas statique, l'objet ne sera pas déplacé, d'où le terme « statique ». Ainsi, les erreurs d'alignement de l'équipement seront présentes dans la métrique de répétabilité dynamique, et non dans celle de répétabilité statique.

II.1.2-B La reproductibilité

Dans le cas de la reproductibilité, seul l'objet est le même ou est similaire. Les cinq facteurs précédemment cités peuvent être différents ou identiques, il convient alors de le préciser afin de connaître les conditions de mesure. Pour la reproductibilité, il est courant de parler de temps long, malgré le fait que ce ne soit pas une nécessité. Il est imaginable d'employer le terme reproductibilité lors de mesures en parallèle, d'objets considérés comme similaires, sur différents équipements. Pour exemple, on peut citer le principe de reproductibilité lors de comparaisons de mesures entre divers instituts [56].

II.2 Application à la rugosité de ligne

La rugosité de ligne est liée au procédé de fabrication et, comme déjà évoqué précédemment, elle a un impact néfaste sur la consommation énergétique ainsi que sur les performances des composants (Chapitre I). Il est donc important de mesurer ce paramètre afin de pouvoir, si ce n'est le réduire, au moins le contrôler. Les métriques liées à la rugosité sont diverses. Il faut tout d'abord définir les types de rugosités de ligne.

II.2.1 Types de rugosités

Dans la définition actuelle de la rugosité de ligne, on considère ce paramètre suivant la longueur de la ligne, sans tenir compte de sa variation possible sur la hauteur. Comme l'illustre la Figure II.3 (vue du dessus), la rugosité de ligne peut être caractérisée suivant deux types. Tout d'abord, la Line Width Roughness (LWR), rugosité de la largeur de ligne en français, qui correspond aux fluctuations de la largeur de la ligne sur toute son étendue. Le deuxième type, la Line Edge Roughness (LER), traduit par rugosité d'un bord de la ligne, qui correspond à la variation de position d'un seul bord, et ce par rapport à une ligne droite moyenne. Ainsi, la LER peut-être dite gauche ou droite, suivant le bord considéré. Dans un grand nombre de cas, les deux bords ont une signature de rugosité de type LER identique, mais si un procédé est asymétrique, comme dans le cas du SADP (Chapitre I) où le motif final n'a pas subi les mêmes procédés sur ses deux côtés, alors les deux LER de gauche et de droite peuvent être différentes.

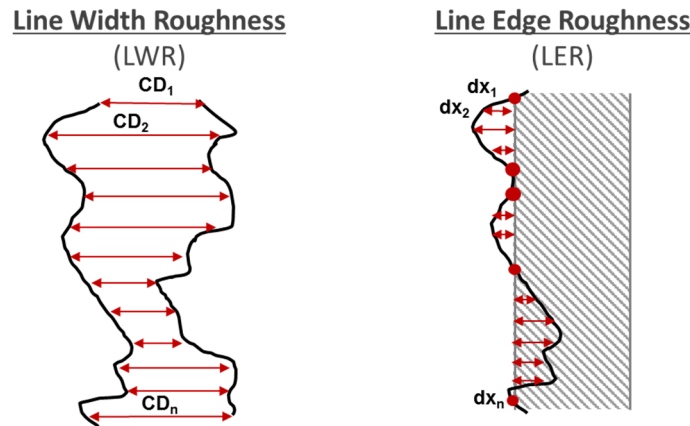


Figure II.3 : La rugosité de ligne vue comme : la variation de dimension de la ligne CD suivant la position sur la longueur de la ligne, à gauche (LWR), et le déplacement d'un bord dx par rapport à une référence, ici une ligne parfaite, à droite (LER).

II.2.2 Métrologie standard

La rugosité de ligne étant vue comme une variation autour d'une valeur visée, elle est naturellement associée à la notion d'écart-type, à savoir la racine de la moyenne des carrés des écarts à la moyenne. Cela se traduit mathématiquement, pour les deux types de rugosité, par :

$$\sigma_{LWR} = \sqrt{\frac{1}{N} \sum_{i=1}^N (CD_i - \overline{CD})^2} \quad (\text{II.1})$$

où N est le nombre de mesures, CD_i les diverses mesures de tailles de lignes, et \overline{CD} la moyenne de ces mesures, et :

$$\sigma_{LER} = \sqrt{\frac{1}{N} \sum_{i=1}^N dx_i^2} \quad (\text{II.2})$$

où dx_i sont les mesures de la distance entre le bord de la ligne et la référence, comme indiqué sur la Figure II.3. Dans le cas LER, la moyenne n'apparaît pas car elle est nulle, puisqu'il s'agit de notre ligne de référence.

II.2.3 La corrélation de bord

Il existe un lien mathématique entre la LER et la LWR. Ce lien est physiquement dû à la corrélation des bords, on peut ainsi trouver les cas spécifiques présentés en Figure II.4. Dans le cas de bords « corrélés » où le CD est toujours identique, on observe un déplacement des bords en phase. Dans le cas où les bords sont dits « anti-corrélés », le déplacement d'un bord dans une direction est relié au déplacement du second dans le sens inverse. Et enfin, dans le cas de bords « non corrélés », où les déplacements d'un bord comme de l'autre sont aléatoires, on considèrera que les bords n'ont pas d'influence l'un sur l'autre.

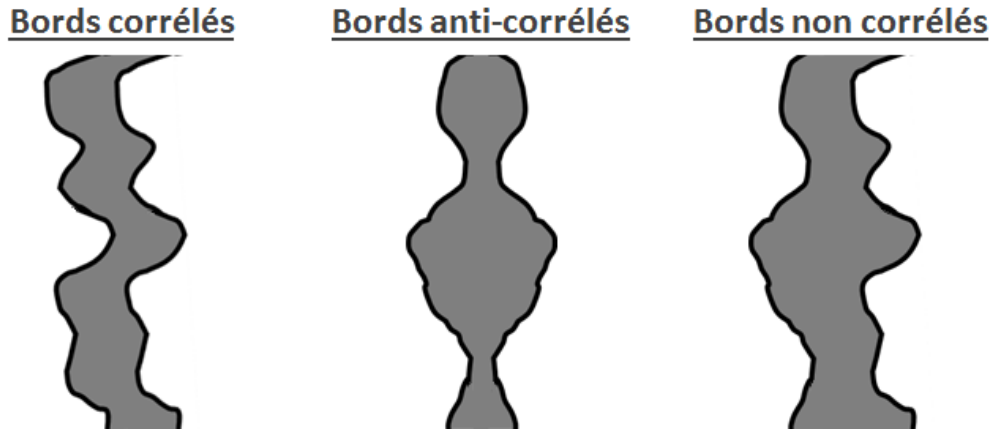


Figure II.4 : Illustration de la corrélation des bords d'une ligne pour les cas remarquables.

A gauche les bords corrélés, les déplacements sont les mêmes à droite et à gauche ; au centre les bords anti-corrélés, les bords se déplacent en sens inverse ; à droite les bords non corrélés qui se déplacent indépendamment l'un de l'autre.

Avec cette notion de corrélation des bords, il est possible de définir une relation mathématique entre la LER et la LWR :

$$\sigma_{LWR}^2 = \sigma_{LERg}^2 + \sigma_{LERd}^2 - 2 \times CC \times \sigma_{LERg} \times \sigma_{LERd} \quad (II.3)$$

Avec $CC \in [-1 : 1]$, le coefficient de corrélation entre les deux bords, σ_{LWR}^2 l'écart-type précédemment explicité dans l'Equation II.1, σ_{LERg}^2 et σ_{LERd}^2 les métriques d'écart-type exprimées respectivement pour le bord gauche et droit. Dans le cas où les deux bords ont une signature similaire, $\sigma_{LER}^2 = \sigma_{LERg}^2 = \sigma_{LERd}^2$, cette équation peut se simplifier par :

$$\sigma_{LWR}^2 = 2 \times \sigma_{LER}^2 \times (1 - CC) \quad (II.4)$$

L'Equation II.4 permet de retrouver les cas atypiques :

- **Bords corrélés**, $CC = 1 \rightarrow \sigma_{LWR} = 0$
- **Bords anti-corrélés**, $CC = -1 \rightarrow \sigma_{LWR} = 2 \sigma_{LER}$
- **Bords non-corrélés**, $CC = 0 \rightarrow \sigma_{LWR} = \sqrt{2} \sigma_{LER}$

Le coefficient de corrélation induit la notion de « wiggling », littéralement traduite par « tortillant », qui reflète le déplacement de la ligne en dehors des variations de dimensions. Ainsi, certaines publications introduisent le terme de Line Center Roughness (LCR), i.e. la rugosité du centre de ligne [57], et appliquent un calcul d'écart-type au métrique du centre de la ligne (moyenne de la position des deux bords), donnant dans le cas le plus général :

$$\sigma_{LCR}^2 = \frac{1}{4} \times \sigma_{LWR}^2 + CC \times \sigma_{LERg} \times \sigma_{LERd} \quad (II.5)$$

pouvant se simplifier grâce à l'Equation II.3 par :

$$\sigma_{LCR}^2 = \frac{1}{4} (\sigma_{LERg}^2 + \sigma_{LERd}^2) + \frac{1}{2} CC \times \sigma_{LERg} \times \sigma_{LERd} \quad (II.6)$$

Soit, dans le cas où les deux bords ont une signature similaire, $\sigma_{LER}^2 = \sigma_{LERg}^2 = \sigma_{LERd}^2$:

$$\sigma_{LCR}^2 = \frac{1}{2} \times \sigma_{LER}^2 \times (1 + CC) \quad (II.7)$$

Ceci induit que la métrique LCR est déjà exprimée dans la métrique LER, à condition de connaître le coefficient de corrélation. Ainsi les mesures de LWR et LER permettent de retrouver la LCR. Plus simplement, il est possible d'écrire dans ce cas l'Equation II.5 sous la forme :

$$\sigma_{LER}^2 = \frac{1}{4} \times \sigma_{LWR}^2 + \sigma_{LCR}^2 \quad (II.8)$$

Cela correspond, finalement, à une description générale où la LER se compose à la fois des variations de la dimension de la ligne LWR et du déplacement de la ligne LCR.

II.3 Vers une description plus complète de la rugosité

Au cours des avancées sur la mesure de rugosité en général, il a été montré que l'expression de la rugosité, comme un simple écart-type, n'était plus suffisante pour refléter la physique. Comme le suggère la Figure II.5, il est possible de mesurer la même valeur d'écart-type pour deux signaux ayant un comportement différent. Or, la rugosité n'a pas la même influence selon son amplitude et sa forme [37].

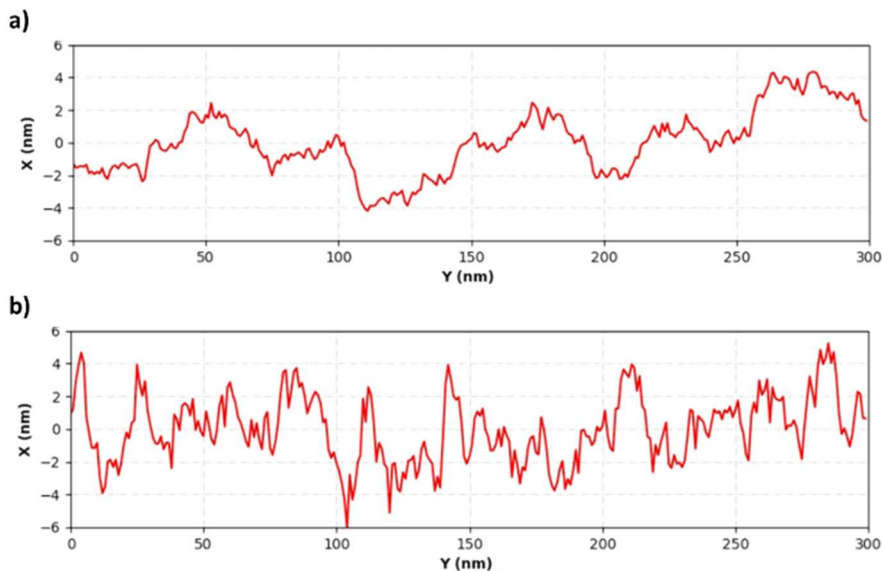


Figure II.5 : Représentation de la position de deux bords de ligne en fonction de la position le long de celle-ci pour une valeur d'écart-type identique $\sigma = 2$ nm mais un comportement fréquentiel différent.

II.3.1 Notion de corrélation et de fractale

Afin d'expliciter le comportement de la rugosité, le principe des fractales est, depuis plus de trente ans, une référence. Ce concept introduit par Mandelbrot [58] est utilisé dans de nombreux domaines. Cette notion est explicitée par le fait que chacune des sous-parties d'un ensemble a des similarités avec le tout, on retrouve cette similarité à toutes les échelles. Le principe de la dimension fractale réside dans le fait que l'objet a une dimension supérieure N , à sa dimension topologique n . Ce concept est généralisable par $N = n^D$ où D est la dimension fractale. On peut ainsi calculer $D = \log(N) / \log(n)$. Pour exemple, dans le cas d'un carré, l'augmentation de sa largeur induit un changement de son aire à la puissance deux ; sa dimension fractale $D = \log(x^2) / \log(x) = 2$ (x étant le facteur de multiplication de sa taille). Dans le cas du triangle de Sierpiński, objet fractal très connu et présenté en Figure II.6, le changement de taille par deux induit la création de trois triangles soit $D = \log(3) / \log(2) \approx 1.58$. Cette

grandeur non triviale représente finalement l'irrégularité d'une structure, d'où son utilisation pour décrire la rugosité. Plus la dimension fractale sera élevée, plus l'aspect du profil sera irrégulier.

Par ailleurs, il est possible de relier la dimension fractale à la notion de processus stochastique dit « auto-affine » [59]. Cette notion implique que, lorsque l'on augmente la dimension d'observation, l'allure du signal semble être dilatée d'un facteur r^α , ainsi :

$$X(rt) \equiv r^\alpha X(t) \quad (\text{II.9})$$

où X est un processus stochastique, \equiv désigne l'égalité au sens des distributions, et $\alpha = d+1 - D$, « d » étant la dimension topologique. Le terme α est ici nommé « exposant de rugosité » (ou « exposant de Hurst »).

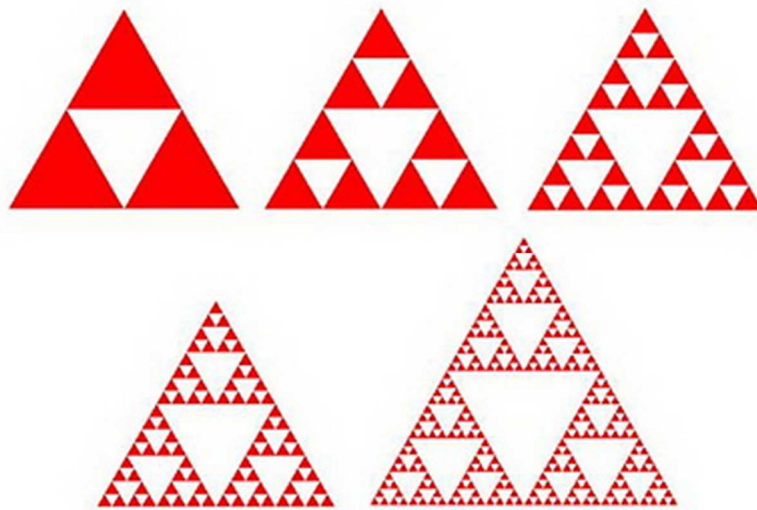


Figure II.6 : Schématisation du triangle de Sierpiński.

Il est remarquable que, dans les cas de la rugosité de ligne, plus la longueur du signal mesuré (L) est grande, plus la valeur de l'écart-type augmente suivant une loi L^α . On retrouve ici le phénomène auto-affine explicité précédemment. Cependant, dans le cas de rugosité de ligne, ceci est limité. En effet, comme le montre la Figure II.7, l'écart-type mesuré augmente jusqu'à une valeur de saturation ($\sigma_{\text{saturation}}$), elle-même associée à une longueur de saturation L_s . Il a été montré que cette longueur est liée à ce que l'on appelle la longueur de corrélation ξ , à savoir le fait que deux points sur le signal peuvent être considérés comme totalement indépendants au-delà de cette longueur [60]. Ceci a mené à la création de divers modèles, basés sur l'utilisation de fonctions de traitement du signal, comme nous allons le voir dans la prochaine partie.

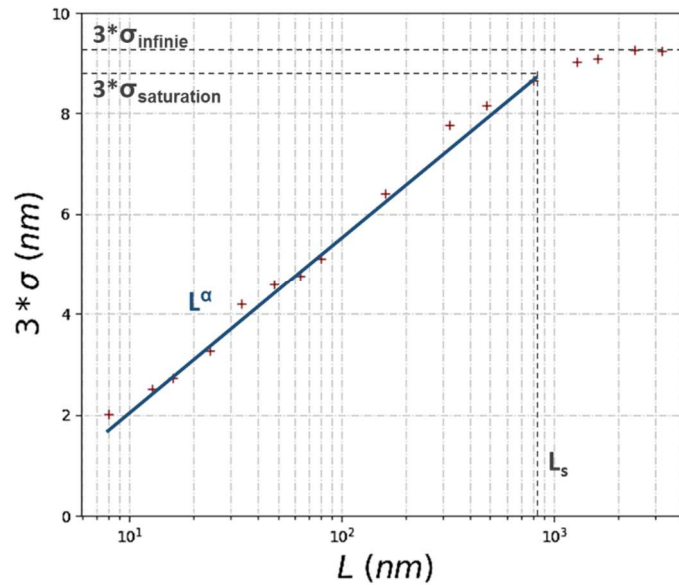


Figure II.7 : Evolution de la mesure de l'écart-type suivant la longueur de la ligne considérée (Mesure CD-SEM sur 9 lignes issues d'une lithographie par NanoImpression).

II.3.1-A Fonction de corrélation des hauteurs

Communément appelée HHCF (pour Height-Height Correlation Function) [60], elle tire son nom de l'analyse de rugosité de surface où la hauteur est le paramètre considéré. Cette fonction peut être appliquée pour d'autres paramètres que la hauteur. Ainsi, de manière générale, pour un processus stochastique X , on a :

$$G(u) = \langle [X(t) - X(t + u)]^2 \rangle \quad (\text{II.10})$$

où $\langle \dots \rangle$ est la fonction moyenne et u est la distance entre deux points de mesure. En appliquant les concepts de saturation et auto-affine, on peut établir que :

$$G(u) \propto \begin{cases} u^{2\alpha} & \text{quand } u \ll \xi \\ \sigma^2 & \text{quand } u \gg \xi \end{cases} \quad (\text{II.11})$$

Ce sur quoi s'appuie Sinha *et al.* [61] pour développer le modèle suivant :

$$G(u) = 2\sigma^2 \left(1 - e^{-|u/\xi|^{2\alpha}} \right) \quad (\text{II.12})$$

où l'exposant de rugosité α est compris entre 0 et 1, les valeurs faibles (proches de 0) correspondent à des surfaces moins lisses, contrairement aux fortes valeurs.

II.3.1-B La fonction d'autocorrélation

Cette fonction, plus connue en anglais par le terme ACF (AutoCorrelation Function) donne, la corrélation d'un signal avec sa copie, décalée d'une distance u . Cette distance est communément appelée « lag ». La fonction d'autocorrélation d'un processus stochastique X (taille de la ligne ou sa position dans nos cas spécifiés précédemment) est définie par :

$$R(u) = \langle [X(t) - \langle X(t) \rangle] [X(t + u) - \langle X(t) \rangle] \rangle \quad (\text{II.13})$$

avec $\langle X(t) \rangle$ la moyenne. On retrouve l'écart-type donné précédemment pour $R(0) = \sigma^2$.

L'ACF et HHCF sont reliées par :

$$G(\mathbf{u}) = 2[\sigma^2 - R(\mathbf{u})] \quad (\text{II.14})$$

Ainsi, l'Equation II.14 associée à l'Equation II.12 permet de trouver un modèle pour l'ACF :

$$R(\mathbf{u}) = \sigma^2 e^{-|u/\xi|^{2\alpha}} \quad (\text{II.15})$$

II.3.1-C La densité spectrale de puissance

Enfin, la Densité Spectrale de Puissance (DSP ou PSD pour Power Spectral Density en anglais) d'un processus aléatoire est, d'après le théorème de Wiener-Khintchine, la transformée de Fourier de la fonction d'autocorrélation.

Pour rappel, la transformée de Fourier spatiale d'un processus aléatoire X est définie par :

$$F(\mathbf{k}) = \frac{1}{L} \int_{-L/2}^{L/2} X(\mathbf{y}) e^{-i\mathbf{k}\mathbf{y}} d\mathbf{y} \quad (\text{II.16})$$

avec L la longueur du signal considéré. On note que la transformée de Fourier donne un résultat dans l'espace des fréquences, et non dans l'espace physique, contrairement aux analyses précédentes. Il est possible d'exprimer la transformée de Fourier dans un cas discret (elle est alors appelée DFT pour Discrete Fourier Transform) :

$$F_n = \frac{1}{N} \sum_{j=0}^{N-1} (X_j - \langle X \rangle_N) e^{-ik_n j \Delta y} \quad (\text{II.17})$$

où N est le nombre de points du signal, Δy le pas entre deux mesures et k_n le nombre d'onde défini par :

$$k_n = \frac{2\pi}{L} n ; n = 0, 1, 2, \dots, N - 1 \quad (\text{II.18})$$

La DFT appliquée à notre problématique, nous donne la densité spectrale de puissance définie par :

$$P_n = F_n \{R_m\} \quad (\text{II.19})$$

où R_m est la fonction de corrélation définie par l'Equation II.15 et discrétisée, avec m correspondant au lag une fois discrétisé. On obtient, de manière discrète, l'expression du lien entre la PSD et l'ACF d'après les travaux de L. Azar-Nouche [62] décrivant la totalité du développement mathématique :

$$P_n = \frac{\Delta y}{2\pi N} \left[\sum_{m=0}^{N-1} (2 - \delta_m) R_m \cos(k_n m \Delta y) (N - m) \right] \quad (\text{II.20})$$

avec δ_m le symbole de Kronecker. En remplaçant R_m par son expression, on a alors le modèle PSD final :

$$P_n = \frac{\Delta y}{2\pi N} \sigma^2 \left[\sum_{m=0}^{N-1} (2 - \delta_m) e^{-\left|\frac{m\Delta y}{\xi}\right|^{2\alpha}} \cos(k_n m \Delta y) (N - m) \right] \quad (\text{II.21})$$

Il existe d'autres modèles développés pour décrire la PSD directement dans l'espace fréquentiel, comme le mentionne Palasantzas [63]. L'un d'eux est repris en 2013 par C. Mack et Al. [64] sous la forme :

$$PSD(\mathbf{k}) = \frac{PSD(0)}{[1 + (k\xi)^2]^{\alpha + \frac{1}{2}}} \quad (\text{II.22})$$

avec :

$$PSD(0) = \frac{2\sigma^2\xi}{2\pi} \left(\frac{\sqrt{\pi}\Gamma(\alpha+\frac{1}{2})}{\Gamma(\alpha)} \right) \quad (II.23)$$

où Γ est la fonction mathématique gamma, prolongation de la fonction factorielle pour les nombres complexes. Ces deux modèles PSD se rejoignent néanmoins dans le cas où $\alpha = 0.5$. De plus, dans les deux cas, il est possible de retrouver la valeur de l'écart-type correspondant à l'aire, sous la courbe visible en Figure II.8.c, et ce grâce au théorème de Parseval qui donne, dans le cas discret :

$$\sigma^2 = \frac{2\pi}{L} \sum_{n=0}^{N-1} P_n \quad (II.24)$$

Finalement, quel que soit le modèle utilisé : ACF, HHCF, PSD II.21 ou II.22, on retrouve comme le montre la Figure II.8, les deux régimes explicités précédemment. De plus, les limites du support (distance de décalage ou fréquence) sont liées aux mêmes paramètres, à savoir la longueur du signal mesuré ainsi que la résolution de la mesure.

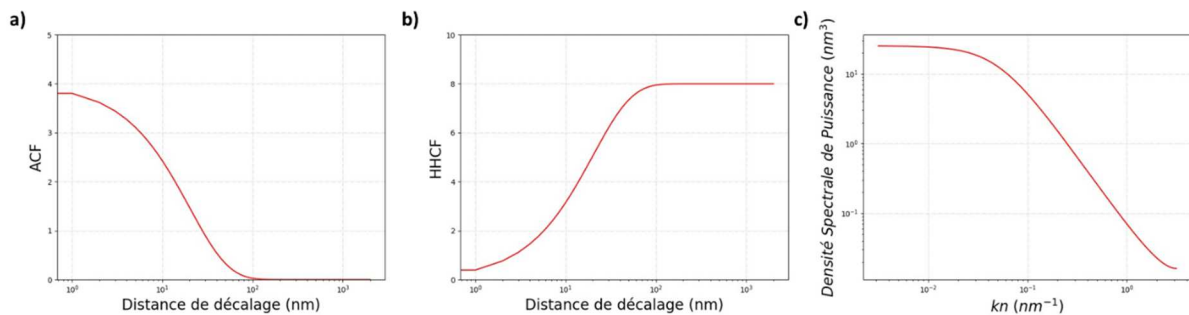


Figure II.8 : Représentation des différents modèles liés à la rugosité, d'un cas de paramétrage identique ($\sigma = 2$ nm, $\alpha = 0.5$, $\xi = 20$ nm), a) ACF (AutoCorrelation Function), b) HHCF (Heigh-Heigh Correlation Function), c) PSD (Power Spectral Density).

Afin de pouvoir conduire la suite de ces travaux, il a fallu adopter un modèle. Nous avons choisi la PSD décrite par L. Azar-Nouche, Equation II.21, pour les raisons décrites ci-dessous.

II.3.1 Le choix de la PSD

Diverses études comparent les modèles ACF, HHCF et PSD, soit deux par deux [60], soit récemment les trois en même temps [65]. Il en ressort que, sur des données dites propres (taille et nombre de lignes suffisants), les résultats sont similaires lors de l'utilisation de fit sur l'intégralité du domaine. Pour exemple, dans le cas où nous avons testé 10 000 signaux simulés (nombre inaccessible sur une vraie mesure), représentant 10 000 lignes d'une longueur de 2 μ m, échantillonnées à 1 nm, l'erreur par rapport aux valeurs nominales sur les trois paramètres, σ , α et ξ , est inférieure à 0.5% pour des ajustements avec les modèles HHCF et PSD. Dans le cas de l'ACF, l'erreur est respectivement de 0.7% 0.2% et 1.1% pour σ , α , ξ .

Cependant, on observe que la réduction du nombre de lignes tend à différencier les modèles. En effet, l'extraction obtenue dans notre cas sur seulement 10 lignes, comme le montre la Table II.1, entraîne des erreurs plus importantes sur le fit des modèles, notamment pour les paramètres α et ξ sur lesquels on obtient pour les modèles : PSD aux alentours de 1% d'erreur, ACF plus de 7% et HHCF près de 15%. L'extraction de σ reste, quant à elle, contenue, avec une erreur aux alentours de 1% pour les trois modèles.

	HHCF	ACF	PSD
Erreur σ	0.6%	1.3%	1.6%
Erreur α	14.8%	7.2%	1.0%
Erreur ξ	15.8%	8.4%	5.1%

Table II.1 : Erreur relative des paramètres obtenus après fit sur le domaine complet, avec les modèles HHCF, ACF et PSD (Explicité Equation II.21 identique à l'Equation II.22 dans ce paramétrage) pour 10 lignes générées ($\sigma = 2$ nm, $\alpha = 0.5$, $\xi = 20$ nm) (voir Chapitre IV).

Par ailleurs, lors de l'ajout de bruit de mesure sur les signaux générés, des différences significatives apparaissent par rapport au résultat précédent. On observe tout d'abord une augmentation d'amplitude sur toutes les distances ou fréquences, et ce pour tous les modèles. Comme le précise C. Mack [65], ce biais étant une constante, il est corrigé. Cependant, d'après cette même étude, d'autres biais tels que la correction de dérive du signal ou encore le filtrage dû à la technique de mesure sont plus aisément corrigés sur le modèle HHCF que sur les autres, car ils se répercutent dans ce cas comme des constantes. Malgré cela, on remarque dans cette publication que l'HHCF n'est pas stable pour des distances de décalage supérieures à la moitié du signal. De plus la représentation fréquentielle de la PSD permet, en dehors de l'utilisation du modèle, une détection de phénomènes tels qu'une fréquence caractéristique d'une erreur systématique, et ce sous la forme d'un pic. On notera aussi, dans le Chapitre III, que la PSD est accessible de manière directe avec la technique SAXS, contrairement aux autres modèles.

En prenant en compte toutes ces considérations, la PSD semble donc être un très bon candidat pour la caractérisation et la représentation de la rugosité. Il reste cependant à comprendre la différence entre les deux modèles de PSD.

Tout d'abord le modèle promu par C. Mack, défini en fréquence, a l'avantage d'être continu dans cet espace de représentation, contrairement à la PSD décrite par L. Azar-Nouche. Il n'y a alors pas de lien avec l'échantillonnage dans le domaine spatial.

Cependant cette force du modèle est aussi sa faiblesse. Dans le cas de mesures réelles, le calcul de la PSD est biaisé par l'échantillonnage ainsi que la fenêtre de mesures, comme le décrit C. Mack lui-même [64]. Après le calcul de la PSD, les données nécessitent donc un grand nombre de corrections [66, 67] permettant de retrouver le modèle défini par C. Mack. A l'opposé de cela, le modèle proposé par L. Azar-Nouche, partant de l'ACF temporelle, inclut les mêmes biais de calculs que lors d'extractions de mesures. Ce modèle est finalement plus proche de ce qui apparaît et la PSD, calculée sur les données expérimentales acquises, nécessite moins de corrections. C'est pour ces différentes raisons que le modèle PSD développé par L. Azar-Nouche a été utilisé dans nos travaux, par exemple lors du fit des données.

Il est possible, par l'Equation II.21, de générer des signaux qui suivent les distributions exprimées par le modèle PSD [62] (explication complète au Chapitre IV), et d'observer l'impact des différents paramètres sur l'aspect spatial des rugosités. La Figure II.9,a) montre qu'à longueur de corrélation constante, le changement de coefficient de rugosité α entraînera un changement sur les hautes fréquences, soit des variations rapides plus abruptes avec la diminution de ce paramètre. De plus, la Figure II.9,b) permet de voir que l'augmentation isolée du paramètre ξ entraînera un aspect de dilatation du signal et donc un changement plus marqué sur les basses fréquences.

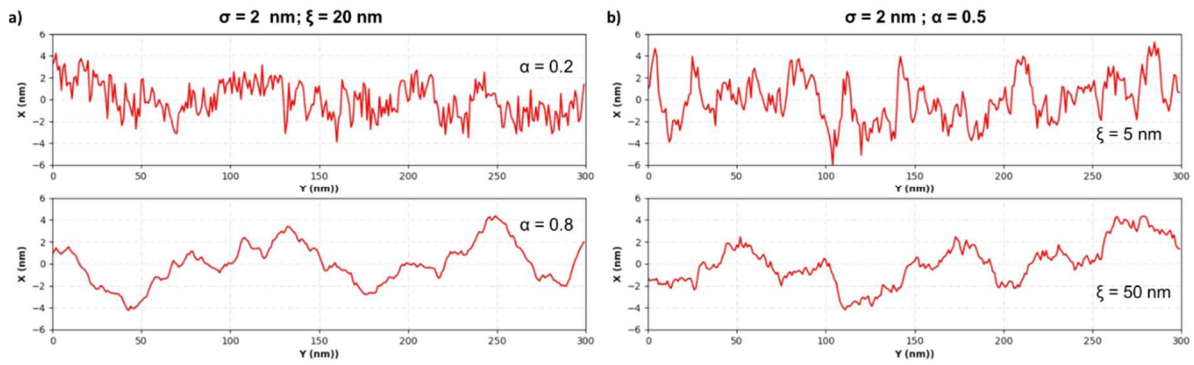


Figure II.9: Exemple de profils de lignes obtenus à partir d'une distribution donnée par l'Equation II.21 , pour les paramètres $\sigma = 2$ nm, a) ξ constant (20 nm) avec $\alpha=0.2$ en haut puis 0.8 en bas ; b) α constant (0.5) avec $\xi = 5$ nm en haut puis $\xi = 50$ nm en bas.

Malgré tous les points positifs cités précédemment, il faut souligner le fait qu'utiliser un modèle induit une simplification de la représentation. Autrement dit, le modèle bien que performant, ne se substituera jamais entièrement à la représentation de la PSD sur la totalité de son domaine. En effet, comme le montre J. Jussot [68], il peut apparaître des densités spectrales de puissance plus complexes avec, par exemple, des pics ou même un modèle ACF « double » (deux longueurs de corrélation, deux coefficients de rugosité) voire plus.

II.3.2 Performance de la PSD

Au modèle de PSD utilisé, Equation II.21, il est possible d'ajouter le terme de bruit de mesure explicité précédemment, tel un bruit blanc, en considérant :

$$\sigma_{mesuré}^2 = \sigma_{réel}^2 + \sigma_{bruit}^2 \quad (II.25)$$

soit le modèle complet considéré :

$$P_n = \frac{\Delta y}{2\pi N} \sigma_{réel}^2 \left[\sum_{m=0}^{N-1} (2 - \delta_m) e^{-\left|\frac{m\Delta y}{\xi}\right|^{2\alpha}} \cos(k_n m \Delta y) (N - m) \right] + \frac{\Delta y}{2\pi} \sigma_{bruit}^2 \quad (II.26)$$

Cette équation permet de prendre en considération le bruit directement dans le fit de la PSD et ainsi, de ressortir le $\sigma_{réel}$ non entaché [62].

Il est important d'observer les limites de ce modèle, à savoir l'influence des paramètres non liés à la rugosité. On en considèrera trois : la longueur de ligne L (à Δy constant), le nombre de lignes mesurées N^* et enfin l'échantillonnage Δy (à L constant). Ces études de performance évaluent la capacité du modèle à obtenir le paramètre de rugosité sans erreur, au travers de la chaîne de création de profils (la génération de profils de ligne est explicitée et validée dans le Chapitre IV) et du fit de ces données.

On souligne que pour l'accélération des traitements, le calcul de la PSD à partir des profils est effectué dans cette partie d'après sa première définition et de manière discrète :

$$P_n = \frac{L}{2\pi} \langle |F_n|^2 \rangle_{N^*} \quad (II.27)$$

F_n étant définie par l'Equation II.17, et N^* étant le nombre de lignes considérées. Comme déjà évoqué, d'après le théorème de Wiener-Khintchine, le résultat est identique.

II.3.2-A Nombre de lignes

Sur la base des recommandations de l'IRDS [11], la première étude a porté sur la modélisation de lignes de 2 μm de long. L'échantillonnage le long de la ligne selon l'axe y a été fixé à $\Delta y = 1 \text{ nm}$, ce qui est petit mais réaliste techniquement (cf. Chapitre III). Enfin, les paramètres de rugosité utilisés sont $\sigma = 2 \text{ nm}$, $\alpha = 0.5$, $\xi = 20 \text{ nm}$ et aucun bruit n'a été intentionnellement ajouté lors de la génération des lignes. On souligne quand même le fait que la génération induit du bruit, à cause de la pixellisation au nanomètre près (cf. Chapitre IV).

Si l'on regarde l'évolution de la PSD, calculée sur un nombre de lignes différentes (Figure II.10 a, b et c, en rouge), on observe que l'utilisation d'un nombre restreint de lignes induit une dispersion importante autour du modèle théorique de la PSD (Figure II.10, en bleu). Ceci implique, comme visible sur la Figure II.10.a, un ajustement des données numériques (en orange) loin du modèle initial (en bleu). Avec l'augmentation du nombre de lignes à 100 (Figure II.10.b), puis à 10000 (Figure II.10.c), on remarque un meilleur accord entre le fit (en orange) et le modèle (en bleu), à une constante près induite par le bruit très visible sur les faibles puissances (à hautes fréquences).

Il n'est pas possible, par cette simple observation, de déterminer le nombre de lignes nécessaires pour obtenir une extraction acceptable des paramètres. Pour aller plus loin, nous avons tracé l'évolution des différents paramètres de rugosité (σ , α , ξ et σ_{bruit}), déduits du fit des données numériques en fonction du nombre de lignes (Figure II.11), pour une gamme comprise entre 50 et 10 000 lignes. Le tracé est ici effectué par pas de 50 lignes traitées.

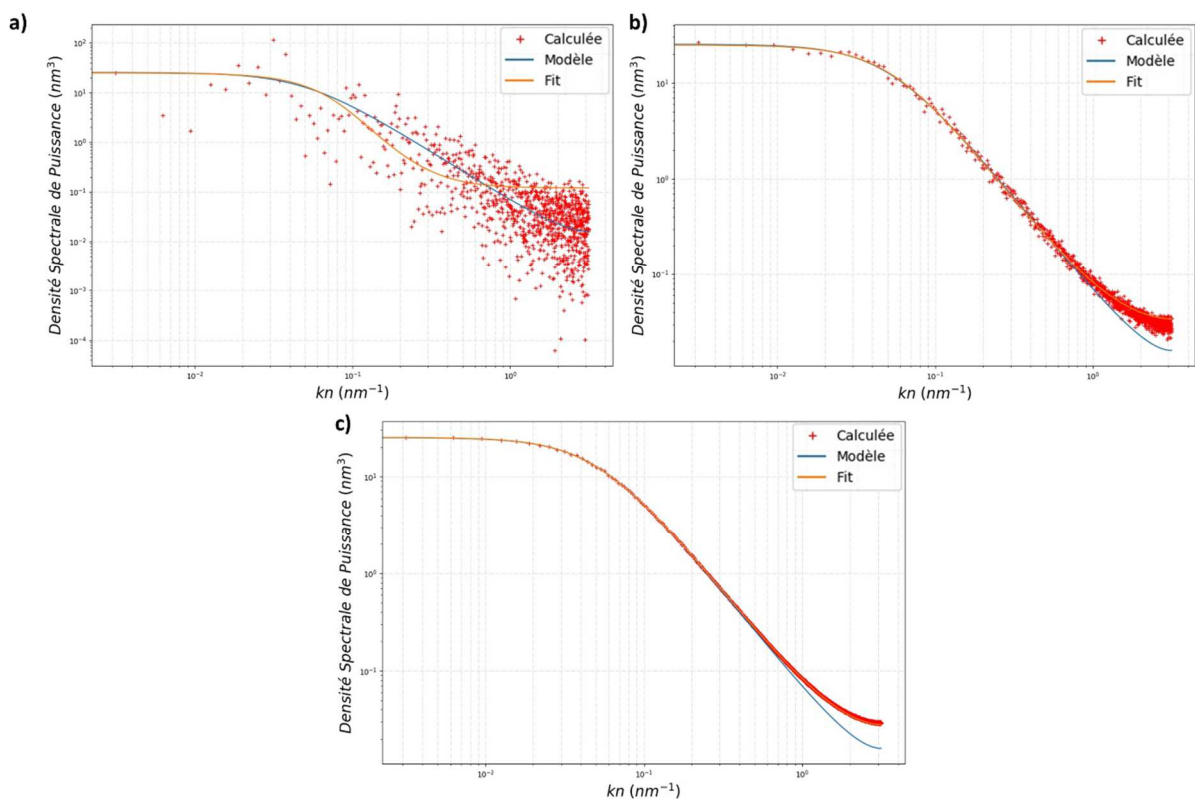


Figure II.10 : Calcul de PSD post génération de profils de lignes, avec le modèle $\sigma = 2 \text{ nm}$, $\alpha = 0.5$ et $\xi = 20 \text{ nm}$, et le fit obtenu pour a) un seul profil, b) 100 profils et c) 10 000 profils.

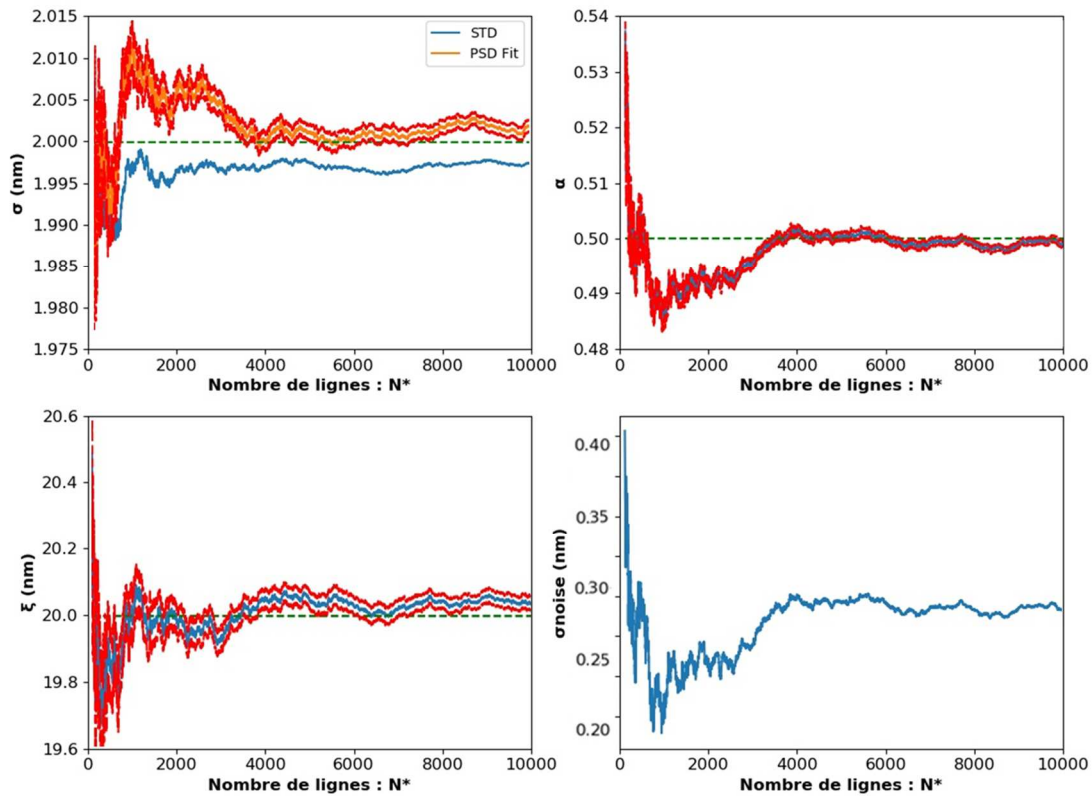


Figure II.11 : Evolution de l'extraction des paramètres de rugosité en fonction du nombre de lignes (50 à 10 000) pour $L = 2 \mu\text{m}$ et $\Delta y = 1 \text{ nm}$ et pour des valeurs nominales de rugosité introduites dans la simulation (en vert) de : $\sigma = 2 \text{ nm}$, $\alpha = 0.5$ et $\xi = 20 \text{ nm}$ (sans bruit). Les incertitudes de fit (à $\pm \sigma$) sont représentées en rouge.
 Courbe σ : STD = calcul direct de l'écart-type sur les profils.

Sur la Figure II.11, on remarque qu'avec plus de 4000 lignes les trois paramètres de rugosité convergent vers leurs valeurs nominales. On souligne cependant le fait que la mesure de 4000 lignes semble irréaliste avec les techniques actuelles de caractérisation (cf. Chapitre III). Le même type d'études, sur la gamme de 1 à 100 lignes résolue à la ligne près, est visible en Figure II.12. On extraira alors qu'au bout de 100 lignes les erreurs sur σ , α et ξ sont respectivement de 0.3%, 3% et 1%, ce qui est déjà très bas. On assumera donc ces niveaux d'erreurs de justesse sur les paramètres de fit, à partir d'un nombre de lignes mesurées supérieur à 100. Des paramétrages de rugosité différents donnent les mêmes résultats dans la limite où ξ reste très inférieur à la longueur de lignes, ceci n'étant pas lié au nombre de lignes.

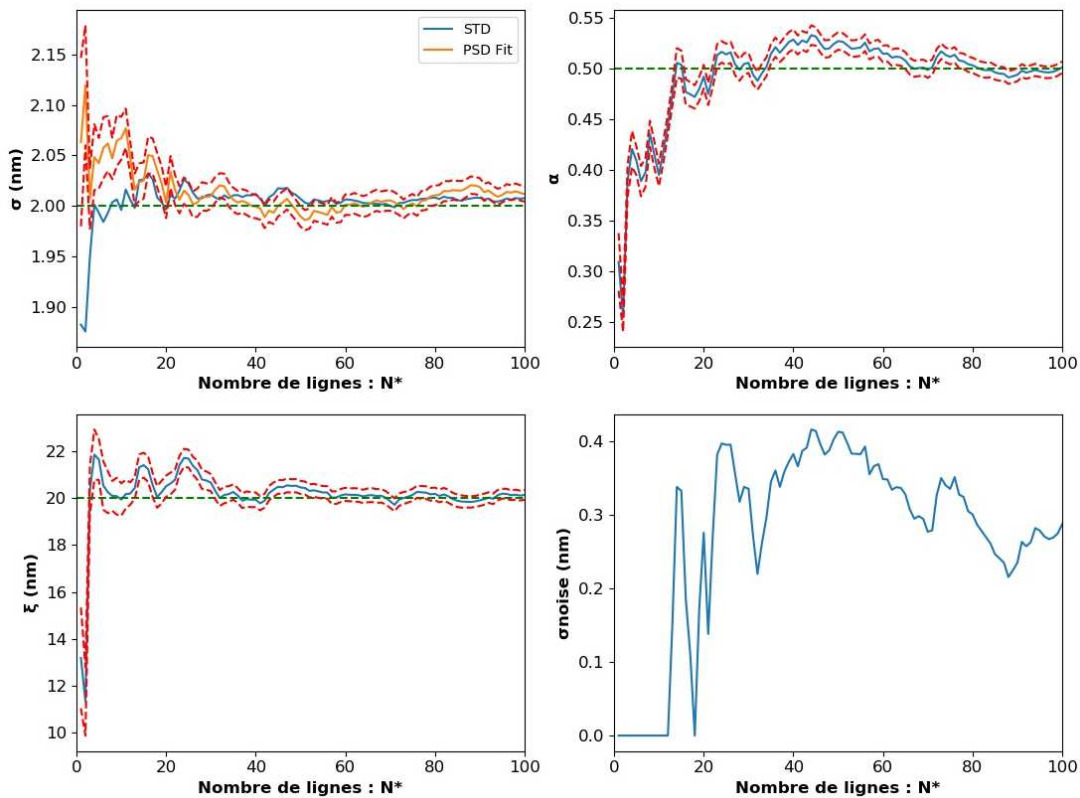


Figure II.12 : Evolution de l'extraction des paramètres de rugosité suivant le nombre de lignes (1 à 100) pour $L=2 \mu\text{m}$ et $\Delta y = 1 \text{ nm}$ et pour des valeurs nominales de rugosité introduites dans la simulation (en vert) de : $\sigma = 2 \text{ nm}$, $\alpha = 0.5$ et $\xi = 20 \text{ nm}$ (sans bruit). Les incertitudes de fit (à $\pm \sigma$) sont représentées en rouge.
 Courbe σ : STD = calcul direct de l'écart-type sur les profils.

II.3.2-B Longueur des lignes à considérer

Pour cette étude de performance liée à la longueur de la ligne, 4000 lignes ont été considérées afin que le résultat ne soit pas impacté par l'erreur citée précédemment. Les paramètres de rugosité considérés sont, dans un premier temps, les mêmes que lors de l'étude sur le nombre de lignes i.e. $\sigma = 2 \text{ nm}$, $\alpha = 0.5$ et $\xi = 20 \text{ nm}$.

On observe tout d'abord (Figure II.13) que la valeur de l'écart-type (STD en bleu), calculée directement sur le profil, a bien la forme attendue reflétant le caractère auto-affine de la rugosité, atteignant le palier équivalent au σ nominal pour des lignes assez grandes (telles que sur la Figure II.7). Il est intéressant de noter (Figure II.13) que le résultat converge rapidement vers la valeur nominale et ce pour tous les paramètres de rugosité, notamment pour α qui a une erreur de moins de 3% même pour une ligne de 50 nm. En regardant plus précisément dans la zone où la longueur de ligne L est inférieure à 200 nm (Figure II.14), il est visible que la valeur de σ se stabilise au-dessus de $L = 75 \text{ nm}$. Cette limite est aux alentours de $L = 175 \text{ nm}$ pour les paramètres ξ et α . Enfin, pour des lignes extrêmement courtes ($< 15 \text{ nm}$) le fit ne fonctionne plus ; il renvoie pour tous les paramètres la valeur 1.

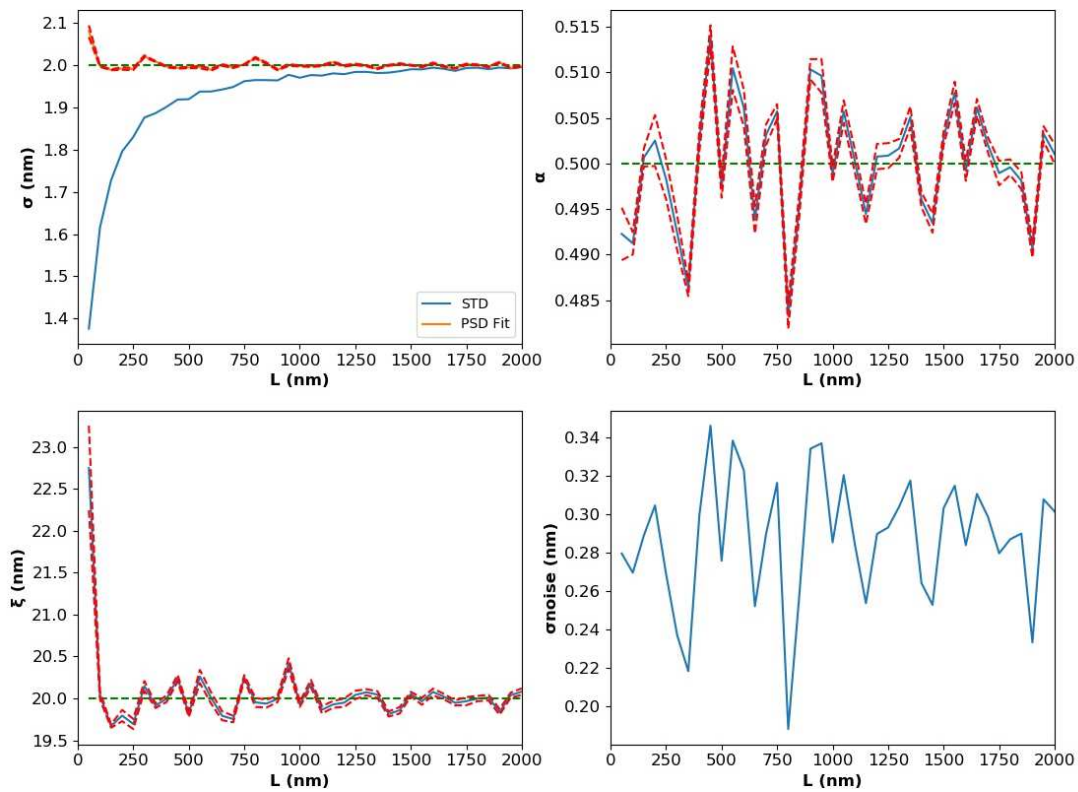


Figure II.13 : Evolution de l'extraction des paramètres de rugosité suivant la longueur des lignes, gamme $L = 50\text{-}2000$ nm par pas de 50 nm à $\Delta y = 1$ nm pour 4000 lignes ; et pour des valeurs nominales de rugosité introduites dans la simulation (en vert) de : $\sigma = 2$ nm, $\alpha = 0.5$ et $\xi = 20$ nm (sans bruit). Les incertitudes de fit (à $\pm \sigma$) sont représentées en rouge.
 Courbe σ : STD = calcul direct de l'écart-type sur les profils.

La Figure II.15 nous permet d'observer que si l'on augmente ξ , dans ce cas égal à 100 nm, la longueur nécessaire afin d'obtenir une stabilisation est elle aussi plus élevée. Ceci est dû au fait qu'en dessous de la valeur de corrélation le plateau de saturation n'est pas atteint, ainsi le fit ne peut donc pas s'effectuer correctement. On note aussi que la valeur de α fluctue encore quelque peu pour des longueurs de lignes proches de 1 000 nm. Par ailleurs, cette limite basse est aussi impactée par la valeur de α , mais à moindre échelle. Une étude [69] montre que pour se prémunir de cette erreur il faut respecter une longueur minimale de L équivalente à $4.\pi.\xi$.

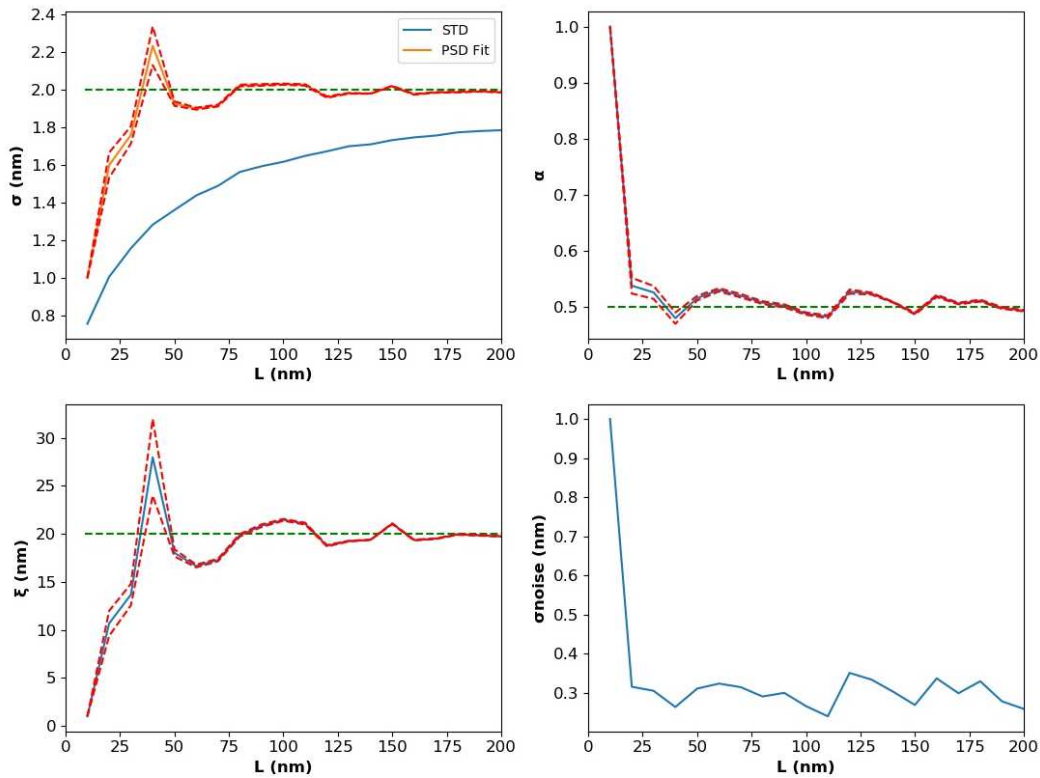


Figure II.14 : Evolution de l'extraction des paramètres de rugosité suivant la longueur des lignes, gamme $L = 10\text{-}200$ nm par pas de 10 nm à $\Delta y = 1$ nm pour 4000 lignes et pour des valeurs nominales de rugosité introduites dans la simulation (en vert) de : $\sigma = 2$ nm, $\alpha = 0.5$ et $\xi = 20$ nm (sans bruit). Les incertitudes de fit (à $\pm \sigma$) sont représentées en rouge.

