THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ GRENOBLE ALPES

École doctorale : EEATS - Electronique, Electrotechnique, Automatique, Traitement du Signal
Spécialité : Nano électronique et Nano technologies
Unité de recherche : Laboratoire des Technologies de la Microélectronique

Mise en oeuvre de méthodes à base d'apprentissage pour une application en métrologie hybride à l'échelle nanométrique

Implementation of learning-based methods for hybrid metrology at nanoscale

Présentée par :

Paul DIGRACI

Direction de thèse :

Jean-Hervé TORTAI CHARGE DE RECHERCHE, CNRS	Directeur de thèse
Maxime BESACIER MAITRE DE CONFERENCES, UNIVERSITE GRENOBLE ALPES	Co-directeur de thèse
Patrice GERGAUD DIRECTEUR DE RECHERCHE, CEA CENTRE DE GRENOBLE	Co-directeur de thèse
Jérôme RECHE Ingénieur Chercheur, CEA-Leti	Co-encadrant de thèse

Rapporteurs :

Bruno AGARD
PROFESSEUR, Polytechnique Montréal
Yves JOURLIN
PROFESSEUR DES UNIVERSITES, Université de Saint-Etienne - Jean Monnet

Thèse soutenue publiquement le **9 janvier 2026**, devant le jury composé de :

Alice CAPLIER, PROFESSEURE DES UNIVERSITES, Grenoble INP - UGA	Présidente
Jean-Hervé TORTAI, DIRECTEUR DE RECHERCHE, CNRS délégation Alpes	Directeur de thèse
Bruno AGARD, PROFESSEUR, Polytechnique Montréal	Rapporteur
Yves JOURLIN, PROFESSEUR DES UNIVERSITES, Université de Saint-Etienne - Jean Monnet	Rapporteur
Tatiana NOVIKOVA, INGENIEURE HDR, Institut Polytechnique de Paris	Examinatrice
Nicolas FELTIN, INGENIEUR DOCTEUR, Laboratoire National de métrologie et d'Essais	Examinateur

Invités :

Maxime BESACIER
LTM
Jérôme RECHE
CEA-LETI

« Le véritable voyage de découverte ne consiste pas à chercher de nouveaux paysages, mais à avoir de nouveaux yeux »

Marcel Proust

Remerciements

Je souhaite tout d'abord adresser mes sincères remerciements aux membres du jury pour l'attention qu'ils ont portée à ce travail. Je remercie en particulier Bruno Agard et Yves Jourlin d'avoir accepté d'être rapporteurs de cette thèse, ainsi que Tatiana Novikova, Alice Caplier et Nicolas Feltin pour leur participation au jury et pour les échanges scientifiques enrichissants qu'ils ont apportés lors de la soutenance.

Je tiens à exprimer ma profonde gratitude à mes encadrants de thèse. Je remercie tout particulièrement Jean-Hervé Tortai pour sa direction, sa confiance et la qualité de son encadrement tout au long de ces travaux. Je souhaite également remercier Jérôme Reche pour son implication constante, sa disponibilité et son accompagnement au quotidien, qui ont été déterminants dans l'avancement de cette thèse. Je remercie Maxime Besacier pour les conseils qu'il a pu m'apporter au cours de ces travaux, ainsi que pour l'opportunité qu'il m'a donnée, avec Jérôme Reche, de réaliser cette thèse. Enfin, je remercie Patrice Gergaud pour son regard extérieur sur ces travaux et les échanges que nous avons pu avoir au cours de la thèse.

Je souhaite également remercier mes collègues du LTM et du CEA-Leti pour les nombreux échanges, discussions et moments partagés au cours de ces années. Je pense en particulier à Cédric Lapeyronie, Rémy Bernardin, Jennifer Not, Nicolas Le Brun et Justine Grasland, avec qui j'ai eu le plaisir de collaborer et d'échanger au quotidien.

Mes remerciements vont aussi à mes amis, pour leur soutien constant tout au long de cette période, ainsi que pour les moments de détente indispensables à l'équilibre.

Enfin, je souhaite adresser mes remerciements les plus sincères à ma famille : mon père, ma belle-mère, mes frères et ma sœur, ainsi que mes grands-parents. Leur soutien, leur patience et leurs encouragements ont été essentiels à l'aboutissement de ce travail.

À toutes celles et ceux qui ont contribué, de près ou de loin, à la réalisation de cette thèse, je tiens à exprimer ma sincère reconnaissance.

Résumé

La miniaturisation des composants électroniques, au cœur des technologies modernes, impose de mesurer des structures de plus en plus petites, à l'échelle du nanomètre. À ces dimensions, obtenir des mesures précises devient particulièrement difficile, car chaque technique de mesure présente ses propres limites et sources d'erreur.

Cette thèse propose une approche dite de métrologie hybride, qui consiste à combiner plusieurs techniques de mesure complémentaires appliquées à un même échantillon, afin d'en améliorer l'exactitude globale. Les données issues de ces différentes techniques sont ensuite analysées à l'aide de méthodes d'intelligence artificielle, en particulier des réseaux de neurones, capables d'exploiter efficacement leur complémentarité.

Un défi majeur de cette approche réside dans le besoin important de données pour entraîner ces modèles. Or, les mesures expérimentales sont coûteuses et longues à obtenir. Pour y remédier, ce travail s'appuie sur la génération de données synthétiques réalistes, produites par des simulateurs intégrant à la fois les propriétés physiques des structures étudiées et les imperfections propres aux instruments de mesure.

Les modèles développés ont été entraînés sur ces données, puis évalués afin d'identifier l'architecture la plus performante. Les résultats obtenus montrent que cette approche permet de réduire significativement les incertitudes de mesure par rapport aux méthodes traditionnelles, ouvrant ainsi des perspectives pour des applications industrielles en microélectronique.

Abstract

The continuous miniaturization of electronic components, which underpins modern technologies, requires the ability to measure structures at the nanometer scale. At such dimensions, achieving accurate measurements becomes particularly challenging, as each measurement technique has its own limitations and sources of error.

This thesis proposes a so-called *hybrid metrology* approach, which consists of combining several complementary measurement techniques applied to the same sample in order to improve overall accuracy. The data obtained from these different techniques are then analyzed using artificial intelligence methods, in particular neural networks, which are able to efficiently exploit their complementarity.

A major challenge of this approach lies in the large amount of data required to train these models. However, experimental measurements are costly and time-consuming. To address this issue, this work relies on the generation of realistic synthetic data, produced using simulators that incorporate both the physical properties of the studied structures and the specific imperfections of the measurement instruments.

The developed models were trained on these data and evaluated to identify the most efficient architecture. The results show that this approach significantly reduces measurement uncertainties compared to traditional methods, thus opening up new perspectives for industrial applications in microelectronics.

Table des matières

REMERCIEMENTS ***5***

RESUME ***7***

ABSTRACT ***9***

TABLE DES MATIERES ***11***

INTRODUCTION ***15***

CHAPITRE I - OUTILS ET TECHNIQUES DE LA METROLOGIE ***19***

I.1 LA METROLOGIE 19

I.1.1 Concept de la métrologie *19*
I.1.2 Distributions statistiques *22*
I.1.3 Diagramme quantile-quantile *25*

I.2 LES TECHNIQUES METROLOGIQUES 27

I.2.1 Contexte de l'étude *27*
I.2.2 Microscopie à force atomique (AFM) *28*
I.2.3 Microscopie électronique à balayage (SEM) *32*
I.2.4 Scatterométrie (Optical Critical Dimension : OCD) *35*
I.2.5 Diffusion des rayons-X aux petits angles (SAXS) *39*
I.2.6 Synthèse des techniques métrologiques *42*
I.2.7 Technique de référence - Microscopie Électronique en Transmission *44*

I.3 LA METROLOGIE HYBRIDE 47

I.3.1 Les liens entre les données *47*
I.3.2 Les architectures d'hybridation *48*

I.4 CONCLUSIONS 51

CHAPITRE II - SIMULATIONS DES TECHNIQUES DE METROLOGIE ***53***

II.1 INTRODUCTION 53

II.2 MODELISATION DES ECHANTILLONS 54

II.2.1 Géométrie des échantillons *54*

II.2.2 Valeurs des paramètres dimensionnels *55*
II.2.3 Variabilité des échantillons *56*

II.3 SIMULATIONS AFM 59

II.3.1 Conception du simulateur *59*
II.3.2 Implémentation des incertitudes de mesure propre à l'AFM *67*
II.3.3 Analyse des paramètres dimensionnels extraits des simulations AFM *69*
II.3.4 Conclusion *71*

II.4 SIMULATIONS CDSEM 72

II.4.1 Principe de fonctionnement du simulateur *72*
II.4.2 Implémentation des incertitudes de mesures du CDSEM *73*
II.4.3 Analyse des paramètres dimensionnels extraits des simulations CDSEM *74*
II.4.4 Conclusion *76*

II.5 SIMULATIONS OCD 76

II.5.1 Principe de fonctionnement du simulateur *76*
II.5.2 Etudes pour le paramétrage du modèle *77*
II.5.3 Implémentation des incertitudes de mesure *85*
II.5.4 Analyse des paramètres dimensionnels extraits des simulations OCD *90*
II.5.5 Conclusion *92*

II.6 SIMULATIONS SAXS 93

II.6.1 Principe de fonctionnement du simulateur *93*
II.6.2 Etudes pour le paramétrage du modèle *94*
II.6.3 Implémentation des incertitudes de mesure *100*
II.6.4 Analyse des paramètres dimensionnels extraits des simulations SAXS *105*
II.6.5 Conclusion *109*

II.7 SYNTHESE DES TRAVAUX DE SIMULATIONS 109

II.7.1 Les résultats *109*
II.7.2 Les corrélations des erreurs *110*

II.8 CONCLUSIONS 111

CHAPITRE III - METROLOGIE HYBRIDE PAR RESEAU DE NEURONES 113

III.1 FONCTIONNEMENT DES RESEAUX DE NEURONES ARTIFICIELS 114

III.1.1 Le neurone artificiel *114*
III.1.2 Organisation en réseau *116*
III.1.3 Préparation des données *116*
III.1.4 Fonction de perte *118*

III.1.5 Processus d'apprentissage *119*
III.1.6 Suivi et évaluation de l'apprentissage *120*

III.2 CHOIX DE L'ARCHITECTURE DU RESEAU DE NEURONES 122

III.2.5 Évaluation initiale des performances de l'hybridation *123*
III.2.6 Architecture Multi-Branches *126*
III.2.7 Analyse comparative des deux architectures *129*
III.2.8 Estimation de la taille minimale du jeu de données *130*

III.3 HYBRIDATION DES DONNEES SIMULEES 133

III.3.5 L'apprentissage *133*
III.3.6 Résultats *134*

III.4 HYBRIDATION DE DONNEES BRUTES 138

III.4.5 Conception du réseau de neurone *138*
III.4.6 Choix des sous-réseaux suivant les données *140*
III.4.7 Application des réseaux de neurones à l'extraction de profils depuis les spectres OCD *143*

III.5 VERS UNE METROLOGIE HYBRIDE FONDEE SUR L'APPRENTISSAGE AUTOMATIQUE 146

CONCLUSIONS ET PERSPECTIVES ***149***

CONCLUSIONS 149

PERSPECTIVES 151

BIBLIOGRAPHIES ***153***

Introduction

La microélectronique occupe aujourd'hui une place centrale dans nos sociétés modernes. Elle est omniprésente dans une multitude d'applications allant des objets connectés du quotidien jusqu'aux satellites en orbite tout en passant par les infrastructures industrielles ou encore les équipements médicaux. Cette omniprésence s'accompagne d'exigences technologiques toujours croissantes, portées par la miniaturisation des dispositifs. La miniaturisation permettant à terme d'accroître les performances et de réduire la consommation énergétique des dispositifs. Depuis plusieurs décennies, cette dynamique a été décrite par la loi de Moore, qui annonçait le doublement du nombre de transistors sur une même surface tous les dix-huit à vingt-quatre mois. Si cette tendance historique atteint aujourd'hui ses limites physiques, la complexité des architectures, la diversité des matériaux et la densité d'intégration continuent de progresser, proposant de nouvelle réponse mais aussi posant des défis inédits à chaque étape de la conception et de la fabrication.

Dans ce contexte de constante évolution, la métrologie joue un rôle clé. En effet, elle constitue un maillon essentiel pour garantir le contrôle, la fiabilité et la reproductibilité des procédés de fabrication. Mesurer avec exactitude les dimensions, la forme, la composition ou les propriétés structurelles est un enjeu majeur pour les industriels. À l'échelle nanométrique, les marges d'erreur tolérables se réduisent drastiquement, tandis que les besoins en rapidité sur les lignes de production augmentent. La métrologie non destructive s'impose ainsi comme une approche incontournable, capable de concilier hautes performances métrologiques et contraintes industrielles.

Malgré les avancées constantes des techniques de mesure, il apparaît de plus en plus clairement qu'aucune d'entre elles ne peut, à elle seule, répondre à l'ensemble des exigences actuelles de la micro et nano-fabrication. La diversité des géométries, la complexité des matériaux, la tridimensionnalité des structures et les impératifs de cadence imposent le recours à plusieurs techniques complémentaires. Chacune d'entre elles ne dispose de points forts mais aussi de limites, qu'il s'agisse : de résolution, de temps de mesure, de caractère invasif ou de capacité à extraire certains paramètres. Cette situation conduit la communauté scientifique à explorer des stratégies plus globales, capables de combiner intelligemment les forces de différentes approches pour accroitre la robustesse des mesures.

C'est dans ce cadre que s'inscrit la métrologie hybride, qui exploite simultanément plusieurs sources d'information issues de techniques distinctes afin de produire des estimations plus robustes, plus fidèle et plus juste des paramètres d'intérêt. Encore relativement récente dans l'industrie microélectronique, cette approche ouvre des perspectives prometteuses, notamment en environnement de production, où elle

permet d'optimiser l'usage des équipements tout en augmentant la pertinence des données collectées. Toutefois, combiner les mesures de techniques variées ne va pas sans soulever de véritables défis comme : assurer la cohérence des données, intégrer les incertitudes propres à chacune des techniques ou encore développer des approches capables de tirer parti de la diversité des approches. L'essor de l'Intelligence Artificielle (IA) et des méthodes d'apprentissage automatique offre de nouveaux outils pour relever ces défis. En effet, l'analyse statistique avancée, l'extraction de caractéristiques, la reconnaissance de motifs ou la prédiction de paramètres à partir de jeux de données complexes sont désormais accessibles grâce à des algorithmes performants et adaptatifs. Le machine learning, et plus particulièrement les réseaux de neurones artificiels, permet d'envisager des approches où le modèle d'estimation est appris directement à partir des données, sans recourir à une modélisation physique explicite.

Cependant, dans le domaine spécifique de la microélectronique, l'un des freins majeurs à l'application de ces approches réside dans la disponibilité limitée de données expérimentales. Contrairement à d'autres secteurs où les bases de données sont abondantes et facilement accessibles, les mesures métrologiques sont souvent soumises à des contraintes de confidentialité ou de disponibilité et obtenir des données dans un but purement d'apprentissage reste très coûteux et long à réaliser. Ce nombre restreint de données constitue un obstacle important au développement et à la validation de modèles d'apprentissage automatique performants.

Pour surmonter cette difficulté, les travaux présentés dans ce mémoire explorent l'utilisation de simulateurs numériques afin de générer artificiellement des jeux de données réalistes. Bien que ces outils soient déjà employés dans l'industrie pour modéliser des systèmes ou prédire des résultats de mesure, leur détournement vers la production massive de données d'entraînement destinées à l'IA reste encore marginal. L'objectif est ici de compenser le manque de données réelles par des données synthétiques enrichies et fidèles aux conditions expérimentales. Cette démarche repose sur une modélisation rigoureuse des échantillons, l'identification des principales sources d'erreurs et leur implémentation dans l'environnement de simulation.

En somme, cette approche, avec d'un côté le développement d'une méthodologie de métrologie hybride fondée sur l'intelligence artificielle et de l'autre la conception d'une chaîne de génération de données simulées réalistes, constitue le cœur de ce travail de thèse. L'un des objectifs majeurs de ces travaux étant de démontrer la faisabilité et l'intérêt de l'hybridation en métrologie, en montrant que des modèles robustes peuvent être entraînés sur des données simulées et l'appliquer à des cas réels. Pour cela, le mémoire est structuré en trois grandes parties. La première présente l'état de l'art des techniques métrologiques utilisées. Ce référencement passe par les principes physiques de chaque technique ainsi que leurs atouts et leurs limites. Un point spécifique sur les concepts clés de la métrologie hybride est aussi référencé. La

deuxième partie est consacrée à la modélisation tant des échantillons que des simulateurs. Elle explicite leur construction et adaptation à notre problématique, et couvre l'implémentation des différentes sources d'erreurs pour produire un jeu de données d'apprentissage réaliste. Enfin, la troisième partie concerne la mise en œuvre des modèles de machine learning pour l'hybridation des données. Dans celle-ci, l'évaluation des performances de l'hybridation est effectuée ainsi qu'une identification des pistes d'amélioration pour de futures applications industrielles.

Chapitre I - Outils et Techniques de la Métrologie

La métrologie est une part importante du domaine de la microélectronique. Il faut noter que dans ce domaine le travail est effectué à l'échelle nanométrique. Ce chapitre propose, dans un premier temps, d'en revoir les concepts ainsi que les outils mathématiques qui y sont liées. Ainsi, nous pourrons appréhender les démarches entreprises au cours de ces travaux de thèse avec une base commune de compréhension. Nous aborderons, dans un second temps, les différentes techniques utilisées dans le cadre de ce projet, en détaillant leurs spécificités, leurs avantages ainsi que leurs limitations. Enfin, nous introduirons les notions de métrologie hybride par le machine learning.

I.1 La métrologie

I.1.1 Concept de la métrologie

Par définition, la métrologie est la science de la mesure [1], elle joue un rôle fondamental dans de nombreux domaines scientifiques, industriels et technologiques. Elle vise à garantir les mesures, en établissant des méthodes normalisées et des références universelles. La métrologie permet aussi le suivi et la validation de processus de production en assurant la conformité aux normes en vigueur. Ainsi, elle constitue un pilier de la recherche scientifique, de l'innovation technologique et du contrôle de qualité.

En métrologie, il est essentiel de faire la distinction entre la « **valeur vraie** » et la « **valeur de référence** ». La valeur vraie désigne la valeur théorique idéale et absolue et, par nature, inatteignable par une mesure expérimentale. En effet, toute technique métrologique comporte diverses sources d'erreurs, aussi minimes soient elles. Par conséquent, nous utilisons une valeur de référence, définie à partir d'étalons, de mesures expérimentales précises ou de calculs théoriques. Cette valeur de référence sert de repère fiable bien qu'elles comportent des erreurs elles-mêmes évaluées en amont.

Pour évaluer la qualité d'une mesure, elle est comparée à sa valeur de référence associée. Plusieurs éléments permettent d'indiquer la qualité d'une mesure, on retrouve la « **justesse** » et la « **fidélité** » (en anglais : accuracy et precision) qui sont des notions clés en métrologie [2].

La **justesse** reflète la capacité de se rapprocher de la valeur de référence, tout en tenant compte des éventuelles sources d'erreur [2]. Une faible justesse indique la présence d'un biais dans les valeurs mesurées, qui devra être corrigé ou, à défaut, pris en considération. Les Figure I.1.A et Figure I.1.B illustrent des mesures présentant une faible justesse, non centrées sur la cible représentant la valeur de référence. Les Figure I.1.C et Figure I.1.D, montrent quant à elles des mesures à forte justesse, bien centrées sur la valeur de référence. Le graphique sous chaque cible représente la distribution des distances entre la valeur de référence et les point de mesure. Un des outils permettant d'évaluer la justesse est la moyenne, car elle permet d'obtenir le centrage d'une distribution et ainsi de calculer le décalage global, aussi appelé biais, d'une population par rapport à une référence :

$$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i \qquad \textit{(I.1)}$$

μ est la moyenne de la population
N est l'effectif de la population
x_i est la valeur de la mesure i

La **fidélité** fait, elle, référence aux écarts observés lors de la répétition d'une même mesure [2]. Contrairement à la justesse, elle est associée à une erreur non systématique et aléatoire. Les Figure I.1.B et Figure I.1.D, présentant une faible dispersion des valeurs mesurées sont des exemples de mesures à forte fidélité. À l'inverse, les Figure I.1.A et Figure I.1.C, marquées par une forte dispersion, représentent des mesures à faible fidélité. Les distributions sous les cibles sont dans ce cas plus ou moins fines en fonction de la fidélité. Un des outils permettant de quantifier cela est l'écart-type. Il permet de calculer la dispersion des points d'une population :

$$\sigma = \sqrt{\frac{\sum(x_i - \mu)^2}{N}} \qquad \textit{(I.2)}$$

σ est l'écart-type de la population
μ est la moyenne de la population
N est l'effectif de la population
x_i est la valeur de la mesure i

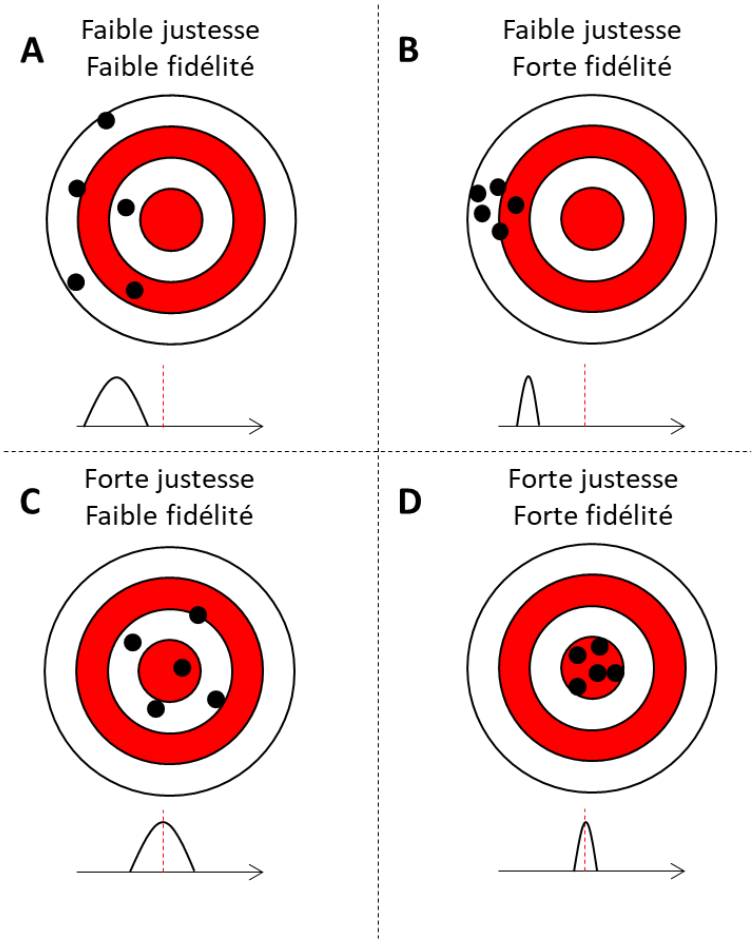

Figure I.1 : Représentation des principes de justesse et de fidélité de mesure au travers de cibles dont le centre correspond à la valeur de référence. Sous chaque cas, la densité de probabilité associée.

D'autre part, on retrouve en métrologie deux notions fondamentales, la **répétabilité** et la **reproductibilité**, car elles permettent d'évaluer la fiabilité et la stabilité d'un système de mesure.

La **répétabilité** désigne l'aptitude d'un système de mesure à produire des résultats très proches lorsque la même mesure est répétée dans des conditions rigoureusement identiques. Cela implique que l'opérateur, le protocole, l'équipement, l'environnement et l'objet mesuré restent constants [2]. En pratique, garantir des conditions parfaitement stables est souvent difficile, en particulier sur de longues périodes. C'est pourquoi les tests de répétabilité sont généralement réalisés dans un laps de temps court, dont la durée est déterminée selon le type de mesure. La répétabilité est généralement quantifiée par des indicateurs statistiques entre plusieurs séries de mesures tels que l'écart-type, l'erreur-type ou l'intervalle de variation des résultats obtenus. On distingue deux types de répétabilité :

- **La répétabilité statique**, qui consiste à effectuer plusieurs mesures successives sans repositionnement de l'échantillon,

- **La répétabilité dynamique**, où l'échantillon est déplacé puis réaligné entre chaque mesure. Dans ce cas, des erreurs supplémentaires peuvent être introduites par les étapes de positionnement. Il est alors nécessaire de définir une métrique spécifique, par exemple une erreur de repositionnement ou un écart de centrage, pour évaluer et corriger ces écarts.

La **reproductibilité**, quant à elle, mesure la capacité à obtenir des résultats comparables lorsque certaines conditions expérimentales varient. Typiquement, cela concerne des changements d'opérateur, d'équipement ou d'environnement. L'unique contrainte est que l'échantillon mesuré reste strictement identique ou, à défaut, équivalent du point de vue métrologique. La reproductibilité est souvent utilisée pour évaluer la robustesse d'une méthode sur le long terme, par exemple lors de comparaisons inter-laboratoires ou d'évaluations croisant différentes techniques de mesure [3].

I.1.2 Distributions statistiques

Les distributions statistiques sont des outils fondamentaux en métrologie pour la description et l'exploitation des données expérimentales. Elles permettent de représenter la répartition des valeurs mesurées, de modéliser les erreurs ou perturbations affectant les mesures (bruit, biais, incertitudes), mais également de quantifier la variabilité observée.

Par ailleurs, les distributions offrent un cadre probabiliste permettant d'estimer la probabilité d'occurrence d'un résultat, d'établir des intervalles de confiance, ou encore de prédire le comportement de futures mesures dans des conditions similaires.

I.1.2.1 Distribution normale

La **distribution normale**, aussi appelée distribution gaussienne, permet de modéliser de nombreux phénomènes naturels. Elle joue un rôle important en métrologie en raison de sa fréquente apparition dans les erreurs de mesures ou dans des probabilités d'apparition d'évènement. Les paramètres de cette distribution sont la **moyenne (μ)** et **l'écart-type (σ)**. Cette distribution est d'écrite par la fonction de probabilité suivante :

$$f_{\mu,\sigma}(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \qquad \textbf{(I.3)}$$

μ est la moyenne de la population
σ est l'écart-type de la population
$\boldsymbol{x}$ est la variable aléatoire continue représentant la valeur mesurée

Une illustration de la distribution normale est donnée par la Figure I.2. Sur cette figure, plusieurs zones sont délimitées par des segments rouges, représentant des intervalles d'écarts-types. En métrologie, ces zones servent de référence pour estimer la dispersion des mesures et quantifier les erreurs potentielles.

- L'intervalle allant de $-\sigma$ à $+\sigma$ contient environ **68 %** des observations.
- L'intervalle compris entre -2σ et $+2\sigma$ englobe **95 %** des données.
- L'intervalle allant de -3σ à $+3\sigma$ couvre près de **99,7 %** des observations.

Ces intervalles et leurs pourcentages associés permettent d'évaluer la probabilité qu'une mesure s'écarte de la moyenne, offrant ainsi un cadre statistique robuste pour l'analyse des incertitudes. Cette distribution est notamment utilisée dans la suite de ces travaux pour cela.

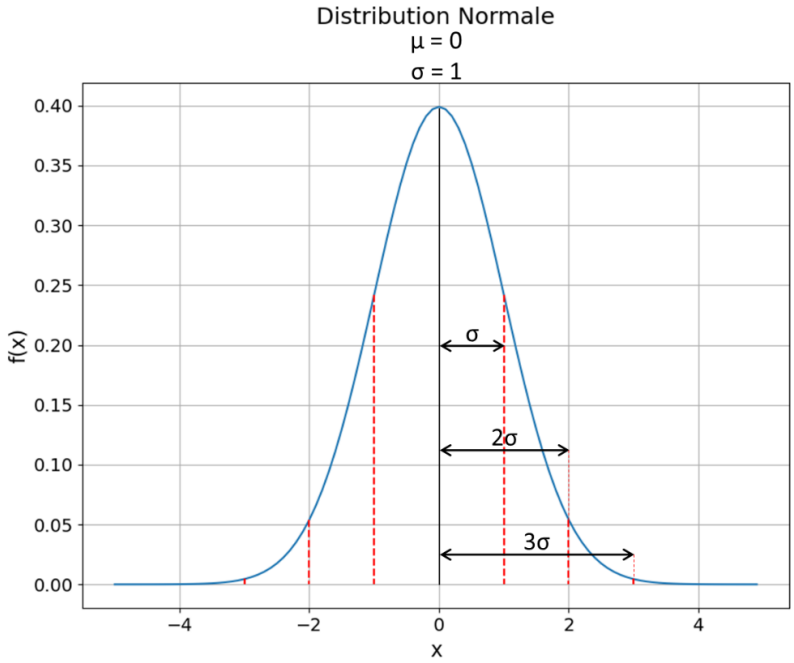

Figure I.2 : Représentation de la distribution normale « centrée réduite » (avec μ=0 et σ=1)

I.1.2.2 Distribution de Poisson

La **distribution de Poisson** est couramment utilisée pour modéliser les événements discrets qui se produisent de manière aléatoire dans le temps ou l'espace. En métrologie, elle est particulièrement utile lorsqu'il s'agit de modéliser le comptage d'événements rares, comme le nombre de défauts dans un processus de fabrication ou les fluctuations aléatoires dans des systèmes de mesure. Elle est également utilisée pour modéliser le bruit de grenaille (bruit quantique), notamment dans la détection de photons et d'électrons, où les fluctuations aléatoires suivent souvent cette distribution,

c'est dans ce cadre que nous avons eu l'occasion de l'utiliser notamment sur la partie simulatrice.

La distribution est donnée par la fonction de probabilité suivante :

$$f_\lambda(k) = \frac{\lambda^k}{k!} e^{-\lambda} \qquad \textbf{(I.4)}$$

k est la variable aléatoire discrète et positive représentant le nombre d'événements observés
λ est le paramètre de la distribution, qui correspond à l'espérance mathématique

Pour illustration, la Figure I.3 représente la distribution de Poisson pour trois valeurs de λ : 1, 6, et 10. On observe qu'avec un λ faible (λ = 1), la distribution est fortement asymétrique et concentrée autour des petites valeurs de k (correspondant au nombre d'évènement) observé). À mesure que λ augmente, la distribution s'élargit et tend vers une forme presque symétrique, se rapprochant de la distribution normale.

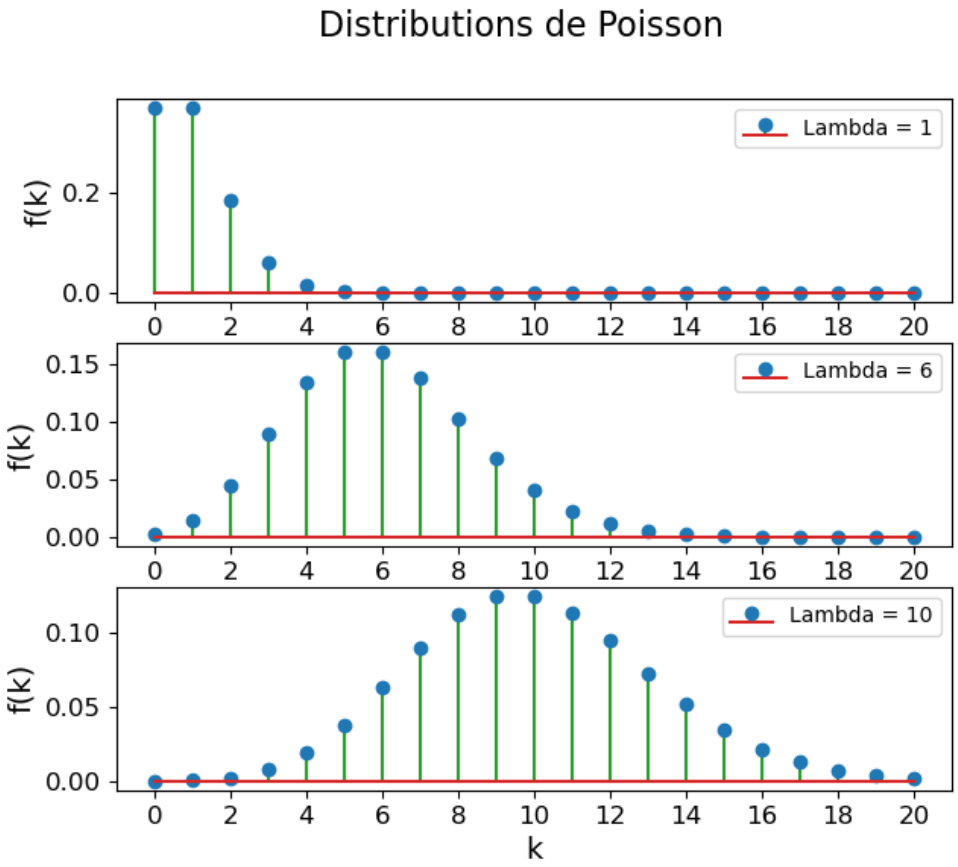

Figure I.3 : Représentation de la distribution de Poisson pour différentes valeurs de λ

I.1.2.3 Distribution Uniforme

La **distribution uniforme** est utilisée pour modéliser un phénomène dans lequel toutes les valeurs possibles dans un intervalle donné ont la même probabilité d'occurrence. Contrairement à la distribution normale ou de Poisson, elle est caractérisée par une probabilité constante dans l'intervalle défini par les bornes a et b. Cette distribution est décrite par l'équation suivante :

$$f_{a,b}(x) = \begin{cases} \dfrac{1}{b-a} & pour\ a \leq x \leq b \\ 0 & sinon \end{cases} \qquad (I.5)$$

La Figure I.4 illustre la distribution uniforme entre a=0 et b=10, où la probabilité est constante sur cet intervalle et nulle en dehors. En métrologie, elle est utilisée pour **générer des mesures aléatoires sans biais** dans une plage de valeurs spécifiée, garantissant que chaque valeur a la même probabilité d'occurrence. Cela permet de simuler des scénarios de test aléatoires, comme nous pourrons le voir plus tard.

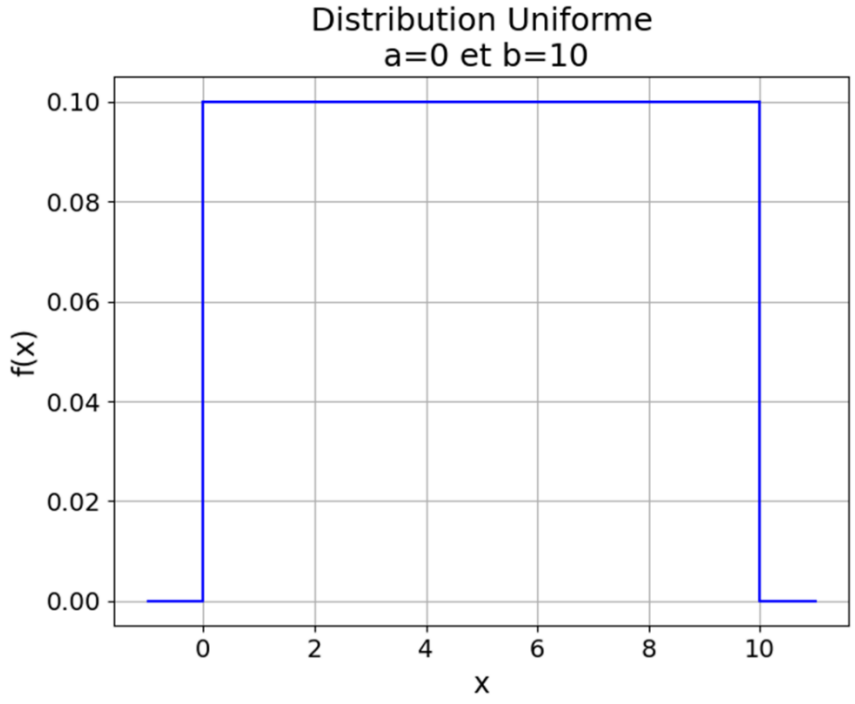

Figure I.4 : Représentation de la distribution uniforme (avec a=0 et b=10)

I.1.3 Diagramme quantile-quantile

L'outil que nous utilisons pour comparer deux distributions entre elles est le diagramme quantile-quantile ou QQ-plot (*quantile-quantile plot* en anglais). Il permet de comparer la répartition d'un ensemble de données à celle d'une distribution théorique de référence, comme la loi normale, la loi de Poisson ou la loi uniforme.

Le principe repose sur la comparaison des **quantiles**, c'est-à-dire des valeurs qui divisent un ensemble de données selon une proportion donnée. Dans un QQ-plot, on place en abscisse les quantiles d'une distribution théorique, et en ordonnée ceux des données observées. Si les deux distributions sont similaires, les points obtenus s'alignent alors selon une droite. Cet outil est particulièrement utile pour détecter des écarts entre la distribution des données et le modèle théorique supposé. Une courbure, une pente différente ou des points très éloignés de la diagonale peuvent indiquer, par exemple, une asymétrie dans les données, une dispersion trop faible ou trop grande, ou la présence de valeurs extrêmes (outliers).

La Figure I.5 illustre un QQ-plot comparant un jeu de données expérimental à une distribution normale théorique. On observe que la majorité des points suivent une droite, ce qui suggère que la loi normale constitue un bon modèle pour ces données. Ce qui est confirmé par la valeur du coefficient de détermination R^2=0,9963, indiquant une correspondance très forte entre les deux distributions (R^2=1 étant une correspondante parfaite).

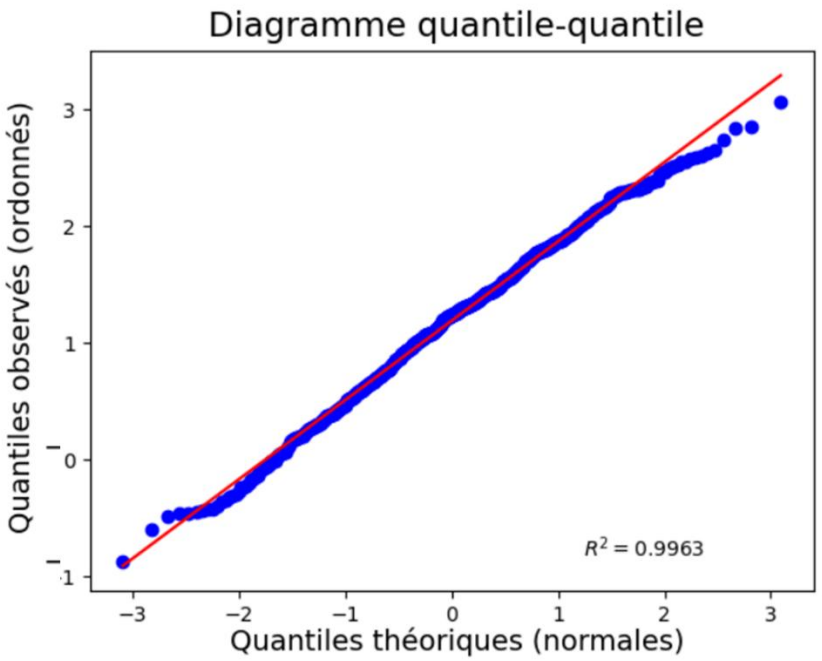

Figure I.5 : Diagramme quantile-quantile entre la distribution d'un jeu de donnée expérimentale et une distribution normale théorique.

L'analyse des distributions statistiques, associée à l'utilisation des QQ-plots, constitue une base méthodologique essentielle pour l'interprétation des mesures. Ces outils nous permettront, dans la suite de ces travaux, d'estimer les lois suivies par les données expérimentales, de modéliser les erreurs et les bruits d'origine stochastique.

I.2 Les Techniques Métrologiques

I.2.1 Contexte de l'étude

Les travaux présentés dans cette thèse portent sur la mesure des paramètres dimensionnels de réseaux de lignes de silicium, structures périodiques gravées qui entrent notamment dans la fabrication des transistors. Les réseaux de lignes sont des structures utilisées pour la métrologie en microélectronique comme des cibles représentant une image de ce qui est fabriqué au niveau des puces. Les paramètres étudiés, illustrés par la Figure I.6, sont :

- La **dimension critique (CD - Critical Dimension)**
- La **hauteur (H)**
- L'**angle de paroi (SWA - Side Wall Angle)**

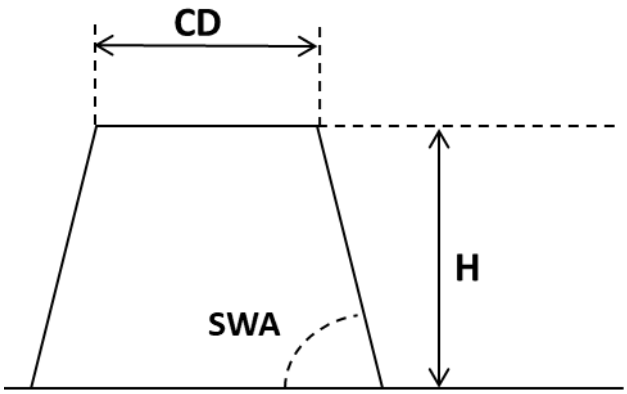

Figure I.6 : Schéma d'une ligne de silicium

Ces paramètres permettent de caractériser la géométrie des structures gravées et de suivre l'évolution des procédés de fabrication. Leur mesure contribue à garantir la conformité des motifs par rapport aux spécifications visées.

Cette étude repose sur l'utilisation des **techniques métrologiques dimensionnelles utilisée couramment en microélectronique**, incluant la microscopie à force atomique (AFM), la microscopie électronique à balayage pour le contrôle dimensionnel (CD-SEM) et l'ellipsométrie spectroscopique par diffraction (OCD) mais aussi d'une technique actuellement moins industrielle la diffusion des rayons X aux petits angles (SAXS) choisie pour son potentiel et son émergence récente dans la littérature de la microélectronique [4]. Ces techniques permettent d'obtenir une description détaillée de la structure des réseaux de lignes, chacune apportant des informations spécifiques et contribuant à une caractérisation plus précise et robuste. L'objectif de ces travaux est d'améliorer la justesse et la fidélité des mesures

dimensionnelles appliquées aux réseaux de lignes, dont les différentes techniques métrologiques sont présentées dans la suite de cette section.

I.2.2 Microscopie à force atomique (AFM)

La microscopie à force atomique (**AFM**, pour *Atomic Force Microscope* en anglais) est une technique de référence en microélectronique. Elle appartient à la famille des microscopies à sonde locale et permet de mesurer diverses propriétés des matériaux, notamment électriques, morphologiques, mécaniques, thermiques, chimiques et magnétiques [5], [6], [7]. Dans le cadre de notre étude, nous nous focaliserons sur l'acquisition de la morphologie des échantillons, en obtenant des représentations directes de leur forme.

I.2.2.1 Principes

Le principe de fonctionnement de l'AFM repose sur l'interaction atomique entre une pointe (ou sonde) et la surface de l'échantillon. La Figure I.7 illustre le balayage de l'échantillon par la pointe, qui suit fidèlement la topographie de sa surface. La flexion de la pointe, plus précisément celle de son support appelé levier, est suivie grâce à un système optique. En effet, un faisceau laser réfléchit au dos du levier est capté par un capteur photodiode dit « quatre-quadrant », qui permet d'analyser les variations du signal lumineux et ainsi de remonter aux déformations de la pointe suivant son déplacement.

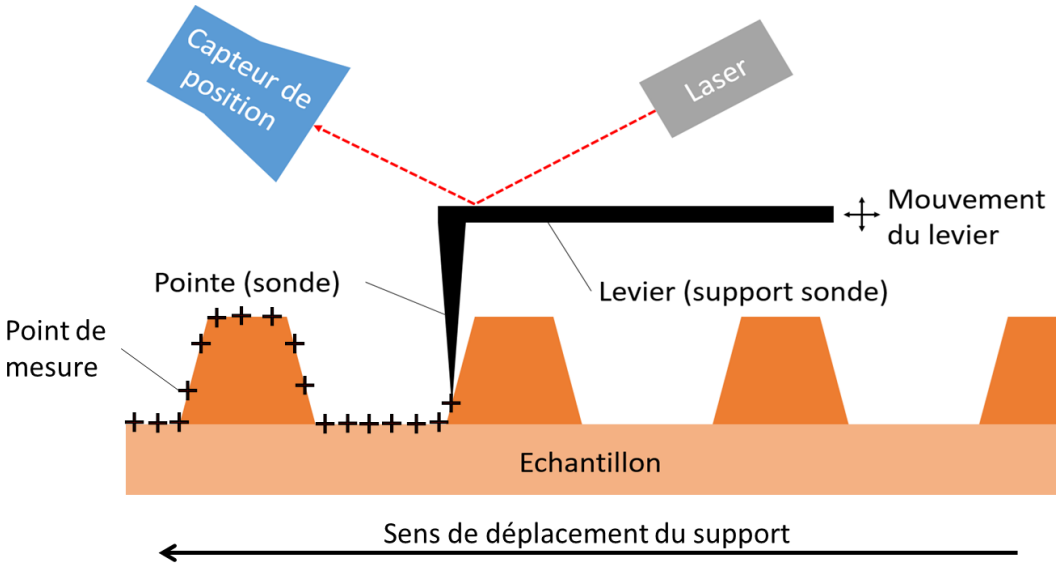

Figure I.7 : Schéma de principe du fonctionnement d'un AFM

Il existe une multitude de modes de fonctionnement de l'AFM, chacun étant adapté à l'analyse de propriétés spécifiques des échantillons. Parmi ces modes, on trouve notamment les modes de contact, de contact intermittent ou encore les modes à force modulée, permettant de caractériser des propriétés mécaniques, électriques ou magnétiques.

Dans le cadre de ces travaux l'AFM a été plus spécifiquement utilisé avec le **mode dit « CD »** (*Critical Dimension*), qui est particulièrement adapté à l'analyse des structures micro et nanométriques dans le domaine de la microélectronique. Ce mode de fonctionnement est basé sur le mode de contact intermittent, également connu sous le nom de "tapping mode". Dans ce mode, le levier supportant la pointe est mis en oscillation à sa fréquence propre, généralement de l'ordre de quelques centaines de kilohertz. L'oscillation est finement régulée de manière que la pointe touche ponctuellement la surface de l'échantillon. Ce mode permet de réduire les risques d'endommagement de l'échantillon, de limiter l'usure prématurée de la pointe et de prévenir l'accumulation de matériaux sur celle-ci. De plus, le fait de relever la pointe après chaque point de contact permet d'éviter des artefacts liés à une éventuelle adhésion de la pointe au matériau, améliorant ainsi la résolution latérale.

Pour exploiter pleinement le mode **CD**, des pointes AFM spécifiques, dotées d'une morphologie calibrée, sont utilisées. Ces pointes, également appelées **pointes CD, schématisées par la** Figure I.8, présentent une particularité essentielle : leur extrémité est **évasée**. Cette conception permet de suivre précisément le profil d'une ligne en évitant que le corps de la pointe n'entre en contact avec celle-ci, contrairement aux pointes standard, qui adoptent généralement une forme conique ou cylindrique.

Le mode CD se distingue par sa capacité à effectuer des déplacements complexes suivant tous les axes, ce qui est essentiel pour la mesure des profils de motifs en trois dimensions. Contrairement aux modes classiques, où la pointe suit simplement la topographie de la surface, le mode CD permet un recul perpendiculaire à la surface mesurée. Ce procédé repose sur la répétition d'une séquence de mouvements spécifiques illustrée par la Figure I.9:

- **Etape 1** : La pointe effectue une première approche jusqu'au contact avec la surface.
- **Etape 2** : Un mouvement de retour est réalisé.
- **Etape 3** : Un déplacement parallèle à la surface est ensuite exécuté

Ce cycle d'étapes est répété pour sonder l'échantillon sur toute une longueur X avant un déplacement sur l'axe Y.