Courbe σ : STD = calcul direct de l'écart-type sur les profils.

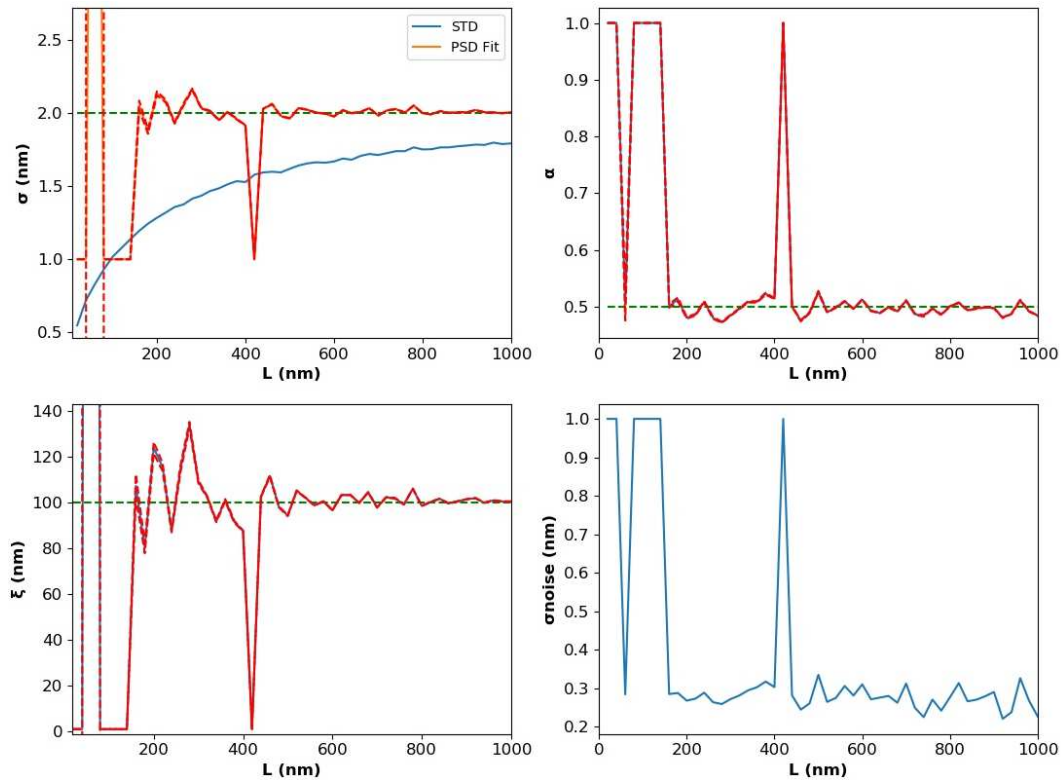


Figure II.15 : Evolution de l'extraction des paramètres de rugosité suivant la longueur des lignes gamme $L = 20\text{-}1000$ nm pas de 20 nm à $\Delta y = 1$ nm pour 4000 lignes ; et pour des valeurs nominales de rugosité introduites dans la simulation (en vert) de : $\sigma = 2$ nm, $\alpha = 0.5$ et $\xi = 100$ nm (sans bruit). Les incertitudes de fit (à $\pm \sigma$) sont représentées en rouge.

Courbe σ : STD = calcul direct de l'écart-type sur les profils.

II.3.2-C Influence de l'échantillonnage

Dans l'étude précédente sur la longueur de la ligne, c'est le nombre de mesures sur une ligne N qui a été modifié à pas d'échantillonnage Δy constant. Il a été observé un impact significatif en dessous d'une certaine longueur. Cependant les paramètres L , N et Δy sont reliés, il est donc nécessaire désormais d'étudier l'influence de Δy . Pour cette étude, la longueur de ligne L est constante et assez grande ($2 \mu\text{m}$) pour ne pas reproduire l'erreur vue précédemment.

Sur la base du modèle considéré et à partir de l'Equation II.21, il est possible de générer des PSD avec différents échantillonnages Δy . On peut ainsi voir, comme sur la Figure II.16 avec un paramétrage σ , α , ξ constant, que l'effet de ce paramètre est directement visible sur la PSD.

Premièrement comme cité précédemment, Δy qui correspond à l'échantillonnage dans l'espace spatial détermine la limite d'analyse en haute fréquence. En effet, au vu du théorème de Shannon concernant l'échantillonnage : « l'échantillonnage d'un signal exige un nombre d'échantillons par unité de temps supérieur au double de l'écart entre les fréquences minimale et maximale qu'il contient » appliqué dans notre cas, une limite d'analyse en haute fréquence et donc un nombre d'onde maximum atteignable, k_{max} apparaît. Cette limite correspond à $n=N/2$, ce qui appliqué à l'Equation II.18 donne un $k_{\text{max}} = \pi / \Delta y$, ceci étant vrai quel que soit N pour les hautes fréquences.

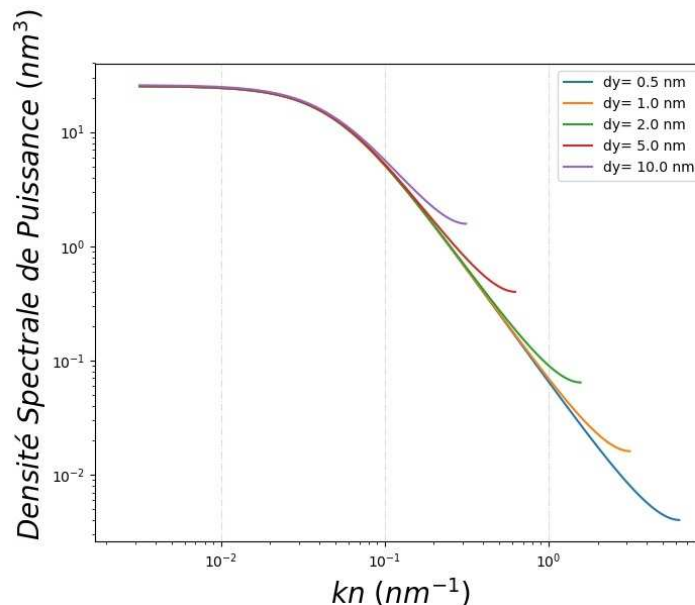


Figure II.16 : Modèle PSD généré pour $\sigma = 2 \text{ nm}$, $\alpha = 0.5$ et $\xi = 20 \text{ nm}$ avec différents Δy .

Le changement de Δy ne fait pas que tronquer la PSD au nombre d'onde donné par k_{max} . Ce paramètre impacte la PSD sur une plus large gamme de fréquences : pour exemple d'après la Figure II.16, il semble impossible d'analyser correctement la PSD dans la gamme $kn = 0,5$ à $1,5 \text{ nm}^{-1}$ avec une valeur de $\Delta y = 2 \text{ nm}$ (représentée en vert). On observe la création, dans des hautes fréquences, d'un « pied de PSD » sur une largeur d'au moins $k_{\text{max}}/3$, soit $0,52 \text{ nm}^{-1}$ dans l'exemple précédent. Cependant il faut souligner que cet impact, visible sur le modèle, peut être inclus dans le calcul sur les profils de lignes. Pour cela il est nécessaire que le paramètre Δy dans le modèle, lors de l'ajustement aux données, soit identique au paramètre d'échantillonnage des mesures.

On retiendra qu'il est nécessaire d'utiliser un Δy le plus faible possible pour obtenir une gamme de fréquences analysable plus grande, ce qui est aujourd'hui limité par les techniques nous le verrons Chapitre III.

II.4 Résumé du chapitre

Au cours de ce chapitre nous avons pu voir les concepts métrologiques de justesse et de fidélité, et ainsi comprendre l'intérêt d'effectuer plus d'une mesure pour obtenir une information qualitative sur les métriques obtenues.

Par suite nous avons explicité les métriques considérées ici en exprimant les différents types de rugosité de ligne, à savoir la LWR et la LER. Cela nous a menés à la notion d'écart-type et à montrer que cette notion n'est pas assez complète pour décrire correctement la rugosité de ligne.

Nous avons donc vu qu'il existe de multiples modèles, basés sur la notion de fractale, afin de décrire de manière plus exacte la rugosité de ligne. Les modèles les plus généraux : HHCF, ACF et PSD ont été explicités. Ainsi nous avons pu expliquer, en calculant l'erreur résiduelle des différents modèles notamment sur un nombre de lignes réduit, le choix dans le cadre de ces travaux de l'utilisation du modèle de PSD développé par L. Azar-Nouche.

Enfin les performances de ce modèle, en termes de justesse, ont été testées. L'impact du nombre de lignes mesurées et de la longueur de lignes sur l'extraction correcte des paramètres du modèle, a été évalué. Désormais nous savons qu'il est possible d'extraire la rugosité de ligne sur un minimum de 100 lignes et connaissons l'erreur associée, due à la fois à la statistique mais aussi à la chaîne de calculs et au modèle utilisé. Un point sur l'effet de l'échantillonnage a finalement été fait. Nous avons alors vu son impact ainsi que celui de différents paramètres, sur notre extraction, et conclu qu'il est nécessaire de diminuer cette distance d'échantillonnage le plus possible.

Grâce à cela nous allons pouvoir étudier les différentes techniques et équipements disponibles, pour les mettre en regard de ces différents points. Nous verrons donc si chacun d'entre eux est capable d'extraire la rugosité suivant notre modèle en fréquence, mais aussi s'ils respectent les paramètres fixés à savoir : le nombre de lignes nécessaires, leur longueur et l'échantillonnage de la mesure.

Chapitre III - Les techniques de caractérisation

Ce chapitre est consacré à la description des quatre techniques de caractérisation utilisées lors de ces travaux, elles sont détaillées au travers de leur principe de fonctionnement général. Un point particulier est fait sur les équipements accessibles pour ces recherches et leurs spécificités par rapport aux moyens mis en œuvre. Ensuite, l'exploitation des mesures est détaillée pour chacune des techniques, dans un cas général, puis appliquée au cas spécifique de la rugosité de ligne. Les limitations des techniques et des équipements sont décrites, ainsi que les progrès faits pour tenter de les repousser.

Dans une dernière partie, les capacités ces techniques et des équipements sont comparées suivant plusieurs axes : leurs capacités fréquentielles, de résolution en amplitude ou de sensibilité, mais aussi de bruits qui les impactent ou encore de leur maturité industrielle. Cette comparaison est, entre autres, effectuée grâce à la mesure d'un set d'échantillons compatibles avec toutes les techniques.

III.1 Microscopie électronique à balayage

La Microscopie Electronique à Balayage (MEB) est une des techniques de métrologie les plus utilisées en microélectronique. Elle est également connue sous le nom de SEM pour « Scanning Electron Microscopy » en anglais. Cette technique permet d'obtenir une image dite directe de l'objet considéré, avec un temps d'acquisition inférieur à la minute.

La microscopie électronique à balayage repose sur la collecte d'électrons issus de l'interaction entre un faisceau d'électrons incidents et la matière.

III.1.1 Interaction électrons-matière

Il existe divers types d'interactions entre les électrons et la matière, comme le montre la Figure III.1. Cette figure met en évidence le fait que ces interactions se produisent à différentes profondeurs du matériau et correspondent à des énergies différentes. A chacune des interactions est associée une poire d'interaction qui peut varier avec le matériau considéré et l'énergie incidente des électrons. Toutes les interactions représentées peuvent être classées dans deux catégories : élastiques ou inélastiques.

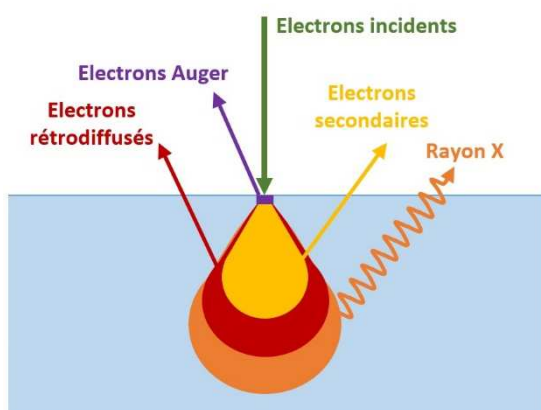


Figure III.1 : Représentation des différentes interactions électrons-matière sous forme de poire d'interaction.

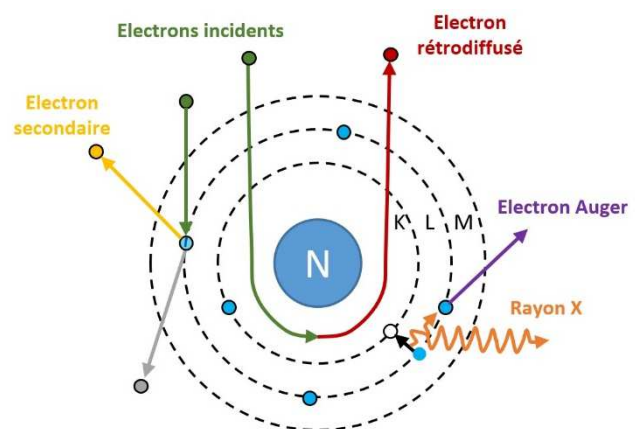


Figure III.2 : Différentes interactions électrons-matière à l'échelle d'un atome (atome de carbone).

Dans le cas des interactions élastiques, les électrons ne perdent pas d'énergie lors de leur interaction avec la matière, on trouve dans cette catégorie les électrons rétrodiffusés. Comme le montre la Figure III.2 (en rouge), ces électrons correspondent à des électrons incidents qui sont déviés par le noyau des atomes, suivant le numéro atomique du matériau considéré, les électrons vont être plus ou moins rétrodiffusés. Les électrons qui ne subissent que des interactions élastiques ressortent avec quasiment la même énergie que celle du faisceau incident E_0 , comme cela est représenté en Figure III.3, avec un pic de détection du nombre d'électrons à cette énergie-là. Ces électrons peuvent être décélérés par un changement de direction, ou encore d'autres interactions.

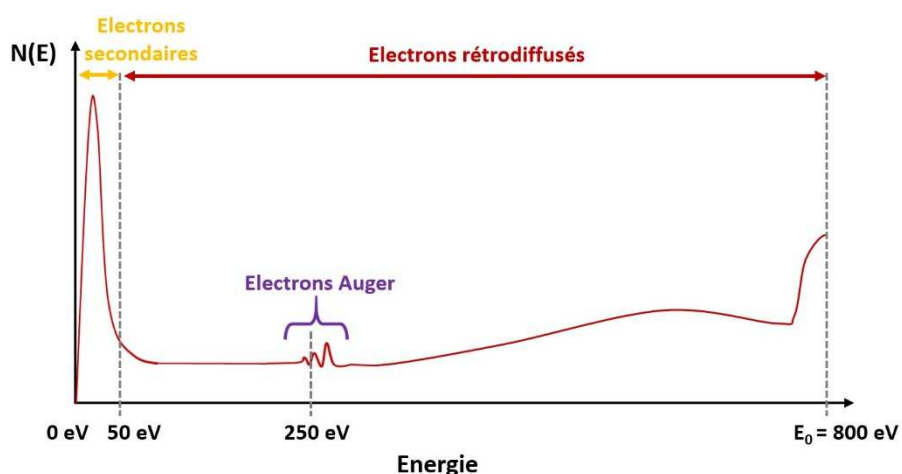


Figure III.3 : Nombre d'électrons récoltés en fonction de leur énergie, pour une tension d'accélération incidente de 800 eV sur du carbone.

Les interactions qui induisent une perte d'énergie sont les interactions dites inélastiques. Trois grandes catégories d'émissions de ce type existent et sont représentées Figure III.1 et Figure III.2 :

- **Les électrons secondaires** (en jaune) correspondent à l'expulsion, hors du nuage électronique, d'électrons contenus dans la matière lors de l'impact avec des électrons incidents. Ces électrons ont une énergie de sortie inférieure à 50 eV (limite arbitraire) ; par conséquent seuls les électrons secondaires proches de la surface ressortent de la matière et sont détectés. Leur quantité est importante, comme le montre-la Figure III.3, ce sont eux qui sont généralement utilisés pour reconstituer une image.
- **Les électrons Auger** (en violet) sont dépendants du matériau bombardé. Comme le montre la Figure III.2, ces électrons sont émis lors de la désexcitation d'autres électrons apportant l'énergie nécessaire pour qu'ils puissent se libérer de l'interaction du noyau. Ces électrons ont donc l'énergie d'un changement de niveau électronique, pour exemple dans le cas de l'atome de carbone, il existe trois changements de niveau d'énergie aux alentours de 250 eV, créant ainsi trois pics d'émission Auger comme visible sur la Figure III.3
- **Les rayons-X** (en orange) sont des photons dont la poire d'interaction est plus en profondeur dans le matériau (Figure III.1). Il s'agit là aussi, principalement, de la désexcitation d'un électron qui émet un rayonnement X non capté par un autre électron. Il existe aussi des rayons-X dits « de freinage » qui sont, eux, émis lors du changement de direction d'électrons incidents.

Malgré des cas d'émissions différenciés, on notera que les électrons peuvent subir diverses interactions successives, ainsi un électron secondaire peut être émis sous l'impact d'un électron rétrodiffusé. De même, un électron capté comme rétrodiffusé a pu subir une perte d'énergie due à des interactions inélastiques, on voit par exemple sur la Figure III.2, en gris, qu'après interaction l'électron incident continue sa course.

III.1.2 Le principe de fonctionnement

Les outils de microscopie électronique à balayage ont un fonctionnement de base similaire à la lithographie électronique (voir Chapitre I). Comme le montre la Figure III.4, un faisceau d'électrons produit par un canon à électrons est mis en forme grâce à des jeux de lentilles. Ce faisceau peut être focalisé sur la surface de l'échantillon et contrôlé en direction, de manière à balayer cette surface. Au cours de ce balayage, des électrons (principalement secondaires) sont réémis. Ces derniers sont détectés en quantité par un détecteur Everhart-Thornley fonctionnant sur le principe d'un scintillateur qui, suite à l'impact d'électrons, émet de la lumière elle-même amplifiée et détectée par une photocathode. Il existe d'autres détecteurs moins courants dans ce type d'équipement car ils ne sont pas adaptés aux fortes énergies [70]. C'est la corrélation entre la position du faisceau et l'intensité du faisceau d'électrons secondaires réémis qui permet de reproduire une image (Figure III.5).

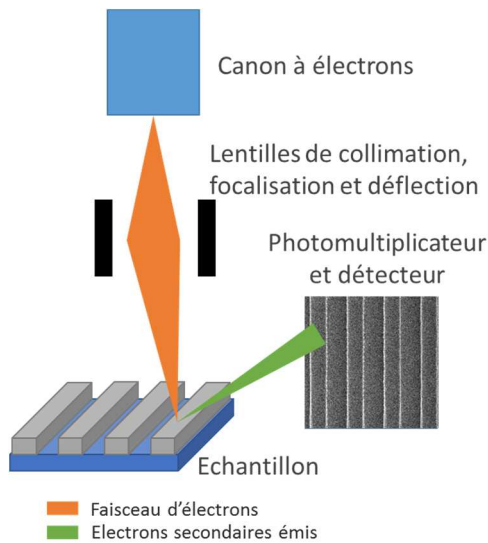


Figure III.4 : Schéma de principe du fonctionnement d'un microscope électronique à balayage.

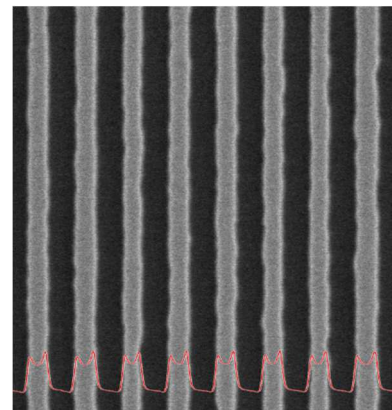


Figure III.5 : Exemple d'image obtenue avec un microscope électronique à balayage en vue du dessus. En rouge est superposé le signal moyen sur la longueur des lignes.

L'image obtenue est en niveaux de gris (codée sur un seul canal, en opposition au système RGB de l'anglais « red, green, blue » où elle serait codée sur 3 canaux, un par couleur). Le niveau de gris est principalement le reflet de la morphologie de l'objet sondé, car la majorité des électrons correspondant à l'image sont des électrons secondaires et donc surfaciques. Il existe également des méthodes de filtrage d'énergie permettant de récolter seulement les électrons rétrodiffusés ; on parle alors de mode BSE (pour BackScattered Electrons), qui donne ainsi une information sur le type de matériau analysé.

III.1.3 Les équipements de microscopie électronique

Il existe différents types de microscopes électroniques. Tout d'abord, les microscopes « top-view » qui permettent d'obtenir une image de l'objet en vue de dessus, comme illustré en Figure III.5 pour un réseau de lignes. Ils peuvent être utilisés pour une analyse qualitative, et alors ne fournir qu'une image de l'objet servant par exemple à faire de la revue de défauts, mais ils peuvent aussi être plus développés et intégrer un système de mesure. Dans ce cas, l'équipement doit être calibré afin de pouvoir fournir une métrique de dimension, on parle alors de CD-SEM (Critical Dimension Scanning Electron Microscope). C'est ce type d'équipement qui a été utilisé dans le cadre de nos mesures de rugosité de ligne.

Il existe enfin des microscopes électroniques « cross-section » qui permettent d'obtenir une vue en coupe de l'objet, comme illustré en Figure III.6, ce type de microscope nécessite une préparation en coupe transverse de l'objet et de son support. Cette méthode est destructive, cependant elle permet d'obtenir des informations non accessibles en vue de dessus telles que la hauteur (Voir Chapitre I). Ce type d'équipement a été utilisé au cours de cette thèse pour valider le profil des lignes mesurées, information nécessaire pour d'autres techniques comme la scatterométrie.

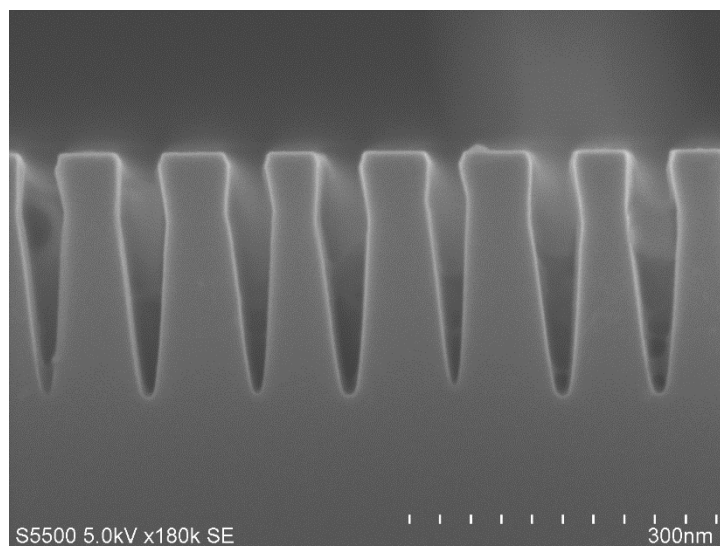


Figure III.6 : Exemple d'image obtenue avec un microscope électronique à balayage dit "cross-section", sur un réseau de lignes gravées dans le silicium.

Dans le cas de « cross-section », l'information est obtenue à une position fixe le long des lignes et ne permet donc pas l'extraction de la rugosité de ligne. De plus, l'information extraite est locale car ce type d'équipement n'est pas automatisé pour la prise d'images, contrairement au CD-SEM, ce dernier permet de balayer des surfaces de près d'une dizaine de μm^2 par zone d'analyse et de réaliser, sans intervention humaine, plusieurs centaines de captures.

Au cours des travaux présentés ici, deux microscopes électroniques « cross-section » ont été utilisés : Hitachi 5000 et Hitachi 5500, ils ont des caractéristiques similaires. Le plus récent des deux (Hitachi 5500) permet un refroidissement de l'échantillon, ce qui limite la destruction des matériaux ou les dépôts carbonés sous l'effet du faisceau d'électrons. La tension d'accélération des électrons, qui est une caractéristique importante des microscopes électroniques à balayage, est de l'ordre de quelques milliers à quelques dizaines de milliers d'électronvolts pour les équipements cités, suivant l'application.

Le CD-SEM utilisé pour la mesure de rugosité de ligne est le Verity4i+ de la compagnie Applied Materials. Cet outil a servi pour une prise d'image calibrée, les mesures ayant été effectuées *a posteriori* sur divers logiciels (voir partie suivante), et non directement sur le module de mesure de l'équipement. Ce CD-SEM, contrairement à ceux de la marque Hitachi utilisés dans de nombreuses publications liées à la rugosité de ligne [66, 71], ne fonctionne pas sur le principe de grossissement (tel un microscope optique). En effet, les paramètres d'entrée du Verity4i+ sont : la taille de l'image voulue en nanomètres ou en micromètres, et la taille des pixels (soit l'échantillonnage) en nanomètres. L'impact de tels paramètres sur la mesure de rugosité de ligne a été explicité dans le Chapitre II. On retiendra le fort avantage du Verity4i+ qui est de ne pas corrélérer la taille de l'image prise à la résolution, autrement dit, de différencier la longueur de mesure par rapport à l'échantillonnage. Cependant, la résolution maximale (capacité à distinguer deux objets adjacents), dans les conditions de qualification de l'équipement faite sur des particules d'or, ne descend pas en dessous de 1.4 nm à cause de, la taille du faisceau d'électrons utilisé et de la largeur de la poire d'interaction des électrons secondaires. Cet équipement fonctionne sur une gamme d'énergie d'accélération de quelques centaines d'électronvolts, ce qui est bien plus faible que sur les équipements de cross-section.

III.1.4 Extraction de la rugosité en CD-SEM

Pour extraire la mesure de rugosité d'une image SEM en vue de dessus, le principe est de déterminer le contour de l'objet, comme cela est visible en Figure III.7 (en rouge). Sur une image, il peut y avoir différents objets caractérisables, ligne ou espace entre deux lignes, il est donc important de savoir si le signal le plus clair correspond à l'un ou l'autre afin de mesurer le bon motif. Une méthode simple est de visualiser, dans un premier temps, un bord du réseau de lignes permettant ainsi de distinguer l'intensité obtenue sans motif.

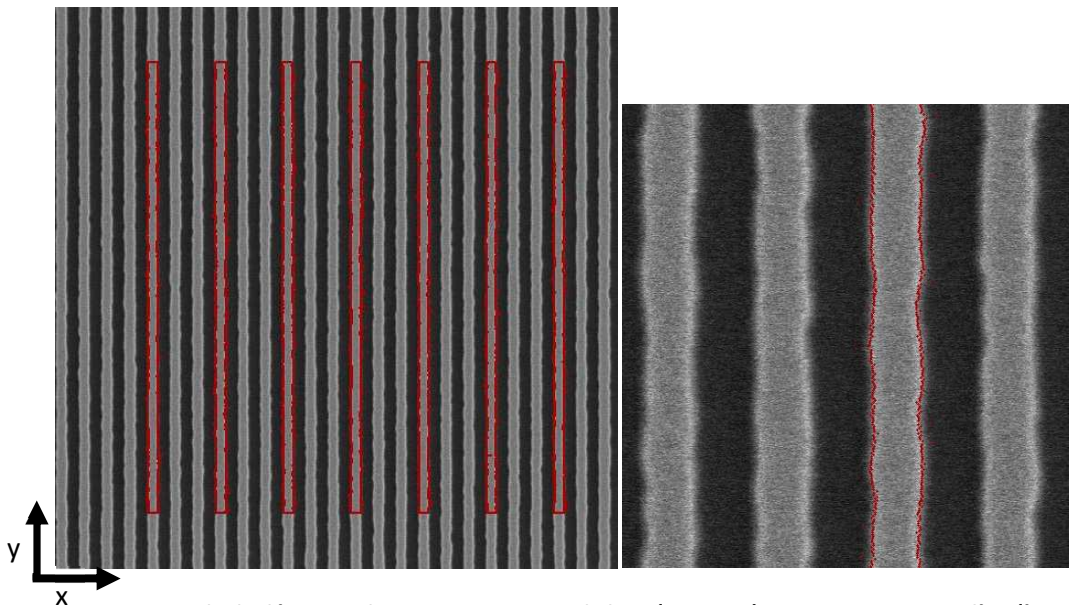


Figure III.7 : Exemple de détection de contours sur 2 µm de long (en rouge), sur une image complète (à gauche) et sur un zoom (à droite), réalisé dans les deux cas sur un réseau de lignes de silicium (en blanc).

Pour effectuer la détection des bords, il existe divers algorithmes, qui peuvent être considérés suivant deux familles. Premièrement, les algorithmes à détection ponctuelle détectent et positionnent un point de mesure unique sur le signal, en niveaux de gris. Ainsi sur la Figure III.8.a on observe, sur un signal moyenné, la détection : « 1 » du maximum d'intensité, « 2 » de 80% de maximum du pic appelé

mode « threshold » (« palier » en français), et « 3 » la pente maximale (égale au maximum de la dérivée). Bien que la position ne soit pas la même, ces différents algorithmes peuvent être utilisés car ce sont les variations de cette position le long de la ligne (dans le cas LER), et non la valeur nominale de chacune des positions, qui vont être étudiées (idem dans le cas LWR avec le CD). Cependant comme le montre la Figure III.8.b, dans le cas d'un signal non moyenné, ces paramètres bien qu'aisément positionnables sont très sensibles au bruit de mesure. Dans notre exemple, on retrouve un positionnement de l'extraction « 2 » et « 3 » au même endroit, et l'extraction « 1 » ne semble pas centrée par rapport au signal.

Pour améliorer ce type d'extraction il est possible d'utiliser un filtre, par exemple une moyenne sur quelques lignes, c'est notamment ce qui est fait pour la mesure du CD ou du pitch. Cependant, cette méthodologie impacte fortement la PSD finale pour l'analyse de rugosité. En effet, l'utilisation d'un filtre suivant l'axe y est équivalente à une augmentation de la distance d'échantillonnage Δy (voir Chapitre II), ce qui impacte l'intensité de la PSD et fausse donc les extractions de rugosité [65, 72, 73].

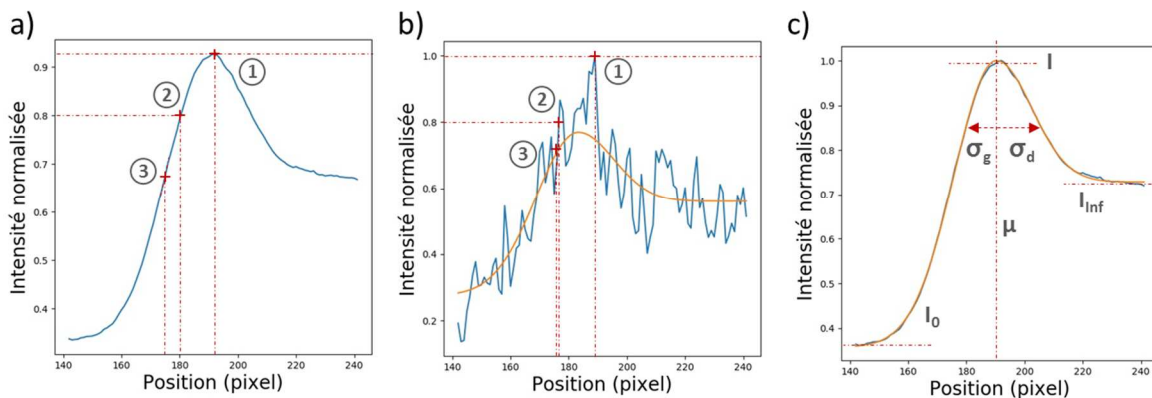


Figure III.8 : Exemple d'extraction d'un bord (en bleu) sur le profil d'une demi-ligne : a) et b) moyenné sur toute la hauteur de la ligne, b) en une position sur la ligne. "1", "2" et "3" représentent respectivement l'extraction pour le maximum du signal, 80% du maximum et la pente maximale. En c) sont explicités les paramètres du modèle de T. Verduin, la convergence du modèle (en orange) sur le profil (en bleu).

Un second type d'extraction sans filtrage existe, il permet de ne pas considérer un seul point dans l'extraction mais un profil complet, choisi par un modèle. Pour ce type d'extraction il existe diverses approches [74, 75] et nous avons utilisé celle développée par T. Verduin [74], au vu des résultats présentés par ce dernier sur la robustesse obtenue et de par la maturité cette méthode. Cet algorithme correspond à une segmentation du profil en quatre parties, à savoir un premier plateau, une demi-gaussienne, une seconde demi-gaussienne et un deuxième plateau. Ce modèle comporte donc plusieurs paramètres, représentés sur la Figure III.8.c, qui sont les niveaux des plateaux (I_0 et I_{inf}), l'intensité des gaussiennes (I), leur centrage (μ) ainsi que leur demi-largeur respective (σ_g et σ_d).

Le principe de cet algorithme est, premièrement d'ajuster le modèle sur le signal moyen selon l'axe y, résultat visible sur la Figure III.8.c, afin de déterminer une première fois l'ensemble des paramètres. Dans un second temps, chaque position y sur la ligne sera approchée, comme visible pour une coupe (Figure III.8.b), par la modification du centrage (μ) et un facteur d'échelle. Ainsi, à partir de ce modèle il sera possible d'extraire (μ), ou n'importe quel autre point d'intérêt du profil, pour chaque position le long de la ligne. La prise en compte d'une multitude de points pour détecter un contour permet d'être beaucoup moins sensible au bruit de mesure, sans pour autant utiliser un filtrage qui dégrade le résultat.

Il est important de mentionner qu'aujourd'hui, ce second type d'extraction n'est pas encore disponible dans les logiciels des fournisseurs d'équipements, ni même dans les logiciels commerciaux tels DGmage (utilisé en partie dans le cadre de ces travaux). Cette méthodologie d'extraction a donc été mise en place en parallèle de ce logiciel. Dans le manuscrit sera mentionné, pour chacune des parties (ou chaque résultat), la méthode employée sous les termes « Verduin » pour l'extraction paramétrée, ou « Threshold » dans le cas de l'utilisation de DGmage.

Quelle que soit la méthode, une fois la détection des bords effectuée, il est possible sur ce type d'image en vue de dessus d'extraire le CD, le pitch, et d'effectuer le calcul de la PSD comme explicité dans le Chapitre II, soit pour une extraction LER, soit pour une extraction LWR. Un exemple de calcul de LWR, avec des images SEM traitées par une extraction Verduin, est donné en Figure III.9. Outre l'effet statistique du nombre de lignes, déjà exploré dans la présentation du modèle PSD, on peut remarquer que sur cette extraction, contrairement au modèle PSD présenté dans le Chapitre II, un large plateau est présent au niveau des hautes fréquences ($> 0.3 \text{ nm}^{-1}$), Ceci correspond au bruit de mesure nous amenant donc aux limites de cette technique.

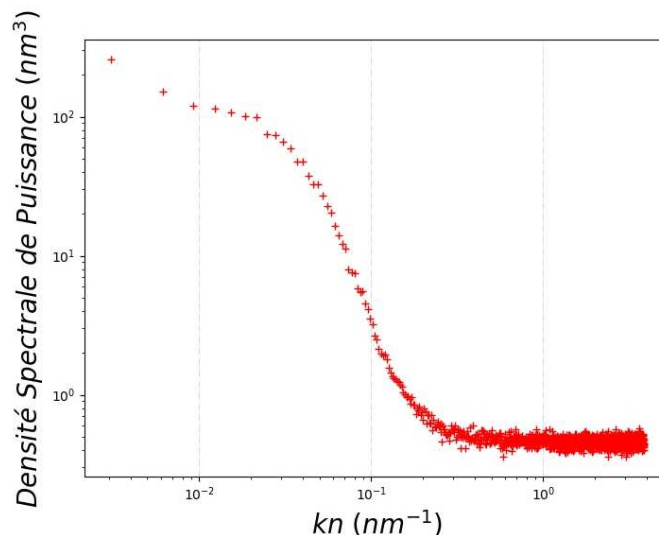


Figure III.9 : PSD LWR d'un échantillon NIL, extraite sur 209 lignes par la méthodologie Verduin;

III.1.5 Limitations

Il existe de multiples limites à la technique MEB, et dans cette partie, seules les limites qui impactent l'extraction de la rugosité de ligne seront explicitées. Dans la majorité des cas, ces limitations n'empêchent pas la détection de bords, grâce notamment aux méthodes évoluées présentées précédemment, mais elles déforment le signal et se répercutent donc sur la PSD.

- **Le bruit type grenaille** est un bruit de mesure aléatoire, il provient du fait que l'image est formée par un nombre fini d'électrons ayant une charge élémentaire (statistique de comptage). Comme le montre L. Azar-Nouche [62], ce type de phénomène induit sur la PSD un bruit blanc, ce dernier a la particularité d'être de même puissance pour toutes les fréquences. Ainsi cette limitation, comme présenté sur la Figure III.9, correspond à l'ajout d'une constante qui apparaît plus visible sur les hautes fréquences où l'intensité de rugosité est la plus faible. Il est possible, soit de prendre en compte ce bruit en l'insérant comme dans le modèle PSD utilisé pour ces travaux (Equation II.26), ou de le supprimer directement sur les données PSD brutes. Malgré le fait que ceci améliore l'extraction au niveau du « pied arrondi » de la PSD,

comme discernable en Figure III.9 entre 0.1 et 0.3 nm⁻¹, ces deux types de corrections possibles ne permettent pas de récupérer correctement l'information potentiellement présente dans le plateau des hautes fréquences. C'est la principale limitation de cette technique, elle induit dans cet exemple une perte de toute l'information en dessous de 10 nm de période. Deux méthodes permettent de réduire ce bruit aléatoire : (i) la prise de plusieurs images qui sont ensuite sommées et, (ii) une déflexion du faisceau plus lente, ce qui revient dans les deux cas à un temps d'acquisition plus long pour détecter plus d'électrons ; mais l'effet positif de ces solutions est limité par le point présenté dans le paragraphe suivant.

- **L'impact destructif du faisceau** est un point problématique pour la mesure de rugosité de ligne. Comme le montrent diverses études [32, 76, 77], sous l'effet du faisceau certains matériaux tels que les résines de lithographie réagissent même après développement. Cet effet, dit de « shrinkage », correspond à la fois à une diminution non contrôlée des dimensions de l'objet mais aussi à son changement de densité. Ces études indiquent que ce phénomène est très important, même pour des énergies déposées faibles et ce avec un impact fort, mais qu'au-delà d'une certaine dose d'énergie apportée, le phénomène se stabilise et le motif alors déformé ne change plus. Ce changement de dimension n'est pas homogène dans tout l'objet, ainsi la rugosité est réduite par cet impact du faisceau [62, 68, 78]. On retrouve ici la notion de métrologie destructive mais à une échelle bien plus fine, l'impact sur la mesure LER montant jusqu'à plus de 15% [72].

Pour ce type de matériaux, il est donc nécessaire d'apporter le minimum d'énergie lors de la prise d'image, ce qui va à l'encontre des corrections du bruit aléatoire. Une autre méthode vise à corriger, *a posteriori*, l'effet de ce « shrinkage » sur la mesure [79]. Cependant, dans notre cas d'étude cet effet ne sera pas nécessairement stable suivant la variation des paramètres de rugosité (σ , α et ξ voir Chapitre II). Ainsi, une telle correction nécessiterait, en plus de la mesure de différentes tailles de lignes (CD et hauteur) pour un matériau donné, la caractérisation ces effets en fonction des paramètres de rugosité.

- **Le chargement électrique de l'objet** entraîne le déplacement des électrons sous l'effet de la répulsion des charges. Ceci induit, au moment de l'acquisition, un déplacement de l'image obtenue lors de la sommation des différents balayages, introduisant un « flou ». Il devient donc compliqué sur ce type de cliché, de détecter correctement la position d'un bord, ce qui est très impactant dans le cadre de mesure de rugosité, mais ceci peut être pris en compte. En effet, pour éviter le chargement électrique de l'objet, on peut réduire le nombre d'images moyennées et accélérer le faisceau, mais cela augmente le bruit aléatoire. Pour ne pas choisir entre l'une ou l'autre des limitations, il est possible de jouer sur la stratégie de déplacement du faisceau. Comme le montre la Figure III.10, dans le cas d'un balayage « standard » (a), le faisceau parcourt l'échantillon ligne par ligne. Le balayage étant plus fin que la taille du faisceau lors du passage à la ligne inférieure, la ligne chargée électriquement et de manière temporaire (temps d'évacuation des charges) lors du premier passage, influence la trajectoire des électrons lors du second passage. L'idée est donc de ne pas passer directement sur la seconde ligne, mais de faire un entrelacement. Par exemple, comme illustré sur la Figure III.10.b une ligne sur trois est balayée puis, lors d'un second et d'un troisième passages, les lignes intermédiaires sont complétées, donnant du temps aux charges induites par le

faisceau pour être évacuées (ce temps de décharge est dépendant des matériaux considérés). Cependant, ce balayage induit aussi un artefact d'ombrage sur la prise d'image, comme décrit dans le point suivant.

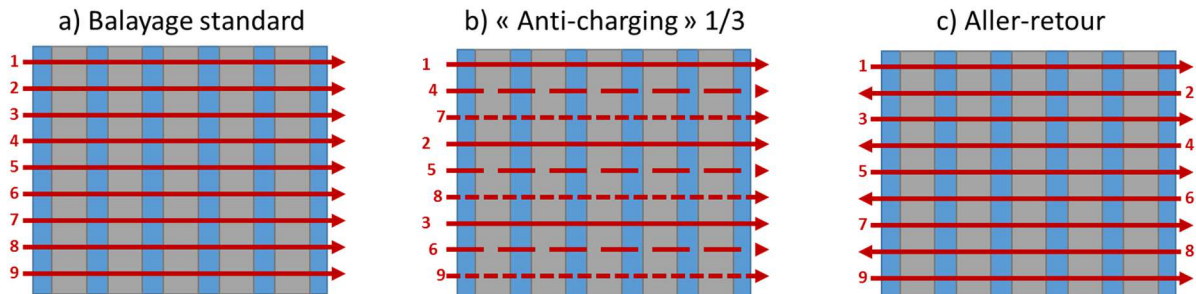


Figure III.10 : Stratégie de balayage de l'échantillon (ligne en gris) par le faisceau, a) standard, b) pour limiter l'effet du chargement électrique du faisceau, c) pour limiter les artefacts d'ombrage.

- **L'ombrage** intervient lors d'un balayage unidirectionnel, comme dans le cas standard de la Figure III.10.a. En effet, il a été remarqué que le passage d'une ligne à un espace (bord droit dans cet exemple) est toujours moins bien défini. Il est supposé que cela provienne soit de la dynamique de déplacement du faisceau ainsi que du temps de sortie des électrons secondaires, soit de la charge locale des motifs. Ce défaut induit, comme dans le cas de chargement électrique, un flou sur le bord. Pour pallier à ce problème, il est possible d'utiliser un mode « aller-retour » illustré en Figure III.10.c, l'image ainsi acquise, dans une direction puis dans une autre, permet une meilleure définition du bord droit. On note que cette option n'est pas disponible sur le CD-SEM utilisé lors de ces travaux, mais seulement sur les versions ultérieures.
- **La taille du faisceau** est, comme il a été dit dans la partie précédente, très limitante. En effet, la taille du faisceau fixe la fréquence maximale analysable à partir du moment où le pixel est plus petit que celle-ci. Plus précisément, le faisceau ayant une forme gaussienne, il vient à la fois éclairer la zone du pixel mais aussi son pourtour, il moyenne donc le signal sur une plus grande surface. Ce sur-échantillonnage (taille de pixel plus petite que la taille atteignable physiquement) permet de réduire le bruit de grenaille. Cependant dans ce cas, l'analyse doit être adaptée. Il n'est pas possible d'analyser les fréquences jusqu'à un Δy correspondant à la taille de pixel mais seulement jusqu'à celui équivalent à la taille du faisceau. Pour l'outil utilisé, l'équipementier parle de résolution (vue précédemment) avec une valeur de 1.4 nm, limite qui n'est pas contournable. C'est la plus grande des problématiques liées à l'utilisation de cet équipement pour la mesure de rugosité.

Malgré un nombre important de limitations, le CD-SEM est aujourd'hui une des techniques de référence pour la mesure de rugosité de ligne, par l'extraction de la PSD qu'elle permet et le nombre important de lignes (supérieur à la centaine) pouvant être mesurées. Cette technique a ainsi été développée au fil des ans pour repousser ces limitations et, récemment, les diverses recommandations liées à la mesure de rugosité, avec la technique CD-SEM, ont été regroupées en vue d'une standardisation des méthodes [35].

III.2 Microscopie à force atomique (AFM)

La microscopie à force atomique, plus connue sous son nom anglais d'« Atomic Force Microscopy » (AFM), est une technique de référence en microélectronique pour la mesure de profils (Hauteur, SWA, CD). Elle fait partie de la famille des microscopies à sonde locale. Cette technique polyvalente offre la possibilité, selon la configuration de l'équipement, de faire différents types d'analyses donnant accès à une grande variété de propriétés : électriques, morphologiques, mécaniques, thermiques, chimiques, magnétiques. L'AFM permet, dans notre cas, d'obtenir une représentation directe d'objet par un sondage mécanique de sa surface.

Cette technique est considérée comme peu rapide par rapport aux besoins industriels de mesure de plusieurs plaques par heure, et nous le verrons dans le cadre de ses limitations, mais elle est la métrologie de référence du fait du contact direct entre la sonde (nommée pointe) et l'objet considéré.

III.2.1 Le principe de fonctionnement

Le principe de l'AFM réside dans l'interaction atomique entre une pointe et la surface de l'objet à mesurer. L'objet est balayé par cette pointe (Figure III.11) qui va donc suivre la topographie de la surface. La position de la pointe, plus précisément de son support nommé levier, est suivie grâce à la réflexion d'un laser ; une photodiode vient alors collecter et positionner ce signal lumineux. On peut dès lors remonter aux variations de la position de la pointe AFM.

Lié à ce principe de fonctionnement de l'AFM, il existe de nombreux modes de mise en œuvre par contact continu, ou encore oscillation de la pointe, mettant ainsi en jeu les diverses forces d'interaction pointe / objet. Dans notre cas, l'outil utilisé est l'Insight3D de la société Bruker qui permet une reconstruction 3D et nécessite pour cela l'emploi d'un mode dédié nommé « CD ». Le fonctionnement se base sur le principe du mode de contact intermédiaire (« tapping » en anglais) qui consiste à faire osciller le système support-pointe à sa fréquence propre, aux alentours de quelques centaines de kilos hertz. L'oscillation (amplitude et phase) est réglée pour venir toucher ponctuellement la surface. Ce fonctionnement permet de limiter l'endommagement de l'échantillon ainsi que la détérioration de la pointe (usure ou dépôt de matériaux sur celle-ci) ; de plus le fait de relever la pointe avant un nouveau point permet de restreindre les artefacts dus au collage de celle-ci dans le matériau, et ainsi d'augmenter la résolution latérale.

Le principe du tapping étant d'éloigner la pointe de la surface considérée à chaque contact, il est nécessaire, dans le cas du profil de motif, que l'outil puisse faire se déplacer la pointe suivant tous les axes. Le mode « CD » a la particularité de pouvoir effectuer ce type de déplacement et d'oscillation et donc de permettre un recul perpendiculairement à la surface mesurée. Ainsi après un aller-retour, comme illustré par les étapes 1 et 2 de la Figure III.12, un déplacement est effectué de manière parallèle à la surface, schématisé par l'étape 3 de la Figure III.12. Entre chaque point de mesure, la pente locale est recalculée afin de garantir la direction de ces différents mouvements par rapport à l'objet sondé. Finalement, ces opérations sont répétées périodiquement sur la distance souhaitée.

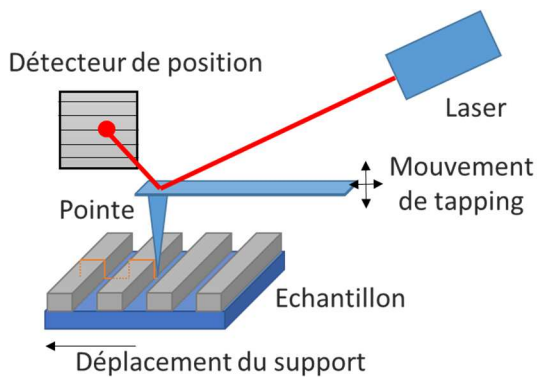


Figure III.11 : Schéma de principe du fonctionnement d'un microscope à force atomique.

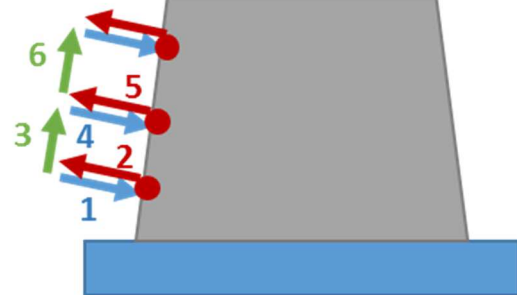


Figure III.12 : Principe de déplacement de la pointe pour le mode "CD".

Pour pouvoir exploiter ce mode « CD » des pointes AFM, adaptées avec une morphologie calibrée, sont utilisées. Ces pointes, nommées aussi « CD », ont la particularité d'être évasées à leur extrémité (partie en contact avec l'objet, cf. Figure III.13) afin de pouvoir suivre le profil d'une ligne sans que le corps de la pointe ne touche celle-ci, contrairement aux pointes standards qui sont coniques ou cylindriques (nanotube de carbone). Notre réseau de lignes est donc sondé avec cette méthodologie et grâce à ce type de pointes, en diverses positions sur la longueur des lignes, suivant un pas de déplacement donné. Au final, la surface du motif complet est obtenue (Figure III.14). Les données acquises correspondent à une suite de points dans un repère orthonormé à 3 axes, de manière ordonnée. Ce dernier élément est nécessaire car, suivant le motif pour une même position (x, y), il peut y avoir différentes valeurs de z.

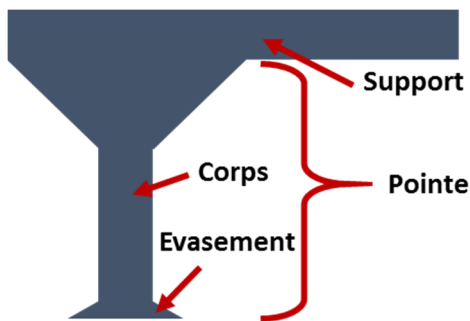


Figure III.13 : Représentation schématique d'une pointe AFM "CD".

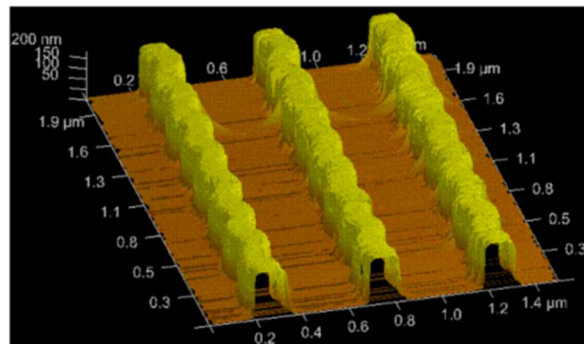


Figure III.14 : Image AFM de $1.5 \times 2 \mu\text{m}^2$ (issue des travaux de G. Freychet [80]).

III.2.2 Extraction de la rugosité en CD-AFM

L'AFM fournissant le profil complet des lignes, il est possible de mesurer la rugosité à diverses hauteurs du motif. Sur le principe, il est primordial d'éviter les zones angulaires telles que les arêtes de la ligne (appelées « Corner Rounding » en anglais). Ainsi on utilisera les données contenues entre 20% et 80% de la hauteur maximum. Par ailleurs, il faut noter qu'il n'existe pas forcément une mesure pour une hauteur donnée et qu'il est donc nécessaire de définir la prise de mesure sur une plage de hauteurs, ou d'interpoler les données. Dans le cas où les hauteurs des lignes sont faibles, par exemple une quarantaine de nanomètres pour une résine utilisée en lithographie e-beam, ce traitement numérique n'est pas possible et le nombre de points de mesure (contact pointe / objet) est parfois limité à un ou deux sur la hauteur de la ligne. Pour cette étude, nous avons fait le choix de n'utiliser qu'une seule mesure par position le long de la ligne avec un placement de bord correspondant à la moyenne sur la

zone 20 - 80% de la hauteur maximum, puisque l'on caractérise la rugosité longitudinale des lignes et non celle sur la hauteur, un exemple de résultat est donné en Figure III.16. Ici, seule la LWR est considérée car, lors de la mesure AFM, on note une dérive non linéaire suivant X, comme visible dans la Figure III.15, due au chargement du système piézoélectrique permettant le déplacement XY de la pointe, ce qui rend impossible la caractérisation de la LER. Ceci nous amène donc à considérer les différentes limites de cette technique.

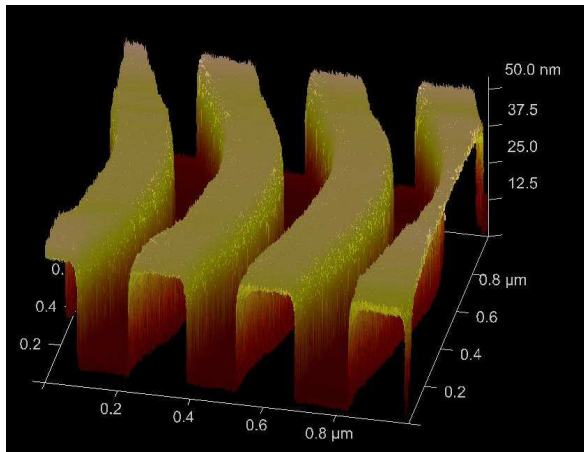


Figure III.15 : Image AFM de $1 \times 1 \mu\text{m}^2$ avec un échantillonnage $\Delta y = 1 \text{ nm}$ et un pas de déplacement théorique de $0,5 \text{ nm}$.

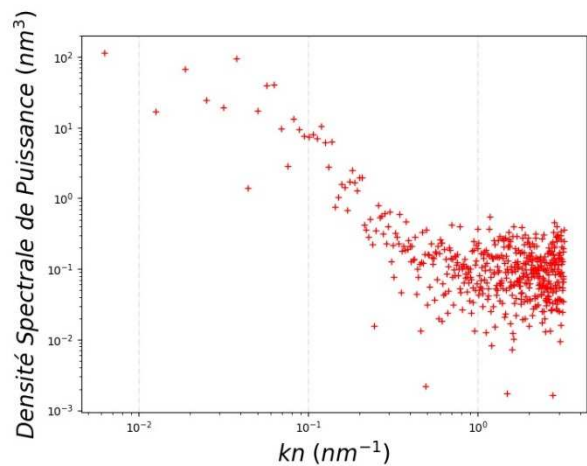


Figure III.16 : PSD LWR extraite grâce à la technique AFM pour 2 lignes de $1 \mu\text{m}$ avec un échantillonnage $\Delta y = 1 \text{ nm}$ correspondant à la Figure III.15.

III.2.3 Limitations

Outre la limitation du type de rugosité extraite, on remarque que le plateau de bruit, au niveau des hautes fréquences, est plus bas en valeur de densité spectrale sur l'AFM que sur le CD-SEM, ceci étant principalement dû à l'utilisation d'une moyenne sur la hauteur de la ligne, soit une limite d'extraction en fréquence moins importante mais qui reste limitante en dessous d'une période de 7 nm .

Par ailleurs, cela est contrasté par le fait que l'AFM est un équipement lent : par exemple pour une acquisition d'une largeur de 200 nm sur $2 \mu\text{m}$ échantillonnée au nanomètre, une heure de temps d'acquisition est nécessaire (le CD-SEM en mesure plus de 200 dans ce même temps). Ce temps d'acquisition peut être réduit, mais il faut alors trouver un compromis entre : la taille de la ligne, la distance d'échantillonnage et le nombre de lignes mesurées. On peut par exemple, avec un temps similaire d'une heure, augmenter raisonnablement Δy à 5 ou 10 nm et pouvoir ainsi mesurer 5 à 10 lignes mais en limitant l'accès aux hautes fréquences. On remarque alors que cette limitation de vitesse est extrêmement liée au nombre de lignes mesurées et donc à la dispersion sur la PSD, qui est importante sur l'exemple visible en Figure III.16. Pour rappel, nous avons vu dans le Chapitre II la nécessité d'un grand nombre de lignes, d'une longueur suffisante, et d'un échantillonnage Δy faible. C'est là, la plus grande difficulté avec la technique AFM.

Dans le cas du mode CD, les pointes adaptées, bien qu'offrant une meilleure définition des profils, sont tout de même limitées sur l'accès aux profils « rentrants » (voir Figure III.17.a), suivant leur largeur d'évasement. On pourrait alors penser qu'un évasement plus grand est préférable ; néanmoins celui-ci entraîne aussi des contraintes sur l'accessibilité entre deux motifs proches (voir Figure III.17.b). Cette limitation est un point bloquant de la technique pour son développement en microélectronique, compte tenu de la réduction continue de la taille des motifs et de la distance qui les sépare.

Par ailleurs, une autre limite due à la pointe peut être atteinte lors du sondage de motifs profonds, lorsque la pointe n'est pas suffisamment longue. Comme représenté dans la Figure III.17.c, le support de la pointe entre alors en contact avec l'objet avant que l'extrémité de la pointe ait atteint le fond du motif. Il existe des pointes longues, mais elles sont alors fragiles. Si une pointe casse au cours d'une acquisition, il faut effectuer à nouveau la totalité de la mesure après un calibrage de la nouvelle pointe.

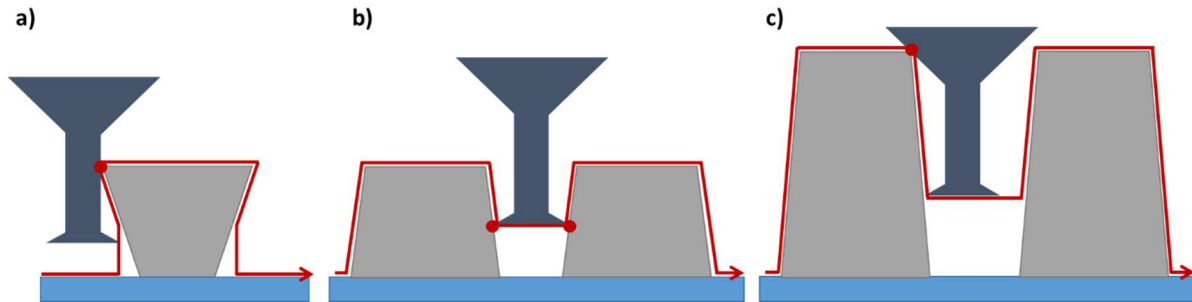


Figure III.17 : Limite d'accès des pointes évasées dans le cas a) d'un SWA trop important, b) un espacement entre deux motifs trop faible, c) une longueur de pointe plus petite que la hauteur des motifs.

La calibration de la pointe est une procédure primordiale dans la technique AFM, dite 3D. La pointe n'étant pas ponctuelle, notamment à son extrémité, il est donc nécessaire après acquisition de déconvoluer sa forme afin d'obtenir les motifs réels. Un exemple de pointe calibrée est donné en Figure III.18, on remarque sur celui-ci que l'on est loin du cas théorique d'évasement présenté en Figure III.13. La différence entre la forme attendue et la réalité est d'autant plus importante qu'elle impacte l'extraction de la rugosité de ligne. En effet, la convolution obtenue par cette pointe n'est pas totalement réversible. Lors de la déconvolution, les oscillations d'une largeur spatiale plus faible que la taille de la pointe (on parle ici de rayon de courbure) ne peuvent pas être reconstruites. On a ici une coupure au niveau des hautes fréquences ; plus important encore, cette convolution impacte la totalité de la plage fréquentielle bien qu'elle ne soit réellement visible que pour les faibles intensités, soit dans le domaine des moyennes et hautes fréquences. Ce phénomène peut être mis en parallèle avec les résultats obtenus sur l'impact du paramètre de mesure Δy (Chapitre II).

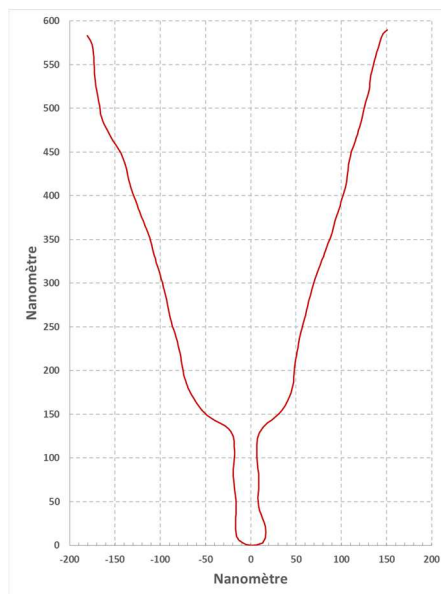


Figure III.18 : Profil réel d'une pointe AFM type CD, la pointe représentant la partie inférieure de 0 à 130 nm, la partie supérieure étant le support.

III.3 La scatterométrie

La scatterométrie, aussi nommée OCD (pour Optical Critical Dimension en anglais), est une technique qui a été fortement développée dans l'industrie de la micro-électronique depuis une quinzaine d'années. En effet, poussée par l'attrait de sa vitesse d'exécution, une statistique de mesure importante (des surfaces micrométriques), et à sa très forte sensibilité de mesure aux dérives morphologiques, cette technique fait l'objet de grandes attentions. La plupart du temps, l'OCD est utilisée dans une logique de mesure de contrôle et plus rarement dans l'objectif d'une extraction de forme d'un objet. Ceci est dû, en grande partie, au fait qu'il s'agit d'une métrologie indirecte : elle nécessite de passer par une étape de modélisation du signal obtenu.

On note néanmoins que la scatterométrie peut permettre d'extraire de nombreuses métriques, notamment celles des profils, soit une vision 3D, ce qui est un de ses atouts.

III.3.1 Le principe de fonctionnement

La scatterométrie est une technique de métrologie par voie optique utilisant l'information contenue dans la diffusion et diffraction d'un rayon lumineux sur un motif périodique, pour retrouver notamment les caractéristiques géométriques de ce dernier. Il existe de multiples paramètres en scatterométrie, comme les angles d'incidence et d'observation, les longueurs d'ondes de lumière utilisées, ou encore la polarisation de ces dernières. Au vu de ces paramètres, il est donc possible de faire des analyses : en angle, en polarisation ou en longueur d'onde. Dans le cas de nos recherches, l'outil utilisé est le Nova T600 qui utilise la technique de réflectométrie, s'intéressant aux variations de l'intensité de la lumière. Comme représenté en Figure III.19, cet outil fonctionne avec une lumière blanche (spectre de longueurs d'ondes continu dans la gamme 235-970 nm), polarisée suivant deux axes, p parallèle au plan d'incidence et s perpendiculaire au plan d'incidence. Ici, le détecteur récupère l'intensité suivant la polarisation et la longueur d'onde pour une position angulaire de détection symétrique à celle de la source. Trois positions angulaires d'incidence existent sur cet outil : (i) normale au support, (ii) suivant un angle donné perpendiculairement à la longueur des lignes, puis (iii) suivant le même angle mais parallèlement à la longueur des lignes.

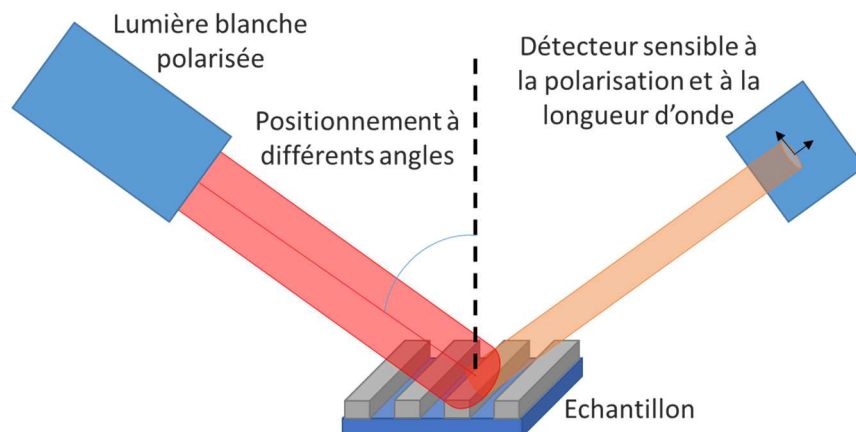


Figure III.19 : Schéma de principe du fonctionnement de la scatterométrie pour l'équipement utilisé lors de ces recherches.

Pour obtenir chacun des spectres (Figure III.20), l'intensité obtenue (pour chaque longueur d'onde, angle et polarisation donnés) est divisée par l'intensité de la lumière incidente. C'est sur la base de tous ces spectres que la mesure est extraite.

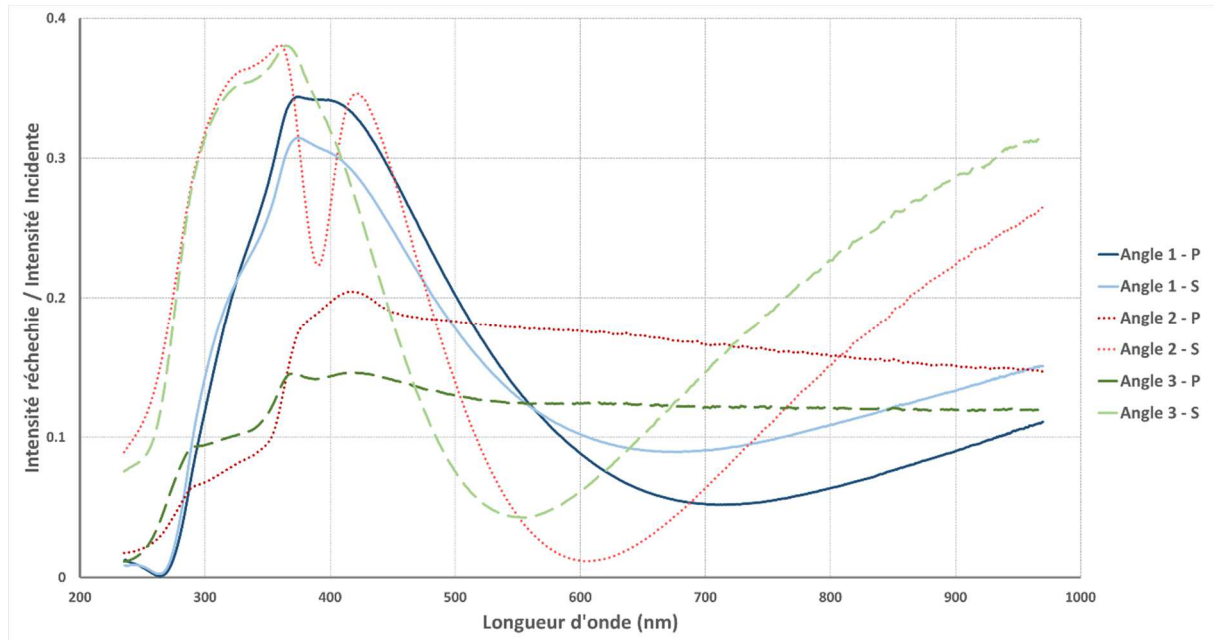


Figure III.20: Spectres obtenus pour un réseau de lignes de résine.

Pour la technique OCD, l'exploitation de la mesure se fait par résolution de problème inverse. Cela consiste, à simuler au travers d'un modèle la réponse optique d'une structure, comme celle visible en Figure III.21.a, et la comparer à la réponse optique expérimentale. Un des modèles largement utilisé dans la littérature est « la théorie rigoureuse des ondes couplées » (en anglais RCWA pour Rigorous Coupled Wave Analysis) [81]. Il s'agit d'une méthode qui décompose en série de Fourier la permittivité diélectrique d'une structure élémentaire périodique. La condition d'utilisation est que la constante diélectrique ϵ soit invariante dans la direction z . La forme de la structure élémentaire est donc un créneau. La modélisation d'une structure complète se réalise par empilement de structures élémentaires en créneau. Pour une structure simple possédant des flancs assez raides, la modélisation demande peu de structures élémentaires, à l'inverse, si la topographie contient des parties formées de pentes (plus ou moins raides), le découpage en structures élémentaires sera plus important. Une fois cette modélisation réalisée, le champ électromagnétique réfléchi et transmis peut être déterminé aux frontières de la structure complète. Il est donc possible de simuler les signatures optiques en fonction de la longueur d'onde et de l'angle d'incidence de l'onde incidente. Les écarts entre les signatures issues du modèle simulé et les courbes expérimentales sont extraits, pour chaque point des spectres, puis les paramètres morphologiques de la structure modélisée sont modifiés. Suivant les variations d'écarts obtenues, chacun des paramètres du modèle est ajusté suivant un nombre d'itérations donné ou jusqu'à un point de convergence (lorsque l'erreur cesse de diminuer). Nous avons utilisé le logiciel « Nova Mars » pour créer les modèles et la simulation de leurs réponses optiques. Ce logiciel est associé aux équipements Nova qui ont été utilisés pour les analyses expérimentales. Les motifs représentés en Figure III.21 sont considérés comme étant le profil moyen des diverses lignes, la largeur du support représentant le pitch.

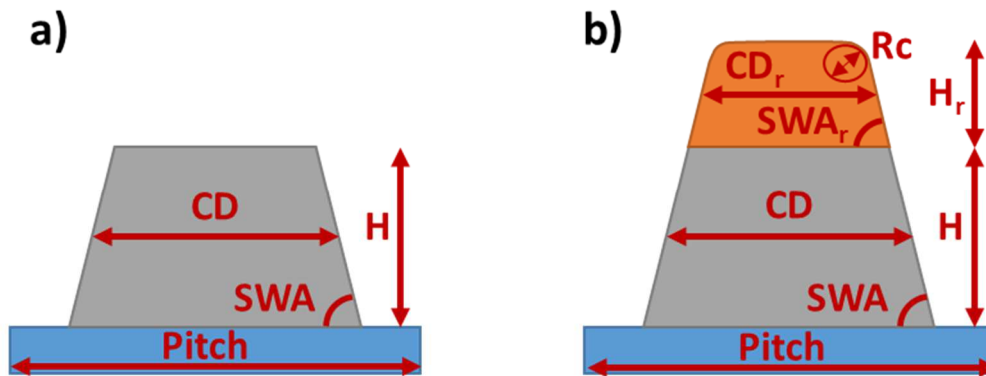


Figure III.21 : Exemple de structure, a) dans le cadre d'un seul matériau, b) d'un motif présent sur deux couches de matériaux différents.

Pour déterminer correctement la structure, il est nécessaire d'obtenir lors de la convergence, le minimum global. Or, si les paramètres d'entrées sont trop nombreux ou que leur plage de variations sont trop importantes, le résultat final risque de converger vers un minimum local. L'affinement des paramètres du modèle est donc primordial. C'est une procédure qui peut être complexe et qui nécessite une bonne connaissance des procédés et matériaux mis en œuvre, avant de relâcher ou contraindre certains paramètres du modèle. De plus, ces derniers peuvent parfois être corrélés entre eux ceci amène aussi à des extractions faussées. Dans ce cas il faut alors fixer un des deux paramètres s'il est connu, ou lorsqu'il est possible de le déterminer, au travers d'une autre mesure. Plus le nombre de couches de matériaux différents est important, plus le nombre de paramètres nécessaires augmente, comme visible en Figure III.21.b avec deux couches, et plus il sera compliqué d'affiner ces paramètres. Il est donc important, sur le modèle, de définir les paramètres d'intérêt (ce que l'on souhaite mesurer), en parallèle des paramètres ayant un fort impact sur les spectres et pouvant donc varier. A titre de contre-exemple, le pitch est très important dans les analyses de scattérométrie mais, une fois extrait, il ne varie pratiquement pas d'une plaque à l'autre compte tenu de la physique liée aux procédés, il peut donc être fixé dans la poursuite d'une analyse.

III.3.2 Extraction de la rugosité par OCD

Dans le cas de la rugosité, celle-ci n'étant pas un phénomène périodique ni même correctement représentable à partir d'une seule ligne, et du fait de la statistique, elle fut très longtemps considérée comme inaccessible par ce type de technique. Par la suite, une étude [82] a montré que l'erreur résiduelle de l'affinement pouvait être considérée comme proportionnelle à la rugosité de ligne et donc utilisée dans ce sens. Depuis une dizaine d'années maintenant, de multiples études ont vu le jour sur l'extraction des défauts non périodiques, dont la rugosité, pour les différents types de scatterométrie [83–87]. Cette technique a donc énormément évolué avec diverses méthodes associées et peut être considérée comme une technique permettant d'obtenir la rugosité de ligne. Dans le cadre de l'équipement utilisé, la latitude de déplacement du détecteur et de la source de lumière étant restreinte, une seule méthode peut être mise en place pour ce type d'extraction.

Cette méthode se nomme Effective Medium Approximation (EMA) en anglais, soit « approximation par un milieu effectif » [84]. Le principe est de considérer (Figure III.22) que dans cette région de bord de ligne, la densité moyenne du matériau constituant le motif sur toute la longueur des lignes est moindre. On peut alors assimiler ce matériau de bord de ligne à un matériau poreux dont la densité

est déterminée par un ratio entre la densité du matériau massif et celle de l'environnement (ici l'air). On notera que cette densité, et la largeur considérée pour cette zone « poreuse », ne sont dépendantes que du paramètre de rugosité σ . Il est donc impossible, par la méthode EMA induite par l'analyse 2D, d'obtenir la totalité des paramètres notamment α et ξ , et aucune donnée concernant les fréquences n'est alors disponible.

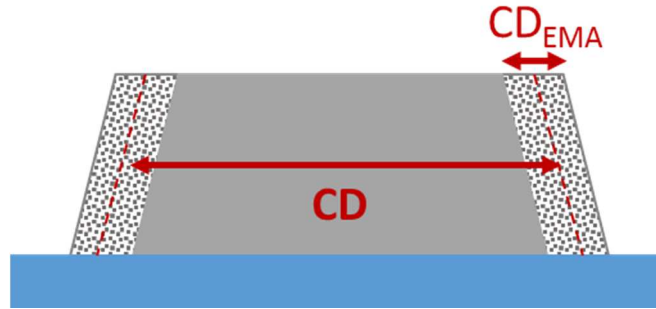


Figure III.22 : Schématisation d'un motif ligne avec une rugosité représentée grâce à la méthode EMA

Pour la mise en place d'un matériau de type EMA, divers modèles existent. Celui considéré ici est le modèle de Bruggeman [88] car il a l'avantage d'être symétrique et donc plus proche d'une réalité physique. Cela signifie que le mélange air-silicium est équivalent au mélange silicium-air. Il donne la permittivité du matériau EMA, ϵ_{eff} suivant :

$$\sum_i^n f_i \frac{\epsilon_i - \epsilon_{eff}}{\epsilon_{eff} + \gamma(\epsilon_i - \epsilon_{eff})} = 0 \quad (III.1)$$

avec i le numéro de chaque matériau considéré, ϵ_i et f_i respectivement la permittivité et la proportion de chaque matériau et γ le facteur de dépolariation, généralement de 1/3 [84]. La valeur de permittivité diélectrique permet par la suite de déterminer l'indice optique n par l'expression suivante :

$$\frac{\epsilon}{\epsilon_0} = n^2 \quad (III.2)$$

Avec ϵ_0 la permittivité diélectrique du vide et l'indice optique k par :

$$k = 1 - \frac{1}{n} \quad (III.3)$$

Lors de l'utilisation d'un EMA il est nécessaire, en plus de connaître les indices optiques des différents matériaux, de déterminer leur fraction dans le matériau final dit « mixte ». La Figure III.23 présente, pour exemple, les spectres simulés dans le cas d'une proportion d'air, f_{air} , dans le matériau mixte allant de 10% à 90% par pas de 10% pour une structure fixe. Pour comparaison, on peut faire varier le CD_{EMA} pour une même proportion air/silicium. La Figure III.24 montre donc les spectres pour $f_{air} = 50\%$, avec des variations de CD_{EMA} . En comparant les Figure III.23 et Figure III.24, on remarque que ces deux paramètres n'ont pas le même effet, on voit par exemple que pour une longueur d'onde aux alentours de 290 nm, les polarisations représentées en gris et violet sont quasi invariables (superposition des courbes d'une même gamme de couleur) pour les variations de densité, contrairement au cas de variation de CD_{EMA} . Une étude de calibration de la proportion de matériaux pourrait permettre d'obtenir une valeur de CD_{EMA} équivalente à la valeur de σ , et non plus proportionnelle seulement. Dans notre cas, nous sommes restés sur une extraction de valeurs proportionnelles, en fixant $f_{air} = 50\%$.

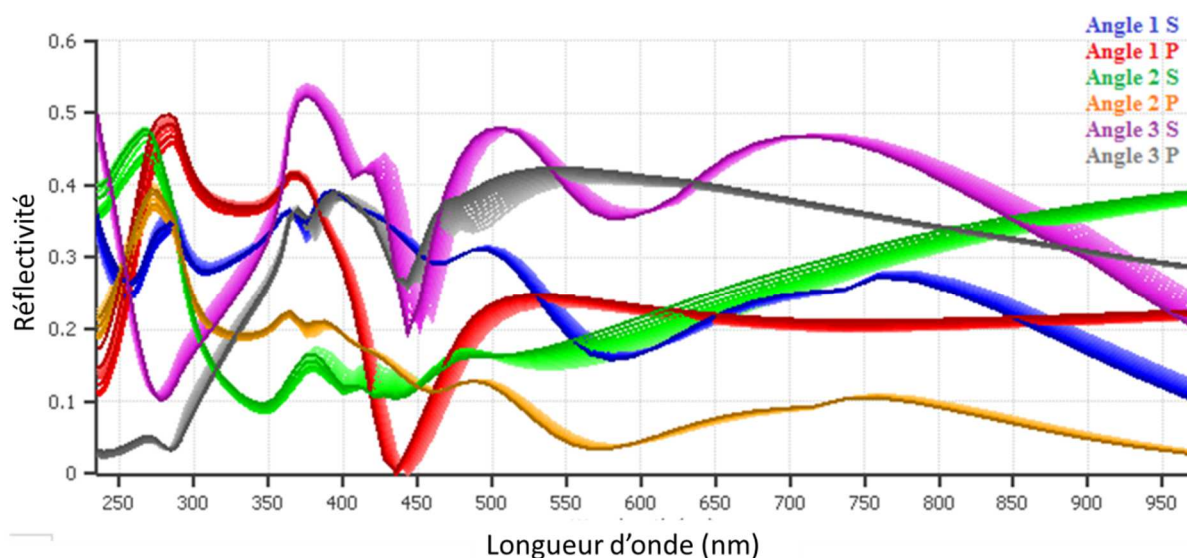


Figure III.23 : Spectres obtenus pour une structure de Pitch = 200 nm , CD= 50 nm , SWA=88° , H = 150 nm , CDEMA = 2 nm , pour une proportion d'air dans le matériau allant de 10% à 90% par pas de 10%. Chaque gamme de couleurs correspondant à une polarisation, les couleurs claires étant la réponse à 10%.

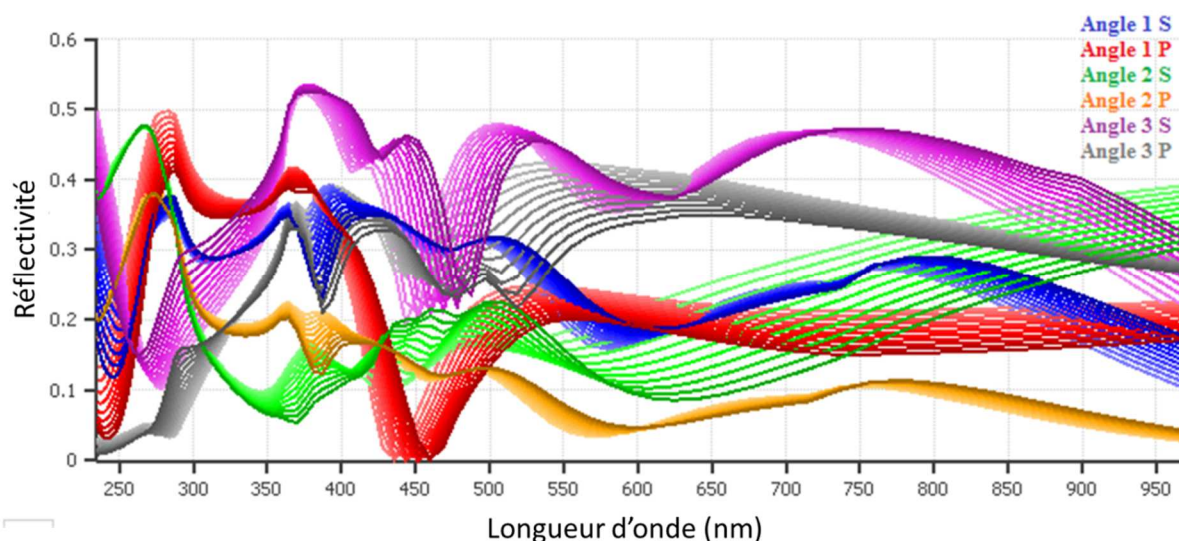


Figure III.24 : Spectres obtenus pour une structure de Pitch = 200 nm , CD= 50 nm , SWA=88° , H = 150 nm , matériau mixte 50%/50%, pour une variation de CD_{EMA} de 0 nm à 10 nm par pas de 1 nm. Chaque gamme de couleurs correspondant à une polarisation, les couleurs claires étant la réponse à 0 nm.

III.3.3 Limitations

Diverses limitations apparaissent à l'utilisation de la scatterométrie, en général, et certaines sont plus spécifiques à l'extraction de rugosité. Premièrement, en scatterométrie il est primordial de connaître les propriétés optiques de chacun des matériaux considérés, incluant tous ceux sous le motif mesuré et sur toute la gamme de longueur d'onde utilisée. Ainsi pour extraire ces coefficients optiques, des analyses ellipsométriques peuvent être réalisées sur des substrats séparés contenant chacun un dépôt uniforme d'un des matériaux considérés, pour une épaisseur proche de la gamme de mesure. Néanmoins, les divers matériaux observés pour nos études subissent les contraintes des étapes de

fabrication qui peuvent altérer leurs propriétés physico-chimiques et donc optiques. Ainsi, cette technique est souvent associée à d'autres caractérisations, lors de sa mise en place.

Une seconde limitation, en lien avec la première, concerne le temps de développement des modèles qui contrebalance une acquisition de données très rapide (quelques secondes). Le temps d'analyse quant à lui est très variable suivant le nombre de paramètre, logiciel utilisé ou encore la puissance de calcul disponible

La scatterométrie est considérée comme une technique statistique nous l'avons vu, donc en plus de ne mettre que peu de temps à acquérir un signal, celui-ci est représentatif d'un grand nombre de lignes. Ceci est intéressant dans le cadre de la rugosité de ligne au vu de l'effet du nombre de lignes mesurées (Chapitre II). Cependant, il faut noter que la surface structurée avec des lignes doit être largement supérieure à la taille du faisceau qui est d'une trentaine de micromètres de large sur l'équipement utilisé. De plus sur certains angles, le faisceau peut couvrir une surface (empreinte du faisceau) beaucoup plus importante. Le fait que le faisceau couvre une grande surface est contraignant mais permet à cette technique d'avoir en une mesure un résultat statistiquement représentatif car issu d'une multitude de ligne. Par ailleurs, on notera que dans l'industrie ce n'est pas les motifs constituant une puce qui sont directement mesurés mais d'autres, présent dans les chemins de découpe des puces par exemple, adapté pour le test afin de suivre le bon déroulement des procédés.

On retiendra que la technique mise en place ici, à travers l'utilisation d'un modèle EMA, ne permet que d'obtenir le paramètre σ et on ajoutera à cela qu'il n'est pas forcément possible de différencier les rugosités LER et LWR dans ce cas d'utilisation.

III.4 La diffusion des rayons-X aux petits angles

La diffusion des rayons-X aux petits angles (SAXS pour Small-Angle X-Ray Scattering en anglais) est une technique encore peu envisagée par l'industrie de la microélectronique. Elle est essentiellement utilisée pour la caractérisation structurelle des nanomatériaux. Cependant, du fait que cette technique utilise des photons de courte longueur d'onde (de l'ordre de 0,1 nm), elle a le potentiel pour analyser des objets de très petites dimensions. Par ailleurs, la taille du faisceau de rayons X, de l'ordre de la centaine de microns, permet de sonder un grand nombre d'objets simultanément, donnant un caractère statistique à cette mesure. Ces deux caractéristiques font que cette technique suscite un intérêt grandissant dans la communauté de la microélectronique [11]. Depuis maintenant une dizaine d'années, de nombreuses études visent à déterminer, grâce à cette technique, la morphologie de lignes dans des réseaux [89–92]. Cela a fait naître une nouvelle terminologie, le CD-SAXS ou aussi XCD.

III.4.1 Interactions rayons-X, matière

De manière similaire au cas des électrons, il existe de multiples interactions entre les rayons-x et la matière. Une partie de ces différentes interactions est présentée en Figure III.25. On peut voir que les rayons-X incidents se répartissent en trois groupes : absorbés (en orange pastel), diffusés (en orange foncé) et transmis (en orange, identique aux rayons-x incidents).

- **L'absorption de photons** correspond à l'excitation d'électrons. Ces derniers sont, par la suite, soit éjectés tels que les électrons Auger, soit tendent à revenir à l'état fondamental en libérant de l'énergie sous forme de photons X de plus faible énergie : photons de fluorescence d'énergies caractéristiques des atomes qui constituent l'échantillon [93, 94].

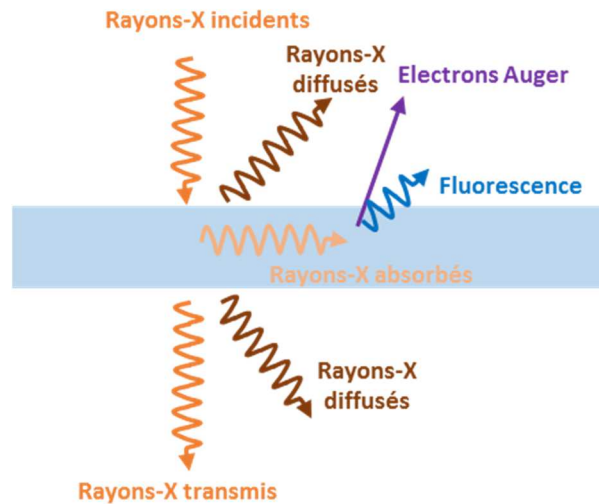


Figure III.25 : Principales interactions rayons-x matière.

- **Les rayons-X transmis** sont ceux qui n'interagissent pas avec la matière lorsqu'ils la traversent, il n'y a alors changement ni d'énergie ni même de direction, il s'agit là d'une « non-interaction ».
- **La diffusion** est la dernière interaction qui nous intéresse pour cette étude. Ce type d'interaction se compose d'une partie inélastique, avec perte d'énergie, appelée diffusion de Compton [95], et d'une partie élastique, sans perte d'énergie, appelée diffusion Thomson [96]. En deçà d'une énergie de 45 keV, la diffusion de Compton peut être négligée. Nos mesures ayant toujours été réalisées à des énergies inférieures à 20 keV, on a négligé ce phénomène et considéré seulement la partie élastique de la diffusion X.

Interaction avec un électron

Pour comprendre plus en détail l'interaction rayons X-matière dans le domaine de la diffusion de Thomson, il faut se pencher sur le principe d'Huygens-Fresnel. Celui-ci exprime que lors de l'illumination de la matière par un faisceau de rayons-X, ce sont principalement les électrons qui interagissent avec l'onde électromagnétique. Sous l'impact de cette dernière, les électrons se mettent à osciller et émettent alors une onde à la même énergie que le faisceau incident et ce dans toutes les directions de l'espace. Thomson donne l'intensité diffusée à une distance r par un électron comme :

$$I_p = \left(\frac{r_e}{r}\right)^2 P(2\theta)I_0 \quad (\text{III.4})$$

avec r_e le rayon de l'électron ($2,818 \cdot 10^{-15}$ m), I_0 est l'intensité du faisceau incident, 2θ l'angle entre le faisceau incident et la direction d'observation et $P(2\theta)$ le facteur de polarisation ($= \frac{1+\cos(2\theta)}{2} \approx 1$ dans le cas de diffusion aux petits angles, i.e. pour $\theta < 3^\circ$)

Interaction avec un atome

On considère alors l'interaction d'un atome avec l'onde électromagnétique X, comme la somme des amplitudes diffusées par chaque électron libre, ce qui définit son facteur de diffusion [80] :

$$f_0(\vec{q}) = \int_V \rho_a(\vec{r}) e^{i\vec{q}\cdot\vec{r}} dV \quad (\text{III.5})$$

où $\rho_a(r)$ est la densité électronique de l'atome sur son volume V , \vec{r} le vecteur position et \vec{q} le vecteur de diffusion défini comme la différence entre le vecteur d'onde diffusé et le vecteur d'onde incident, comme visible en Figure III.26 et qui a pour norme :

$$\mathbf{q} = \frac{4\pi}{\lambda} \sin(\theta) \quad (\text{III.6})$$

avec θ le demi-angle de diffusion définit précédemment et λ la longueur d'onde du faisceau de rayons-X incident. Il est ainsi possible de parler de diffusion de manière angulaire avec θ , ou dans l'espace nommé « réciproque », formé des composantes du vecteur \vec{q} suivant l'espace de l'objet x , y et z .

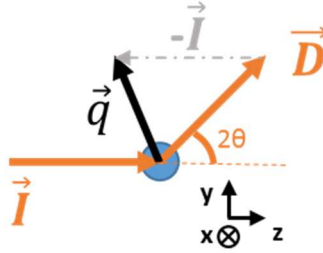


Figure III.26 : Représentation schématique 2D des vecteurs : incident \vec{I} , d'onde diffusée \vec{D} , de diffusion \vec{q} , ainsi que l'angle de diffusion 2θ .

Il faut noter que les électrons, dans le cas d'un atome, ne sont pas complètement libres car liés au noyau de l'atome et sur des niveaux d'énergie quantique discrets. Ainsi, des termes de correction de dispersion anormale $f'(E)$ et $f''(E)$, fonction de l'énergie du faisceau incident et de la nature du matériau irradié, sont à ajouter à f_0 :

$$f_\alpha(\vec{q}, E) = f_0(\vec{q}) + f'(E) + i.f''(E) \quad (\text{III.7})$$

Ces coefficients correctifs sont non négligeables, notamment lorsque l'énergie des photons incidents est proche du seuil d'excitation de l'élément chimique considéré.

On retiendra finalement que : toute particule diffuse dans toutes les directions, suivant sa densité électronique ; ainsi il sera possible en mesurant la variation d'intensité obtenue, de déterminer des variations de densité électronique.

III.4.1 Principe et équipements

Dans le cas d'une particule isolée centro-symétrique, on définit donc le facteur de forme, variation d'intensité selon q , par :

$$F^2(q) = \int_{V_2} \int_{V_1} \rho_{r1} \rho_{r2}(r) \frac{\sin(q.r)}{q.r} dV_1 dV_2 \quad (\text{III.8})$$

où ρ_{r1} et ρ_{r2} sont les densités électroniques, respectivement aux positions 1 et 2, et r la distance entre les deux positions. Dans le cas où plusieurs particules sont présentes, on voit apparaître des interactions entre les diverses diffusions, sous forme d'une figure de diffusion dont l'analyse permet de remonter à la forme de l'objet. Si les particules sont positionnées régulièrement (ou seulement deux particules), on peut définir l'effet de la structure sous le terme de fonction d'interférence ou facteur de structure $S(q)$ donnant lieu à des taches de diffraction. On obtient alors l'intensité finale comme :

$$I(q) = N_{part} |F(q)|^2 S(q) \quad (\text{III.9})$$

où N_{part} est le nombre de particules considérées.

Il faut noter que compte tenu de la longueur d'onde des rayons-X (0,1 nm environ) et des dimensions caractéristiques des lignes que nous souhaitons analyser (100 nm environ), d'après l'Equation III.6, la gamme d'angle de diffusion autour du faisceau incident à observer pour extraire une information est petite et comprise entre 0 et au plus quelques degrés, ce qui justifie le nom de diffusion des rayons-X aux petits angles pour cette technique de caractérisation.

Par ailleurs, on distingue deux configurations expérimentales pour mesurer la diffusion des rayons-X : le Grazing-Incidence SAXS (GiSAXS) qui, comme son nom l'indique, observe la diffusion du faisceau incident après réflexion des photons à la surface d'un échantillon plan, et le T-SAXS ou plus simplement SAXS qui considère la diffusion du faisceau incident après transmission de celui-ci au travers de l'échantillon. La configuration SAXS par transmission est schématisée sur la Figure III.27, et la configuration GiSAXS sur Figure III.28. Les deux techniques fonctionnent sur l'acquisition par un détecteur bidimensionnel d'une image de diffusion des rayons X autour du faisceau direct. L'image obtenue est représentative d'une coupe 2D dans l'espace réciproque de l'objet analysé et peut être convertie dans cet espace à partir des angles d'incidences et de diffusion :

$$\vec{q} = \frac{2\pi}{\lambda} \begin{bmatrix} \cos(\alpha_f) \sin(2\theta) - \cos(\alpha_i) \\ \cos(\alpha_f) \sin(2\theta) \\ \sin(\alpha_f) - \sin(\alpha_i) \end{bmatrix} \quad (\text{III.10})$$

où α_i est l'angle d'incidence et α_f et 2θ les angles de diffusion, comme représenté suivant la configuration sur la Figure III.27 ou Figure III.28.

On peut constater, par exemple dans le cas de la transmission, qu'une rotation de l'échantillon autour de l'axe y permet de faire varier les coordonnées l'espace de l'échantillon et donc dans le réseau réciproque d'obtenir une série de coupes 2D de ce réseau, permettant donc une reconstruction tridimensionnelle du réseau réciproque de l'échantillon et donnant la possibilité d'extraire la forme 3D de l'objet ou du réseau d'objets sous le faisceau de rayons X. Ce type d'approche a fait l'objet de plusieurs publications ces dernières années [91, 97].

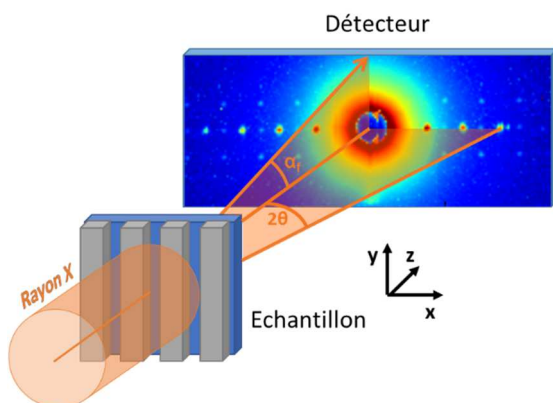


Figure III.27 : Schéma de principe de la technique SAXS par transmission, image de diffraction issue des travaux de G. Freychet [80], ici $\alpha_i=0$.

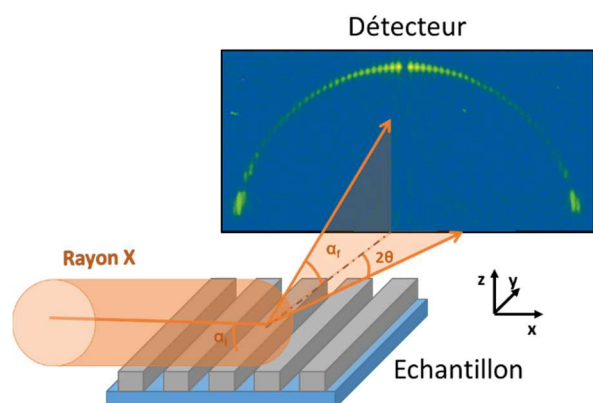


Figure III.28 : Schéma de principe de la technique GiSAXS, image de diffraction issue des travaux de G. Freychet [80].

Au cours de ces travaux, pour des raisons de taille d'échantillon (dimension du réseau de lignes de la centaine de microns) et de taille du faisceau X (centaine de microns), mais aussi pour limiter nos

simulations à une approche purement cinématique de la diffusion des rayons qui est explicitée ultérieurement dans ce manuscrit, nous avons essentiellement réalisé des mesures en utilisant la configuration SAXS i.e. en mode transmission. Cela a eu des conséquences sur le choix des équipements compatibles avec cette configuration.

En effet, Les équipements de laboratoires accessibles pour cette étude (Diffractomètre SmartLab de la société Rigaku ou l'équipement de A. De Geyer du département Recherche Fondamentale du CEA Grenoble) fonctionnent avec des sources de rayons-X constituées d'une anode en cuivre et ont donc la particularité d'émettre un faisceau X à une énergie de 8 keV. Cette énergie est entièrement absorbée par les substrats silicium de plus de 700 μm (transmission $< 10^{-4}\%$ des photons incidents), ce qui exclut l'utilisation de ces équipements de laboratoire, sauf en amincissant sensiblement l'épaisseur du substrat, posant d'autres difficultés en termes de préparation d'échantillons.

Une alternative aux équipements de laboratoire peut être trouvée sur les grands instruments que sont les sources synchrotrons. Ces grands instruments ont la particularité de fournir un faisceau blanc avec une grande bande passante en énergie (quelques 100 eV à plusieurs 100 de keV) que l'on peut ensuite monochromatiser grâce à des optiques adaptées aux rayons-X. Nous avons ainsi pu travailler sur une gamme d'énergie allant de 11 keV jusqu'à 18 keV (soit plus de 36% de transmission pour un substrat de 750 μm d'épaisseur). Des mesures SAXS en transmission, mais aussi GISAXS, ont été faites sur les synchrotrons de Berkeley (ALS), de New-York (Brookhaven National Laboratory) ou de Grenoble (ESRF). En fonction des lignes de lumière utilisées, divers paramètres instrumentaux peuvent changer tels que la distance entre l'échantillon et le détecteur, la taille et la sensibilité du détecteur, ou encore l'empreinte du faisceau ou son flux (nombre de photons par unité de surface). La calibration des équipements est nécessaire afin de pouvoir normaliser les données, mais aussi pour permettre de convertir les images de diffraction dans l'espace réciproque et de pouvoir ensuite extraire des informations quantitatives de ces images.

III.4.1 Extraction de la rugosité de ligne par CD-SAXS

Avant de démontrer où et comment extraire l'information de rugosité d'une mesure SAXS, il est important de rappeler les grandes lignes de l'analyse de CD, ou plus généralement l'analyse morphologique 2D ou 3D par cette technique.

Pour l'extraction de la morphologie, dans le cas de nos structures en réseaux de lignes, on peut expliquer la figure de diffusion totale en considérant les deux cas suivants :

- La figure de diffusion d'une seule ligne qui est donnée par le facteur de forme de la ligne comme représenté en Figure III.29. Ce facteur de forme est semblable au cas de la particule, mais pour une somme des intensités suivant la longueur des lignes (axe y) et non pas radiale. On observe, sur cette même figure, que la variation de la taille de la ligne (CD) fait fluctuer la position du minimum d'intensité.
- La diffraction due à l'organisation en réseau d'éléments. Soit, d'après la loi de Bragg, l'apparition de pics d'intensité aussi appelés pics de Bragg suivant :

$$2 \times d \times \sin \theta = n \times \lambda \quad (\text{III.11})$$

avec d la distance entre deux éléments (pas du réseau), θ le demi-angle de diffusion décrit précédemment, n l'ordre de diffraction et λ la longueur du faisceau de rayons-X. On voit apparaître ici la notion de « facteur de structure » mentionnée sur l'Equation III.9.

En considérant la structure complète qu'est un réseau de lignes on obtient alors, comme visible Figure III.30, des pics suivant les positions définies par la loi de Bragg et ayant une intensité correspondant au facteur de forme, on peut donc à partir de ces données extraire les informations moyenne du réseau de ligne.

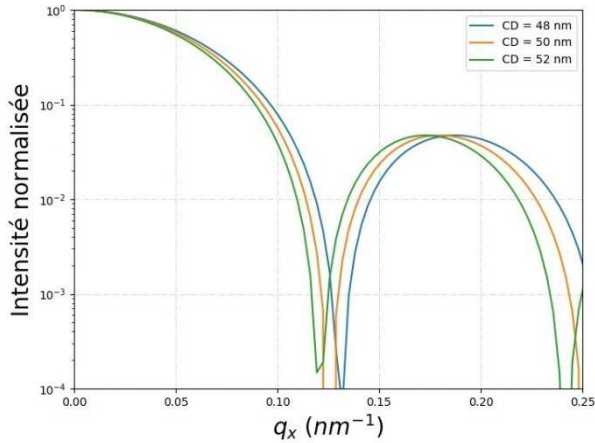


Figure III.29 : Intensité de diffusion de diverses lignes isolées correspondant alors au facteur de forme.

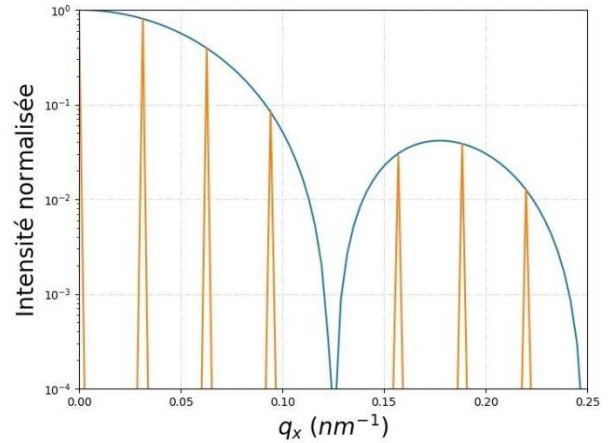


Figure III.30 : Intensité décrite Equation III.9 pour un réseau de lignes (orange) et son enveloppe (bleu) correspondant au facteur de forme moyen des lignes.

En effet, la position des pics de diffraction (facteur de structure) nous donne un accès direct à la période du réseau de lignes (pitch) selon la relation :

$$Pitch = \frac{2\pi n}{q_{x,n}} \quad (III.12)$$

où n est l'ordre de la tâche et $q_{x,n}$ la position associée à cette tâche de diffraction [80].

La forme moyenne d'une ligne dans le réseau est donnée par l'intensité relative des pics de diffraction, dans laquelle se retrouve le facteur de forme d'une ligne. Dans une approche purement analytique proposée par C. Wang [98], et basée sur la théorie cinématique de la diffusion des rayons-X, il est possible de remonter à la largeur moyenne (CD) des lignes grâce à :

$$CD = \frac{Pitch}{n\pi} \arccos \sqrt{\frac{I_{2n}}{I_n}} \quad (III.13)$$

I_n et I_{2n} sont respectivement les intensités intégrées sous les pics de diffraction de l'ordre n et $2n$ pour $q_y = 0 \text{ nm}^{-1}$, défini par :

$$I_n = \sum_{n=-\infty}^{\infty} \delta\left(q_x - \frac{2n\pi}{L}\right) 16\lambda^2 \left(\frac{\sin\left(q_x \frac{CD}{2}\right)}{q_x}\right) \quad (III.14)$$

avec λ la longueur d'onde du faisceau de rayons-X et δ un peigne de Dirac.

Cependant, cette approche analytique ne prend pas en compte d'autres paramètres tels que la rugosité de ligne ou le SWA. Aussi depuis quelques années, la méthode d'analyse la plus employée consiste en la résolution du problème inverse, dans une approche de type Monte Carlo Markov [80, 89] assez similaire à la technique OCD. On peut citer par exemple le logiciel Xi-Cam, développé par l'ALS [92], le synchrotron de Berkeley, qui sur la base d'une première analyse 2D extrait analytiquement le pitch. Ensuite, sur des clichés mesurés pour différents angle de rotation de l'échantillon et donnant accès à la composante q_z du réseau réciproque, la morphologie 3D de

l'échantillon est extraite en schématisant une ligne comme la superposition d'une multitude de trapèzes de même hauteur, de SWA et CD variables. Ce modèle est utilisable à la fois en mode GISAXS et en mode SAXS. On note que dans le cas du GISAXS [99], cette approche est similaire à celle utilisée en scattérométrie, avec notamment la prise en compte des réflexions multiples et des effets dynamiques présents en incidence rasante dans les empilements considérés.

Dans le cas de la technique SAXS en transmission, en considérant que la source de rayons-X est ponctuelle et que la distance d'observation est assez éloignée de l'objet considéré, on se trouve dans le cas d'une diffraction dite de Fraunhofer, les effets dynamiques et de diffusion multiples sont alors négligeables et une approche cinématique de la diffusion peut alors être considérée. Ce type de diffraction est modélisable par le module au carré de la Transformée de Fourier (TF) appliquée à une représentation des lignes. Plus précisément, dans la suite de ces travaux, nous avons appliqué cette approche à une image binaire représentant un réseau de lignes en vue du dessus, comme celle représentée sur la Figure III.31. Cette représentation a été nommée « échantillon virtuel », nous y reviendrons plus amplement dans les prochains chapitres. Dans la suite de cette étude sur l'utilisation du SAXS, nous limiterons la simulation des figures de diffusion à une simulation 2D dans le plan q_x, q_y dans le réseau réciproque. Un exemple de résultat, obtenu par ce traitement, est donné sur la Figure III.32. On observe, sur cette simulation, une partie du résultat dans l'espace réciproque avec une échelle de niveau de gris en \log_{10} pour une meilleure lisibilité, les pics de Bragg du réseau de lignes sont visibles le long de q_x à $q_y = 0 \text{ nm}^{-1}$. On peut voir qu'un signal est diffusé en dehors des pics de Bragg. Dans le cadre de ces travaux, nous verrons qu'une partie importante de l'information de rugosité est justement présente dans cette diffusion. Or, quelle que soit la méthodologie utilisée dans la littérature pour extraire l'information de morphologie, le principe reste basé sur l'intensité max ou intégrée de chaque tâche de Bragg sans prendre en compte la diffusion autour. Deux de ces méthodes pour extraire la rugosité de ligne en SAXS suivant les cas considérés sont explicité ci-après avant la nouvelle méthode que nous avons développée.

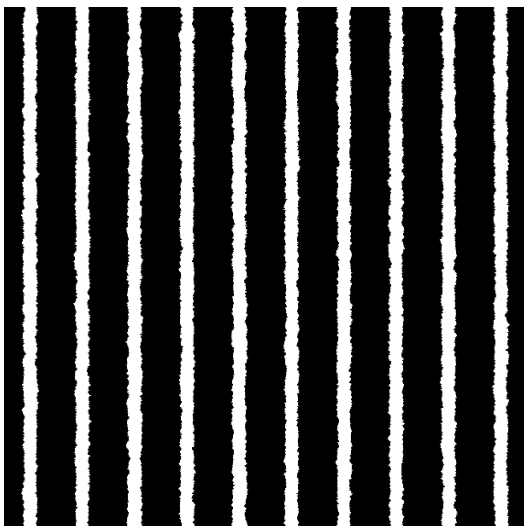


Figure III.31 : Exemple d'un échantillon virtuel $2 \times 2 \mu\text{m}^2$ comportant 10 lignes $\text{CD}=50 \text{ nm}$, $\text{Pitch}=200 \text{ nm}$, aux bords non corrélés ; généré avec $\sigma = 3 \text{ nm}$, $\alpha = 0.5$, $\xi = 10 \text{ nm}$, $\Delta y=1 \text{ nm}$.

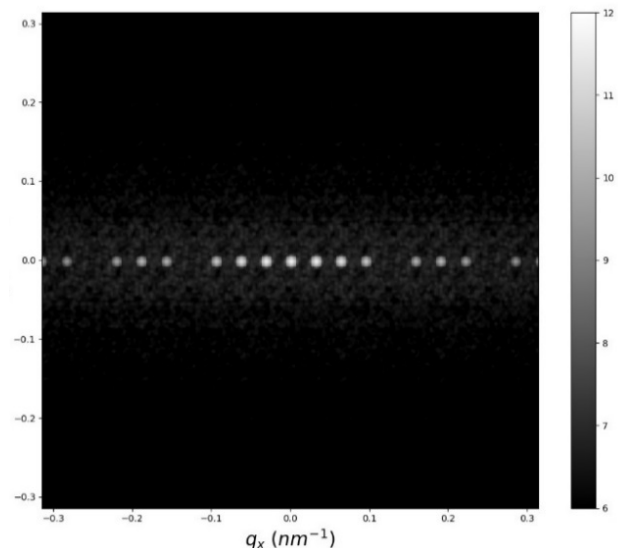


Figure III.32 : Exemple de simulation d'un cliché de diffusion en SAXS à partir de la Figure III.31.

Méthode 1 : Extraction analytique

Des études précédentes ont montré la faisabilité dans le cas de lignes aux bords corrélés ou anti-corrélés [34, 100–102]. Ces études se basent sur des rugosités périodiques, ce qui signifie qu'en plus de la structure suivant le pitch des lignes, il existe un aspect de structure dans le sens de la longueur des lignes, il apparaît donc des pics de Bragg suivant q_y comme visible en Figure III.33.