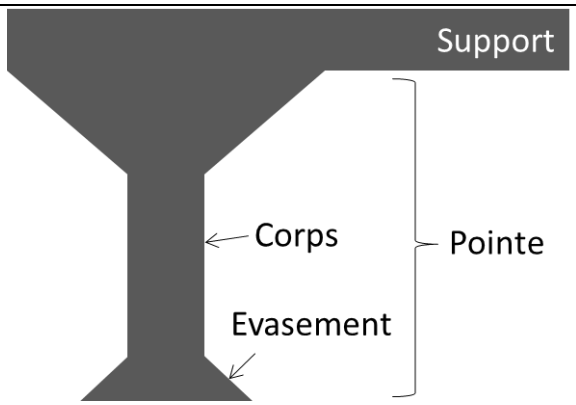

Figure I.8 : Représentation schématique d'une pointe AFM pour le mode CD.

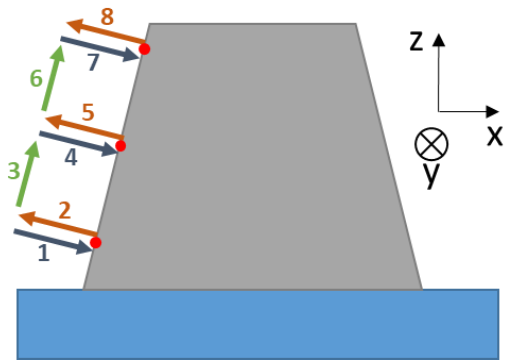

Figure I.9 : Etapes du principe de déplacement de la pointe pour le mode CD.

Afin d'optimiser l'exactitude des mesures, la pente locale de l'échantillon est recalculée à chaque point, garantissant ainsi une correction dynamique du déplacement de la sonde. Ce processus est répété sur l'ensemble de la zone d'analyse afin d'obtenir une reconstruction 3D détaillée et fidèle de la surface des structures étudiées. Un exemple de forme complète obtenue sur un réseau de lignes est visible sur la Figure I.10.

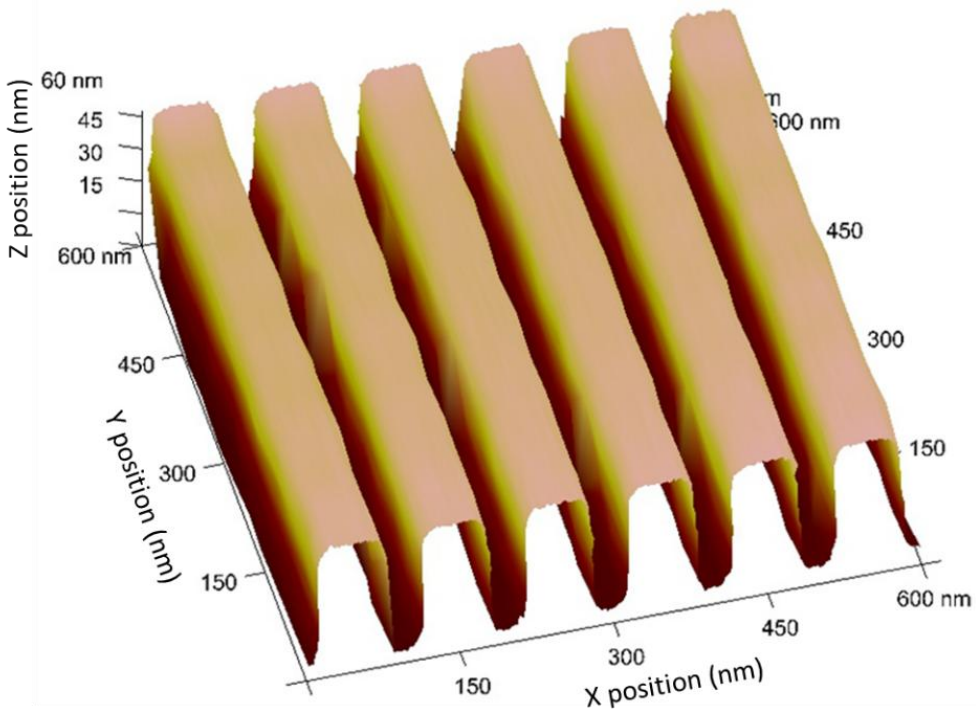

Figure I.10 : Reconstruction 3D d'un échantillon mesuré par AFM (en mode CD).

I.2.2.2 Avantages et Limites

Le microscope à force atomique (AFM) présente plusieurs avantages et inconvénients en tant que technique de métrologie en nanoélectronique. En effet, c'est une technique de **métrologie dite directe**, ce qui signifie que l'extraction des paramètres dimensionnels peut être effectuée directement à partir des données brutes, à l'aide de logiciels de traitement d'image. Cela permet une interprétation immédiate des résultats et garantit une grande exactitude, ce qui en fait une technique de référence dans de nombreux travaux [8], [9].

Par ailleurs, l'AFM est une technique relativement lente, avec des durées d'acquisition pouvant facilement dépasser une heure. Cette durée dépend bien entendu de la surface à analyser ainsi que de la résolution souhaitée. Ce temps d'acquisition élevé constitue un frein majeur à son intégration dans un environnement de production.

L'AFM présente également des limitations liées à la morphologie des structures. En effet, comme illustré sur la Figure I.11, en fonction de la géométrie de l'échantillon, la pointe peut ne pas atteindre certaines zones, ce qui entraîne des imprécisions ou une impossibilité de mesurer certaines topographies.

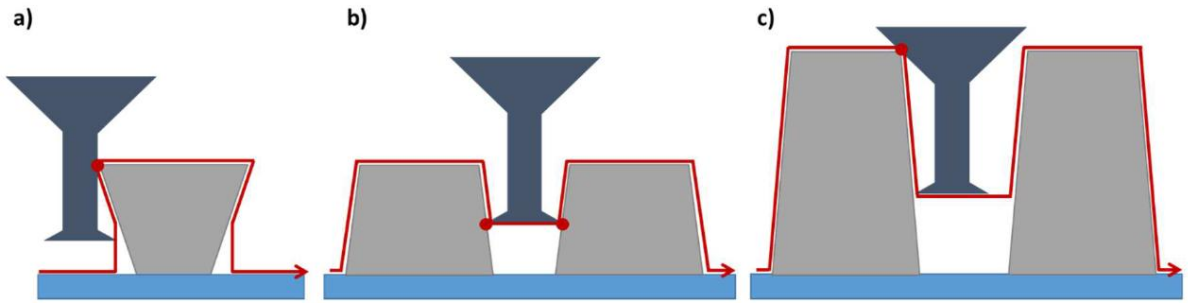

Figure I.11 : Limite d'accès des pointes évasées dans le cas a) d'un SWA trop important, b) d'un espacement entre 2 motif trop faible, c) une longueur de pointe plus petite que la hauteur des motifs. Figures issues des travaux de J. Rêche [10]

In fine, bien que l'AFM soit une méthode de métrologie précise et directe, son temps d'acquisition long et ses contraintes d'accessibilité doivent être pris en compte dans son utilisation pour la caractérisation des nanostructures.

Dans un contexte de chaîne de production, cette technique est utilisée de manière ponctuelle afin de valider un procédé. Pour les besoins du contrôle en continu, d'autres techniques métrologiques sont privilégiées, telles que la scatterométrie, qui sera présentée plus loin dans ce chapitre.

I.2.3 Microscopie électronique à balayage (SEM)

Le **microscope électronique à balayage (SEM - Scanning Electron Microscope)** permet d'obtenir des images haute résolution d'un échantillon grâce au balayage de sa surface à l'aide d'un faisceau focalisé d'électrons [11]. En interagissant avec la matière, ce faisceau génère divers signaux fournissant des informations sur la morphologie ou la composition chimique. Grâce à sa résolution nanométrique, le SEM est largement utilisé en science des matériaux, biologie et nanotechnologies.

Une variante du SEM, le **microscope électronique à balayage de dimension critique (CDSEM - Critical Dimension Scanning Electron Microscope)** [8], [9]**,** est spécifiquement conçu pour la métrologie dimensionnelle dans l'industrie des semi-conducteurs , plus spécifiquement le terme dimension critique étant utilisé pour désigner les largeurs minimales de motifs dont le respect est essentiel au bon fonctionnement des dispositifs. Contrairement au SEM classique, utilisé pour l'analyse qualitative, le CDSEM permet des mesures précises et reproductibles de structures nanométriques, et ce grâce à un référencement au système de mesure international. Dans notre étude, l'analyse des structures est menée à l'aide du mode « vue du dessus » (top view), configuration principalement répandue sur les CDSEM.

I.2.3.1 Principes

L'imagerie en SEM repose sur les interactions entre les électrons et la matière, illustrées en Figure I.12. Ces interactions, se divisent en deux catégories : élastiques, où les électrons changent de direction sans perte d'énergie, et inélastiques, où une partie de l'énergie est transférée à la matière. Ces deux types d'interactions donnent lieu à des signaux détectables différents, chacun apportant des informations spécifiques sur la composition ou la morphologie de l'échantillon. Parmi les émissions on retrouve comme illustré sur la Figure I.12 :

- **Les électrons secondaires (jaune)**, émission inélastique, expulsés lors de l'impact des électrons incidents (Figure I.12), sont majoritairement utilisés pour la formation des images.
- **Les électrons Auger (violet)**, émission inélastique, spécifiques au matériau, résultent de transitions électroniques internes (Figure I.12).
- **Les rayons X (orange)**, émission inélastique, émis plus en profondeur, résultent de la désexcitation électronique (Figure I.12) et peuvent aussi provenir d'un phénomène de freinage des électrons incidents.
- **Les électrons rétrodiffusés (rouge)**, émission élastique, issus d'interactions élastiques entre les électrons incidents et les noyaux atomiques, sont réfléchis vers la surface avec une énergie relativement élevée

(Figure I.12). Ils fournissent des informations sur la **composition chimique** de l'échantillon.

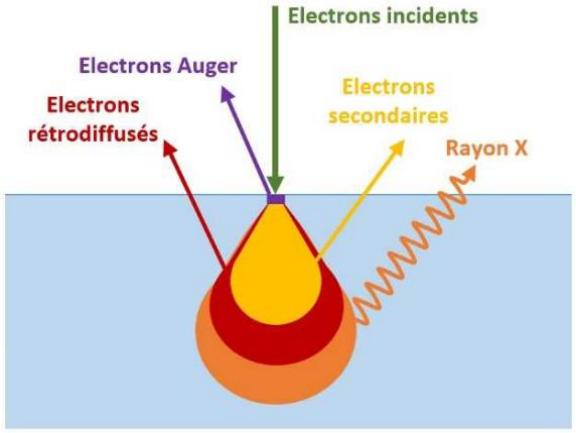

Figure I.12 : Représentation des différentes interactions electrons-matiere sous forme de poire d'interaction. Figure issue des travaux de j. Reche [9]

Dans les CDSEM, comme illustré en Figure I.13, le faisceau d'électrons est focalisé sur la surface de l'échantillon grâce à des lentilles électroniques. Ce faisceau balaie la surface, entraînant l'émission d'électrons secondaires, détectés par un détecteur Everhart-Thornley. Ce dernier est couramment utilisé, il repose sur un scintillateur qui convertit l'impact des électrons en photons. Ces derniers sont ensuite reconvertis en électrons par une photocathode, ce qui permet l'amplification et le traitement du signal.

L'image obtenue est en niveaux de gris, comme visible sur la Figure I.14 et repose principalement sur la détection des électrons secondaires, particulièrement sensibles à la morphologie de la surface de l'échantillon. L'intensité de chaque pixel de l'image correspond à la quantité d'électrons détectés à une position donnée lors du balayage du faisceau. Ainsi, plus le nombre d'électrons secondaires émis et collectés est élevé, plus le pixel associé apparaît clair, ce qui permet de reconstituer visuellement les reliefs et les détails topographiques de la surface analysée.

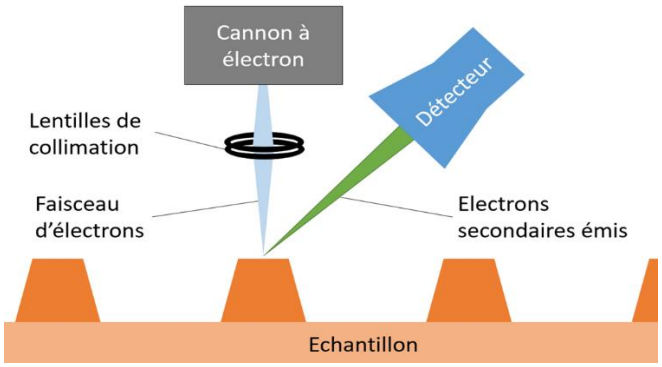

Figure I.13 : Schéma de principe du fonctionnement d'un microscope électronique à balayage.

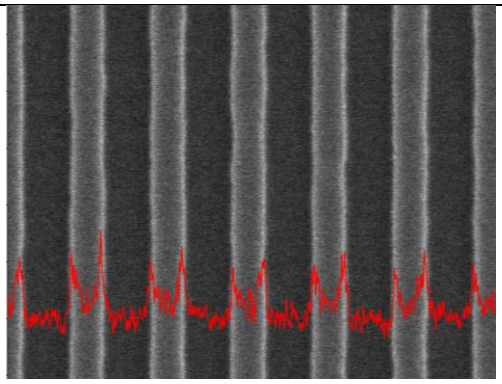

Figure I.14 : Exemple d'image obtenue par CDSEM en mode « Top View ». En rouge est superposé le signal moyen sur la longueur des lignes.

L'exploitation des données issues du CDSEM repose sur l'analyse de contraste obtenu. Grâce aux électrons secondaires, l'imagerie met en évidence la morphologie et les variations topographiques. Les logiciels de traitement d'image permettent d'extraire des paramètres clés tels que la largeur des motifs, leur homogénéité et leurs éventuelles déformations.

I.2.3.2 Avantages et Limites

L'un des principaux atouts du CDSEM est qu'il constitue une méthode d'analyse directe, ne nécessitant ni préparation complexe de l'échantillon ni traitement indirect des données. Son temps d'acquisition de quelques secondes par image permet une visualisation immédiate et une mesure précise des structures à l'échelle nanométrique, un atout essentiel pour le contrôle qualité et la métrologie des dispositifs avancés en microélectronique. Toutefois, la visualisation "Top View" (vue du dessus) ne permet d'extraire, dans les paramètres qui nous intéressent, que le CD (Critical Dimension), sans accès aux informations sur la hauteur et le SWA des structures analysées.

Malgré ses avantages, le CDSEM présente certaines limitations techniques pouvant affecter l'exactitude des mesures. Le bruit de grenaille [14], lié à la nature discrète des électrons, peut induire des variations aléatoires dans l'image, réduisant la fiabilité des mesures. De plus, cette technique peut être invasive, notamment pour les matériaux sensibles aux irradiations électroniques, entraînant des modifications structurelles indésirables, telle que la génération de défauts ponctuels [15], [16].

Un autre défi est le chargement électrique de l'échantillon, particulièrement problématique sur les matériaux isolants. L'accumulation de charges peut perturber le faisceau et dégrader la qualité des images [17], nécessitant parfois des solutions comme des revêtements conducteurs. La taille du faisceau électronique impose également une

limite à la résolution de quelques nanomètres et à la précision des mesures, car la diffusion des électrons peut altérer la netteté de l'acquisition.

Après avoir présenté le CDSEM, méthode directe à haute résolution mais localisée et potentiellement destructive, nous abordons l'OCD (Optical Critical Dimension). Cette technique indirecte et non destructive permet des mesures rapides et globales.

I.2.4 Scatterométrie (Optical Critical Dimension : OCD)

L'**OCD** pour **Optical Critical Dimension** en anglais, aussi appelé **scatterométrie**, est une technique métrologie indirecte. Les données en sortie d'équipement ne sont pas des images des échantillons. Elles nécessitent la modélisation de la structure afin d'extraire les paramètres désirés. Cette méthode s'est largement développée dans le domaine de la microélectronique au cours des vingt dernières années.

I.2.4.1 Principes

L'OCD est une technique de métrologie optique exploitant les phénomènes de diffusion et de diffraction d'un faisceau lumineux sur une structure périodique afin d'en extraire des informations. Plusieurs paramètres sont à prendre en compte lors de l'analyse des données de sortie : l'angle d'incidence (**Φ**), sa polarisation et la longueur d'onde de la lumière utilisée (**λ**).

Cette technique métrologique dispose de deux modes principaux. Le premier mode est la réflectométrie, qui consiste à analyser les coefficients de réflexion de la lumière. Cette méthode permet une analyse rapide, mais fournit une quantité d'informations limitée, en particulier dans le cas de structures complexes ou multicouches. Le second mode est l'ellipsométrie, exploite les variations de l'état de polarisation de la lumière réfléchie ou diffractée, généralement sous différents angles d'incidences, en mesurant précisément les paramètres elliptiques Ψ et Δ. Dans le cadre de notre étude, nous privilégierons l'approche ellipsométrique.

Comme illustré en Figure I.15, cette technique repose dans notre cas sur l'utilisation d'une source de lumière blanche couvrant un spectre continu sur une plage définie de longueurs d'onde, typiquement de l'ultraviolet au proche infrarouge. On définit le plan d'incidence comme le plan géométrique formé par le rayon lumineux incident et la normale à la surface d'échantillon. Par la suite, la lumière est polarisée selon deux composantes : **p**, parallèle à ce plan, et **s**, perpendiculaire à celui-ci. Après interaction avec la structure périodique de l'échantillon, l'état de polarisation du faisceau réfléchi

est modifié. Ces modifications sont alors analysées par l'OCD, qui mesure les variations relatives d'amplitude et de phase entre les composantes p et s.

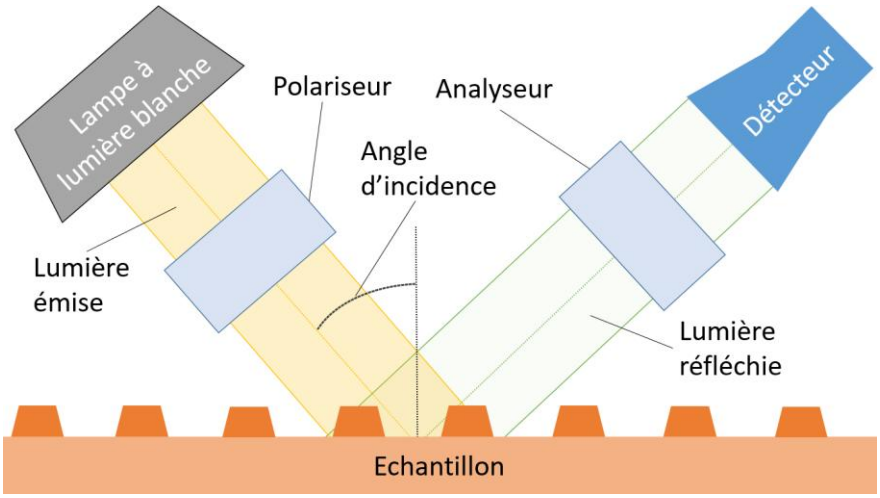

Figure I.15 : Schéma de principe du fonctionnement de la scatterométrie

La lumière blanche incidente polarisée et celle réfléchie par l'échantillon peuvent être décrites respectivement à l'aide des champs électriques E^i (incident) et E^r (réfléchi). Ces champs peuvent être décomposés selon les axes de polarisation **s** et **p**, comme l'illustrent la Figure I.16 ainsi que les équations suivantes :

$$E^i = E_s^i + E_p^i \qquad \textit{(I.6)}$$

$$E^r = E_s^r + E_p^r \qquad \textit{(I.7)}$$

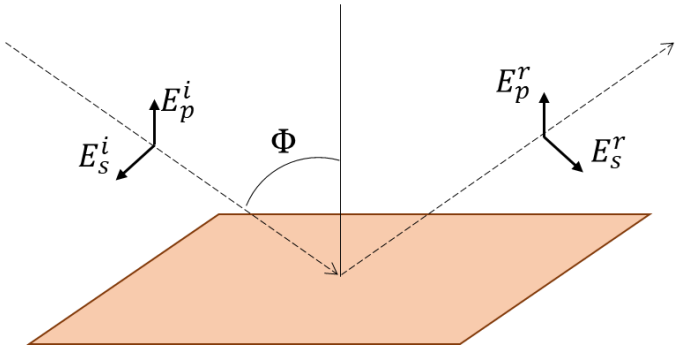

Figure I.16 : Représentation des composantes parallèles et perpendiculaires des faisceaux incidents et réfléchis

L'ellipsométrie qui repose sur la mesure des variations d'amplitude, calcul les rapports entre les champs électriques associés aux deux composantes de polarisation. Ces variations peuvent être décrites par l'expression suivante :

$$r_p = \frac{E_p^r}{E_p^i} \quad \textbf{(I.8)}$$

$$r_s = \frac{E_s^r}{E_s^i} \quad \textbf{(I.9)}$$

Les paramètres $\mathbf{r_p}$ et $\mathbf{r_s}$ appelés coefficients de Fresnel, permettent de calculer divers couples de spectres décrivant les variations d'amplitude et de phase des champs électriques selon les axes de polarisation. Dans le cadre de ce travail, les grandeurs optiques utilisées sont les composantes de réflexion **Rp** et **Rs**. Ce choix est motivé par la nature des données expérimentales que nous avons à disposition, mesuré par l'ellipsomètre du CEA-Leti, illustré par l'exemple de la Figure I.17 et calculé selon l'expression suivante :

$$R_p = |r_p|^2 \quad \textbf{(I.10)}$$

$$R_s = |r_s|^2 \quad \textbf{(I.11)}$$

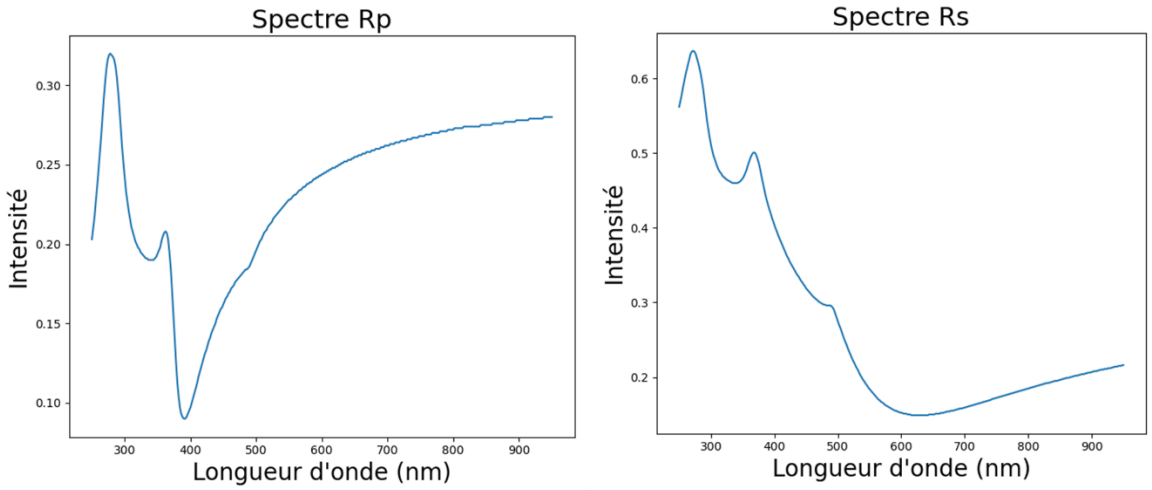

Figure I.17 : Exemple de spectres Rp, Rs pour les longueurs d'onde allant de 250 nm à 950 nm, obtenu via simulateur

Les spectres obtenus après mesure constituent les données brutes. Afin de les exploiter efficacement, ils sont comparés à des spectres de référence. En l'absence de bibliothèque de référence, nous avons recours à des algorithmes de résolution de problème inverse couplé à des simulateurs [18], [19], [20].

Ces algorithmes permettent d'estimer les paramètres initiaux du système, en l'occurrence les paramètres dimensionnels, à partir des grandeurs de sortie, c'est-à-dire les spectres mesurés. Ce processus repose sur la comparaison entre les données expérimentales et des spectres simulés, obtenus à partir d'un modèle géométrique donné via un simulateur. Un algorithme d'optimisation ajuste alors les paramètres du

modèle afin de minimiser l'erreur entre la simulation et l'expérience, selon une métrique définie. A la suite d'une multitude de cycle de simulations et d'évaluation, nous retrouvons les paramètres dimensionnels associés aux spectres Rp et Rs.

I.2.4.2 Avantages et Limites

L'Optical Critical Dimension (OCD) est une technique de métrologie extrêmement rapide (pour la mesure), ce qui la rend particulièrement utile pour le contrôle lors des différentes phases de production. En effet, sa rapidité d'exécution permet d'effectuer un grand nombre de mesures en un temps réduit, optimisant ainsi l'efficacité des processus industriels, notamment dans le domaine de la microélectronique et de la nanofabrication. De plus, cette méthode est non destructive, ce qui représente un atout majeur par rapport à d'autres techniques nécessitant un contact physique avec l'échantillon ou une préparation invasive.

Cependant, bien que la mesure de l'OCD soit rapide, l'exploitation des données mesurées requiert des temps de traitement pouvant être lourds. Allant de quelques minutes dans le cas de l'utilisation de bibliothèque de références à plusieurs dizaines de minutes dans le cas de la résolution de problème inverse. Cela dépendant du nombre de paramètres recherchés et donc laissés libres sur le modèle. Comme dans toutes résolutions par problème inverse, si le nombre de paramètres est trop important ou avec des plages de variation trop importantes, de multiples cas, avec une erreur très faible, peuvent répondre correctement et donc en théorie il est possible d'obtenir de multiples solutions.

Un autre inconvénient majeur de l'OCD réside dans sa résolution spatiale limitée. Contrairement à des techniques telles que l'AFM ou au CDSEM, qui permettent d'obtenir des mesures locales précises sur un élément spécifique, l'OCD ne permet pas de caractériser une structure individuelle. Le faisceau lumineux utilisé ayant un diamètre de plusieurs micromètres, il englobe simultanément plusieurs motifs, rendant impossible la distinction d'une ligne unique au sein du réseau. Les mesures obtenues reflètent donc une moyenne des caractéristiques globales du réseau de lignes.

Ainsi, bien que l'OCD soit une technique rapide, non destructive et bien adaptée au suivi de production, elle présente certaines limitations en termes d'exactitude locale et nécessite une modélisation avancée pour interpréter correctement les données obtenues. Par ailleurs, sa capacité à détecter de petites variations dimensionnelles est intrinsèquement limitée par les longueurs d'ondes utilisées : une réduction de celles-ci permet d'améliorer la sensibilité métrologique, en particulier pour les dimensions critiques nanométriques.

C'est dans ce contexte que la diffusion des rayons X aux petits angles (SAXS) constitue une méthode complémentaire pertinente, en offrant une sensibilité accrue

aux variations fines de géométrie grâce à l'utilisation de rayons X à très courte longueur d'onde.

I.2.5 Diffusion des rayons-X aux petits angles (SAXS)

Le **SAXS** pour Small-Angle X-Ray Scattering en anglais (ou diffusion des rayons-X aux petits angles en français) est une technique métrologique indirecte récente dans l'industrie de la microélectronique [21]. En raison de l'emploi de photons de faible longueur d'onde (rayon X), cette méthode se révèle particulièrement adaptée à l'étude d'objets de très petite taille. Depuis une dizaine d'années, diverses études exploitent cette technique afin de caractériser la morphologie des lignes dans les réseaux, donnant ainsi naissance à une nouvelle terminologie : le CD-SAXS [21], [22], [23].

I.2.5.1 Principes

Le SAXS fait intervenir les interactions entre les rayons X et la matière. Il existe une multitude d'interactions, représentée par la Figure I.18. Nous retrouvons trois types :

- **Les photons absorbés** : Ils entraînent l'excitation des électrons, qui peuvent alors suivre deux cas distincts. Ils peuvent soit être expulsés, comme dans le cas des **électrons Auger**, soit retourner à leur état fondamental en libérant de l'énergie. Cette restitution d'énergie se manifeste sous la forme de photons X de moindre énergie, appelés **photons de fluorescence**, dont les niveaux d'énergie sont caractéristiques des atomes présents dans l'échantillon [24].
- **Les rayons X transmis** : Ils sont ceux qui traversent la matière sans subir d'interaction. Dans ce cas, ni leur énergie ni leur direction ne sont altérées, un phénomène que l'on qualifie de « non-interaction ».
- **Les photons diffusés** : Cette dernière interaction se divise en deux composantes : une diffusion inélastique, impliquant une perte d'énergie et connue sous le nom de diffusion de Compton [25], et une diffusion élastique, sans perte d'énergie, désignée sous le terme de diffusion Thomson [26]. Ces diffusions peuvent se produire selon différentes directions, notamment vers l'avant (forward scattering) ou vers l'arrière (backscattering), selon l'angle d'interaction entre le photon incident et l'électron.

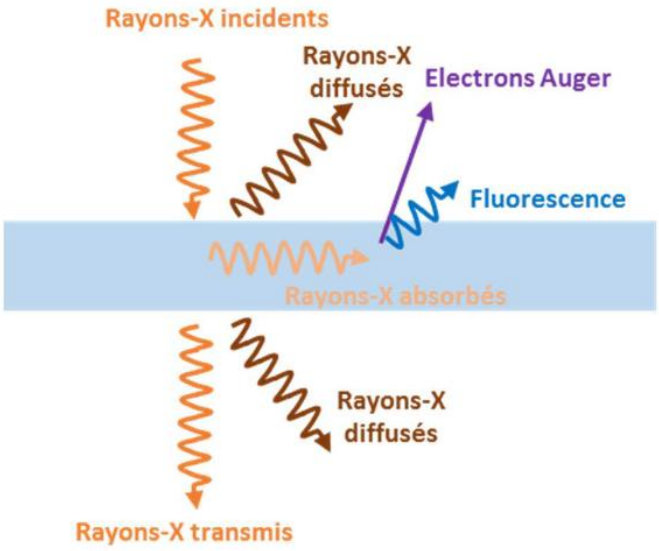

Figure I.18 : Principales interactions rayons-x matière, provenant des travaux de thèse de J. Rêche.

Dans le cadre de nos travaux, nous avons utilisé le SAXS en transmission du fait du mode de fonctionnement de l'équipement utilisé et des échantillons déjà mesurés. La configuration SAXS en transmission est illustrée à la Figure I.19. Un faisceau de rayons X est envoyé sur l'échantillon sous un angle d'incidence **ω**. Les rayons transmis à travers l'échantillon sont collectés sur un détecteur bidimensionnel, permettant de former une **image de diffraction selon les coordonnées q_x et q_y**.

En faisant varier l'angle ω, on accède à différentes composantes du vecteur de diffusion $\vec{q}$, permettant de reconstruire une portion de l'**espace réciproque**, qui contient l'information tridimensionnelle de la diffraction et donc par analogie de la géométrie de l'échantillon.

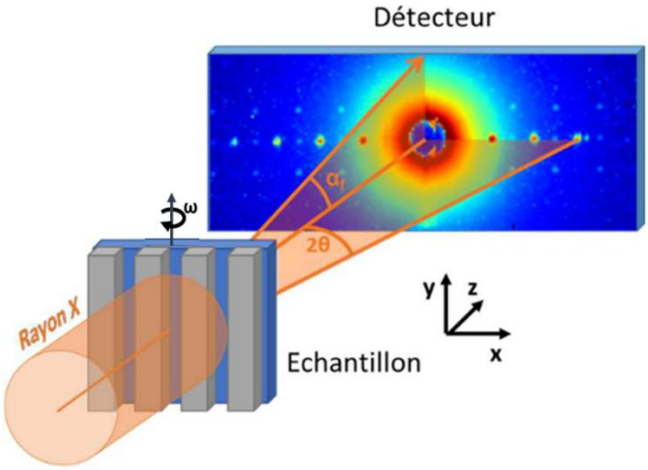

Figure I.19 : Schéma de principe de la technique SAXS par transmission, pour ω=0°, adapté des travaux de G. Freychet [27]

Nos analyses sont restreintes au plan (q_x, q_z), en fixant q_y=0 nm^{-1}. En effet, les échantillons sont constitués de lignes supposées infinies selon l'axe y, celui-ci n'apporte donc aucune information sur les paramètres dimensionnels d'intérêts pour notre étude. Ainsi, seule la diffraction selon x (axe de la périodicité) et z (l'axe de la hauteur des lignes) est pertinente pour la caractérisation.

L'espace réciproque est donc reconstruit à partir d'une série d'images acquises à différents angles ω, chaque image contribuant à une coupe dans le plan (q_x, q_z), comme l'illustre la Figure I.20.

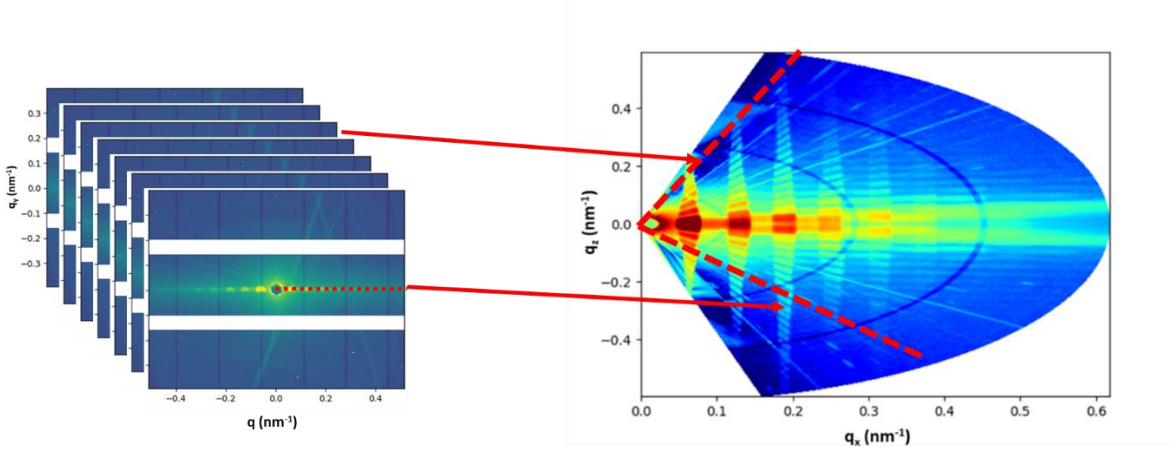

Figure I.20 : Construction d'un espace réciproque à partir de mesures de figures de diffraction, issue des travaux de J. Rêche.

Pour extraire les informations recherchées, un prétraitement peut s'avérer nécessaire afin de minimiser les erreurs et le bruit présents dans les figures de diffraction. Comme pour l'OCD, l'extraction des paramètres géométriques repose sur la résolution d'un problème inverse. L'algorithme est itératif : à chaque étape, il génère un ensemble de simulations du motif de diffraction, reconstruit l'espace réciproque correspondant, puis le compare aux mesures expérimentales. Ce processus permet d'ajuster les paramètres du modèle géométrique jusqu'à convergence avec les données observées.

I.2.5.2 Avantages et Limites

La durée de mesure en SAXS est directement liée au temps d'exposition défini par l'opérateur. Une meilleure exactitude nécessite un temps d'exposition plus long, car la qualité du signal dépend du nombre de photons récoltés et donc diffusés par l'échantillon, lui-même conditionné par la brillance du faisceau incident. Plus cette brillance est élevée, plus le flux de photons collectés est important, ce qui améliore le rapport signal sur bruit et réduit le temps nécessaire à l'acquisition.

Il existe différents types de sources de rayons X, chacun présentant des avantages et des inconvénients. Les sources synchrotron, par leur très forte brillance, permettent des temps d'acquisition très courts (de quelques secondes à quelques minutes) avec une excellente qualité de signal. À l'inverse, les sources de laboratoire conventionnelles offrent une brillance plus faible, impliquant des temps d'exposition plus longs (allant jusqu'à plusieurs heures). Pour jouer sur ce temps d'acquisition il est possible de changer la taille du faisceau pour optimiser l'éclairage de l'échantillon. Un faisceau trop étroit réduit la structure analysée et le nombre de photons interagissant, limitant ainsi la quantité d'informations recueillies. À l'inverse, un faisceau trop large permet une meilleure interaction mais risque d'introduire du bruit parasite dans les figures de diffraction et d'engendrer l'apparition d'artefacts si des motifs annexes présents dans l'échantillon sont éclairés.

La distance entre l'échantillon et le détecteur joue également un rôle important, car elle influence la résolution angulaire et la plage de vecteurs de diffusion q accessibles. Une distance plus grande permet d'observer des petits angles de diffusion (faibles q), utiles pour détecter des structures de grande taille, tandis qu'une distance échantillon-détecteur plus courte favorise l'accès aux grands angles (forts q), plus sensibles aux détails fins. Ainsi, un compromis doit être trouvé afin d'illuminer la structure de manière optimale, en maximisant l'intensité du signal sans compromettre la fiabilité des mesures.

Afin de situer cette technique parmi les méthodes métrologiques disponibles, il est proposé une synthèse des approches abordées, en mettant en évidence leurs complémentarités, leurs limites respectives et les contextes dans lesquels elles sont les plus adaptées.

I.2.6 Synthèse des techniques métrologiques

Plusieurs techniques métrologiques ont été mobilisées dans le cadre de ces travaux de thèse afin de caractériser les structures nanométriques de type lignes. Bien qu'elles partagent un objectif commun : l'extraction de paramètres dimensionnels (CD, H et SWA), ces techniques reposent sur des principes de mesure fondamentalement différents. Chacune présente des atouts et des limitations, en termes de résolution, de temps d'acquisition, de temps de traitement, ou de disponibilité.

Nous pouvons distinguer 2 types de technique :

- Les techniques directes : L'AFM et le CDSEM, dont les données brutes sont exploitables directement ou nécessitant peu de traitement.

- Les techniques indirectes : L'OCD et le SAXS, dont les données brutes ne peuvent être utilisées directement, nécessitant de lourd traitement tel que la résolution de problème inverse.

Le Tableau I.1 présente un comparatif des principales techniques de métrologie utilisées pour l'extraction de la Critical Dimension (CD), en indiquant l'ordre de grandeur des écarts-types (σ) des erreurs de mesure rapportés dans la littérature [9], [27-31], ainsi que les temps d'acquisition et de traitement associés [9], [32-35]. Les valeurs d'écart-type correspondent à la composante aléatoire de l'incertitude obtenue dans des conditions expérimentales représentatives et doivent être interprétées comme des ordres de grandeur plutôt que comme des performances absolues.

On observe que les techniques directes, comme l'AFM et le CD-SEM, offrent une mesure physique du profil des structures. L'AFM présente des écarts-types compris entre 0,75 et 2,6 nm, mais son temps d'acquisition peut varier de quelques minutes à plusieurs heures selon la zone balayée et la résolution souhaitée, avec un traitement souvent rapide (quelques secondes à quelques minutes) lorsque l'analyse est automatisée. Le CD-SEM, avec des écarts-types de l'ordre de 1,2 à 1,5 nm, affiche généralement des temps d'acquisition et de traitement de quelques minutes.

On peut voir dans le Tableau I.1 que les techniques indirectes, telles que l'OCD et le SAXS, se distinguent par des acquisitions plus rapides. L'OCD présente typiquement des écarts-types compris entre 0,78 et 1,6 nm. Son exactitude dépend fortement de la qualité de la modélisation optique et de la justesse des paramètres utilisés. Les acquisitions se font souvent en quelques secondes, mais l'étape de traitement, qui repose sur la résolution d'un problème inverse, peut durer de quelques minutes à plusieurs heures selon la complexité des modèles et l'exactitude recherchée. Le SAXS peut atteindre des écarts-types aussi faibles que 0,25 nm. Le temps d'acquisition varie de quelques secondes à quelques minutes, et le traitement, basé sur l'ajustement de modèles physiques aux données, requiert également de quelques minutes à plusieurs dizaines de minutes.

La connaissance de ces ordres de grandeur constitue ainsi une base de travail et une référence pour évaluer, comparer et combiner les techniques dans le cadre d'une approche de métrologie hybride visant à réduire les incertitudes globales.

Technique métrologique	Ecart-type des erreurs (σ) sur CD	Ordre de grandeur du temps d'acquisition	Ordre de grandeur du temps de traitement
AFM	0,75 – 2,6 nm	Quelques minutes à heures	Quelques secondes à minutes
CDSEM	1,2 - 1,5 nm	Quelques minutes	Quelques minutes
OCD	0,78 – 1,6 nm	Quelques secondes	Quelques minutes à plusieurs heures
SAXS	0,25 nm	Quelques secondes à minutes	Quelques minutes à plusieurs heures

Tableau I.1 : Comparaison des techniques de métrologie pour l'extraction du CD : écarts-types des erreurs, temps d'acquisition et de traitement, d'après la littérature [9], [27-35]

I.2.7 Technique de référence - Microscopie Électronique en Transmission

La Microscopie Électronique en Transmission (**TEM**, pour *Transmission Electron Microscopy*) est une technique d'imagerie de haute résolution permettant d'explorer la structure interne des matériaux à l'échelle nanométrique, voire atomique [37], [38].

Dans le cadre de cette thèse, le TEM est employé afin d'obtenir les **valeurs de référence** pour les paramètres dimensionnels. Parmi l'ensemble des techniques mises en œuvre, le TEM est considéré comme **la plus précise** en raison de sa résolution élevée et de sa capacité à visualiser des structures à l'échelle atomique. Toutefois, cette méthode présente une contrainte majeure : elle nécessite la **destruction des échantillons.** En effet, elle nécessite une étape de découpage en lamelles fines des échantillons. Par conséquent, les mesures en TEM doivent être réalisées en dernier, après l'acquisition des données obtenues par d'autres techniques non destructives telles que l'AFM, le CD-SEM, l'OCD et le SAXS. Cette approche garantit une corrélation optimale entre les différentes méthodes de caractérisation tout en préservant la possibilité d'effectuer plusieurs analyses sur un même échantillon avant sa destruction. Une limitation importante du TEM, dans le cadre de cette thèse, réside dans le faible nombre de lignes analysées par échantillon, souvent une seule, ce qui limite la représentativité des mesures obtenues.

I.2.7.1 Principes

Le TEM utilise un faisceau d'électrons accélérés qui traverse un échantillon ultrafin pour produire une image à haute résolution de sa structure interne. Comme illustré sur la Figure I.21, le faisceau est émis par un canon à électrons, condensé par un système de lentilles électromagnétiques, puis focalisé sur l'échantillon. Ce dernier doit être préparé avec une extrême exactitude, son épaisseur étant généralement inférieure à une centaine de nanomètres, afin de permettre la transmission d'un nombre suffisant d'électrons.

Dans le domaine de la microélectronique, la préparation d'échantillons pour la TEM repose le plus souvent sur la technique du faisceau d'ions focalisés (Focused Ion Beam – FIB). Celle-ci permet d'extraire localement une lamelle dans la zone d'intérêt, puis de l'amincir progressivement jusqu'à l'épaisseur requise, tout en limitant les dommages induits par le bombardement ionique.

Lors de l'analyse, une partie des électrons traverse l'échantillon sans interaction significative (électrons transmis), tandis que d'autres sont diffusés ou absorbés. Les électrons transmis sont collectés par la lentille objectif, qui forme une image intermédiaire ensuite agrandie par les lentilles de projection avant d'être observée sur un écran fluorescent.

Le TEM est particulièrement adapté aux observations en coupe transversale (*cross-section*), permettant de visualiser directement le profil des structures et de mesurer avec exactitude des paramètres dimensionnels tels que le Critical Dimension (CD), la hauteur (H) ou l'angle des flancs (Side Wall Angle, SWA). L'exemple présenté en Figure I.22 illustre ce type d'acquisition : la coupe nanométrique d'un motif y révèle un niveau de détail suffisant pour permettre des mesures directes et fiables de ses dimensions.

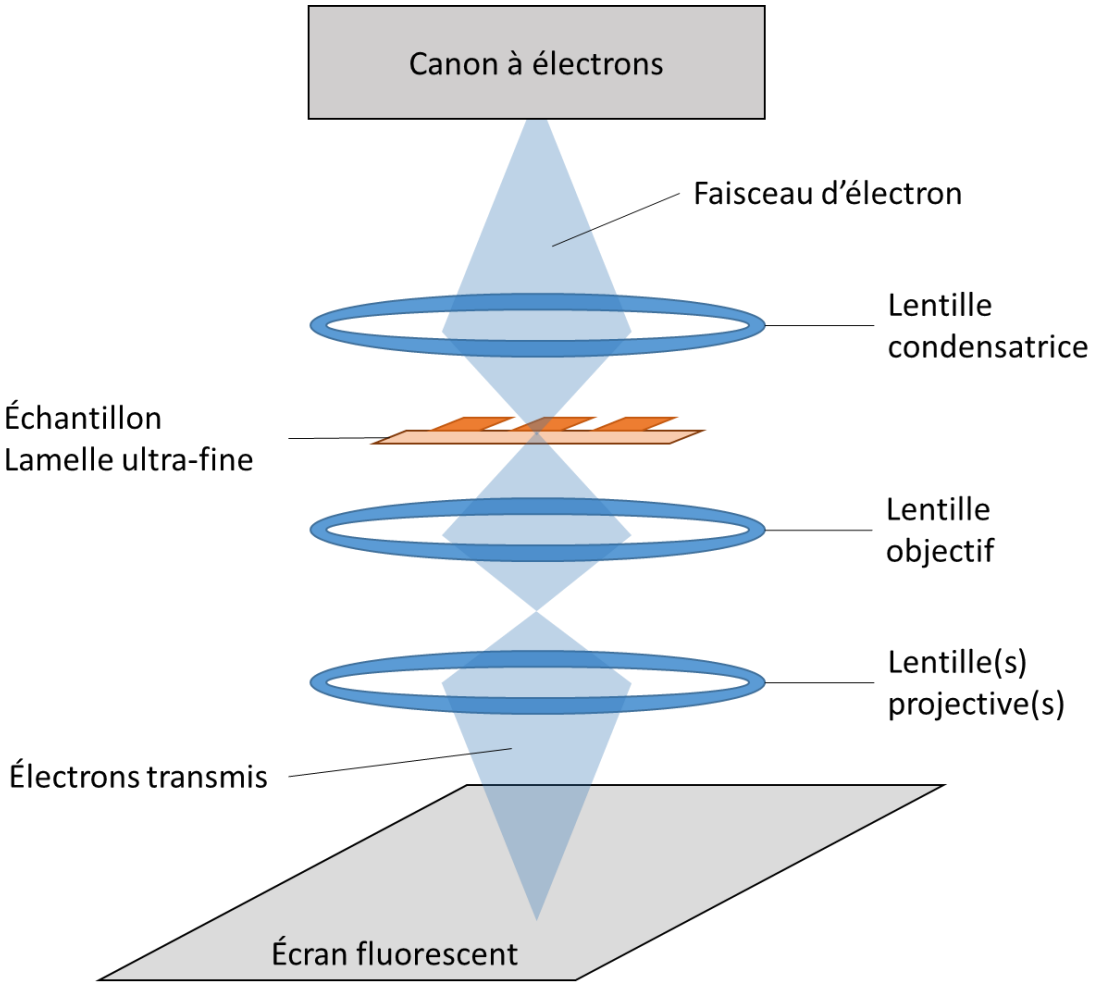

Figure I.21 : Schéma de principe de la technique TEM

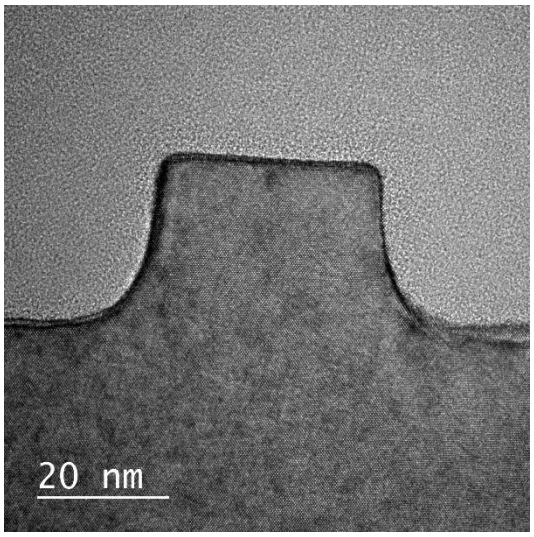

Figure I.22 : Mesure expérimentale d'une ligne de silicium réalisé par coupe TEM lors du projet MadeIn4.