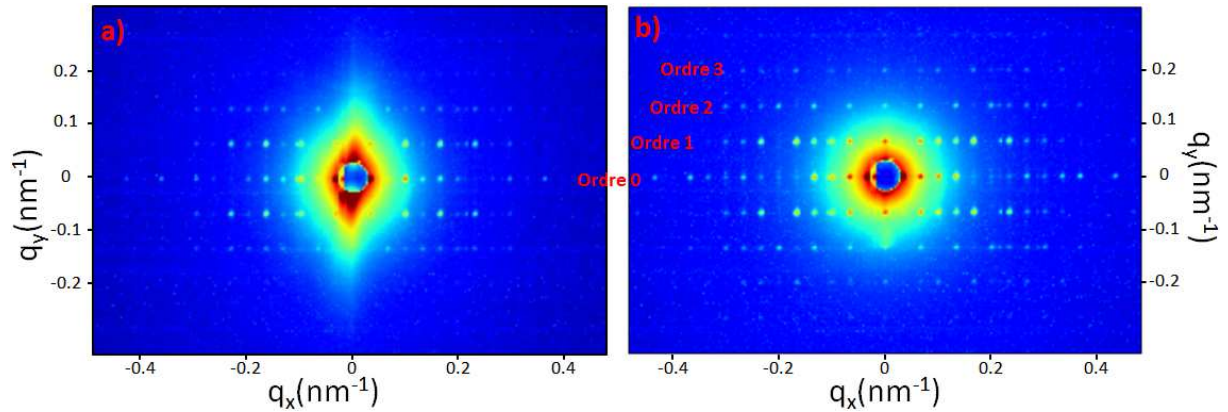


Figure III.33 : Clichés CD-SAXS issus des travaux de G. Freychet [80], obtenus sur des échantillons de faible épaisseur comportant : a) une rugosité de bords corrélés, b) une rugosité de bords anticorrélés ; mesures réitérées sur le synchrotron de Berkeley (ALS) en avril 2017 à une énergie de 10 keV

A partir d’une extraction ponctuelle sur ces pics, il est possible comme dans le cas d’extraction analytique de pitch et de CD, d’obtenir une information de rugosité séparant LER et LWR. En effet, d’après C. Wang [100], on a une relation entre la amplitude de rugosité l et le ratio de l’intensité des pics suivant q_y , dans le cas LER par :

$$\frac{I_0}{I_{\pm 1}} \sim \frac{\cos^2(q_x \frac{l}{2})}{\sin^2(q_x \frac{l}{2})} = \cot^2(q_x \frac{l}{2}) \quad (\text{III.15})$$

Et dans le cas LWR :

$$\frac{I_0}{I_{\pm 1}} \sim \frac{\sin^2(q_x \frac{CD}{2}) \times \cos^2(q_x \frac{l}{2})}{\cos^2(q_x \frac{CD}{2}) \times \sin^2(q_x \frac{l}{2})} = \tan^2(q_x \frac{l}{2}) \times \cot^2(q_x \frac{l}{2}) \quad (\text{III.16})$$

Un ajustement sur des données réelles, grâce à ce dernier modèle est donné en Figure III.34. On observe une tendance correcte, mais avec une erreur qui croît avec l’ordre des pics de diffraction, ceci est en partie imputable au bruit de comptage. Dans cet exemple, on extrait une l’amplitude de rugosité de 2.1 nm à ± 0.7 nm ce qui est proche de la valeur nominale de 2 nm, cependant il faut noter que ce type d’approche ne reflète pas la totalité de la rugosité présente, mais seulement l’amplitude pour une seule fréquence, à savoir celle de la structure selon y .

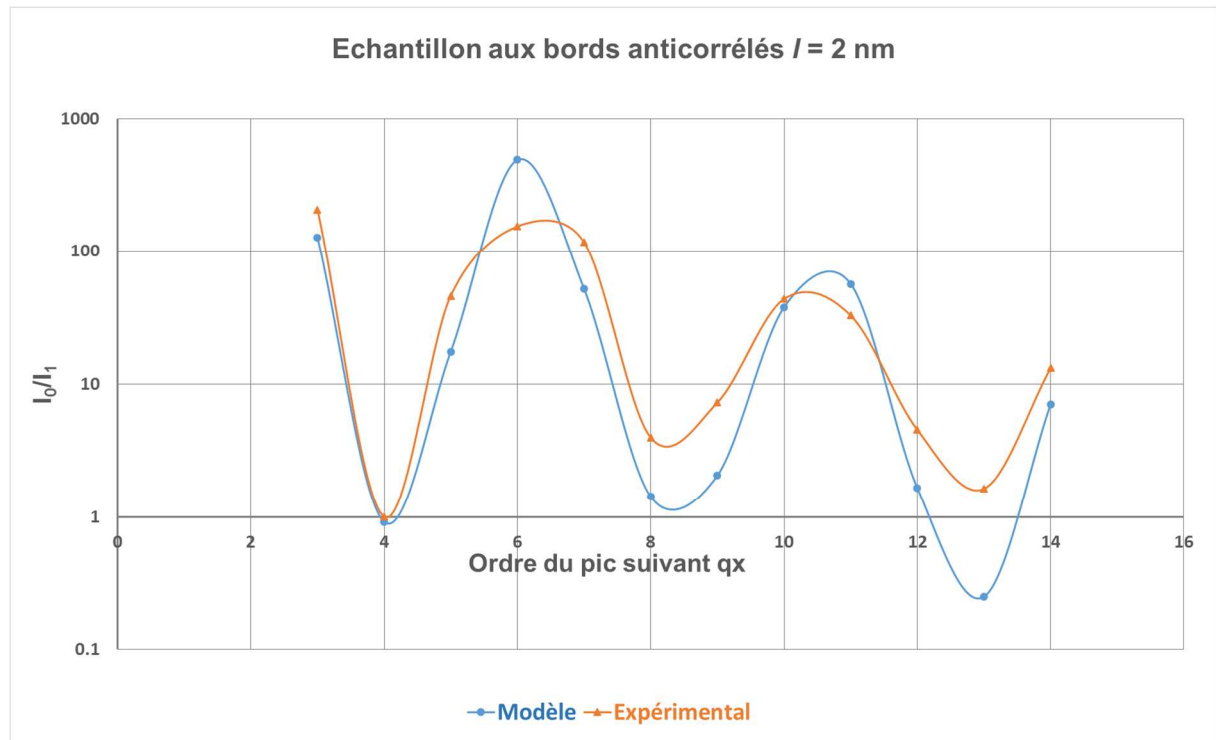


Figure III.34 : Extraction ponctuel du ratio I_0 / I_1 dans le cas d'un échantillon étalon avec une rugosité périodique de bords anticorrélés d'amplitude $l = 2$ nm (en orange), voir création Chapitre IV, et ajustement avec avec le modèle Equation III.16 (en bleu). Données acquises sur la ligne BM32 à l'ESRF, avec une énergie de 17keV à plus de 3 mètres (pixel du détecteur de 55 μ m de large)

Méthode 2 : Facteur de Debye-Waller

Dans le cas où les rugosités ne sont plus corrélées ou anti-corrélées, les pics de Bragg selon q_y ne sont plus observables. On peut alors se servir de la décroissance des pics de Bragg selon q_x pour extraire une amplitude de rugosité [103, 104]. Ces études ont montré que la rugosité peut être extraite comme une atténuation de l'intensité des taches de Bragg, suivant un facteur de Debye-Waller (DW) qui dans notre cas s'exprime ainsi :

$$I_{mesuré} = I_{ideal} \times \exp(q_x^2 \sigma_{DW}^2) \quad (III.17)$$

avec I_{ideal} l'intensité obtenue sans rugosité, σ_{DW} le paramètre de rugosité (standard déviation) dans ce cas aussi appelé facteur de DW. On retrouve ce résultat d'atténuation des pics, présenté Figure III.35, en simulant la diffusion d'échantillons virtuels ayant une rugosité de différentes amplitudes dans le cas de bords corrélés. On observe que cela correspond avec l'Equation III.13.

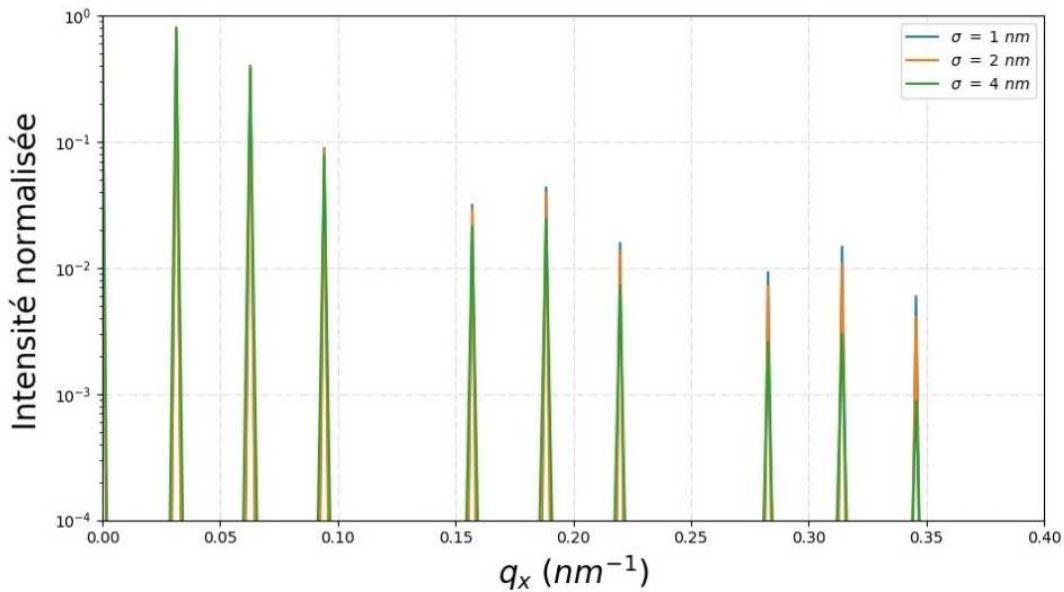


Figure III.35 : Signal de diffraction obtenu en $q_y = 0$, pour des réseaux : pitch 200 nm, CD 50 nm avec diverses amplitudes de rugosité de type LER (bords corrélés).

Dans cette approche particulière, nous avons voulu tester quel type de rugosité (LER, LWR) était mesuré au travers du facteur de DW. Nous avons alors généré des lignes avec σ_{LWR} variant en conservant σ_{LER} constant, la rugosité LWR induisant forcément de la rugosité LER comme mentionné Chapitre II avec l'Equation II.4, mais aussi des lignes à σ_{LER} variant ($\sigma_{LWR} = 0$). Pour effectuer cela nous avons additionné des générations de bords anti-corrélés à celles de bords corrélés. Comme le montre la Figure III.36, pour l'ordre 10, nous avons pu mettre en évidence que seule la composante σ_{LER} a un impact sur l'intensité des pics de diffraction. Autrement dit, le fait que les bords soient corrélés ou non ne change pas le facteur de DW que l'on peut extraire, ainsi σ_{DW} correspond finalement à σ_{LER} .

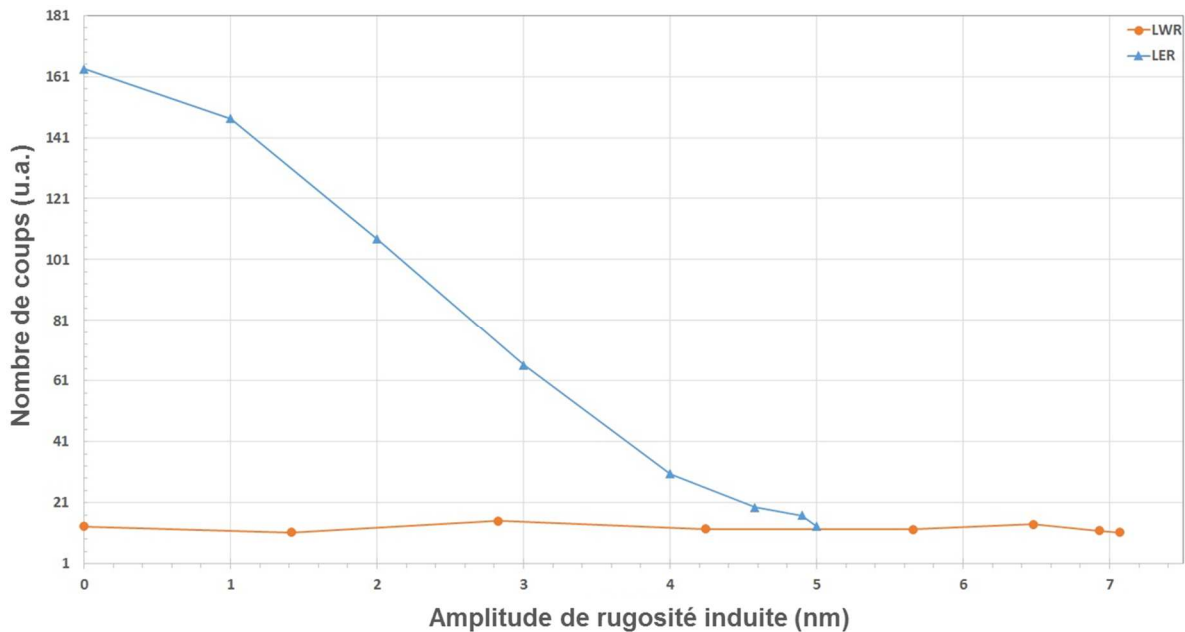


Figure III.36 : Nombre de coups obtenu sur l'ordre 10 pour des variations d'amplitude de rugosité LER ou LWR

Méthode 3 : Extraction PSD

Les simulations 2D de la figure de diffusion, comme celles visibles en Figure III.36, nous montrent que dans le cas de réseaux de lignes rugueux (b, c et d), un signal non nul est présent en dehors de $q_y = 0$, contrairement au cas sans rugosité (a). Pour une quantité identique de matière (nombre de pixels à 1 dans le cas de simulation), la totalité de l'intensité diffusée étant identique, on comprend que l'intensité non présente sur les pics de Bragg est répartie sur l'espace réciproque entier. De plus, si l'on compare les clichés de diffusion visibles en Figure III.36, de lignes aux bords anti-corrélés (b) avec celui des lignes aux bords corrélés (c), on remarque une répartition de la diffusion différente. En effet, des zones (q_x, q_y) distinctes, où le signal est faible voir nul, apparaissent. Plus exactement, dans le cas de lignes aux bords corrélés (c) on observe, aux positions q_x correspondantes à l'extinction du facteur de forme, que l'intensité est nulle pour toute valeur de q_y (idem en $q_x = 0$ excepté pour $q_y=0$). Ce phénomène n'est pas visible sur des bords anti-corrélés (b), ni même dans le cas où les bords de ligne ne sont pas corrélés (d). Des observations similaires ont été faites dans l'étude de A. Fernandez Herrero [105] ; cependant le cas spécifique d'un CD égale au demi-pitch induit l'extinction, de par le facteur de forme, d'un pic de Bragg sur deux, cette extinction complexifie la décorrélation des divers phénomènes en présence. De plus, ces travaux utilisent des réseaux dont la rugosité est formée à partir de blocs sur la longueur de la ligne de 100 nm de long. Cela induit malgré un placement aléatoire, une structure périodique se rapprochant donc d'un résultat de rugosité périodique. Ainsi, sur les résultats présentés par A. Fernandez Herrero on observe, non pas des pics de diffraction (dus à une rugosité périodique) mais des taches diffuses. Dans notre cas, l'utilisation sur les échantillons virtuels d'un échantillonnage (taille des blocs verticaux) $\Delta y = 1$ nm permet de ne plus avoir cet aspect de structure selon y et donc d'observer les intensités diffusées suivant cet axe.

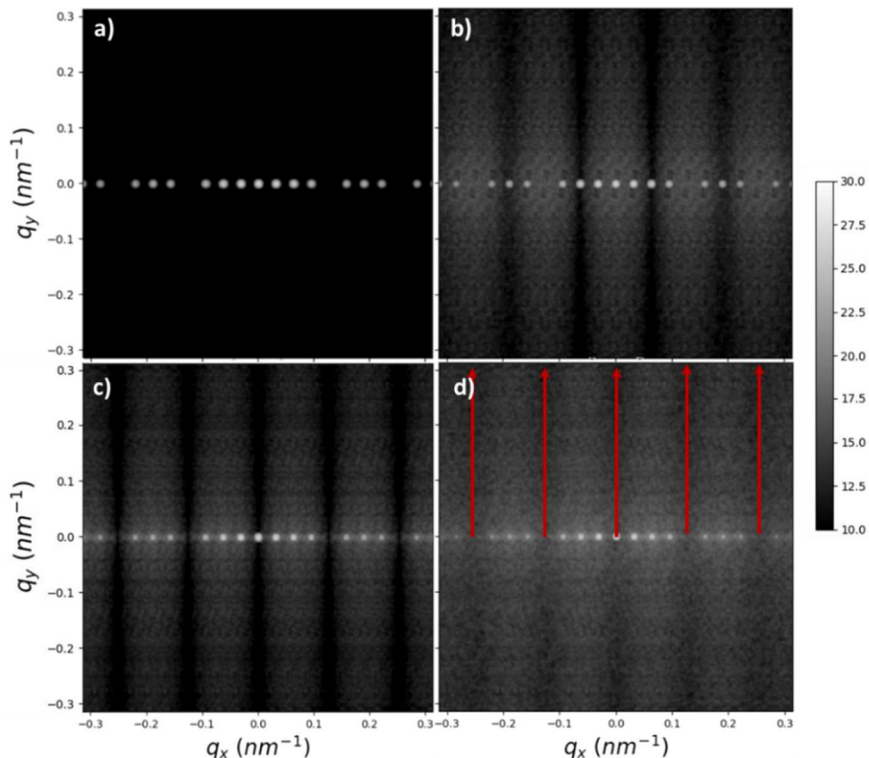


Figure III.37 : Clichés de diffusion obtenus par simulation sur un réseau de 10 lignes CD = 50 nm, pitch = 200 nm d'une longueur de 2 µm dans les cas : a) lignes parfaites (sans rugosité), b) avec une rugosité de bords anti-corrélés, c) avec une rugosité de bords corrélés, d) avec une rugosité de bords non-corrélés . Pour tous ces cas le paramétrage est $\sigma = 2$ nm, $\alpha = 0.5$, $\xi = 20$ nm.

Au vu des remarques précédentes, dans le cas de bords corrélés, la rugosité de type LWR est nulle (voir Chapitre II) parallèlement à un signal nul au point d'extinction. Nous avons donc émis l'hypothèse que le signal présent en ces positions correspond uniquement à la rugosité du type LWR sur les autres clichés. Le modèle de calcul de la PSD d'après le théorème de Wiener–Khinchin (cf. Chapitre II) et celui de la diffraction énoncé précédemment sont étroitement reliés par la TF. Ainsi, les intensités obtenues aux positions d'extinction du facteur de forme (à $q_y = 0 \text{ nm}^{-1}$), sur le cliché de diffusion, sont proportionnelles à la PSD LWR. En effectuant des coupes, comme visible sur la Figure III.34.d par les flèches rouges orientées suivant l'augmentation du nombre d'onde k_n , il est possible de représenter la PSD comme sur la Figure III.37.

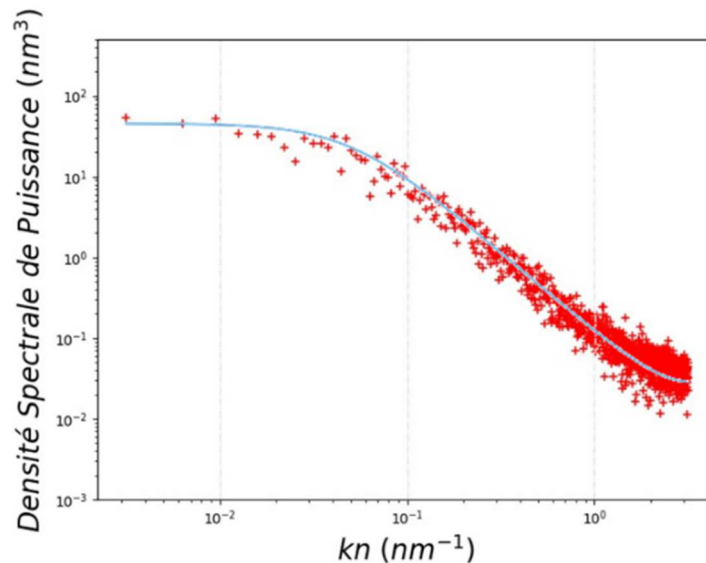


Figure III.38 : PSD moyenne extraite en un point d'extinction (en rouge) sur la simulation par TF d'une diffraction de 10 réseaux comportant chacun 10 lignes de $2\mu\text{m}$ de long, PSD de paramètres d'entrée (en bleu) $\sigma = 2.5 \text{ nm}$, $\alpha = 0.5$, $\xi = 20 \text{ nm}$, $\Delta y = 1 \text{ nm}$.

Dans le cas de simulations, les niveaux de puissance réels de la PSD sont accessibles en normalisant par le nombre d'objets qui diffusent (dans notre cas équivalent au nombre de pixels à 1 de l'image représentant notre réseau de lignes). Divers tests ont alors été réalisés en induisant des rugosités avec des bords de lignes plus ou moins corrélés, comme les cas extrêmes présentés dans la Figure III.36 b, c, d où la corrélation est respectivement de -1, 1 et 0. Ces tests ont permis de vérifier l'hypothèse de l'obtention de la PSD LWR seule, lors de l'extraction au niveau des extinctions du facteur de forme. Ce point est aujourd'hui protégé par un brevet. De plus, une étude sur des échantillons ayant des paramètres de rugosité de lignes LWR différents a montré la faisabilité d'extraction sur des clichés réels de diffusion [106].

Par ailleurs, on remarque que la simulation d'une PSD, obtenue comme le résultat d'une diffusion d'un réseau de N lignes, n'est pas identique à la moyenne des PSD obtenues sur chaque ligne séparément. En générant plusieurs diffractions de réseaux (tels plusieurs clichés) on retrouve l'effet de la statistique avec une PSD lissée, comme visible sur la Figure III.37 avec 10 générations, soit l'équivalent de 10 lignes et non de 100. Sur un cliché réel de diffusion, Figure III.38, obtenu par la moyenne de 150 clichés de 10 sec, et soustrait de la moyenne de 14 clichés de 10 sec de référence sans structure (soit près de 30 min d'acquisition), il est possible d'extraire la PSD LWR suivant le premier point d'extinction du facteur de forme, comme explicité précédemment (des artefacts dus au setup empêchant l'extraction en $q_x = 0 \text{ nm}^{-1}$). On obtient alors la Figure III.39. Dans ce cas d'acquisition réelle, le faisceau illumine 50×50

μm^2 soit environ 500 lignes sur une longueur de $50 \mu\text{m}$, ce qui est bien plus que la simulation et permet d'obtenir une PSD plus lisse. Cette taille est donc un atout, mais c'est aussi une limitation, comme expliqué dans la suite de ce chapitre.

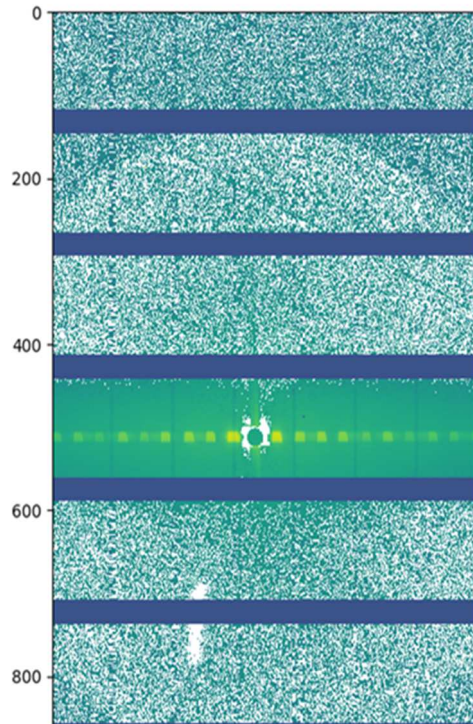


Figure III.39 : Cliché de diffusion, suivant les pixels du détecteur, obtenu par moyenne de 150 clichés de diffusion de notre échantillon NIL et soustraction de la moyenne de 14 clichés de silicium sans structure. Cliché obtenu sur la ligne BM02 de l'ESRF, à 10 keV et une distance de plus de 8 m (pixel camera de $172 \mu\text{m}$ de large)

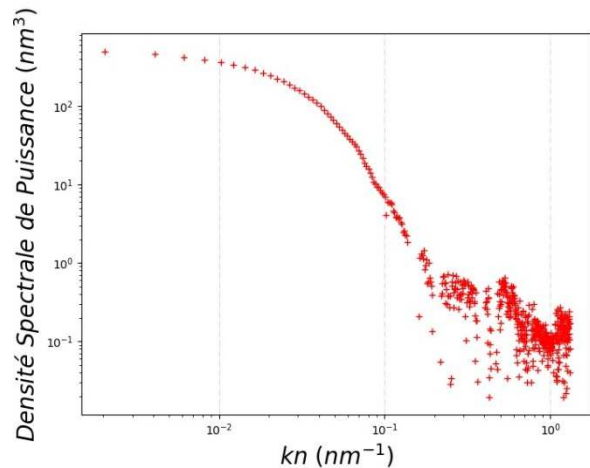


Figure III.40 : Extraction PSD obtenue sur le cliché de diffusion Figure III.38, d'un échantillon gravé (lithographie NIL), pitch nominal = 100 nm et CD nominal= 50 nm. Résultat recalé en puissance (à partir d'un résultat CD-SEM).

III.4.2 Limitations

La principale limitation a déjà été évoquée, et elle est instrumentale. En fait, la qualité du signal et la vitesse d'acquisition sont essentiellement pilotées par la quantité de photons diffusés par l'échantillon et donc par la brillance (définie ici comme le nombre de photons /s / mm^2 en supposant que la qualité spectrale et la divergence du faisceau sont imposées par la résolution de la mesure) du faisceau incident. Ceci signifie qu'un faisceau plus grand permet d'obtenir plus de photons et donc plus de signal. En somme, il faut donc éclairer un maximum de la structure avec un maximum de photons, sans dépasser toutefois ses dimensions car cela ajouterait du bruit et, potentiellement, la signature de motifs annexes dans le cliché de diffusion. Cette limitation est particulièrement vraie dans le cas de la technique GiSAXS. En effet, cette technique se positionne sur des angles d'incidence généralement proches de 1° , ce qui induit que l'empreinte du faisceau (longueur éclairée) est au moins 60 fois plus grande que sa taille. Cela se traduit à la fois par la perte de beaucoup de photons (près de 2 ordres de grandeur dans l'exemple présent) et par un niveau de bruit très élevé.

Dans la même idée, et comme déjà évoqué, l'absorption des rayons-X par le substrat est une des problématiques aujourd'hui présentes dans le cadre de l'application de la méthodologie en transmission pour la microélectronique. En effet, les substrats de silicium ont une épaisseur standard entre 700 et 800 μm (suivant le diamètre utilisé et l'empilement considéré), ce qui induit comme le

montre la Figure III.40 (pour 750 μm d'épaisseur de substrat de silicium), qu'en dessous d'une énergie de 10 keV, moins de 0.5% des photons traversent le substrat. Ce phénomène n'est pas linéaire avec l'épaisseur, ainsi la diminution de l'épaisseur d'un tiers (soit un substrat de 500 μm) induit encore une absorption de plus de 97% pour la même énergie et n'est donc pas une solution envisageable. On peut cependant voir que le passage de 15 à 18 keV permet, dans le cas de substrats standards, de passer de 20% à près de 40% de transmission.

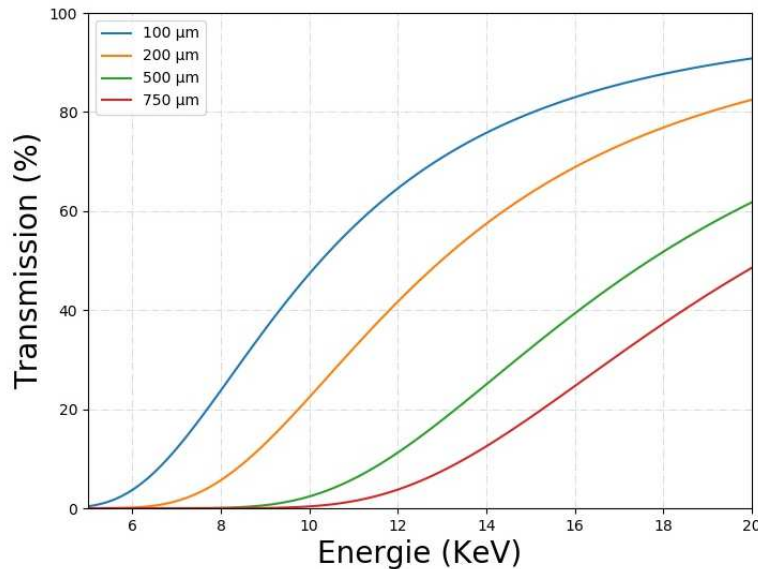


Figure III.41 : Transmission obtenue pour des substrats de silicium de différentes épaisseurs, à différentes énergies de faisceau incident.

Pour obtenir un nombre de photons plus important sur le cliché, il est aussi possible d'utiliser des temps d'acquisition plus longs. Il faut souligner le fait que ceci est limité, notamment par le courant d'obscurité (signal émis par le détecteur même lorsqu'il ne reçoit aucun photon) qui ne se réduit pas lorsque plusieurs clichés sont moyennés par exemple. Dans le cas de l'extraction Figure III.39, nous supposons que c'est cette limite qui est atteinte dans les hautes fréquences ($>0.3 \text{ nm}^{-1}$), avec un bruit ayant un niveau plus important que le signal.

Une deuxième limitation peut être évoquée dans la résolution de problème inverse, de manière identique à l'OCD. Cependant nous avons vu que dans le cas de l'analyse de la rugosité au travers de la PSD LWR, l'obtention se fait de manière directe sur le cliché de diffraction ; dans ce cas précis cette limite est donc contournée.

Par ailleurs, bien qu'il soit aisé sur des simulations de normaliser l'intensité par le nombre de lignes considérées dans la diffraction, ceci est moins simple dans le cas réel. En effet, pour remonter à la valeur de σ_{LWR} lors de la diffraction, il faut à la fois connaître la taille du faisceau incident et son flux, afin d'estimer le nombre de photons ayant éclairé le réseau de lignes et en quelle proportion. A cela s'ajoute la nécessité de déterminer la morphologie complète des structures (Pitch, CD, SWA, hauteur) afin de pouvoir déterminer la quantité de matière interagissant et prendre en compte sa densité électronique. Dans la suite ces travaux, nous nous sommes donc limités aux informations α et ξ dans le cadre de l'extraction de la rugosité par la technique SAXS en transmission.

En termes de métrologie, on peut souligner que la résolution obtenue en SAXS est liée à la fois à la capacité de résolution du détecteur (taille des pixels), mais surtout à la distance entre ce dernier et l'échantillon. La diffusion X mesurant des angles (2θ), plus la distance entre l'objet et le détecteur sera

grande, plus un angle donné représentera une distance grande sur le détecteur. Appliquée à la PSD, pour une distance donnée, la taille du détecteur limitera donc $k_{n,max}$ et fixera aussi l'échantillonnage en fréquence. Il est ainsi nécessaire de trouver un compromis entre la fréquence à laquelle on veut accéder et l'échantillonnage. Il peut aussi être envisagé d'effectuer des mesures à des distances différentes, pour obtenir le résultat sur diverses plages de fréquences, sans être limité par l'échantillonnage, mais bien sûr au détriment du temps total d'acquisition.

La qualité des clichés de diffraction obtenu en SAXS est reliée au nombre de photons diffusés et récoltés.

III.5 Comparaison expérimentale des techniques

Afin de comparer les différentes techniques de mesure de rugosité, un jeu d'échantillons pouvant être caractérisé sur les divers équipements disponibles a été réalisé. Ce jeu d'échantillons doit répondre aux contraintes de chaque technique, par exemple à la prise en compte des contraintes de dimensionnement annexes, comme la hauteur des structures ou encore les matériaux utilisés. La création d'échantillon est détaillée dans le chapitre suivant. On retiendra ici que la rugosité utilisée est périodique et que diverses corrélations de bords ont été testées.

De plus pour pouvoir comparer les techniques, il faut connaître la gamme en fréquence mesurable par chacune d'elle pour définir une périodicité de rugosité sur notre échantillon accessible à toutes ces techniques.

III.5.1 Gamme de travail en fréquences

Les gammes de fréquences auxquelles chaque technique peut accéder sont représentées en Figure III.42. Elles correspondent à la physique de chaque technique et donnent un ordre de grandeur.

Accessibilité aux faibles fréquences : On peut voir que l'OCD et l'AFM permettent d'obtenir une information sur les très basses fréquences, soit des distances importantes jusqu'à des dizaines de microns, du fait respectivement de la longueur d'onde utilisée en OCD et du déplacement possible sur une longue distance en AFM. A l'inverse, le CD-SEM subit une limitation sur les basses fréquences due aux aberrations des déflexions électroniques, ainsi il ne sera pas pertinent pour une période supérieure à quelques micromètres. Dans le cas du SAXS, des objets ayant une dimension caractéristique supérieure à quelques centaines de nanomètres sont difficilement mesurables, l'information de diffusion se retrouvant aux très petits angles de diffusion et donc très proche du faisceau direct ou du beamstop.

Accessibilité aux grandes fréquences : La technique SAXS, au vu des énergies utilisées et donc des longueurs d'ondes courtes (18 keV \rightarrow 0.069 nm), peut être poussée jusqu'à une échelle sub-nanométrique. Le CD-SEM et l'AFM sont limités par le théorème de Shannon, i.e. par la taille du faisceau ou la taille de la pointe. Enfin dans notre cas, l'OCD travaillant jusqu'à une longueur d'onde de 235 nm, pourra caractériser une réponse correspondant à une fréquence de quelques dizaines de nanomètres.

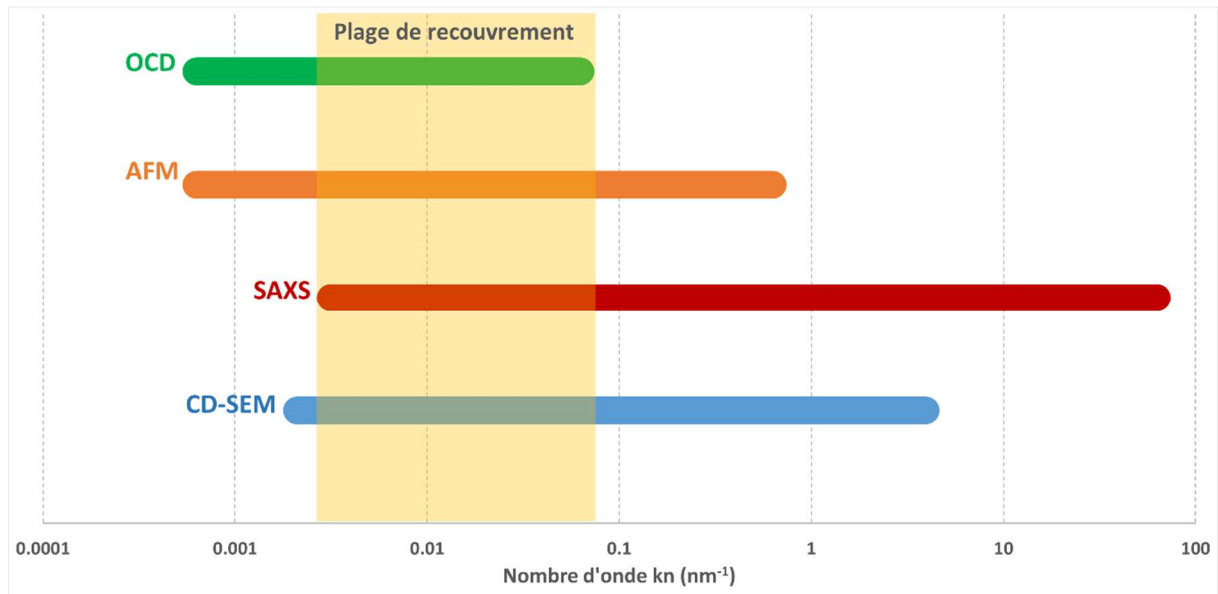


Figure III.42 : Plage de fréquence spatiale accessible pour chacune des techniques utilisées.

A partir de ces plages de fréquence spatiale, il est possible de voir la complémentarité des différents équipements, afin d'avoir une caractérisation de la rugosité sur une plage de fréquences la plus grande possible. *A contrario*, on peut déterminer comme visible sur la Figure III.42, une plage de recouvrement où tous les équipements sont capables de détecter une rugosité de ligne si celle-ci à une amplitude suffisante. C'est sur cette dernière plage que la comparaison des techniques va pouvoir se faire afin de tester la capacité de résolution pour l'extraction de la rugosité.

Nous avons défini, compte tenu de ces paramètres, un jeu d'échantillons constitués de réseaux de lignes avec un pitch de 200 nm et un CD nominal visé de 100 nm (le CD réel étant plus proche de 150 nm). Afin d'observer les capacités réelles des équipements, ce jeu comporte une rugosité périodique dont l'amplitude va, suivant les échantillons, de 1 à 16 nm. Un exemple d'échantillon obtenu par lithographie e-beam puis gravé est présenté en Figure III.43, sur l'image CD-SEM en vue du dessus.

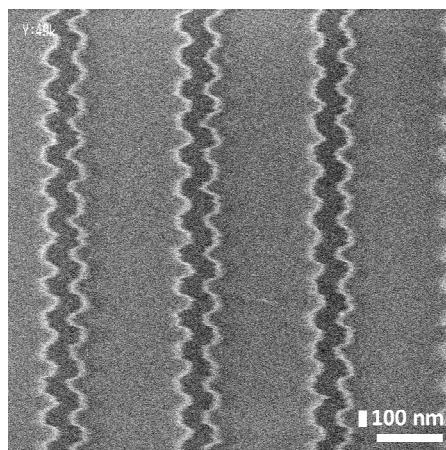


Figure III.43 : Image CD-SEM rectangulaire (HCG4000) d'un échantillon aux bords corrélés ayant une amplitude de rugosité de ligne égale à 16 nm (les espaces étant représentés en noir et les lignes en gris clair).

Pour compléter cette étude, des échantillons de référence, sans rugosité programmée, ont été mesurés pour voir les différences obtenues sur les extractions de PSD.

III.5.2 Capacité d'extraction de la rugosité

Lors de cette étude, l'extraction de la rugosité par la technique SAXS a été effectuée grâce à la méthode analytique (méthode 1) présentée précédemment. Une des extractions est visible Figure III.34, la méthode PSD n'ayant pas encore été mise au point à ce moment de l'étude. Dans le cadre du CD-SEM, l'Equation II.24 de Parseval a été utilisée suite à l'extraction des PSD, de par la spécificité de l'échantillon périodique qui ne suit pas le modèle PSD (pic en fréquence). Le résultat des extractions, par ces deux techniques, est présenté Figure III.44. On remarque, dans le cas du SAXS, que la pente des régressions est proche de 1 et l'ordonnée à l'origine de 0 ce qui montre une bonne extraction de cette rugosité périodique. Dans le cadre de la campagne de mesure SAXS, l'échantillon de rugosité normale de 1 nm n'a pas été mesuré faute de temps, cependant lors d'une seconde campagne du même type, le passage de cet échantillon a permis de vérifier la présence de pic de Bragg selon l'axe y, et donc la possibilité de détecter cette amplitude avec la technique SAXS. Pour le CD-SEM, on note un décalage important entre la mesure et la valeur nominale, l'ordonnée à l'origine étant à près de 5 nm. Cependant, malgré le fait que la pente de la courbe de tendance ne soit pas à 1, on remarque que la tendance linéaire correspond bien avec un $R^2 = 0.9962$.

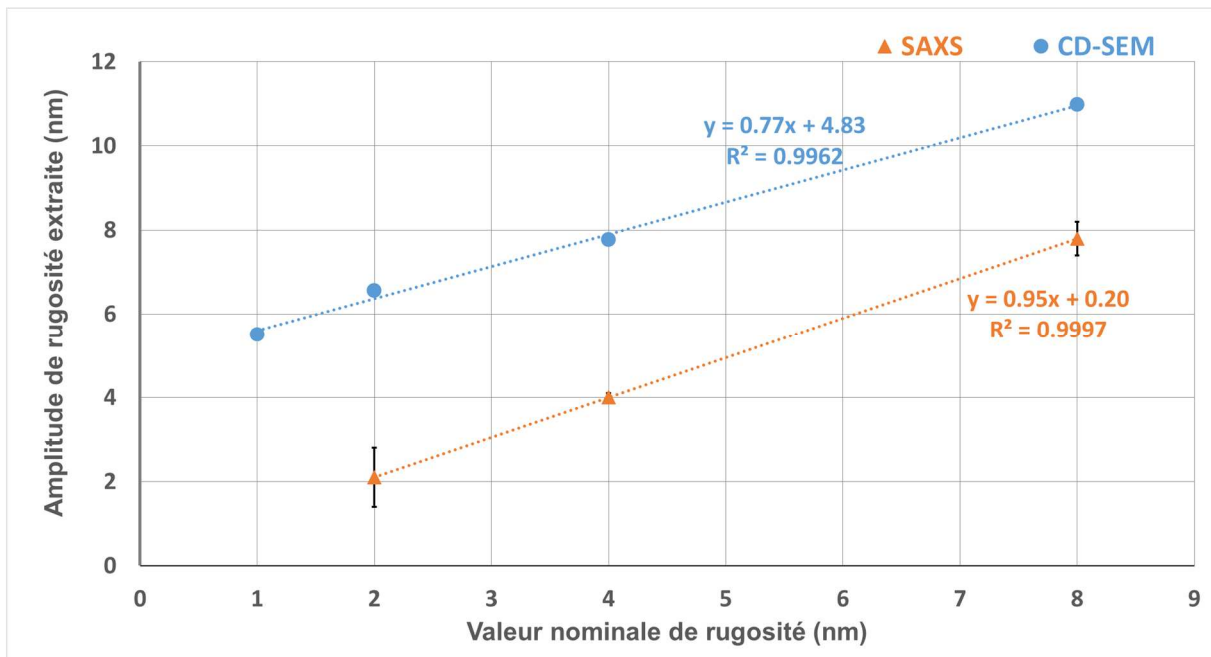


Figure III.44 : Extraction d'amplitude de rugosité de ligne LWR obtenue pour les techniques CD-SEM et SAXS, en fonction de la valeur nominale de rugosité induite. Les incertitudes SAXS sont issues du fit, nécessaire à la méthode d'extraction analytique.

Par ailleurs, la Figure III.45 montre le résultat de l'extraction avec l'OCD grâce à la méthodologie EMA, soit une valeur de CD_{EMA} en fonction de la valeur nominale de rugosité. Le paramétrage EMA est le même que cité précédemment à savoir un facteur de dépolarisation de 1/3 et une fraction d'air dans le matériau de 50%. On remarque que malgré la détection de la rugosité d'amplitude la plus faible, 1 nm, la réponse du modèle n'est pas linéaire avec la variation de la rugosité. Ceci entraîne qu'une rugosité en deçà de cette limite de 1 nm risque d'être encore moins détectable. Pour obtenir la valeur exacte de la rugosité, une calibration est nécessaire, cela peut passer par le changement du matériau EMA, notamment la fraction air/matériau qui a été mise dans notre cas à 50/50.

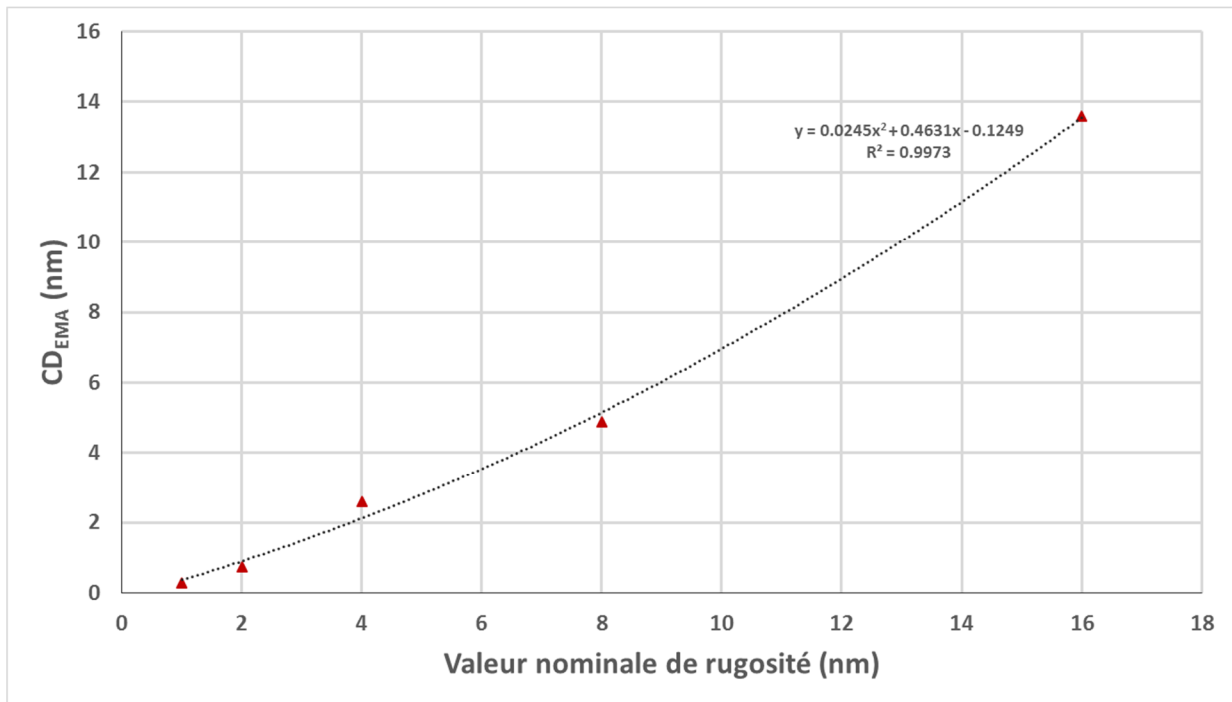


Figure III.45 : CD_{EMA} obtenue pour les réseaux ayant une rugosité périodique de différentes amplitudes nominales.

On retient finalement, comme visible en Figure III.46, que toutes les techniques exceptée l'AFM (cet outil ne disposant pas de pointe adaptée lors de l'étude) ont la capacité d'accéder aux amplitudes de rugosité jusqu'à 1 nm, même si un décalage important est visible sur l'extraction CD-SEM, nous y reviendrons dans le chapitre suivant. De plus, on peut voir que les méthodologies mises en place permettent aux techniques, hormis l'OCD, d'accéder aux informations fréquentielles puisqu'elles ont la capacité, dans cette étude, d'extraire la période de la rugosité et donc *a posteriori* d'extraire des PSD.

Technique	Rugosité mesurée	Données fréquentielles (PSD)
SEM	1 nm ✓	✓
AFM	Non traitée	✓
OCD	1 nm à calibrer ~	✗
SAXS	1 nm ✓	✓

Figure III.46 : Tableau récapitulatif des résultats obtenus pour les diverses techniques, concernant leur capacité à mesurer une amplitude de rugosité (la plus faible possible) et l'accès aux informations fréquentielles.

Pour aller plus loin dans cette comparaison de performances liées à la rugosité de ligne, la PSD LWR a été extraite sur un échantillon gravé, en silicium, issu d'une lithographie NIL grâce aux techniques SEM (méthodologie Verduin sur des images de $2 \times 2 \mu\text{m}^2$) et SAXS. Le résultat, présenté en Figure III.47, laisse voir que la limite de bruit obtenue en SAXS est plus faible que celle en SEM. Ce point permet une plage d'analyse en fréquences plus grande, notamment sur les hautes fréquences, jusqu'à $0.2 - 0.3 \text{ nm}^{-1}$ contre 0.1 nm^{-1} en SEM. Par ailleurs, concernant les basses fréquences, la limite obtenue en SAXS est plus éloignée que celle en SEM. Cependant, la taille maximale de l'image SEM n'a pas été utilisée, cela mènerait à l'accession d'une fréquence identique au SAXS, soit l'analyse d'une longueur de $3 \mu\text{m}$ de long. On rappelle néanmoins que seul la PSD LWR sans l'information de σ est extraite en SAXS, ce qui est une limitation non négligeable dans la considération de la rugosité de ligne. Outre les performances

vis-à-vis de l'extraction de la rugosité, chacune des techniques utilisées se différencie suivant des critères liés à l'industrie.

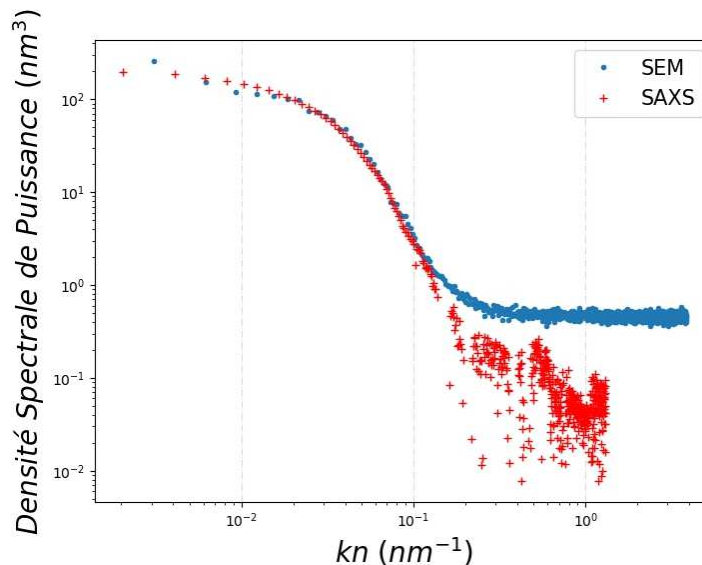


Figure III.47 : PSD LWR extraite dans le cadre des deux techniques SEM et SAXS sur un réseau de lignes gravé, issu d'une lithographie NIL. La PSD SAXS ayant été recalée en amplitude sur celle du SEM.

III.5.3 Maturité de développement pour une métrologie en ligne

Parfois, des techniques peuvent être mises de côté par les industriels qui les jugent comme des techniques de laboratoire, non adaptées à leurs contraintes. Cette position s'explique vis-à-vis des performances en termes de rendement, et donc d'efficacité et de coût. Pour exemple, dans le cas de la mesure CD, le comparatif fait par B. Bunday en 2017 [107] et visible en Figure III.48, montre pour défaut le temps pour une mesure SAXS par transmission.

Technique	Nœud technologique accessible (nm)	Temps	Statistique	Destructif
CD-SEM	7	=	+	=
AFM	10	-	+	+
OCD	7	+	=	+
SAXS	< 5	--	=	+

Figure III.48 : Tableau comparatif des techniques pour l'extraction de CD, extrait des travaux de B. Bunday [107]. Les divers signes traduits de la publication représentent : + une qualité de la technique, - un défaut, -- un défaut prononcé et = un cas neutre.

Afin d'observer la maturité des techniques utilisées dans le cadre de la mesure de rugosité, ainsi que leur possible utilisation en industrie, nous avons sélectionné, à l'instar de B. Bunday, divers critères et les avons évalués. D'un point de vue général, le résultat obtenu Figure III.49 ne permet pas de dire qu'une technique est préférable à une autre. Cependant, contrairement au résultat pour le CD présenté Figure III.48, on distingue la vitesse d'acquisition et celle de traitement, ce qui est très important notamment pour les techniques OCD et SAXS qui sont rapides en acquisition, ce qui n'apparaît pas forcément dans la littérature. Par ailleurs, on remarque que dans le cas de la rugosité, la statistique de mesure est bien meilleure avec ces deux dernières techniques qu'avec le CD-SEM ou l'AFM, contrairement au résultat montré Figure III.48 pour le CD.

Technique	Vitesse d'acquisition	Statistique	Vitesse de traitement	Déployée	Limite théorique
CD-SEM	= (> minute)	– 500 lignes	=	+	Taille de faisceau ~1 nm
AFM	- (> heure)	1 – 5 lignes	=	=	Taille de la pointe ~3 nm
OCD	++ (~ seconds)	> 1 000 lignes	-	+	Å
SAXS	+ (< minute)	> 1 000 lignes	-	-	< Å

Figure III.49 : Tableau comparatif des techniques pour l'extraction de rugosité sur des critères liés à l'industrie et aux besoins futurs. Les divers signes représentent : + une qualité de la technique, – un défaut et = un cas neutre.

Plus particulièrement, pour l'extraction de rugosité, on retrouve en Figure III.50 que les techniques historiques telles que le SEM et l'AFM ont une vitesse d'acquisition de quelques minutes à quelques heures, avec un nombre restreint de lignes. Ces techniques, bien que limitées en résolution à une échelle nanométrique, profitent du fait d'être déjà très présentes dans le secteur de la microélectronique. Par là même, les traitements liés ont été développés depuis de nombreuses années et sont aujourd'hui, bien que souvent lents, automatisés et robustes. Ces techniques restent encore les références pour l'extraction de la rugosité de ligne.

A l'opposé, la technique SAXS par transmission est actuellement restreinte à des caractérisations en laboratoire ou sur synchrotron. Cependant cette technique permet, en moins d'une minute, d'obtenir une statistique sur un très grand nombre de lignes et avec une résolution bien meilleure que toutes les autres techniques présentées ici. La technique SAXS souffre d'un manque de développement de fond qui induit des temps de traitement importants, également lié à la résolution de problème inverse. Cette résolution reste nécessaire malgré une analyse directe par la PSD, afin de déterminer les positions exactes des points d'extinction du facteur de forme (liées au CD et au pitch).

Dans le même cas, la technique OCD à une faible vitesse de traitement, néanmoins cette technique a su, grâce à sa vitesse d'acquisition et sa résolution, s'imposer dans le secteur de la microélectronique. Ainsi, on retrouve divers équipements de scatterométrie, aujourd'hui déployés dans l'industrie, principalement sur des applications de contrôle de variation morphologique.

III.6 Résumé du chapitre

Dans ce chapitre ont été présentées les diverses techniques mises en place pour extraire la rugosité de ligne : SEM, AFM, OCD et SAXS. Un point sur les limites de chaque technique a été fait.

On retiendra que le SEM permet d'obtenir la totalité des paramètres de rugosité par la PSD, pour les types LER et LWR. L'AFM ne permet l'extraction des paramètres, suivant la PSD, que pour la rugosité LWR. La technique SAXS permet d'accéder à α et ξ pour la LWR et σ dans le cas de la LER, au travers du facteur de Debye-Waller. Enfin, la technique OCD accède, par une analyse EMA, au paramètre σ mais ne permet pas de différencier les deux types de rugosité (LER et LWR), ni même en l'état, de donner d'autres informations sur la rugosité.

Par ailleurs, un point sur les capacités des techniques à être utilisées dans l'industrie, ou pour une métrologie en ligne, a été fait. Il en ressort qu'aucune des techniques ne permet, seule, de répondre à la problématique de mesure de rugosité de ligne sub-nanométrique. Finalement, chaque technique ayant des capacités ainsi que des plages d'accès en fréquence différentes, on retrouve ici l'intérêt de faire de l'hybridation de données.

Chapitre IV - Développement d'étalons de rugosité

Dans le cadre des concepts de la métrologie (Chapitre II), nous avons vu que les étalons permettent de calibrer les outils de mesure et ainsi, de garantir la justesse des mesures faites par la suite. Il existe aujourd'hui des étalons pour certaines métriques tels que le CD, ou encore la hauteur des motifs.

Pour la rugosité de ligne il n'existe pas d'étalon actuellement, ce qui, en plus d'entraîner une dérive des mesures et des méthodes [35], limite une partie des études sur la rugosité. Ces dernières correspondent aux développements ou validations, (i) de procédés de fabrication engendrant moins de rugosité de ligne, (ii) de mise en place de nouvelles techniques ou méthodologies de mesure. Ainsi, les recherches effectuées sur la rugosité de ligne se basent, pour la plupart, soit sur des motifs comportant une rugosité périodique ou assimilable [82, 100, 105, 108], soit sur un procédé qui est, par la suite, dégradé par le changement des paramètres lithographiques (dose d'insolation, focus, ...) ou de gravure (pourcentage des espèces chimiques, aspect temporel, ...) afin de générer plus de rugosité [44]. On voit que ces méthodes ne permettent pas, dans le cas de rugosité périodique, d'obtenir une rugosité représentative de celle, réelle, pouvant apparaître et, dans le cas de dégradation d'un procédé, de balayer toute la gamme de paramétrage de la rugosité (voir Chapitre II).

Pour répondre à cette problématique nous avons proposé, durant ces travaux, de développer des étalons de rugosité de ligne dits « virtuels », par simulation, mais également « réels » sur substrat de silicium.

IV.1 Rugosité périodique

Pour fabriquer ces étalons expérimentaux, nous avons premièrement généré des motifs ayant une rugosité périodique, ce cas est déjà connu mais permet de comprendre les concepts, les enjeux et les limites principales liés à la génération de rugosité. Dans un second temps, nous avons extrapolé ce principe pour générer des rugosités respectant la statistique d'une PSD. Comme nous l'avons vu, il y a différents types de rugosité (LER, LWR, bord corrélés...), ainsi il est important de générer différents types d'étalons.

IV.1.1 Les différents types

La Figure IV.1 présente les différents types d'étalons que nous avons générés. On retrouve les cas de rugosité vus précédemment avec les bords corrélés (a), ou anti-corrélés (b). A ces cas s'ajoute celui de l'asymétrie (c) qui permet d'obtenir un $LWR = LER_g$ (et $LER_d = 0$), ceci afin de tester à la fois la différence d'extraction pouvant avoir lieu sur le LWR et le LER, mais aussi de voir les capacités que peuvent avoir les outils d'analyse sur une différenciation des bords gauche et droit. Cela permet aussi d'avoir, sur un même motif, la programmation d'une rugosité d'un côté et d'une référence de l'autre (sans ajout de rugosité programmée).

Par ailleurs, sur cette même Figure IV.1, on voit apparaître les paramètres pris en compte pour la rugosité périodique, à savoir sa période λ et son amplitude l . Dans le cas anti-corrélé, la rugosité générée sur la métrique σ_{LWR} est égale à $2l$.

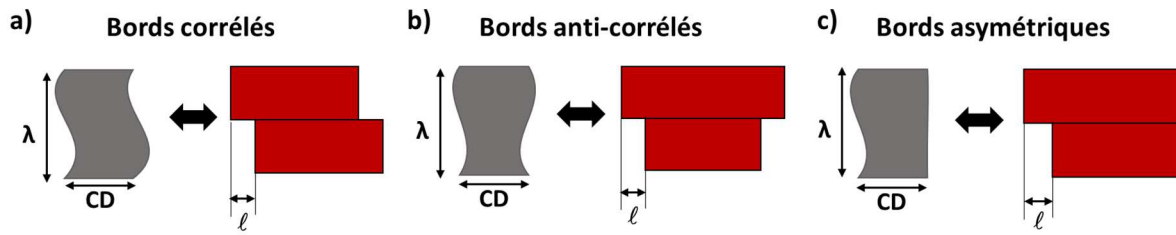


Figure IV.1 : Schématisation des différents types de rugosité, pour le cas de rugosité de période λ et d'une amplitude l . Dessins théoriques en rouge, résultats attendus sur le substrat en gris.

De plus, la Figure IV.1 montre que l'on souhaite générer un bord de ligne suivant une sinusoïde (en gris). Pour réaliser cela, les dessins informatiques générés sont deux rectangles (en rouge) pour une période. La transformation de ces rectangles (lissage) lors de la réalisation dépend des capacités résolventes de la lithographie (outils et procédé). En effet, comme l'illustre la Figure IV.2, suivant la hauteur des blocs ($\lambda/2$ dans ce cas), le résultat expérimental dans la résine (en bleu) n'est pas le même (représentation simplifiée issue de simulation de lithographie). Par exemple, dans le cas où les motifs ne peuvent pas être résolus, Figure IV.2.a, l'amplitude de rugosité est atténuée et ne correspond plus à celle visée. A l'inverse si la période est trop grande, Figure IV.2.c, on voit apparaître des zones « plates » où le bord de ligne ne suit plus une courbe sinusoïdale. La solution idéale serait d'être dans la condition b, forme sinusoïdale avec l'amplitude souhaitée, soit λ aux alentours de 5 nm. Cependant, comme il a été précisé dans le chapitre précédent, les équipements ne mesurent pas tous dans la même gamme de fréquences. Ainsi, on fera l'hypothèse d'être dans le cas c (forme non sinusoïdale) avec une fréquence $\lambda = 200$ nm qui est bien supérieure à la capacité de résolution des équipements de lithographie disponibles et des procédés associés. Cette fréquence est accessible par tous les équipements, comme présenté dans le Chapitre III. Cette forme spécifique, mais périodique, entraînera la création d'harmoniques, nous le verrons sur l'analyse PSD.

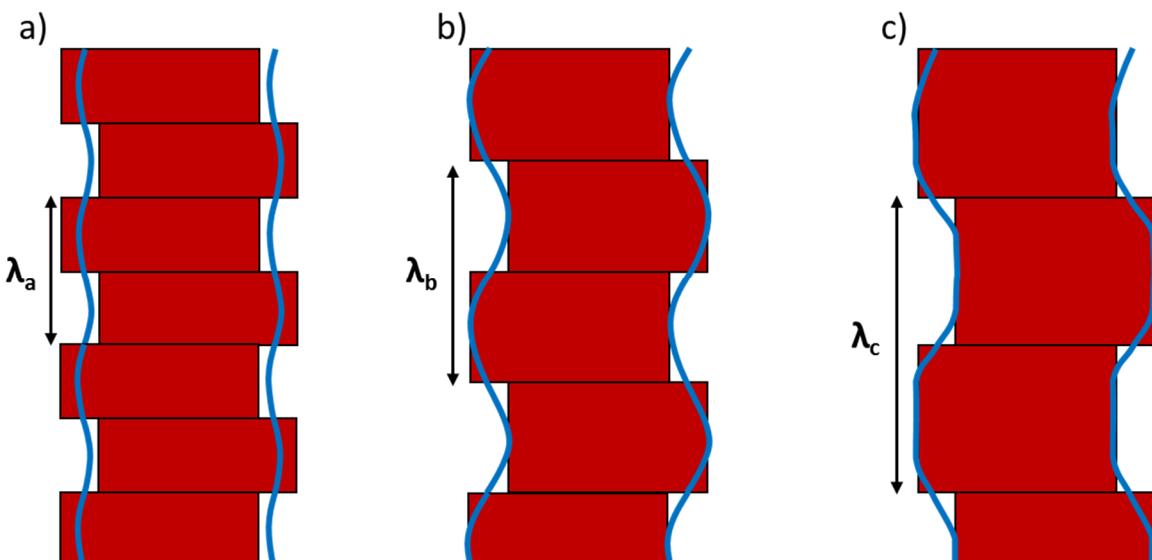


Figure IV.2 : Illustration schématique de l'évolution des bords de lignes lithographiées suivant la hauteur des blocs créant la ligne $\lambda/2$, pour des rugosités de bords corrélés, $\lambda_b = 1.5 * \lambda_a$ et $\lambda_c = 2 * \lambda_a$.

La plage de fréquence de fonctionnement des équipements et des techniques n'est pas la seule limitation à laquelle il faut répondre pour la création d'un étalon de rugosité.

IV.1.2 Contraintes multi-équipements

Il existe de nombreuses contraintes pour chacun des équipements et il est nécessaire, afin d'obtenir des étalons mesurables par toutes les techniques, de répondre à chacune d'entre elles. Cette section présente des travaux réalisés à une fréquence unique, ne sera donc pas fait de référence à la problématique de plage de fréquences, présentée précédemment.

La surface du motif

Dans le cas des techniques OCD et SAXS, les faisceaux de lumière utilisés ont une taille comprise entre une dizaine et une centaine de micromètres, en fonction des équipements et des angles d'incidence utilisés. Il est alors nécessaire, pour limiter le bruit de mesure, d'avoir un réseau de lignes s'étendant sur une surface au moins aussi grande que la surface illuminée. Par ailleurs, un plus grand réseau (surface patternée) permet de s'aligner plus rapidement, notamment en SAXS, car à la première approche d'alignement des motifs et de le positionnement du faisceau sur cette zone, se fait visuellement. Les autres techniques de mesure n'ont pas de contraintes sur ce paramètre, cependant plus le réseau sera grand, plus il sera long à exposer en lithographie e-beam. Nous avons fixé dans un premier temps la taille des réseaux à $250 \times 250 \mu\text{m}^2$, ce qui permet une durée d'écriture raisonnable par motif, inférieur à 15 min, avec cette périodicité $\lambda_0=200\text{nm}$.

Le pas du réseau (pitch)

Pour le SAXS et le SEM, on favorisera des pitches de petites dimensions ($\leq 200 \text{ nm}$). En effet en SAXS, comme déjà évoqué dans le Chapitre III, un pitch plus grand va se traduire par un signal de diffraction plus proche du faisceau direct ou du beamstop, rendant l'analyse des premiers ordres de diffraction difficile, voire impossible. Pour une image SEM d'une taille donnée, on peut remarquer que plus le pitch sera réduit, plus il y aura de lignes à mesurer sur une image. Un pitch plus petit permet alors de réduire le nombre d'images à prendre et donc le temps d'acquisition.

Pour l'AFM, on utilise une pointe physique qui doit pouvoir se déplacer entre les lignes. Les pointes de mesure CD les plus petites commercialisées sont les « CDR15 », avec une largeur théorique de $15 \pm 5 \text{ nm}$. De plus, l'amplitude nécessaire de recul de la pointe, généralement considérée, est de deux fois la taille de la pointe. Ainsi, la taille de l'espacement requis, dans le plus défavorable des cas, est d'environ 40 nm (pour une pointe « CDR20 » un peu plus courante, cette limite monte à 50 nm). Dans le cas de l'AFM, on note également que la taille du pitch ne peut pas être considérée seule, puisque l'espace entre deux lignes dépend à la fois du pitch mais aussi de la largeur des lignes.

La largeur des lignes

Nous avons vu que ce paramètre est important pour l'AFM, mais il faut aussi prendre en compte les amplitudes de rugosité considérées. En effet, si l'amplitude de rugosité dépasse la largeur des lignes cela peut, par exemple pour le cas de bords corrélés, créer des discontinuités sur la longueur de la ligne (les deux blocs rouges consécutifs ne se touchent plus), ou dans le cas des bords anti-corrélés, supprimer complètement le bloc dont la taille est réduite. Ainsi, dans le cas où l'on souhaite obtenir des rugosités allant jusqu'à 16 nm , il nous faut considérer une ligne d'une largeur minimale de 20 nm .

En considérant ces deux types de contraintes opposés, sur le CD et le Pitch, sans être dans les conditions limites des équipements de mesure, mais aussi de lithographie, nous avons choisi d'utiliser

un pitch de 200 nm avec un CD moyen théorique de 50 nm. On obtient alors, pour les différents types de motifs, les résultats présentés en Figure IV.3.

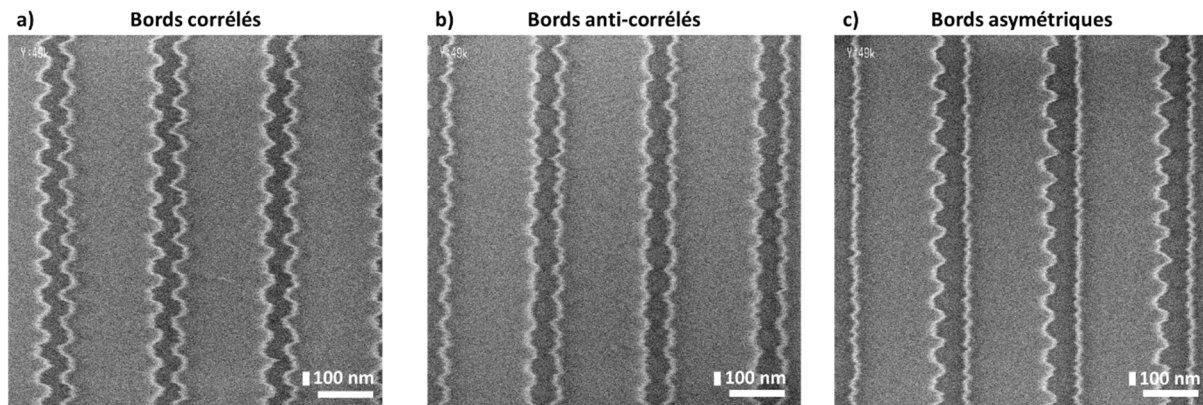


Figure IV.3 : Image CD-SEM rectangulaire (HCG4000) des trois types de motifs pour une amplitude de défaut similaire a) bords corrélés ($l = 16$ nm), b) bords anti-corrélés ($l = 8$ nm), c) bords asymétriques ($l = 16$ nm).

Afin de valider les différentes amplitudes de rugosité obtenues, une analyse avec le CD-SEM a été effectuée, cette technique étant la référence actuelle pour la mesure de rugosité.

IV.1.3 Mesure de validation

Les extractions ont été effectuées sur des images obtenues avec le CD-SEM Hitachi CG4000, images dites "rectangulaires", à savoir qu'elles ont une résolution de pixel de 2,49 nm en y (axe défini le long de la ligne) et de 0,45 nm en x (axe défini orthogonalement à la ligne). Le logiciel utilisé pour l'exploitation est DGMage, logiciel commercial développé par une association indépendante « FreRo Apps ».

Sur les PSD représentées en Figure IV.4, on observe des pics habituellement non présents sur les extractions PSD. Ces pics sont le reflet de la périodicité de notre rugosité. Un pic à $k_n = 0.031 \text{ nm}^{-1}$ (correspondant à une longueur d'onde de 200 nm) apparaît donc pour l'échantillon sur lequel une rugosité de bords anti-corrélés d'amplitude $l = 8$ nm a été induite. Cependant, on remarque que même sur l'extraction PSD du motif n'ayant, en principe, pas de rugosité périodique programmée, des pics apparaissent en $k_n = 0.062 \text{ nm}^{-1}$ et $k_n = 0.126 \text{ nm}^{-1}$. Ces pics peuvent être expliqués par la méthode de génération des motifs. En effet, les lignes sans rugosité sont créées de la même manière que celles avec rugosité, grâce à de multiples rectangles, mais en utilisant un défaut d'amplitude $l = 0$ nm. Ainsi, le fait d'écrire des blocs d'une dimension $\lambda/2 = 100$ nm, même sans rugosité, induit des défauts d'écriture (raccords de structures) qui induisent le pic à $k_n = 0.062 \text{ nm}^{-1}$, le second pic étant, quant à lui, la deuxième harmonique de cette fréquence, $k_n = 0.0628 * 2 = 0.1256 \text{ nm}^{-1}$. En dehors des pics de rugosité explicites, le signal obtenu est dû à la rugosité naturelle du procédé et au bruit des équipements de mesure (voir Chapitre III). Ainsi, si l'on extrait l'amplitude de rugosité σ du motif de référence, avec le théorème de Parseval (Equation II.24). On obtient une valeur qui inclue à la fois la rugosité du procédé mais aussi la rugosité ajoutée par le défaut de raccordement de champs. De la même manière, si l'on calcule l'amplitude de rugosité σ du motif ayant une rugosité programmée, on englobe les mêmes phénomènes (rugosité réelle, raccords de structures), en plus de la rugosité programmée, comme ce qui a été fait dans le chapitre précédent. Ainsi, afin de valider la rugosité induite par le dessin seulement, la PSD de référence est soustraite à celle de chaque échantillon

rugueux, comme dans le cas de traitement de spectre sonore [109]. On obtient ainsi une PSD résiduelle sans la rugosité du procédé et les défauts d'écriture, comme présentée en Figure IV.5. On note qu'il aurait été possible d'effectuer la soustraction sur les valeurs de σ^2 , cependant l'opération sur les PSD permet de garder l'information de répartition fréquentielle.

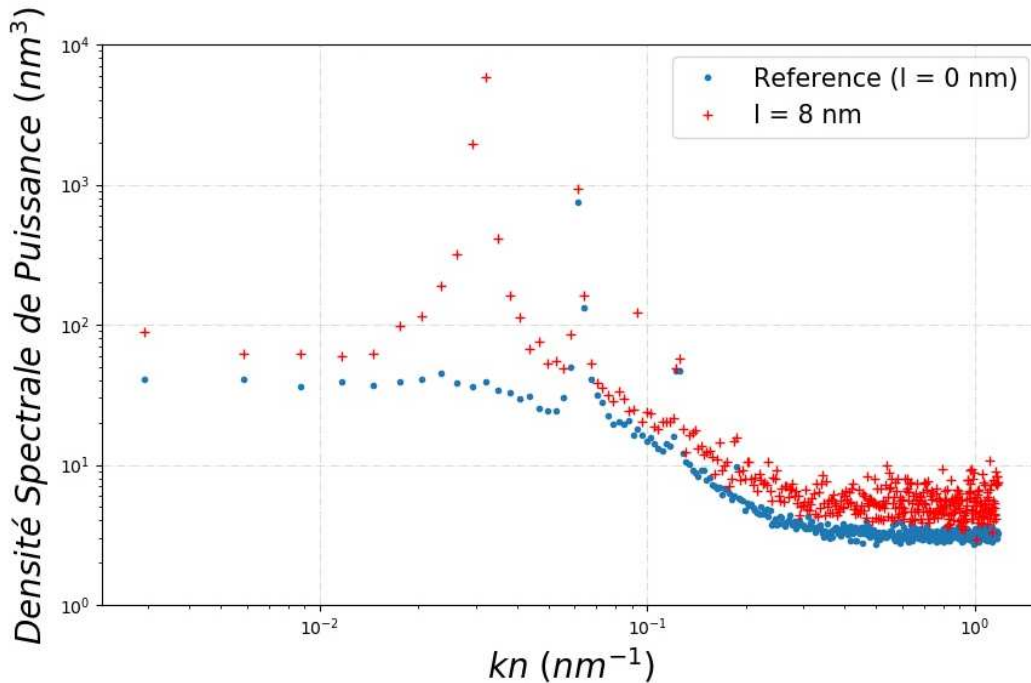


Figure IV.4 : Les PSD LWR, extraites par CD-SEM sur plus de 200 lignes, dans les cas : d'un réseau sans ajout de rugosité (référence, rond bleu), et avec une rugosité de bords anti-corrélés $l = 8$ nm comme représenté en Figure IV.3.b ($l = 8$ nm, croix rouge).

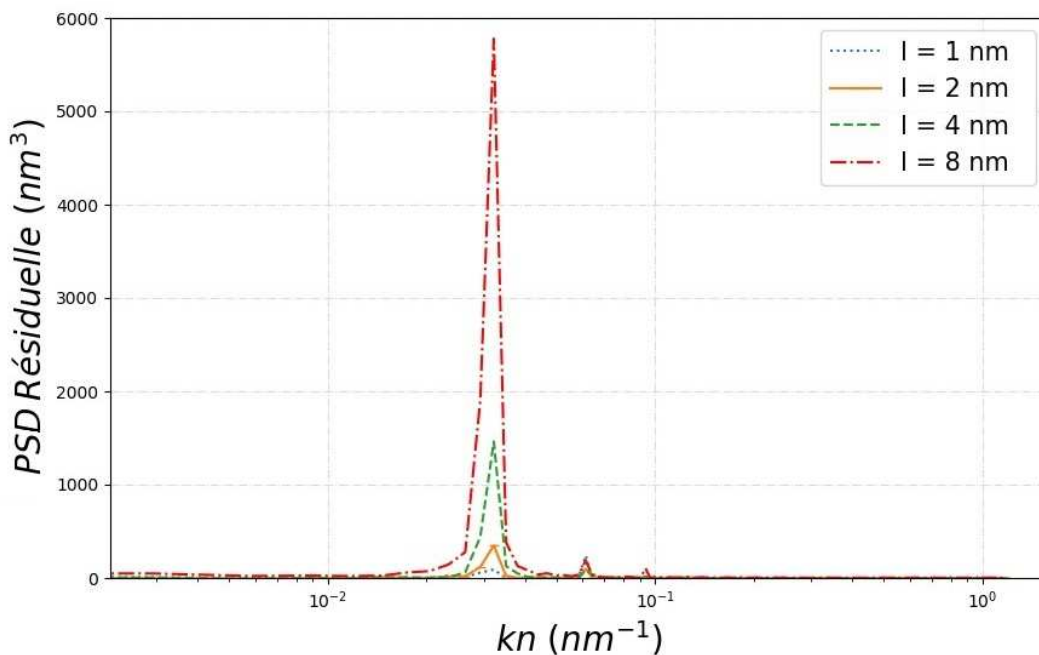


Figure IV.5 : Les PSD résiduelles ($PSD - PSD_{référence}$) obtenues pour les cas de rugosité de bords anti-corrélés de diverses amplitudes.

Sur les PSD résiduelles de l'échantillon aux bords anti-corrélés, Figure IV.5, on retrouve tout d'abord le pic correspondant à la fréquence de rugosité induite ($k_n = 0.0314 \text{ nm}^{-1}$), ainsi que les seconde et troisième harmoniques de cette fréquence ($k_n = 0.0314 * 2 = 0.0628 \text{ nm}^{-1}$ et $k_n = 0.0314 * 3 = 0.0942 \text{ nm}^{-1}$), d'une puissance bien plus faible, due à la forme non parfaitement sinusoïdale du bord de ligne.

En intégrant l'aire sous le pic principal, on trouve alors l'amplitude de rugosité induite par le dessin, les résultats obtenus sont visibles en Figure IV.6 pour le CD-SEM. Sur cette figure, on observe l'ancienne mesure de rugosité faite dans le chapitre précédent par l'intégration de toute la courbe par le théorème de Parseval (en bleu). Par ailleurs, en intégrant seulement le pic en fréquence, après soustraction de la PSD naturelle, on trouve une bonne corrélation avec les valeurs nominales (pente très proche de 1). On retrouve des résultats très proches de ceux extraits en SAXS, cette dernière technique ne reflétant pas la rugosité naturelle du fait de la méthodologie. En effet, la méthode analytique ne prend en compte qu'un seul point d'intensité (soit une fréquence), de plus elle inclue par le ratio d'intensités, Equations III.15 et III.16, un référencement.

Après traitement de référencement sur le CD-SEM la différence la plus significative entre la rugosité extraite et la valeur nominale est pour $l = 1 \text{ nm}$ avec une erreur de 0.06 nm, soit 5.5%. On obtient, par fit linéaire, un biais constant de 0.02 nm pouvant être dû au calibrage de l'équipement ou encore l'échantillonnage de la PSD, ce qui est bien plus faible que les 4.83 nm extrait dans le chapitre précédent. On notera quand même que grâce à cette nouvelle extraction visible en Figure IV.6, le biais est un peu plus important en SAXS (0.2 nm), ce que l'on associe à une extraction moins automatisée et donc moins robuste.

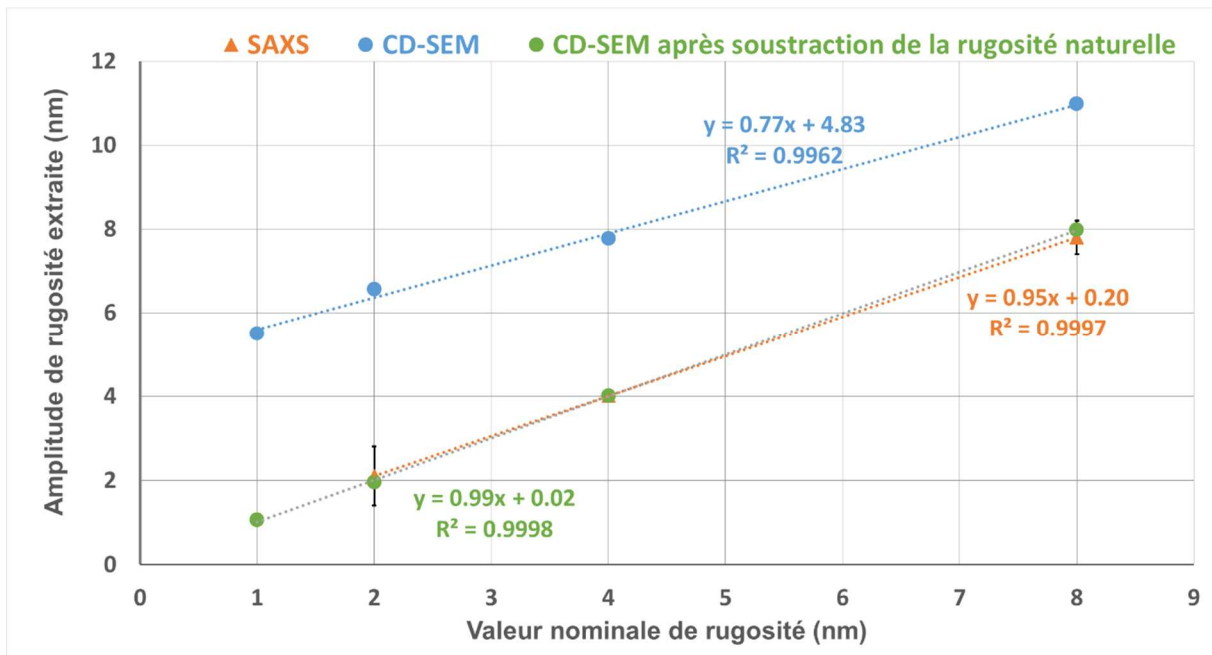


Figure IV.6 : Amplitude de rugosité extraite des échantillons de bords anti-corrélés : avec le CD-SEM sur la PSD complète (rond bleu), avec le CD-SEM après soustraction de la PSD naturelle et intégration de l'aire sous le pic correspondant à la période 200 nm (rond vert), et avec la technique SAXS par la méthode analytique de J. Wang [110] (triangle orange).

Nous venons de montrer dans cette partie qu'il est possible d'induire de la rugosité, dans ce cas périodique, au travers d'un dessin sous forme de blocs créant un défaut. Nous avons vérifié que nous pouvons induire, détecter et quantifier une rugosité jusqu'à une résolution de 1 nm. Par ailleurs, une méthodologie de référencement par soustraction de PSD a été démontrée et validée

expérimentalement afin de ne prendre en compte, dans notre extraction, que la rugosité appliquée sur le dessin et non celle naturellement induite par le procédé. Il est intéressant, maintenant, de se pencher sur un cas de rugosité plus réelle correspondant au modèle PSD explicité précédemment.

IV.2 Programmation de rugosité suivant une PSD

Dans le cadre de la programmation de rugosité plus proche de cas réels, le principe est, non plus de programmer un objet périodique, mais de se baser sur des principes de dessin similaires en induisant une rugosité issue de notre modèle PSD (Chapitre II). De plus, les principes de dessin définis ci-dessus et développés sur la rugosité périodique, tels que les tailles des réseaux sur contrainte multi-équipements par exemple, restent à prendre en compte pour cette programmation.

IV.2.1 Génération du signal de bord

Le principe mathématique du calcul de la PSD permet de passer de l'espace physique à l'espace fréquentiel. Il est donc en principe possible d'effectuer le parcours inverse [62]. Ainsi, en partant du modèle de la PSD défini par P_n et en générant des nombres de manière aléatoire Φ_n (distribution suivant une loi normale standard), on obtient :

$$F_n = \sqrt{\frac{2\pi}{L} P_n} \times \Phi_n \quad (\text{IV.1})$$

avec F_n les facteurs d'amplitude de la Transformée de Fourier discrète pour chaque kn . En effectuant une Transformée de Fourier inverse discrète (TF^{-1}) sur ces facteurs, on obtient un signal de rugosité de ligne. Pour une même PSD de départ, on peut donc générer une multitude de signaux différents qui respecte la distribution statistique de la PSD. Un signal correspond, dans notre cas, aux positions du bord d'une ligne. Pour créer une ligne on applique cette variation suivant deux positions séparées de la taille du CD voulu. Pour les rugosités de types bords : (i) corrélés, les deux flancs bougent simultanément, on applique alors le même déplacement aux deux cotés ; (ii) anti-corrélés, les variations sont appliquées d'un côté et de l'autre avec un changement de signe ; (iii) asymétriques, les variations ne sont induites que sur un bord. Ainsi pour ces cas, une seule génération par ligne est nécessaire ; *a contrario*, si l'on souhaite obtenir des lignes aux bords non corrélés, comme vu Chapitre II, il est nécessaire d'effectuer deux générations (une pour chaque bord) à partir de la même PSD.

IV.2.2 Paramétrage du dessin de lignes

Tout d'abord, il faut noter qu'un signal de bord est généré avec un échantillonnage Δy , celui-ci doit être égal ou sous-multiple de la hauteur « Dy » des rectangles utilisés pour la génération du dessin, comme visible en Figure IV.7. De manière similaire au cas des rugosités périodiques, où Dy était égal à la demi-période, il est important de vérifier si cette valeur peut être résolue correctement par l'outil de lithographie. Ainsi, des tests sur des rugosités périodiques, analysables facilement, ont été effectués à nouveau en diminuant la valeur de la période. L'outil utilisé (e-beam gaussien) ayant un pas de grille (résolution de déplacement) minimal de 1.25 nm, les cas $Dy = n \cdot 1.25$ nm ont été testés pour $n = 1$ à 20 soit $Dy = 1.25$ à 25 nm. Il a été établi, empiriquement, que l'amplitude de rugosité peut aller jusqu'à un maximum de 2 nm avec une taille Dy minimum de 5 nm. Cette valeur de 2 nm laisse une latitude suffisante pour la création de rugosité de ligne avec un $\sigma = 2$ nm, car les pas d'amplitude de rugosité non périodique (voir Figure IV.7), ne font pas des sauts abrupts comme dans le cas de rugosités

périodiques. Dans le cas des écritures sur plaque, c'est donc le paramétrage $Dy = 5 \text{ nm}$ qui a, par suite, été utilisé. Ce dernier induit, nous le verrons, une analyse jusqu'à une fréquence équivalente à une longueur de 10 nm .

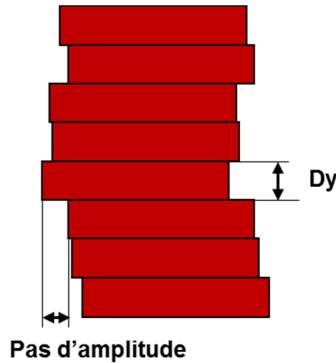


Figure IV.7 : Exemple de dessin de rugosité non périodique par empilement de rectangles pour le type corrélé.

D'autre part nous avons vu, dans le Chapitre II, qu'il est nécessaire d'avoir un nombre de lignes important et sur une longueur suffisante, afin de pouvoir retrouver la PSD d'origine. Il est donc important de générer plusieurs lignes soit un réseau.

IV.2.3 De la ligne au réseau

Dans le cas des rugosités périodiques, nous avons vu que le dessin de périodicité créé était le même sur toutes les lignes. De la même manière, il est possible de générer des réseaux de lignes identiques, comme visible en Figure IV.8, les lignes étant représentées en blanc (soit un 1 binaire le reste étant à zéro). Ce type de réseau a plusieurs avantages. Tout d'abord, le motif ne contient qu'un seul et unique signal de bord de ligne à récupérer, le fait de prendre plusieurs lignes identiques permet alors de diminuer le bruit sur chaque P_n de la PSD (effet de statistique). De plus, d'un point de vue lithographique, il est simple de gérer ce type de réseau car la description d'une seule ligne et du pitch suffisent, ce qui rend le fichier de données contenant le dessin léger, mais aussi parce que les effets de proximité (variations dimensionnelles liées aux fluctuations de dose en un endroit spécifique) sont les mêmes partout. Cependant, les lignes identiques ne permettent pas de récupérer un modèle PSD représentatif d'une rugosité naturelle. En effet, les lignes générées ayant exactement le même signal de bords, il n'y a pas d'effet statistique permettant de converger sur le modèle.



Figure IV.8 : Réseau de lignes identiques générées sur le type corrélé pour un paramétrage $\sigma = 2.0 \text{ nm}$ $\alpha = 0.5$ et $\xi = 20 \text{ nm}$, $Dy = 1 \text{ nm}$, $CD = 50 \text{ nm}$ et $\text{pitch} = 200 \text{ nm}$. Seule une infime partie de la longueur de la ligne est visible ici.

A l'inverse, on peut alors concevoir des réseaux de lignes différentes, comme montré en Figure IV.9, qui sont plus représentatifs de procédés réels du fait de la statistique de nombre. C'est donc ce type

de réseaux que l'on choisira pour les étalons. Il est alors nécessaire de définir un nombre de lignes différentes (nombre de signaux de bords générés) et leur longueur propre, afin d'obtenir une statistique suffisante (voir Chapitre II). Cependant, le nombre total de rectangles générés ne doit pas être trop important, pour que l'outil de lithographie puisse le gérer. Le nombre total de rectangles est calculable par :

$$Nbr_{rec} = \frac{L}{D_y} \times N^* \quad (IV.2)$$

avec L la longueur des lignes qui serait équivalente à la hauteur totale du réseau et :

$$N^* = \frac{\text{Largeur du réseau}}{\text{Pitch}} \quad (IV.3)$$

Au vu des tests de convergence faits dans le Chapitre II, il semble nécessaire d'avoir une taille de ligne d'au moins 2 μm de long, et au moins 100 lignes différentes. C'est donc ce paramétrage qui sera utilisé, aboutissant donc à la réalisation de 40 000 rectangles par réseau de lignes, ce qui reste raisonnable d'un point de vue procédé. Ce bloc est ensuite répété sur la largeur et la longueur afin d'obtenir une taille totale de réseau compatible avec les techniques utilisées dans ces travaux, notamment OCD et SAXS.



Figure IV.9 : Réseau de lignes différentes, générées grâce à la somme des contributions de signaux de type anti-corrélé : $\sigma = 2.0 \text{ nm}$ $\alpha = 0.5$ et $\xi = 20 \text{ nm}$ et corrélé : $\sigma = 5.0 \text{ nm}$ $\alpha = 0.8$ et $\xi = 50 \text{ nm}$ avec $D_y = 1 \text{ nm}$, $CD = 50 \text{ nm}$ et $\text{pitch} = 200 \text{ nm}$. Seule une infime partie de la longueur de la ligne est visible ici.

On notera par ailleurs, comme c'est le cas sur la Figure IV.9, qu'il est possible de générer des lignes à partir de plusieurs signaux : un par bord par exemple, ou même une ligne composée d'un signal créant des bords corrélés, ajouté à un second créant des bords anti-corrélés. On a dans ce cas l'ajout des deux amplitudes et on obtient $\sigma_{total}^2 = \sigma_1^2 + \sigma_2^2$. Dans ce dernier exemple, suivant les amplitudes des deux signaux, les bords obtenus seront plus ou moins corrélés. Le dessin de lignes étant finalement généré, il est possible d'effectuer les étapes de lithographie pour obtenir nos échantillons.

IV.3 La fabrication des étalons et leurs biais

Dans cette partie est réalisée la programmation PSD, comme explicité précédemment, pour divers paramétrages α et ξ . Les différents types de motifs (corrélés, anti-corrélés et asymétriques) ont été générés. Ces divers dessins de rugosité de ligne sont ensuite mis en fabrication avec nos procédés d'élaboration. Les analyses LER et LWR ont été effectuées sur les échantillons obtenus en utilisant le CD-SEM. Les PSD obtenues étant identiques (superposables) suivant le type de rugosité, seul le cas des bords anti-corrélés avec une analyse de la métrique LWR est présenté comme exemple.

Pour rappel, le biais dû à Dy a déjà été explicité lors de la création des dessins et du paramétrage de la PSD, sous la forme de Δy . Cependant, ceci n'est pas la seule déformation que l'on voit apparaître lors de l'écriture des motifs.

IV.3.1 Du modèle à la réalité

De manière générale, on peut considérer que les PSD mesurées expérimentalement sont ici déterminées par :

$$PSD_{mesurée} = F_{Métro} \times (F_{Litho} \times F_{Dessin} \times PSD_{Programmée} + PSD_{Procédé}) \quad (IV.4)$$

Sur la $PSD_{Programmée}$ s'appliquent : une fonction de transfert liée à la création du dessin F_{Dessin} (pixellisation, création de rectangles, ...), puis une fonction de transfert liée à l'outil de lithographie F_{Litho} , qui modifie la PSD. Enfin, il faut ajouter la rugosité due au procédé, $PSD_{Procédé}$. Le tout est mesuré au travers d'une fonction métrologie $F_{Métro}$.

La Figure IV.10 montre les PSD LWR obtenues sur le motif programmé ainsi que sur un motif sans rugosité programmée. On remarque, comme attendu, que la rugosité programmée est plus importante en amplitude que la référence sans rugosité programmée. De la même manière que dans le cas précédent de la rugosité périodique, la programmation est venue s'ajouter à la rugosité naturelle $PSD_{Procédé}$. Ainsi, par la soustraction de la PSD mesurée sur un échantillon de référence, il est possible de supprimer $F_{Métro} \times PSD_{Procédé}$, il reste alors :

$$PSD_{extraite} = F_{Métro} \times F_{Litho} \times F_{Dessin} \times PSD_{Programmée} \quad (IV.5)$$

On peut alors comparer la $PSD_{extraite}$ sur la Figure IV.11, par rapport au modèle numérique de départ.

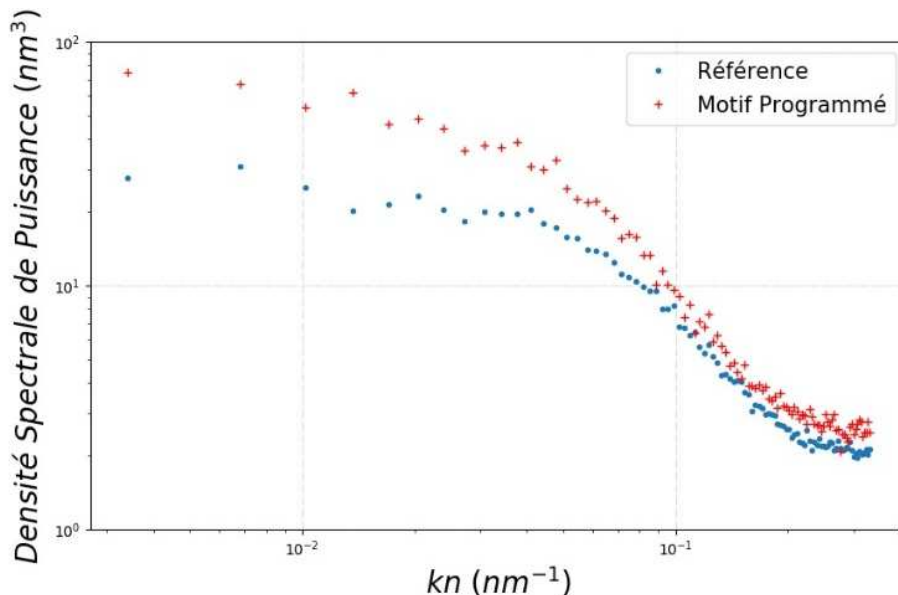


Figure IV.10 : PSD LWR extraite sur plus de 500 lignes (CD-SEM), sur un réseau : sans rugosité ajoutée (référence, rond bleu), et avec de la rugosité programmée $\sigma = 2.0$ nm, $\alpha = 0.5$, $\xi = 20$ nm et $Dy = 5$ nm (motif programmé, croix rouge).

Sur le résultat, visible en Figure IV.11, on peut voir en analysant la PSD des courtes vers les hautes fréquences, tout d'abord que l'intensité basse fréquence est plus forte que dans le cas du modèle, soit une différence d'aire entre les deux courbes pour $k_n < 0.01$ nm⁻¹ équivalent à 0.3 nm (voir théorème de Parseval, Equation II.24). De plus, un décalage du plateau suivant k_n (d'environ 0.04⁻¹ à 0.05 nm⁻¹) et donc de la valeur de ξ (environ 20 nm) est visible. Enfin, on remarque que le niveau en haute

fréquence est plus bas que sur le modèle avec une différence d'aire équivalente à une rugosité de 0.6 nm.

Malgré le fait qu'on observe bien une modification de la rugosité expérimentalement, ce résultat montre des différences importantes entre la rugosité souhaitée et celle obtenue sur plaque. Dans la suite de cette partie, nous allons donc mettre plus en évidence les différences entre le modèle et le résultat expérimental en fonction des divers paramètres de programmation de la rugosité.

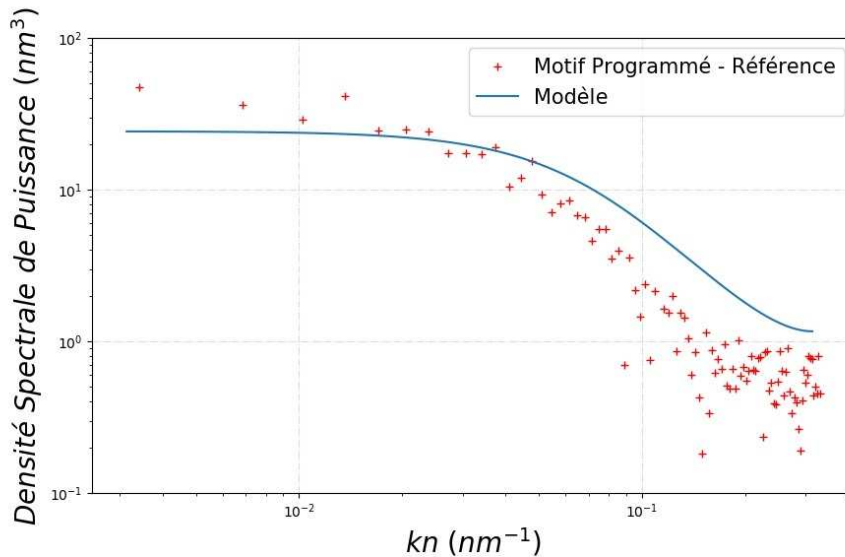


Figure IV.11 : PSD obtenue après soustraction de la PSD LWR de référence à celle d'un motif programmé, ainsi que la PSD modèle introduite initialement.

IV.3.1-A Variation de α

Si l'on compare à ξ constant et α variable, les courbes théoriques visibles en Figure IV.12, et celles obtenues expérimentalement après soustraction de la rugosité naturelle présentées en Figure IV.13, on observe de nombreuses similitudes.

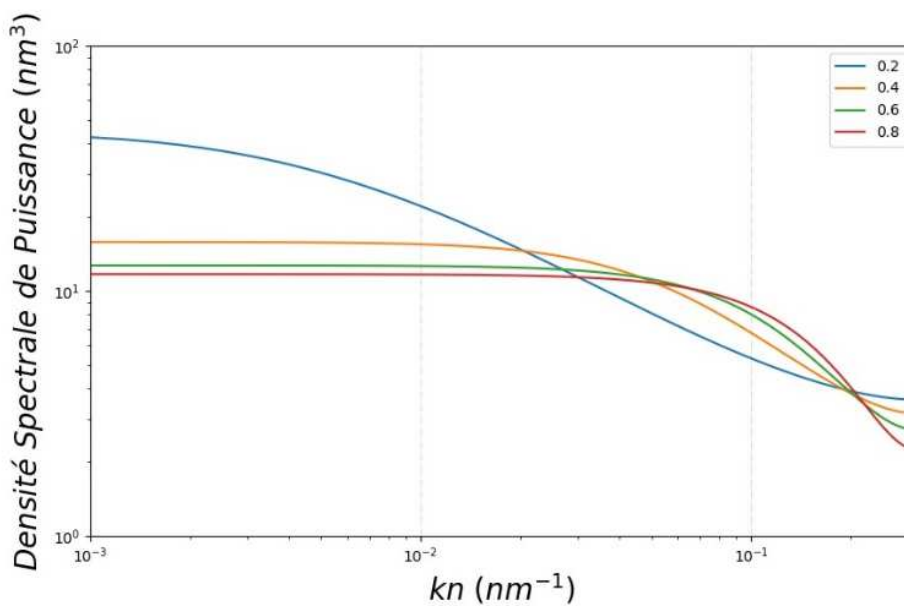


Figure IV.12 : Modèle numérique de PSD pour diverses valeurs de α , avec un paramétrage identique $\sigma = 2.0$ nm, $\xi = 20$ nm et $Dy = 5$ nm.

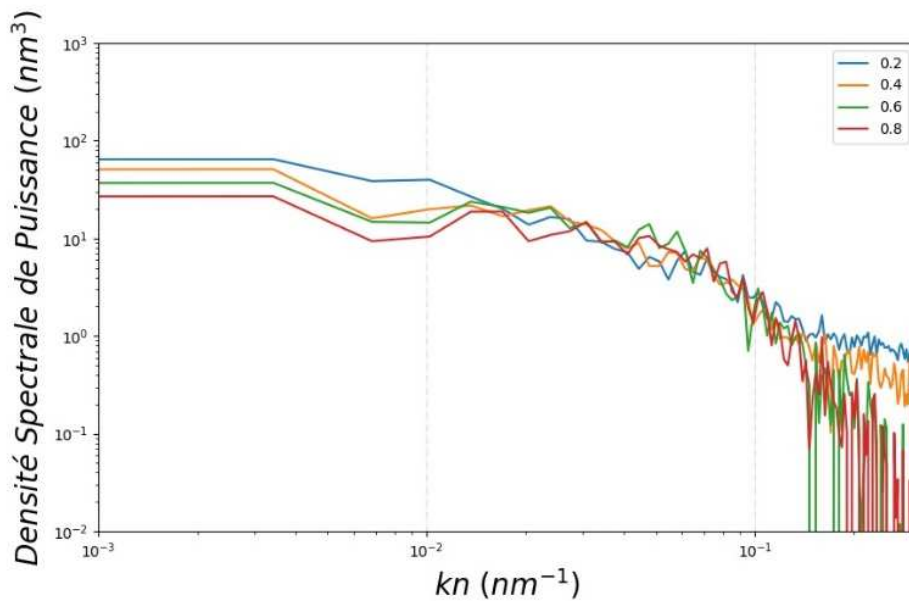


Figure IV.13 : PSD LWR extraite après référencage, pour diverses valeurs de α avec un paramétrage identique $\sigma = 2.0$ nm, $\xi = 20$ nm et $D_y = 5$ nm.

Premièrement l'ordonnancement en amplitude, notamment en basse fréquence, est identique entre le modèle théorique et les mesures expérimentales. La valeur nominale d'amplitude est différente, comme évoqué, mais l'ordre de grandeur est conservé. On retrouve aussi un ordonnancement correct dans les hautes fréquences, très visible pour de faibles valeurs de α (0.2, 0.4), les PSD ayant un α plus important (0.6 et 0.8) étant, elles, confondues dans le bruit.

Par ailleurs, si l'on regarde les pentes (hors plateaux) des courbes PSD qui sont (comme explicité Chapitre II) proportionnelles à $1/\alpha$, on observe par exemple que la courbe obtenue pour $\alpha = 0.2$ est moins abrupte que pour $\alpha = 0.8$. Plus exactement, en calculant la pente γ dans l'espace log10 soit :

$$\gamma = \frac{d(\log_{10}(P_n))}{d(\log_{10}(k_n))} \quad (\text{IV.6})$$

pour $k_n > 0.05$ nm⁻¹, on obtient pour les différentes courbes extraites suivant $\alpha = 0.2, 0.4, 0.6$ et 0.8 respectivement $\gamma = -1.37, -1.88, -3.06, -3.12$ u.a., ce qui est cohérent qualitativement avec le résultat attendu.

On note cependant que la courbe $\alpha = 0.2$ semble atteindre les limites de la programmation avec ses paramètres de génération et de réalisation, car dans le modèle nominale sa forme est déjà remarquable par rapport aux autres. Ceci peut être expliqué par le fait que ce paramétrage de rugosité, α très faible, crée un bord avec des discontinuités en hautes fréquences très importantes. L'hypothèse posée ici est que le procédé ne peut écrire de telles discontinuités (effet de lissage).

Une analyse, similaire à celle faite pour les variations de α par comparaison des courbes théoriques avec les PSD obtenues expérimentalement, peut aussi être faite pour les variations du paramètre ξ à α constant.

IV.3.1-B Variation de ξ

En comparant les courbes PSD théoriques (Figure IV.14), et celles obtenues expérimentalement après suppression de la PSD de référence (Figure IV.15), dans le cas de variation de paramètre ξ , on peut tirer les mêmes conclusions que pour les variations de α , notamment sur les amplitudes aux basses

fréquences. On remarquera, plus spécifiquement dans ce cas, que la variation de la position de la fin du plateau (en basse fréquence) est obtenue, comme attendu par la théorie (voir Chapitre II).

On souligne quand même le fait que la programmation d'un $\xi = 5$ nm n'est pas réellement faisable avec un $D_y = 5$ nm. Ceci est visible directement sur la courbe PSD nominale qui est presque plate, on voit cependant qu'elle apparaît plus marquée sur la mesure (Figure IV.15), les hautes fréquences ayant donc été lissées lors de la réalisation, cette partie est explicitée plus en détail dans la partie V.4. La PSD avec $\xi = 10$ nm est aussi en limite, ce paramétrage correspond au dessin de deux rectangles seulement, pour générer un caractère de rugosité redondant ce qui est insuffisant (voir théorie sur la rugosité de ligne et le paramètre ξ au Chapitre II).

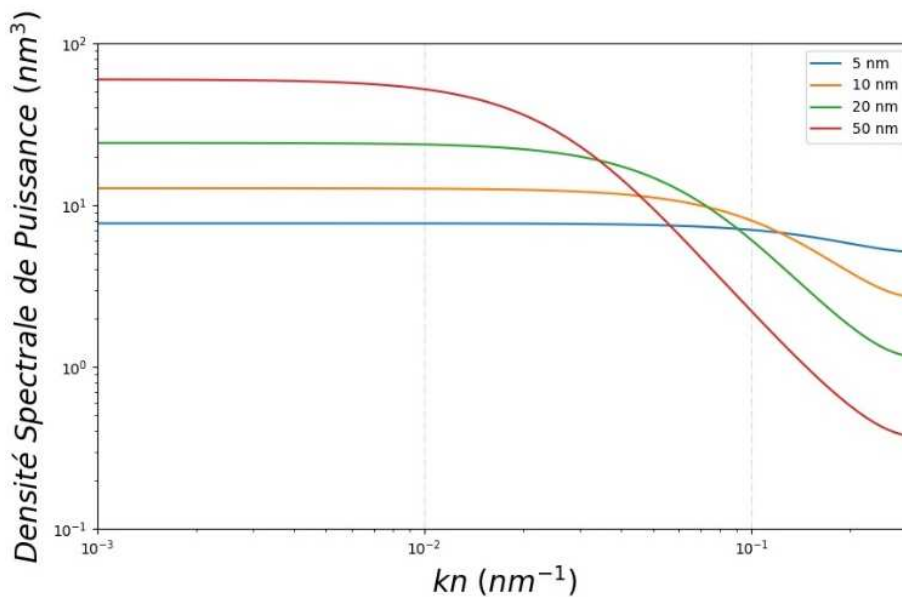


Figure IV.14 : Modèle numérique de PSD pour diverses valeurs de ξ avec un paramétrage identique $\sigma = 2.0$ nm, $\alpha = 0.6$ et $D_y = 5$ nm.

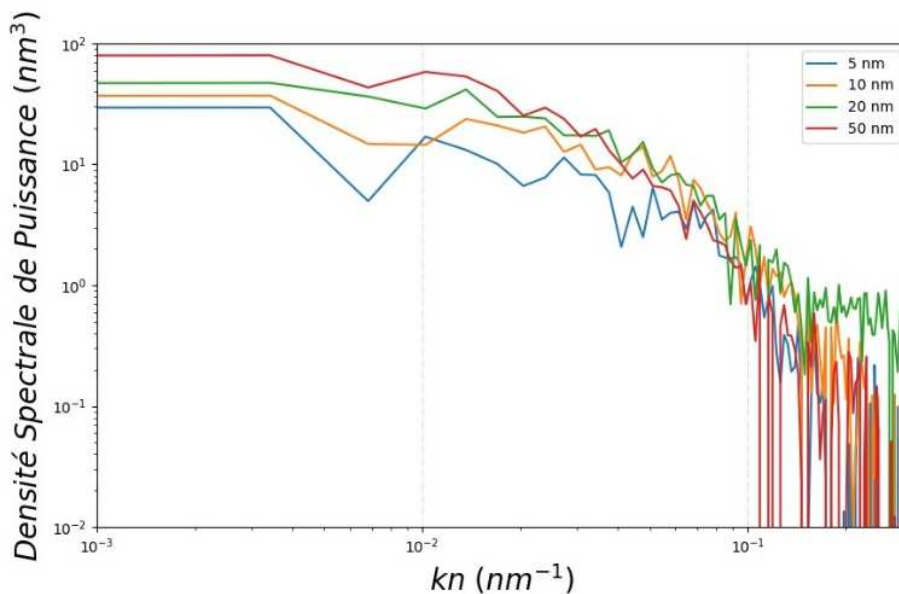


Figure IV.15 : PSD LWR extraite après référencement, pour diverses valeurs de ξ avec un paramétrage identique $\sigma = 2.0$ nm, $\alpha = 0.6$ et $D_y = 5$ nm.

Afin de comprendre plus en détail les modifications de la PSD, du modèle à l'extraction expérimentale, et pouvoir par la suite générer des étalons plus proches du motif visé, nous nous sommes basés sur

l'Equation IV.5 pour différencier les causes. Le premier biais alors exploré est la fonction de transfert liée à la génération du dessin de rugosité, F_{Dessin} .