I.3 La Métrologie Hybride

Comme nous l'avons vu, les différentes techniques de métrologie présentent chacune des avantages spécifiques, mais aussi des limitations qui peuvent affecter la justesse et la fidélité des mesures. La **métrologie hybride** vise à combiner plusieurs méthodes afin de maximiser leurs bénéfices tout en réduisant leurs contraintes. En fusionnant les données issues de diverses sources, elle permet d'obtenir des résultats plus justes et fidèle qu'une technique isolée. Cette approche est d'ailleurs identifiée comme un axe stratégique par l'IRDS, qui souligne que l'hybridation ou la combinaison métrologique aura un rôle plus important dans le futur (*« Hybrid or combined metrology will play a bigger role moving forward. »* [4]). Elle fait déjà l'objet de nombreux travaux de recherche [39], [40], [41], [42]

I.3.1 Les liens entre les données

En fonction des capteurs utilisés, de la nature des données et des objectifs visés, il est essentiel de comprendre la relation existante entre les informations issues de chaque technique. Ces relations peuvent être classées en trois catégories principales : **complémentaire**, **coopérative** et **redondante** [39] [43].

- **Lien complémentaire** : Les données sont obtenues par des capteurs ou des techniques mesurant le même sujet avec un point de vue ou des approches différentes, sans recouvrement. Par exemple, dans le cas de la voiture autonome, illustré par la Figure I.23.A, le radar et la caméra sont utilisés ensemble pour estimer le type d'obstacle dans un cas et sa distance dans l'autre. En métrologie, ce lien est souvent utilisé lorsque les techniques sont sensibles à des aspects ou conditions différentes.
- **Lien coopératif** : Les capteurs mesurent différents aspects d'un même phénomène afin d'estimer une information dérivée ou composite, inaccessible directement par un seul capteur pris isolément. Dans l'exemple de la voiture autonome, illustré par la Figure I.23.B, le radar mesure la distance à l'obstacle, tandis que le compteur de vitesse fournit la vélocité du véhicule. En combinant ces deux informations, il devient possible d'estimer une grandeur non directement mesurable, comme le temps avant impact. En métrologie, ce lien est utilisé lorsque l'exploitation conjointe de plusieurs techniques permet de déduire un paramètre complexe ou non observable directement.
- **Lien redondant** : Plusieurs capteurs mesurent exactement le même phénomène pour garantir la fiabilité et détecter d'éventuelles anomalies. Dans le cas de la voiture autonome, illustré par la Figure I.23.C, un radar et un lidar peuvent mesurer la même distance pour garantir l'absence de mesures erronées.

En métrologie, ce type de lien est utilisé pour réduire les erreurs en s'appuyant sur des données redondantes.

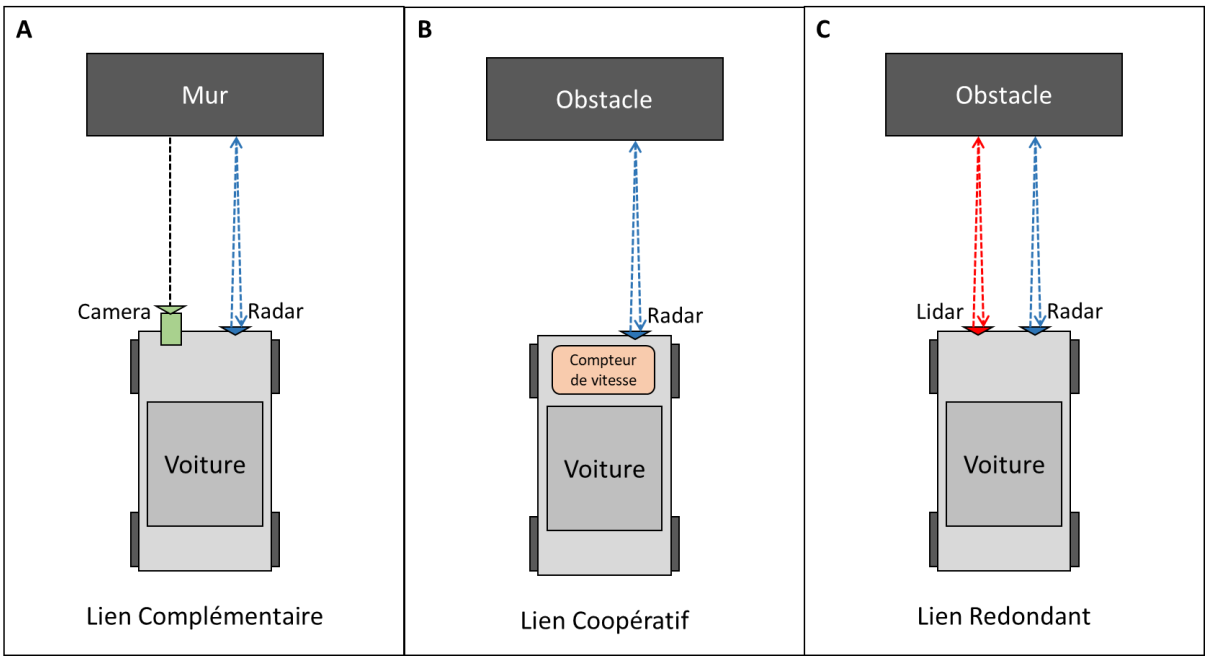

Figure I.23 : Illustration des liens entre les données en métrologie hybride (A) complémentaire, (B) coopératif et (C) redondant

Dans notre cas d'étude, les mesures dimensionnelles des lignes de silicium obtenues par **AFM, CDSEM, OCD et SAXS** relèvent d'un **lien redondant**. Chaque technique permet de mesurer les mêmes grandeurs (paramètres dimensionnels : **CD**, **H** et **SWA**) à l'aide de principes physiques différents. Cette redondance permet de croiser les résultats obtenus par différentes techniques, renforçant ainsi la robustesse de l'analyse et permettant une meilleure estimation des incertitudes associées aux paramètres mesurés.

I.3.2 Les architectures d'hybridation

Pour tirer pleinement parti de la complémentarité des données issues de différentes techniques de mesure, plusieurs types d'architectures d'hybridation peuvent être envisagés. Selon la manière dont les données sont collectées, traitées et fusionnées, on distingue quatre grandes architectures principales : **centralisée**, **distribuée**, **hiérarchique** et **décentralisée** [43], chacune offrant des avantages et des inconvénients selon les objectifs de la métrologie hybride.

- **Architecture centralisée** (Figure I.24) : Dans cette architecture, toutes les données issues des capteurs convergent vers un unique centre d'hybridation. Ce centre centralise le traitement global, ce qui peut offrir une vue d'ensemble

complète des données. Cependant, cette méthode nécessite de transférer et de traiter un grand volume de données simultanément, ce qui peut entraîner une surcharge de calcul et des délais si les données sont volumineuses.

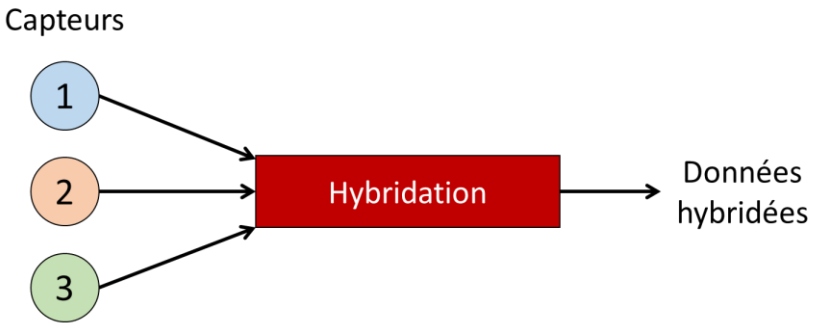

Figure I.24 : Architecture de type centralisée, d'après J.RECHE [10]

- **Architecture distribuée** (Figure I.25) : Ici, les données sont d'abord post-traitées localement par des sous-systèmes ou des capteurs individuels avant de converger vers un centre unique d'hybridation. Ce prétraitement permet de réduire la charge de travail du centre principal en filtrant les données ou en effectuant des analyses préliminaires. Toutefois, cette approche présente un risque de perte d'information, car le prétraitement local limite la vision globale des phénomènes mesurés à ce stade.

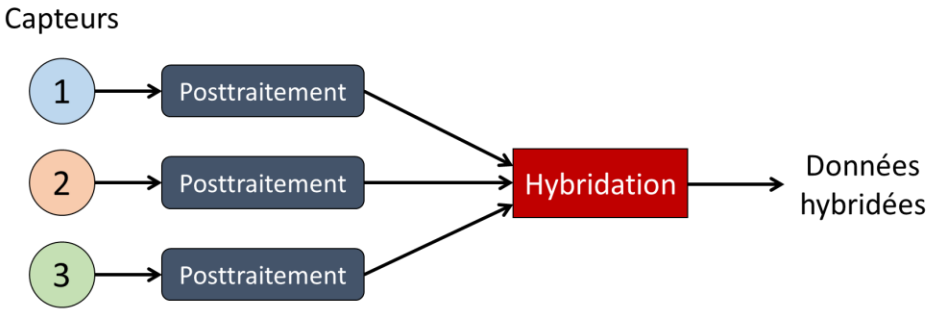

Figure I.25 : Architecture de type distribuée, d'après J.RECHE [10]

- **Architecture hiérarchique** (Figure I.26) : Les données sont fusionnées de manière progressive selon une structure en plusieurs niveaux. Par exemple, les données des capteurs 1 et 2 sont d'abord hybridées localement, puis les résultats intermédiaires sont fusionnés avec les données du capteur 3 pour une hybridation finale. Cette approche permet de réduire la charge de travail au niveau de l'hybridation globale tout en suivant l'évolution des données au fil des étapes, ce qui peut être bénéfique pour affiner les résultats.

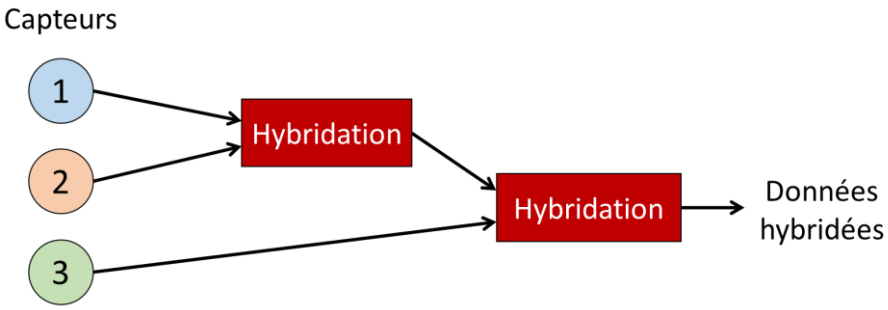

Figure I.26 : Architecture de type hiérarchique, d'après J.RECHE [10]

- **Architecture décentralisée** (Figure I.27) : Dans ce cas, l'hybridation des données s'effectue en plusieurs points répartis dans un réseau. Chaque point d'hybridation effectue son propre traitement des données, et les résultats sont disponibles indépendamment sans qu'un centre unique ne gère l'ensemble du processus. Bien que cette approche permette une plus grande résilience et une flexibilité dans le traitement, elle est complexe à mettre en place en raison des nombreuses voies de communication nécessaires et des traitements simultanés à gérer.

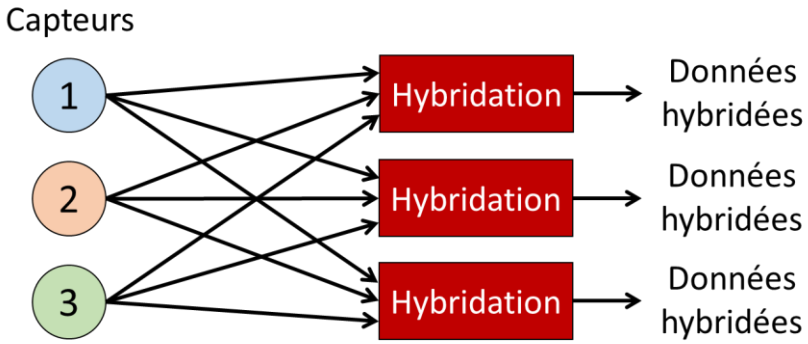

Figure I.27 : Architecture de type décentralisée, d'après J.RECHE [10]

L'hybridation des données en métrologie hybride vise à combiner les informations issues de plusieurs techniques de mesure afin de tirer parti de leurs complémentarités. Les méthodes bayésiennes, par exemple, s'appuient sur la théorie des probabilités pour fusionner les données de manière rigoureuse [39], [41], [40]. Elles nécessitent toutefois une modélisation explicite des incertitudes et une bonne connaissance des phénomènes physiques en jeu.

Une autre voie repose sur l'apprentissage automatique, et notamment sur les réseaux de neurones artificiels, qui offrent une alternative flexible permettant d'apprendre directement les relations entre variables à partir d'exemples. Cependant, cette approche requiert la disponibilité de jeux de données suffisamment grands, représentatifs et de bonne qualité, afin de capturer correctement les variabilités expérimentales et les interactions entre paramètres.

Dans le cadre de ces travaux, nous avons retenu cette seconde voie. Les principes et architectures de réseaux de neurones utilisés seront détaillés dans le Chapitre 3.

I.4 Conclusions

Ce chapitre a introduit les concepts essentiels à la compréhension des travaux développés dans cette thèse. Nous avons présenté les différentes techniques métrologiques utilisées, à savoir l'**AFM**, le **CD-SEM**, l'**OCD**, le **SAXS**, en mettant en avant leurs complémentarités et leurs limitations dans le cadre de la métrologie dimensionnelle. L'IRDS souligne qu'aucun instrument ne possède la capacité ou la résolution nécessaires pour mesurer le grand nombre de paramètres et de grandeurs requis pour caractériser certaines des nouvelles structures (*« No single instrument has the capability or resolution needed to measure the large number of parameters and measurands needed to characterize some of the new device structures »* [4]). Cette observation met en évidence la nécessité de mettre en place des approches de métrologie hybride afin de répondre à cette problématique.

Pour optimiser cette approche et exploiter pleinement les données issues de différentes sources métrologiques, nous avons pris le parti d'utiliser l'intelligence artificielle, ce qui permettra d'exploiter ces informations de manière efficace et robuste. L'intégration de réseau de neurones dans la métrologie hybride constitue ainsi un axe central de cette thèse.

Le prochain chapitre portera sur l'**acquisition des données** nécessaires à l'apprentissage des modèles de machine learning, une étape clé pour assurer la performance des algorithmes développés et leur capacité à améliorer l'exactitude des mesures métrologiques.

Chapitre II - Simulations des techniques de métrologie

II.1 Introduction

Dans le chapitre précédent, nous avons évoqué que l'hybridation des données issues des techniques de métrologie peut s'effectuer par le biais d'un réseau de neurones, méthode que nous utiliserons mais dont la performance dépend de la phase d'apprentissage et des données fournies. Cela nécessite un volume important de données diversifiées afin que le réseau puisse « apprendre » et généraliser efficacement le problème posé. Par des travaux préliminaires [44], nous avons déterminé qu'il fallait un minimum de 700 échantillons distincts pour permettre un bon apprentissage, ne dégradant pas les prédictions effectuées le réseau. Or, la création d'échantillons et leur caractérisation par diverses techniques de métrologie engendrent un coût financier et en temps trop important et limite le nombre de données nécessaires à l'apprentissage. Par conséquent, la solution que nous avons développée est la génération massive de données réalistes par l'utilisation de simulateurs.

Les simulateurs de techniques de métrologie ont pour objectifs de générer un volume important de données à faible coût, en temps et en ressources. Cela s'aligne avec notre objectif de génération massive de données réalistes, tout en pouvant accéder à des cas difficiles à créer expérimentalement. Cependant, leurs utilisations apportent aussi certains défis : les données produites par les simulateurs sont dérivées de modèles mathématiques et sont donc au départ exemptes d'erreurs de mesure propre à chaque technique. Dans ce chapitre, nous détaillerons les méthodes employées pour implémenter des incertitudes représentatives des phénomènes observés expérimentalement. Pour ce faire, nous nous sommes appuyés sur des travaux antérieurs [45],[46] portant sur l'estimation des incertitudes et leurs sources. De plus, des données expérimentales collectées, dans le cadre du projet européen « MadeIn4 », par mesures AFM, CDSEM, OCD, SAXS et TEM pourront être utilisées pour s'assurer de la qualité des simulations effectuées.

L'objectif principal de ce chapitre est donc de générer un jeu de données réalistes à partir de simulations, en intégrant les sources d'erreurs représentatives des conditions expérimentales. Ce jeu de données constitue une étape clé dans le cadre de cette thèse, dont l'objectif global est de développer une méthodologie d'hybridation des mesures par réseau de neurones.

II.2 Modélisation des échantillons

Avant de pouvoir recourir aux simulateurs de techniques métrologiques, il est essentiel de définir la géométrie de nos échantillons. Comme évoqué au chapitre précédent, ce projet de thèse se concentre sur l'étude et la mesure de réseaux périodiques de lignes en silicium, en mettant l'accent sur les paramètres dimensionnels que nous cherchons à mesurer : la dimension critique (Critical Dimension, CD), la hauteur (H) et l'angle de paroi latérale (Side Wall Angle, SWA). Comme le montre la Figure II.1, nous modélisons ces lignes tels des profils en deux dimensions (2D) en considérant leur longueur comme infinie.

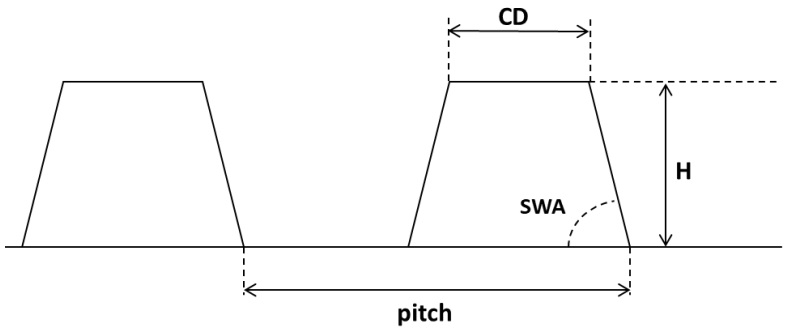

Figure II.1 : Schéma 2D simplifié d'un réseau de lignes vu en coupe avec la représentation des paramètres d'intérêts (CD, H et SWA) ainsi que la période du réseau (pitch)

II.2.1 Géométrie des échantillons

Initialement, nos échantillons étaient modélisés de manière simplifiée, en utilisant une seule forme trapézoïdale pour représenter les lignes du réseau. Cependant, afin d'obtenir des résultats plus représentatifs des structures réelles, nous avons adopté une modélisation plus détaillée. La Figure II.2.A nous présente une mesure de coupe TEM nous dévoilant la forme réelle de nos échantillons. Nous pouvons y observer des arrondis au niveau des angles (corners rounding). Ainsi, dans le but d'être plus représentatif de la réalité, nous avons opté pour la modélisation présentée dans la Figure II.2.B, où l'on voit une modélisation trapézoïdale ayant des corners rounding en haut et en bas de la ligne.

- Afin de ne pas surcharger le nombre de paramètres, il a été décidé de rester focalisé sur les paramètres dimensionnels à l'étude, quelques simplifications ont donc été appliquées : Le pitch des réseaux de lignes est défini comme uniforme et invariable, fixé à 113 nm.
- Les lignes sont supposées symétriques, avec un angle de paroi latérale (SWA) identique à gauche et à droite.

- Les Corners rounding (CR) implémentées sont de tailles fixes avec une distinction entre le CR du haut (CR top) et le CR du bas (CR bottom).

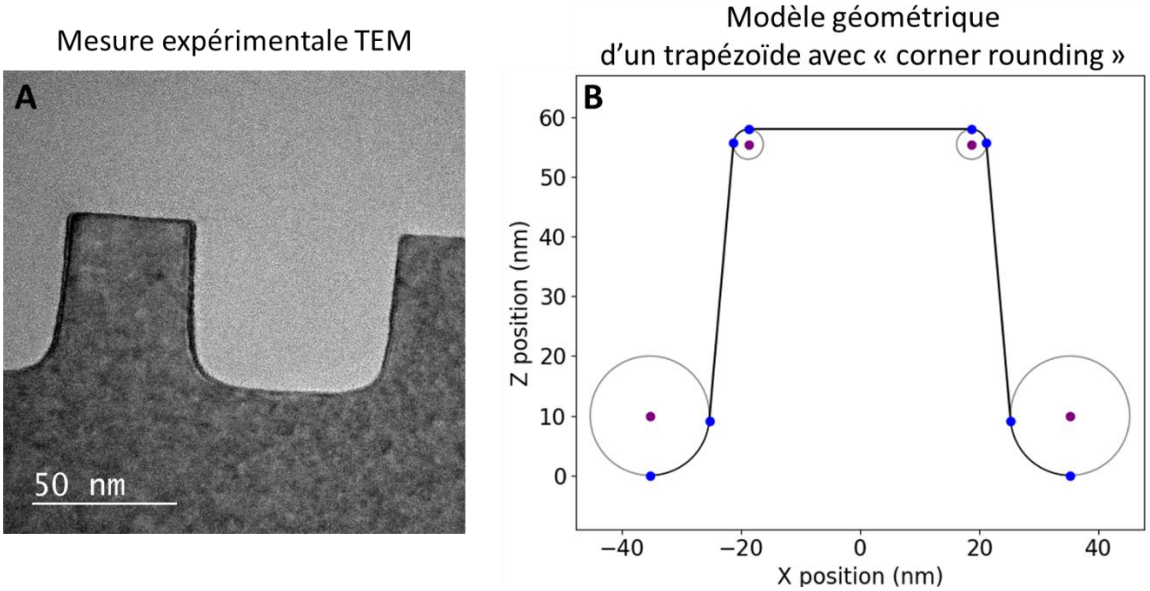

Figure II.2 : (A) Image TEM en coupe d'un échantillon de réseau de lignes. (B) Modèle géométrique trapézoïde avec « Corner Rounding » top et bottom fixés respectivement à 2.5 et 10 nm de rayon.

II.2.2 Valeurs des paramètres dimensionnels

Dans le but de pouvoir générer notre jeu de données, il nous faut choisir les plages de variation des valeurs des paramètres dimensionnels (CD, H et SWA). Afin d'avoir un ensemble de données expérimentales pour la validation, nous avons choisi des plages de valeurs incluant les dimensionnels des échantillons du projet européen MadeIn4, qui précède ces travaux de thèse. La génération de chaque paramètre dimensionnel s'effectue dans les gammes :

$$\mathbf{30\ nm \leq CD \leq 65\ nm}$$
$$\mathbf{20\ nm \leq H \leq 100\ nm}$$
$$\mathbf{80\ ° \leq SWA \leq 90\ °}$$

De plus, les exactitudes choisies pour les valeurs sont de 10^{-2} nm pour la hauteur (H) et la dimension critique (CD) et 10^{-2} degré pour le Side Wall Angle (SWA). Ce choix a été fait pour des raisons physiques liées à la taille de la maille du silicium.

Afin de garantir un échantillonnage équilibré, nécessaire à l'apprentissage de manière non ciblée du réseau de neurones, les valeurs des paramètres CD, H et SWA sont déterminées à l'aide d'une fonction aléatoire uniforme. Nous pouvons observer des répartitions homogènes sur toutes les plages de valeurs, que cela soit entre CD et

H (Figure II.3.A), CD et SWA (Figure II.3.B) et pour SWA et H (Figure II.3.C). Cela permet d'affirmer de l'absence de corrélation par l'observation des relations croisées entre les paramètres.

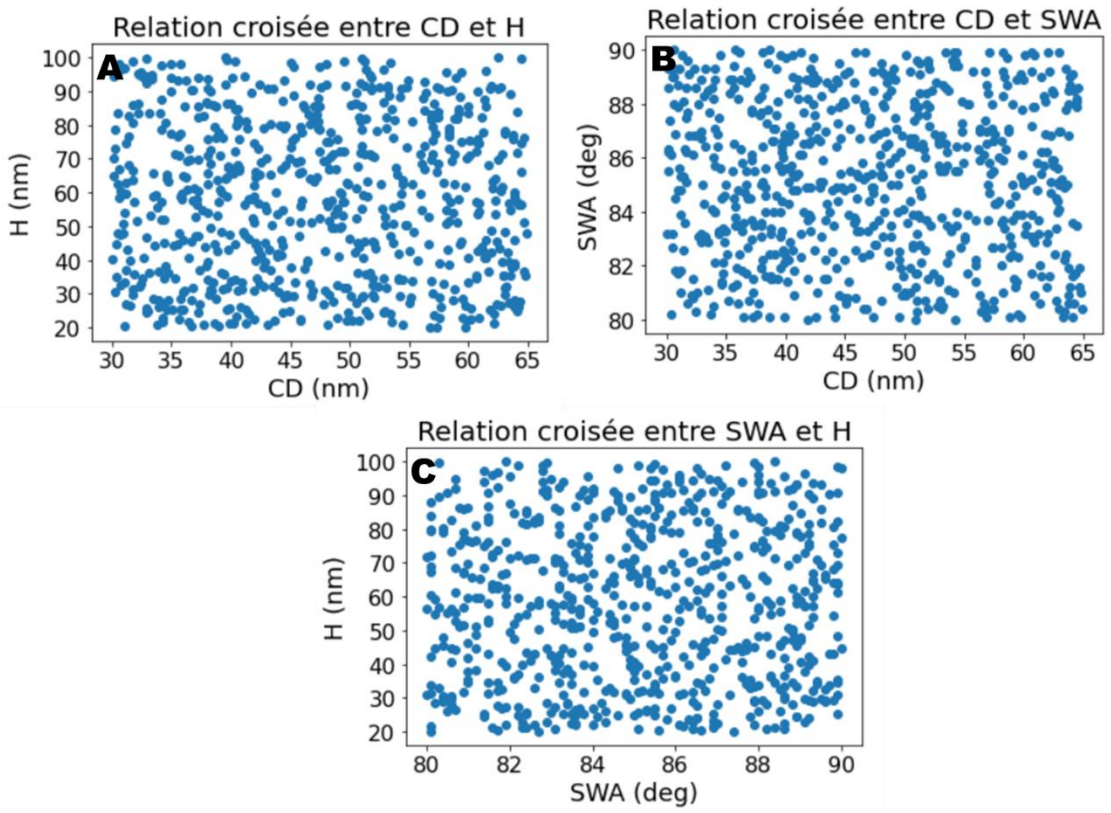

Figure II.3 : Relations croisées des paramètres dimensionnels : (A) entre CD et H ; (B) entre CD et SWA ; (C) entre SWA et H, pour les 700 paramétrisations générées.

II.2.3 Variabilité des échantillons

Par l'observation de mesures AFM, nous avons pu observer la présence de variations dimensionnelles entre les lignes d'un même réseau, comme nous pouvons l'observer sur la Figure II.4.

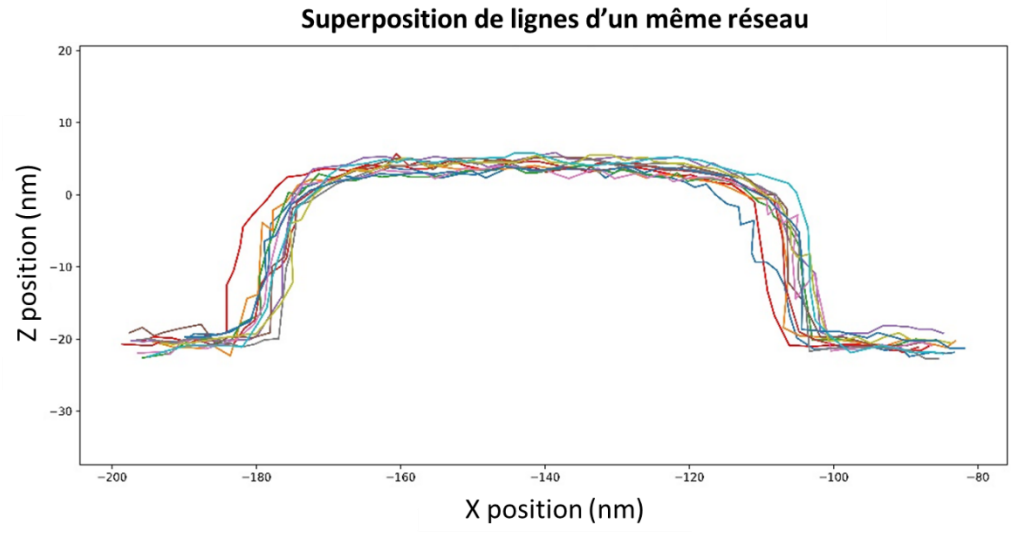

Figure II.4 : Superposition de mesures de lignes de 65 nm par AFM, où chaque couleur correspond à une ligne différente du même réseau

Cette variabilité est l'une des sources d'incertitudes dans les mesures des paramètres dimensionnels. Il est donc important de les quantifier afin de les modéliser. Ainsi, à partir d'ensembles de mesures AFM dont les dimensions correspondent aux plages de valeurs étudiées, nous avons calculé les écarts-types des paramètres dimensionnels pour plusieurs populations. Pour chaque paramètre, les valeurs finales présentées dans le Tableau II.1 correspondent aux moyennes des écarts-types obtenus pour l'ensemble des populations analysées.

Paramètre dimensionnel	**CD**	**H**	**SWA**
Ecart-type	1,90 nm	0,99 nm	3,8 °

Tableau II.1 : Ecart-types moyens des paramètres dimensionnels observées sur 68 lignes de 8 réseaux de lignes différents

Sur la base de ces observations, nous avons envisagé une approche de modélisation dite « multi-lignes », inspirée des recherches sur la modélisation de la variabilité entre lignes pour des simulations OCD [47]. Cette méthode permet l'implémentation des variations observées directement dans les modèles géométriques impactant alors les données brutes obtenues après simulation. Enfin, Le choix des valeurs des paramètres dimensionnels de chaque ligne k de cette modélisation est effectué par une fonction aléatoire gaussienne, comme décrite par les trois équations suivantes :

$$CD_k = Aléatoire_{normale}(\mu = CD, \sigma = 1.9\ nm) \quad \textbf{(II.1)}$$

CD correspond à la valeur de l'échantillon équilibré
CD_k correspond à la valeur du CD de la ligne k
$Aléatoire_{normale}$ est une fonction aléatoire gausienne ayant comme paramètres :

- μ : la valeur moyenne
- σ : l'écart-type

$$H_k = Aléatoire_{normale}(\mu = H, \sigma = 0.99\ nm) \quad \textbf{(II.2)}$$

H correspond à la valeur de l'échantillon équilibré
H_k correspond à la valeur de la hauteur de la ligne k

$$SWA_k = Aléatoire_{normale}(\mu = SWA, \sigma = 3.8\ °) \quad \textbf{(II.3)}$$

SWA correspond à la valeur de l'échantillon équilibré
SWA_k correspond à la valeur du side wall angle de la ligne k

Cette nouvelle stratégie a été appliquée à tous les simulateurs de techniques métrologiques. Plusieurs études ont été effectuées afin de déterminer le nombre total de ligne que l'on doit modéliser. Il est important de prendre en compte certaines limitations techniques liées aux simulateurs. En effet, plus il y a de lignes dans le modèle multi-lignes plus les calculs sont nombreux et couteux en temps de calculs, poussant certains simulateurs dans leurs limites d'utilisation. Ainsi, après une multitude de tests, nous avons décidé d'adopter une modélisation avec 3 lignes. Ce choix permet d'avoir à la fois une certaine diversité dans les variabilités avec un faible nombre de lignes.

Afin de calculer les erreurs de mesure pour chacune des techniques métrologiques, il nous faut créer des valeurs de références. Celles-ci sont calculées en moyennant les valeurs des paramètres dimensionnels des 3 lignes du modèle, comme décrit dans le cas du CD:

$$CD_{ref} = moyenne(CD_1, CD_2, CD_3) \quad \textbf{(II.4)}$$

L'étude de la modélisation des échantillons nous a conduit à adopter un modèle à trois lignes, dont la variabilité des paramètres dimensionnels (CD, H et SWA) a été quantifiée à l'aide de mesures AFM. Ce modèle vise à représenter de manière réaliste les phénomènes observés tout en intégrant certaines simplifications pour se concentrer sur l'analyse de ces paramètres. Afin d'assurer la cohérence des simulations, cette modélisation sera appliquée à chaque simulateur de technique métrologique.

II.3 Simulations AFM

La génération massive de données repose sur l'utilisation d'outils de simulation. Bien que plusieurs simulateurs soient disponibles pour les différentes techniques métrologiques, nous n'avons pas de simulateur AFM à disposition au sein du LTM et du CEA-Leti. Il a donc été nécessaire d'avoir une compréhension suffisante des principes de mesure AFM afin d'en concevoir un simulateur dédié, indispensable pour la génération de données.

II.3.1 Conception du simulateur

Nous avons pu voir le principe de fonctionnement d'un AFM dans le chapitre précédent. Ainsi, pour concevoir le simulateur, il nous faut dans un premier temps modéliser l'impact de la pointe sur les échantillons. Les données brutes générées ne peuvent pas être traitées avec les outils des équipementiers à cause de leur format encrypté. Par conséquent, nous avons fait le choix de développer un outil d'analyse AFM afin de les traiter. Nous avons alors 2 parties dans le développement de l'outil : une partie simulation de données brutes et une partie extraction des paramètres dimensionnels à partir des profils simulés. Ainsi nous pourrons générer massivement des données en modifiant le modèle géométrique. Deux approches ont été étudiées, la première approche se concentre sur la correction analytique des métriques à l'aide des caractéristiques géométriques de la pointe lors de la mesure de l'échantillon. La seconde approche est une approche par corrélation 2D permettant de savoir les positions possibles de la pointe sur l'échantillon.

II.3.1.1 Approche géométrique

Conception du simulateur

Cette approche consiste à modifier les métriques (CD, H, SWA) mesurés par un AFM à l'aide de relations paramétriques prenant en compte l'encombrement approché de la pointe AFM, usuellement modélisé par une convolution.

L'équipement utilisé au sein du CEA-Leti permet l'acquisition de la forme géométrique de la pointe lors de sa calibration à partir d'étalon. Ainsi nous pouvons obtenons un ensemble de points pour définir sa géométrie, comme illustré par la Figure II.5.A. De plus, nous pouvons appliquer une simplification géométrique permettant de définir la pointe à partir de 2 arcs de cercles (définis par les rayons R_{gauche} et R_{droit}) et de sa largeur L. La Figure II.5.B permet d'illustrer l'interaction entre la pointe et l'échantillon au niveau géométrique. Ainsi, le profil brut obtenu lors de

mesures AFM correspond aux pointillés, ce qui diffère de la forme réelle de l'échantillon.

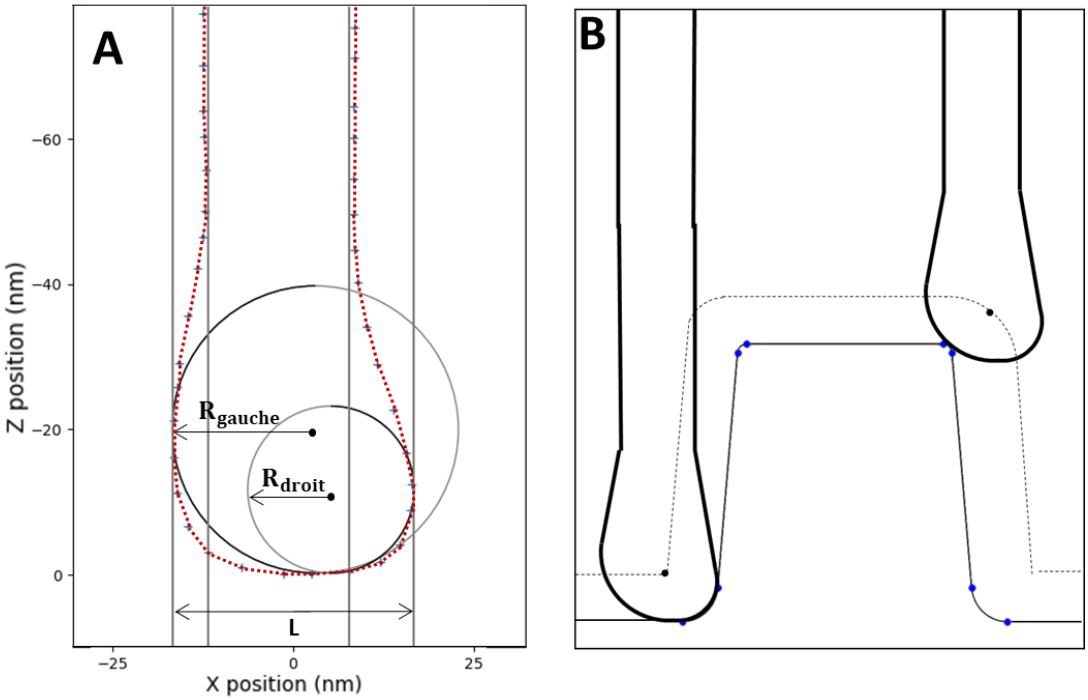

Figure II.5 : (A) Forme de la pointe acquise par l'équipement AFM (en pointillés) et estimation de sa forme par 2 arcs de cercle de rayons R_{gauche} et R_{droit}. (B) Influence géométrique de la pointe sur un échantillon lors d'une mesure. (--) correspond au profile mesuré par AFM, convolué par la pointe. (-) Correspond à la forme réelle de l'échantillon.

L'approche géométrique consiste à modifier analytiquement les métriques mesurées par AFM en fonction des paramètres dimensionnels de la pointe. Ainsi en prenant en compte les interactions géométriques et simplifications de la pointe, nous en avons déduit les équations suivantes :

$$CD_{AFM} = CD + L \quad \textit{(II.5)}$$

CD_{AFM} : Valeur de CD obtenue par mesure AFM
CD : Valeur réelle du CD
L : largeur de la pointe

$$H_{AFM} = H \quad \textit{(II.6)}$$

H_{AFM} : Valeur de H obtenue par mesure AFM
H : Valeur réelle de H

$$SWA_{AFM} = SWA \quad \textit{(II.7)}$$

SWA_{AFM} : Valeur du SWA obtenue par mesure AFM
SWA : Valeur réelle du SWA

$$CR^{AFM}_{haut,gauche} = CR_{haut} + R_{droit} \quad \textbf{(II.8)}$$

$CR^{AFM}_{haut,gauche}$: Rayon du corner rounding du haut gauche par mesure AFM
CR_{haut} : Valeur réelle du rayon du corner rounding haut
R_{droit} : Valeur du rayon de l'arc de cercle droit de la pointe

$$CR^{AFM}_{haut,droit} = CR_{haut} + R_{gauche} \quad \textbf{(II.9)}$$

$CR^{AFM}_{haut,droit}$: Rayon du corner rounding du haut droit par mesure AFM
CR_{haut} : Valeur réelle du rayon du corner rounding haut
R_{gauche} : Valeur du rayon de l'arc de cercle gauche de la pointe

$$CR^{AFM}_{bas,gauche} = \begin{cases} si\ CR_{bas} \leq R_{droit}\ alors\ 0 \\ sinon\ CR_{bas} - R_{droit} \end{cases} \quad \textbf{(II.10)}$$

$CR^{AFM}_{bas,gauche}$: Rayon du corner rounding du bas gauche par mesure AFM
CR_{bas} : Valeur réelle du rayon du corner rounding bas
R_{droit} : Valeur du rayon de l'arc de cercle droit de la pointe

$$CR^{AFM}_{bas,droit} = \begin{cases} si\ CR_{bas} \leq R_{gauche}\ alors\ 0 \\ sinon\ CR_{bas} - R_{gauche} \end{cases} \quad \textbf{(II.11)}$$

$CR^{AFM}_{bas,droit}$: Rayon du corner rounding du bas droit par mesure AFM
CR_{bas} : Valeur réelle du rayon du corner rounding bas
R_{gauche} : Valeur du rayon de l'arc de cercle gauche de la pointe

Ainsi, en ayant un modèle géométrique des échantillons définissant les paramètres des corner rounding, du CD et de la hauteur H, comme illustré sur la Figure II.6.A, nous pouvons modéliser l'influence de la pointe sur échantillon, nous permettant de simuler la mesure d'échantillon par AFM, illustré par la Figure II.6.B.

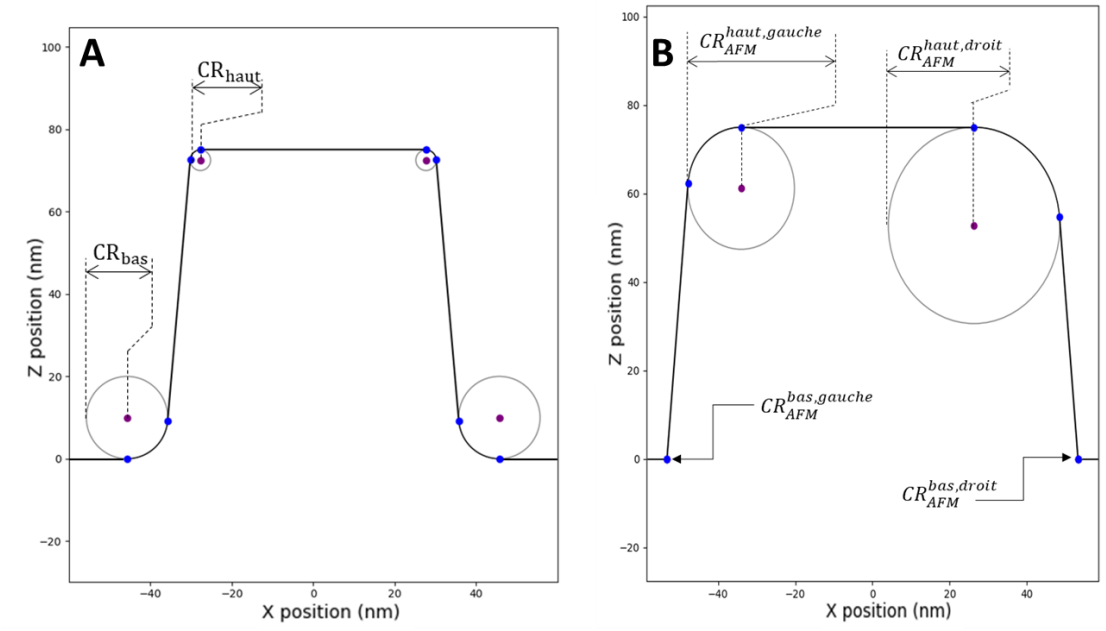

Figure II.6 : (A) Modèle géométrique d'un échantillon avec corners rounding. (B) Modèle géométrique d'un échantillon mesuré par AFM.

Cette approche analytique nous permet ainsi de générer des données brutes de mesures AFM à partir des informations géométriques de la pointe et d'échantillons. Il nous faut donc concevoir les outils permettant d'extraire les paramètres dimensionnels de l'échantillon de façon automatique à partir des profils ainsi générés.

Conception de l'extracteur des paramètres dimensionnels à partir d'un profil

Nous avons vu qu'à partir des équations précédentes, il est possible de directement calculer la structure géométrique des échantillons afin de simuler la mesure par AFM. Ainsi, en réalisant l'opération inverse, il est possible de retrouver les paramètres d'origines de l'échantillon. Il est important de comprendre qu'il y aura une implémentation de bruit et rugosité de surface inhérent à l'échantillon ainsi qu'à l'AFM. Ceux-ci ont été observés expérimentalement (Figure II.4) et quantifiés (Tableau II.1). La présence de ces bruits permettra d'introduire des erreurs métrologiques afin de générer des profils réalistes à l'aide du simulateur.

L'extracteur des paramètres dimensionnels se caractérise par l'utilisation d'outils de reconnaissance de forme, tel que OpenCV [48]. Cette reconnaissance de forme permet l'identification des zones correspondant aux corners rounding, des surfaces, des creux et des parois des lignes, en fonction de leur forme et position spatiale afin de déterminer les paramètres des dimensionnels de l'échantillon mesuré (ou simulé) par AFM.

Dans un premier temps, cette méthode est appliquée à des cas simples. Nous pouvons observer sur la Figure II.7.A la segmentation des éléments en fonction de leur forme, en vert les éléments rectilignes et en rouge les éléments qui ne le sont pas. La Figure II.7.B montre comment identifier le centre d'un arc de cercle (élément non rectiligne) à partir des droites normales. Ainsi, il est possible de retrouver les valeurs des rayons des corners rounding, comme l'illustre la Figure II.7.C

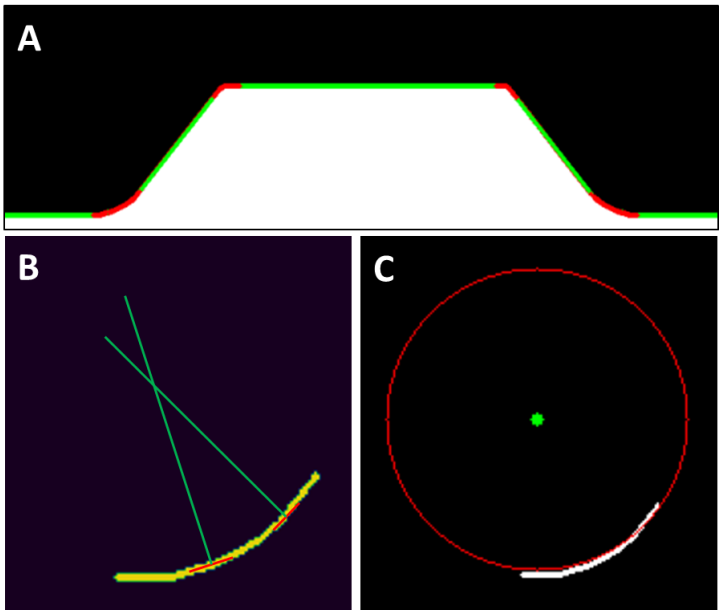

Figure II.7 : (A) Les différents éléments identifiés sont : en vert correspond à des éléments plats (surfaces, creux et parois) et en rouge à des éléments courbés (corners rounding). (B) Le centre et le rayon d'un arc de cercle (élément courbé) sont identifiés par l'utilisation de ses droites normales. (C) Visualisation du cercle trouvé associé à l'arc de cercle permettant le calcul du rayon CR (B).

Cette approche s'avère efficace pour ces cas idéaux, sans bruit et rugosité de surface. Cependant, comme expliqué précédemment, les mesures AFM de lignes de silicium présentent une rugosité de surface. Cette rugosité est cruciale à modéliser, non seulement pour représenter fidèlement la réalité, mais aussi parce qu'elle génère des incertitudes importantes. En effet, si la rugosité de surface n'est pas modélisée, nos outils appliqueront uniquement la correction de l'influence de la pointe sur la géométrie de l'échantillon, sans tenir compte des erreurs supplémentaires engendrées par la rugosité elle-même. Il est alors indispensable de l'implémenter après la simulation de la mesure mais avant l'extraction des paramètres dimensionnels.

Afin d'évaluer la robustesse de notre algorithme face aux perturbations, nous avons introduit un bruit plus faible que celui observé sur les données expérimentales. Dans la Figure II.8.A, on observe une sensibilité excessive au bruit, ce qui complique l'identification des segments rectilignes et courbes en raison de la prise en compte de fluctuations indésirables. À l'inverse, la Figure II.8.B illustre une sensibilité trop faible au bruit, rendant l'identification des éléments du profil incertaine et peu fiable.