IV.3.2 Les biais du dessin

Tout d'abord, pour obtenir un dessin formé de pixels (partie élémentaire), sur la grille de lithographie, les valeurs du signal généré sont arrondies. Pour comprendre l'impact de cet arrondi, le calcul de la PSD moyenne sur 1000 lignes a été effectué, avec un arrondi sur le signal de bord (arrondi) et sans arrondi sur ce même signal (brut), visibles en Figure IV.16. Sur cette figure est aussi représentée la valeur absolue de la différence entre ces deux PSD. On peut voir que cette valeur est relativement constante sur la totalité des fréquences, ce qui est assimilable à un bruit blanc. La puissance de ce bruit est en moyenne inférieure à 0.02 nm^3 , soit 5 fois moins que la résolution sub-nanométrique attendue, il peut donc être négligé.

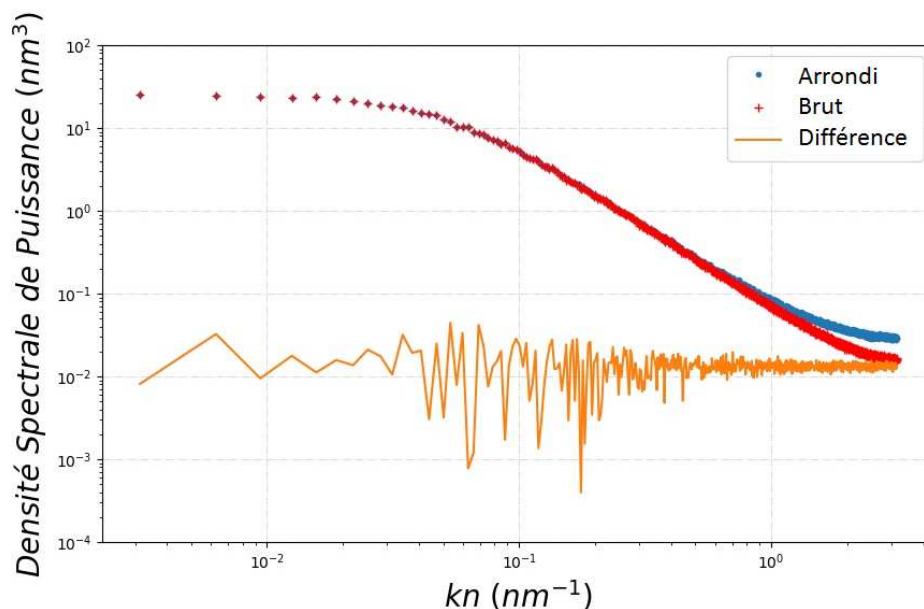


Figure IV.16 : PSD de 1000 lignes dans le cas : de valeurs arrondies au pixel 1 nm (rond bleu) et sans arrondi de valeur (brut, croix rouge) ; ainsi que le tracé de la valeur absolue de la différence entre les deux PSD (en orange) pour un paramétrage $\sigma = 2.0 \text{ nm}$, $\alpha = 0.5$, $\xi = 20 \text{ nm}$ et $\Delta y = 1 \text{ nm}$.

De plus le dessin étant composé, comme précédemment expliqué, de 100 lignes de $2 \mu\text{m}$ de long répétées sur la surface totale de l'échantillon, il existe lors de la mesure expérimentale un biais dû à la statistique, non pas de 500 lignes mesurées (500 signaux de bords différents), mais de 100 lignes différentes seulement. Ceci donne des erreurs sur σ , α et ξ respectivement de 0.3%, 3% et 1% (voir Chapitre II), c'est ce niveau d'erreur qu'il est important de prendre en compte ici.

Au-delà du dessin, le passage par les équipements de métrologie et par la technique d'extraction associée peut induire le biais $F_{\text{Métro}}$, comme présenté dans l'Equation IV.5.

IV.3.3 Les biais métrologiques

Pour considérer l'impact de la chaîne de mesure, nous sommes partis d'images binaires de lignes rugueuses, l'une d'entre elles est visible en Figure IV.17. Ce type d'échantillon, que nous avons nommé "échantillon virtuel", est caractérisé par des transitions espaces-lignes qui sont clairement marquées 0 à 1, comme le montre la Figure IV.18.

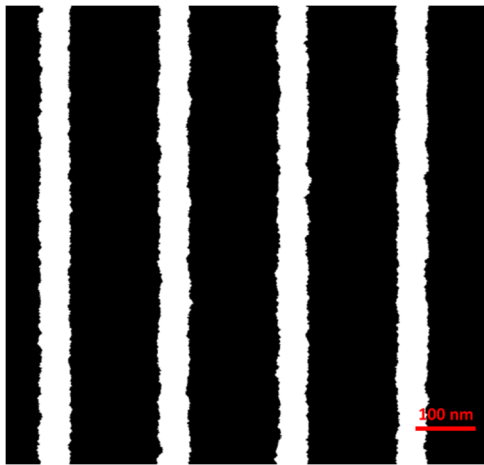


Figure IV.17 : Image d'un morceau d'échantillon virtuel (binaire).

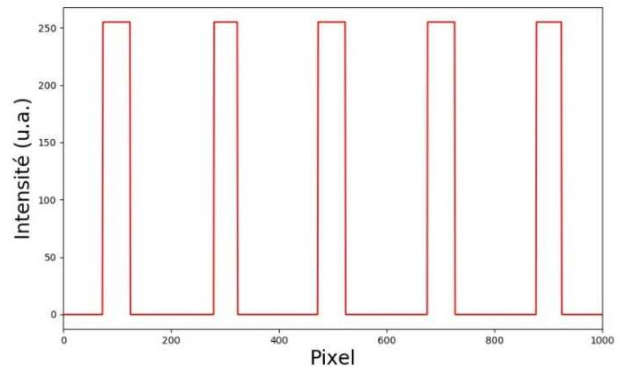


Figure IV.18 : Signal obtenu sur une coupe horizontale d'un échantillon virtuel (binaire).

Par le biais d'un logiciel interne (SEMABOAR), calibré sur la réponse d'un faisceau d'électrons impactant du silicium [111], il est possible de transformer ces représentations binaires en images sur niveaux de gris, typiques d'acquisitions SEM réelles. Un exemple d'image sans ajout de bruit (type aléatoire) est visible en Figure IV.19. Ces images, similaires à celles pouvant être obtenues avec un CD-SEM, ont contrairement aux représentations binaires, des transitions espaces-lignes plus compliquées à déterminer, comme présenté sur la Figure IV.20 (voir la partie extraction de la mesure CD-SEM présentée au Chapitre III). Grâce à ces simulations SEM, appliquées à nos motifs de rugosité de ligne, la chaîne de mesure peut être évaluée.

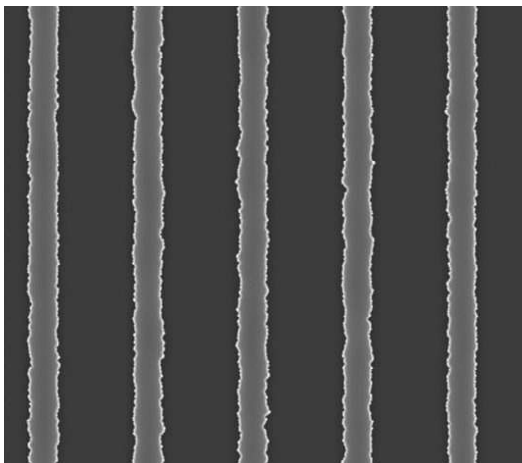


Figure IV.19 : Image d'un morceau d'échantillon après simulation SEM (niveaux de gris).

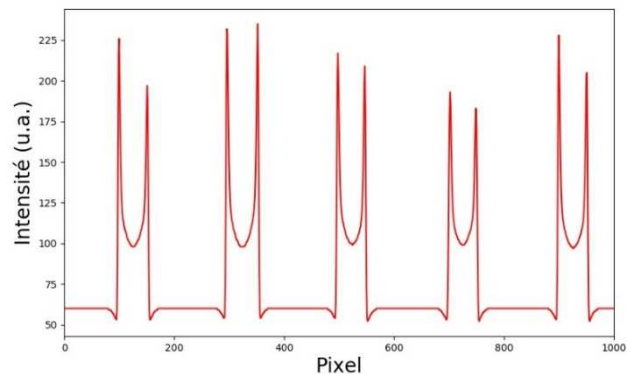


Figure IV.20 Signal obtenu sur une coupe horizontale d'un échantillon après simulation SEM (niveaux de gris).

Si l'on effectue l'extraction de la PSD sur les images de simulations SEM, on peut la comparer au même type d'extraction faite sur les images binaires. La Figure IV.21 montre donc les extractions effectuées dans ces deux cas. On remarque, sur cette figure, une très forte diminution des puissances pour les fréquences supérieures à $k_n = 0.1 \text{ nm}^{-1}$. Cette diminution est représentative d'un lissage de la rugosité au-delà de cette fréquence. Ce lissage provient du modèle SEM et traduit l'effet de lissage dû au faisceau d'électrons de forme gaussienne. Bien que visible seulement en hautes fréquences, à cause des puissances faibles ($<1 \text{ nm}^3$) dans cette plage, cet impact est cependant présent sur la totalité des fréquences.

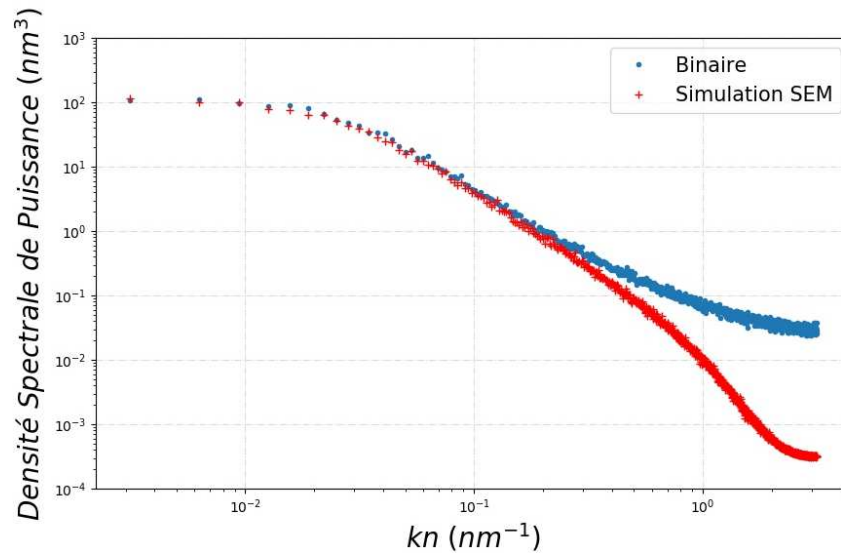


Figure IV.21 : Extractions PSD LWR effectuées pour 140 lignes, dans le cas d'images binaires (rond bleu) et de simulations SEM (croix rouge) pour un paramétrage, $\sigma = 2.0$ nm, $\alpha = 0.5$, $\xi = 20$ nm et $\Delta y = Dy = 1$ nm

Ce biais $F_{\text{Métro}}$ est important, il est donc nécessaire de le considérer comme une limite de l'extraction SEM. On notera cependant que le modèle induit un biais qui n'est pas exactement celui de l'équipement. En effet, la simulation a été calibrée pour une énergie de faisceau de 5 keV (10 fois supérieure à celle utilisée pour les extractions faites précédemment), ce qui signifie que le biais du modèle est sans doute plus important que celui expérimental, expliquant ainsi une déformation moindre sur les hautes fréquences sur les mesures à comparer des simulations. De plus, on note que le bruit du CD-SEM n'a pas été traité dans cette partie de simulation. Le dernier biais restant à considérer dans notre équation IV.5 est celui de la lithographie F_{Litho} , ceci nécessiterait de calibrer le procédé dans sa globalité et de l'inclure dans un modèle. Ce type de recherches nécessite des temps de calculs importants, pour un résultat (en termes de rugosité de ligne) très variable au vu de la diversité des motifs. Nous nous baserons donc sur le fait que le biais lithographique est limité, grâce au paramétrage de Dy à 5 nm, comme vu dans les tests réalisés précédemment.

En prenant en compte les diverses contributions, grâce aux outils disponibles au moment de l'étude, nous avons donc pu effectuer la simulation de la génération de rugosité périodique.

IV.3.4 Extraction de rugosité sur images simulées

Dans la simulation des motifs on prendra en compte, sur la base de l'Equation IV.5, que :

$$PSD_{\text{Simulée}} = F_{\text{Métro}} \times F_{\text{Dessin}} \times PSD_{\text{Programmée}} \quad (\text{IV.7})$$

ainsi on obtient :

$$F_{\text{Litho}} = PSD_{\text{extraite}} / PSD_{\text{Simulée}} \quad (\text{IV.8})$$

aux erreurs près du modèle métrologique.

Sur les images simulées, les PSD LWR ont donc été extraites comme dans le cas des mesures expérimentales. On retrouve alors les PSD extraites pour des variations de paramètres α , présentées sur la Figure IV.22, et celles permettant de visualiser l'influence du paramétrage de ξ , sur la Figure

IV.24. Il est ainsi possible de voir l'évolution des PSD entre le modèle numérique, le dessin après simulation, et le résultat réel obtenu expérimentalement.

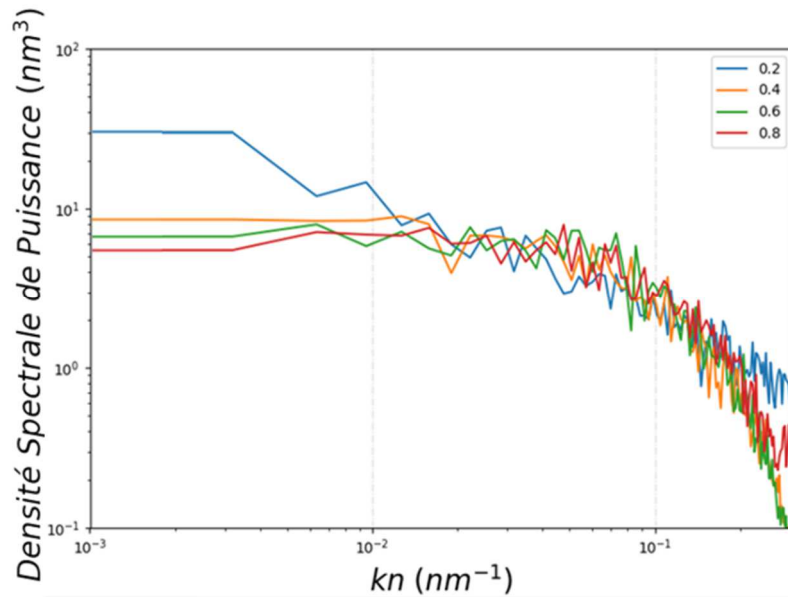


Figure IV.22 : PSD LWR extraite sur les images SEM simulées, pour diverses valeurs de α , avec un paramétrage identique $\sigma = 2.0 \text{ nm}$, $\xi = 20 \text{ nm}$ et $Dy = 5 \text{ nm}$.

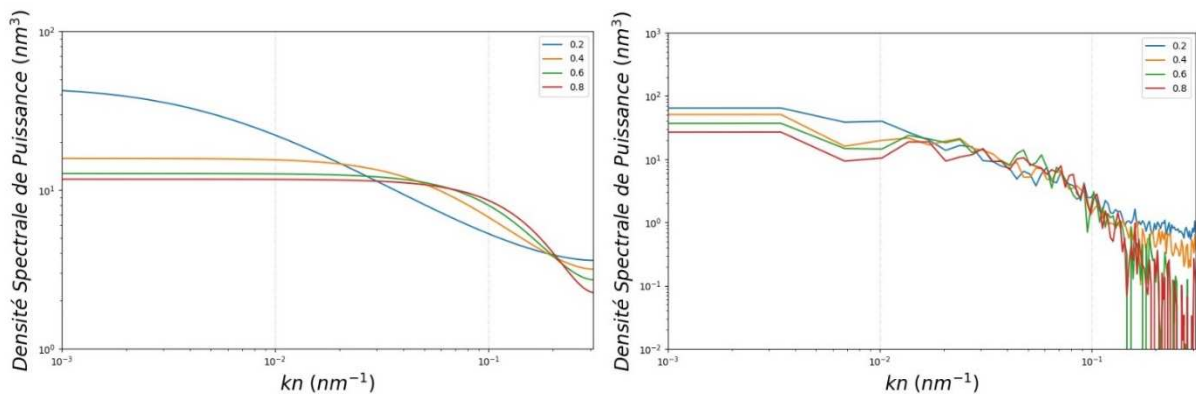


Figure IV.23 : Réédition des Figure IV.12 (à gauche, modèles) et Figure IV.13 (à droite, extractions expérimentales) pour comparaison des PSD LWR représentées par rapport à celles sur les images CD-SEM, simulées dans le cadre de la variation du paramètre α .

Si l'on compare dans le détail ces trois cas de PSD, on note tout d'abord une forte modification des puissances dans les hautes fréquences sur les PSD_{Simulées}, par rapport au modèle numérique, cette modification est plus importante que celle lors du passage du modèle au cas expérimental. Ce changement de puissance a un effet sur les pentes des PSD générées suivant une variation du paramètre α , les cas 0.4, 0.6 et 0.8 venant à se confondre au-delà de $k_n = 0,1 \text{ nm}^{-1}$. Cette modification excessive peut être expliquée par (i) une modélisation inappropriée de l'impact du CD-SEM avec, comme vu précédemment, l'énergie du faisceau qui n'est pas bien prise en compte dans nos paramétrages de simulation, (ii) par l'absence de bruit sur les images simulées, ce qui augmenterait la puissance du plateau hautes fréquences.

Par ailleurs, dans les basses fréquences les simulations restent cohérentes avec le modèle. On retrouve l'écart atypique du cas $\alpha = 0.2$ par rapport aux autres, ce qui n'est pas visible sur la PSD obtenue sur

plaque ; ceci confirme l'hypothèse précédente évoquée dans la partie IV.3.1-A, à savoir un lissage dû au procédé de lithographie et non à la création du dessin.

Dans le cas des variations du paramètre ξ , les différentes PSD sont bien distinguables en hautes fréquences (variations de 1 à 2 décades), contrairement au cas de la variation de α . L'ordonnement des PSD reste similaire au modèle numérique, seuls les cas proches de $\xi = 10$ nm et 20 nm se confondent.

Au niveau des basses fréquences, tout comme dans le cas des PSD extraites expérimentalement, l'ordre est conservé. On note quand même, de façon identique au cas des variations de α , que les PSD_{Simulées} ont une puissance moindre que le modèle, on a ici un lissage global de la rugosité.

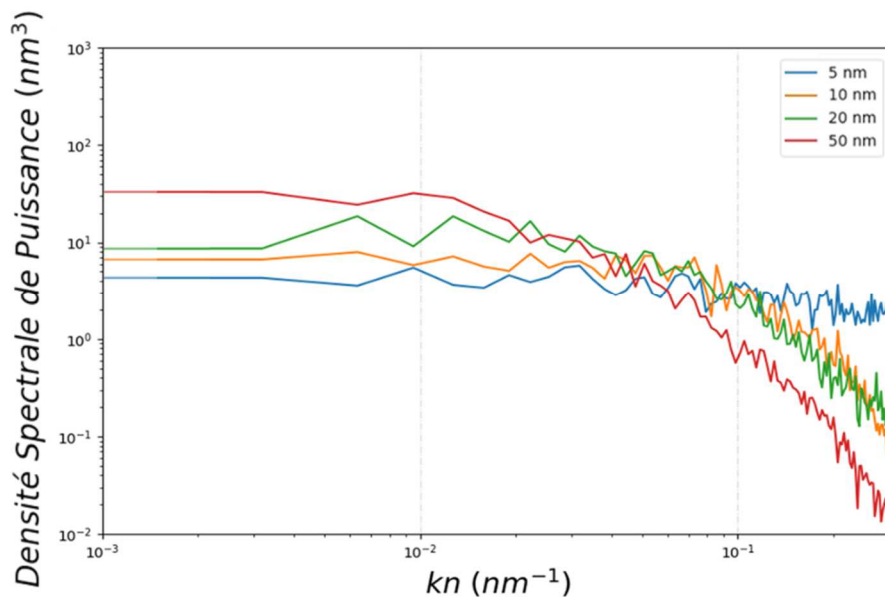


Figure IV.24 : PSD LWR extraites sur les images SEM simulées pour diverses valeurs de ξ , avec un paramétrage identique $\sigma = 2.0$ nm, $\alpha = 0.6$ et $Dy = 5$ nm.

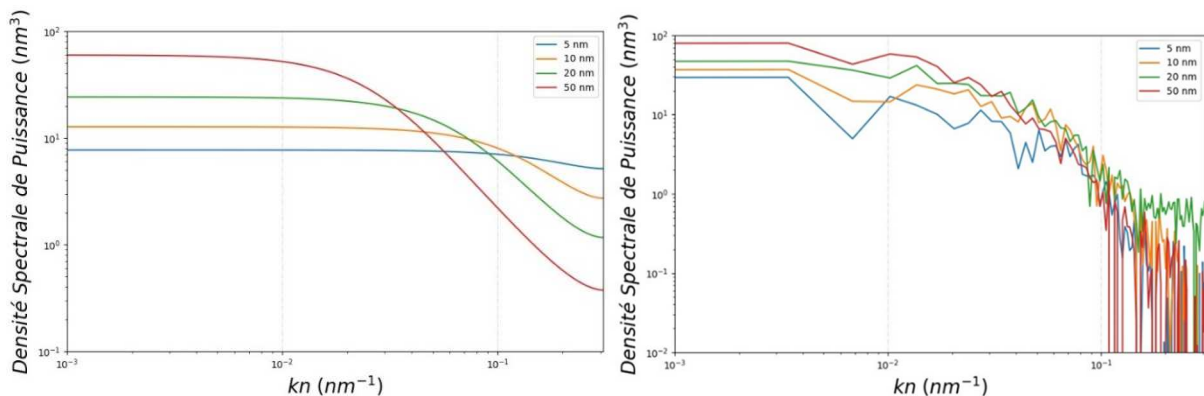


Figure IV.25 : Réédition des Figure IV.14 (à gauche, modèle) et Figure IV.15 (à droite, extraction sur plaque) pour comparaison des PSD LWR représentées par rapport à celles sur les images CD-SEM, simulées dans le cadre de la variation du paramètre ξ .

De manière plus générale, on remarque que les PSD_{Simulées} ont des formes très proches de celles expérimentales, notamment dans les basses fréquences, ainsi la modification dans cette zone est principalement induite par la chaîne de mesure. *A contrario*, la partie haute fréquence diverge entre

les PSD_{Simulée}, et les cas pratiques, on déduit donc que cette déformation est principalement due à la fonction de l'outil de lithographie d'après l'Equation IV.8.

On retiendra ici que la programmation est possible, mais que des biais existent. Connaissant l'impact de chacun, nous allons maintenant voir si l'on peut éviter ces biais et les possibles leviers pour améliorer la programmation de rugosité, et donc nos étalons de rugosité de ligne finaux.

IV.4 Améliorations et limites de la programmation

Divers problèmes ont été soulevés dans les parties précédentes en reliant les causes et les effets sur des plages de fréquences spécifiées, notamment au travers des diverses contributions mentionnées Equation IV.4. A partir de là, il est possible de mettre en évidence différentes solutions pour améliorer cette programmation, tout en ayant conscience des limites de ces solutions et de leur réalisation.

IV.4.1 Amélioration du signal généré

La méthode de génération de rugosité par la transformée de Fourier inverse (TF⁻¹) a la problématique d'inclure le biais dû à la discrétisation du signal de bord et à la taille de fenêtre, comme dans le cas de la TF explicité dans le calcul de la PSD au Chapitre II. Il existe une autre méthode de génération, par matrice de corrélation, se basant sur la même théorie sans passage par le domaine fréquentiel.

IV.4.1-A Génération par matrice de corrélation

La matrice de corrélation correspond à la corrélation entre m variables aléatoires X_m. Dans notre cas on considère une variable m=0 comme le signal de bord de ligne (ou de CD de ligne) dans une position, et la variable suivante m=1 comme le même signal décalé d'un point d'échantillonnage sur son support. On retrouve alors la notion d'ACF décrite dans le Chapitre II. Ainsi, sur la base du modèle ACF décrit Equation II.14 et réécrit ci-dessous :

$$R(m) = \sigma^2 e^{-|m/\xi|^{2\alpha}} \quad (IV.9)$$

où m est le décalage entre le signal et sa copie, il est possible de remplir une matrice de corrélation comme visible sur l'exemple de la Figure IV.26, sous forme d'image. Plus généralement, la matrice se représente sous forme de tableau comme en Figure IV.27. On retrouve ici les caractéristiques principales d'une matrice de corrélation C, à savoir qu'elle est carrée de taille N x N, symétrique et semi-définie positive à diagonale constante R_{m=0} = σ² (la matrice de corrélation correspondant à une matrice de covariance au facteur σ² près).

Connaissant la matrice C, il est alors possible d'obtenir un signal Y ayant les propriétés de distribution (μ, C), μ étant la moyenne du signal, à partir d'une génération aléatoire normale X par :

$$Y = C^{1/2}X + \mu \quad (IV.10)$$

Or, d'après la factorisation de Cholesky [112], il existe une matrice triangulaire L (les valeurs en dessous ou au-dessus de la diagonale étant toutes nulles), telle que :

$$C = LL^T \quad (IV.11)$$

L^T étant la transposée de L soit :

$$C^{1/2} = L \quad (IV.12)$$

Il est donc possible de générer, grâce à la matrice L (de taille N*N) et une génération aléatoire normale X (de taille N), un signal Y (de taille N) respectant la covariance induite par C :

$$Y = LX + \mu \tag{IV.13}$$

Dans notre cas μ est égal à 0 lors de la génération, on ajoute, *a posteriori*, l'information de CD ou de position moyenne du bord. Durant ces travaux, les calculs des matrices L et Y sont effectués avec une seule fonction de la librairie « numpy » du langage python.

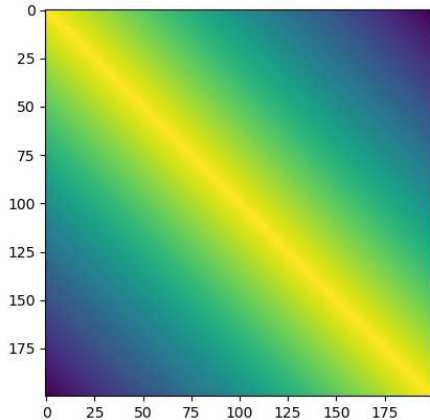


Figure IV.26 : Représentation imagée d'une matrice de covariance avec le modèle ACF pour $\sigma = 2 \text{ nm}$, $\alpha = 0.5$ et $\xi = 20 \text{ nm}$, $N = L = 200 \text{ nm}$.

j \ i =	0	1	...	N
0	$R_0 = \sigma^2$	$R_{m=1}$...	$R_{m=N}$
1	$R_{m=1}$	$R_0 = \sigma^2$...	$R_{m=N-1}$
...	$R_0 = \sigma^2$...
N	$R_{m=N}$	$R_{m=N-1}$...	$R_0 = \sigma^2$

Figure IV.27 : Matrice de covariance générique i et j représentant les décalages respectivement sur le signal et sa copie, ainsi dans l'Equation IV.9 $|i - j| = m$.

Cette génération-là est plus proche du modèle car elle n'est pas entachée du biais induit par la Transformée de Fourier. Mais si l'on recalcule la valeur de la PSD post-génération, on peut voir dans le paragraphe suivant, par comparaison, qu'elle n'est pas non plus identique au modèle, même après un nombre de générations de plus de 1000 lignes.

IV.4.1-B Différence des deux modes de générations

Sur la Figure IV.28 est tracée la différence entre le modèle et les générations, par TF^{-1} ou par matrice de corrélation. Ces différences correspondent à l'erreur de chacune d'elles. On remarque, sur ce graphique, que l'erreur lors de l'utilisation de la TF^{-1} est très supérieure à celle de la matrice de corrélation, dès que $k_n > 0.03 \text{ nm}^{-1}$. On pourrait penser que l'erreur induite par la génération TF^{-1} est constante sur toute la plage de fréquences et qu'il s'agit donc d'un bruit blanc, simple à traiter. Cependant, lorsqu'on déduit de cette erreur le bruit blanc (constante équivalente à la moyenne des valeurs), on obtient une erreur résiduelle non nulle et encore supérieure à celle de la génération par matrice de corrélation. Ceci induit que l'erreur faite lors de la génération par TF^{-1} n'est pas un bruit blanc, ou pas seulement un bruit blanc, mais aussi que la génération par matrice de corrélation génère moins d'erreurs.

On retiendra que lors de génération par TF^{-1} , l'erreur induite est d'une puissance moyenne de 0.7 nm^3 , comme visible en Figure IV.28, ce qui est trente fois plus important que le biais du dessin dû à l'arrondi du signal mentionné précédemment. Ainsi, c'est la méthode de génération qui reste le plus grand des biais dans notre signal généré. Il serait donc pertinent de faire la programmation avec la méthode passant par la matrice de corrélation comportant une erreur plus faible.

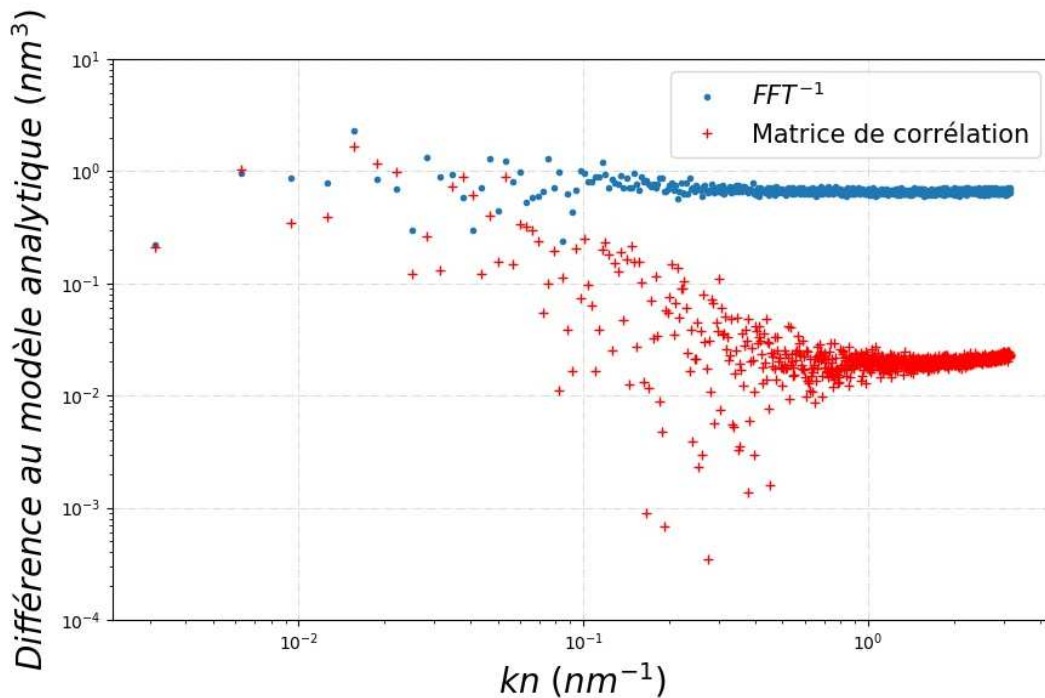


Figure IV.28 : Différence entre les générations effectuées sur 1 000 Lignes : au travers du modèle TF^{-1} (FFT^{-1}) ou par matrice de corrélation, par rapport au modèle PSD.

Néanmoins, il est important de souligner qu'intrinsèquement le calcul d'une TF^{-1} est beaucoup plus rapide que celui de la matrice triangulaire nécessaire à la génération par matrice de corrélation, dont la taille est proportionnelle au carré de la taille N du profil généré. On note cependant que le calcul TF^{-1} doit être effectué sur chaque génération, une à une, tandis que le calcul de la matrice L ne se fait qu'une seule fois. Pour ce qui est de la variable aléatoire, ce sont strictement les mêmes générations qui sont réalisées dans les deux cas, soit un temps identique. En conclusion, pour un nombre très important de lignes on s'attend à ce que la génération par matrice de corrélation devienne plus rapide que celle en TF^{-1} .

La méthode de génération par matrice de corrélation, et les résultats liés, ont été obtenus au cours des avancées sur ces travaux de rugosité. Désormais, toutes les générations sont faites suivant cette génération.

Un des autres biais observés, qui intervient en second dans la réalisation des échantillons, est la modification des motifs soit F_{Litho} lors de l'écriture sur plaque, mais il est possible d'améliorer cela.

IV.4.2 Les outils de lithographie et les procédés

Pour améliorer concrètement la réalisation des étalons de rugosité, il est nécessaire de mieux prendre en compte la fonction lithographique F_{Litho} , ou de réduire son impact.

Il est possible, pour cela, d'utiliser des procédés plus résolus. C'est-à-dire optimiser les matériaux comme la résine de lithographie, ainsi que son développement, ou encore les paramètres d'exposition des outils tels que le pas de grille, ou encore la forme de la source de lumière (voir Chapitre I sur la limite de résolution de la photolithographie). Pour exemple, dans le cas de la lithographie e-beam par faisceau gaussien, utilisée dans ce travail, il est possible de définir des sens de balayage du faisceau mais aussi son énergie. Pour tester l'amélioration de la résolution, l'utilisation des rugosités périodiques avec une analyse de l'amplitude obtenue suivant la taille des blocs Dy , peut-être

envisagée. On note que la production conforme des objets dessinés informatiquement, ainsi que la résolution ultime de lithographie sont des sujets vastes [113, 114, 19] liées aux lithographies avancées. Dans un second temps, il serait intéressant d'intégrer les modèles de lithographie type OPC (correction des effets de proximité de l'anglais Optical Proximity Correction), afin de prendre en compte la déformation restante notamment liée à la taille du faisceau, ou encore les effets chimiques des résines utilisées. Il est possible d'utiliser les modèles actuels déjà développés, mais leur précision est aujourd'hui restreinte à une tolérance d'erreur de 3 nm. Pour la rugosité, il faudrait descendre cette limite afin d'aller le plus loin possible dans les hautes fréquences.

IV.4.3 Prise en compte en amont de la rugosité naturelle

Nous avons pu voir que les échantillons créés sur plaque comportent, sur la PSD de rugosité de ligne, à la fois la rugosité programmée mais aussi celle naturellement induite par le procédé. Pour retrouver la rugosité programmée, nous avons dû soustraire la PSD de rugosité présente sur un échantillon de référence. Nos échantillons n'ont donc pas, au final, la rugosité voulue et ne peuvent donc pas être considérés comme des étalons de rugosité. Pour certaines techniques, ceci n'est pas un problème comme on l'a vu avec le CD-SEM (Chapitre III) puisqu'il est possible par référencement, de distinguer la rugosité induite de celle du procédé. Cependant, d'autres techniques comme l'OCD ont une sensibilité non linéaire et peuvent, par exemple, détecter une variation de rugosité sur une valeur nominale de rugosité importante, mais pas sur une valeur nominale faible [82]. Or, dans notre cas, les amplitudes induites sont additionnées à la rugosité naturelle. Il n'est donc pas possible de calibrer et de qualifier correctement la réponse de ce type de technique avec des étalons de ce genre.

Pour pallier à cela, il est possible de prendre en compte la rugosité naturelle, en amont de la programmation. Le principe est de caractériser une première fois le procédé, afin d'en extraire la rugosité dite de référence. Par la suite, dans le cas de génération PSD, la distribution obtenue est soustraite à celle voulant être créée. On se base ainsi sur cette nouvelle distribution pour la génération des signaux de rugosité de bords. Lors de l'écriture des motifs sur plaque, la rugosité dessinée va être additionnée à la rugosité naturelle, on retrouvera ainsi la rugosité que l'on souhaite programmer.

Une limite importante est l'amplitude minimale souhaitée. En effet, lors de la programmation de rugosité, il sera impossible par cette méthodologie de programmer une puissance de rugosité pour une fréquence donnée (si l'on se trouve dans le domaine fréquentiel), inférieure à celle induite par le procédé lui-même. On retiendra que la rugosité naturelle n'est pas supprimée par cette méthode, mais prise en compte et englobée dans la programmation. Par ailleurs, pour effectuer ce type de correction en amont, il est nécessaire que les procédés (incluant les outils de lithographie) soient les plus stables possibles afin que la rugosité naturelle ne change pas entre la première réalisation (de référence) et la création des étalons de rugosité.

Par ailleurs, cette méthode visant à améliorer la programmation de rugosité, nécessite d'être confiant dans la mesure de la rugosité de référence, ce qui nous amène à la dernière amélioration pour le contrôle de nos étalons, à savoir le choix de la métrologie.

IV.4.4 La métrologie de référence

Nous avons pu voir précédemment que nous utilisons la technique CD-SEM pour référence, avec la méthodologie d'extraction PSD. Cette méthodologie est en accord avec la programmation de rugosité

faite. Cependant aujourd'hui, au vu des travaux explicités ici, nous avons montré que la technique SAXS peut extraire la PSD, notamment de manière plus résolue sur les hautes fréquences. Malgré tout, cette technique n'étant pas entièrement développée, l'extraction du paramètre σ nécessite l'utilisation d'une autre technique comme le CD-SEM. Or, dans le cas de la méthodologie décrite dans la partie précédente, la programmation de rugosité est dépendante de la qualité de mesure de la rugosité naturelle du procédé. Ainsi, une meilleure mesure de cette référence mène à une meilleure programmation de la rugosité. On notera aussi que la technique OCD ne peut être utilisée comme référence puisqu'elle ne permet pas d'obtenir les informations fréquentielles liées à la rugosité.

De manière plus générale, si l'on regarde l'Equation IV.4, plus la fonction métrologique sera réduite, plus l'on pourra caractériser avec précision la fonction lithographique qui déforme nos motifs, la rugosité induite par le procédé, et ainsi les prendre en compte en amont. Ceci nous ramène donc à la nécessité d'une méthode de métrologie la plus aboutie possible.

IV.5 Résumé du chapitre

Au cours de ce chapitre, nous avons pu voir qu'il est possible de réaliser des échantillons se rapprochant d'étalons de rugosité. En effet ces réalisations, bien qu'impactées par de multiples facteurs reflétant la chaîne complète d'écriture et de mesure, comportent bien les variations liées à notre paramétrage. Ces résultats sont prometteurs mais nécessitent encore des améliorations.

Pour chaque étape de la création de ces échantillons, les défauts ont été ciblés. Nous avons associé à chacun d'eux les limites, physiques ou mathématiques, posant problème. En réponse à cela, nous avons proposé des alternatives ou avancées permettant de meilleures réalisations finales, tout en expliquant les limites de ces solutions. Grâce à ce type d'échantillons, il est possible de comparer les techniques avec plus de détails, mais aussi de prendre en compte les variations d'une technique à l'autre pour se rapprocher d'une valeur plus juste et plus fidèle. Ceci servira, pour la suite de ces travaux, à valider l'hybridation de données.

Par ailleurs, nous avons montré qu'il est possible d'obtenir des échantillons dits virtuels qui peuvent, au travers de différentes simulations, refléter le caractère des métrologies utilisées, notamment ici le CD-SEM. Ces échantillons virtuels seront des éléments essentiels de notre méthodologie lors du passage par les réseaux de neurones, explicité dans le chapitre suivant.

Chapitre V - Hybridation et simulations

Nous avons vu précédemment les avantages et les limitations des différentes techniques de métrologie. L'hybridation de données, que nous mettons en place dans ce chapitre, a pour but de garder le bénéfice de chacune des techniques, tout en supprimant ou réduisant leurs contraintes.

Nous allons donc aborder les principes de base de l'hybridation de données et étudier de quelle manière ce type de méthodologie a déjà fait ses preuves dans le domaine de la microélectronique, pour différentes métriques et applications.

Dans une première partie, deux techniques d'hybridation seront présentées, nous expliquerons notre choix pour celle finalement retenue, qui sera alors plus précisément détaillée afin de bien comprendre son fonctionnement ainsi que ses limites. Ceci nous mènera aux choix faits sur l'application de cette méthodologie d'hybridation, vis-à-vis des différentes données pouvant être extraites des techniques de métrologie considérées dans ces travaux. Ensuite, sur la base d'une génération de données aléatoires, les paramètres principaux liés à la méthodologie pour cette application, seront présentés et leur valeur optimale recherchée.

Enfin, la simulation des différents équipements de métrologie sera présentée comme point d'entrée nécessaire à l'hybridation, l'apprentissage futur sur ces données simulées restant à effectuer afin de permettre l'application de cette approche dans des cas réels.

V.1 Principe de base de l'hybridation

Comme vu en introduction générale, l'hybridation peut permettre la prise en compte de diverses mesures, que ce soit autour de paramètres identiques ou différents. Il n'existe donc pas un seul et unique type d'hybridation de données, on les différencie alors suivant plusieurs critères tels que décrits ci-dessous.

V.1.1 Le lien entre les données

Sur ce critère de lien entre les données, trois catégories distinctes peuvent être énoncées : complémentaires, coopératives, et redondantes [54, 115].

Données complémentaires

Les données obtenues par les capteurs sont celles d'un même objet étudié, mais vu de manière différente afin d'obtenir une vue plus globale. Un exemple est proposé en Figure V.1, deux caméras permettent de recréer, avec deux images séparées, une image plus grande comme un panorama. Ici, aucune donnée du premier capteur n'est introduite dans celles du second, on a un simple raccordement de données, contrairement au type suivant.

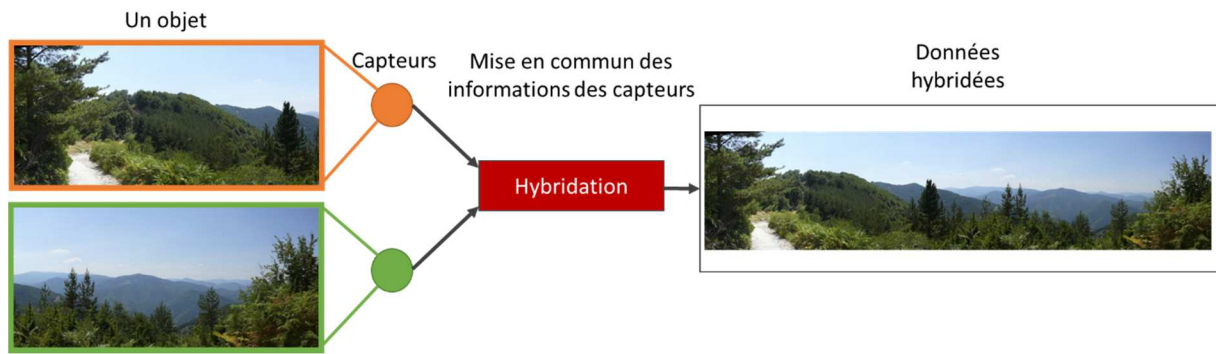


Figure V.1 : Principe d'hybridation de données complémentaires, exemple du panorama.

Hybridation coopérative

Cette stratégie permet d'obtenir une information plus grande que celle contenue dans les deux images. Par exemple dans le cas de la Figure V.2, une image sans couleur mais hybridée à une partie de la colorimétrie, permet d'obtenir une image entièrement colorisée même pour des couleurs non contenues dans la colorimétrie donnée. Il aurait été impossible de remonter à cela à partir de la donnée d'un seul capteur, ceci n'est pas le cas dans la dernière catégorie.

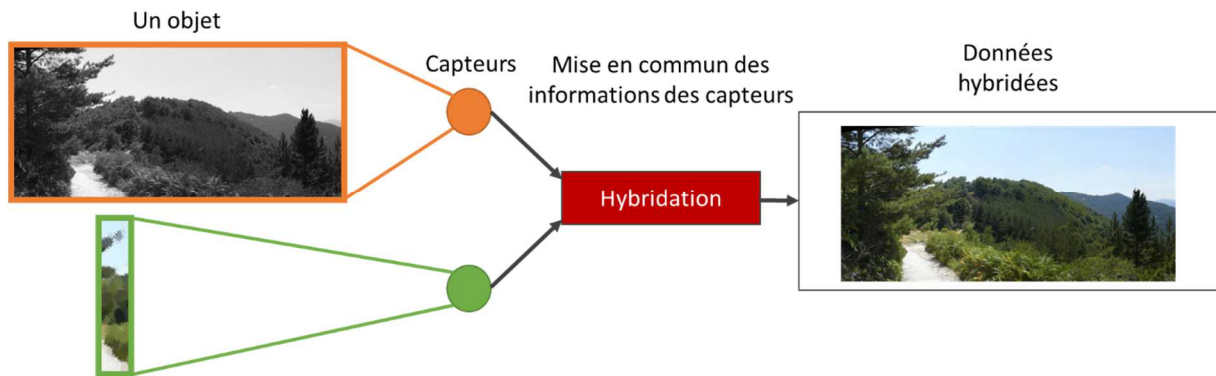


Figure V.2 : Principe illustré de l'hybridation coopérative, exemple d'une image résolue et complète en noir et blanc, et d'une image petite et peu résolue avec la colorimétrie.

Données redondantes

Comme le montre la Figure V.3, cette hybridation consiste à prendre la même information avec deux capteurs. Ainsi, le résultat obtenu contient une information identique et de même taille que celle des capteurs utilisés, et il n'y a pas ici d'augmentation de l'information totale, mais une amélioration de la fiabilité. Cette stratégie peut permettre de réduire les biais, ou d'obtenir par exemple un système avec une détection d'erreur. Ce type d'hybridation peut ainsi être utilisé pour des cas sensibles tels que des systèmes de vote, ou encore des applications liées à la sécurité.

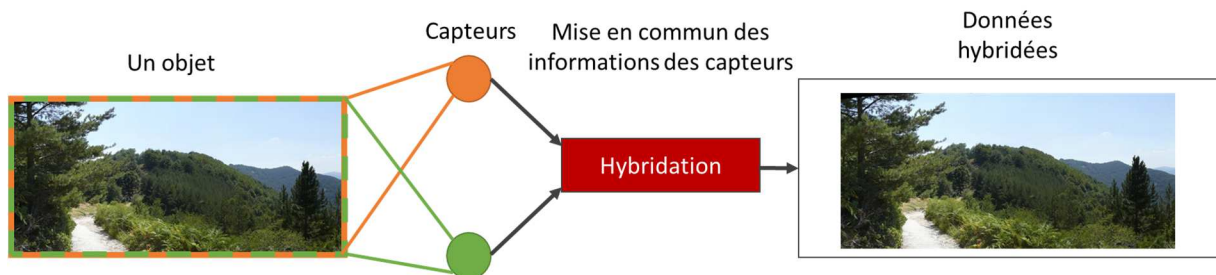


Figure V.3 : Illustration de l'hybridation de données redondantes.

Nous venons donc de voir qu'il est possible de caractériser l'hybridation suivant le lien entre les données sources. Nous avons illustré ces différences avec des images, mais cela peut être mis en place pour de multiples types de données. Ainsi, il est possible de classer l'hybridation suivant le type de données d'entrée et de sortie.

V.1.2 Type de données

Trois types de données peuvent être traités par hybridation. Le premier concerne les données brutes i.e., ce que lit et transmet directement le capteur, il peut s'agir de chiffres, de lettres, de couleurs etc. Le second type de données, caractéristiques ou « Feature » en anglais, correspond à une description ou une extraction faite à travers les données de base. Pour exemple, il peut s'agir d'une métrique extraite d'une image. Enfin, il est possible de traiter des décisions, une donnée déjà pré-analysée et dont on a déjà un résultat binaire, ou au moins catégorisé.

Ces types de données sont tous traitables comme entrées d'hybridation, la sortie étant soit du même type, soit de type supérieur proche (brute < feature < décision). On obtient donc, d'après la classification de B.V. Dasarathy [116], 5 cas différents d'hybridation, Entrée → Sortie :

- **Brute → Brute**
- **Brute → Feature**
- **Feature → Feature**
- **Feature → Décision**
- **Décision → Décision**

Ceci nous amène au dernier cas de classification du type hybridation, assez proche de celui par type de données mais prenant en compte plus de paramètres, qui est la classification par niveau d'abstraction.

V.1.3 Niveau d'abstraction

Cette classification a été explicitée suivant divers modèles tels que celui de R.C. Luo [117], allant d'un niveau « signal » à celui « symbole ». Elle est généralisée par quatre niveaux, proches de la classification de type de données [115], qui sont explicités ci-dessous :

Bas : données brutes en entrée pour obtenir, en sortie, des données brutes.

Moyen : caractéristiques en entrée, donnant des caractéristiques en sortie.

Haut : décisions en entrée et en sortie.

Multiple : les données d'entrée peuvent être mixtes et, en sortie, la donnée est de type supérieur ou égal à celui de la plus haute des entrées.

C'est bien ce dernier niveau qui fut la grande nouveauté de cette classification puisque l'on interprète des données pouvant être variées, ce qui n'est pas pris en compte dans la classification par type de données.

Une autre classification d'hybridation répandue, aussi sur la base de niveaux d'abstractions, est nommée « JDL Data Fusion Classification ». Cette catégorisation prend en compte de multiples facteurs, à savoir le type de données d'entrée mais aussi l'impact humain sur celles-ci et, enfin, la manière dont elles sont assemblées. Les niveaux d'abstractions ne prennent donc pas seulement en compte le type de données d'entrée/sortie, comme dans les cas précédents, mais aussi la manière

dont les données sont traitées [115]. Ceci nous amène au dernier cas, la classification suivant l'architecture déployée.

V.1.4 L'architecture d'hybridation

On trouve dans la littérature [115] quatre grands types d'architectures pouvant être déployés pour faire de l'hybridation de données :

- **Centralisée** : comme le montre la Figure V.4, toutes les données convergent vers un seul et même centre d'hybridation. Ceci nécessite le transfert et le traitement de beaucoup de données en même temps.

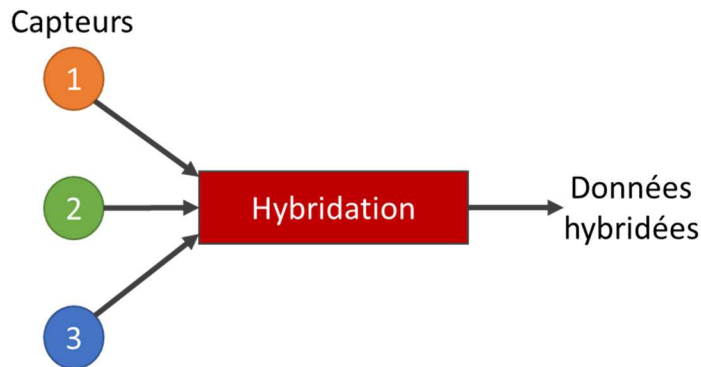


Figure V.4 : Architecture de type centralisée.

- **Distribuée** : avant d'être introduites dans un seul centre d'hybridation, les données sont post-traitées individuellement. Cette architecture a l'avantage du prétraitement des données qui empêche la surcharge d'une seule hybridation en termes de données, mais ceci peut induire une perte d'information due à la vue, seulement locale, de l'objet à ce moment-là.
- **Hiérarchique** : les données sont hybridées sur un nombre réduit de sources, avant une hybridation finale. Un exemple est donné en Figure V.5, avec une première hybridation des données provenant des capteurs 1 et 2 ; ensuite l'hybridation est faite sur ce résultat et les données fournies par le capteur 3, dans ce cas les hybridations en amont permettent de décharger l'hybridation globale mais aussi d'obtenir une vue de l'amélioration des données au fil des hybridations.

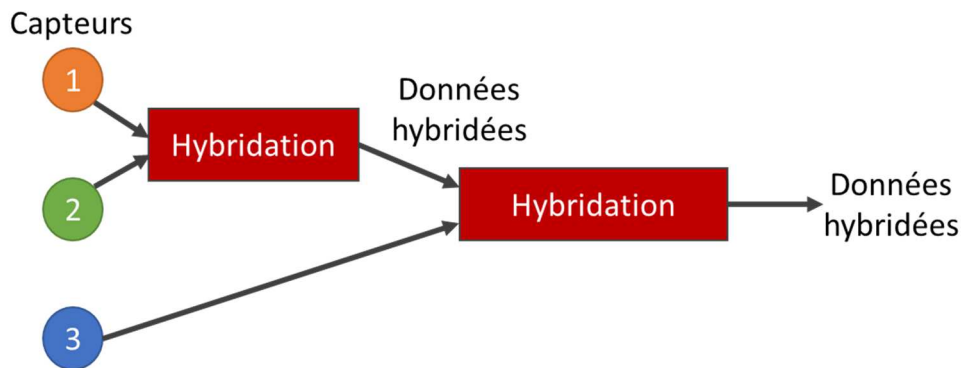


Figure V.5 : Exemple d'architecture d'hybridation hiérarchique.

- **Décentralisée** : dans ce cas il s'agit d'un réseau composé de multiples points d'hybridation localisés. Il y a donc plusieurs résultats d'hybridation disponibles au même moment, et aucun n'a la connaissance de tout le réseau. Ce type d'architecture est la plus compliquée à mettre en place du fait d'un grand nombre de voies de communications et des traitements importants qui sont nécessaires.

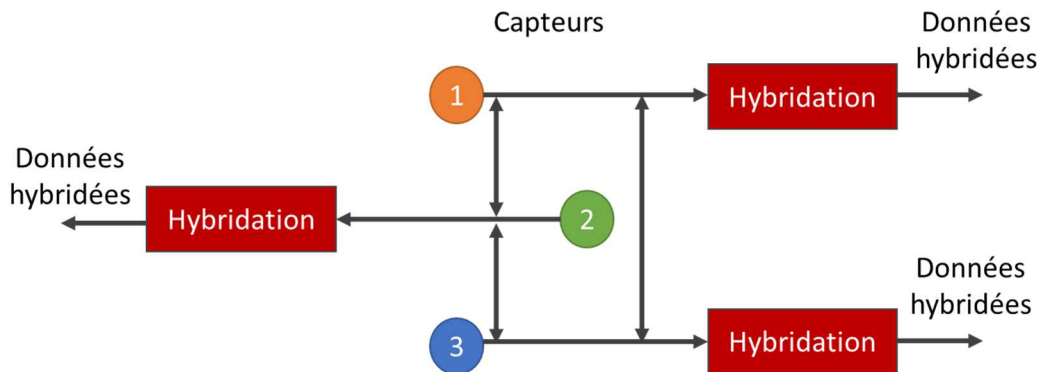


Figure V.6 : Représentation d'une architecture d'hybridation décentralisée.

En somme, une même hybridation peut être classée dans diverses catégories selon l'angle d'observation, ceci allant des données à l'architecture. Cela va, en grande partie, dépendre de la problématique et des contraintes matérielles et logicielles. Associées à ces catégories, il est possible de mettre en place diverses techniques d'hybridation, ce point est présenté dans la section suivante.

V.2 Les techniques d'hybridation

Il existe un grand nombre de techniques d'hybridation. Elles peuvent être basées sur une connaissance préalable telle que l'incertitude des capteurs utilisés, ou encore leur biais. *A contrario*, ces techniques peuvent utiliser une référence, comme un équipement jugé plus juste [53] ou des simulations [118]. Nous détaillons ici seulement deux des techniques possibles. Tout d'abord, le réseau bayésien, car cette technique a fait l'objet de multiples travaux antérieurs et publications dans le domaine de la microélectronique [54, 119–121] ; ensuite le réseau de neurones qui a aussi été utilisé dans ce secteur, dans un but de résolution de problème inverse par exemple pour la scatterométrie [122–124].

V.2.1 Réseau Bayésien

V.2.1-A Principe théorique

Comme explicité dans la littérature [125, 120, 54, p. 141], cette technique d'hybridation se base sur le théorème de Bayes [126], à savoir :

$$P(Y | X) = \frac{P(X | Y) P(Y)}{P(X)} \quad (V.1)$$

où $P(Y | X)$ représente la probabilité de Y connaissant l'information X, $P(Y)$ la probabilité de Y, et $P(X | Y)$ la probabilité de Y connaissant X, et $P(X)$ la probabilité de X (utilisée pour la normalisation). Le principe est donc d'estimer la probabilité d'un résultat en connaissant une information, *a priori*. Ainsi le facteur de Bayes est défini par :

$$B = \frac{P(Y | X)}{P(Y | \bar{X})} \quad (V.2)$$

où $P(Y | \bar{X})$ est la probabilité de Y connaissant la probabilité « non X ». De manière simplifiée, on obtient une probabilité de résultat sur un système ; égale à la multiplication du facteur de Bayes par la probabilité originelle.

Dans le cas de fusion de données, si on considère deux sources de données (ou équipement), A et B, répondant chacune à une distribution gaussienne de la forme :

$$P(v | D) \propto e^{-\frac{(v-\mu)^2}{2\sigma^2}} \quad (\text{V.3})$$

où σ est l'écart-type, μ la moyenne estimée de la distribution pour les données D et v la valeur à évaluer, et si ces distributions sont indépendantes, il est possible de les combiner selon :

$$P(v | D_A, D_B) \propto P(v | D_A) \times P(v | D_B) \quad (\text{V.4})$$

avec D_A et D_B les données provenant respectivement des équipements A et B. Ceci donne par injection de leur forme gaussienne (Equation V.3) :

$$P(v | D_A D_B) \propto e^{-\frac{(v-\mu_A)^2}{2\sigma_A^2}} \times e^{-\frac{(v-\mu_B)^2}{2\sigma_B^2}} \propto e^{-\frac{(v-\mu_F)^2}{2\sigma_F^2}} \quad (\text{V.5})$$

(μ_A, σ_A), (μ_B, σ_B) et (μ_F, σ_F) étant la valeur moyenne et l'écart-type, respectivement obtenus pour les équipements A, B et leur hybridation F. En résulte alors l'expression des paramètres après hybridation comme :

$$\frac{1}{\sigma_F^2} = \frac{1}{\sigma_A^2} + \frac{1}{\sigma_B^2} \quad (\text{V.6})$$

$$\mu_F = \frac{\sigma_F^2}{\sigma_A^2} \mu_A + \frac{\sigma_F^2}{\sigma_B^2} \mu_B \quad (\text{V.7})$$

Il est notable, d'après l'Equation V.7, que l'hybridation bayésienne revient à une moyenne pondérée par l'incertitude des sources.

L'hybridation de données faisant partie des points ciblés par l'IRDS [11] pour le développement des futures métrologies, on retrouve cette technique pour la réponse à de multiples problématiques de la microélectronique, avec des développements ayant pour pilier l'approche bayésienne.

V.2.1-B Les applications en microélectronique

Un des premiers exemples en microélectronique, où l'hybridation a fait ses preuves, est l'accès à une mesure de CD de ligne plus fiable. Divers travaux ont été détaillés sur la prise en compte de couples de techniques tels que le SEM et l'OCD [121] ou l'AFM et OCD [127], et même plus récemment dans le cas de trois techniques i.e. : SEM, OCD et AFM [54, 120]. Dans tous ces cas l'extraction était effectuée sur la largeur de ligne, mais également sur d'autres paramètres comme le SWA ou encore la hauteur de ligne. C'est à la fois la prise en compte de techniques différentes, mais aussi de plus de paramètres, qui ont permis d'obtenir une amélioration importante de la mesure.

De plus, l'hybridation par l'approche de réseaux bayésiens a aussi été effectuée sur des structures autres que des lignes. En effet, les travaux de R.M. Silver et al. [119] ont porté sur des structures de type contact. Il s'agit là d'un réseau de trous suivant les deux directions X et Y. La spécificité notable de ces recherches est la prise en compte des quatre techniques identiques à celles des travaux présentés ici, à savoir : SEM, AFM, OCD et SAXS. Le résultat exposé montre qu'après hybridation, l'incertitude de mesure de CD est réduite de moitié, approchant la limite du nanomètre.

Par ailleurs, on notera l'utilisation d'une approche dérivée de celle bayésienne (nommée Gaussian Process) dans une application plus proche de notre problématique, à savoir la rugosité de surface [128]. Ces travaux montrent qu'avec la technique AFM et la microscopie à balayage laser, les erreurs sur l'extraction des paramètres de rugosité peuvent être réduites jusqu'à une limite basse de 1.5%.

Enfin on peut citer, pour l'extraction de rugosité LER, des travaux de fusion comme ceux de A. Vaid [129] par CD-SEM, OCD et des références AFM et TEM. On note plus particulièrement que ces résultats ne se basent pas sur les réseaux bayésiens mais sur une approche dérivée de celle-ci, du fait d'un grand nombre de prérequis nécessaires pour cette technique d'hybridation.

V.2.1-C Les prérequis de la technique

Tout d'abord, les réseaux bayésiens supposent d'avoir des données qui répondent à des distributions gaussiennes. Ce prérequis est à l'origine du formalisme de l'Equation V.3 menant à la forme des Equations V.6 et V.7. C'est aussi un des points fondamentaux de la technique car il permet, entre autres, de répondre au second prérequis, i.e. la connaissance de l'incertitude des mesures. Dans ce cas, l'incertitude est estimée comme étant la dispersion de chacune des distributions gaussiennes. Ces prérequis nécessitent souvent des premières études, en amont, pour chaque système considéré. On note par exemple le développement de stratégies de suppression de tendances, liées aux objets ou caractéristiques mesurées, ou encore la suppression de données aberrantes [54], permettant ainsi de retrouver des distributions gaussiennes. Appliqué à notre cas qui est basé sur l'ajustement d'un modèle PSD sur notre extraction, ce retour à une distribution gaussienne n'est pas trivial à réaliser, car l'ajustement tend toujours à minimiser l'erreur quadratique et que la variation des différents paramètres n'est pas entièrement décorrélée. C'est là une des limites de l'utilisation des réseaux bayésiens à notre problématique.

Le dernier prérequis est l'absence de biais entre les techniques. Ceci est toutefois quasi impossible dans les cas réels, et des méthodes de suppression de biais ont vu le jour [120]. Cependant ce type de stratégie suppose de choisir une technique comme référence, et ce principe représente un non-sens dans le cas de la rugosité de ligne. Même si une technique donne une valeur plus juste que les autres dans un cas spécifié, ce n'est pas forcément le cas dans toute la gamme des fréquences mesurées ou sur tous les paramètres à extraire. Par exemple dans le cas de l'AFM, la valeur du paramètre de rugosité σ est considérée comme juste, cependant la puissance obtenue en haute fréquence est fortement détériorée par l'impact de la taille de la pointe, explicité au Chapitre III. Ainsi dans notre cas, aucune technique ne peut être considérée comme référente. Une alternative, proposée dans les travaux de N. Rana [118], porte sur l'utilisation de simulations des métrologies à partir du dessin théorique des lignes. Cette méthodologie permet d'appréhender le biais de chacune des techniques et de le prendre en compte tel un apprentissage. On voit alors, ici, une convergence avec une autre technique permettant l'hybridation des données, qui est l'utilisation des réseaux de neurones, objet du paragraphe suivant.

V.2.2 Les réseaux de neurones

Comme son nom l'indique, la technique des réseaux de neurones correspond à l'assemblage de parties élémentaires au travers de liaisons de communication, à l'instar des neurones composant, entre autres, le système cérébral humain. Avant la mise en réseau, il nous faut cependant définir la notion du neurone unique.

V.2.2-A Neurone biologique ou mathématique

Les premiers travaux [130] des années cinquante sur cette technique, ont donné le nom de « neurone » à ce concept mathématique alors créé parce qu'il s'inspire du fonctionnement d'un neurone biologique. Le neurone biologique schématisé en Figure V.7 est composé de quatre parties :

Les dendrites : réceptrices des divers signaux parvenant au neurone.

Le corps cellulaire avec le noyau : il traite localement les signaux pour n'en renvoyer qu'un.

L'axone : il permet de propager le signal généré par le corps cellulaire.

Les synapses : interfaces avec les muscles auxquels elles vont donner l'information sous forme de signal électrique. Les synapses peuvent être activatrices ou inhibitrices de signal, telle une pondération.

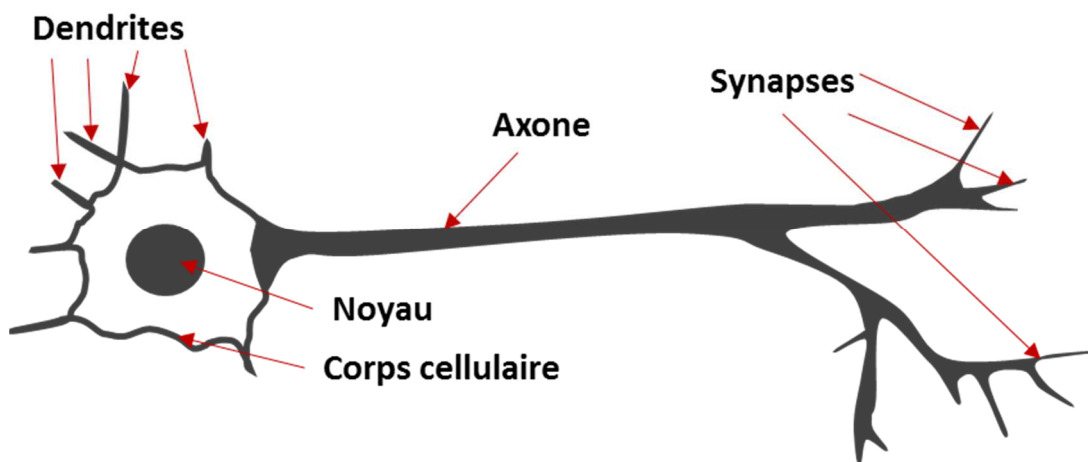


Figure V.7 : Schématisation d'un neurone biologique.

Ces diverses parties peuvent se résumer en 3 termes : « entrées », « traitement » et « sorties » avec un aspect directionnel. C'est exactement ce que l'on retrouve dans le neurone artificiel mathématique. En effet, comme le présente la Figure V.8, ces parties sont représentées de gauche à droite avec les diverses entrées qui passent dans un cœur de traitement, et finalement en résulte une sortie.

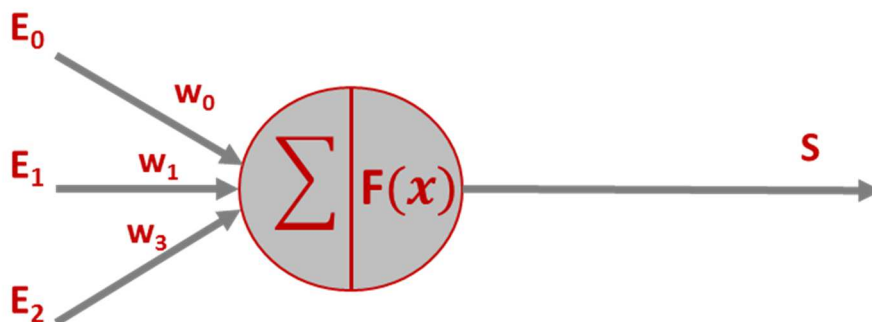


Figure V.8 : Représentation d'un neurone artificiel mathématique.

Quelques différences de représentation, entre la biologie et les mathématiques sont tout de même à noter [122, 131–133]. Dans le cas des neurones artificiels, la pondération est faite en entrée et on considère une sortie unique. De plus, la représentation mathématique du traitement est effectuée en deux parties. La première est la somme des entrées pondérées :

$$v = \sum_{i=1}^E w_i e_i \quad (\text{V.8})$$

où v est le potentiel du neurone, e_i les différentes valeurs d'entrée associées à leur poids w_i . Parfois une entrée supplémentaire est considérée comme un biais directement apporté par le neurone avec un poids égal à 1. Dans notre cas, la méthode de prise en compte des entrées implique qu'il s'agit d'un neurone linéaire, il en existe d'autres types (neurone polynomial, neurone de type noyau, etc.) [122].

Une fois le potentiel du neurone calculé, il va être traité par le biais d'une fonction d'activation F , dont on obtiendra la sortie s du neurone telle que :

$$s = F(v) = F\left(\sum_{i=1}^E w_i e_i\right) \quad (\text{V.9})$$

Il existe une multitude de fonctions d'activation, les plus courantes sont représentées dans la Table V.1 avec leur équation respective. On note qu'elles sont toutes continues et différentiables (sauf la fonction binaire). Les plus utilisées aujourd'hui, d'après la littérature [134–136], sont les fonctions à base sigmoïdale (Sigmoide et TanH) car elles introduisent un caractère non linéaire au modèle.

Les bases d'un seul neurone étant posées, on peut comprendre que le neurone artificiel seul peut répondre à certaines problématiques simples, telles qu'un classement par un seuillage (sortie supérieure ou non, à une valeur). Cependant, la force de ce système depuis quelques années est la possibilité d'assembler plusieurs neurones entre eux, c'est ce principe que nous allons explorer dans la section suivante.

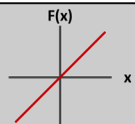
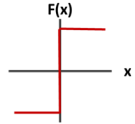
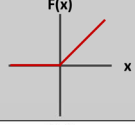
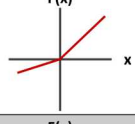
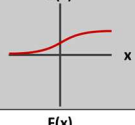

Nom	Représentation	Equation
Identité		$F(x) = x$
Binaire		$F(x) = \begin{cases} 0 & \text{si } x < 0 \\ 1 & \text{si } x \geq 0 \end{cases}$
Linéaire Rectifié (ReLU)		$F(x) = \begin{cases} 0 & \text{si } x < 0 \\ x & \text{si } x \geq 0 \end{cases}$
Linéaire Rectifié paramétrique (PReLU)		$F(x) = \begin{cases} ax & \text{si } x < 0 \\ x & \text{si } x \geq 0 \end{cases}$
Logistique ou Sigmoide		$F(x) = \frac{1}{1 + e^{-x}}$
Tangente Hyperbolique (TanH)		$F(x) = \frac{2}{1 + e^{-2x}} - 1$

Table V.1 : Fonctions d'activation les plus courantes [137, 138].

V.2.2-B Principe du Perceptron Multi-Couches [124]

Le perceptron est un neurone artificiel capable d'apprendre ; ceci se traduit de manière technique par une boucle de rétroaction qui permet de changer le poids de chaque entrée du neurone.

Il existe une multitude d'architectures pour assembler les neurones entre eux. On retrouve, dans la littérature, les explications détaillées sur la construction de ces architectures et leurs utilités [122, 131]. Dans le cadre de nos travaux, nous avons utilisé une architecture multi-couches au vu du type de données à traiter (Chapitre II). Cette architecture est techniquement appelée Perceptron Multi-Couche (PMC). Comme le montre la Figure V.9 et comme son nom l'indique, elle est composée de différentes couches de neurones avec pour règle principale que les neurones d'une même couche ne soient pas reliés entre eux. D'autre part, dans le cadre formel du PMC, chaque neurone d'une couche est relié à tous les neurones de la couche précédente et de la suivante. Il n'y a pas de saut de couche ni de connexion manquante.

La première couche, à gauche, est celle nommée « entrée », elle comporte autant de neurones que le nombre de paramètres d'entrées. De la même manière, la couche représentée le plus à droite Figure V.9, est celle de sortie et a autant de neurones que de sorties attendues. Les couches intermédiaires sont nommées couches cachées. Leur nombre L ainsi que le nombre de neurones qui composent chacune d'elle N_x (x étant le numéro de la couche cachée) est la clef de la résolution par réseau de neurones. Ces paramètres sont dits « hyper-paramètres », ils doivent être déterminés suivant chaque application. De plus, on notera que chacune des couches peut avoir sa propre fonction d'activation, mais cette dernière reste la même pour tous les neurones d'une même couche. On voit, de par sa structure, que cette architecture a un sens de propagation de l'information de gauche à droite dans le cas de la représentation en Figure V.9. Lorsque le réseau est utilisé, aucune information ne transite dans l'autre sens.

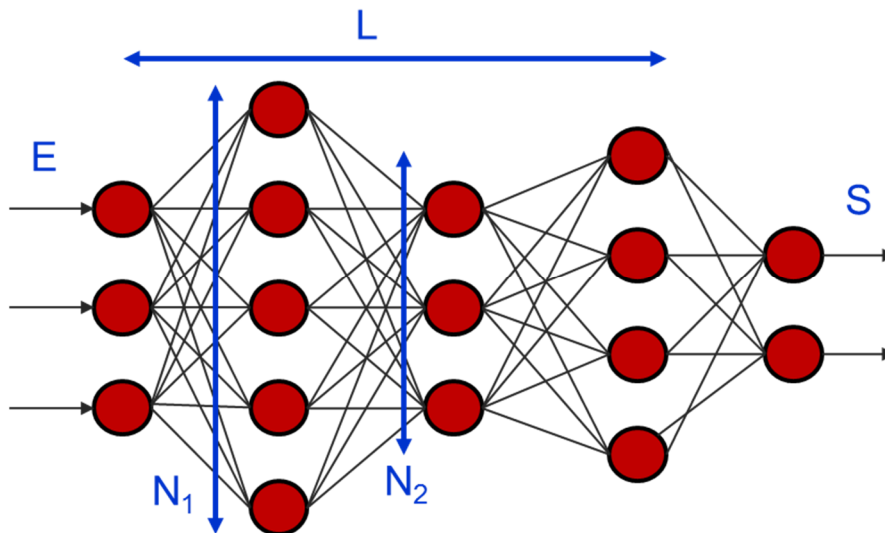


Figure V.9 : Représentation schématique d'une architecture multi-couches.

Cette architecture de réseau a été utilisée dans le domaine de la micro-électronique et, plus particulièrement pour la scatterométrie, à de multiples reprises [122, 124, 131]. Les problématiques liées à la détermination des hyper-paramètres et à la notion d'apprentissage se retrouvent dans chaque développement de réseau de neurones, suivant l'application abordée. Ces problématiques font l'objet de la partie suivante.

V.2.2-C Utilisation et problématique du PMC

Les étapes dans l'utilisation d'un PMC

L'utilisation d'un PMC se déroule temporellement en trois étapes :

- **L'apprentissage** : il s'agit de fournir au réseau de neurones un certain nombre de données d'entrées dont les résultats théoriques finaux sont connus. Celles-ci vont traverser le réseau qui va calculer les sorties, suivant une initialisation aléatoire. Ces résultats sont comparés aux résultats théoriques, ainsi une erreur est calculée et peut donc être répercutée sur chaque couche du réseau afin de déterminer le poids correct de chaque connexion. On parle de rétropropagation de l'erreur. Ce processus est effectué de manière itérative, afin que le réseau puisse apprendre sur tout l'espace des valeurs à explorer. On peut faire l'analogie de ce processus avec le calcul d'un fit polynomial sur un jeu de données, avec pour but de fixer le bon paramétrage. Les algorithmes de rétropropagation sont explicités dans les travaux de I. Gereige [122].
- **La validation** permet de calculer le résidu final (« Loss » en anglais) soit l'erreur restante en sortie du réseau. Pour cela, on utilise aussi des données dont on connaît les résultats théoriques, mais différentes de celles utilisées pour l'apprentissage. Pour un jeu de données connues, il faut donc établir un ratio permettant l'apprentissage puis la validation du réseau. Le ratio le plus souvent cité est de 80% des données pour l'apprentissage, et 20% pour la validation.
- **L'utilisation** : durant cette phase le réseau est fixé. Il n'y a plus lieu de changer les poids sur les connexions. On note cependant que le réseau a appris sur un domaine de données défini, il doit donc être utilisé sur ce même espace. On retrouve ce concept sur la partie suivante qui concerne la normalisation des données.

La normalisation

Dans le cadre du réseau de neurones, il est conseillé de normaliser les données des entrées. Ceci n'est pas une obligation mais, si les données ont des amplitudes très variées, il faudra que les poids sur les différentes connexions soient aussi très variés, ce qui déséquilibrerait rapidement le réseau. On remarque que les logiciels, ou bibliothèques, de mise en place de réseaux de neurones, indiquent aussi cette recommandation. Nous effectuons donc la normalisation entre 0 et 1 à partir du set d'apprentissage des données, et ce sur chacune des entrées, suivant :

$$E_{Ni} = \frac{E_i - \text{Min}(E)}{\text{Max}(E) - \text{Min}(E)} \quad (\text{V.10})$$

avec E_{Ni} la donnée d'entrée normalisée pour E_i , E la totalité des données pour une entrée. De manière inverse, la dénormalisation se fait grâce au même coefficient, suivant :

$$S_i = S_{Ni} \times (\text{Max}(E) - \text{Min}(E)) + \text{Min}(E) \quad (\text{V.11})$$

où S_i est la sortie dénormalisée de la sortie S_{Ni} .

On voit clairement ici que le réseau a appris sur une plage de données [minimum, maximum] définie. Chercher une donnée au-delà de ces limites revient à une extrapolation, ce qui n'est pas toujours fondé sur des modèles non linéaires. On note qu'il existe d'autres normalisations prenant en compte, par

exemple, la moyenne et l'écart-type du jeu de données, mais cela suppose que ces dernières suivent une loi gaussienne ce qui n'est pas toujours le cas.

Les hyper-paramètres

Comme dit précédemment, les hyper-paramètres sont le centre névralgique des réseaux de neurones. C'est grâce à eux que le réseau fonctionnera à partir des données d'apprentissage et permettra de traiter l'information. On notera, tout d'abord, que l'ajout d'un perceptron dans une couche implique l'ajout d'un nombre de connexions et donc de poids associés équivalant à la somme du nombre de perceptrons de la couche précédente et la suivante. Le fait d'avoir un trop grand nombre de paramètres, à l'intérieur même du réseau, est équivalent à un degré de liberté trop important. On parle alors de « sur-ajustement », pour illustration cela revient à augmenter le degré d'un polynôme pour passer par tous les points de données, quitte à déformer complètement la réponse de la fonction. Ainsi on veillera, dans le cas où deux réseaux auraient un résultat de perte similaire, à toujours favoriser celui ayant un nombre de paramètres plus faible.

Le premier hyper-paramètre est le nombre L de couches cachées du réseau. Les publications les plus anciennes font état de peu de couches cachées, avec des conclusions stipulant que seulement deux couches cachées suffiraient à résoudre toutes les problématiques, incluant celles non linéaires [139]. D'autres publications vont même jusqu'à dire qu'une seule couche cachée, comportant assez de neurones [140], serait suffisante. *A contrario*, les dernières recherches montrent qu'une quinzaine de couches est plus favorable [141], mais ces travaux traitent des données brutes de gros volumes (images), contrairement aux cas cités précédemment d'une ou deux couches. De manière plus générale, on remarque un lien entre le nombre de couches et le nombre de perceptrons par couche, le second hyper-paramètre.

Le nombre de neurones ou perceptrons, par couche, a été plusieurs fois étudié avec des résultats qui divergent souvent. On retrouve donc diverses recommandations : des couches ayant une taille identique à celle de la couche d'entrée, ou à celle de la couche de sortie, ou à 0.75% de la taille de la couche d'entrée, ou encore à la racine du produit du nombre d'entrées par le nombre de sorties [142]. De la même manière, on retrouve une multitude de recommandations contradictoires sur le choix de la fonction d'activation. En effet ce dernier hyper-paramètre, constitué par la ou les fonctions d'activation, est très controversé également. On rappelle que cette fonction peut être différente sur chacune des couches. Ainsi la mise en place d'une fonction, ou plus, peut dépendre de la complexité des données mises en commun, mais aussi du but de chaque couche.

En résumé, on note que ces hyper-paramètres sont fortement liés à l'application mise en jeu et qu'il n'existe donc pas de règles spécifiées pour les déterminer. Le test empirique reste aujourd'hui l'approche la plus commune. Cependant, on voit apparaître de plus en plus d'études portant sur l'explication, au travers de leur application, du choix du nombre de couches cachées, de perceptrons, ou encore des fonctions d'activation [143]. Dans la section suivante est explicitée l'application du PMC, dans le cadre de notre problématique, afin de déterminer les hyper-paramètres de notre réseau.

V.3 Application du PMC au cas d'étude

Dans le cadre de ces travaux, le but est de prendre en considération les quatre techniques disponibles afin d'obtenir une information sur la rugosité de ligne. Pour rappel, les techniques sont : le SEM, l'AFM, l'OCD et la SAXS qui fournissent, dans des conditions pouvant être différentes, des informations complémentaires ou identiques sur la rugosité. Il est nécessaire de définir un cadre d'utilisation du réseau de neurones. Nous nous sommes donc appuyés sur l'utilisation de la librairie Keras en Python, cette dernière pouvant se baser sur plusieurs sous-librairies de traitement, nous avons arbitrairement choisi de la faire interagir avec l'outil TensorFlow développé par Google. Le cadre étant posé, les premiers points à définir pour l'utilisation d'un réseau de neurones sont les données.

V.3.1 Les données d'entrées et de sorties

Dans le cas d'un premier développement appliqué à la rugosité de ligne, nous nous sommes basés sur une hybridation d'abstraction de niveau moyen. On souligne ici le fait que le réseau de neurones va donc être utilisé avec un but prédictif, i.e. sur l'obtention de valeurs, à l'inverse du cas de classification qui se rapporte plus à une sortie de type décision.

Le principe est d'entrer des données de rugosité déjà extraites, telles que le paramètre σ , et d'obtenir une caractérisation complète, soit les trois caractéristiques : σ , α et ξ . La Table V.2 résume l'accessibilité des différents paramètres de rugosité de ligne par les techniques mises en place (plus de détails disponibles au Chapitre III). Cette approche a pour but d'évoluer afin de se rapprocher d'une hybridation bas niveau, et donc de limiter les analyses intermédiaires qui ont leurs propres biais et erreur.

Techniques\Paramètres	σ	α	ξ
SEM			
AFM			
OCD			
SAXS			

Table V.2 : Représentation des caractéristiques accessibles par chacune des techniques prises en compte en vert = extraite, en orange = valeur proportionnelle, en rouge = non accessible en l'état.

Nous avons donc un système à 9 entrées (cases vertes et orange) et à 3 sorties. Connaissant cela, il est possible de tester les hyper-paramètres en générant de manière stochastique des données autour de valeurs théoriques. Pour ce faire, des données aléatoires ont été produites à partir des valeurs de paramètres mentionnées en Table V.3 et leur incertitude respective dans la Table V.4. Au total, plus de 210 cas sont générés cinquante fois, ce qui correspond à 10500 jeux de données, afin d'obtenir une population représentative de notre problématique globale.

Paramètres	σ (nm)	α	ξ (nm)
Min à Max / Pas	0,5 à 2,5 / 0,5	0,4 à 1 / 0,1	10 à 110 / 20

Table V.3 : Plage considérée des paramètres et leur pas de résolution.

Technique\Paramètres	σ (nm)	α	ξ (nm)
SEM	20	20	10
AFM	10	10	5
OCD	10		
SAXS		20	20

Table V.4 : Incertitude arbitraire sur chaque paramètre et pour chacune des techniques dans le cas d'une génération aléatoire.

Grâce à cette multitude de données, il est possible de tester les hyper-paramètres à utiliser dans le réseau de neurones.

V.3.2 Sélection des hyper-paramètres

Au vu de la problématique considérée avec des données sous forme de caractéristiques, nous restreindrons les recherches à un maximum de 3 couches cachées, limitant ainsi les temps de traitement. De même, au vu des différentes théories explicitées précédemment, le nombre de neurones par couche sera compris entre le nombre d'entrées, 9 et le nombre de sorties, 3.

Fonction d'activation

Lors de la sélection des hyper-paramètres, la plus grande difficulté est la relation entre ces derniers. Le fait de commencer par un paramètre ou un autre peut donc engendrer des résultats différents. Nous avons choisi, arbitrairement, de commencer par regarder quelle fonction d'activation utiliser en prenant une configuration fixe, à savoir : aucune couche cachée et la totalité des données, rangées par paquet de 200 jeux, pour l'apprentissage. Le résultat, présenté dans la Table V.5, nous montre que les erreurs sur les paramètres sont très importantes, celles-ci devraient être réduites au fur et à mesure de la détermination des hyper-paramètres. De plus, ce résultat nous oriente sur l'utilisation de la fonction « TanH » (voir la Table V.1 pour sa représentation et forme mathématique) pour la suite des études.

Fonction	Résidu total	σ	α	ξ
Relu	0.255	55%	37%	60%
Tanh	0.088	61%	27%	53%
Sigmoid	0.114	60%	28%	103%
Softmax	0.143	52%	26%	93%

Table V.5 : Erreur obtenue en termes de résidu et d'erreur sur les paramètres de sortie, lors de la validation pour diverses fonctions d'activation.

Nombre d'époques

L'apprentissage du réseau doit être fait sur un certain nombre d'itérations, nommées « époques ». Le nombre d'époques optimal correspond au nombre d'itérations nécessaires pour que l'erreur cesse de

diminuer. Pour déterminer ce nombre optimal, nous avons pris un réseau comportant seulement la couche d'entrée et celle de sortie, avec la fonction d'activation déterminée précédemment « TanH ».

Le résultat, présenté en Figure V.10, montre une stabilisation du résidu à partir de 50 époques. Plus exactement, au bout de la 50^{ième} itération, le résidu est de 0.0836 et de 0.0832 au bout de la 55^{ième} itération, soit une diminution de moins de 0.1% de le résidu par itération. Nous fixerons donc, pour la suite, le nombre d'époques à 50.

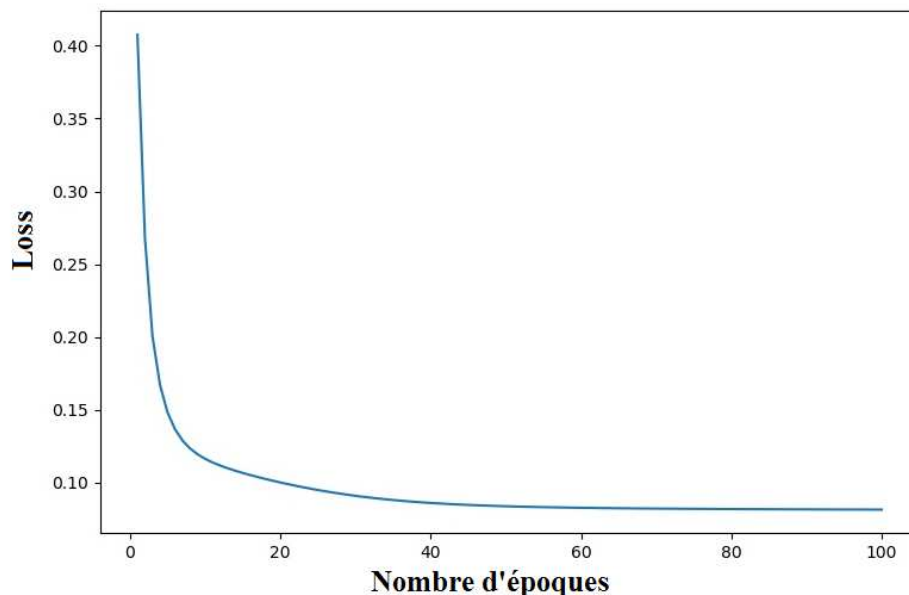


Figure V.10 : Evolution de la loss ou residu suivant le nombre d'iterations d'apprentissage (époques).

Taille des paquets de données

Pour réaliser les tests, il est nécessaire de définir pour le réseau une taille de paquets de données. Ceci correspond au nombre de cas différents d'entrées utilisés pour l'apprentissage, avant d'appliquer la rétroaction. Si le nombre de cas par paquet est trop petit, le réseau va orienter son résultat sur un minimum local ; à l'inverse, si la taille du paquet est trop importante le réseau n'arrivera pas à converger. On considère cette taille de paquet comme un hyper-paramètre présent uniquement lors de l'apprentissage, cependant il est tout aussi important que les autres hyper-paramètres. Pour choisir la taille de paquet, nous avons basé le réseau sur la configuration précédente, avec les paramètres prédéterminés, i.e. la fonction d'activation « TanH » et un nombre d'époques égal à 50.

La Figure V.11 montre le résultat obtenu en termes de résidu, pour différentes tailles de paquets. On observe un minimum pour une taille de paquet de 100 jeux de données, c'est donc cette valeur qui sera considérée dans la suite de la mise en place du réseau de neurones. On souligne le fait que les variations présentées sur cette même figure sont relativement faibles, cela étant dû au paramétrage précédent. Désormais, il faut considérer les derniers hyper-paramètres, le nombre de couches et leur taille respective.

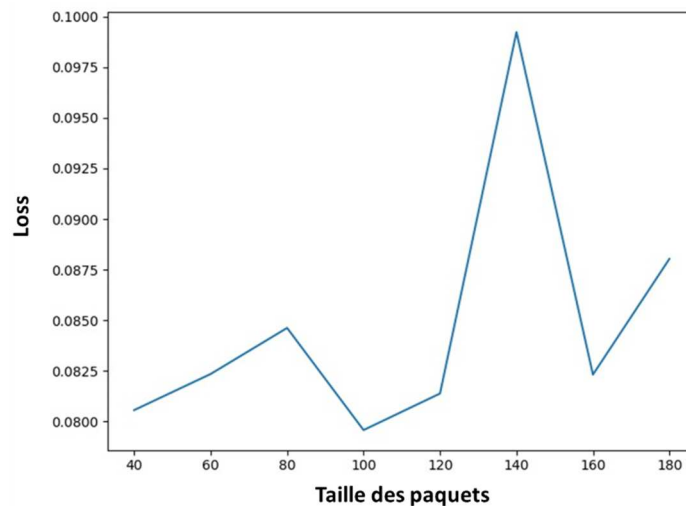


Figure V.11 : Evolution de la loss ou résidu, suivant la taille des jeux de données, sur une plage de 40 à 180 jeux par pas de 20.

Nombre de couches cachées et nombre de neurones par couche

Ces hyper-paramètres, déjà présentés précédemment, sont le cœur du réseau. Ils serviront non seulement lors de l'apprentissage, mais aussi par la suite lors de l'utilisation du réseau. Comme explicité précédemment, ces deux paramètres ne peuvent pas être considérés séparément et, pour rappel, certaines publications proposent peu de couches avec beaucoup de neurones, et d'autres l'inverse. Nous avons donc fait varier les deux paramètres en même temps.

Les résultats de ce test sont présentés en Table V.6. Le résultat de le résidu sans couche cachée est déductible de la Figure V.11, soit environ 0.080. Ces extractions montrent que les architectures ne se différencient que peu en termes de résidu. Au global il ressort seulement deux structures ayant un résidu plus faible que les autres. Ces dernières sont celles avec une et deux couches cachées, pour des tailles de 9 neurones, soit un nombre de neurones identique à la couche d'entrée. En regardant plus en détail le résultat de ces deux architectures, on remarque que les erreurs sur les paramètres sont aussi très proches ; le cas $L=1$ est meilleur que $L=2$ pour les paramètres σ et ξ mais, moins performante pour l'extraction de α . Dans ce cas-là, l'architecture $L=1$ sera privilégiée car elle comporte moins de degrés de liberté.

Afin d'optimiser encore les différents hyper-paramètres, il faudrait repartir de cette dernière architecture et chercher à nouveau chacun des paramètres pour se rapprocher d'un optimum. On soulignera, cependant, le fait que cet optimum sera local et qu'un autre système, entièrement différent, pourrait avoir des performances encore meilleures.

Par ailleurs, deux points n'ont pas été abordés dans cette étude. Le premier est le nombre de données minimum à traiter, cependant nous supposons au vu d'autres études de ce type [122], que le nombre pris est au-delà de celui nécessaire. Dans un second temps, la gamme de variation des paramètres et le pas de résolution sur celle-ci n'ont pas non plus été étudiés. Au vu du premier jeu de données, présenté dans la Table V.3, on pense qu'un pas de variation plus faible lors de l'apprentissage permettrait de gagner sur l'erreur finale, mais ceci reste à tester. Pour aller plus loin sur le sujet, nous avons préféré nous attarder sur la problématique de l'obtention de données plus réalistes, pour l'apprentissage.

L = 1

Nx	Résidu	σ	α	ξ
3	0.089	13%	5%	6%
5	0.096	9%	6%	10%
7	0.090	11%	6%	8%
9	0.074	10%	6%	4%

L = 2

Nx	Résidu	σ	α	ξ
3	0.087	11%	4%	10%
5	0.089	13%	4%	9%
7	0.085	13%	6%	4%
9	0.075	11%	5%	6%

L = 3

Nx	Résidu	σ	α	ξ
3	0.085	7%	6%	10%
5	0.087	13%	6%	4%
7	0.089	13%	6%	8%
9	0.086	10%	5%	9%

Table V.6 : Erreurs obtenues en termes de résidu et sur les paramètres de sortie, pour diverses architectures ayant un nombre de couches cachées L et un nombre de neurones, par couche, constant et de valeur N_x .

V.4 Vers des données plus proches des métrologies

Afin de faire l'apprentissage du réseau dans des cas au plus près de la réalité, sans pour autant devoir fabriquer expérimentalement et caractériser un nombre important d'échantillons, nous nous sommes tournés, comme dans le cas des travaux de N. Rana [118], vers un apprentissage grâce à des données simulées correspondant aux différentes métrologies considérées. Pour ce faire, il faut pouvoir effectuer la simulation de chacune des métrologies dites virtuelles, sur les mêmes échantillons virtuels. Pour rappel, un échantillon virtuel est représenté par une image du type de celle en Figure III.31. Dans les parties suivantes, nous expliquons donc les générations des paramètres issues des techniques utilisées et les paramètres clefs qu'il nous est possible, aujourd'hui, de modifier pour l'apprentissage. Les exemples de générations seront donnés sur le même échantillon virtuel présenté en Figure III.31.

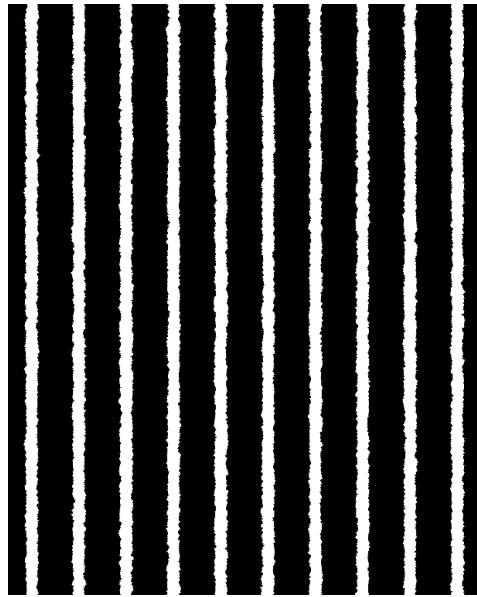


Figure V.12 : Exemple d'un échantillon virtuel $2 \times 3 \mu\text{m}^2$ comportant 10 lignes $\text{CD}=50 \text{ nm}$ $\text{Pitch}=200 \text{ nm}$, bords non corrélés, généré avec $\sigma = 3 \text{ nm}$, $\alpha = 0.5$, $\xi = 10 \text{ nm}$, $\Delta y = 1 \text{ nm}$.

V.4.1 Génération SEM

Comme explicité dans le Chapitre IV, ce type de génération est effectué au travers d'un logiciel nommé SEMABOAR, développé en interne au laboratoire de lithographie du LETI sur la base de la calibration d'un modèle de faisceau gaussien. Ainsi, on peut obtenir l'image en niveau de gris comme présentée Figure V.13, à partir d'un fichier de données géométriques équivalentes à notre échantillon virtuel (les deux étant générés en même temps, seul le format change, image ou données numériques).

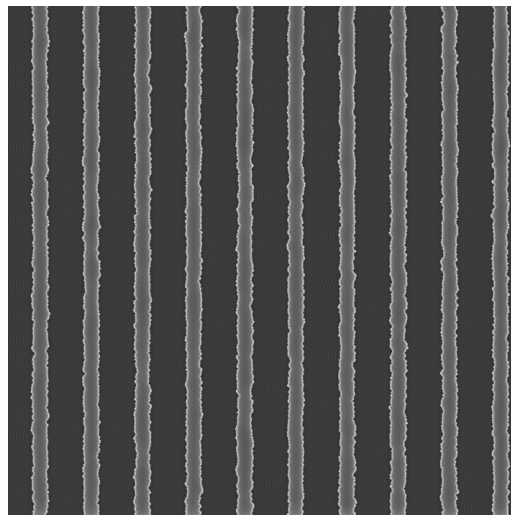


Figure V.13 : Génération SEM à partir de l'échantillon virtuel présenté Figure III.31, l'image finale fait $2 \times 2 \mu\text{m}^2$.

Dans le cadre de cette génération, nous avons accès à plusieurs points, à savoir le SWA nominal des lignes (qui ne change pas la position du bord en tant que telle, mais seulement son intensité) et le bruit de l'image. Ce bruit est défini à partir d'une génération aléatoire dont on fixe la dispersion. Un exemple de génération, avec différents bruits à partir de l'échantillon virtuel Figure III.31, est donné sur la Figure V.14.

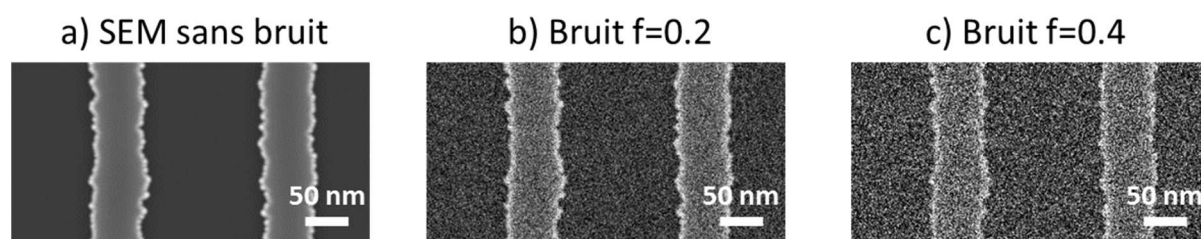


Figure V.14 : Génération zoomée sur le haut des deux premières lignes dans le cas : a) sans bruit, ou un bruit de distribution gaussienne avec une dispersion b) 0.2 et c) 0.4.

Si l'on analyse ces images (détection de bord, calcul de PSD et fit par notre modèle) on obtient les résultats présentés dans la Table V.7. On remarque un biais important entre les valeurs nominales et l'extraction, jusqu'à près de 20% sur σ par exemple. Pour rappel, ces résultats sont obtenus sur une seule image pour illustration, alors que nous avons déterminé qu'il fallait au minimum extraire 100 lignes pour obtenir un résultat correct.

Facteur de bruit	σ (nm)	α	ξ (nm)
0	3.59	0.47	9.55
0.2	3.06	0.42	8.63
0.4	2.75	0.44	8.81
Valeur nominale	3.00	0.50	10.0

Table V.7 : Resultat d'extraction des paramètres PSD pour une image SEM générée en Figure V.13, avec différents niveaux de bruit, et la valeur nominale des paramètres de rugosité en entrée.

Par ailleurs, on voit à partir de ces résultats que le bruit influe de manière importante sur les paramètres extraits. Ainsi suivant le système étudié en réalité, en termes de bruit sur les images, il sera nécessaire de faire un apprentissage du réseau sur des données simulées comportant un même niveau de bruit afin que le réseau puisse le prendre en compte.

V.4.2 Génération AFM

La génération AFM doit prendre en compte, principalement, la taille de la pointe AFM qui est le paramètre limitant pour la mesure de rugosité de ligne avec cette technique. Dans le cadre de nos équipements virtuels, l'AFM nous donne une image en vue du dessus, ce qui est équivalent à la mesure lors du calcul de la moyenne de l'information sur la hauteur du flanc de ligne. Un exemple de résultat de simulation est donné en Figure V.15, pour le même échantillon virtuel que précédemment.

Pour une analyse plus en détail de la problématique AFM, nous avons décidé de pouvoir changer le diamètre de la pointe dans la simulation. La Figure V.16 montre donc l'évolution des lignes suivant la taille de la pointe considérée. On observe une augmentation de la taille de la ligne qui peut être facilement supprimée, comme dans le cas pratique, mais qui n'influe pas sur l'extraction de rugosité ; un lissage important est cependant notable sur les bords des lignes. Notre simulation prend donc bien en compte ce défaut de convolution par la pointe, déjà montré expérimentalement lors de l'explication de la technique.

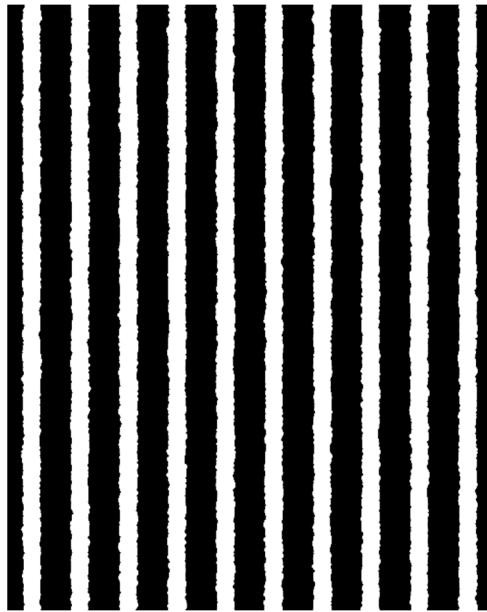


Figure V.15 : Génération AFM à partir de l'échantillon virtuel présenté Figure III.31, de dimension $2 \times 3 \mu\text{m}^2$

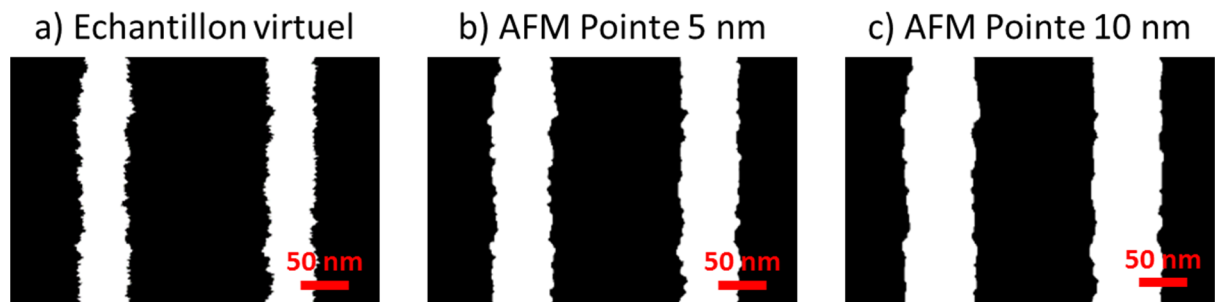


Figure V.16 : Zoom sur le haut des deux premières lignes d'une génération dans le cas : a) de l'échantillon virtuel, d'une génération avec une pointe AFM d'un diamètre de b) 5 nm et c) 10 nm.

Si l'on fait l'extraction complète de la PSD, comme celle présentée en Figure V.17, et que l'on approche les données avec le modèle choisi (voir Chapitre II), on obtient (Table V.8) des données avec un biais important par rapport aux données nominales. De plus, l'écart de résultat entre les deux tailles de pointes est important, jusqu'à près de 30% sur le paramètre ξ . Il est donc indispensable, pour l'apprentissage du réseau de neurones, d'utiliser des données de simulation AFM notamment sur la taille de pointe, en adéquation avec celles extraites de l'équipement.

Taille de la pointe (nm)	σ (nm)	α	ξ (nm)
5	3.73	0.77	9.76
10	3.58	0.79	13.14
Valeur nominale	3.00	0.50	10.00

Table V.8 : Paramètres de PSD obtenue par fit sur les PSD d'images AFM simulées pour deux tailles de pointes différentes, et la valeur nominale des paramètres de rugosité en entrée.

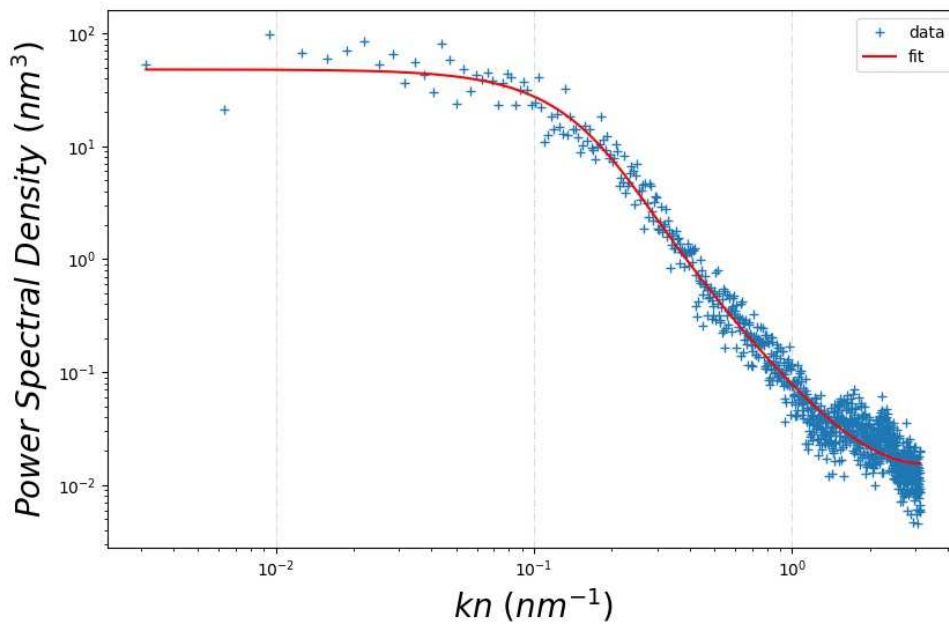


Figure V.17 : Extraction de PSD faite sur la simulation AFM ayant une taille de pointe de 10 nm.

V.4.3 Génération OCD

Le logiciel de simulation OCD utilisé, opère dans notre cas, dans l'espace 2D (comme présenté au Chapitre III), avec le principe déjà explicité de l'EMA. Pour rappel, cela correspondant à l'utilisation d'un matériau mixte entre celui du motif et l'environnement autour des lignes, ainsi il ne faut considérer que la densité moyenne de la structure le long de la ligne. Or comme le montre la Figure V.18, cette densité est identique pour tous les paramétrages ayant le même σ , il ne sera donc pas utile de générer des spectres de réflectivité pour toutes les variations de paramètres, mais seulement pour les changements de σ . Ceci limite donc le nombre de générations OCD nécessaires.

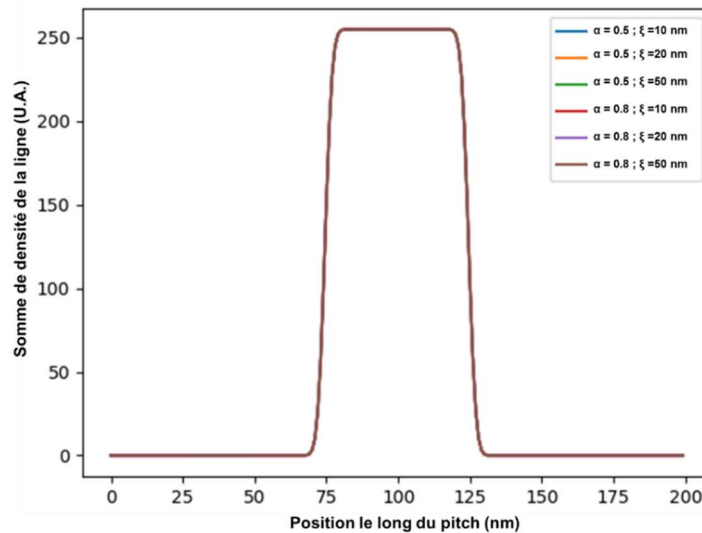


Figure V.18 : Somme des densités le long de la ligne pour différents paramétrages de rugosité, avec pour constante $\sigma = 2.0$ nm, obtenue à partir de 1000 lignes de chaque cas.

V.4.4 Génération SAXS

Comme nous avons pu le voir précédemment, il est possible de simuler la diffusion des rayons X dans notre configuration, comme la valeur absolue du carré de la TF d'un échantillon virtuel. Un exemple de diffusion obtenue avec l'échantillon virtuel présenté en Figure III.31, est donné en Figure V.19.

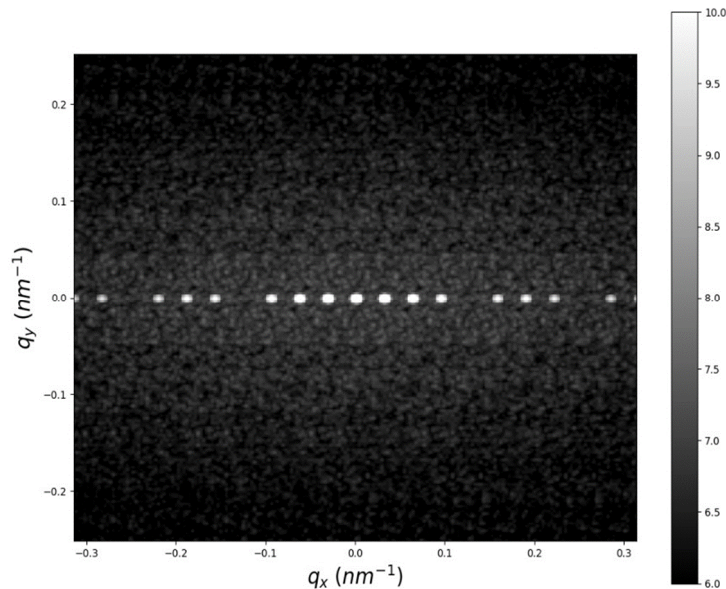


Figure V.19 : Cliché de diffusion simulé, obtenu à partir de l'échantillon virtuel de la Figure V.19

En se basant sur cette diffusion, il est possible comme explicité au Chapitre IV, de retrouver la forme de la PSD, i.e. les paramètres α et ξ . Afin de se rapprocher de cas réels pour l'apprentissage, nous avons mis en place un paramétrage de l'intensité équivalant à un comptage de photons. Le maximum d'intensité étant fixé par un simple facteur de multiplication, cela entraîne le fait qu'une partie des intensités se retrouve être négative. Ces dernières sont alors mises à zéro. De plus, les valeurs sont arrondies dans l'idée où les particules élémentaires ont une énergie donnée, un photon est donc présent ou non, mais ne peut pas être scindé.