Ainsi, notre approche basée sur la détection de contours se révèle particulièrement sensible aux perturbations, entraînant des défaillances dans l'identification des différents éléments du profil. En effet, certaines zones non rectilignes sont détectées de manière fragmentée et discontinue, tandis que des régions correspondant à des segments rectilignes sont également mal interprétées.

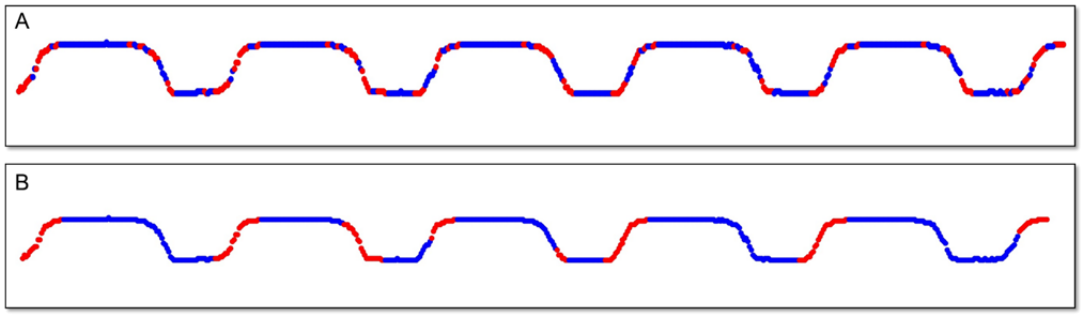

Figure II.8 : Illustration de l'impact de la sensibilité au bruit sur la détection des éléments du profil. (A) Une sensibilité excessive conduit à l'identification erronée des segments rectilignes et courbes en raison du bruit. (B) Une sensibilité insuffisante entraîne une détection incomplète et incertaine des éléments du profil.

De nombreux essais ont été menés afin d'optimiser le paramétrage du traitement automatique des contours issus d'images AFM bruitées. Toutefois, il apparaît que l'extraction des paramètres dimensionnels à partir de ces mesures reste sujette à des erreurs, notamment dans l'identification des différents éléments du profil. Ces imprécisions compromettent la fiabilité de l'analyse et limitent l'exploitabilité de cette approche dans le cadre de nos travaux. Ainsi, bien que cette méthode ait été explorée, elle ne se révèle pas suffisamment robuste pour être retenue comme solution viable.

II.3.1.2 Approche par corrélation croisée 2D

Conception du simulateur

L'outil mathématique utilisé dans cette approche est la corrélation croisée en deux dimensions (2D). Cet outil permet de calculer la similitude entre deux matrices. Nous transformons alors les points définissant la pointe et l'échantillon en deux images binaires représentées par des matrices de 0 et de 1, indiquant respectivement l'absence et la présence de matière. Ainsi, il est possible d'obtenir la similitude entre ces images. La Figure II.9 permet d'illustrer cette opération. Nous pouvons voir à gauche de la figure une représentation très simplifiée d'une ligne (matrice 6x6) où des pixels symétriques ont été ajoutés aux bords afin de conserver la même taille (6x6) en fin d'opération. De même, une matrice 3x3 représente une pointe utilisée comme noyau de filtrage (kernel). L'image résultante, situé à droit, montre le profil obtenu après l'application de la corrélation croisée entre ces deux matrices, permettant ainsi d'extraire les caractéristiques du contour traité.

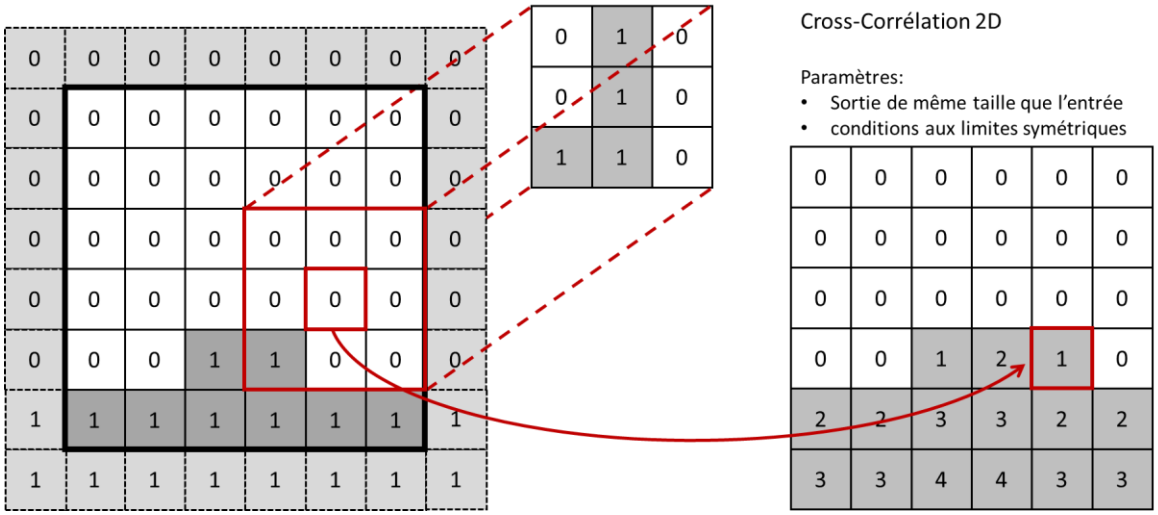

Figure II.9 : Principe de fonctionnement de la cross-corrélation 2D [49] appliqué à des modélisations simplifiées de la pointe et de la ligne.

Cette approche présente l'avantage de ne pas avoir besoin de connaitre les équations modélisant l'influence de la pointe sur l'échantillon. Ainsi, son application à une large variété de types d'échantillons est possible. Toutefois, cette méthode nécessite une puissance de calcul importante, particulièrement lorsque des résolutions élevées d'images sont requises. La cross-corrélation nécessite environ 20 secondes pour une résolution de 1 nm/px (corrélation croisée de deux images de dimensions 35 x 80 px pour la pointe et 113 x 105 px pour le profil réel considéré) à l'aide de l'utilisation de CPU. Cependant, notre objectif étant d'atteindre une résolution de 0,1 nm/px, le nombre de calculs requis serait multiplié par 100. L'utilisation d'unités de calcul plus spécialisées dans le calcul matriciel est alors requise afin d'avoir des temps de calcul raisonnables. Nous pouvons voir sur le Tableau II.2, que l'utilisation de carte graphique (GPU) permet une diminution significative du temps de calcul. Ainsi nous pouvons nous permettre d'avoir une résolution de 0.1 nm/px.

Résolution de 1nm/pixel		**Résolution de 0,1nm/pixel**
Cross-corrélation **CPU**	Cross-corrélation **GPU**	Cross-corrélation **GPU**
≈ 20 secondes	< 1 seconde	≈ 3 secondes

Tableau II.2 : Temps de calculs en fonction de la résolution et de l'unité de calcul utilisé.

Nous disposons ainsi d'une méthode plus flexible pour modéliser l'interaction pointe-échantillon que l'approche géométrique, ne nécessitant que quelques secondes. Néanmoins, cette méthode requière une transformation d'une modélisation par points à une représentation par image. Cependant cette transformation est réversible.

Conception de l'extracteur des paramètres dimensionnels à partir d'une image

L'approche par corrélation croisée 2D utilisée dans la simulation des mesures AFM facilite la mise en œuvre de certains algorithmes destinés à l'extraction des paramètres dimensionnels. En s'inspirant de l'extraction de ces paramètres par d'autres techniques métrologiques que nous verrons plus loin, une méthode par résolution de problème inverse a été adoptée. Par ailleurs, à l'instar de l'approche géométrique, il est essentiel d'introduire la rugosité de surface entre la simulation des mesures et l'extraction des paramètres afin de mieux refléter les conditions expérimentales.

La Figure II.10 illustre le fonctionnement de l'algorithme de résolution de problème inverse. Tout d'abord, un échantillon mesuré par AFM (qu'il soit expérimental ou simulé) est requis. Un modèle théorique est ensuite généré à partir de paramètres initiaux choisis aléatoirement. Une soustraction est effectuée entre l'échantillon mesuré (bruité) et le modèle théorique (parfait). Le critère clé est le nombre de pixels non nuls résultant de cette soustraction, qui représente le score du modèle théorique. Un algorithme d'optimisation cherche à minimiser ce score en ajustant les paramètres dimensionnels pour produire un nouveau modèle théorique. En répétant ce processus un grand nombre de fois, les paramètres dimensionnels finissent par converger vers des valeurs stables. Ces valeurs de convergence correspondent alors aux paramètres dimensionnels déterminés par l'algorithme.

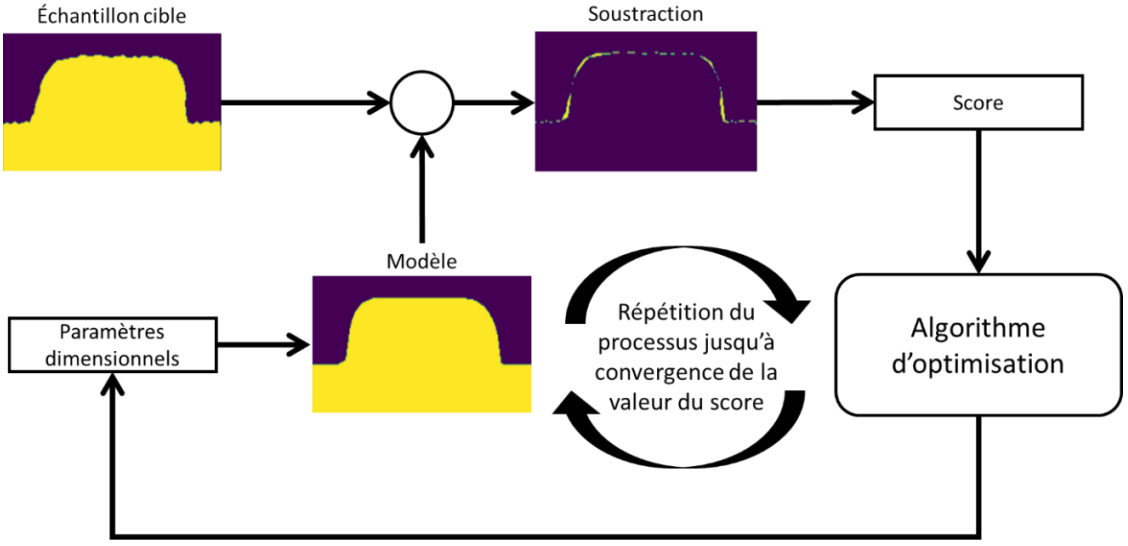

Figure II.10 : Schéma du principe de fonctionnement de la résolution de problème inverse

L'un des algorithme d'optimisation les plus couramment utilisés pour la résolution de problèmes inverses est la méthode de Monte Carlo [50]. Cependant, la présence de nombreux minima locaux pose problème à cette méthode. Pour remédier à cela, nous

avons mis en œuvre une méthode plus robuste, bien que plus coûteuse en temps de calcul : la méthode de Basin-Hopping [51], [52], [53]. Cette approche effectue plusieurs recherches de minimum dans l'espace des solutions grâce à des sauts aléatoires. Lorsqu'un minimum est identifié de manière répétée (paramètre à définir par l'opérateur), il est considéré comme le minimum global. Cela réduit ainsi le risque de trouver un minimum local. Cependant, cette méthode requiert un nombre de calculs significativement plus élevé, ce qui allonge les temps de traitement.

Les méthodes présentées opèrent dans un domaine continu. Or, nous travaillons dans un domaine discret, dont l'exactitude est limitée par la résolution des images. Cela entraîne des calculs redondants et non optimisés. Pour pallier cette limitation, nous avons intégré l'approche de « recherche tabou » [54], [55]. En mémorisant les résultats déjà obtenus, cette méthode permet d'éviter des calculs redondants. De plus, pour contourner les contraintes du domaine continu, nous avons arrondi les paramètres selon la résolution souhaitée et stocké les résultats associés. Grâce à cette optimisation, nous avons réduit le nombre de calculs nécessaires, diminuant ainsi le temps de résolution du problème inverse, qui est a été divisé par 4 pour finir proche des 5 minutes.

L'utilisation de la recherche tabou présente également des limites. En effet, au-delà d'un certain seuil, le temps nécessaire pour vérifier la liste des calculs précédemment effectués dépasse celui requis pour exécuter le calcul lui-même. Pour maintenir l'efficacité de cette approche, il est essentiel de limiter la taille de cette liste en la structurant sous forme de « pile » : les éléments les plus anciens sont supprimés afin de libérer de l'espace pour de nouvelles entrées, garantissant ainsi la pertinence et la rapidité de la recherche tabou.

Un très grand nombre de calculs est nécessaire pour permettre la génération des modèles théoriques. Ainsi, l'importance de la corrélation croisée 2D accéléré par GPU est primordiale. Après optimisation de la simulation de mesure AFM et de l'extraction des paramètres dimensionnels des échantillons simulés, les temps de calculs typiques sont de l'ordre d'une dizaine de minutes. Nous avons également la possibilité de produire de façon massive des données AFM synthétiques en utilisant cette approche (pour rappel, quelques secondes pour une image synthétique d'un profil).

II.3.2 Implémentation des incertitudes de mesure propre à l'AFM

Une étape importante entre simulation de mesure AFM et extraction des paramètres dimensionnels est l'implémentation de la rugosité de surface que l'on observe sur les données expérimentales, illustrée par la Figure II.11. En effet, l'algorithme compare

des images entre elles, il est alors sensible aux variations géométriques liées à la rugosité.

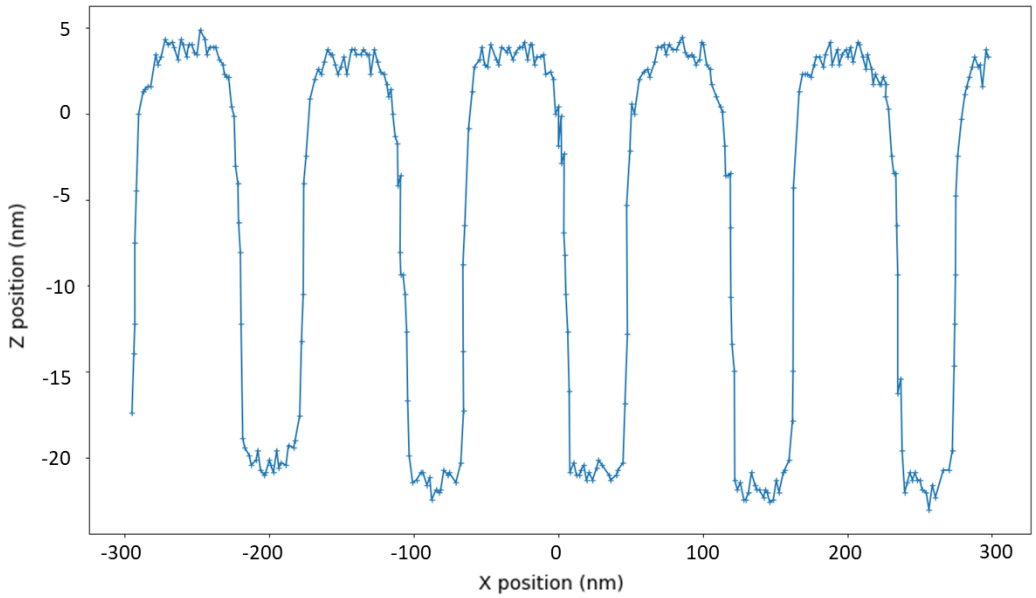

Figure II.11 : Mesure expérimentale d'un réseau de ligne de silicium par un AFM (pitch = 113 nm, H = 23 nm et CD = 72 nm avec l'influence de la pointe)

Nous avons estimé l'écart-type de la rugosité à 1,5 nm. Cette estimation a été réalisée en analysant les positions des points des surfaces et des creux des lignes obtenues expérimentalement. Pour chaque surface et chaque creux, nous avons d'abord calculé la valeur moyenne des positions des points. Ensuite, nous avons déterminé l'écart-type en mesurant la dispersion des valeurs par rapport à cette moyenne. Afin d'implémenter cette rugosité, nous avons opté pour une modélisation par bruit gaussien, en utilisant cette valeur estimée. Ainsi, après avoir simulé l'interaction pointe-échantillon par corrélation croisée, nous appliquons la rugosité de surface, avant de réaliser la résolution de problème inverse. Cette approche nous permet d'obtenir des résultats visuellement très proches des données expérimentales, comme illustré en Figure II.12.

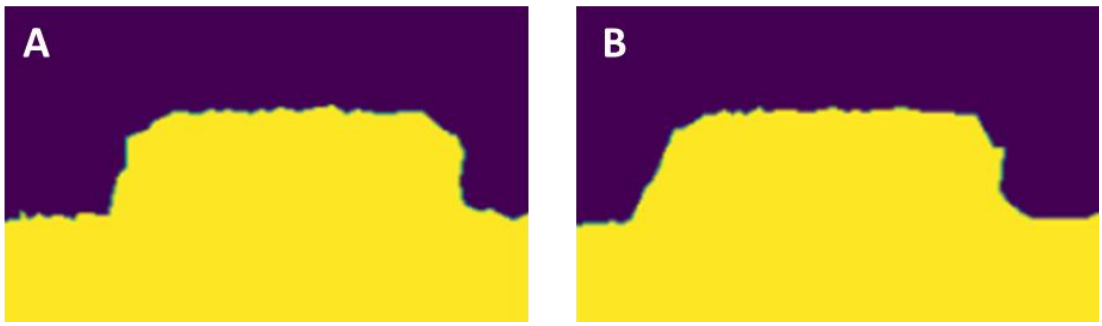

Figure II.12 : Image binaire d'échantillons simulé (A) et image binaire issue d'un profil AFM expérimental (B)

II.3.3 Analyse des paramètres dimensionnels extraits des simulations AFM

Toutes les étapes précédentes nous ont permis de générer notre jeu de données simulées. Pour introduire de la variabilité entre les lignes d'un même réseau, nous avons utilisé un modèle basé sur trois lignes, présenté dans la partie II.2, Cette approche a été appliquée à chaque technique.

Une autre source d'erreur aux simulations AFM est la rugosité. Nous avons en effet constaté sa présence dans nos données expérimentales, ce qui induit des erreurs lors de l'extraction des paramètres dimensionnels.

Ainsi, pour calculer les valeurs de CD, H et SWA des échantillons, nous moyennons les résultats obtenus pour chacune des trois lignes, puis nous déterminons les erreurs en les comparant aux valeurs de référence :

$$X_{AFM} = moyenne(x_1, x_2, x_3) \quad \textbf{(II.12)}$$

x_1, x_2 et x_3 sont les valeurs d'un paramètre dimensionnel extraites des lignes 1, 2 et 3

$$\varepsilon_{X,AFM} = X_{AFM} - X_{ref} \quad \textbf{(II.13)}$$

$\varepsilon_{X,AFM}$ est l'erreur du paramètre dimensionnel X (CD, H ou SWA) de simulation AFM

Après avoir simulé l'ensemble des 700 échantillons et extrait les paramètres dimensionnels, nous avons calculé les erreurs de mesures AFM simulées. Afin d'analyser la distribution des erreurs, nous avons effectué une comparaison avec une distribution normale à l'aide de diagrammes quantile-quantile (qq-plots), également connus sous le nom de droites de Henry [56]. Ainsi, en comparant notre distribution des erreurs à une distribution normal théorique, illustrée par la Figure II.13.A pour les erreurs sur le paramètre CD, nous observons un ensemble de points alignés le long d'une droite, avec un coefficient de détermination R^2 de 0,983. Cela signifie que la distribution obtenue s'apparente à une distribution normale. Nous pouvons confirmer cette analyse par l'observation de l'histogramme des erreurs sur CD, présenté par la Figure II.13.B. Nous y voyons notre distribution (en orange) et la distribution normale associée. Celle-ci est obtenue par l'utilisation d'une fonction d'estimation de Scipy [57]. Nous pouvons alors avoir une estimation des valeurs des paramètres σ et μ de la gaussienne associée aux erreurs obtenues.

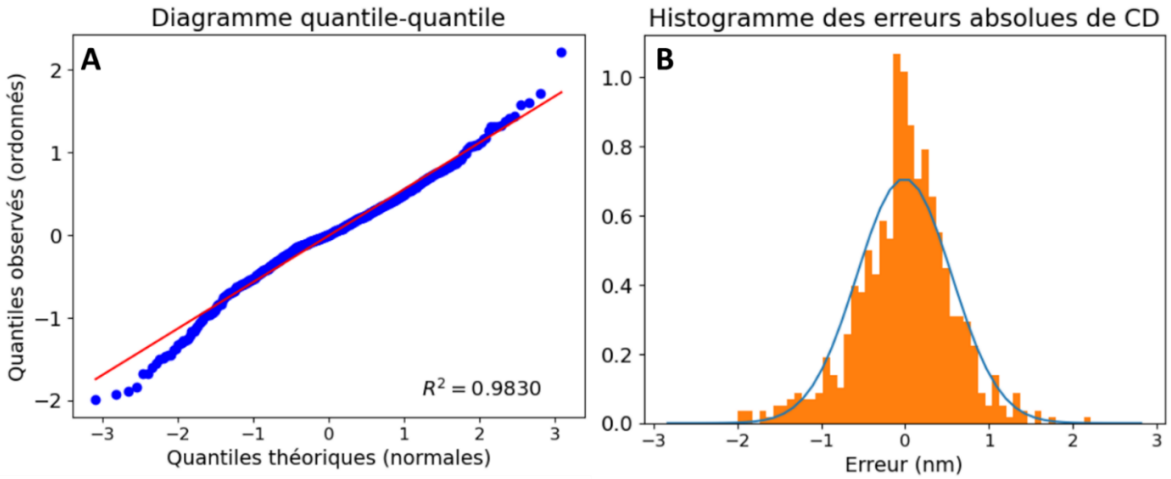

Figure II.13 : (A) Diagramme quantile-quantile des erreurs absolues de CD. (B) Histogramme des erreurs absolues de CD (orange) et distribution normale associée (bleu).

Nous avons effectué les mêmes analyses pour les paramètres H et SWA. Les diagrammes quantile-quantile ont également mis en évidence un comportement gaussien, avec un coefficient de détermination R^2 de 0,9964 pour H et 0,9855 pour SWA. Ainsi nous avons pu estimer les paramètres de ces distributions, comme l'illustrent la Figure II.14.A pour H et la Figure II.14.B pour SWA.

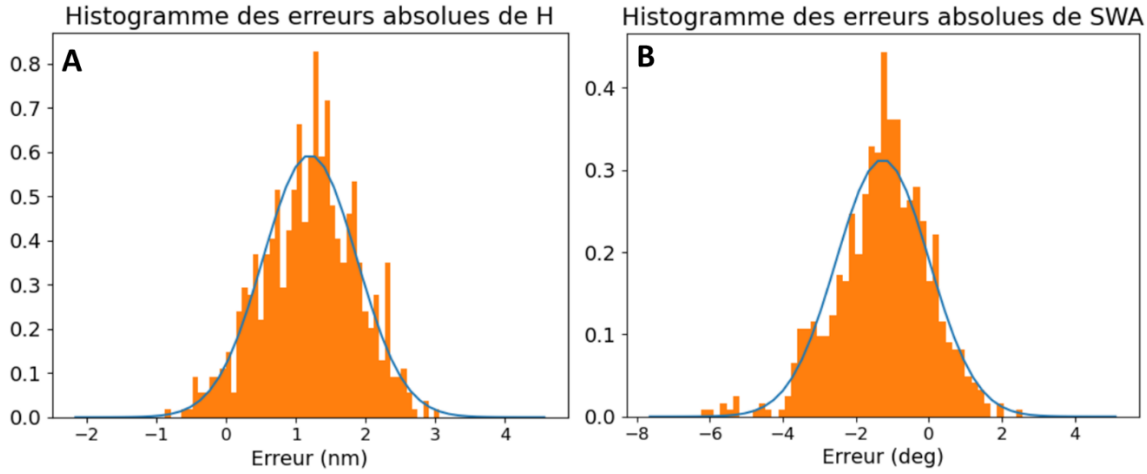

Figure II.14 : (A) Histogramme des erreurs absolues de H (orange) et distribution normale associée (bleu). (B) Histogramme des erreurs absolues de SWA (orange) et distribution normale associée (bleu).

Les incertitudes de mesure associées aux simulations AFM ont été évaluées pour les trois paramètres d'intérêt : CD, H et SWA. Les résultats, présentés dans le Tableau II.3, indiquent une incertitude de σ = 0,65 nm sur le CD, avec une erreur moyenne μ = -0,05 nm, ce qui suggère l'absence de biais significatif. Pour les paramètres H et SWA, les incertitudes sont respectivement de σ = 0,68 nm et σ = 1,29°, avec des erreurs moyennes de μ = 1,19 nm et μ = -1,29°, indiquant la présence d'un biais dans les deux cas.

	Paramètres des distributions gaussiennes des erreurs des simulations	
	μ (moyenne)	**σ** (écart-type)
CD	-0,05 nm	0,65 nm
H	1,19 nm	0,68 nm
SWA	-1,29 °	1,29 °

Tableau II.3 : Estimation des paramètres des distributions normales des erreurs sur les paramètres dimensionnels des échantillons simulés.

Nous formulons l'hypothèse que l'erreur systématique observée sur H est liée à la rugosité de surface introduite dans les modèles, qui pourrait perturber la détection précise de ce paramètre lors de l'analyse. De même, le biais observé sur SWA pourrait résulter du fonctionnement de l'algorithme de résolution inverse, qui tend à ajuster ce paramètre pour compenser d'éventuels écarts sur d'autres variables, afin de minimiser le nombre de pixels non-nuls.

Enfin, les faibles différences entre les incertitudes issues des simulations et les valeurs attendues, telles que présentées dans le chapitre 1, peuvent s'expliquer par la non-prise en compte de certains phénomènes, ou par des simplifications volontairement introduites dans notre approche :

- L'usure de la pointe AFM, qui modifie légèrement sa géométrie au fur et à mesure des mesures.
- L'aspect symétrique des modèles utilisés, avec des valeurs de SWA identiques à gauche et à droite.
- Le possible ajustement imparfait de certains paramètres par l'opérateur, comme l'asservissement de la pointe.

II.3.4 Conclusion

Nous avons pu concevoir un simulateur AFM, simulant l'influence géométrique de la pointe sur des échantillons. Nous avons pu implémenter les sources d'erreur métrologique sur nos modèles et données brutes. Enfin, nous avons développé un outil permettant l'extraction des paramètres dimensionnels retirant l'influence de la pointe de façon automatique. Ces résultats valident l'approche de simulation AFM et démontrent la capacité du simulateur à reproduire des caractéristiques expérimentales.

II.4 Simulations CDSEM

II.4.1 Principe de fonctionnement du simulateur

Contrairement à l'AFM, nous disposons d'un simulateur CDSEM : Nebula [58], [59]. Il s'agit d'un simulateur open source basé sur une méthode Monte Carlo, conçu pour modéliser les interactions électron-matière caractéristiques des mesures effectuées par CDSEM. Ce simulateur est optimisé pour exploiter pleinement les architectures multicœurs des CPU ainsi que les GPU, offrant ainsi des temps de calcul réduits tout en conservant une grande exactitude.

Pour que *Nebula* puisse effectuer des simulations, plusieurs éléments sont nécessaires :

- **Géométrie de l'échantillon** : géométrie représentée par un ensemble de points définissant le contour des 3 lignes (dans une approche multi lignes).
- **Propriétés du faisceau d'électrons** : plusieurs caractéristiques du faisceau doivent être définies. En s'appuyant sur les conditions expérimentales, l'énergie des électrons est fixée à 800 eV. Les mesures étant effectuées en mode « Top View », l'angle du faisceau est réglé perpendiculairement à l'échantillon (angle normal).
- **Matériaux de l'échantillon** : propriétés des matériaux constituant les lignes (silicium)

Une fois les paramètres de Nebula définis, les simulations ont été réalisées par une équipe dédiée aux outils numériques en lien avec la salle blanche au sein du CEA-Leti. La Figure II.15.A est une image CDSEM produite par simulation et la Figure II.15.B une image CDSEM obtenues expérimentalement. Nous pouvons voir que le simulateur permet de produire des images réalistes mais n'incorpore par de bruit de fond.

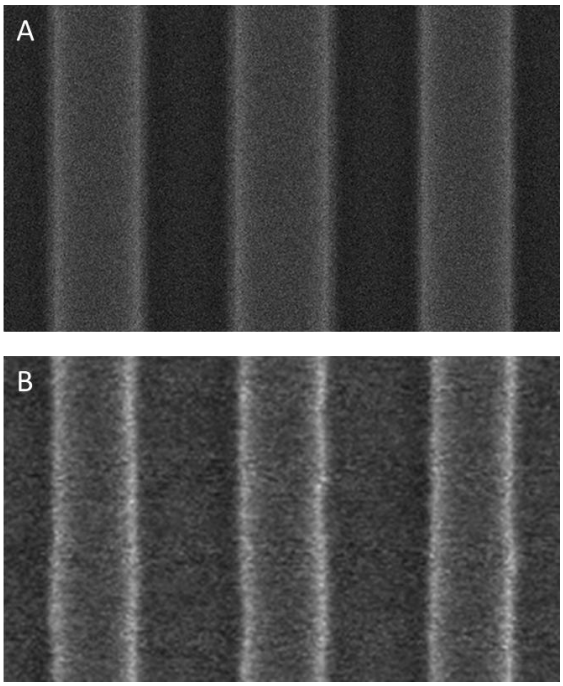

Figure II.15 : Image de mesure CDSEM, (A) mesure simulée, (B) mesure expérimentale. La mesure simulée est ici de plus haute résolution et sans implémentation de bruit.

II.4.2 Implémentation des incertitudes de mesures du CDSEM

La principale source d'incertitude dans les mesures réalisées par CDSEM provient des fluctuations quantiques des électrons, un phénomène décrit par Walter Schottky [60], [61], donnant son nom à ce phénomène, le bruit de Schottky. Ce bruit de fond, observé expérimentalement dans les images CDSEM, est généralement modélisé par une distribution de Poisson. Pour reproduire cette incertitude dans les simulations, une variable aléatoire suivant une loi de Poisson est associée à chaque pixel des images générées par Nebula, en utilisant comme paramètre de la distribution (λ) la valeur d'intensité du pixel considéré.

L'évolution induite par l'ajout du bruit est observable sur les images de simulation CDSEM, avant (Figure II.16.A) et après l'ajout du bruit de Poisson (Figure II.16.B), comparées à l'image obtenue expérimentalement (Figure II.16.C). On constate une plus grande similarité entre l'image bruitée et l'image expérimentale qu'entre la simulation initiale et la mesure réelle. Cette étape de bruitage permet ainsi de rendre les images simulées plus fidèles aux conditions expérimentales.

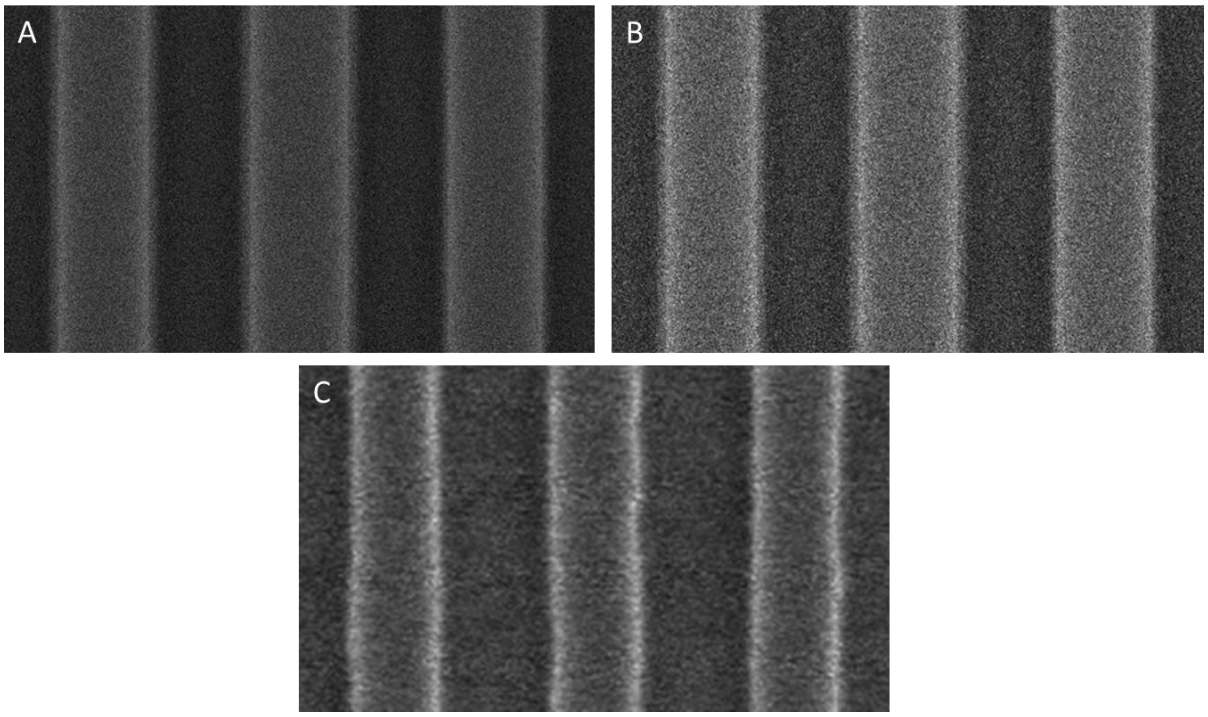

Figure II.16 : (A) Image de simulation CDSEM sans bruit. (B) Image de simulation CDSEM avec bruit de poisson. (C) Image de mesure expérimentale CDSEM

II.4.3 Analyse des paramètres dimensionnels extraits des simulations CDSEM

Les extractions du paramètre dimensionnel CD à partir des images CDSEM sont effectuées à l'aide du logiciel SIMPL développé par la société ASELTA Nanographics. Ce logiciel exploite un algorithme dit « model-based », qui offre une meilleure robustesse face aux perturbations, comme le bruit, par rapport aux techniques traditionnelles d'extraction basées sur des seuils [62]. Ainsi, SIMPL permet l'extraction des CD pour chacune des 3 lignes du modèle. Ces valeurs sont ensuite moyennées, avant d'être comparées aux valeurs de références respectives.

Nous avons confronté la distribution obtenue, Figure II.17.A, à une distribution normale au moyen d'un diagramme quantile-quantile. Nous avons un R^2 de 0,9815, montrant ainsi la similarité entre la distribution des erreurs et une distribution gaussienne. Cela permet d'estimer cette distribution, illustrée par la Figure II.17.B, permettant alors de quantifier précisément les paramètres de la distribution normale associée aux erreurs de mesures des simulations CDSEM, (Tableau II.4).

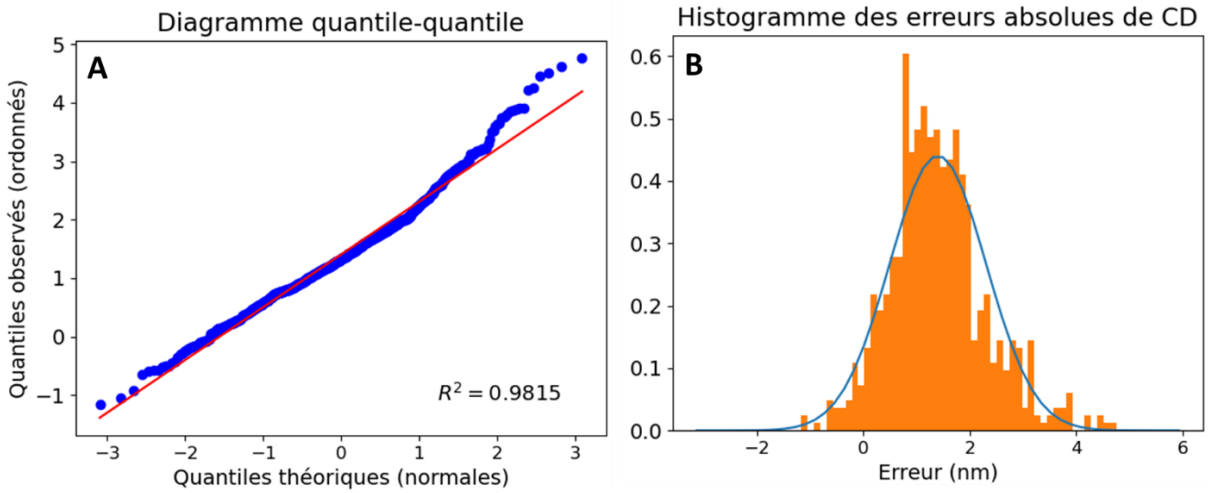

Figure II.17 : (A) Diagramme quantile-quantile des erreurs absolues de CD. (B) Histogramme des erreurs absolues de CD (orange) et distribution normale associé (bleu).

Les incertitudes de mesures ont pu être générées par une modélisation 3 lignes et l'ajout du bruit de poisson, permettant une modélisation des phénomènes observés expérimentalement. Ainsi, les erreurs associées aux profils simulés de CD-SEM ont été évaluées par rapport aux valeurs de références, définies en début de chapitre. Les résultats, présentés dans le Tableau II.4, montrent un écart-type des erreurs de σ = 0,91 nm et une erreur moyenne de μ = 1,40 nm. Cette erreur moyenne significative traduit la présence d'un biais systématique dans la méthode d'extraction du CD.

Erreurs des simulations (CD)	**σ** (écart-type)	**μ** (moyenne)
	0,91 nm	1,40 nm

Tableau II.4 : Estimation des écart-types et moyennes des erreurs sur CD des échantillons simulés.

Ce biais peut s'expliquer par plusieurs facteurs. En premier lieu, l'extraction du « CD top » (dimension critique au sommet des lignes) peut être rendue incertaine en raison de la difficulté à identifier précisément la transition entre la surface des motifs et leurs flancs, notamment en présence d'un arrondi (corner rounding) en haut de ligne. Cette incertitude est également sensible à la valeur du SWA, ce qui complique davantage la détection des contours.

Enfin, il convient de souligner que certaines sources d'erreur n'ont pas été prises en compte dans les simulations, notamment la rugosité longitudinale (selon l'axe y) et la rugosité de surface (selon l'axe z), ce qui pourrait expliquer la faible différence observée entre les erreurs obtenues et celles attendues, telles que présentées dans le chapitre 1.

II.4.4 Conclusion

Les différentes étapes mises en œuvre nous ont permis de générer un jeu de données CDSEM simulées intégrant des sources de variabilité réalistes. L'introduction de variations géométriques contrôlées, combinée à l'ajout de bruit de Poisson, a permis de reproduire des artefacts et des distributions comparables à celles observées expérimentalement. Afin d'assurer une continuité entre simulation et expérimentation, les images générées ont été analysées à l'aide des mêmes outils de traitement et d'extraction que ceux employés dans le cadre des mesures réelles. Cette approche garantit une cohérence méthodologique et renforce la représentativité des données simulées. Ainsi, le jeu de données obtenu constitue une base fiable pour l'entraînement de réseau de neurones.

II.5 Simulations OCD

II.5.1 Principe de fonctionnement du simulateur

La génération du jeu de données pour la technique métrologique dite **OCD** a été réalisée grâce à un simulateur développé en interne au **LTM**, nommé **S-Genius** [63]. Cet outil utilise la Méthode Modale par Expansion de Fourier (MMFE) [64] pour simuler les interactions onde-matière. Cette approche repose sur une représentation des champs électromagnétiques en termes de **modes propres locaux** dans chaque région d'une structure périodique. Ces modes sont ensuite utilisés pour établir un système d'équations linéaires décrivant l'interaction entre les différentes régions via les conditions de continuité des champs électromagnétiques aux interfaces.

Pour effectuer une décomposition en modes propres, plusieurs prérequis doivent être remplis :

- Périodicité : La structure doit présenter une périodicité dans une ou plusieurs directions (ici selon l'axe x) permettant une décomposition en série de Fourier des champs électromagnétiques et des propriétés matérielles (comme la permittivité).
- Stratification : La structure doit être stratifiée en couches ayant des constantes diélectriques homogènes dans la direction normale au plan de la périodicité (ici selon l'axe z). Si la structure ne respecte pas cette homogénéité, elle doit être approximée par une superposition de structures élémentaires respectant cette condition, comme illustré par la Figure II.18.

Dans chaque couche, les équations de Maxwell sont résolues en supposant une dépendance spatiale périodique des champs dans la direction x. Ensuite, par des

conditions de continuité aux interfaces, les champs sont calculés dans toute la structure. Ainsi, nous pouvons avoir les spectres décrivant la structure étudiée.

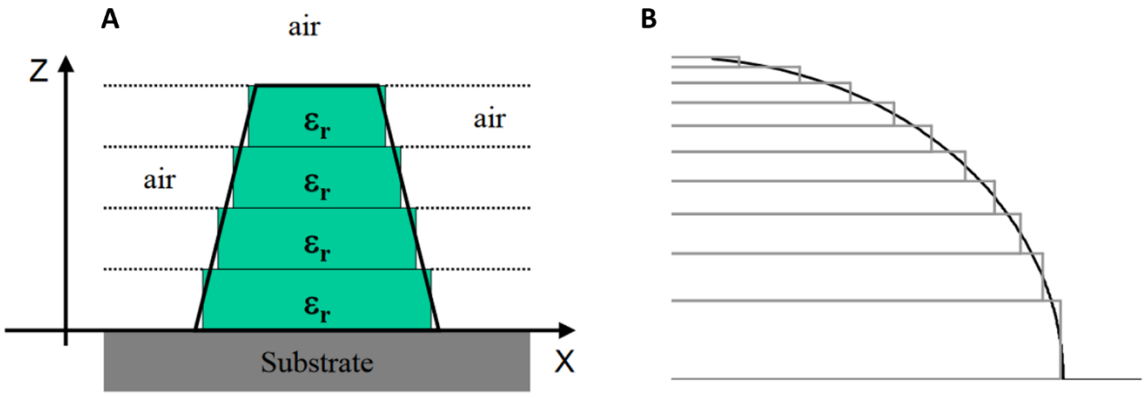

Figure II.18 : Figure extraite des travaux de David FUARD [63]. (A) Exemple de découpage lamellaire pour une section de ligne trapézoïdale avec une permittivité ε_r. (B) Vue du découpage lamellaire optimisé (en gris) pour un motif arrondi (en noir).

II.5.2 Etudes pour le paramétrage du modèle

Comme mentionné précédemment, il est nécessaire de découper la structure de notre modèle en structures élémentaires. Or, le modèle utilisé comporte des arrondis sur les coins du motif (appelés corner rounding) nécessitant une décomposition en structures élémentaires appropriés pour garantir une représentation fidèle.

Il est donc nécessaire de déterminer le nombre minimal de couches nécessaires pour obtenir une représentation suffisamment précise c'est-à-dire un modèle dont les résultats de simulation ne sont pas significativement affectés par la simplification géométrique. Pour évaluer cette exigence, nous avons défini un **modèle de référence**, qui représente le trapézoïde avec *corners rounding* à l'aide d'un **nombre élevé de couches**, ici fixé à **30 couches**. Ce modèle sert de base pour comparer les simulations obtenues avec des modèles décomposés en un nombre inférieur de couches.

Des tests réalisés avec S-Genius ont été menés afin d'explorer la manière de concevoir un modèle, notamment en faisant varier la forme et le nombre de couches. Ces tests ont également servi à analyser le fonctionnement de l'algorithme de résolution de problème inverse et connaitre quel est le score utilisé. Ainsi, le score est le calcul des erreurs quadratique des spectres Rp et Rs (gSE2) :

$$gSE2 = \sum \left(\left(Rp_{ref}\,(x) - Rp(x) \right)^2 + \left(Rs_{ref}\,(x) - Rs(x) \right)^2 \right) \qquad \textbf{(II.14)}$$

x est la longueur d'onde mesurée

L'analyse des scores obtenus par résolution de problème inverse lors des tests a permis d'identifier un seuil minimal atteignable. Cette valeur est 10^{-3} et correspond au plus bas niveau d'erreur observé, déterminé de façon empirique. Ce seuil sert de référence pour évaluer la pertinence des différents modèles testés et orienter la sélection de la structure la plus adaptée.

Différentes modélisations ont été simulées et comparées à un modèle de référence. Les modèles étudiés ont été volontairement sur-décomposés en un nombre de couches supérieur à celui attendu, afin d'isoler l'impact du *corner rounding* indépendamment de la capacité du modèle à être résolu. Le score calculé correspond ici à l'erreur entre les spectres simulés des différents modèles. Ainsi, si cette erreur est inférieure à 10^{-3}, on peut considérer que la modélisation n'impacte pas significativement les résultats attendus lors d'une résolution inverse.

Trois configurations ont été comparées :

- Modèle simplifié : ne comporte aucun corner rounding, illustré par la Figure II.19.A.
- Modèle intermédiaire : inclut uniquement les corners rounding en bas, illustré par la Figure II.19.B.
- Modèle de référence : inclut les corners rounding à la fois en haut et en bas des lignes, illustré par la Figure II.19.C.

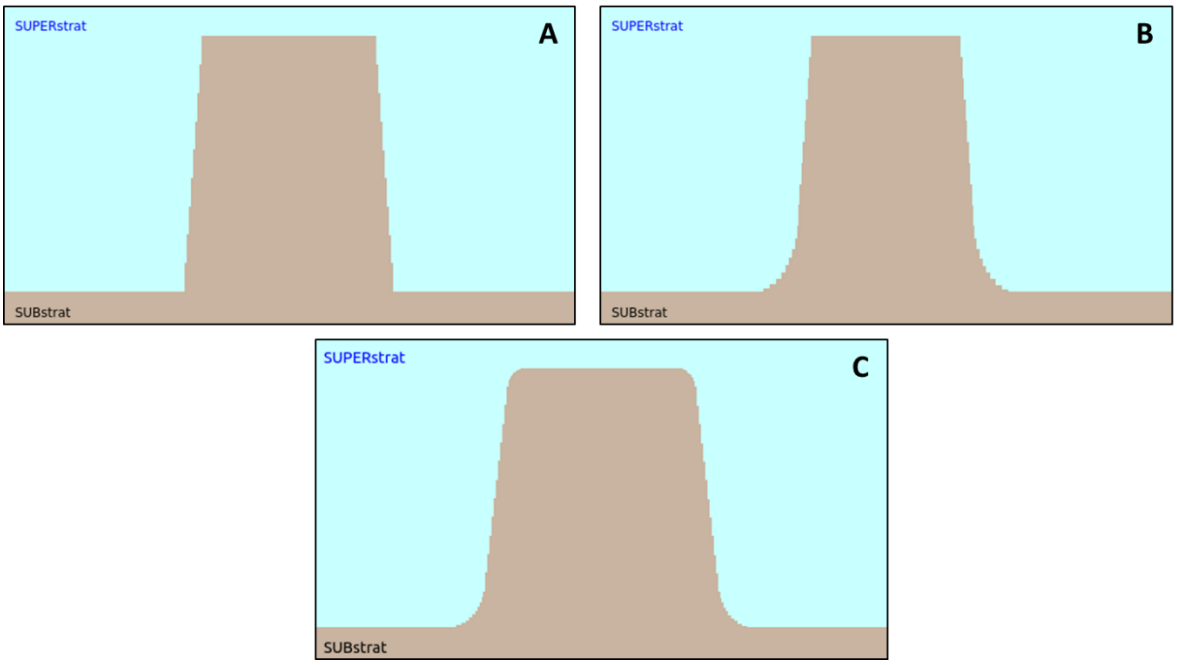

Figure II.19 : Représentation d'une ligne suivant le modèle (A) simplifié, (B) intermédiaire, (C) Complet (de référence).

Nous avons réalisé une étude similaire sur une autre technique métrologique, en simulant cette fois une série de 48 échantillons, pour lesquels les paramètres dimensionnels ont été variés dans leur plage de valeurs respectives. Pour chaque échantillon, nous avons calculé l'erreur entre le modèle de référence (correspondant au modèle complet) et le modèle intermédiaire, ainsi qu'entre le modèle de référence et le modèle simplifié.

La Figure II.20.A présente, sous forme de boîte à moustaches, les erreurs entre le modèle complet et le modèle simplifié. Ces erreurs dépassent très largement le seuil de tolérance fixé à 10^{-3}, avec une valeur maximale atteignant 16,64. À l'inverse, la Figure II.20.B montre les erreurs entre le modèle complet et le modèle intermédiaire. Les écarts restent ici nettement plus faibles, avec une erreur maximale de $3{,}4 \times 10^{-3}$, légèrement supérieure au seuil défini.

Ces résultats montrent qu'aucun des deux modèles réduits ne respecte strictement le critère initial. Toutefois, le modèle intermédiaire présente un comportement bien plus fidèle au modèle complet que ne le fait le modèle simplifié. La Figure II.21, qui compare les spectres Rp et Rs obtenus pour chacun des trois modèles, illustre clairement cette différence : le modèle simplifié génère des écarts notables vis-à-vis du modèle de référence, alors que le modèle intermédiaire en reste beaucoup plus proche.

Malgré le léger dépassement du seuil, plusieurs arguments justifient l'adoption du modèle intermédiaire pour la suite des travaux. D'une part, le seuil fixé découle d'un critère empirique, établi à partir des erreurs maximales observées lors de multiples résolutions du problème inverse. D'autre part, les études expérimentales menées par l'équipe métrologie du CEA-LETI ne prennent pas en compte certains détails géométriques tels que les *corner roundings* en haut de motifs, considérés comme négligeables pour cette configuration, ce qui correspond à une modélisation intermédiaire. Enfin, les simulations envisagées nécessitent l'introduction de variations inter-lignes dans un cadre multi-ligne, ce qui impose des modifications au simulateur. Bien que des ajustements restent nécessaires même avec le modèle intermédiaire, ils s'avèrent bien plus légers que ceux qu'exigerait l'emploi du modèle complet.

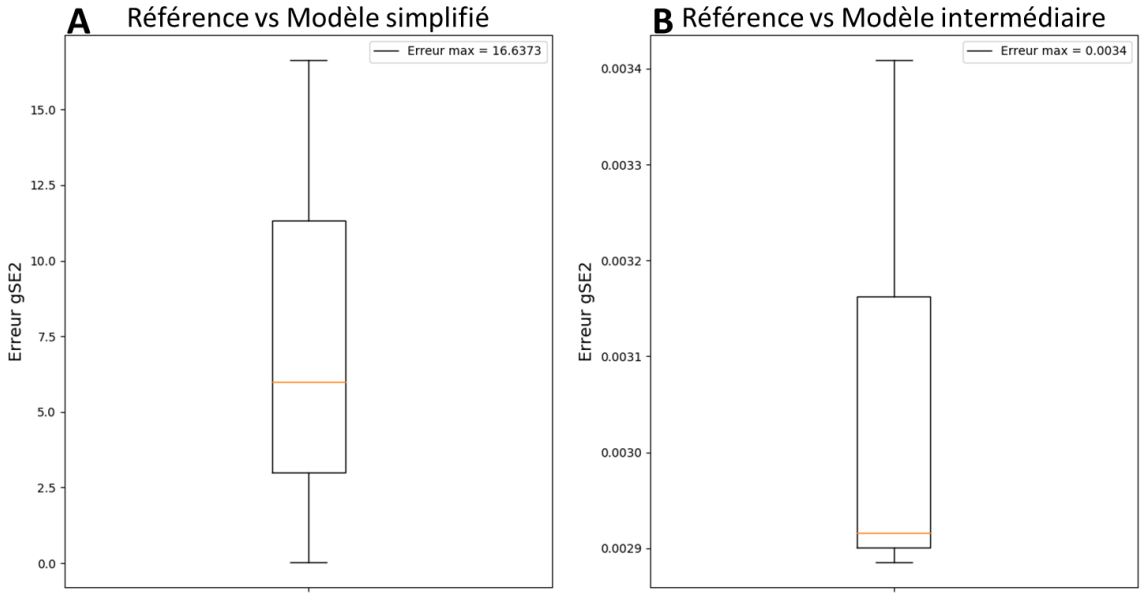

Figure II.20 : Boite à moustache des erreurs, (A) entre le modèle de référence et le modèle simplifié, (B) entre le modèle de référence et le modèle intermédiaire

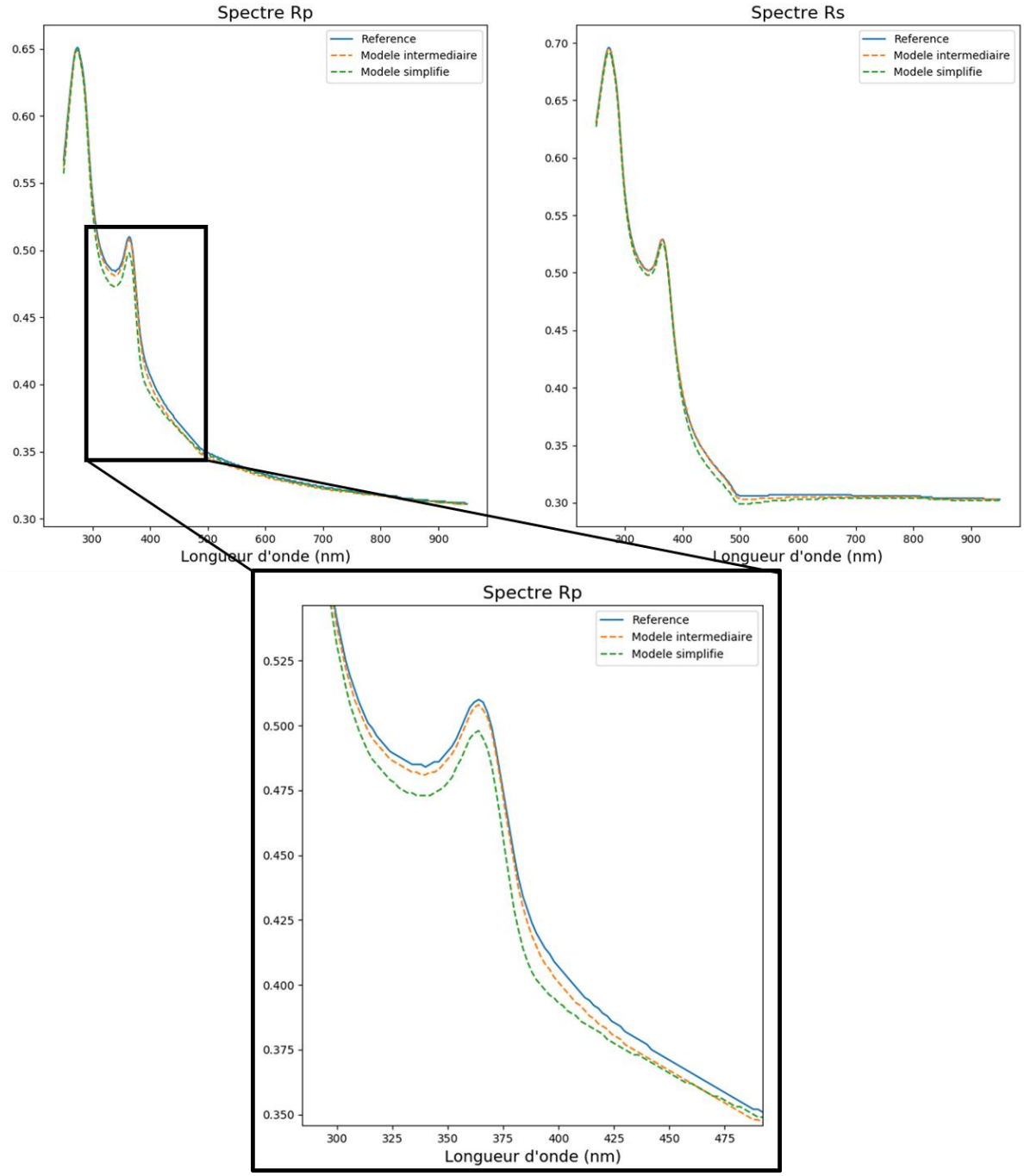

Figure II.21 : Spectres Rp et Rs des modèles de référence (en bleu), intermédiaire (en orange) et simplifié (en vert)

Ces observations permettent de conclure que la modélisation des corners rounding en haut des lignes n'est pas nécessaire, car elle n'a pas d'impact significatif sur les résultats des simulations. Par conséquent, l'utilisation du modèle intermédiaire constitue un compromis efficace entre exactitude et complexité de modélisation.

Ainsi, pour la suite des travaux, nous utiliserons exclusivement le modèle intermédiaire. Ce modèle permet de distinguer deux zones :

- Zone L1, correspondant aux corners rounding présents en bas des lignes.
- Zone L2, correspondant à la partie trapézoïdale de la ligne.

Ces deux zones sont représentées dans la Figure II.22. Nous devons alors déterminer le nombre minimal de couches nécessaires pour modéliser ces deux zones. Afin que le nombre de couches n'ait pas d'impact significatif sur les résultats nous utiliserons le critère précédemment utilisé (gSE2 < 10^{-3}).

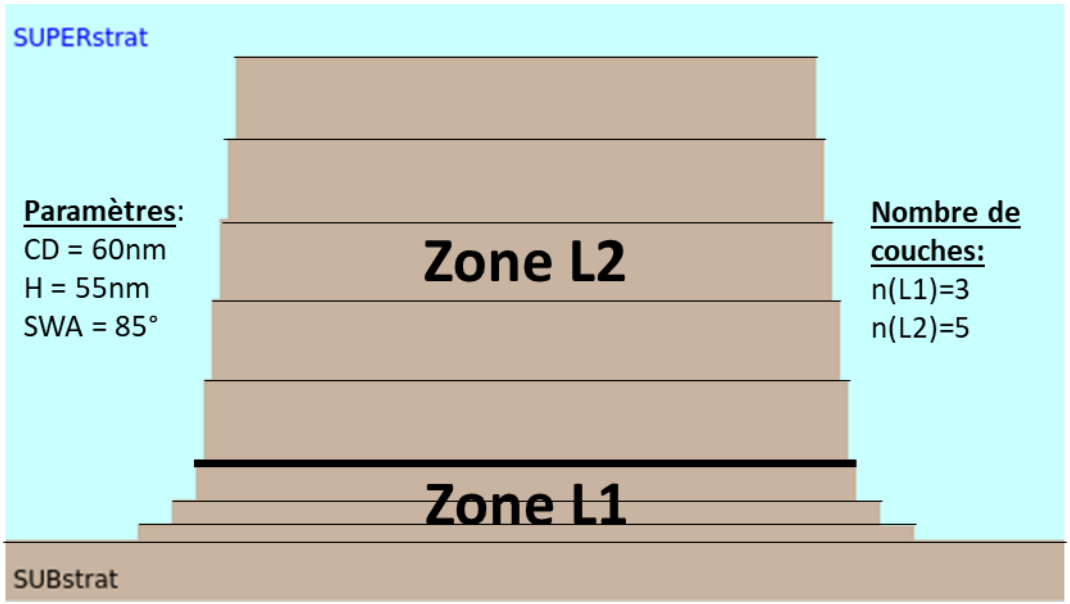

Figure II.22 : Représentation graphique du simulateur SGenius du modèle intermédiaire

Le paramètre dimensionnel clé pour déterminer le nombre de couches nécessaires est la hauteur des échantillons. Cependant, il est important de noter que la zone L1 est indépendante de la hauteur globale de la ligne. Elle est influencée principalement par le corner rounding et le side wall angle (SWA). En considérant les plages de valeurs des échantillons étudiés, la hauteur de la zone L1 varie entre 9.1 nm et 10 nm. La Figure II.23 représente l'évolution de l'erreur quadratique en fonction du nombre de couche dans la zone L1 pour 3 valeurs de hauteur différentes. Cette Figure II.23 montre que la hauteur totale de la ligne n'a pas d'impact significatif sur les résultats de simulation pour la zone L1. Ainsi, nous déterminons qu'il est nécessaire d'utiliser au moins 3 couches dans la zone L1 afin de respecter le critère défini, à savoir une erreur gSE2 inférieure à 10^{-3}.

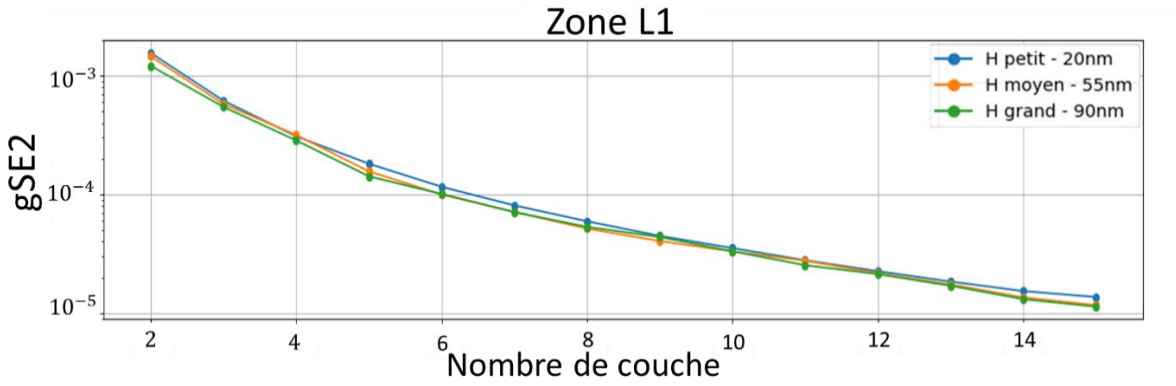

Figure II.23 : Erreurs gSE2 en fonction du nombre de couches dans la zone L1 entre un modèle de références (ayant 30 couches par zone) et des modèles d'études (ayant 30 couches dans la zone L2)

Contrairement à la zone L1, la hauteur de zone L2 est fortement dépendante de la hauteur de la ligne. Ainsi, la Figure II.24 montre que le nombre de couches nécessaires pour respecter le critère d'une erreur inférieure à 10^{-3} varie en fonction de la hauteur de la ligne. Pour résoudre ce problème, nous avons opté pour une représentation alternative qui ne dépend pas du nombre total de couches, mais plutôt de la hauteur individuelle des couches, illustré par la Figure II.25. Cette représentation permet une modélisation plus adaptée, mettant en évidence que le critère important est la hauteur des couches et non leur nombre.

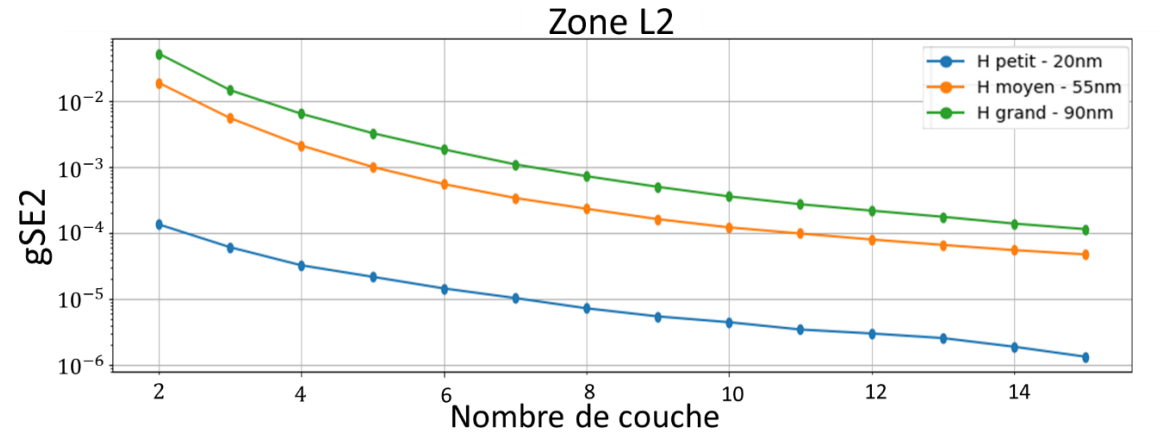

Figure II.24 : Erreurs gSE2 en fonction du nombre de couches dans la zone L2 entre un modèle de références (ayant 30 couches par zone) et des modèles d'études (ayant 30 couches dans la zone L1)

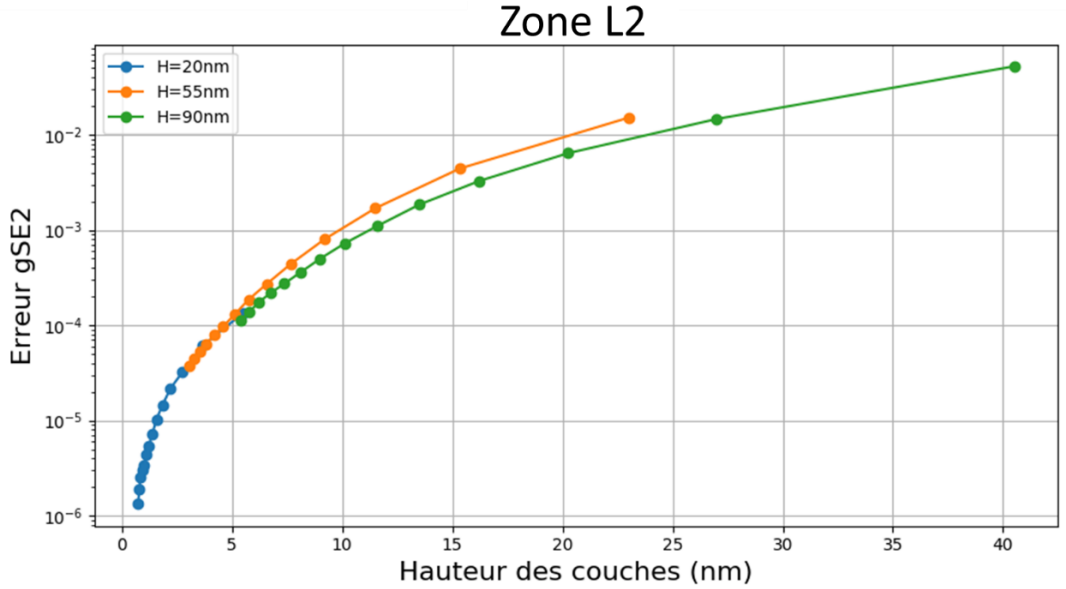

Figure II.25 : Erreurs gSE2 en fonction de la hauteur des couches dans la zone L2 entre un modèle de références (ayant 30 couches par zone) et des modèles d'études (ayant 30 couches dans la zone L1)

Afin de garantir que la valeur de gSE2 reste inférieure à 10^{-3}, la hauteur des couches dans la zone L2 doit être inférieure à 10 nm. Le nombre de couches dans cette zone dépend donc de la hauteur totale de l'échantillon, afin de satisfaire ce critère. À l'inverse, la zone L1 comporte systématiquement trois couches, indépendamment de la hauteur des lignes, puisque cette dernière n'influence pas la hauteur de la zone L1. Il en résulte que les couches de la zone L1 sont plus fines que celles de la zone L2.

Une fois le modèle optimisé en fonction des nombres de couches déterminés pour chaque zone, il a été constaté que les erreurs générées dans les zones L1 et L2 se cumule. Cette accumulation d'erreurs a conduit à un dépassement marginal du critère de 10^{-3}. Pour résoudre ce problème, une couche supplémentaire a été ajoutée dans chaque zone, assurant ainsi que l'erreur globale respecte le seuil fixé (4 pour L1 et des couches de 10 nm pour L2 avec une supplémentaire).

Grâce à ces études, nous avons défini un découpage qui optimise à la fois l'exactitude des résultats et la complexité des modèles. En ajustant finement le nombre et la hauteur des couches dans les zones L1 et L2, nous avons déterminé les paramètres optimaux permettant de minimiser le temps de simulation tout en respectant les critères d'erreur fixés. Cette optimisation garantit un équilibre entre la fidélité des simulations et l'efficacité computationnelle.

II.5.3 Implémentation des incertitudes de mesure

La modélisation et l'implémentation des incertitudes dans la technique OCD est nettement plus complexe que pour les techniques directes précédemment abordées. Cela est dû au caractère « indirect » de l'OCD, où les données brutes sont des signaux spectraux, nécessitant une compréhension approfondie des phénomènes physiques sous-jacents. En nous appuyant sur les travaux de thèse de Mohamed El Kodadi portant sur le développement des techniques de scatterométrie en temps réel pour le suivi des procédés de gravure plasmas [45], il apparaît qu'il existe une multitude de sources de bruits et d'erreurs. Ces incertitudes peuvent être regroupées en deux grandes catégories :

- Incertitudes de type A : liées aux bruits des capteurs, affectant directement la qualité des signaux mesurés.
- Incertitudes de type B : liées aux imprécisions dans les paramètres utilisés, comme les angles d'incidence ou les indices de réfraction, qui peuvent varier de façon significative.

Concernant les incertitudes de type A, une étude a été menée en estimant le bruit présent sur les spectres de mesures expérimentales. Cette estimation a été réalisée à partir d'un ensemble de mesures effectuées sur des wafers vierges dont les spectres ont été moyennés pour identifier le bruit caractéristique. Nous pouvons observer sur Figure II.26 deux courbes, en bleu le bruit caractéristique l'équipement utilisé par les équipe du LETI, s'alignant avec d'autres travaux [9]. Une estimation du bruit a été réalisée à l'aide d'un ajustement polynômial. Le choix de l'ordre du polynôme s'est appuyé de manière empirique sur les valeurs du coefficient de détermination R^2. Ainsi, la courbe rouge correspond à un polynôme d'ordre 5, ajusté sur le bruit observé du spectre Rp. Les coefficients R^2 obtenus sont respectivement de 0,989 pour le spectre Rp et de 0,998 pour le spectre Rs. Les polynômes ainsi déterminés ont ensuite été appliqués à des spectres simulés, préalablement à la résolution des problèmes inverses.

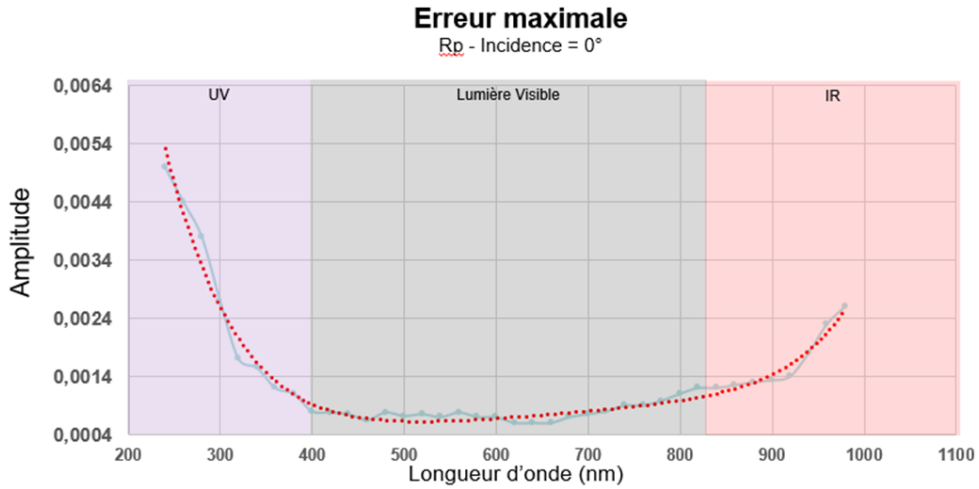

Figure II.26 : Estimation de l'erreur maximale pour le spectre Rp en fonction de la longueur d'onde entre une série de mesures et leur valeurs moyennes. En bleu, le bruit caractéristique de l'équipement. En rouge le polynôme d'ordre 5 associé à ce bruit.

Cependant, les résultats illustrés par la Figure II.27 montrent que l'ajout de ce bruit aux spectres simulés n'a qu'un impact limité sur les résultats finaux, ne générant pas un niveau d'incertitudes de mesure suffisant. Ce constat est cohérent avec les travaux de Mohamed El Kodadi [45], qui avait déjà souligné que les incertitudes de type A, bien que présentes, ont une influence relativement faible sur les résultats en scatterométrie, comparées aux sources d'incertitudes de type B.

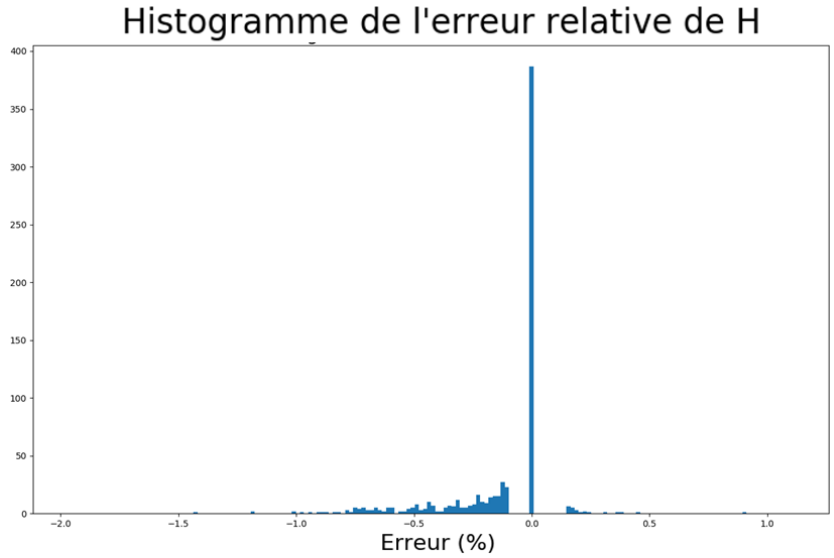

Figure II.27 : Histogramme des erreurs relatives sur le paramètre H entre des échantillons simulés puis bruit et le modèle trouvé par résolution de problème inverse

Concernant les incertitudes de type B, il a été démontré [45] que leur principale origine est une connaissance imparfaite de l'indice de réfraction, comme dans le cas de la résine étudiée par Mohamed El Kodadi [45]. Cependant, dans notre cas, les lignes

étant composées de silicium, un matériau dont l'indice de réfraction est bien connu et maîtrisé, ce type d'incertitudes n'a qu'un impact négligeable sur nos simulations. Il ne semble donc pas pertinent de se concentrer sur ce type de bruit, car il ne génère pas d'erreur significative.

Ainsi, nous concluons que les incertitudes, qu'il s'agisse des types A ou B, sont négligeables dans notre contexte. Ce constat est cohérent avec d'autres études [9], qui soulignent que la dégradation de la exactitude des mesures est avant tout liée aux choix des modèles utilisés. C'est donc sur cet aspect que nous allons focaliser nos efforts, afin de mieux intégrer les incertitudes dans nos simulations et de générer des mesures plus représentatives des données expérimentales.

Afin de prendre en compte les incertitudes de mesure liées à la modélisation, nous nous sommes appuyés sur des travaux portant sur la prise en compte de la rugosité et des variations géométriques inter-lignes à l'aide d'une modélisation multi-lignes (ou super-cell en anglais) [47]. Dans un premier temps, des simulations ont été réalisées à l'aide de cette approche, sans introduire de variations inter-lignes, dans le but de vérifier la concordance des spectres avec ceux issus d'un modèle mono-ligne. Ces simulations ont toutefois révélé des écarts significatifs entre les spectres Rp obtenus et ceux générés à partir d'une modélisation mono-ligne classique, comme l'illustre la Figure II.28.A. Toutefois, nous n'observons que peu de différences sur les spectres Rs, comme nous le montre la Figure II.28.B.

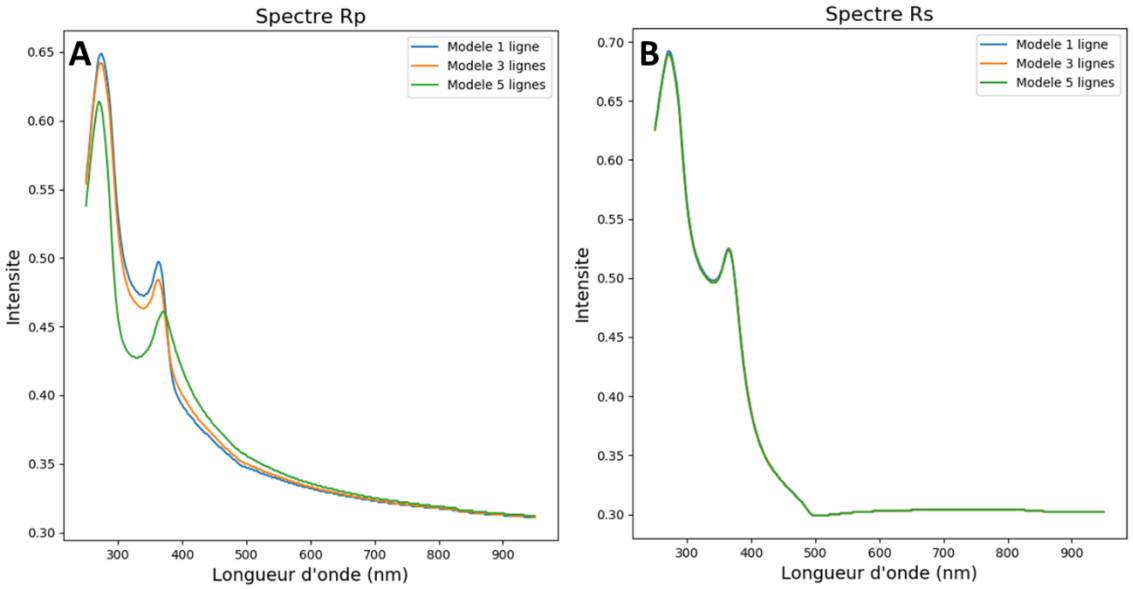

Figure II.28 : Comparaison de spectres Rp (A) et Rs (B) de simulations de modèle 1 ligne, 3 lignes et 5 lignes

Pour comprendre les différences observées entre les spectres, il convient d'examiner les calculs réalisés par la méthode MMFE [65], [66], et plus

particulièrement l'équation suivante, qui met en évidence les simplifications introduites par le simulateur concernant le nombre d'harmoniques prises en compte :

$$\epsilon(x) = \sum_{m=-M}^{M} \epsilon_m e^{im\frac{2\pi}{\Lambda}x} \qquad \textbf{(II.15)}$$

M est le nombre d'harmonique maximal à prendre en compte
$\epsilon(x)$ est la fonction de permittivité en fonction de la position spatiale x
ϵ_m est le coefficient de Fourier pour l'harmonique m
Λ est la période du modèle

En effet, dans le simulateur S-Genius, une approximation est faite pour limiter le nombre d'harmoniques m dans la série de Fourier à une plage de ±5 pour des raisons de coûts computationnels. Cette limitation devient particulièrement importante lorsque la période Λ augmente, comme c'est le cas avec la modélisation multi-lignes. Une augmentation de Λ sans ajustement du nombre d'harmoniques m entraîne une diminution du rapport m/Λ, ce qui impacte directement l'exactitude des résultats.

Pour compenser cet effet, nous avons ajusté m en maintenant un rapport constant m/Λ. Ce réglage permet de corriger les écarts observés, ayant un plus grand impact sur les spectres Rp, illustré par la Figure II.29.A que sur les spectres Rs, illustré par la Figure II.29.C. Nous pouvons voir que le modèle 3 lignes ne prenant en compte que les 15 premières harmoniques est celui correspondant le mieux aux spectres du modèle de référence, comme le montre les Figure II.29.A et Figure II.29.B. Ainsi, nous rendons les spectres obtenus avec une modélisation multi-lignes identiques avec ceux issus d'une modélisation mono-ligne. Cette stratégie garantit que les variations dimensionnelles sont bien modélisées sans introduire d'artéfacts dus à une approximation numérique inadéquate.

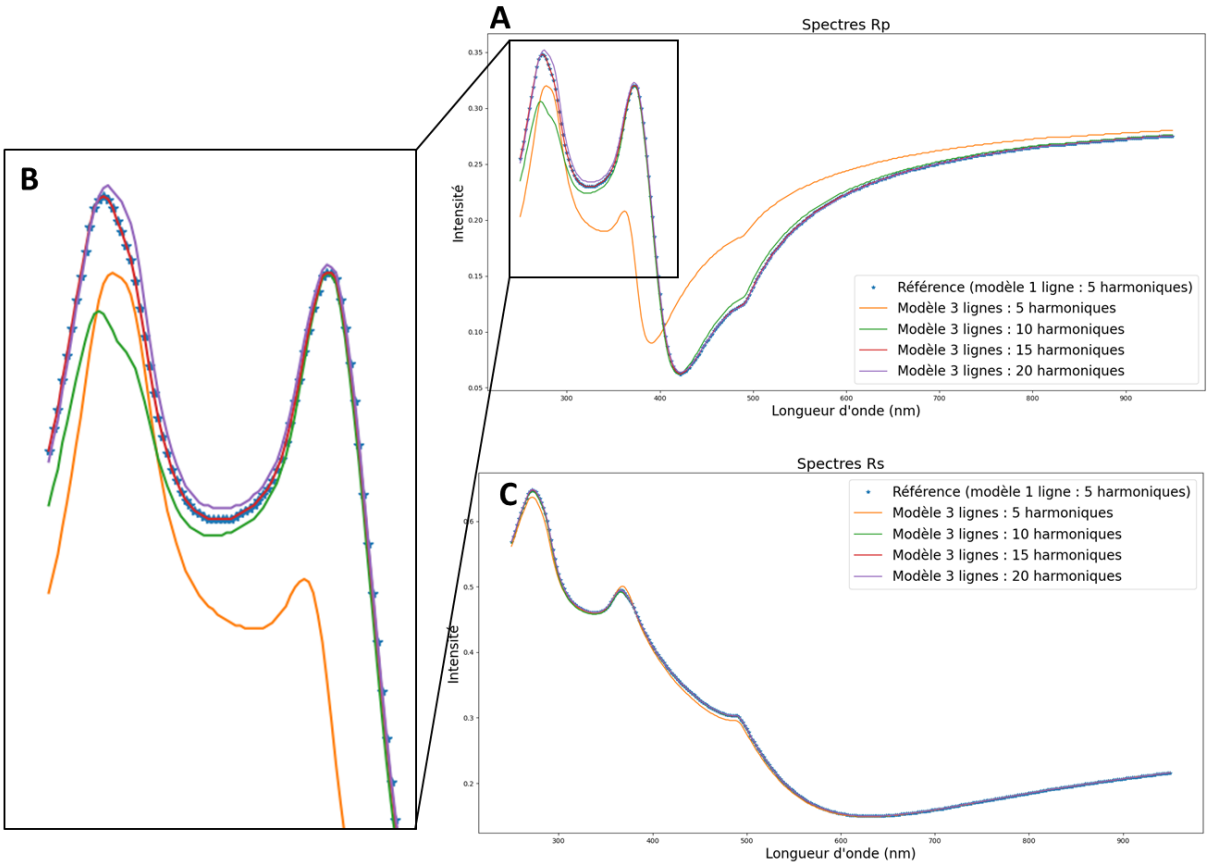

Figure II.29 : Simulation de modèle 1 ligne (en prenant en compte que les 5 première harmonique) et de modèles 3 lignes en prenant en compte un nombre différent d'harmonique). (A) Spectres Rp de 250 nm à 950 nm. (B) Spectres Rp de 250 nm à 400 nm. (C) Spectres Rs de 250 nm à 950 nm

Pour la suite des travaux, nous avons décidé d'utiliser une modélisation à trois lignes avec un nombre d'harmoniques fixé à $m = \pm 15$, pour la génération des échantillons. Ce choix équilibre les besoins en temps de calcul et l'exactitude des simulations, tout en conservant une diversité minimale des variations inter-lignes. Ce réglage garantit que les spectres obtenus sont représentatifs des phénomènes réels tout en étant économes en ressources.

Afin de modéliser les variations de hauteur entre les lignes, et en raison des limitations imposées par le simulateur, deux zones supplémentaires (zones L3 et L4) ont été ajoutées, comme illustré dans la Figure II.30. Ces nouvelles zones, qui prolongent la zone L2, doivent conserver une hauteur de couche inférieure à 10 nm afin de respecter le critère d'erreur fixé. Or, les variations de hauteur inter-lignes observées sont largement inférieures à cette valeur. Par conséquent, une seule couche est suffisante pour modéliser chacune de ces zones.

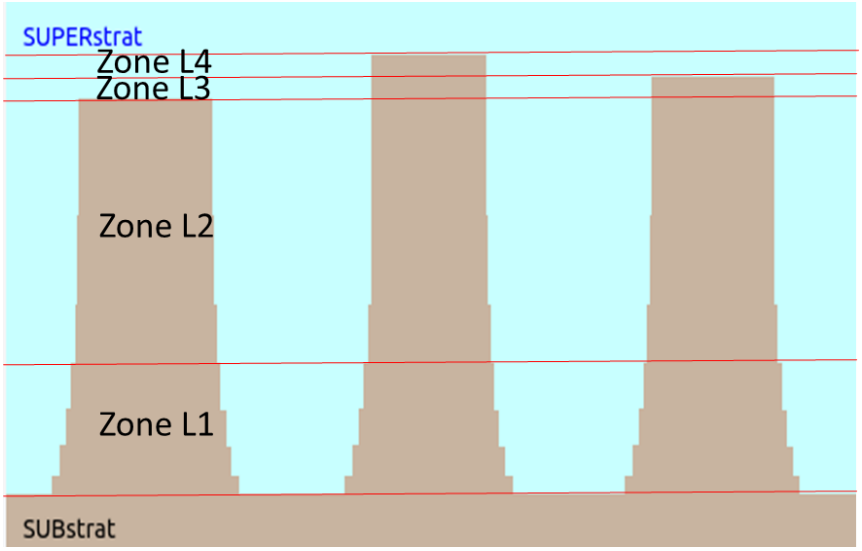

Figure II.30 : Représentation graphique de S-Genius du modèle 3 lignes

Ainsi, nous avons déterminé le modèle optimal pour les simulations des échantillons, en trouvant un équilibre entre justesse des résultats et temps de calcul. Ce travail a impliqué la définition précise de la forme du modèle et son découpage en formes élémentaires, ainsi que l'ajustement des paramètres nécessaires pour permettre l'utilisation efficace du modèle à trois lignes. Ces avancées nous permettent désormais d'intégrer des incertitudes de mesure représentatives des conditions expérimentales, renforçant ainsi la fiabilité et la pertinence des simulations.

II.5.4 Analyse des paramètres dimensionnels extraits des simulations OCD

Comme expliqué précédemment, nous avons défini les paramètres nécessaires pour établir les modèles à trois lignes permettant de simuler les 700 échantillons. Suite à ces simulations, des résolutions de problèmes inverses ont été réalisées en utilisant l'algorithme de Gauss-Newton [67] et un modèle à une seule ligne. Cette approche reproduit les étapes appliquées dans le cadre du traitement de données expérimentales pour extraire les paramètres dimensionnels, à savoir le CD (Critical Dimension), la hauteur (H), et l'angle de paroi latérale (SWA).

Suite à l'extraction des paramètres dimensionnels des échantillons par résolutions de problème inverse, nous avons pu calculer les erreurs de mesures liées aux simulations. Ainsi, nous avons obtenu les distributions des erreurs. Nous avons confronté la distribution des erreurs du paramètre CD obtenue à une distribution normale via un diagramme quantile-quantile, Figure II.31.A. Nous obtenons un R^2 de 0,9797, nous permettant d'affirmer que notre distribution des erreurs est assimilable à une distribution normale. Cela permet d'estimer cette distribution, illustrée par la

Figure II.31.B, permettant alors de quantifier précisément les paramètres de la distribution normale associée aux erreurs de mesures des simulations OCD.

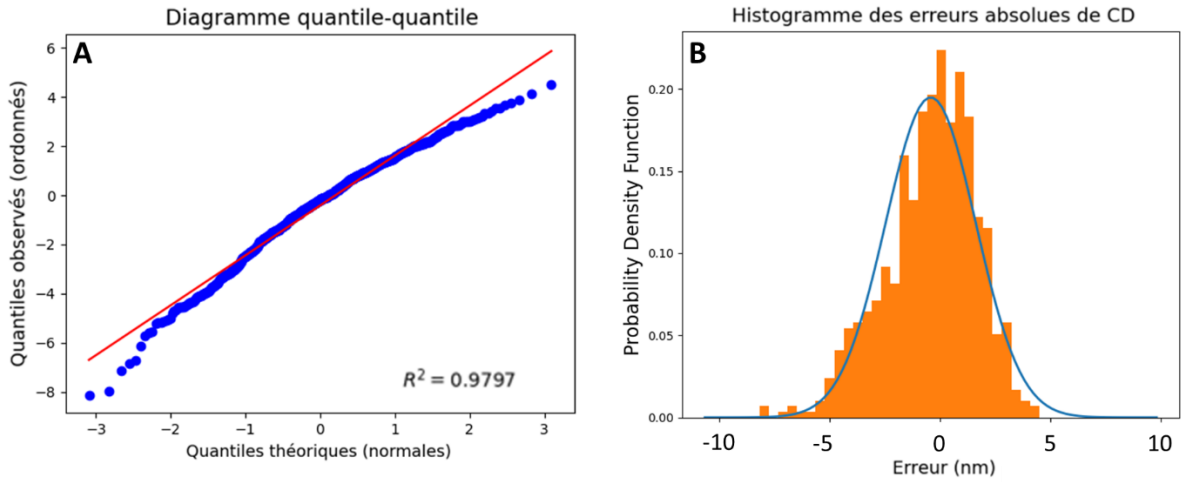

Figure II.31 : (A) Diagramme quantile-quantile des erreurs absolues de CD. (B) Histogramme des erreurs absolues de CD (orange) et distribution normale associé (bleu).

Nous avons par la suite confronté nos distributions des erreurs des paramètres H et SWA avec des distributions normales. Nous obtenons respectivement un R^2 de 0,9971 pour H et un R^2 de 0,981 pour SWA. Ainsi, nous pouvons estimer les distributions gaussiennes associé, illustré par les Figure II.32.A et Figure II.32.B.

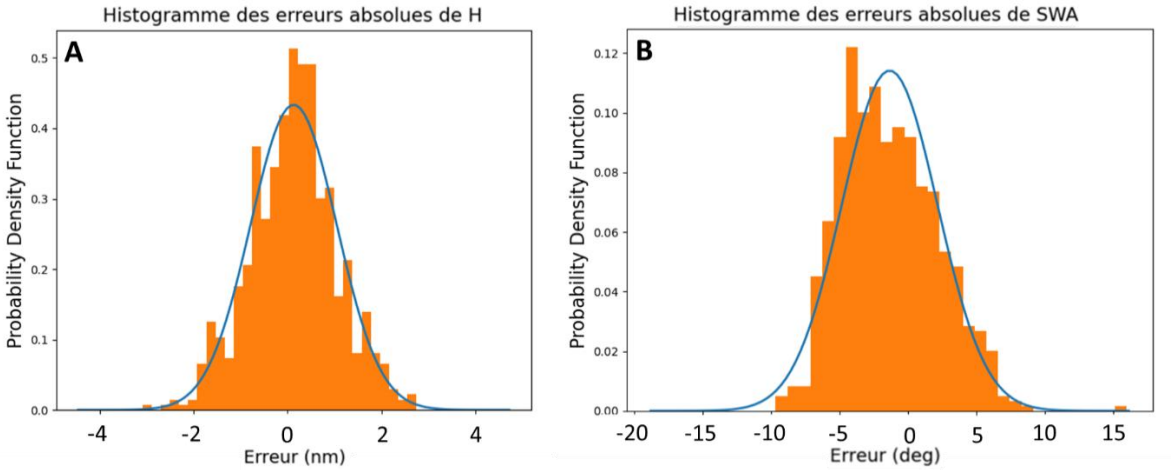

Figure II.32 : (A) Histogramme des erreurs absolues de H (orange) et distribution normale associée (bleu). (B) Histogramme des erreurs absolues de SWA (orange) et distribution normale associée (bleu).

Les incertitudes de mesure associées aux profils simulés par OCD ont été évaluées pour les paramètres CD, H et SWA, par rapport aux valeurs de référence définies en début de chapitre. Les erreurs obtenues, résumées dans le Tableau II.6, présentent un écart-type de σ = 2,05 nm pour le CD, avec une erreur moyenne de μ = -0,41 nm. Pour la hauteur H, les valeurs sont de σ = 0,92 nm et μ = 0,13 nm, tandis que pour le SWA,

une variabilité plus importante est observée, avec un écart-type de σ = 3,49° et une erreur moyenne de μ = -1,32°.

	Paramètres des distributions gaussiennes des erreurs des simulations	
	μ (moyenne)	**σ** (écart-type)
CD	-0,41 nm	2,05 nm
H	0,13 nm	0,92 nm
SWA	-1,32 °	3,49 °

Tableau II.5 : Estimation des paramètres des distributions normales des erreurs sur les paramètres dimensionnels des échantillons simulés.

Ces incertitudes apparaissent plus élevées que celles attendues, notamment pour les paramètres CD et SWA tels que définis au chapitre 1. Cet écart peut s'expliquer par plusieurs limitations du modèle utilisé. En effet, une modélisation à trois lignes ne permet probablement pas de capturer l'ensemble des subtilités liées aux variations inter-lignes. Ces variations, lorsqu'elles sont modélisées sur un nombre restreint de lignes, peuvent avoir un impact plus important que dans le cas d'un modèle à grand nombre de lignes. Ainsi, un modèle plus complexe serait nécessaire pour mieux représenter la réalité physique du système, ce qui a pu être fait dans certains travaux [47]. En revanche, cela impliquerait l'utilisation d'un simulateur capable de gérer des calculs nettement plus lourds.

II.5.5 Conclusion

Les observations menées mettent en évidence les limites actuelles des simulations, soulignant la nécessité d'affiner les modèles ou d'optimiser les algorithmes afin d'améliorer la représentativité des résultats vis-à-vis des mesures expérimentales. Cependant, il est important de noter qu'une tendance cohérente a pu être reproduite à partir des simulations, confirmant la validité des approches utilisées. Ainsi, l'approche étudiée, permettant l'intégration des incertitudes dans les données simulées, s'aligne avec les orientations actuelles de la recherche en métrologie dans la modélisation de la rugosité et de la variabilité inter-ligne, ce qui renforce la pertinence et le potentiel de cette stratégie pour des applications futures.

II.6 Simulations SAXS

II.6.1 Principe de fonctionnement du simulateur

La technique métrologique SAXS (Small-Angle X-ray Scattering) est une méthode relativement récente, dont l'utilisation dans le domaine de la métrologie dimensionnelle s'est développée grâce à des avancées technologiques significatives. Au CEA-Leti, un simulateur a été spécialement conçu pour traiter les données brutes expérimentales à l'aide d'un algorithme de résolution de problème inverse. Ce simulateur est un outil essentiel pour relier les données obtenues par SAXS aux caractéristiques dimensionnelles des échantillons étudiés.

Contrairement à des mesures expérimentales SAXS où l'on obtient le facteur de structure modulé par le facteur de forme. Ainsi, le facteur de forme (en bleu sur la Figure II.33) correspond à l'enveloppe du facteur de structure (en rouge sur Figure II.33). Il est important de noter que le simulateur utilisé ne peut produire que le facteur de forme. Cette distinction est importante car elle met en évidence la différence entre les informations obtenues expérimentalement et celles générées par simulation.