Trois résultats, issus de la simulation de diffusion de la Figure V.19, sont présentés Figure V.20. On remarque ainsi qu'un faible nombre de coups (en orange), 10 000 fois moins que sur le maximum (en bleu), pouvant être dû à un temps d'acquisition moindre ou un faisceau de rayons X avec une faible brillance, entraîne une perte de l'information en haute fréquence. Ceci aura un impact direct sur l'extraction des paramètres de rugosité de ligne. Il est donc important de caractériser notre équipement afin d'être en adéquation avec la simulation et donc l'apprentissage du réseau de neurones.

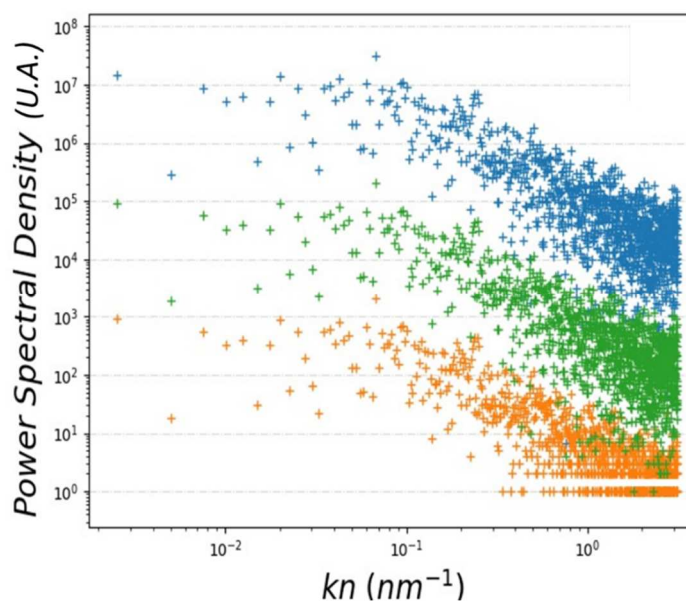


Figure V.20 : Extraction de PSD sur la diffraction simulée d'un échantillon virtuel, pour différents nombres de coups maximum détectés. En bleu le maximum possible, en vert= $\text{total}/10^2$ et orange= $\text{total}/10^4$

V.5 Résumé du chapitre

Au cours de ce chapitre, nous avons pu voir les principes de base de l'hybridation en termes de type de données, ou encore de niveau d'abstraction, ou d'architecture. Nous avons remarqué qu'il n'y avait pas une seule hybridation de données, mais une pluralité.

Ceci nous a conduits à aborder deux techniques d'hybridation de données différentes. La première, le réseau bayésien qui a été détaillé théoriquement afin d'en comprendre ses limites ; les prérequis nécessaires à cette technique, nous ont poussés à l'exploration de la seconde technique, les réseaux de neurones. Cette technique a été expliquée de son origine, en se basant sur un concept biologique et en passant par les différentes formes qu'elle prend jusqu'à son principe d'application.

Une fois les principes posés, nous avons choisi d'utiliser le réseau de neurones, et donc pu commencer à le mettre en place en étudiant les différents paramètres permettant son utilisation et, plus particulièrement son apprentissage. Les paramètres explorés ont été les fonctions d'activation, le nombre de couches cachées, leur taille et la manière de gérer des données pour l'apprentissage. Ceci nous a menés à une première architecture pouvant encore être améliorée.

Par ailleurs, nous nous sommes attachés à mettre en place les données de simulation d'équipement qui permettront l'apprentissage du réseau, à terme. Nous avons, dans ce cadre, exprimé les paramètres sur chacune des techniques virtuelles, permettant de répondre aux divers cas de paramétrage des équipements réels. Il restera donc, par la suite, à évaluer la valeur de ces paramètres pour les équipements réellement utilisés. Ainsi l'apprentissage du réseau sera possible sur les données simulées, et une vérification avec un set d'étalon de rugosité sera possible.

Conclusions et perspectives

Nous avons vu que la réduction de la rugosité de ligne est l'un des enjeux actuels en microélectronique, en effet ce défaut augmente la consommation des composants et dégrade leurs performances. Afin de pouvoir contrôler l'amélioration de procédé vis-à-vis de la rugosité, il est nécessaire d'avoir une métrologie adaptée. Cette métrologie doit répondre à la fois aux problématiques de résolution, de fiabilité et de reproductibilité. Les techniques de mesure seules ne répondant pas à cela, ces travaux ont visé à en hybrider quatre (SEM, AFM, OCD, SAXS), avec sur chacune l'utilisation ou le développement de méthodologie pour l'extraction de la rugosité de ligne.

1. Principaux résultats obtenus

En nous basant sur un état de l'art de la problématique de rugosité de ligne, nous avons défini que les meilleures métriques d'analyse passent par le calcul de la PSD. Pour cela, nous avons mis en place une génération de signaux de rugosité afin d'identifier le modèle adéquat, à savoir celui de la PSD formulée par L. Azar-Nouche. Sur cette base, nous avons mené une étude statistique sur le nombre de lignes et leur taille nécessaire pour l'extraction de la rugosité de ligne, i.e. 100 lignes sur 2 μm de long menant à une erreur après ajustement du modèle de 0.3%, 3% et 1% respectivement pour les paramètres σ , α et ξ .

Nous avons alors revu les techniques disponibles pour la mesure de rugosité et analysé leurs capacités à fournir les métriques de la PSD. Les principaux enseignements de cette étude, sur la base de l'état de l'art, sont : (i) la capacité du CD-SEM et de l'AFM à mesurer la rugosité par une analyse PSD, (ii) la possibilité d'accéder à une information de rugosité non complète (seulement σ) par l'OCD. Un focus sur la technique SAXS par simulation nous a permis de montrer la capacité de cette technique, à extraire la forme de la PSD LWR, par l'utilisation d'une nouvelle méthode que nous avons brevetée. Cependant la difficulté dans l'extraction de σ avec la différentiels LER et LWR reste entière. Des mesures expérimentales ont permis de comparer et classer ces techniques, suivant leurs capacités et leurs performances que nous avons mis ces résultats en regard d'étude déjà faite sur l'extraction du CD. Pour la problématique de rugosité de ligne en ressort que chacune des techniques fonctionne sur des plages de fréquences spatiales différentes (avec toutefois une zone de recouvrement), mais également que les points forts et points faibles de chaque technique sont distincts. Aucune technique ne répond donc, toute seule, à la problématique de mesure de rugosité de ligne, assez résolue et fiable. Nous avons mis en évidence un besoin d'hybridation de données.

Pour valider une métrologie et méthodologie de mesure de rugosité de ligne, notamment celle de l'hybridation de données, il est nécessaire d'avoir des références, ici des étalons de rugosité. Nous avons montré qu'en se basant sur un dessin de rugosité, dans un premier temps périodique, puis finalement créé à partir du modèle PSD choisi, il est possible d'induire de la rugosité contrôlée expérimentalement. Valider l'amplitude et la forme de cette rugosité, avec les outils actuels, a nécessité un développement spécifique afin de retirer de l'extraction finale l'impact de la programmation de rugosité dans le dessin, par rapport à la rugosité du procédé lui-même. De ces travaux sont aussi ressorties les diverses problématiques de création de formes géométriques en microélectronique, telles que la résolution, ou encore la modification, des formes par les procédés. La prise compte de ces divers paramètres et leurs limites ont été explicitées. On retiendra de cette étude

qu'il est possible de programmer de la rugosité de ligne, et donc de fabriquer des étalons de rugosité ce qui n'avait jamais été fait auparavant.

Nous avons, enfin, mis en place des outils pour l'hybridation de données de rugosité de ligne. Après une étude des techniques existantes et au vu des contraintes sur les données extraites, nous avons choisi d'utiliser la technique de réseau de neurones avec laquelle de nombreuses perspectives sont possibles. La mise en place du réseau de neurones s'est faite au travers de la définition de ses divers paramètres (nombre de couches, fonctions associées...) à l'aide de données aléatoires et en tentant de réduire l'erreur globale en sortie de réseau. La solution finale choisie, se base sur un apprentissage supervisé. Des outils de simulations pour chacune des techniques ont été développés avec la création d'échantillons virtuels. Ceci permet de limiter le nombre d'échantillons expérimentaux nécessaires à l'apprentissage du réseau, ceux-ci serviront seulement de données de validation.

En somme, ces travaux ont permis une amélioration dans l'extraction de la rugosité de ligne par les quatre techniques, en les utilisant au maximum de leurs capacités, mais aussi en développant d'autres méthodologies, en SAXS notamment ce qui est un progrès important. Une uniformisation des sorties de ces techniques, en termes de modèle, a permis de les comparer et nous donne la possibilité par suite de les hybrider dans des conditions optimales. De plus, l'évaluation des temps de mesure, ainsi que le référencement des capacités de chacune des techniques, ont permis de rendre ces travaux compatibles avec le secteur de la microélectronique. L'utilisation concrète de cette nouvelle méthode de mesure par hybridation pouvant, par suite, être appliquée suivant les capacités de chaque industriel et de ses besoins, avec 2, 3 ou 4 équipements, en connaissant, grâce aux simulations, l'erreur finale attendue sur les métriques de rugosité.

2. Perspectives

Différents axes de travaux peuvent être, désormais, mis en œuvre pour finaliser ces premiers résultats et pour aller plus loin dans ces recherches.

2.1 Amélioration des techniques individuellement

Actuellement, les deux techniques les plus matures pour l'extraction de rugosité de ligne sont le SEM et l'AFM. Le premier est déjà en limite d'amélioration, après plus de 10 ans de travaux spécifiques, seule une réduction de la taille de faisceau ainsi que la correction de certaines aberrations pourraient être envisagées pour un gain en résolution. Quant à l'AFM, cette technique doit être optimisée, comme dit précédemment, en sélectionnant la bonne distance d'échantillonnage tout en augmentant le nombre de lignes mesurées. Cependant, ceci fait plus partie d'une amélioration méthodologique que de la technique en elle-même.

Pour ce qui est de l'OCD, la technique comporte déjà des voies d'améliorations, telles que la prise de mesures dans les zones dites « Dark Field », ou encore l'analyse par matrices de Muller, toutefois ceci nécessite un équipement adapté. En effet, il est nécessaire d'acquérir le spectre diffusé sur un angle différent de celui d'incidence, ce qui n'est pas prévu, notamment sur l'outil utilisé durant ces travaux. L'accès à ces données permettrait d'obtenir des informations comme la forme de la PSD par exemple, à l'instar de la technique SAXS, mais aussi une différenciation LER et LWR.

Enfin, c'est la technique SAXS qui nécessite aujourd'hui le plus d'améliorations. Ses performances sont très intéressantes mais elles manquent de maturité, de manière générale cette technique doit sortir

des laboratoires afin de permettre une meilleure accessibilité au niveau de l'industrie. Ceci est en cours avec des sources X installées dans des équipements (hors synchrotron), celles-ci ayant une brillance suffisante et une source de haute énergie pour traverser un substrat de silicium. Si l'on regarde de manière plus appliquée à la problématique de rugosité de ligne, la technique SAXS a besoin d'être encore étudiée pour permettre de sortir les paramètres de la PSD LER par exemple, ou encore trouver la normalisation dans le cas d'une mesure réelle pour l'obtention de σ_{LWR} .

2.2 Test fonctionnel du premier réseau

A l'heure actuelle, le réseau a été testé sur des données générées aléatoirement, suivant une valeur cible et une estimation d'erreur des techniques. Maintenant, tout est prêt pour qu'il puisse être testé par l'apprentissage d'un jeu de données simulées, à partir d'un jeu d'échantillons virtuels unique à toutes les techniques, puis validé sur des étalons de rugosité de ligne.

Pour ce faire, les étapes suivantes sont :

- Génération de jeu d'échantillons virtuels (fait)
- Simulation du jeu d'échantillons sur toutes les techniques
- Création du réseau et apprentissage sur les données simulées
- Mesure d'un échantillon commun sur les quatre techniques (SEM, SAXS, OCD fait)
- Extraction des paramètres et passage dans le réseau.

Outre cette mise en place, il reste la problématique du temps total d'acquisition des données, c'est la perspective suivante.

2.3 Vitesse de traitement globale

Afin d'obtenir une extraction par hybridation d'un temps équivalent à une extraction actuelle, comme celle d'un CD-SEM, il serait intéressant d'accélérer chaque prise de mesure, quitte à avoir un ratio signal sur bruit moins bon. Cela induit un apprentissage du réseau avec plus de bruit sur les images SEM, donc une moins bonne détermination des paramètres de rugosité, idem pour les clichés SAXS. Par ailleurs, dans le cadre de l'AFM, cela revient à une augmentation de la distance d'échantillonnage Δy . Il est donc nécessaire de caractériser l'erreur faite par le réseau dans ces cas précis, et donc sa capacité à rejeter le bruit.

Dans le cas de l'extraction réelle, une augmentation de la vitesse d'acquisition, notamment avec un SEM et l'AFM, peut aussi correspondre à la mesure d'un nombre plus restreint de lignes et donc plus de dispersion sur la PSD, ceci est aussi à tester directement lors de l'apprentissage. Ce point n'est pas un facteur à prendre en considération avec l'OCD et la SAXS, puisque le temps d'acquisition est indépendant du nombre de lignes au vu des tailles des faisceaux utilisés.

2.4 Changement du niveau d'hybridation

Le changement de niveau d'hybridation sera la clef pour aller encore plus loin sur cette problématique. Dans un premier temps, il peut s'agir d'une hybridation des PSD sans l'extraction des paramètres. Comme on a pu le voir, les paramètres permettent de caractériser simplement la rugosité, mais ne se substitueront jamais à la courbe PSD. Ainsi sans extraction, on limite les erreurs dues à l'ajustement des données par le modèle, cependant on peut aussi être très dépendant de la statistique, comme

Conclusions et perspectives

celle du nombre de lignes analysées. Par ailleurs, on retrouve le principe de plage de fréquences, qui serait plus étendue par la mise en commun des données, donc une hybridation complémentaire.

Enfin, aller vers un niveau d'abstraction encore plus bas (sans prétraitement), en prenant des données brutes telles les images ou les spectres, permettrait de s'affranchir d'un réseau paramétré pour un type de bruit ou une condition d'acquisition spécifique. De plus, cela limiterait les biais d'extraction, comme la détection d'un bord ou encore le positionnement d'une coupe sur un cliché de diffraction.

Ces deux améliorations nécessiteront quand même une refonte du réseau de neurones décrit. Un nombre de couches beaucoup plus important peut être envisagé. On peut aussi penser à l'utilisation de centres d'hybridation dispersés, comme visible sur la Figure 2, qui travailleraient en deux temps, prenant en compte (i) les données morphologiques puis (ii) la rugosité ; ou encore, comme en Figure 3, une structure avec un réseau non plus PMC mais CNN (Convolutional Neural Network) réseau de neurones convolutif en français, générant directement le traitement d'images.

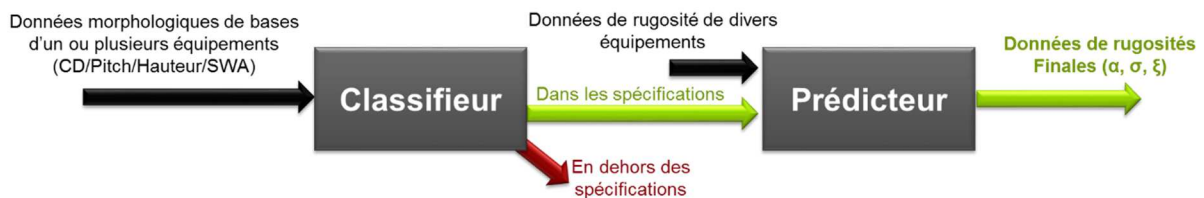


Figure 2 : Deux réseaux de neurones permettant un classement des données avant le traitement de rugosité.



Figure 3 : Réseau de neurones dit "CNN" pouvant traiter des données brutes.

Glossaire

2D	Deux dimensions
3D	Trois dimensions
ACF	Auto-Correlation Function (Fonction d'autocorrélation)
AFM	Atomic Force Microscopie (Microscopie à force atomique)
BSE	BackScattered Electrons (Electrons rétrodiffusés)
CD	Critical Dimension (Dimension critique)
CEA	Commissariat à l'Énergie Atomique et aux énergies alternatives
DFT	Discret Fourier Transform (Transformé de Fourier discrète)
DoF	Deep of Focus (Profondeur de champ)
DSA	Direct Self Assembly (Auto-alignement Direct)
DSP	Densité Spectrale de Puissance
DW	Debye-Waller
EMA	Effective Medium Approximation (approximation d'un milieu effectif)
ESRF	European Synchrotron Radiation Facility
EUV	Extrem Ultra-Violet
GiSAXS	Grazing incidence -SAXS (Lumière rasante SAXS)
HHCF	Height-Height Correlation Function
IoT	Internet of Things (Internet des objets)
IRDS	International Roadmap for Devices and Systems (Feuille de route de la microélectronique)
Kn	Nombre d'onde
LCR	Line Center Roughness (Rugosité de centre de ligne)
LER	Line Edge Roughness (Rugosité d'un bord de ligne)
LWR	Line Width Roughness (Rugosité de variation de la taille de la ligne)
MEB	Microscope Electronique à Balayage
NA	Numerical Aperature (Ouverture numérique)
NIL	Nano Impression Lithography
OCD	Optical Control Dimension (Scattérométrie)
Pattern	Motif
Pitch	Pas de répétition d'une ligne
PMC	Perceptron Multi-Couche
PSD	Power Spectrale Density (voir DSP)
RCWA	Rigorous Coupled Wave Analysis
RGB	Red Green Blue (Rouge vert bleu)
SADP	Self-Aligned Double Patterning
SAXS	Small-Angle X-Ray Scattering (Diffusion des rayons-X au petit angle)
SEM	Scanning Electron Microscopie (voir MEB)
STD	STandard Deviation (Ecart-type)
SWA	Side-Wall Angle (Angle du bord de la ligne)
SWR	Side-Wall Roughness (Rugosité sur la hauteur de la ligne)
TF	Transformé de Fourier
TF ⁻¹	Transformé de Fourier Inverse

Bibliographie

- [1] B. Lojek, *History of semiconductor engineering*. Berlin: Springer, 2007, ISBN [978-3-540-34258-8] .
- [2] Z. Ma et D. Seiler, Éd., *Metrology and Diagnostic Techniques for Nanoelectronics*. Taylor & Francis Group, 6000 Broken Sound Parkway NW, Suite 300, Boca Raton, FL 33487-2742: CRC Press, 2016, <https://doi.org/10.1201/9781315185385>.
- [3] L. Łukasiak et A. Jakubowski, « History of Semiconductors », *J. Telecommun. Inf. Technol.*, vol. 3 -9, n° 1, p. 7, 2010.
- [4] C. Saint-Etienne, *La troisième révolution industrielle*. CreateSpace Independent Publishing Platform, 2017, ISBN [978-1-5429-7790-6] .
- [5] M. Croisy, « Etude de procédés plasmas pour le retrait de résine implantée pour les filières CMOS et photonique », Université Grenoble Alpes, 2016.
- [6] G. E. Moore, « Cramming more components onto integrated circuits, Reprinted from Electronics, volume 38, number 8, April 19, 1965, pp.114 ff. », *IEEE Solid-State Circuits Soc. Newsl.*, vol. 11, n° 3, p. 33-35, sept. 2006, <https://doi.org/10.1109/N-SSC.2006.4785860>.
- [7] « 10 nm lithography process - WikiChip ». [En ligne]. Disponible à: https://en.wikichip.org/wiki/10_nm_lithography_process#Industry. [Consulté le: 26-août-2019].
- [8] « Semiconductor device fabrication », *Wikipedia*. 20-août-2019.
- [9] H. N. Khan, D. A. Hounshell, et E. R. H. Fuchs, « Science and research policy at the end of Moore's law », *Nat Electron*, vol. 1, no 1, p. 14-21, janv. 2018, <https://doi.org/10.1038/s41928-017-0005-9>
- [10] L. Bi-Qiu *et al.*, « 28 nm poly-cut layer lithography process developments », dans *2015 China Semiconductor Technology International Conference*, 2015, p. 1-3, <https://doi.org/10.1109/CSTIC.2015.7153345>.
- [11] « 2017 Edition Reports - IEEE International Roadmap for Devices and Systems », 2017. [En ligne]. Disponible à: <https://irds.ieee.org/roadmap-2017>. [Consulté le: 09-mai-2019].
- [12] S. W. Jones, « A Simulation Study Of 450mm Wafer Fabrication Costs », p. 19.
- [13] « Le monde des semi-conducteurs bascule dans les usines à 10 milliards de dollars », *Les Echos*, 14-juin-2012. [En ligne]. Disponible à: <https://www.lesechos.fr/2012/06/le-monde-des-semi-conducteurs-basculer-dans-les-usines-a-10-milliards-de-dollars-358514>. [Consulté le: 06-mai-2019].
- [14] M. Lapedus, « EUV tool costs hit \$120 million », *EDN Network*, 19-nov-2010. [En ligne]. Disponible à: <https://www.edn.com/electronics-news/4210901/EUV-tool-costs-hit--120-million->. [Consulté le: 10-mai-2019].
- [15] M. Lapedus, « ASML, Canon, Nikon tip immersion tools », *EETimes*. [En ligne]. Disponible à: https://www.eetimes.com/document.asp?doc_id=1162381. [Consulté le: 10-mai-2019].
- [16] A. Pizzagalli *et al.*, « More than Moore Devices: The Wind of Change ». [En ligne]. Disponible à: <https://blog.semi.org/business-markets/more-than-moore-devices-the-wind-of-change>. [Consulté le: 06-mai-2019].
- [17] A. H. Gabor *et al.*, « Overlay error statistics for multiple-exposure patterning », *J. MicroNanolithography MEMS MOEMS*, vol. 18, n° 02, p. 1, avr. 2019, <https://doi.org/10.1117/1.JMM.18.2.021202>.

Bibliographie

- [18] C. A. Mack, « Will stochastics be the ultimate limiter for nanopatterning? », dans *Novel Patterning Technologies for Semiconductors, MEMS/NEMS, and MOEMS 2019*, San Jose, United States, 2019, p. 2, <https://doi.org/10.1117/12.2517598>.
- [19] S. Landis *et al.*, Éd., *Lithography: main techniques*. London : Hoboken, NJ: ISTE ; Wiley, 2011, ISBN [978-1-84821-202-2] .
- [20] « Lithography k1 coefficient », *Semiconductor Engineering*. [En ligne]. Disponible à: https://semiengineering.com/knowledge_centers/manufacturing/lithography/lithography-k1-coefficient/. [Consulté le: 29-avr-2019].
- [21] B. J. Lin, « The k3 coefficient in nonparaxial λ/NA scaling equations for resolution, depth of focus, and immersion lithography », *J. MicroNanolithography MEMS MOEMS*, vol. 1, n° 1, p. 7, avr. 2002, <https://doi.org/10.1117/1.1445798>.
- [22] S. Landis, *Technologies de base en lithographie*. Paris; Lavoisier: Hermès Science publications ;, 2010, ISBN [978-2-7462-2445-2] .
- [23] M. Bossard, « Développement de moules intrinsèquement antiadhésifs pour l'étude du collage en nano-impression », Université Grenoble Alpes, Grenoble, France, 2016.
- [24] S. V. Sreenivasan, « Nanoimprint lithography steppers for volume fabrication of leading-edge semiconductor integrated circuits », *Microsyst. Nanoeng.*, vol. 3, n° 1, p. 17075, déc. 2017, <https://doi.org/10.1038/micronano.2017.75>.
- [25] F. Delachat *et al.*, « Evaluation of anti-sticking layers performances for 200mm wafer scale smart NILTM process through surface and defectivity characterizations », dans *Advances in Patterning Materials and Processes XXXV*, San Jose, United States, 2018, p. 41, <https://doi.org/10.1117/12.2298407>.
- [26] Y. Ban *et al.*, « Self-aligned double-patterning layout decomposition for two-dimensional random metals for sub-10-nm node design », *J. MicroNanolithography MEMS MOEMS*, vol. 14, n° 1, p. 011004, déc. 2014, <https://doi.org/10.1117/1.JMM.14.1.011004>.
- [27] Y. Ma *et al.*, « Directed Self Assembly (DSA) compliant flow with immersion lithography: from material to design and patterning », présenté à SPIE Advanced Lithography, San Jose, California, United States, 2016, p. 97770N, <https://doi.org/10.1117/12.2219505>.
- [28] A. P. Lane *et al.*, « Directed Self-Assembly and Pattern Transfer of Five Nanometer Block Copolymer Lamellae », *ACS Nano*, vol. 11, n° 8, p. 7656-7665, août 2017, <https://doi.org/10.1021/acsnano.7b02698>.
- [29] C. Valade *et al.*, « Tilted beam SEM, 3D metrology for industry », dans *Metrology, Inspection, and Process Control for Microlithography XXXIII*, San Jose, United States, 2019, p. 32, <https://doi.org/10.1117/12.2514977>.
- [30] A. Chen *et al.*, « Utilization of Insitu Metrology Capability of ASML Lithography Scanner to Improve Overall Process Control », dans *2006 IEEE International Symposium on Semiconductor Manufacturing*, Tokyo, Japan, 2006, p. 356-359, <https://doi.org/10.1109/ISSM.2006.4493106>.
- [31] M. T. I. Sales, « Inline Offline Measurement », *MTI Instruments*. [En ligne]. Disponible à: <https://www.mtiinstruments.com/knowledge-center/inline-offline-measurement/>. [Consulté le: 02-mai-2019].
- [32] T. Hoffmann *et al.*, « Investigation of 193 nm Resist Shrinkage During CD-SEM Measurements », *FOCUS*, p. 6, sept. 2014.
- [33] P. Oldiges *et al.*, « Modeling line edge roughness effects in sub 100 nanometer gate length devices », dans *Simulation of Semiconductor Processes and Devices, 2000. SISPAD 2000. 2000 International Conference on*, 2000, p. 131–134.

Bibliographie

- [34] J.-Y. Lee *et al.*, « Effect of line-edge roughness (LER) and line-width roughness (LWR) on sub-100-nm device performance », dans *Microlithography 2004*, 2004, p. 426–433.
- [35] A. Moussa *et al.*, « The need for LWR metrology standardization: the imec roughness protocol », dans *Metrology, Inspection, and Process Control for Microlithography XXXII*, San Jose, United States, 2018, p. 12, <https://doi.org/10.1117/12.2294617>.
- [36] A. Yamaguchi *et al.*, « Characterization of line-edge roughness in resist patterns and estimations of its effect on device performance », dans *Microlithography 2003*, 2003, p. 689–698.
- [37] G. F. Lorusso *et al.*, « Impact of line width roughness on device performance », présenté à SPIE 31st International Symposium on Advanced Lithography, San Jose, CA, 2006, p. 61520W, <https://doi.org/10.1117/12.656128>.
- [38] C. Gustin *et al.*, « Impact of line width roughness on the matching performances of next-generation devices », *Thin Solid Films*, vol. 516, n° 11, p. 3690-3696, avr. 2008, <https://doi.org/10.1016/j.tsf.2007.08.117>.
- [39] W. Li *et al.*, « Strain Consideration of Charge Plasma Structure on Junctionless Transistor With Line Edge Roughness », dans *2018 IEEE International Conference on Electron Devices and Solid State Circuits (EDSSC)*, Shenzhen, 2018, p. 1-2, <https://doi.org/10.1109/EDSSC.2018.8487066>.
- [40] S.-W. Yoon *et al.*, « Line-edge roughness on fin-field-effect-transistor performance for below 10nm patterns », dans *Extreme Ultraviolet (EUV) Lithography X*, San Jose, United States, 2019, p. 54, <https://doi.org/10.1117/12.2515224>.
- [41] R. Baert *et al.*, « System-Level Impact of Interconnect Line-Edge Roughness », dans *2018 IEEE International Interconnect Technology Conference (IITC)*, Santa Clara, CA, USA, 2018, p. 67-69, <https://doi.org/10.1109/IITC.2018.8430429>.
- [42] T. Marschner *et al.*, « Advanced edge roughness measurement application for mask metrology », 2007, p. 65181R, <https://doi.org/10.1117/12.712530>.
- [43] A. R. Pawloski et P. F. Nealey, « Effect of photoacid generator concentration on sensitivity, photoacid generation, and deprotection of chemically amplified resists », *J. Vac. Sci. Technol. B Microelectron. Nanometer Struct.*, vol. 20, n° 6, p. 2413, 2002, <https://doi.org/10.1116/1.1525811>.
- [44] J. Foucher *et al.*, « Impact of acid diffusion length on resist LER and LWR measured by CD-AFM and CD-SEM », présenté à Advanced Lithography, San Jose, CA, 2007, p. 65181Q, <https://doi.org/10.1117/12.712186>.
- [45] M. Carcasi *et al.*, « EUV resist sensitization and roughness improvement by PSCAR with in-line flood exposure system (Conference Presentation) », dans *Advances in Patterning Materials and Processes XXXV*, San Jose, United States, 2018, p. 5, <https://doi.org/10.1117/12.2297498>.
- [46] C. Cutler *et al.*, « Roughness power spectral density as a function of resist parameters and its impact through process », dans *Optical Microlithography XXXI*, San Jose, United States, 2018, p. 6, <https://doi.org/10.1117/12.2297690>.
- [47] V. Constantoudis *et al.* « Quantification of line-edge roughness of photoresists. II. Scaling and fractal analysis and the best roughness descriptors », *J. Vac. Sci. Technol. B Microelectron. Nanometer Struct.*, vol. 21, n° 3, p. 1019, 2003, <https://doi.org/10.1116/1.1570844>.
- [48] T. Wallow *et al.*, « Line-edge roughness in 193-nm resists: lithographic aspects and etch transfer », présenté à Advanced Lithography, San Jose, CA, 2007, p. 651919, <https://doi.org/10.1117/12.712319>.
- [49] L. Azarnouche *et al.*, « Plasma treatments to improve line-width roughness during gate patterning », *J. MicroNanolithography MEMS MOEMS*, vol. 12, n° 4, p. 041304, sept. 2013, <https://doi.org/10.1117/1.JMM.12.4.041304>.

Bibliographie

- [50] J. Jussot *et al.*, « Line width roughness reduction strategies for patterns exposed via electron beam lithography », présenté à SPIE Advanced Lithography, San Jose, California, USA, 2014, p. 905405, <https://doi.org/10.1117/12.2047972>.
- [51] C. A. Mack, « Understanding the efficacy of linewidth roughness postprocessing », *J. MicroNanolithography MEMS MOEMS*, vol. 14, n° 3, p. 033503–033503, 2015.
- [52] « ITRS Metrol Roadmap 2015 ». [En ligne]. Disponible à : <https://eps.ieee.org/images/files/Roadmap/ITRSMetrol2015.pdf>. [Consulté le: 26-janv-2018].
- [53] A. Vaid *et al.*, « A holistic metrology approach: hybrid metrology utilizing scatterometry, CD-AFM, and CD-SEM », présenté à SPIE Advanced Lithography, San Jose, California, USA, 2011, p. 797103, <https://doi.org/10.1117/12.881632>.
- [54] N. G. Schuch, « Hybrid Metrology Applied to dimensional Control in Lithography », Université Grenoble Alpes, Grenoble, France, 2017.
- [55] Bureau International des Poids et Mesures, « International vocabulary of metrology – Basic and general concepts and associated terms ». [En ligne]. Disponible à : <https://www.bipm.org/en/publications/guides/vim.html>. [Consulté le: 07-mai-2019].
- [56] G. A. Werdmuller et W. G. de Ruig, « Precision of test methods : determination of repeatability and reproducibility by inter-laboratory tests : application in development and assessment of dairy methodology », Wageningen : RIKILT (Verslag / RIKILT 83.23) - 14, Research report 463239, 1983.
- [57] V. Constantoudis *et al.*, « Challenges in line edge roughness metrology in directed self-assembly lithography: placement errors and cross-line correlations », *J. MicroNanolithography MEMS MOEMS*, vol. 16, n° 2, p. 024001, avr. 2017, <https://doi.org/10.1117/1.JMM.16.2.024001>.
- [58] B. Mandelbrot, « The Fractal Geometry of Nature », *SIAM Rev.*, vol. 26, no 1, p. 131-132, janv. 1984. <https://doi.org/10.1137/1026020>.
- [59] J. Lopez, G. Hansali, J. C. Le Bossé, et T. Mathia, « Caractérisation fractale de la rugosité tridimensionnelle d'une surface », *J. Phys. III*, vol. 4, n° 12, p. 2501-2219, déc. 1994, <https://doi.org/10.1051/jp3:1994294>.
- [60] V. Constantoudis *et al.*, « Line edge roughness and critical dimension variation: Fractal characterization and comparison using model functions », *J. Vac. Sci. Technol. B Microelectron. Nanometer Struct.*, vol. 22, n° 4, p. 1974, 2004, <https://doi.org/10.1116/1.1776561>.
- [61] S. K. Sinha *et al.*, « X-ray and neutron scattering from rough surfaces », *Phys. Rev. B*, vol. 38, n° 4, p. 2297, 1988, <https://doi.org/10.1103/PhysRevB.38.2297>.
- [62] L. Azar-Nouche, « Défis liés à la réduction de la rugosité des motifs de résine photosensible 193 nm », Université de Grenoble, 2012.
- [63] G. Palasantzas, « Roughness spectrum and surface width of self-affine fractal surfaces via the K-correlation model », *Phys. Rev. B*, vol. 48, n° 19, p. 14472-14478, nov. 1993, <https://doi.org/10.1103/PhysRevB.48.14472>.
- [64] C. A. Mack, « Generating random rough edges, surfaces, and volumes », *Appl. Opt.*, vol. 52, n° 7, p. 1472, mars 2013, <https://doi.org/10.1364/AO.52.001472>.
- [65] C. A. Mack, « Biases and uncertainties in the use of autocovariance and height–height covariance functions to characterize roughness », *J. Vac. Sci. Technol. B Nanotechnol. Microelectron. Mater. Process. Meas. Phenom.*, vol. 34, n° 6, p. 06K701, nov. 2016, <https://doi.org/10.1116/1.4961445>.
- [66] C. A. Mack, « Systematic errors in the measurement of power spectral density », *J. MicroNanolithography MEMS MOEMS*, vol. 12, n° 3, p. 033016–033016, 2013.

Bibliographie

- [67] C. A. Mack, « More systematic errors in the measurement of power spectral density », *J. MicroNanolithography MEMS MOEMS*, vol. 14, n° 3, p. 033502, juill. 2015, <https://doi.org/10.1117/1.JMM.14.3.033502>.
- [68] J. Jussot, « Lithographie directe à faisceaux d'électrons multiples pour les noeuds technologiques sub-20nm », Université Joseph-Fourier-Grenoble I, 2015.
- [69] A. Hiraiwa et A. Nishida, « Discrete power spectrum of line width roughness », *J. Appl. Phys.*, vol. 106, n° 7, p. 074905, oct. 2009, <https://doi.org/10.1063/1.3226883>.
- [70] A. Šakić *et al.*, « Boron-layer silicon photodiodes for high-efficiency low-energy electron detection », *Solid-State Electron.*, vol. 65-66, p. 38-44, nov. 2011, <https://doi.org/10.1016/j.sse.2011.06.042>.
- [71] G. F. Lorusso *et al.*, « Unbiased roughness measurements: Subtracting out SEM effects », *Microelectron. Eng.*, vol. 190, p. 33-37, avr. 2018, <https://doi.org/10.1016/j.mee.2018.01.010>.
- [72] T. Verduin, « Quantum Noise Effects in e-beam Lithography and Metrology », TU Delft, 2017.
- [73] B. D. Bunday et C. A. Mack, « Influence of metrology error in measurement of line edge roughness power spectral density », présenté à SPIE Advanced Lithography, San Jose, California, USA, 2014, p. 90500G, <https://doi.org/10.1117/12.2047100>.
- [74] T. Verduin *et al.*, « Determination of line edge roughness in low dose top-down scanning electron microscopy images », 2014, p. 90500L, <https://doi.org/10.1117/12.2046493>.
- [75] L. Zhang *et al.*, « Pattern quality and defect evaluation based on cross correlation and power spectral density methods », *J. Vac. Sci. Technol. B*, vol. 36, n° 5, p. 052902, sept. 2018, <https://doi.org/10.1116/1.5040391>.
- [76] T. Kudo *et al.*, « Mechanistic Studies on the CD Degradation of 193nm Resists during SEM Inspection. », *J. Photopolym. Sci. Technol.*, vol. 14, n° 3, p. 407-417, 2001, <https://doi.org/10.2494/photopolymer.14.407>.
- [77] C. J. Wu *et al.*, « Investigation on the mechanism of the 193-nm resist linewidth reduction during the SEM measurement », présenté à 26th Annual International Symposium on Microlithography, Santa Clara, CA, 2001, p. 190, <https://doi.org/10.1117/12.436848>.
- [78] M. Fouchier *et al.*, « Atomic force microscopy study of photoresist sidewall smoothing and line edge roughness transfer during gate patterning », *J. MicroNanolithography MEMS MOEMS*, vol. 12, n° 4, p. 041308, oct. 2013, <https://doi.org/10.1117/1.JMM.12.4.041308>.
- [79] T. Ohashi *et al.*, « Correction of EB-induced shrinkage in contour measurements », présenté à SPIE Advanced Lithography, San Jose, California, USA, 2014, p. 90500J, <https://doi.org/10.1117/12.2047563>.
- [80] G. Freychet, « Analyses morphologiques et dimensionnelles de nanostructures organisées par diffusion centrale des rayons X », Université Grenoble Alpes, 2016.
- [81] M. G. Moharam et T. K. Gaylord, « Rigorous coupled-wave analysis of metallic surface-relief gratings », *J. Opt. Soc. Am. A*, vol. 3, no 11, p. 1780, nov. 1986. <https://doi.org/10.1364/JOSAA.3.001780>
- [82] Y. Cohen, B. Yaakovovitz, Y. Tsur, et D. Scheiner, « A novel method for pushing the limits of line edge roughness detection by scatterometry », 2008, p. 692220, <https://doi.org/10.1117/12.772117>.
- [83] D.-M. Shyu, Y.-S. Ku, et N. Smith, « Angular scatterometry for line-width roughness measurement », 2007, p. 65184G, <https://doi.org/10.1117/12.711734>.
- [84] B. C. Bergner *et al.*, « Effect of line-width roughness on optical scatterometry measurements », 2009, p. 72720U, <https://doi.org/10.1117/12.813770>.

- [85] H. Gross *et al.*, « Modelling line edge roughness in periodic line-space structures by Fourier optics to improve scatterometry », *J. Eur. Opt. Soc.-Rapid Publ.*, vol. 9, 2014. <https://doi.org/10.2971/jeos.2014.14003>.
- [86] J. Bischoff et K. Hehl, « Scatterometry modeling for gratings with roughness and irregularities », 2016, p. 977804, <https://doi.org/10.1117/12.2219019>.
- [87] M. Heusinger *et al.*, « Diffuse scattering due to stochastic disturbances of 1D-gratings on the example of line edge roughness », *Opt. Express*, vol. 26, n° 21, p. 28104, oct. 2018, <https://doi.org/10.1364/OE.26.028104>.
- [88] B. C. Bergner *et al.*, « Effective medium approximations for modeling optical reflectance from gratings with rough edges », *JOSA A*, vol. 27, n° 5, p. 1083-1090, mai 2010, <https://doi.org/10.1364/JOSAA.27.001083>.
- [89] C. Settens *et al.*, « Critical dimension small angle X-ray scattering measurements of FinFET and 3D memory structures », 2013, vol. 8681, p. 86810L-86810L-9, <https://doi.org/10.1117/12.2012019>.
- [90] D. F. Sunday, M. R. Hammond, C. Wang, W. Wu, R. J. Kline, et G. E. Stein, « Three-dimensional x-ray metrology for block copolymer lithography line-space patterns », *J. MicroNanolithography MEMS MOEMS*, vol. 12, n° 3, p. 031103–031103, 2013.
- [91] D. F. Sunday *et al.*, « Determination of the Internal Morphology of Nanostructures Patterned by Directed Self Assembly », *ACS Nano*, vol. 8, n° 8, p. 8426-8437, août 2014, <https://doi.org/10.1021/nn5029289>.
- [92] G. Freychet *et al.*, « Critical-dimension grazing incidence small angle x-ray scattering », 2018, p. 37, <https://doi.org/10.1117/12.2297518>.
- [93] W. A. Coghlan et R. E. Clausing, « Auger catalog calculated transition energies listed by energy and element », *At. Data Nucl. Data Tables*, vol. 5, n° 4, p. 317-469, janv. 1973, [https://doi.org/10.1016/S0092-640X\(73\)80005-1](https://doi.org/10.1016/S0092-640X(73)80005-1).
- [94] S. Seltzer, « Tables of X-Ray Mass Attenuation Coefficients and Mass Energy-Absorption Coefficients, NIST Standard Reference Database 126 ». National Institute of Standards and Technology, 1995, <https://doi.org/10.18434/t4d01f>.
- [95] A. H. Compton, « A Quantum Theory of the Scattering of X-rays by Light Elements », *Phys. Rev.*, vol. 21, n° 5, p. 483-502, mai 1923, <https://doi.org/10.1103/PhysRev.21.483>.
- [96] J. J. Thomson, *Conduction of electricity through gases*. Cambridge University Press, 1906, ISBN [978-1-107-41428-0].
- [97] G. Freychet *et al.*, « Reconstructing the three-dimensional latent image of extreme ultraviolet resists with resonant soft x-ray scattering », *J. MicroNanolithography MEMS MOEMS*, vol. 18, n° 02, p. 1, mai 2019, <https://doi.org/10.1117/1.JMM.18.2.024003>.
- [98] C. Wang *et al.*, « Small angle x-ray scattering measurements of lithographic patterns with sidewall roughness from vertical standing waves », *Appl. Phys. Lett.*, vol. 90, n° 19, p. 193122, 2007, <https://doi.org/10.1063/1.2737399>.
- [99] M. A. Singh et M. N. Groves, « Depth profiling of polymer films with grazing-incidence small-angle X-ray scattering », *Acta Crystallogr. A*, vol. 65, n° Pt 3, p. 190-201, mai 2009, <https://doi.org/10.1107/S0108767309007508>.
- [100] C. Wang *et al.*, « Characterization of correlated line edge roughness of nanoscale line gratings using small angle x-ray scattering », *J. Appl. Phys.*, vol. 102, n° 2, p. 024901, 2007, <https://doi.org/10.1063/1.2753588>.

Bibliographie

- [101] C. Wang *et al.*, « Line edge roughness characterization of sub-50nm structures using CD-SAXS: round-robin benchmark results », 2007, p. 65181O, <https://doi.org/10.1117/12.725380>.
- [102] G. Freychet *et al.*, « A study of lateral roughness evaluation through critical-dimension small angle x-ray scattering (CD-SAXS) », présenté à Metrology, Inspection, and Process Control for Microlithography XXX, 2016, vol. 9778, p. 97783V, <https://doi.org/10.1117/12.2235102>.
- [103] M. Jergel *et al.*, « X-ray scattering study of interface roughness correlation in Mo/Si and Ti/C multilayers for X-UV optics », *Phys. B Condens. Matter*, vol. 253, n° 1-2, p. 28–39, 1998.
- [104] D. K. Bowen et B. K. Tanner, *X-ray metrology in semiconductor manufacturing*. Boca Raton: CRC/Taylor & Francis, 2006, ISBN [978-0-8493-3928-8] .
- [105] A. Fernández Herrero *et al.*, « Characteristic diffuse scattering from distinct line roughnesses », *J. Appl. Crystallogr.*, vol. 50, n° 6, p. 1766-1772, déc. 2017, <https://doi.org/10.1107/S1600576717014455>.
- [106] J. Reche *et al.*, « Application of PSD for the extraction of programmed line roughness from SAXS », dans *Metrology, Inspection, and Process Control for Microlithography XXXIII*, San Jose, United States, 2019, p. 28, <https://doi.org/10.1117/12.2514919>.
- [107] B. Bunday *et al.*, « Metrology capabilities and needs for 7nm and 5nm logic nodes », présenté à SPIE Advanced Lithography, San Jose, California, United States, 2017, p. 101450G, <https://doi.org/10.1117/12.2260870>.
- [108] B. D. Bunday *et al.*, « CD-SEM measurement line-edge roughness test patterns for 193-nm lithography », dans *Microlithography 2003*, 2003, p. 674–688, <https://doi.org/10.1117/12.485234>.
- [109] K. Fabrice et D. Cheikh, « Rehaussement d'un signal de parole », p. 11.
- [110] C. Wang *et al.*, « Linewidth roughness and cross-sectional measurements of sub-50 nm structures with CD-SAXS and CD-SEM », 2008, p. 69221Z, <https://doi.org/10.1117/12.773558>.
- [111] J. Belissard et J. Hazart, « Limits of model-based CD-SEM metrology », dans *34th European Mask and Lithography Conference*, Grenoble, France, 2018, p. 6, <https://doi.org/10.1117/12.2323696>.
- [112] « Factorisation de Cholesky : définition et explications », *Techno-Science.net*. [En ligne]. Disponible à : <https://www.techno-science.net/definition/5125.html>. [Consulté le: 31-juill-2019].
- [113] K. Barnham et D. Vvedensky, Éd., *Low-Dimensional semiconductor structures: Fundamentals and device applications*. Cambridge: Cambridge University Press, 2001, <https://doi.org/10.1017/CBO9780511624247>.
- [114] M. Cheng *et al.*, « Improving resist resolution and sensitivity via electric-field enhanced postexposure baking », *J. Vac. Sci. Technol. B Microelectron. Nanometer Struct.*, vol. 20, n° 2, p. 734, 2002, <https://doi.org/10.1116/1.1464835>.
- [115] F. Castanedo, « A Review of Data Fusion Techniques », *Sci. World J.*, vol. 2013, p. 1-19, 2013, <https://doi.org/10.1155/2013/704504>.
- [116] B. V. Dasarathy, « Sensor fusion potential exploitation-innovative architectures and illustrative applications », *Proc. IEEE*, vol. 85, n° 1, p. 24-38, janv. 1997, <https://doi.org/10.1109/5.554206>.
- [117] R. C. Luo *et al.*, « Multisensor fusion and integration: approaches, applications, and future research directions », *IEEE Sens. J.*, vol. 2, n° 2, p. 107-119, avr. 2002, <https://doi.org/10.1109/JSEN.2002.1000251>.

- [118] N. Rana et C. Archie, « Hybrid reference metrology exploiting patterning simulation », présenté à SPIE Advanced Lithography, San Jose, California, 2010, p. 76380W, <https://doi.org/10.1117/12.848311>.
- [119] R. M. Silver *et al.*, « Improving optical measurement accuracy using multi-technique nested uncertainties », 2009, p. 727202, <https://doi.org/10.1117/12.816569>.
- [120] J. Hazart *et al.*, « Data fusion for CD metrology: heterogeneous hybridization of scatterometry, CDSEM, and AFM data », 2014, p. 90502L, <https://doi.org/10.1117/12.2046484>.
- [121] M.-A. Henn *et al.*, « Optimizing hybrid metrology: rigorous implementation of Bayesian and combined regression », *J. MicroNanolithography MEMS MOEMS*, vol. 14, n° 4, p. 044001–044001, 2015.
- [122] I. Gereige, « Contribution des réseaux de neurones dans le domaine de l'ellipsométrie: Application à la scatterométrie », Université de Saint-Etienne, 2009.
- [123] C. Gourgon et G. Granet, « Classification par réseaux de neurones dans le cadre de la scattérométrie ellipsométrique ».
- [124] I. E. Kalioubi, « Développement de la technique de scattérométrie neuronale dynamique », Université Grenoble Alpes, 2015.
- [125] R. E. Kass et A. E. Raftery, « Bayes Factors », *J. Am. Stat. Assoc.*, vol. 90, n° 430, p. 773-795, juin 1995, <https://doi.org/10.1080/01621459.1995.10476572>.
- [126] L. Mejlbro, « Introduction to Probability », *Introd. Probab.*, p. 61. ISBN: 978-87-7681-515-8
- [127] N. F. Zhang *et al.*, « Improving optical measurement uncertainty with combined multitool metrology using a Bayesian approach », *Appl. Opt.*, vol. 51, n° 25, p. 6196, sept. 2012, <https://doi.org/10.1364/AO.51.006196>.
- [128] Y. Chen, « Data fusion for accurate microscopic rough surface metrology », *Ultramicroscopy*, vol. 165, p. 15-25, juin 2016, <https://doi.org/10.1016/j.ultramic.2016.03.012>.
- [129] A. Vaid *et al.*, « Hybrid metrology solution for 1X node technology », 2012, p. 832404, <https://doi.org/10.1117/12.916940>.
- [130] T. F. Brain *et al.*, *Chapter 7 WHAT THE FROG'S EYE TELLS THE FROG'S BRAIN from « The Mind: Biological Approaches to its Functions »*, William C. Corning, Martin Balaban., vol. The Mind: Biological Approaches to its Functions. 1968, ISBN [978-0-470-17493-7] .
- [131] S. F. P. Zaki, « Classification par réseaux de neurones dans le cadre de la scattérométrie ellipsométrique », Université de Lyon, 2016.
- [132] « INTRODUCTION AUX RESEAUX DE NEURONES ». [En ligne]. Disponible à: https://www.lrde.epita.fr/~sigoure/cours_ReseauxNeurones.pdf. [Consulté le: 11-oct-2018].
- [133] « Neural Network FAQ, part 1 of 7: Introduction ». [En ligne]. Disponible à: <ftp://ftp.sas.com/pub/neural/jargon>. [Consulté le: 07-août-2018].
- [134] K. Hornik *et al.*, *Degree of Approximation Results for Feedforward Networks Approximating Unknown Mappings and Their Derivatives*. 1995.
- [135] Y. LeCun, Y. Bengio, et G. Hinton, « Deep learning », *Nature*, vol. 521, n° 7553, p. 436-444, mai 2015, <https://doi.org/10.1038/nature14539>.
- [136] N. Kang, « Multi-Layer Neural Networks with Sigmoid Function— Deep Learning for Rookies (2) », *Medium*, 04-févr-2019. [En ligne]. Disponible à: <https://towardsdatascience.com/multi-layer-neural-networks-with-sigmoid-function-deep-learning-for-rookies-2-bf464f09eb7f>. [Consulté le: 01-août-2019].

Bibliographie

- [137] « Activations - Keras Documentation ». [En ligne]. Disponible à: <https://keras.io/activations/#usage-of-activations>. [Consulté le: 01-août-2019].
- [138] S. Sharma, « Activation Functions in Neural Networks », *Medium*, 14-févr-2019. [En ligne]. Disponible à: <https://towardsdatascience.com/activation-functions-neural-networks-1cbd9f8d91d6>. [Consulté le: 01-août-2019].
- [139] E. D. Sontag, « Neural Networks for Control », dans *Essays on Control*, H. L. Trentelman et J. C. Willems, Éd. Boston, MA: Birkhäuser Boston, 1993, p. 339-380, https://doi.org/10.1007/978-1-4612-0313-1_10.
- [140] K. Hornik, « Approximation capabilities of multilayer feedforward networks », *Neural Netw.*, vol. 4, n° 2, p. 251-257, 1991, [https://doi.org/10.1016/0893-6080\(91\)90009-T](https://doi.org/10.1016/0893-6080(91)90009-T).
- [141] N. Chaudhary *et al.*, « Automated rough line-edge estimation from SEM images using deep convolutional neural networks », dans *Photomask Technology 2018*, Monterey, United States, 2018, p. 56, <https://doi.org/10.1117/12.2501723>.
- [142] C. M. Bishop, « Neural networks and their applications », *Neural Netw.*, vol. 65, n° 6, p. 30, 1994, <https://doi.org/10.1063/1.1144830>.
- [143] N. Chaudhary *et al.*, « Line roughness estimation and Poisson denoising in scanning electron microscope images using deep learning », *J. MicroNanolithography MEMS MOEMS*, vol. 18, n° 02, p. 1, avr. 2019, <https://doi.org/10.1117/1.JMM.18.2.024001>.

Résumé

La détermination de la rugosité sub-nanométrique sur les flancs des motifs, dont les dimensions critiques atteignent une taille inférieure à 10nm, devient une étape primordiale. Mais à ce jour, aucune technique de métrologie n'est suffisamment robuste pour garantir un résultat juste et précis. Une voie, actuellement en cours d'exploration pour la mesure dimensionnelle, consiste à hybrider différentes techniques de métrologie. Pour ce faire, des algorithmes de fusion de données sont développés afin de traiter les informations issues de multiples équipements de métrologie, le but étant donc d'utiliser ce même type de méthode pour la mesure de rugosité de ligne. Ces travaux de thèse explicitent tout d'abord les progrès de méthodologie de mesure de rugosité de ligne, au travers de la décomposition fréquentielle et des modèles associés. Les différentes techniques utilisées pour la mesure de rugosité de lignes sont présentées avec une nouveauté importante concernant le développement et l'utilisation de la technique SAXS pour ce type de mesure. Cette technique possède un potentiel élevé pour la détermination de motifs sub-nanométriques. Des étalons de rugosités de ligne sont fabriqués, sur la base de l'état de l'art comportant des rugosités périodiques, mais aussi, des rugosités plus complexes déterminées par un modèle statistique, utilisé normalement pour la mesure. Ces travaux se focalisent finalement sur les méthodes d'hybridation, et plus particulièrement sur l'utilisation de réseaux de neurones. Ainsi, la mise en place d'un réseau de neurones est détaillée au travers de la multitude de paramètres qu'il comporte. Le choix d'un apprentissage du réseau de neurones sur simulation mène à la nécessité de savoir générer les différentes métrologies en présence.

Abstract

Roughness at Sub-nanometric scale determination becomes a critical issue, especially for patterns with critical dimensions below 10nm. Currently, there is no metrology technique able to provide a result with high precision and accuracy. A way, based on hybrid metrology, is currently explored and dedicated to dimensional measurements. This hybrid metrology uses data fusion algorithms in order to address data coming from different tools. This thesis presents some improvements on line roughness analysis thanks to frequency decomposition and associated model. The current techniques used for roughness determination are explained and a new one SAXS (Small Angle X-rays Scattering) is used to push again limits of extraction of roughness. This technique has a high potential to determine sub-nanometric patterns. Moreover, the design and manufacturing of reference line roughness samples is made, following the state of art with periodic roughness, but also more complex roughness determined by a statistical model usually used for measurement. Finally, this work focus on hybridization methods and more especially on neural network utilization. Thus, the establishment of a neural network is detailed through the multitude of parameters which must be set. In addition, training of the neural network on simulation leads to the capability to generate different metrology.