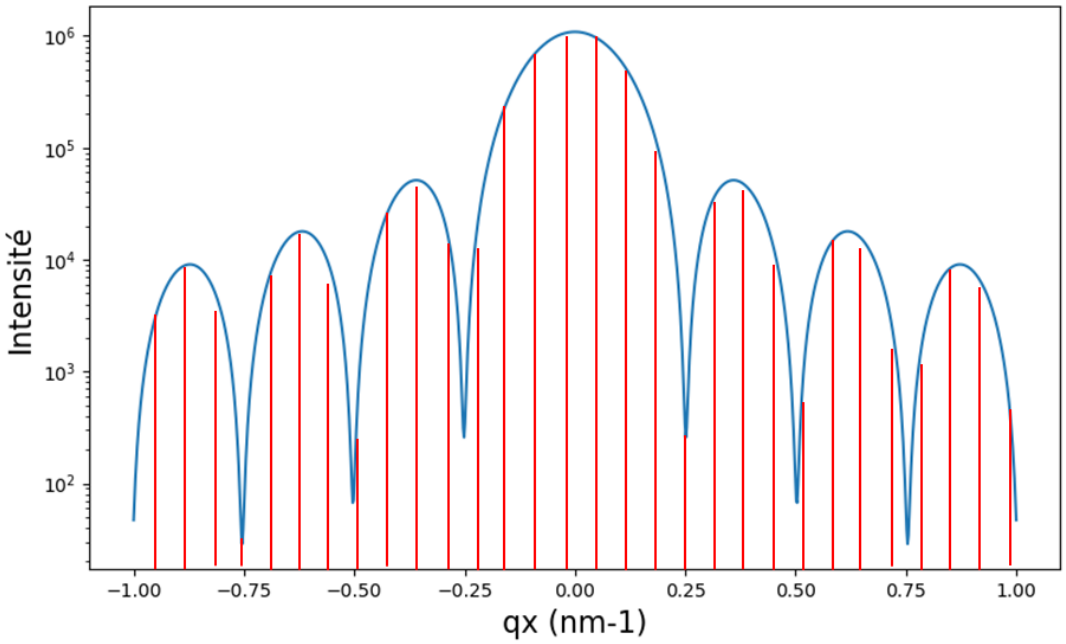

Figure II.33 : Représentation du facteur de structure (en rouge) obtenue lors de mesure expérimentale et modulé par le facteur de forme (en bleu) que l'on obtient par simulation.

Le simulateur développé au CEA-Leti adopte une approche simplifiée. Contrairement à une approche MMFE, il s'appuie sur des transformées de Fourier sous forme analytique pour produire les modulations liées au facteur de forme. Cette méthode présente plusieurs avantages significatifs :

- Réduction du coût computationnel : Les transformées de Fourier analytiques demandent moins de ressources de calcul.
- Capacité à gérer des structures complexes : L'approche par transformée de Fourier permet de modéliser des formes géométriques plus variées, comme des structures trapézoïdales, sans compromettre l'exactitude des simulations. Cela rend le simulateur plus flexible pour représenter des modèles réalistes des échantillons.

En réduisant la complexité des calculs, les simulations peuvent être effectuées à grande échelle tout en conservant une bonne fidélité dans la représentation des structures complexes des échantillons. Ainsi, nous pourrons générer les espaces réciproques à partir d'un ensemble de simulations de facteur de forme pour un coût computationnel réduit. Ceux-ci sont fondamentaux pour extraire des informations précises sur les paramètres dimensionnels des structures étudiées.

II.6.2 Etudes pour le paramétrage du modèle

Afin de garantir l'exactitude des simulations, il est indispensable de mener des études similaires à celles réalisées pour les modélisations OCD. Ces études visent à évaluer :

- L'impact des corners rounding : La prise en compte ou l'exclusion des arrondis aux coins peut influencer significativement les résultats.
- Le nombre de trapèze nécessaires : La partie centrale des lignes étant un trapèze, seules les parties représentant les corners rounding ont besoins d'un découpage en multiples trapèzes comme illustré par la Figure II.34.

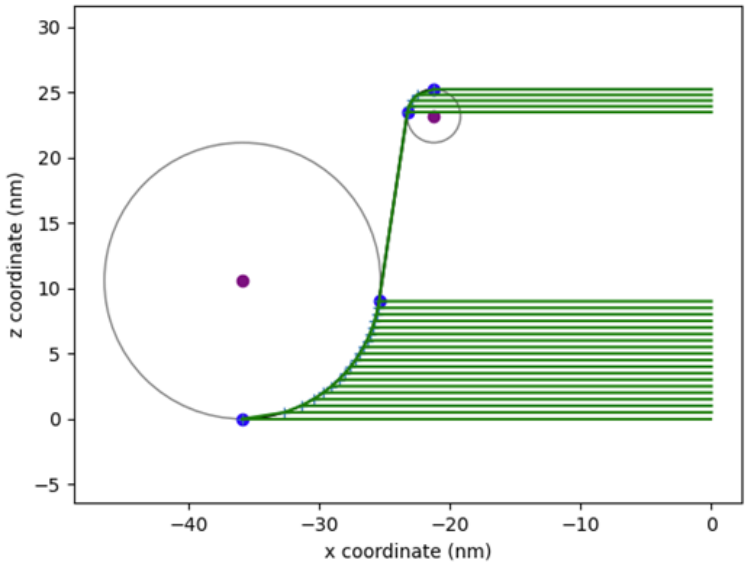

Figure II.34 : Décomposition en éléments trapézoïdale d'une ligne avec corners rounding

Bien que le simulateur développé par le CEA-Leti soit plus rapide qu'un simulateur basé sur MMFE, il demeure essentiel de réduire davantage les temps de calcul afin de permettre la génération et l'extraction des paramètres dimensionnels sur un large volume de données dans un délai raisonnable. Ainsi, ces analyses permettront de déterminer la segmentation optimale en trapèzes des modèles géométriques pour maintenir un bon équilibre entre exactitude des résultats et coût computationnel. Il est alors essentiel d'établir un critère permettant de déterminer si l'erreur calculée est suffisamment faible pour ne pas affecter de manière significative les résultats. Nous avons choisi d'utiliser la fonction d'erreur employée par l'algorithme de résolution de problème inverse. Cette fonction repose sur l'évaluation des facteurs de forme extraits des espaces réciproques mesurés, décrite par l'équation suivante :

$$Erreur = \sum_{N=1}^{10} \frac{|\log_{10}(Data(N)) - \log_{10}(Model(N))|}{nb(N)} \qquad \textbf{(II.16)}$$

N est le numéro de l'ordre de diffractions
nb(N) est le nombre de points de l'ordre N
Data(N) est le vecteur des points de l'échantillon à l'ordre N
Model(N) est le vecteur des points du modèle trouvé par résolution de problème inverse à l'ordre N

Cette méthode s'appuie sur les informations obtenues à partir des 10 premiers ordres de diffraction présents sur l'espace réciproque, présenté par la Figure II.35. Ces ordres de diffraction fournissent des données essentielles sur la distribution spatiale des intensités de diffusion, qui sont directement liées à la géométrie et aux dimensions des structures étudiées.

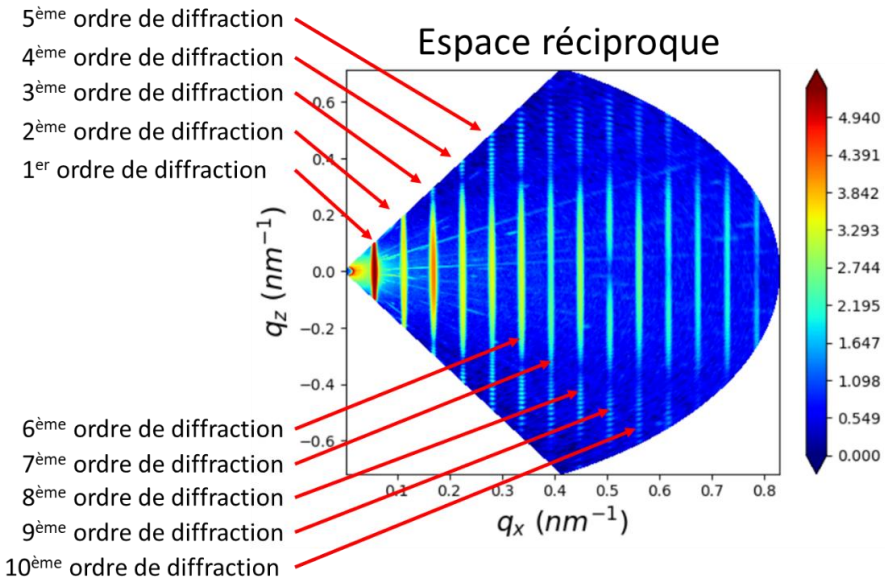

Figure II.35 : Espace réciproque d'une mesure expérimentale SAXS

Afin d'obtenir une représentation centrée sur les facteurs de forme des 10 premiers ordres de diffraction, une extraction de celles-ci à partir de l'espace réciproque est nécessaire. Ainsi, nous obtenons la représentation suivante (Figure II.36) en découpant l'espace réciproque en dix plans verticaux à q_x fixé. Par souci de lisibilité, nous effectuons artificiellement un décalage de leur intensité de 10^{-2*N}. L'extraction de ces facteurs de forme est aussi nécessaire pour l'algorithme de résolution de problème inverse. En effet, il compare les vecteurs des points de chaque ordre avec ceux modélisés.

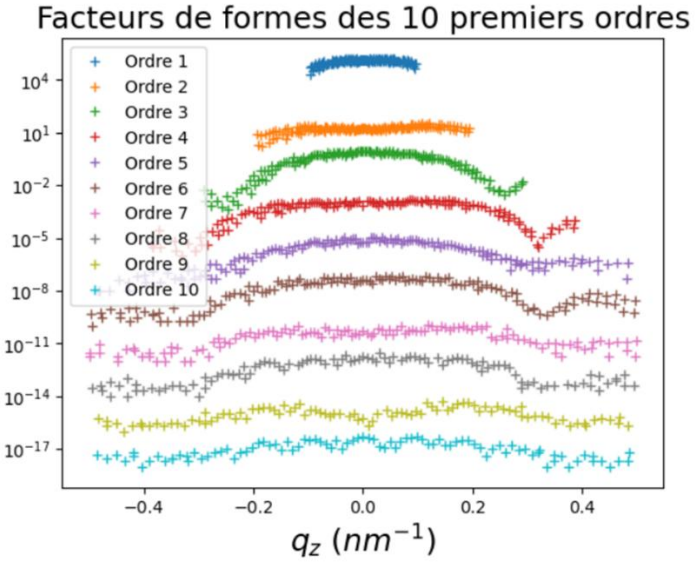

Figure II.36 : Représentation des 10 premiers ordres de diffraction extraits de l'espace réciproque de la Figure II.354.

Un grand nombre de simulations ont été réalisées afin de mieux comprendre le fonctionnement global du simulateur. À partir des extractions de paramètres dimensionnels obtenues par résolution de problèmes inverses sur les données simulées, nous avons pu identifier un seuil minimal de l'erreur (eq II.19), fixé à 0,49. Ce seuil permet de déterminer si un modèle ou une configuration donnée satisfait les exigences d'exactitude définies pour les simulations et les résolutions de problème inverse.

Pour choisir la complexité du modèle géométrique optimale, trois modèles différents ont été proposés :

- Le modèle simplifié n'incluant pas de corner rounding, illustré par la Figure II.37.A
- Le modèle intermédiaire incluant uniquement les corners rounding du bas, illustré par la Figure II.37.B
- Le modèle complet (qui sert de référence pour le calcul d'erreur) incluant les corners rounding du haut et du bas, illustré par la Figure II.37.C

Une étude a été menée pour évaluer l'erreur résultante associée à chaque modèle. Afin de garantir que les résultats soient uniquement influencés par les choix de modélisation et non par la discrétisation géométrique, le nombre de trapèzes représentant les corners rounding a été défini de manière à obtenir une hauteur de 0,01 nm par trapèze, correspondant à l'exactitude maximale visée. Ce niveau de discrétisation, bien que non validé par une étude de convergence formelle, a été considéré comme suffisamment fin pour minimiser l'impact de ce paramètre sur les résultats comparatifs entre modèles.

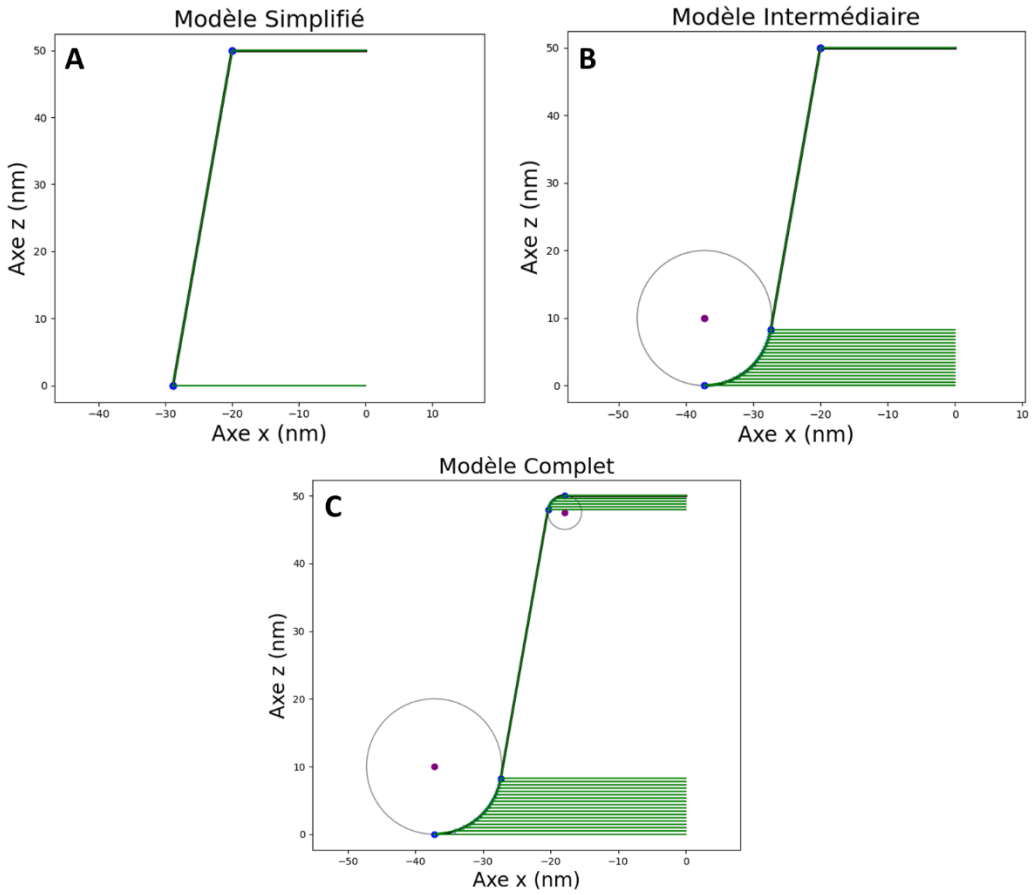

Figure II.37 : (A) Modèle simplifié décomposé en d'un seul trapèze d'une ligne de 50nm de haut. (B) Modèle intermédiaire décomposé en 17 trapèzes pour le corner rounding et en 1 trapèze pour le reste de la ligne. (C) Modèle complet (de référence) décomposé en 17 trapèzes et 5 trapèzes pour le bottom et top rounding respectivement et 1 trapèzes pour le reste de la ligne.

Nous avons simulé une série d'échantillons en faisant varier les paramètres dimensionnels dans leur plage de valeurs respectives, pour les trois modélisations considérées. Pour chaque échantillon, nous avons calculé l'erreur entre le modèle de référence (correspondant au modèle complet) et le modèle intermédiaire, ainsi qu'entre le modèle de référence et le modèle simplifié. L'analyse a été menée sur un total de 48 échantillons présentant des valeurs différentes de CD, H et SWA.

La Figure II.38.A illustré, sous forme de boîte à moustaches, les erreurs observées entre le modèle complet et le modèle simplifié. Ces erreurs sont largement supérieures au seuil de 0,49 (valeur maximale acceptable), atteignant jusqu'à 8,438. La Figure II.38.B présente quant à elle les erreurs entre le modèle complet et le modèle intermédiaire. Bien que les erreurs soient globalement plus faibles, elles dépassent néanmoins le seuil fixé, avec une valeur maximale de 0,866. Ces résultats montrent que ni le modèle intermédiaire ni le modèle simplifié ne permettent de respecter le critère de sélection défini, à savoir une erreur inférieure au seuil de 0,49.

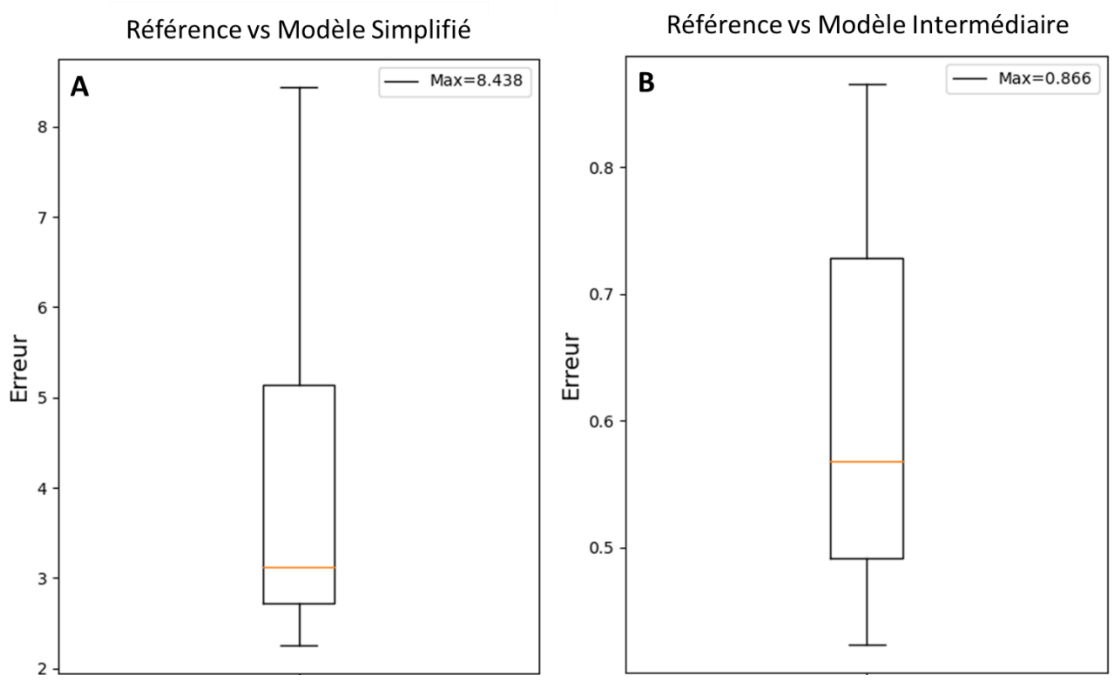

Figure II.38 : Boite à moustache des erreurs, (A) entre le modèle de référence et le modèle intermédiaire, (B) entre le modèle de référence et le modèle simplifié

Nous devons utiliser un modèle complet, en revanche nous ne connaissons pas le nombre minimum de trapèzes pour que cette modélisation respecte le critère d'erreur. Afin de réduire le coût computationnel tout en respectant le critère d'erreur établi (0,49), une étude comparative a été menée entre un modèle de référence, basé sur une modélisation complète ayant des hauteurs de trapèze extrêmement faibles (0,01 nm) et des modèles complets avec diverses hauteurs de trapèze supérieures à 0.01 nm. Pour chacune de ces modélisations, des simulations ont été réalisées pour 48 configurations de paramètres dimensionnels.

Les erreurs entre notre référence et les modèles ayant des hauteurs de trapèze, définissant les corners rounding, de 2 nm, 1,5 nm et 1 nm ont pu être calculées. Pour une hauteur de 2 nm, l'erreur maximale observée est de 0,9298, ce qui dépasse largement le seuil acceptable fixé (0,49). En revanche, des hauteurs de 1,5 nm et 1 nm donnent respectivement des erreurs maximales de 0,4378 et 0,2590, toutes deux inférieures au seuil. Ainsi, la modélisation optimale possède des hauteurs de trapèze aux corners rounding de 1,5 nm, illustré par la Figure II.39. Ce choix permet de réduire significativement le nombre de trapèzes nécessaires, typiquement à 7 pour la partie inférieure du motif et 2 pour la partie supérieure. Par conséquent, le temps de simulation. La modélisation adoptée dans la suite repose donc sur des trapèzes de 1,5 nm de hauteur pour les *corners rounding*, comme illustré par la Figure II.39.

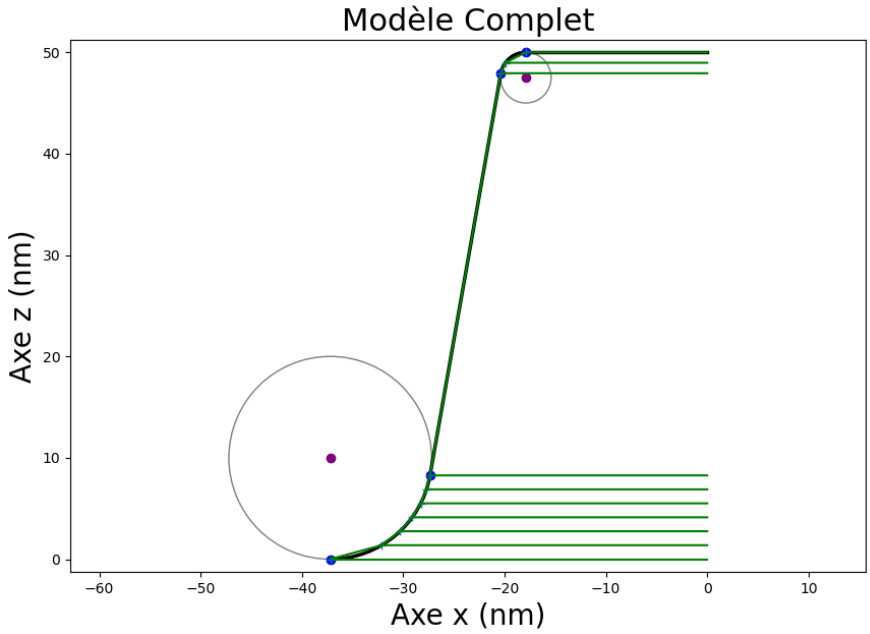

Figure II.39 : Modèle complet d'une ligne de 50 nm dont la hauteur des trapèzes des corners rounding est de 1,5 nm (6 trapèzes en bas et 2 en haut)

Par ces études, une modélisation optimisée a pu être définie, avec une décomposition en trapèzes adaptée aux différentes zones du motif. Les zones de corners rounding sont modélisées par plusieurs trapèzes ayant une hauteur de 1,5 nm, tandis que la zone centrale est représentée par un unique trapèze. Cette approche équilibre la simplicité du modèle et l'exactitude, en assurant que les erreurs restent négligeables vis-à-vis des résultats finaux.

II.6.3 Implémentation des incertitudes de mesure

Une similarité peut être observée entre les techniques métrologiques OCD et SAXS, notamment à travers l'adoption d'une approche multi-lignes dans les simulations SAXS. Cependant, en raison des spécificités du simulateur SAXS, qui diffère d'une approche de type MMFE, la modélisation ne repose pas sur trois lignes juxtaposées. À la place, les signatures individuelles sont superposées numériquement, en ne conservant qu'un tiers du signal de chaque ligne.

Or, les valeurs des paramètres dimensionnels extraites des simulations par résolution de problème inverse correspondent aux valeurs moyennes des lignes modélisées, qui constituent les valeurs de référence. Ainsi, cette modélisation n'introduit que très peu d'erreur en elle-même. La Figure II.40 illustre ce phénomène pour le paramètre hauteur H des lignes, nous y voyons une concentration d'échantillons n'ayant pas d'erreur. Nous avons le même comportement pour les

paramètres CD et SWA des lignes. Nous en concluons que ce n'est pas la variabilité inter-ligne des échantillons qui limite l'extraction des paramètres dimensionnels.

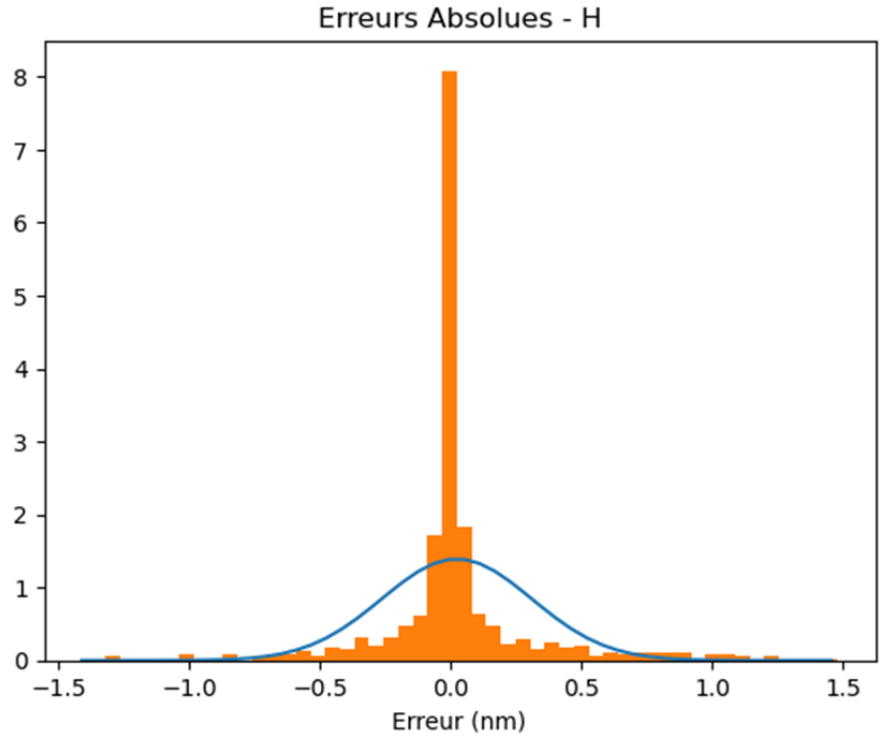

Figure II.40 : (Orange) Histogramme des erreurs de H des mesures des simulations SAXS par rapport aux valeurs de références. (Bleu) Gaussienne estimé à partir de l'écart-type et de la moyenne de la distribution des erreurs selon H (cette distribution n'étant pas gaussienne)

Une autre approche dans la modélisation du bruit de mesure a été étudiée, centrée sur les spécificités de la technique SAXS. Contrairement aux simulations idéales, les données expérimentales présentent des incertitudes à la fois sur la position et l'intensité des photons enregistrés. Pour pallier cela, une **opération de projection et d'interpolation pondérée** est utilisée. Elle consiste à regrouper les photons localisés autour d'une direction de diffraction donnée (définie par un intervalle de q_x) et à en extraire un **profil moyen**, en prenant en compte leur position et leur intensité relative, comme l'illustre la Figure II.41. Cette opération permet de reconstituer un facteur de forme plus lisse en atténuant les effets du bruit de mesure. L'approche permet de synthétiser le signal dans une forme exploitable pour les traitements inverses, prenant ainsi en compte les limitations instrumentales réelles.

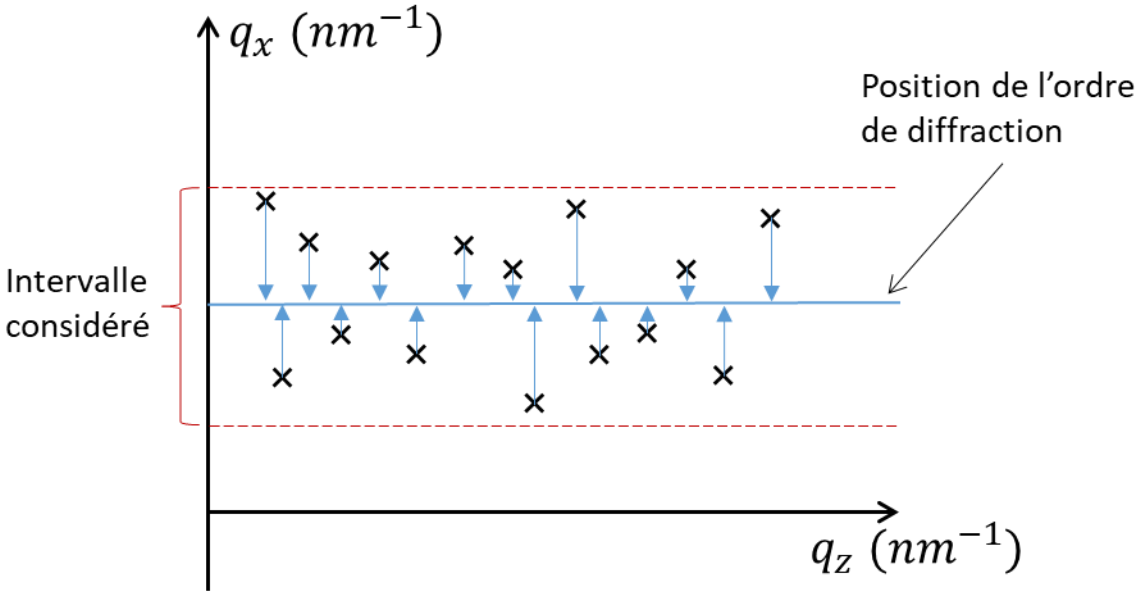

Figure II.41 : Projection et interpolation des photons détectés dans la fenêtre centrée sur l'ordre de diffraction selon l'axe q_z

Afin de générer des données nécessitant un tel traitement, nous avons mis en place un processus permettant l'acquisition des ordres de diffraction qui comporte plusieurs paramètres à prendre en compte. Nous en avons considéré trois :

En premier lieu, la largeur des intervalles de projection et d'interpolation des différents ordres de diffraction est à prendre en compte. Le choix de ces paramètres est effectué par un opérateur. Une bande trop large présentera des signaux plus bruités. Inversement, un signal ayant une bande trop fine, aura bien moins de points, représentant plus difficilement le facteur de forme.

Ensuite, les incertitudes associées au nombre de photons reçus à une position donnée induisent un bruit spécifique qui doit également être introduit. Nous nous sommes basés sur des travaux sur l'évaluation statistique des erreurs en SAXS [68]. L'écart-type de ce bruit est défini comme étant la racine carrée de l'intensité du point à bruiter, ce qui reflète les variations statistiques naturelles observées dans les mesures expérimentales.

Enfin, un autre aspect crucial dans la modélisation des incertitudes SAXS concerne la durée de la mesure. Le simulateur, en l'état, ne prend pas en compte ce paramètre et génère un grand nombre de points avec des intensités souvent supérieures à celles observées expérimentalement. Pour remédier à cela, les intensités ont été ajustées afin de correspondre à celles issues des mesures expérimentales du projet MadeIn4, deux ajustements ont été introduits :

- Les intensités ont été réajustées pour reproduire une intensité moyenne cohérente avec les mesures, puis arrondies à l'entier le plus proche afin de refléter le caractère discret du comptage photonique.
- Le nombre de points a été réduit afin d'obtenir une densité comparable à celle des jeux de données expérimentaux. Cela a été réalisé via une méthode d'agrégation spatiale locale : l'espace réciproque est discrétisé en zones, et pour chacune, un point unique est généré à partir du barycentre des photons présents, avec une intensité moyenne arrondie à l'entier.

Ces opérations permettent de réduire la complexité des données tout en maintenant une représentation fidèle du signal, exploitable pour les traitements inverses.

Après l'implémentation de ces ajustements, incluant les différents bruits et la réduction du nombre de points, Les simulations obtenues se rapprochent davantage des données expérimentales. En effet, nous pouvons observer sur la Figure II.42.A les facteurs de formes expérimentaux et sur la Figure II.42.B les facteurs de formes par simulations. Ainsi, nous avons pu, par différents procédés, implémenter certaines sources d'erreurs de mesure propre au SAXS. Une différence peut être observée dans la répartition des points entre les ordres de diffraction. Contrairement aux données expérimentales, où les ordres faibles présentent une densité de points plus élevée que les ordres plus élevés, la simulation génère une répartition relativement uniforme. Cette différence provient de la manière dont les données sont acquises expérimentalement et générées en simulation.

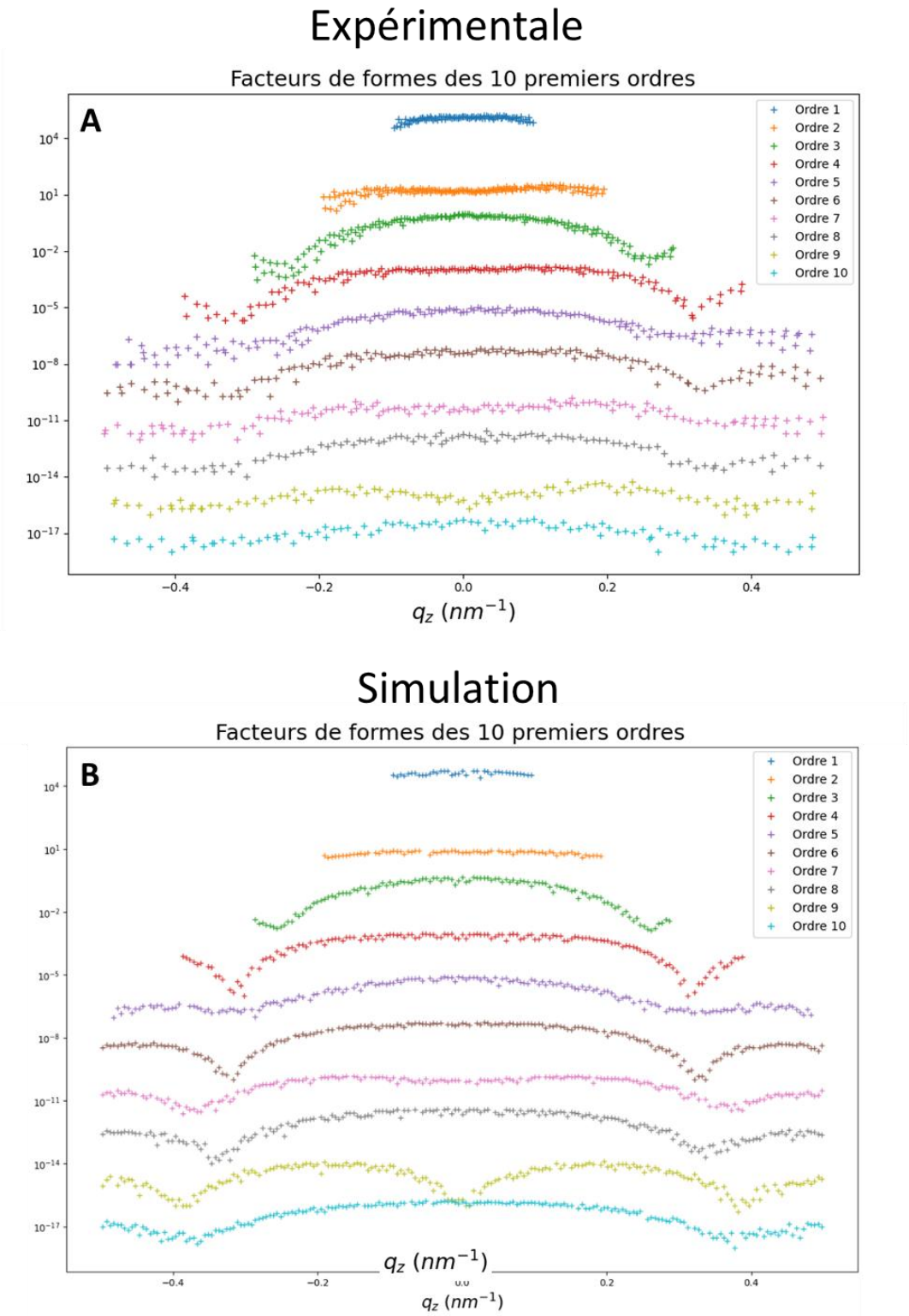

Figure II.42 : Facteur de forme des 10 premiers ordres de mesure expérimentale (A) et de mesure simulé (B)

II.6.4 Analyse des paramètres dimensionnels extraits des simulations SAXS

Les techniques développées afin d'implémenter des incertitudes de mesures réalistes dans les simulations ont été appliquées aux 700 échantillons simulés Ensuite, nous avons pu extraire les paramètres dimensionnels des échantillons par résolution de problème inverse. Les erreurs résultantes ont alors été calculées par rapport aux valeurs des paramètres dimensionnels de références.

Un aspect notable des résultats est la présence d'un grand nombre de valeurs aberrantes (outliers). Ces valeurs peuvent être identifiées en appliquant la méthode développée par John Tukey [69] à l'aide des équations suivantes :

$$IQR = Q3 - Q1 \qquad \textbf{(II.17)}$$

$$Outlier\ si \begin{cases} Valeur\ < Q1 - 1.5 \times IQR \\ Valeur\ > Q3 + 1.5 \times IQR \end{cases} \qquad \textbf{(II.18)}$$

Q1 est le 1er quartile de la suite des valeurs
Q3 est le 3ème quartile de la suite des valeurs
IQR est écart interquartile

Parmi les 700 ensembles de paramètres dimensionnels étudiés, un nombre significatif de valeurs aberrantes a été identifié. En détail, 101 valeurs de CD, 68 valeurs de H, et 113 valeurs de SWA ne respectent pas le critère de Tukey. Une fois ces outliers éliminés, les données restantes (soit entre 84% et 90% des données en fonction du paramètre dimensionnel) ont pu être analysées. De la même manière qu'avec les données des simulateurs précédent, nous commençons par comparer la distribution des erreurs avec une distribution normale. La Figure II.43.A met en évidence, par la présence d'une droite (avec un R^2 de 0,9833), l'aspect gaussien de la distribution des erreurs du paramètre CD (Figure II.43.B). Nous observons les mêmes comportements avec les paramètres H (avec un R^2 de 0,9948) et SWA (avec un R^2 de 0,9839). Ainsi nous pouvons estimer les valeurs des gaussienne associées aux distributions obtenues, comme l'illustre la Figure II.44.

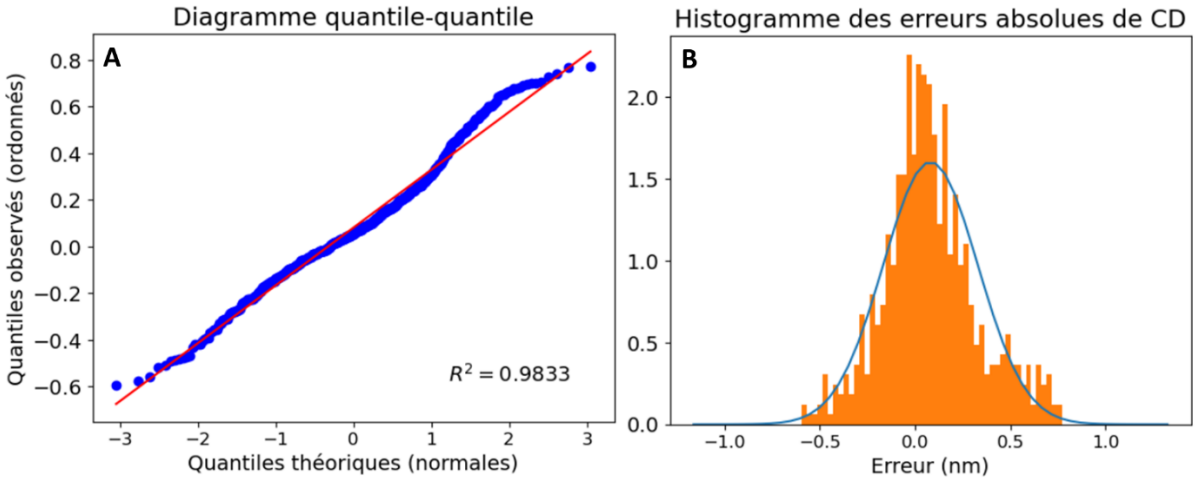

Figure II.43 : (A) Diagramme quantile-quantile des erreurs absolues de CD. (B) Histogramme des erreurs absolues de CD (orange) et distribution normale associé (bleu).

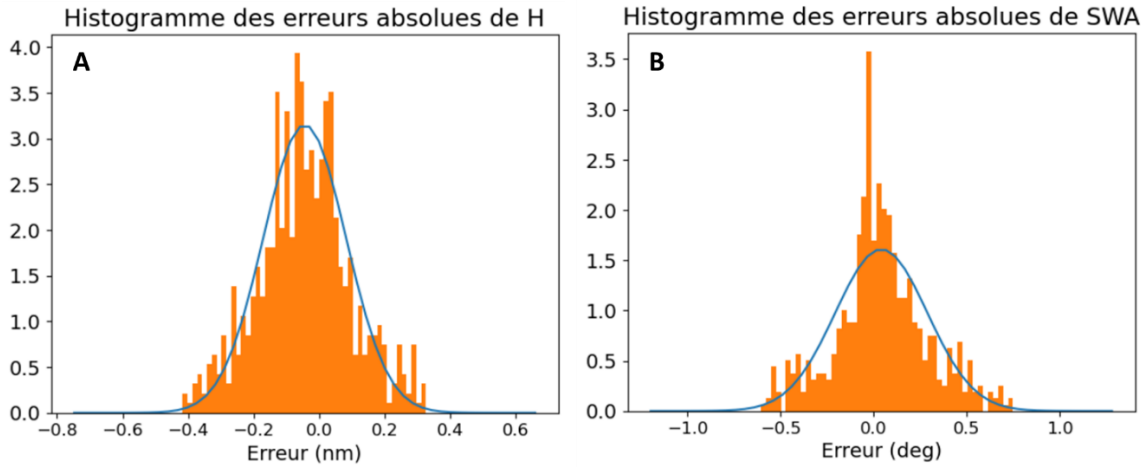

Figure II.44 : (A) Histogramme des erreurs absolues de H (orange) et distribution normale associé (bleu). (B) Histogramme des erreurs absolues de SWA (orange) et distribution normale associé (bleu).

Au vu du nombre important de données aberrantes, nous avons décidé de les analyser. Ainsi, à l'instar des données précédentes, les distributions des erreurs de ces valeurs aberrantes ont été comparées à des distributions normales via des diagrammes quantile-quantile, ce qui est illustré par la Figure II.45.A, la Figure II.46.A et la Figure II.47.A. Ces représentations révèlent une répartition gaussienne des valeurs aberrantes, nous permettant d'estimer les valeurs des paramètres gaussiens de ces distributions, illustrés par la Figure II.45.B, la Figure II.46.B et la Figure II.47.B. Nous en déduisons que les erreurs extraites des simulations SAXS présentent une distribution qui pourrait être décrite comme une double distribution gaussienne. Toutefois, aucune explication claire n'a pu être déterminée pour justifier ce comportement particulier.

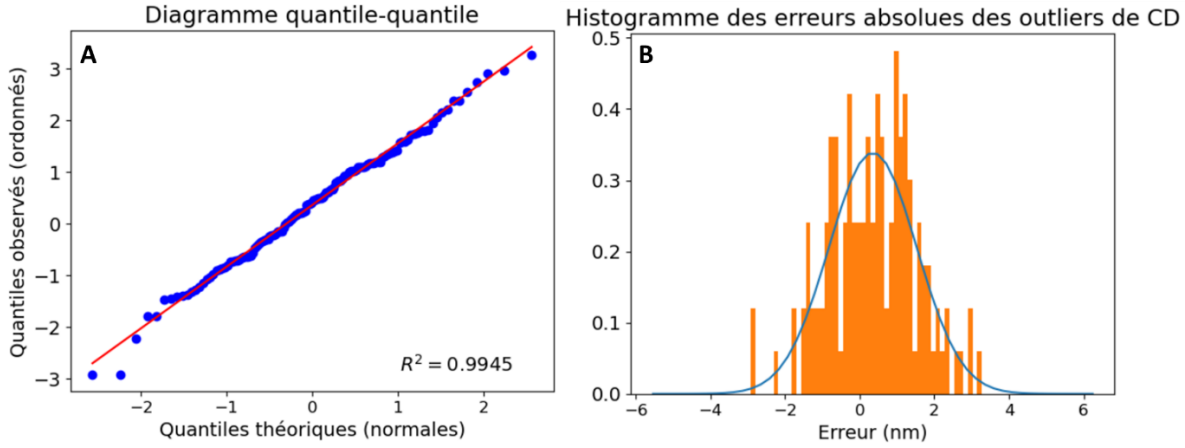

Figure II.45 : (A) Diagramme quantile-quantile des erreurs absolues des outliers de CD. (B) Histogramme des erreurs absolues de CD (orange) et distribution normale associée (bleu).

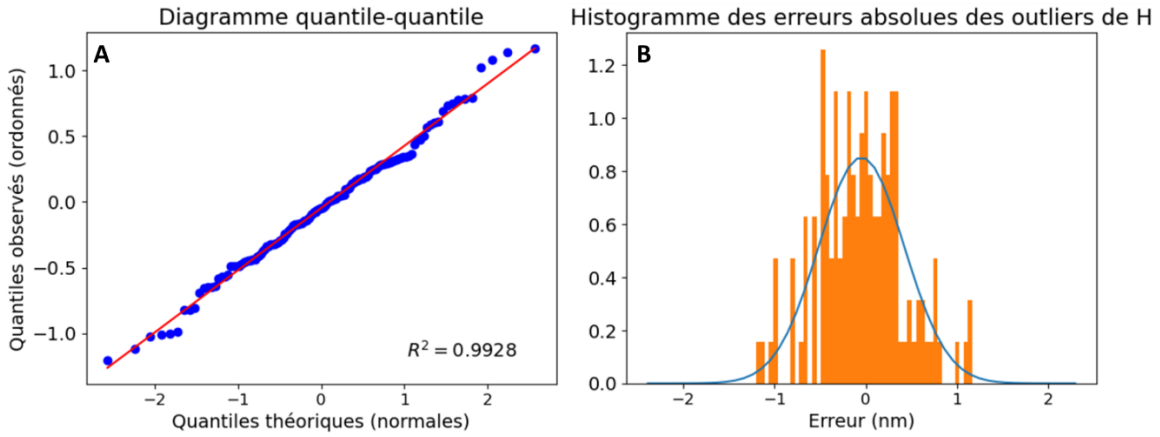

Figure II.46 : (A) Diagramme quantile-quantile des erreurs absolues des outliers de H. (B) Histogramme des erreurs absolues de H (orange) et distribution normale associée (bleu).

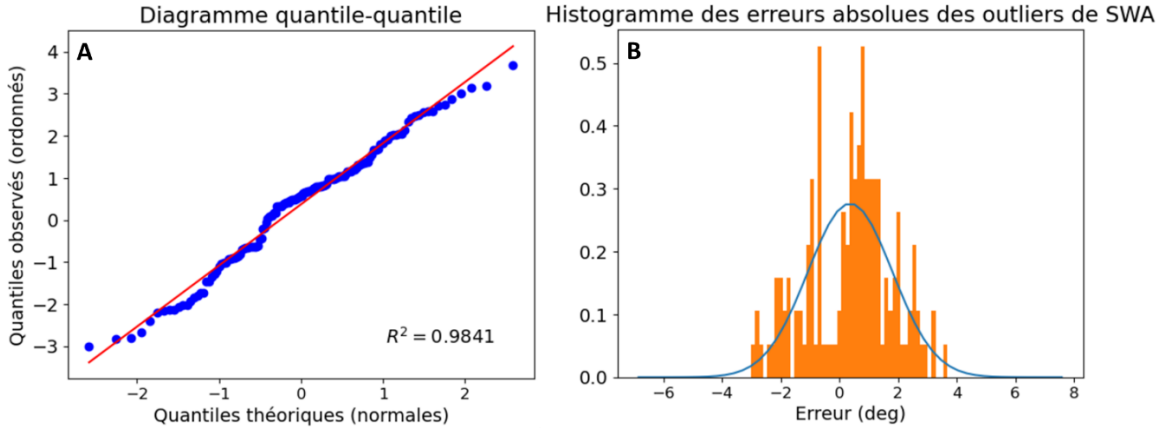

Figure II.47 : (A) Diagramme quantile-quantile des erreurs absolues des outliers de SWA. (B) Histogramme des erreurs absolues de SWA (orange) et distribution normale associée (bleu).

La distribution des valeurs aberrantes, ou outliers, s'apparente à une distribution normale, ce qui permet d'estimer les paramètres qui les définissent. Ainsi, les résultats obtenus ont été analysés et les paramètres estimés sont résumés dans le Tableau II.6 . Ce tableau présente deux types de gaussiennes :

- Les gaussiennes dites "fines" : elles correspondent à environ 84% à 90% des échantillons. Ces gaussiennes caractérisent les données principales, considérées comme non aberrantes.
- Les gaussiennes dites "larges" : elles représentent environ 16% à 20% des échantillons et concernent les données associées aux outliers, c'est-à-dire les paramètres dimensionnels qui s'écartent des critères standards.

Cette distinction entre les deux types de gaussiennes offre une vision plus détaillée des variations et des erreurs dans les simulations, tout en soulignant la proportion significative d'échantillons présentant des valeurs atypiques.

Les erreurs issues des simulations SAXS ont été analysées en distinguant les distributions gaussiennes fines (valeurs non aberrantes) et larges (valeurs aberrantes), telles que présentées dans le Tableau II.8. Pour le paramètre CD, l'erreur moyenne associée aux valeurs non aberrantes est de $\mu = 0{,}08$ nm, avec un écart-type de $\sigma = 0{,}25$ nm. En tenant compte des valeurs aberrantes, l'écart-type atteint 1,21 nm, pour une moyenne de 0,33 nm. L'écart-type global des erreurs sur le CD est ainsi de 0,54 nm.

Concernant la hauteur H, les erreurs sur les valeurs non aberrantes sont de $\mu = -0{,}05$ nm et $\sigma = 0{,}15$ nm, tandis que les valeurs aberrantes présentent un écart-type de 0,49 nm. La valeur de l'écart-type global de H, pour les valeurs aberrantes et non-aberrantes, est de 0,26 nm. Le SWA présente quant à lui une erreur moyenne de 0,05° et un écart-type de 0,27° pour les valeurs non aberrantes, contre un écart-type de 1,44° pour les valeurs aberrantes. Le nombre de valeurs aberrantes est particulièrement élevé, en particulier pour le SWA (113 occurrences), ce qui se traduit par un écart-type global de 0,67°.

La dispersion observée peut être attribuée à la complexité de la résolution inverse mise en œuvre dans les simulations, qui peut favoriser la génération de solutions extrêmes en réponse à de faibles variations du signal.

	Distribution gaussienne fine (valeurs non-aberrantes)		Distribution gaussienne large (valeurs aberrantes)		Nombre de valeur aberrante	Ecarts-types globaux des erreurs
Distribution normale de l'erreur	**μ** (moyenne)	**σ** (écart-type)	**μ** (moyenne)	**σ** (écart-type)		
ΔCD	0.08nm	0.25nm	0.33nm	1.21nm	101	0,54 nm
H	-0.05nm	0.15nm	-0.07nm	0.49nm	68	0,26 nm
ΔSWA	0.05°	0.27°	0.37°	1.44°	113	0,67°

Tableau II.6 : Estimation des distributions gaussiennes des erreur des simulations SAXS

II.6.5 Conclusion

Les simulations SAXS réalisées ont permis la génération de figures de diffraction, représentées dans l'espace réciproque. Des sources d'incertitude ont été introduites, notamment par l'ajout de bruit gaussien sur l'intensité et la position des photons reçus. Bien qu'un modèle multi-lignes ait été implémenté afin de prendre en compte la variabilité inter-ligne, son impact sur les résultats s'est révélé limité. Néanmoins, les données obtenues intègrent des incertitudes de mesure, et présentent des caractéristiques proches de celles observées expérimentalement. Ces résultats soulignent la pertinence de la simulation SAXS dans une approche de métrologie avancée, notamment pour la génération massive de données réalistes.

II.7 Synthèse des travaux de simulations

II.7.1 Les résultats

Nous avons pu produire des données réalistes en utilisant des simulateurs à partir des mêmes modèles 3 lignes, enrichis par des études qui ont permis d'intégrer des incertitudes spécifiques à chaque technique métrologique. En analysant les erreurs entre les paramètres dimensionnels (CD, H et SWA) obtenus par extractions à partir de simulations et les valeurs attendues, nous avons également pu estimer les incertitudes de mesure propres à chaque technique simulée. Malgré certaines limites de nos approches, ces travaux ont permis de générer des incertitudes caractéristiques pour chaque méthode, aboutissant ainsi aux résultats synthétisés dans le Tableau II.7.

Nous pouvons noter que les ordres de grandeur des erreurs issues de la littérature (présentés dans le chapitre 1) sont globalement respectés par nos simulations. Toutefois, nous nous attendions à obtenir des erreurs plus faibles pour l'OCD, en

particulier sur les paramètres CD et SWA, ainsi qu'à des erreurs légèrement plus élevées avec l'AFM. Par ailleurs, un nombre important de valeurs aberrantes a été observé dans les simulations SAXS. Malgré ces écarts, notre approche se montre prometteuse pour la génération de données synthétiques réalistes.

		CD	H	SWA
AFM	µ (moyenne)	-0,05 nm	1,19 nm	-1,29 °
	σ (écart-type)	**0,65 nm**	**0,68 nm**	**1,29 °**
CDSEM	µ (moyenne)	1,40 nm		
	σ (écart-type)	**0,91 nm**		
OCD	µ (moyenne)	-0,41 nm	0,13 nm	-1,32 °
	σ (écart-type)	**2,05 nm**	**0,92 nm**	**3,49 °**
SAXS	moyenne	0,12 nm	-0,05 nm	0,08 °
	Ecart-type	**0,54 nm**	**0,26 nm**	**0,67 °**

Tableau II.7 : Estimation des incertitudes générer pour chaque simulateur de technique métrologique

II.7.2 Les corrélations des erreurs

Le principal objectif de ces démarches est la simulation de données réalistes. En cela, elles doivent avoir des incertitudes qui reflètent le plus fidèlement possible ce qui peut être observé expérimentalement. Ainsi, contrairement à des données générées de façon mathématique, nous trouvons des corrélations entre les erreurs des paramètres dimensionnels extraites à partir d'une même technique (Figure II.48). Leur présence démontre que l'on peut générer des données ayant des incertitudes spécifiques propre à chacune des techniques métrologiques.

Nous observons des corrélations modérément fortes entre le CD et le SWA pour l'AFM (Figure II.48.A), ainsi qu'entre la hauteur (H) et le SWA pour le SAXS (Figure II.48.C). En revanche, la majorité des autres paramètres dimensionnels présentent de faibles corrélations, toutes techniques confondues. Notons également que H et SWA ne sont pas corrélés dans le cas de l'AFM.

La présence de ces corrélations nous seront particulièrement utiles lors de l'hybridation des données par réseau de neurones. En effet, au-delà des distributions des erreurs, les relations entre les erreurs des différents paramètres dimensionnels sont une source d'information supplémentaire que le modèle pourra exploiter pour affiner ses prédictions. Il ne faut pas non plus négliger le fait que l'absence (ou la faible présence) de corrélation pour certains paramètres, combinée aux corrélations existantes d'une autre technique métrologique, constitue aussi une information exploitable par le réseau.

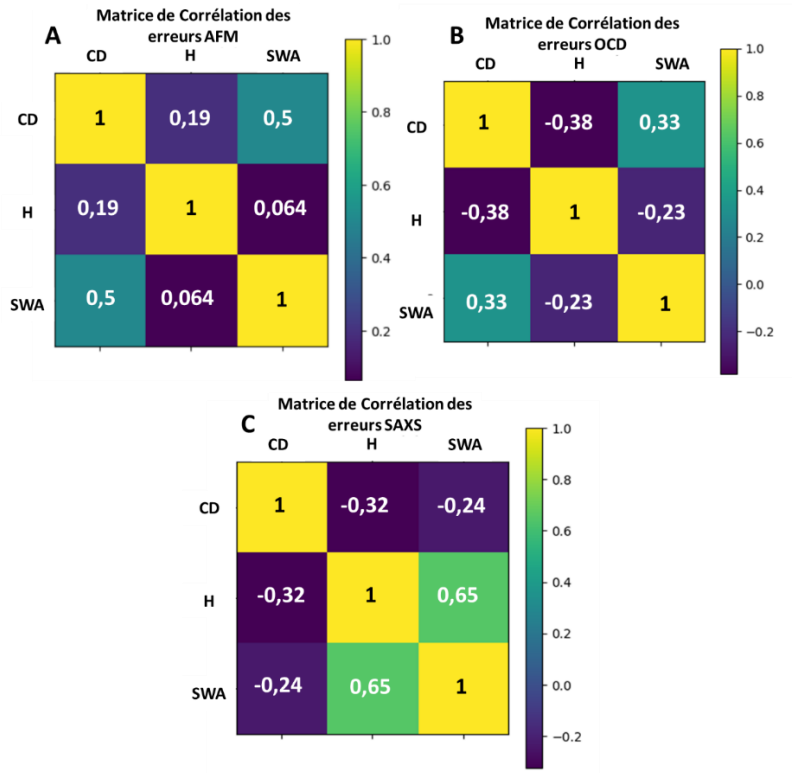

Figure II.48 : Matrices de corrélation entre les paramètres dimensionnels, (A) de l'AFM, (B) de l'OCD, (C) du SAXS.

II.8 Conclusions

Dans ce chapitre, nous avons étudié différentes approches pour simuler des données réalistes pour les différentes techniques de métrologies considérées. Nous avons intégré la variabilité entre les lignes d'un même réseau, propre aux échantillons étudiés, en proposant une approche multi-lignes du modèle. Nous avons également intégré les sources erreurs propres à chaque technique métrologique, comme le bruit de Schottky pour le CDSEM. Ces développements ont permis de nous rapprocher des comportements expérimentaux et d'approfondir notre compréhension des mécanismes responsables des erreurs de mesure. Ces avancées ouvrent la voie à des recherches futures pour affiner encore davantage les modélisations des techniques métrologiques.

Les travaux menés ont abouti à la génération de deux types de jeux de données complémentaires. Les données brutes, constituées des images et signaux directement produits par les simulateurs, et les données numériques, regroupant les paramètres dimensionnels (CD, H et SWA) obtenus après traitement des données brutes. Ces deux jeux de données représentent une ressource essentielle pour la suite de nos travaux.

En conséquence, nous avons atteint l'objectif principal de ce chapitre, à savoir la génération massive de données réalistes. Ces données serviront de base aux travaux du chapitre suivant, où nous évaluerons les performances de deux approches distinctes : l'hybridation basée sur les données numériques et l'hybridation à partir des données brutes. Par cette réussite, nous posons ainsi les fondations pour la mise en œuvre d'un système de métrologie hybride par réseau de neurones.

Chapitre III - Métrologie hybride par réseau de neurones

L'intelligence artificielle (IA) désigne l'ensemble des méthodes et approches qui visent à doter les systèmes informatiques de capacités d'analyse, de décision ou d'adaptation inspirées, dans une certaine mesure, du raisonnement humain. Parmi ces approches, le machine learning (ou apprentissage automatique) occupe une place centrale. Son principe est de permettre à un modèle d'apprendre directement à partir de données, sans définir de règles explicites à l'avance. Le modèle améliore progressivement ses prédictions en s'entraînant sur de nombreux exemples, un peu à la manière d'un élève qui progresse en répétant des exercices. Le deep learning (ou apprentissage profond) correspond à une branche particulière du machine learning. Il se distingue par un nombre beaucoup plus élevé de paramètres internes que les méthodes classiques, ce qui lui confère une plus grande capacité à représenter des relations complexes et non linéaires. C'est cette profondeur (d'où le terme deep) qui rend ces modèles particulièrement adaptés à des problèmes riches en données et difficiles à modéliser analytiquement. Cette hiérarchie entre intelligence artificielle, machine learning et deep learning est représentée par la Figure III.1

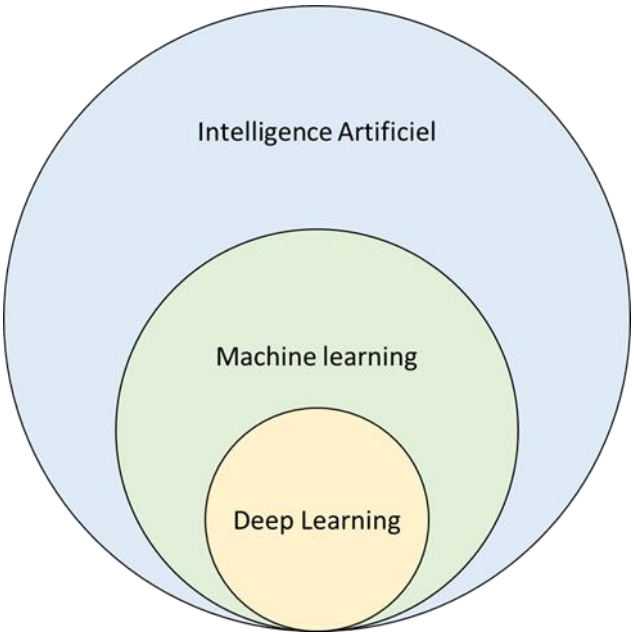

Figure III.1 : Représentation schématique des relations entre intelligence artificielle, machine learning et deep learning.

Dans le cadre de la métrologie hybride, le deep learning, et plus particulièrement les réseaux de neurones artificiels, offrent des outils flexibles pour exploiter la complémentarité de différentes techniques de mesure. Plutôt que de recourir à une modélisation analytique détaillée des phénomènes physiques, ils apprennent

directement à relier les données d'entrée (mesures issues de techniques) aux valeurs de référence.

L'objectif de ce chapitre est de présenter, de manière progressive, les concepts fondamentaux liés à l'apprentissage par réseaux de neurones. La première partie introduit les notions générales nécessaires pour comprendre leur fonctionnement et les principaux paramètres qui gouvernent l'apprentissage. Les parties suivantes porteront sur l'étude des architectures de réseaux adaptées à nos problématiques, l'analyse des résultats obtenus à partir de données numériques, puis des travaux préliminaires consacrés à l'hybridation de données brutes.

III.1 Fonctionnement des réseaux de neurones artificiels

Comme introduit précédemment, un réseau de neurones artificiel cherche à établir un lien entre des données d'entrée et une sortie attendue, en ajustant ses paramètres internes à partir d'exemples connus. Ce mécanisme, appelé apprentissage supervisé, repose sur un processus itératif : prédire, comparer à la référence, puis corriger. Afin de comprendre comment ce principe est mis en œuvre, il est nécessaire de détailler l'unité fondamentale qui compose ces réseaux : le neurone artificiel.

III.1.1 Le neurone artificiel

L'élément de base d'un réseau est le neurone artificiel, le plus basique est le perceptron. Il s'agit d'un modèle mathématique qui imite, de manière très simplifiée, le fonctionnement d'un neurone biologique : il reçoit plusieurs informations, les combine, puis produit une sortie.

Comme illustré par la Figure III.2, chaque donnée d'entrée ($Input_k$) est multipliée par un poids (ω_k) qui traduit l'importance que le neurone accorde à cette information. Un biais (ω_o) est ensuite ajouté, permettant de décaler le résultat indépendamment des entrées. L'ensemble est transformé par une fonction d'activation f, dont le rôle est d'introduire une non-linéarité. Cette étape est essentielle : sans fonction d'activation, le neurone ne pourrait représenter que des relations purement linéaires entre les données.

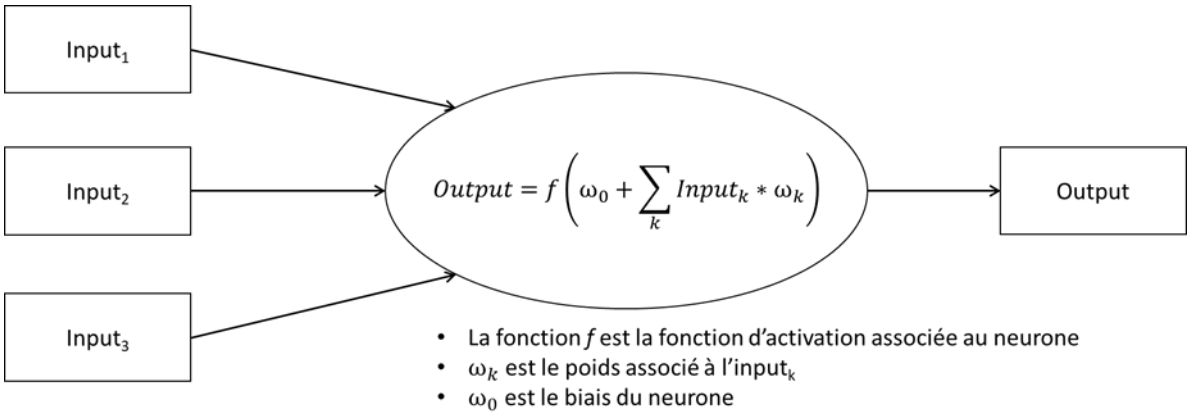

Figure III.2 : Schéma de principe du perceptron.

Il existe une grande variété de fonctions d'activation (par exemple la sigmoïde, la tangente hyperbolique ou encore ReLU), chacune présentant des avantages et des limites. La fonction ReLU (Rectified Linear Unit) est définie par :

$$f(x) = \max(0, x) \qquad \textit{(III.1)}$$

Cette fonction d'activation (Figure III.3) est aujourd'hui l'une des plus utilisées dans le domaine du deep learning. Elle doit sa popularité à sa simplicité de calcul, ainsi qu'à son efficacité dans l'entraînement de réseaux profonds. Contrairement aux fonctions sigmoïde ou tangente hyperbolique, qui saturent pour certaines valeurs et ralentissent la propagation des gradients, ReLU permet de favoriser une convergence plus rapide [70], [71].

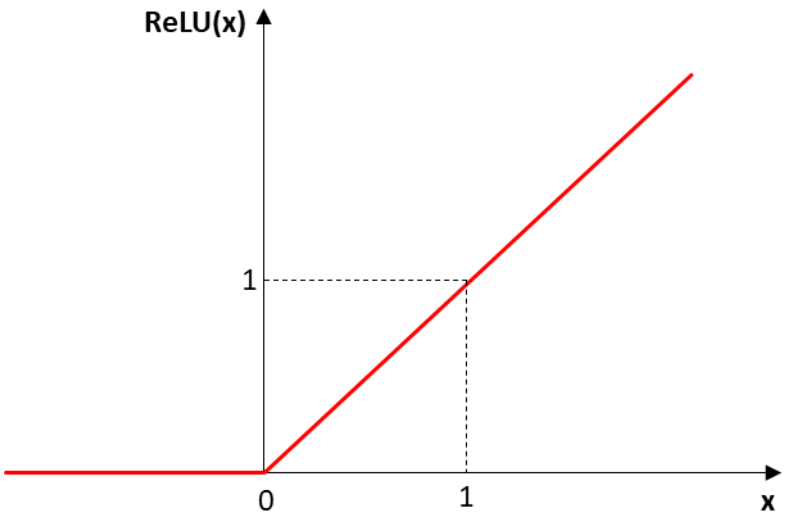

Figure III.3 : Représentation graphique de la fonction d'activation ReLU.

III.1.2 Organisation en réseau

La véritable puissance du deep learning provient de l'organisation de nombreux neurones en réseau, où la sortie de certains devient l'entrée des suivants. Cette structure s'appelle Perceptrons Multi-Couches (ou Multi-Layers Perceptrons, MLP, en anglais). Comme illustré par la Figure III.4, un réseau de neurones est généralement structuré en plusieurs parties :

- Une couche d'entrée, dont le nombre de neurones est ajusté au nombre de paramètre d'entrée.
- Une ou plusieurs couches intermédiaires, encore appelées couches cachées, où les neurones transforment progressivement l'information.
- Une couche de sortie, qui produit le résultat final, souvent une valeur numérique dans le cas d'un problème de régression.

La profondeur d'un réseau, c'est-à-dire le nombre de transformations successives qu'il applique, joue un rôle essentiel : plus elle est importante, plus le réseau dispose de paramètres internes et de capacité pour représenter des relations complexes. En revanche, cette augmentation du nombre de paramètres implique un besoin proportionnellement plus important en données d'apprentissage afin d'assurer une estimation fiable de ces paramètres. C'est précisément cette profondeur qui distingue le deep learning des approches plus classiques, et qui explique son association étroite avec la notion de big data.

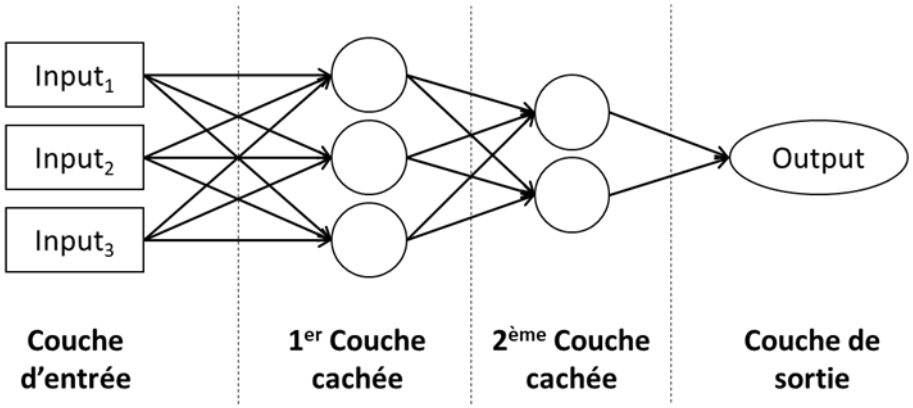

Figure III.4 : Schéma général de l'organisation d'un réseau de neurones artificiels.

III.1.3 Préparation des données

Avant d'entraîner un réseau de neurones, il est essentiel de préparer soigneusement les données qui seront utilisées. Le jeu de données considéré se compose de 10 paramètres d'entrée (CD_{afm}, H_{afm}, SWA_{afm}, CD_{cdsem}, CD_{ocd}, H_{ocd}, SWA_{ocd}, ΔCD_{saxs}, H_{saxs},

ΔSWA_{saxs}) et de 3 paramètres de sortie, correspondant aux valeurs de référence ou cibles d'apprentissage (CD_{ref}, H_{ref}, SWA_{ref}).

Une étape fondamentale est la normalisation, qui consiste à ramener les différentes grandeurs considérées dans des plages de valeurs comparables. Sans cette opération, une grandeur dont l'intervalle de variation est large exercerait une influence disproportionnée par rapport à d'autres.

Dans notre cas, les paramètres dimensionnels mesurés présentent des ordres de grandeur différents : la hauteur H varie de 20 à 100 nm, tandis que l'angle de paroi SWA est compris entre 80° et 90°. Sans normalisation, la variation plus importante de H aurait tendance à dominer le calcul et à réduire le poids relatif du SWA, ce qui fausserait l'apprentissage du réseau.

Les deux approches couramment utilisées sont :

- La standardisation (z-score) : Elle consiste à soustraire la moyenne μ et à diviser par l'écart-type σ, de sorte que les données soient centrées autour de zéro et présentent une variance unitaire :

$$x_{norm} = \frac{x - \mu(X)}{\sigma(X)} \qquad \textbf{(III.2)}$$

Où

- x est un individu de la population X
- μ(X) est la moyenne de la population X
- σ(X) est l'écart-type de la population X

- La normalisation par intervalle interquartile (IQR) : Elle repose sur la médiane m et l'écart interquartile IQR (IQR = Q_3 - Q_1), où Q_1 et Q_3 sont respectivement le premier et le troisième quartile :

$$x_{norm} = \frac{x - m(X)}{Q_3(X) - Q_1(X)} \qquad \textbf{(III.3)}$$

Où

- x est un individu de la population X
- m(X) est la médiane de la population X
- Q_1(X) et Q_3(X) sont les 1er et 3ème quartiles de la population X

Cette dernière est plus robuste que la standardisation vis-à-vis des valeurs extrêmes (outliers), car elle n'est pas influencée par des points atypiques susceptibles de fausser la moyenne ou l'écart-type. Dans le cadre de ces travaux, et en particulier pour les données issues du SAXS, la présence de valeurs aberrantes rend la normalisation par IQR plus appropriée.

La préparation des données ne se limite pas à leur normalisation : elle inclut également leur découpage en ensembles distincts, utilisés à différentes étapes de l'apprentissage :

- Le jeu d'entraînement (training set), utilisé pour ajuster les paramètres internes du réseau (poids et biais).
- Le jeu de validation (validation set), qui permet de suivre l'évolution de l'apprentissage et d'adapter certains réglages pendant son déroulement.
- Le jeu de test (test set), totalement séparé des précédents, qui sert à évaluer la capacité du modèle à conserver de bonnes performances sur des données nouvelles.

Ce découpage, associé à une normalisation appropriée, constitue une étape cruciale pour garantir un apprentissage fiable et une évaluation pertinente du modèle.

III.1.4 Fonction de perte

Dans le cadre de l'apprentissage des modèles de machine learning, il est nécessaire de définir une fonction permettant de quantifier l'écart entre les prédictions du modèle et les valeurs de référence. Cette fonction, appelée fonction de perte, guide l'ajustement des paramètres du réseau de neurones au cours de l'entraînement.

Dans ces travaux, la fonction de perte utilisée est l'erreur quadratique moyenne (Mean Squared Error, MSE), largement employée en régression. Elle est définie comme la moyenne des carrés des écarts entre les valeurs estimées $\hat{x}_i$ et les valeurs de référence x_i, selon l'expression suivante :

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(\hat{x}_i - x_i)^2 \qquad \textit{(III.4)}$$

Où N représente le nombre d'échantillons considérés.

Le MSE fournit une mesure globale de l'erreur en agrégeant l'ensemble des écarts en une unique valeur scalaire. La mise au carré des différences permet de pénaliser davantage les erreurs importantes, ce qui le rend particulièrement adapté aux problèmes où les grandes déviations doivent être évitées. Ainsi, il définit une fonction objectif aisément minimisable par descente de gradient lors de l'entraînement des réseaux de neurones.

III.1.5 Processus d'apprentissage

Une fois les données préparées, l'étape suivante consiste à entraîner le réseau de neurones, c'est-à-dire à ajuster ses paramètres internes (poids et biais) afin qu'il produise des sorties aussi proches que possible des valeurs de référence. Le mécanisme repose sur trois étapes principales :

- Propagation avant (forward pass) : les données d'entrée traversent successivement les différentes couches du réseau jusqu'à produire une prédiction en sortie.
- Évaluation de l'erreur : la prédiction est comparée à la valeur attendue au moyen d'une fonction de perte, qui mesure quantitativement l'écart entre sortie et référence (dans notre cas c'est l'erreur quadratique moyenne MSE)
- Ajustement des paramètres : l'erreur mesurée est utilisée pour corriger les poids et les biais du réseau.

Cet ajustement repose sur l'algorithme de descente de gradient, qui modifie les poids et biais pour minimiser la valeur de la fonction de perte. L'amplitude de cette correction est contrôlée par le taux d'apprentissage (learning rate) λ. La règle de mise à jour s'écrit :

$$\theta^{(t+1)} = \theta^{(t)} - \lambda \cdot \nabla_\theta J(\theta^{(t)}) \qquad \textbf{(III.5)}$$

Où :

- $\theta^{(t)}$ est le vecteur de l'ensemble des poids et biais du réseau à l'itération t
- λ est le taux d'apprentissage (learning rate)
- $\nabla_\theta J(\theta^{(t)})$ est le gradient de la fonction de perte, c'est-à-dire le vecteur des dérivées partielles de la fonction de perte J par rapport à chaque poids du modèle.

En pratique, la mise à jour ne s'effectue pas sur l'ensemble des données d'un coup, mais sur de petits sous-ensembles appelés lots (mini-batches) de taille B, appelée batch size. Le gradient est alors estimé à partir de ce sous-ensemble de données de la manière suivante :

$$\nabla_\theta J(\theta) \approx \frac{1}{B} \sum_{i=1}^{B} \nabla_\theta l(\theta; x_i, y_i) \qquad \textbf{(III.6)}$$

Où :

- (x_i, y_i) représente une paire entrée/sortie du lot considéré
- $l(\theta; x_i, y_i)$ est la fonction de perte du batch, qui mesure l'erreur du modèle sur un exemple donnée

Cette variante constitue la base de la plupart des méthodes d'optimisation utilisées en apprentissage automatique. Plusieurs améliorations (par exemple Adam, RMSProp, Adagrad) existent et permettent d'accélérer la convergence ou d'améliorer la stabilité.

L'entrainement consiste à répéter ce cycle (prédiction → calcul de l'erreur → mise à jour des paramètres) sur l'ensemble du jeu de données, comme l'illustre la Figure III.5. Ces cycles sont appelés epoch en machine learning.

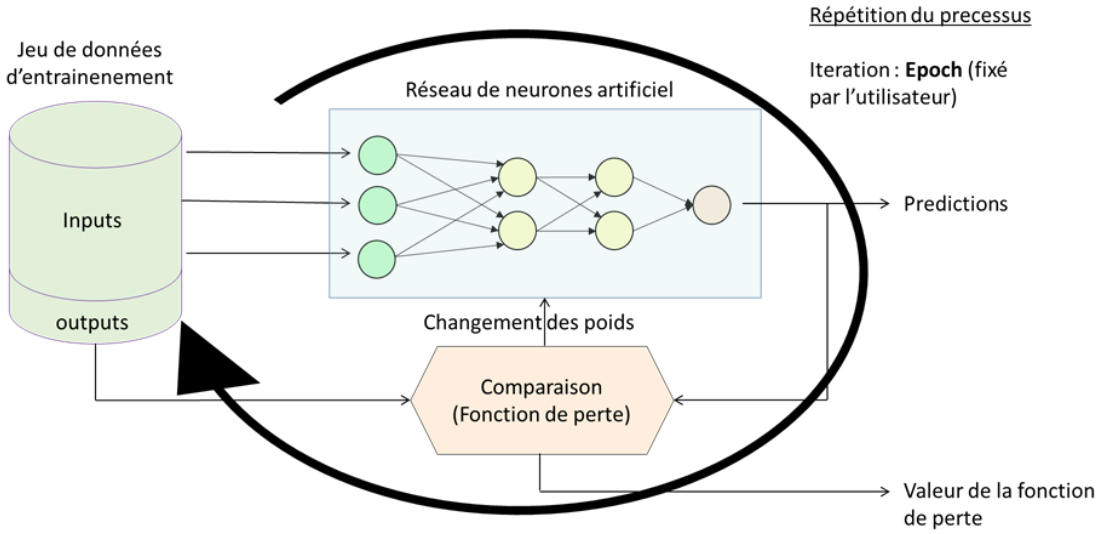

Figure III.5 : Structure générale d'un réseau de neurones artificiels et processus d'apprentissage.

III.1.6 Suivi et évaluation de l'apprentissage

L'évolution de la fonction de perte au cours de l'entraînement constitue un indicateur essentiel de la qualité de la modélisation. En traçant cette fonction en fonction du nombre d'epochs, à la fois sur le jeu d'entraînement et sur le jeu de validation, il est possible de suivre la convergence du modèle et de détecter d'éventuels problèmes lors de l'apprentissage.

En début d'entraînement, la perte diminue simultanément sur les deux ensembles, ce qui traduit une amélioration globale des performances. Si l'entraînement se prolonge, il arrive qu'un écart se forme : la perte sur le jeu d'entraînement continue de baisser, tandis que celle du jeu de validation remonte (Figure III.6). Ce comportement correspond à un sur-apprentissage (overfitting) : le modèle s'adapte trop spécifiquement aux données d'entraînement et perd sa capacité de généralisation. Ce phénomène peut survenir lorsque le réseau est trop complexe, c'est-à-dire qu'il comporte un nombre excessif de couches ou de neurones par rapport à la quantité ou à la variabilité des données. Cela peut également se produire lorsque le nombre d'epochs est trop élevé, conduisant le modèle à affiner exagérément ses paramètres sur les exemples d'entraînement.

À l'inverse, si les deux courbes restent élevées sans diminution notable, cela traduit un sous-apprentissage (underfitting), généralement dû à un modèle trop simple ou à un nombre d'epochs insuffisant.

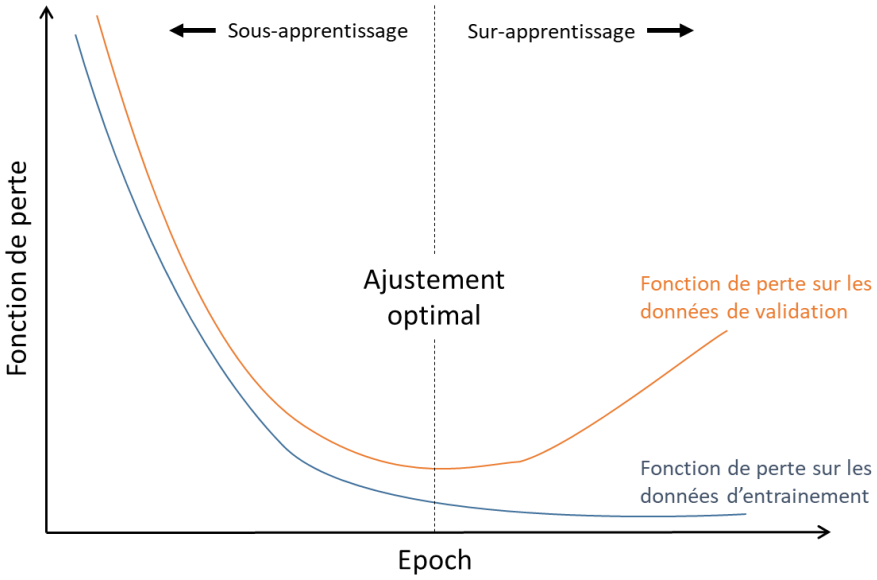

Figure III.6 : Évolution de la fonction de perte sur les jeux d'entraînement et de validation au cours des epochs.

Ces phénomènes de sur-apprentissage et de sous-apprentissage peuvent également s'illustrer en considérant la capacité du modèle à approximer une fonction cible f(x) à partir d'exemples d'entraînement. Dans le cas de sous-apprentissage (Figure III.7.A), le modèle est trop simple et ne capture pas la structure de f(x). Lorsqu'un ajustement satisfaisant est atteint (Figure III.7.B), le modèle restitue correctement la tendance générale sans ajustement excessif aux données d'apprentissage. Enfin, en situation de sur-apprentissage (Figure III.7.C), le modèle suit excessivement les fluctuations du jeu d'entraînement, y compris le bruit, ce qui nuit à sa capacité de généralisation.

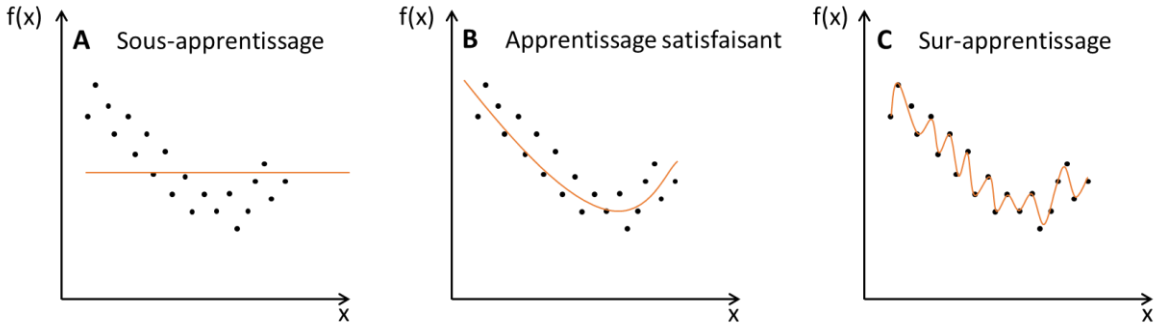

Figure III.7 : : Illustration des régimes d'ajustement d'un modèle de régression visant à approximer de la fonction f(x). (A) sous-apprentissage, le modèle ne capture pas la structure de f(x). (B) : ajustement global satisfaisant. (C) : sur-apprentissage, le modèle reproduit les fluctuations du jeu d'entraînement au détriment de la généralisation.

Dans la suite de ces travaux, l'évaluation de la qualité de l'apprentissage reposera principalement sur l'analyse des courbes de perte en fonction du nombre d'epochs, qui constituent un outil central pour suivre la convergence des modèles et détecter d'éventuels problèmes de sur-apprentissage.

III.2 Choix de l'architecture du réseau de neurones

Dans les premières phases de la thèse, des travaux exploratoires ont été menés afin de valider le principe de la métrologie hybride par apprentissage automatique. Une base de données, dite « virtuelle », a été construite à partir de modèles mathématiques simples. Pour chaque échantillon, les paramètres dimensionnels de référence (CD, H et SWA) ont été choisis aléatoirement à l'aide d'une loi uniforme dans leurs plages respectives :

$$30\ \text{nm} \leq \text{CD} \leq 65\ \text{nm}\ ;\ 20\ \text{nm} \leq \text{H} \leq 100\ \text{nm}\ ;\ 80° \leq \text{SWA} \leq 90°$$

La Figure III.8 montre les distributions obtenues pour chacun de ces paramètres, confirmant leur répartition uniforme dans les intervalles définis.

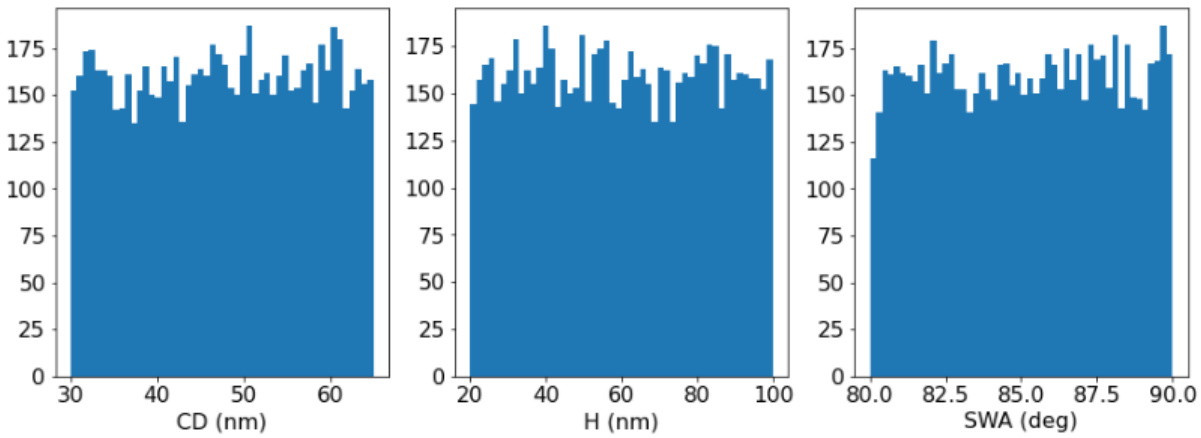

Figure III.8 : Distribution uniforme des paramètres dimensionnels de référence CD, H et SWA utilisés pour générer la base de données virtuelle (8000 échantillons).

Les grandeurs d'entrée associées ont ensuite été obtenues en ajoutant un bruit gaussien sur ces valeurs de référence, avec une incertitude fixée à 5 % de la valeur cible pour un intervalle de confiance de 3σ. Les corrélations entre erreurs n'ont pas été prises en compte et aucun biais fixe n'a été introduit. Ce jeu de données virtuel constitue ainsi un cadre contrôlé permettant d'évaluer le comportement de l'hybridation par réseaux de neurones.

III.2.5 Évaluation initiale des performances de l'hybridation

Nous générons 8 000 échantillons « virtuels » dans un cadre contrôlé. Pour chaque échantillon, nous avons : les paramètres de référence (CD_{ref}, H_{ref} et SWA_{ref}) correspondant aux sorties du réseau. Et les 10 paramètres bruités représentant les techniques métrologiques (CD_{afm}, H_{AFM}, SWA_{AFM}, CD_{CDSEM}, CD_{OCD}, H_{OCD}, SWA_{OCD}, ΔCD_{SAXS}, H_{SAXS} et ΔSWA_{SAXS}). L'ensemble est réparti en 5 000 échantillons pour l'apprentissage, 2 000 pour la validation, et 1 000 pour le test. Cette démarche s'aligne sur l'approche ANN de fusion de données AFM, CDSEM, OCD et CDSAXS décrite par les travaux de Penlap Woguia et al. [72].

Ainsi, en reprenant ces travaux [72] où plusieurs architectures ont été testé, nous utilisons le plus performant : un MLP 10/5/5/3, 10 entrées, deux couches cachées de 5 neurones (activation ReLU), sortie à 3 neurones (activation linéaire) pour prédire les 3 paramètres dimensionnels, comme décrit par la Figure III.9. La fonction de perte d'entraînement est la MSE.

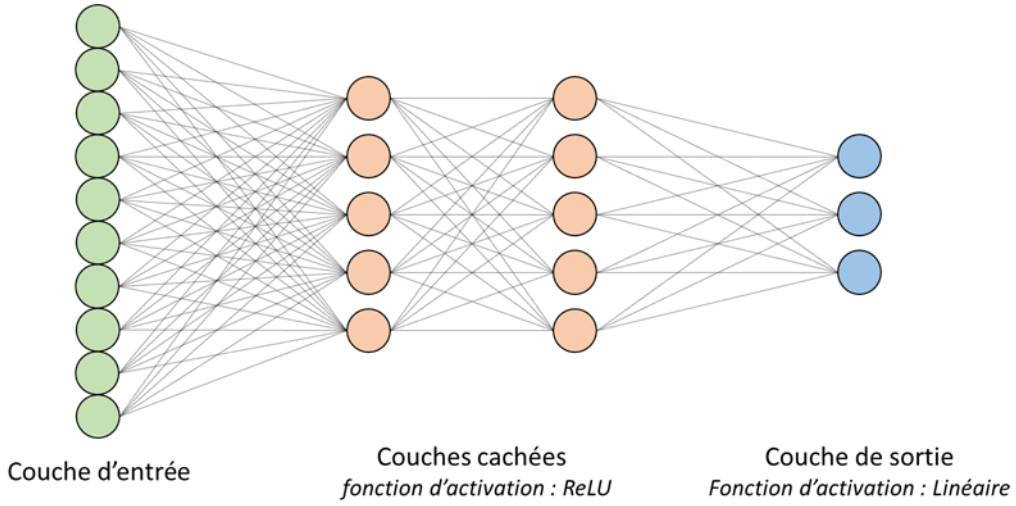

Figure III.9 : Schéma du réseau MLP 10/5/5/3 (fonction ReLU pour les couches cachées et linéaire pour la couche de sortie).

Les pertes entraînement/validation diminuent conjointement et se stabilisent vers ~100 epochs, sans écart caractéristique d'un sur-apprentissage (Figure III.10).

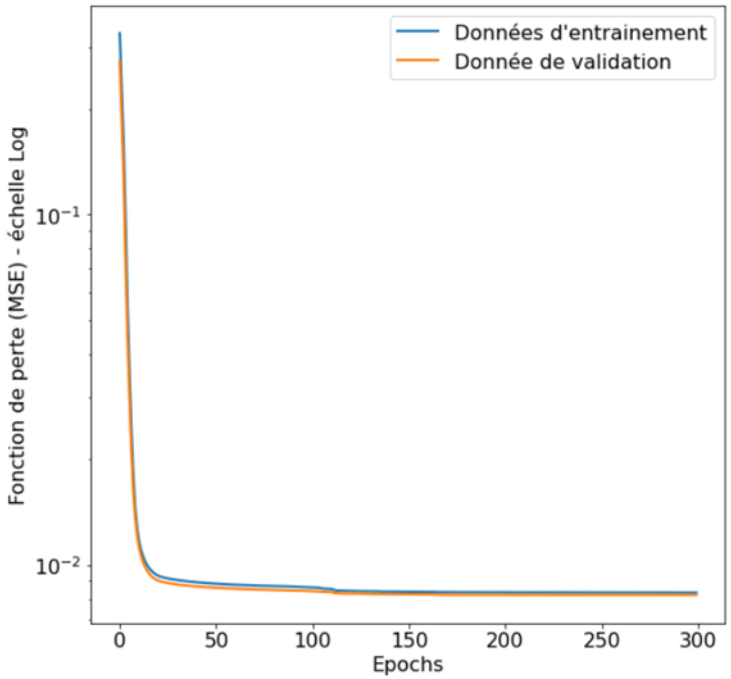

Figure III.10 : Évolution de la fonction de perte pendant l'apprentissage du réseau MLP 10/5/5/3, sur 300 epochs, avec un BatchSize de 50 et un learning rate de 0.001. 5000 données virtuelles pour l'entrainement et 2000 données virtuelles pour la validation.

L'évaluation finale a été réalisée sur le jeu de test (1 000 échantillons). Cette étape permet de vérifier la capacité du modèle à généraliser, c'est-à-dire à fournir des estimations fiables sur des données qu'il n'a jamais vues lors de l'apprentissage. Pour chaque paramètre de sortie (CD, H et SWA), nous avons calculé les erreurs relatives entre les valeurs prédites et les valeurs de référence. Les résultats sont résumés dans le Tableau III.1, où sont reportés à la fois la moyenne et l'intervalle correspondant à 3σ.

Erreur relative en sortie de réseau de neurones	3σ (3 écart-types)	μ (moyenne)
CD	3.5%	0.13%
H	3.8%	0.29%
SWA	2.9%	0.04%

Tableau III.1 : Erreurs relatives après hybridation par réseau de neurones

Ces résultats montrent que le réseau de neurones parvient à réduire les incertitudes initiales appliquées aux entrées, fixées à 5 % pour un intervalle de confiance à 3σ. Les erreurs obtenues restent donc inférieures au bruit ajouté, ce qui confirme que le modèle exploite efficacement la redondance entre les différentes techniques métrologiques pour améliorer l'exactitude des estimations.

Afin de compléter cette analyse, nous avons étudié la distribution des résidus, c'est-à-dire la différence entre les valeurs prédites et les références. Pour évaluer la

distribution des erreurs, nous avons eu recours à des diagrammes quantile-quantile (QQ-plots), qui permettent de comparer visuellement les résidus à une loi normale théorique. La Figure III.11 illustre ce résultat pour le paramètre CD : la bonne superposition des points sur la diagonale indique que les erreurs suivent une loi proche de la normale. Un comportement similaire est observé pour les deux autres paramètres, H et SWA.

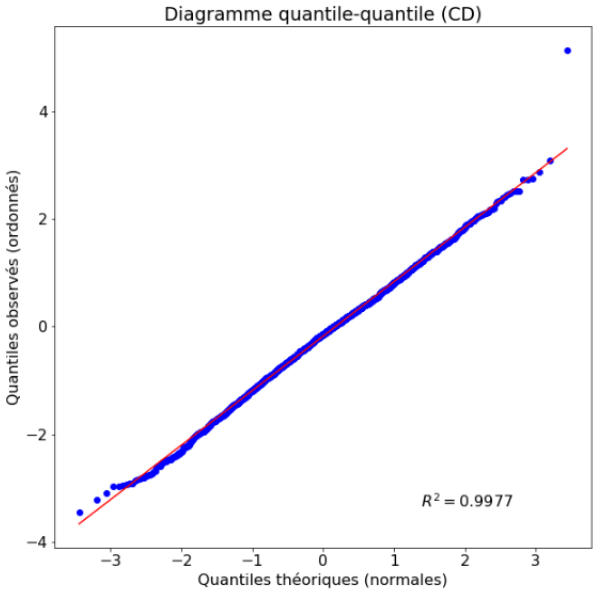

Figure III.11 : Diagramme quantile-quantile (QQ-plot) des erreurs du CD prédit.

Une particularité importante est apparue lors de l'analyse de la fonction de perte utilisée pendant l'apprentissage. Dans le cas d'une régression multi-sorties, la MSE globale calculée par l'algorithme d'apprentissage correspond à la moyenne des erreurs quadratiques obtenues pour chaque paramètre de sortie. Mathématiquement, on peut l'écrire :

$$MSE_{global} = \frac{1}{3}\left(MSE_{CD} + MSE_{H} + MSE_{SWA}\right) \qquad \textbf{(III.7)}$$

Pour mieux comprendre le comportement du réseau, nous avons également examiné les MSE calculées séparément pour chaque paramètre, toujours dans l'espace normalisé. Les résultats, présentés dans le Tableau III.2, mettent en évidence un déséquilibre marqué :

	CD	H	SWA
MSE	10^{-3}	$8*10^{-4}$	$3*10^{-2}$

Tableau III.2 : MSE calculées manuellement pour chaque sortie du réseau de neurone (valeurs normalisées)

On constate que la valeur associée au SWA est nettement plus élevée que celles du CD et du H. Comme la fonction de perte globale correspond à la moyenne de ces trois termes, elle est dominée par la contribution du SWA. En pratique, on peut donc écrire:

$$MSE_{global} \approx \frac{1}{3} MSE_{SWA} \qquad \textbf{(III.8)}$$

Ainsi, lors de l'apprentissage, l'algorithme, cherchant à minimiser la fonction de perte, se focalise davantage sur le SWA, celui-ci ayant un impact plus fort sur le MSE global que les autres paramètres. Cela pose donc un problème d'optimisation lors de la phase d'apprentissage. Ceci nous a amené à étudier une nouvelle architecture.

III.2.6 Architecture Multi-Branches

Plusieurs travaux ont exploré l'utilisation de réseaux de neurones artificiels dits multi-branches [73]. Ces architectures s'inspirent du principe bien connu en algorithmique de « divide and conquer » (diviser pour mieux régner) [74], [75], qui consiste à décomposer un problème complexe en sous-problèmes plus simples, résolus indépendamment avant d'être recombinés. Dans le cas des réseaux multi-branches, chaque branche prend en charge une partie du problème, ce qui favorise la spécialisation et peut améliorer l'exactitude globale du modèle.

Des modèles connus comme ResNeXt [76], Xception [77] ou SqueezeNet [78] s'appuient sur ce type de structure pour optimiser l'efficacité et la exactitude des réseaux de neurones profonds. En effet, lorsqu'une branche du réseau est associée à une seule sortie, il devient possible de calculer une fonction de perte spécifique à chaque grandeur cible. Cela permet d'éviter les effets de domination observés lors de l'optimisation d'une fonction de perte globale partagée par toutes les sorties.

L'architecture mise en place dans ce travail repose sur ce principe : chaque paramètre de sortie (CD_{ref}, H_{ref} et SWA_{ref}) est pris en charge par une branche dédiée du réseau, comme illustré en Figure III.12. Les dix paramètres d'entrée (CD_{AFM}, H_{AFM}, SWA_{AFM}, CD_{CDSEM}, CD_{OCD}, H_{OCD}, SWA_{OCD}, ΔCD_{SAXS}, H_{SAXS} et ΔSWA_{SAXS}) sont partagés entre les différentes branches, permettant au modèle d'exploiter les corrélations inter-paramètres lors de l'apprentissage.

Chaque branche correspond à un perceptron multicouche (MLP) indépendant, composé d'une seule couche cachée suivie d'une couche de sortie spécifique au paramètre prédit. Le nombre de neurones dans cette couche cachée, ainsi que les fonctions d'activation associées, ont été déterminés empiriquement en testant plusieurs configurations. L'architecture retenue correspond à celle offrant le meilleur

compromis entre performance et complexité, c'est-à-dire un nombre minimal de neurones tout en assurant une bonne généralisation sur le jeu de validation.

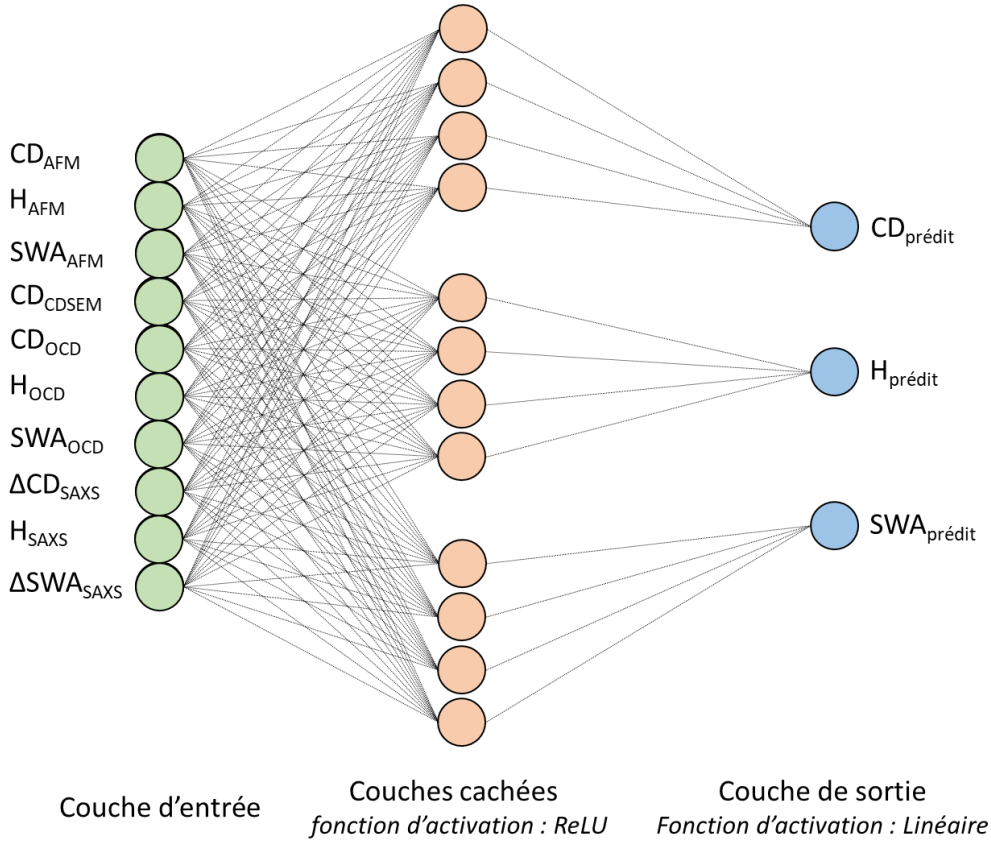

Figure III.12 : Schéma de l'architecture multi-branches 10/4*3/3. Chaque branche est dédiée à un seul paramètre dimensionnel

Une phase d'apprentissage supervisé est réalisée en suivant la même procédure que pour l'architecture précédente (qu'on appellera architecture MLP). Chaque branche est entraînée indépendamment, en utilisant les mêmes bases de données que précédemment. La fonction de perte associée à chaque sortie est évaluée séparément.

Nous avons pu acquérir l'évolution des fonctions de perte pour chacune des branches. On observe une convergence rapide du MSE sur les jeux d'entraînement et de validation. Sur la branche du CD (Figure III.13.A), un très léger début de sur-apprentissage est perceptible. Il est donc nécessaire de sauvegarder les poids du modèle avant l'apparition de ce phénomène, c'est-à-dire vers l'epoch 100. Les branches H et SWA (Figure III.13.B et Figure III.13.C) ne présentent pas de divergence notable entre les courbes d'entraînement et de validation, ce qui indique l'absence de sur-apprentissage. Dans l'ensemble, ces résultats confirment la bonne capacité de généralisation du modèle multi-branches.

On observe une réduction systématique du MSE (normalisé) sur l'ensemble des sorties lorsque l'on utilise l'architecture multi-branches, comme le montre Tableau III.3. La diminution est particulièrement marquée pour le paramètre H, dont l'erreur est réduite d'un facteur supérieur à trois. Pour le CD et le SWA, la baisse est plus modérée (environ 30 %), mais reste significative et confirme que l'architecture multi-branches améliore l'exactitude de l'ensemble des prédictions.

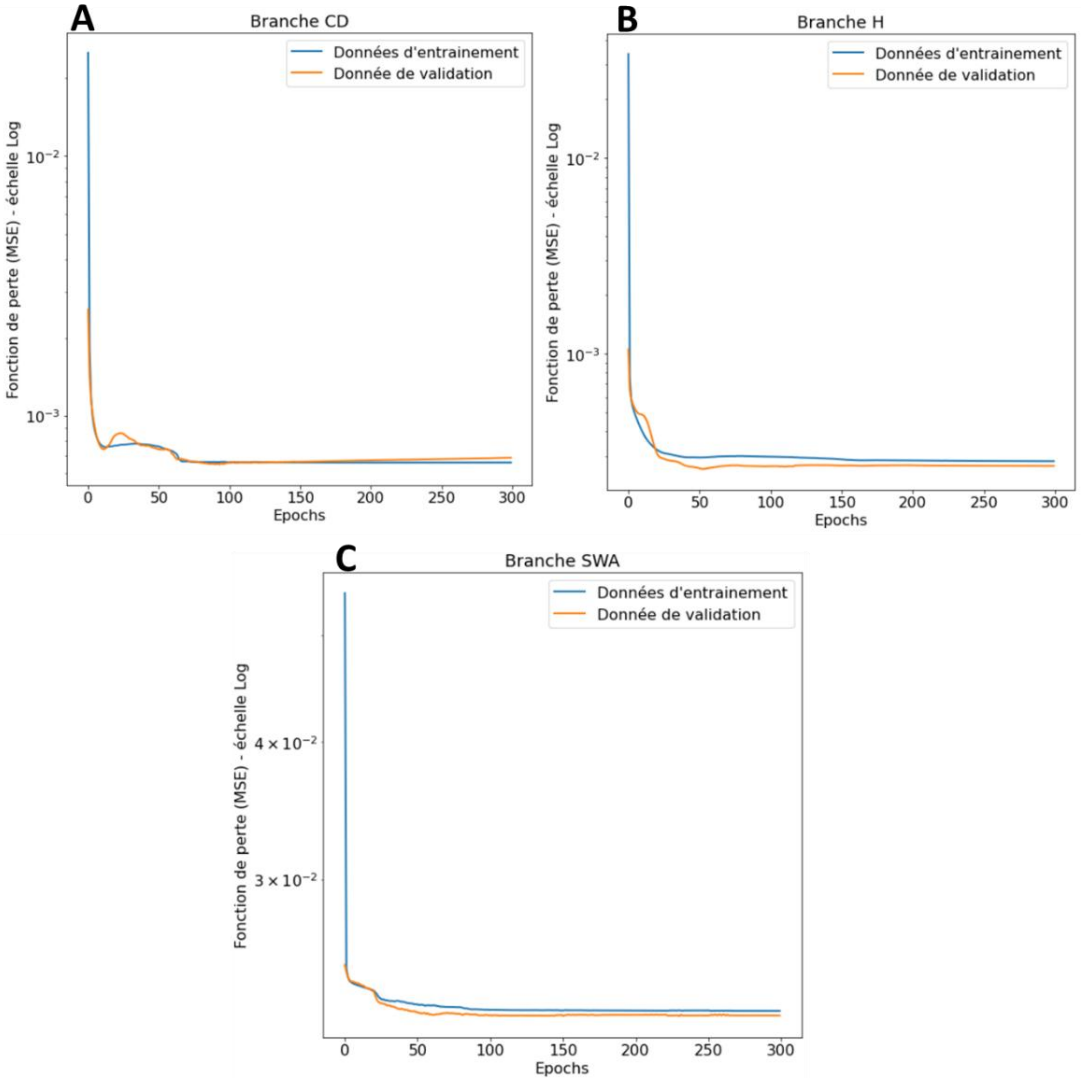

Figure III.13 : Évolution des fonctions de perte de chaque branche pendant l'apprentissage de la branche CD (A), de la branche H (B) et de la branche SWA (C). Sur 300 epochs, avec un BatchSize de 50 et un learning rate de 0.001. 5000 données virtuelles pour l'entrainement et 2000 données virtuelles pour la validation.

MSE	CD	H	SWA
Architecture classique	10^{-3} nm²	$8*10^{-4}$ nm²	$3*10^{-2}$ deg²
Architecture multi-branches	$7*10^{-4}$ nm²	$3*10^{-4}$ nm²	$2*10^{-2}$ deg²

Tableau III.3 : Comparaison des fonctions de perte en fin d'entrainement pour chaque sortie entre l'architecture classique et l'architecture multi-branches

Ces résultats confirment que l'approche multi-branches améliore l'apprentissage sur chacune des grandeurs cibles. En particulier, la spécialisation de chaque sous-réseau permet d'atténuer l'effet de compétition entre les sorties au sein de la fonction de perte globale, observé avec l'architecture classique. Ce comportement valide l'intérêt de cette architecture pour la métrologie hybride.

III.2.7 Analyse comparative des deux architectures

Tout comme avec l'architecture MLP, les erreurs en sortie du modèle multi-branches suivent des distributions gaussiennes, ce que nous avons pu vérifier par l'intermédiaire de QQ-plots. Cela permet une comparaison directe entre les deux approches. Le Tableau III.4 présente une amélioration nette de l'exactitude des valeurs prédites avec l'architecture multi-branches.

En entrée, les incertitudes simulées correspondaient à 5 % de la valeur réelle (pour un intervalle de confiance à 3σ). Avec l'architecture classique, les incertitudes relatives obtenues en sortie étaient de 3.5 % pour le CD, 3.8 % pour le H et 2.9 % pour le SWA. Avec l'architecture multi-branches, ces incertitudes sont réduites respectivement à 2.9 %, 3.0 % et 2.6 % pour les trois mêmes grandeurs. Ces résultats confirment les tendances observées sur les courbes de fonction de perte et soulignent l'influence déterminante de l'architecture du modèle. En particulier, le choix de dissocier les fonctions de perte pour chaque sortie apparaît ici comme un levier efficace pour améliorer la qualité des prédictions.

Erreur relative (à 3σ)	CD	H	SWA
Architecture MLP	3.5%	3.8%	2.9%
Architecture multi-branches	2.9%	3.0%	2.6%

Tableau III.4 : Comparaison des erreurs entre les architectures classique et multi-branches

Au vu des résultats obtenus, l'architecture multi-branches s'impose comme une solution plus adaptée au contexte de la métrologie hybride. Elle permet non seulement

une amélioration globale des performances en termes d'exactitude, mais offre également une meilleure maîtrise de l'apprentissage en limitant les interférences entre les sorties du modèle. Chaque grandeur cible étant traitée par un sous-réseau dédié, il devient possible de suivre l'apprentissage de manière indépendante pour chaque sortie. Concrètement, cela signifie qu'une branche peut atteindre une convergence plus rapide que les autres et ne nécessite pas forcément le même nombre d'epochs. Cette modularité offre une souplesse supplémentaire dans les phases d'entraînement, en permettant d'adapter le suivi et l'arrêt de l'optimisation branche par branche. Ainsi, l'ensemble des travaux présentés dans la suite de ce manuscrit s'appuiera exclusivement sur cette architecture.

III.2.8 Estimation de la taille minimale du jeu de données

Afin de garantir des performances stables et une bonne capacité de généralisation du modèle, une étude a été menée pour évaluer l'influence du nombre de données d'apprentissage sur le comportement de la fonction de perte. Les ensembles de validation et de test ont aussi été ajustés proportionnellement (avec des volumes fixés à respectivement un tiers et un sixième du jeu d'entraînement avec un minimum de 100 données). Cette analyse a été réalisée à partir de données virtuelles, dans les mêmes conditions expérimentales que les précédents tests avec une architecture multi-branches, 3 branches ayant 1 couche cachée de 4 neurones chacune, configuration identique (learing rate = 0.001 et batch size = 50).

Plusieurs éléments sont à prendre en compte lors de cette analyse : les valeurs des fonctions de perte en fin d'apprentissage, l'écart entre les fonctions de perte sur les données d'entraînement et de validation, et, dans une moindre mesure, la vitesse de convergence.

La Figure III.14 présente l'évolution des fonctions de perte d'apprentissage et de validation pour différentes tailles de jeux de données pour la branche se focalisant sur la hauteur (H) des lignes. On observe les mêmes comportements avec les branches CD et SWA. On distingue un écart notable entre les erreurs quadratiques moyennes des données d'entraînement et de validation pour les plus petits jeux. L'étude a été menée sur plusieurs tailles de jeux d'entraînement : 200 données (Figure III.14.A), 300 données (Figure III.14.B), 400 données (Figure III.14.C) et 500 données (Figure III.14.D). On observe que, pour le cas de 200 données, la fonction de perte de validation se stabilise tardivement, vers environ 230 epochs, et atteint une valeur minimale d'environ $3{,}5\times10^{-4}$. À l'inverse, pour les cas avec 300, 400 et 500 données, la convergence est beaucoup plus rapide (dès une cinquantaine d'epochs) et les valeurs finales atteignent un niveau plus bas (environ $2{,}3\times10^{-4}$). Ce résultat montre que l'augmentation de la taille du jeu d'apprentissage améliore à la fois la vitesse et la

qualité de la convergence en validation. L'écart entre les courbes d'entraînement et de validation devient minime à partir de 400 données. La réduction de cet écart constitue un indicateur clé de la capacité de généralisation du modèle : plus les courbes sont proches, plus le réseau est capable de reproduire correctement le comportement sous-jacent sur des données qu'il n'a jamais vues, sans sur-apprentissage ni sous-apprentissage.

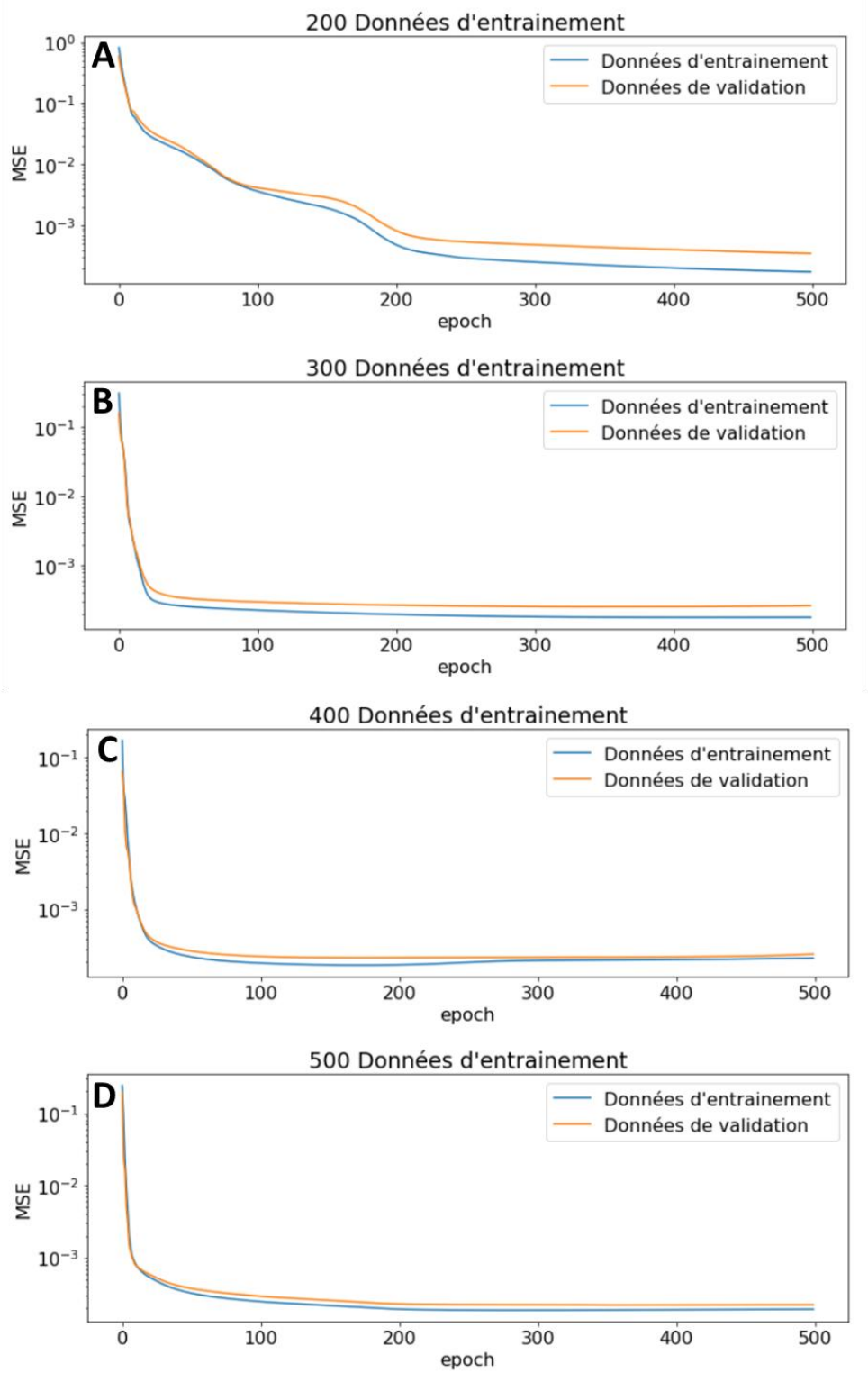

Figure III.14 : Evolution de l'erreur quadratique moyen (fonction de perte) pour la branche H sur les données d'entrainement et de validation en fonction de l'epoch pour des jeux de données de (A) 200 données d'entrainement et 100 données de validation, (B) 300 données d'entrainement et 100 données de validation, (C) 400 données d'entrainement et 134 données de validation et (D) 500 données d'entrainement et 167 données de validation.

Au vu de ces résultats, nous avons estimé qu'un minimum de 400 données d'entraînement est nécessaire pour garantir un apprentissage optimal avec l'architecture actuelle. Conformément aux pratiques couramment admises en apprentissage automatique, une proportion d'environ un tiers des données est allouée à la validation. Nous avons donc retenu une configuration avec 200 données pour la validation et 100 pour le test, permettant d'évaluer les performances du modèle sur des données indépendantes, non vues durant l'apprentissage. Cette configuration, soit un total de 700 échantillons, offre un compromis satisfaisant entre la quantité de données et la capacité de généralisation du modèle.

Bien que basés sur des données virtuelles, ces travaux nous ont permis d'estimer de manière pragmatique l'ordre de grandeur du nombre de données nécessaires pour entraîner correctement le modèle. Cette estimation a ensuite servi de référence pour définir la taille des jeux de données dites simulées générés dans le chapitre précédent, plus représentatives des conditions expérimentales réelles et utilisées dans la suite du travail.

III.3 Hybridation des données simulées

À la suite des études menées sur des données virtuelles, générées à partir de distributions gaussiennes, nous poursuivons notre démarche en introduisant l'hybridation avec les données simulées. Ces dernières, issues des simulations décrites dans le chapitre précédent, permettent un apprentissage sur des données plus proches des observations expérimentales.

III.3.5 L′apprentissage

À l'issue de l'apprentissage, les courbes de perte (entraînement/validation) ont été examinées branche par branche. La Figure III.15 montre que toutes les branches convergent et que l'écart de généralisation (différence validation–entraînement) reste faible. Seule la branche associée à H présente un léger sur-apprentissage à partir d'environ 200 epochs. Une sauvegarde automatique des poids et biais du modèle a toutefois permis de conserver les paramètres correspondant à la meilleure performance de validation, avant l'apparition de cette dérive. Ce comportement global indique que le réseau a généralisé correctement les relations apprises à partir des données d'entraînement.

Sur la branche SWA, on observe qu'en tout début d'entraînement, la perte mesurée sur le jeu de validation est inférieure à celle du jeu d'entraînement. Ce phénomène peut s'expliquer par plusieurs possibilités :

- L'utilisation des mini-batchs rend la perte d'entraînement plus bruitée que celle de validation, généralement évaluée sur l'ensemble du dataset.
- Du fait du découpage aléatoire, le jeu de validation soit initialement légèrement plus favorable que le jeu d'entraînement.

Dans tous les cas, cet écart transitoire disparaît rapidement au cours de l'entraînement, signe que le réseau converge vers une solution stable et généralisable.

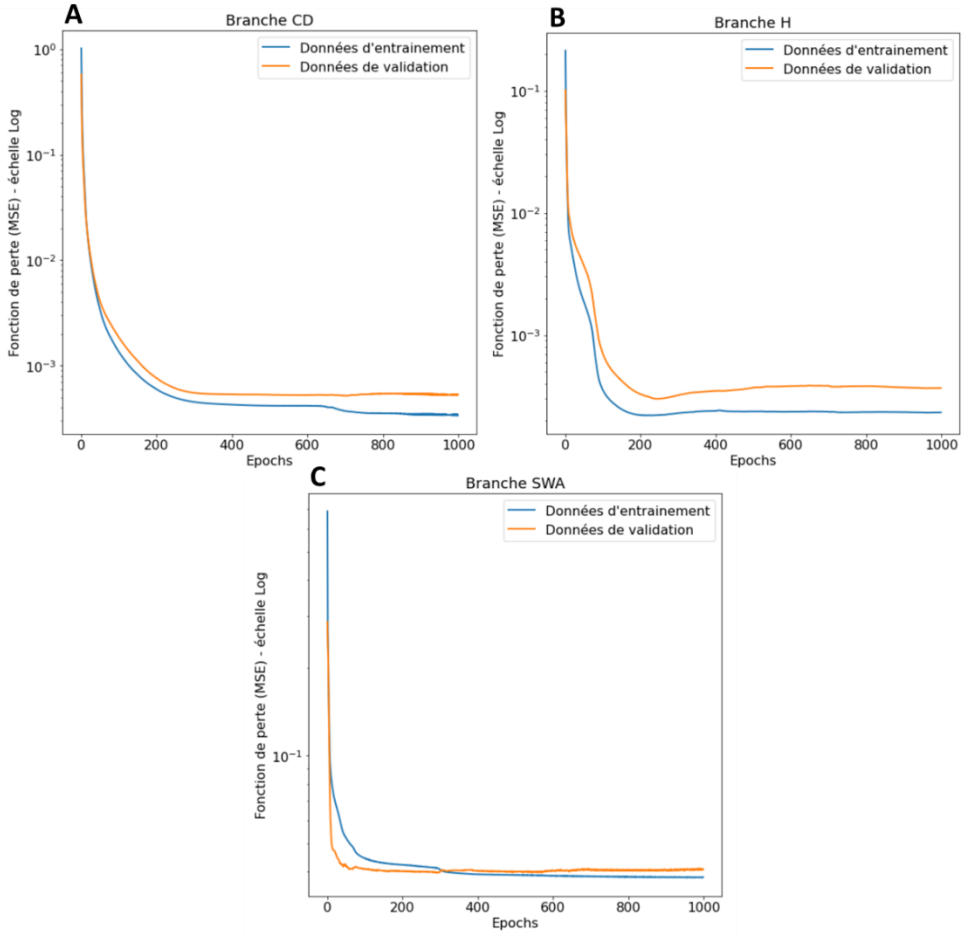

Figure III.15 : Fonctions de pertes des données d'entraînement et de validation pour les branches (A) du CD, (B) de H et (C) du SWA

III.3.6 Résultats

Afin d'évaluer les performances du réseau de neurones, un ensemble de 100 échantillons de test a été utilisé. Ces données, distinctes de celles utilisées lors de l'apprentissage et de la validation, ont été fournies au réseau afin de générer des prédictions pour chacun des paramètres dimensionnels. Les valeurs prédites ont

ensuite été comparées aux valeurs de référence attendues, ce qui a permis de quantifier les erreurs de prédiction.

Pour cette évaluation, deux indicateurs statistiques ont été retenus : la moyenne des erreurs, permettant d'identifier un éventuel biais systématique, et l'écart-type des erreurs, mesurant la dispersion des prédictions autour des valeurs attendues. Les résultats sont présentés dans le Tableau III.5

		CD	H	SWA
AFM	µ (moyenne)	-0,05 nm	1,19 nm	-1,29 °
	σ (écart-type)	**0,65 nm**	**0,68 nm**	**1,29 °**
CDSEM	µ (moyenne)	1,40 nm		
	σ (écart-type)	**0,91 nm**		
OCD	µ (moyenne)	-0,41 nm	0,13 nm	-1,32 °
	σ (écart-type)	**2,05 nm**	**0,92 nm**	**3,49 °**
SAXS	moyenne	0,12 nm	-0,05 nm	0,08 °
	Ecart-type global	**0,54 nm**	**0,26 nm**	**0,67 °**
	Ecart-type fin	0,25 nm	0,15 nm	0,27 °
	Ecart-type large	1,21 nm	0,49 nm	1,44 °
Réseau de neurones	Moyenne	-0,05 nm	0,01 nm	0,01 °
	Ecart-type	**0,35 nm**	**0,47 nm**	**0,43 °**

Tableau III.5 : Comparaison des moyennes et écart-types des différentes techniques et du réseau de neurones réalisant l'hybridation.

On observe que les moyennes des erreurs sont quasi nulles pour l'ensemble des paramètres étudiés post-hybridation, ce qui traduit une absence de biais dans les prédictions du réseau. Ce résultat montre que l'hybridation est capable de compenser les biais propres à chaque technique métrologique.

Concernant la dispersion des erreurs, les écarts-types obtenus pour les paramètres CD et SWA sont inférieurs à ceux observés sur les données des simulations de chaque technique prise individuelles, ce qui témoigne d'une amélioration effective apportée par l'hybridation. Plus précisément, l'écart-type du CD passe de 0,54 nm (valeur minimale obtenue avec la technique SAXS) à 0,35 nm après hybridation, tandis que celui du SWA diminue de 0,67° à 0,43°.

En revanche, pour le paramètre H, cela est plus complexe. En effet, dans le chapitre précédent, nous avons montré que les données simulées SAXS présentent une distribution non-gaussienne, composée de deux populations distinctes. En conséquence, nous avons distingué dans le Tableau III.5, 3 valeurs d'écart-type pour les résultats SAXS :

- Un écart-type « fin », correspondant à la population majoritaire présentant une erreur faible
- Un écart-type « large », correspondant à environ 15 % des données considérées comme aberrantes.
- Un écart-type « global », correspondant à l'écart-type calculé sur l'ensemble des données.

L'écart-type obtenu pour le paramètre H avec le réseau de neurones (0.47 nm) est légèrement inférieur à l'écart-type SAXS « large » (0.49 nm), mais bien supérieur à l'écart-type SAXS « fin » (0.15 nm). Cela suggère que le modèle ne parvient pas à discriminer la population à laquelle un échantillon donné appartient. Pour minimiser la fonction de perte, le réseau semble adopter un compromis, en intégrant implicitement la présence de cas à forte incertitude.

Enfin, il convient de rappeler que les données SAXS sont, parmi les techniques simulées, celles associées aux incertitudes les plus faibles, ce qui explique la difficulté à faire mieux sur certains paramètres.

L'analyse des erreurs commises par le réseau de neurones multibranches sur l'ensemble de test montre que leurs distributions suivent une loi proche de la gaussienne. Cette hypothèse a été vérifiée à l'aide de QQ-plots, pour lesquels les coefficients de détermination (R^2) obtenus sont supérieurs à 0,98 pour l'ensemble des paramètres étudiés (Figure III.16). Bien que ces résultats soient encourageants, il convient de rester prudent quant à leur interprétation, le jeu de test ne comportant que 100 échantillons, ce qui limite la robustesse statistique de l'analyse.

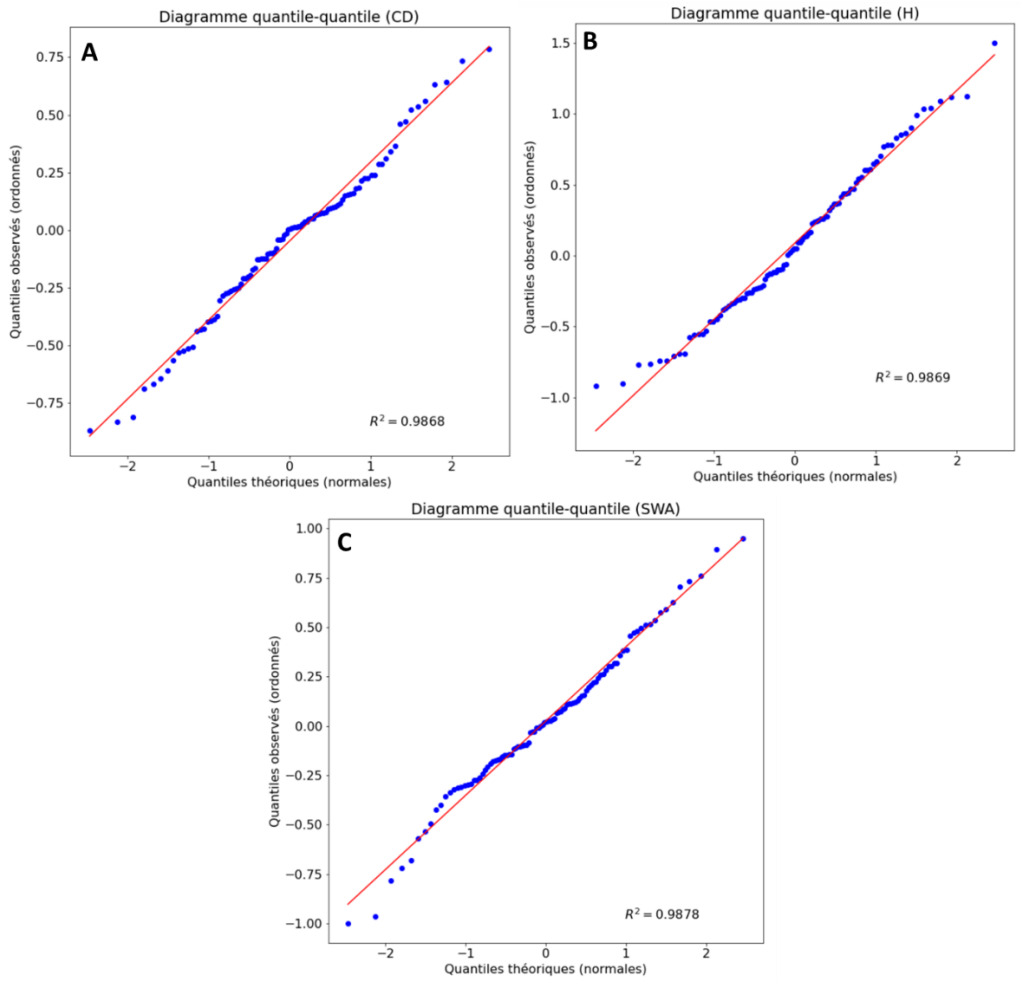

Figure III.16 : Diagramme quantile-quantile entre une distribution normale et les distributions des erreurs des paramètres : (A) CD, (B) H et (C) SWA

Les erreurs étant distribuées selon une loi gaussienne, il est possible d'estimer directement les incertitudes à ±1σ, soit un intervalle de confiance d'environ 68 %. Cette propriété permet également d'étendre l'évaluation aux intervalles ±2σ (≈95 %) et ±3σ (≈99,7 %), offrant ainsi une vision plus complète de la dispersion et des bornes attendues des prédictions.

Selon l'IRDS [4], les incertitudes métrologiques actuelles restent supérieures aux besoins industriels (<1 nm pour certaines métriques) et appellent à la fois à une meilleure quantification statistique et à l'essor de la métrologie hybride pour accroître la fiabilité. Les résultats présentés ici, évalués à 1σ à partir des simulations des différentes techniques de mesure, montrent des écarts-types compris entre 0,26 et 2,05 nm, cohérents avec les performances actuelles mais encore au-dessus des objectifs visés. En revanche, l'approche hybride par réseau de neurones aboutit à des incertitudes nettement réduites (σ ≈ 0,35 nm pour le CD, 0,47 nm pour la hauteur et 0,43° pour l'angle de paroi), tout en garantissant l'absence de biais et de valeurs

aberrantes. Cette robustesse valide la pertinence de la fusion de données issues de différentes techniques métrologiques et ouvre des perspectives concrètes pour une intégration plus large de l'apprentissage automatique dans les chaînes d'analyse, en parfaite cohérence avec les recommandations de l'IRDS.

III.4 Hybridation de données brutes

Au-delà de l'hybridation réalisée à partir de données numériques traitées ou extraites, nous explorons ici l'utilisation directe de données brutes simulées (images, spectres, figures de diffraction, …) comme entrée du processus d'hybridation. Ces données proviennent directement des simulateurs, sans post-traitement ou extraction ni transformation supplémentaire. Cette approche présente deux avantages majeurs : d'une part, les données brutes contiennent l'ensemble des phénomènes physiques modélisés, sans perte d'information liée à une simplification. D'autre part, leur obtention est généralement plus rapide que celle des données extraites, notamment en évitant les étapes itératives de résolution inverse.

III.4.5 Conception du réseau de neurone

Dans le cadre de cette approche, chaque technique métrologique est traitée de manière indépendante à l'aide d'un sous-réseau de neurones dédié, spécifiquement entraîné sur les données brutes associées. L'objectif initial de chaque sous-réseau est d'apprendre à prédire les paramètres dimensionnels à partir des données brutes simulées. Contrairement à l'approche précédente fondée sur des données numériques déjà extraites par résolution de problèmes inverses, il ne s'agit pas ici d'ajouter un étage supplémentaire mais bien de remplacer explicitement cette étape d'inversion par l'apprentissage automatique. Les sous-réseaux apprennent directement une représentation latente des données brutes, laquelle est ensuite utilisée pour l'hybridation.

Le processus de conception envisagé suit trois étapes principales :

1. **Entraînement initial des sous-réseaux** : pour chaque technique (AFM, CDSEM, OCD, SAXS), un réseau de neurones est entraîné indépendamment afin de prédire l'ensemble des paramètres dimensionnels à partir des données brutes spécifiques à la technique considérée.
2. **Extraction des représentations internes** : Une fois l'entraînement terminé, la couche de sortie de chaque sous-réseau est supprimée. Les activations de la dernière couche cachée (valeurs de sortie des neurones), constituant une représentation latente de l'échantillon pour la technique

donnée, sont extraites. Ces vecteurs latents, propres à chaque sous-réseau, sont ensuite concaténés pour former un vecteur global unique, représentatif de l'ensemble des données brutes simulées disponibles pour un échantillon donné. Ce vecteur sert de support à l'hybridation.

3. **Construction du réseau d'hybridation** : Le vecteur global ainsi obtenu est utilisé comme entrée d'un réseau de neurones composé de plusieurs branches, chacune dédiée à la prédiction d'un paramètre dimensionnel. La phase d'apprentissage de ces branches suit le même principe que celui mis en œuvre lors de l'hybridation fondée sur les données numériques. Toutefois, en raison de la richesse accrue des informations contenues dans le vecteur d'entrée, un plus grand nombre de neurones et de couches est généralement requis pour capturer les corrélations pertinentes.

Ce découplage entre l'apprentissage des représentations spécifiques à chaque technique métrologique et l'estimation finale des paramètres permet de tirer parti de la richesse des données brutes tout en garantissant une modularité du modèle. La Figure III.17 présente l'architecture complète proposée, incluant les sous-réseaux spécialisés, la concaténation des représentations latentes, et les branches finales de prédiction.

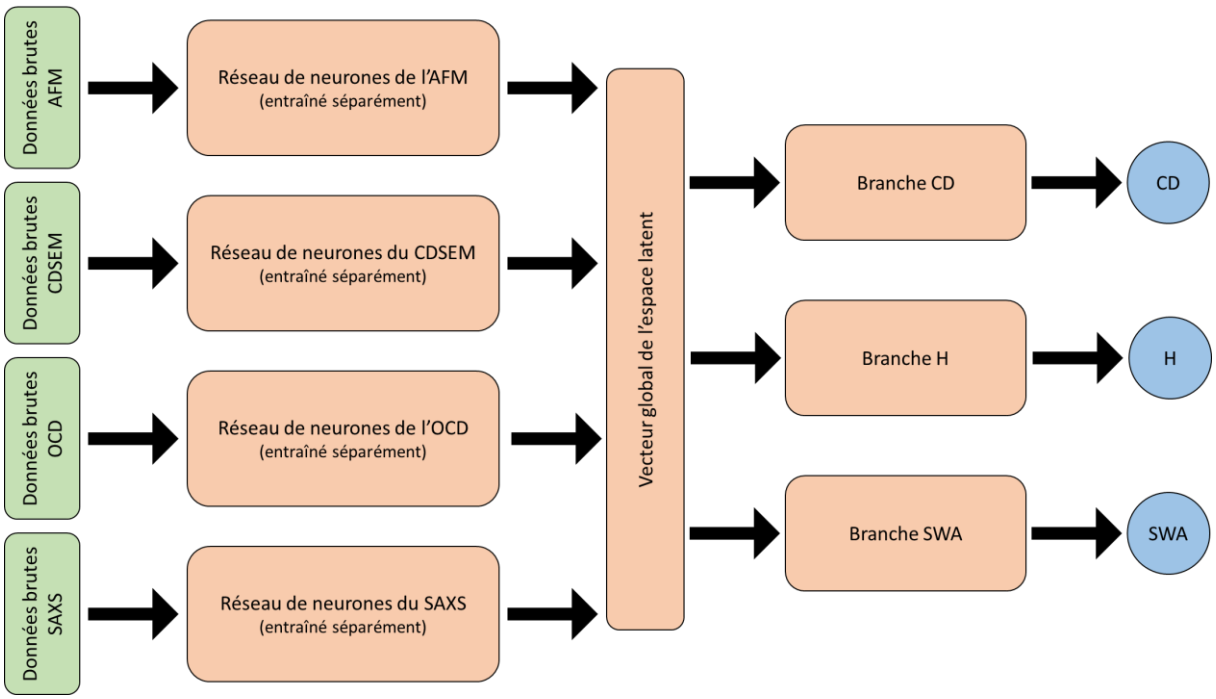

Figure III.17 : Schéma de principe de l'hybridation des données brutes par réseaux de neurones

III.4.6 Choix des sous-réseaux suivant les données

Les techniques métrologiques considérées dans cette étude produisent des données de formats hétérogènes, reflétant la diversité des principes physiques sous-jacents. Ces données peuvent prendre la forme d'images 2D ou 3D, de profils de contours, de nuages de points, ou encore de courbes spectrales. Cette variabilité nécessite une attention particulière lors de l'intégration dans un cadre d'analyse unifiée, notamment dans le contexte de l'hybridation par apprentissage automatique.

III.4.6.1 AFM

Comme présenté dans le chapitre précédent, les données brutes issues de l'AFM se présentent sous forme de nuages de points en 3 dimensions décrivant la topographie des échantillons analysés. Plusieurs types de réseaux de neurones peuvent être envisagés pour leur traitement.

Une première approche consiste à utiliser un réseau de neurones convolutifs (CNN) [79], adapté au traitement d'images. Cela implique toutefois une transformation préalable du nuage de points, soit en image 3D, soit en image 2D où l'intensité des pixels encode la coordonnée altimétrique (axe z).

Une alternative consiste à vectoriser directement l'ensemble des points du nuage, ce qui permet l'utilisation d'un perceptron multicouche (MLP), comme l'illustre la Figure III.18. Cette approche présente plusieurs avantages : elle réduit la complexité du modèle, les MLP comportant généralement moins de paramètres que les CNN, ce qui diminue les besoins en données d'apprentissage et en ressources de calcul.

Dans cette étude, nous privilégions cette seconde méthode. Toutefois, elle impose que chaque échantillon soit représenté par un nombre identique de points, contrairement à l'approche fondée sur les CNN, où seule une taille d'image constante est requise. Cette contrainte peut être contournée en ajoutant et/ou supprimant des points par interpolation afin de respecter le format que l'on aura défini.

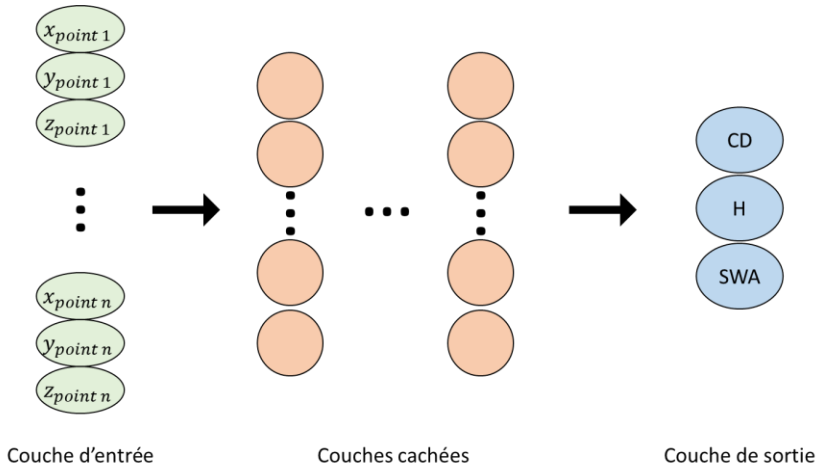

Figure III.18 : Réseau de neurones (MLP) pouvant traiter les données brutes de l'AFM

III.4.6.2 CDSEM

Les données brutes issues du CDSEM se présentent sous forme d'images 2D, dans lesquelles l'intensité des pixels reflète des variations de contraste liées à des variations locales de hauteur, sans correspondre directement à une mesure altimétrique. Par ailleurs, cette technique métrologique permet l'extraction d'un seul paramètre dimensionnel, à savoir la dimension critique (CD). Dans ce contexte, l'utilisation d'un réseau de neurones convolutif (CNN) [79] apparaît naturelle, ce type d'architecture étant particulièrement adapté au traitement d'images.

Les CNN exploitent des filtres convolutifs pour extraire automatiquement des motifs locaux (bords, textures, formes) présents dans l'image, sans nécessiter d'ingénierie manuelle des caractéristiques. Ces représentations sont ensuite combinées au sein de couches entièrement connectées pour effectuer la prédiction finale, comme l'illustre la Figure III.19.

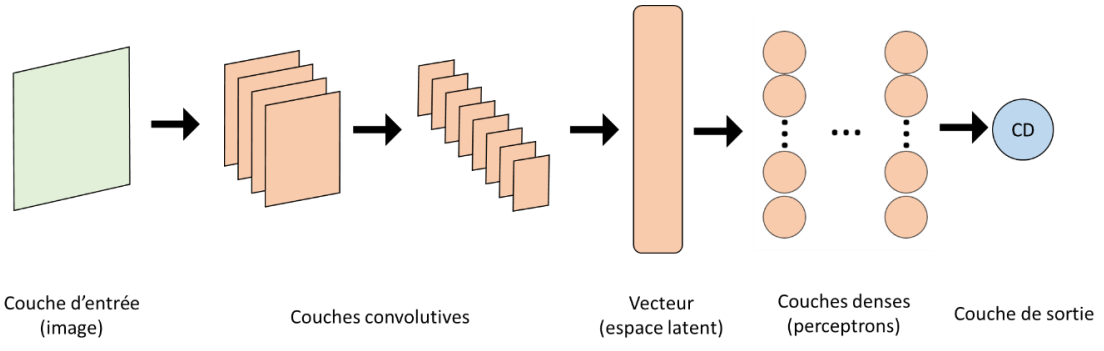

Figure III.19 : Réseau de neurones (CNN) pouvant traiter les données brutes du CDSEM

Cependant, on peut s'interroger sur la pertinence de l'usage d'un réseau de neurones pour ce type d'extraction. En effet, les méthodes classiques d'extraction de CD à partir

d'images CDSEM donnent déjà de très bons résultats. Les principales sources d'incertitude proviennent du bruit de Schottky et de la rugosité des flancs, soit d'éléments inhérents à la technique d'acquisition elle-même, et non au traitement des données. Afin de valider cette hypothèse, il demeure pertinent de comparer les résultats issus de l'approche classique avec ceux obtenus via l'apprentissage automatique.

III.4.6.3 OCD

Les données brutes issues de l'OCD se présentent sous la forme de couples de spectres, mesurés selon différentes configurations angulaires et/ou polarisations. Les réseaux de neurones de type perceptron multicouche (MLP) semble adapté pour traiter des spectres. Ceux-ci doivent être vectorisés, c'est-à-dire transformés en vecteurs de caractéristiques, pour être traités efficacement par un MLP, comme l'illustre la Figure III.20.

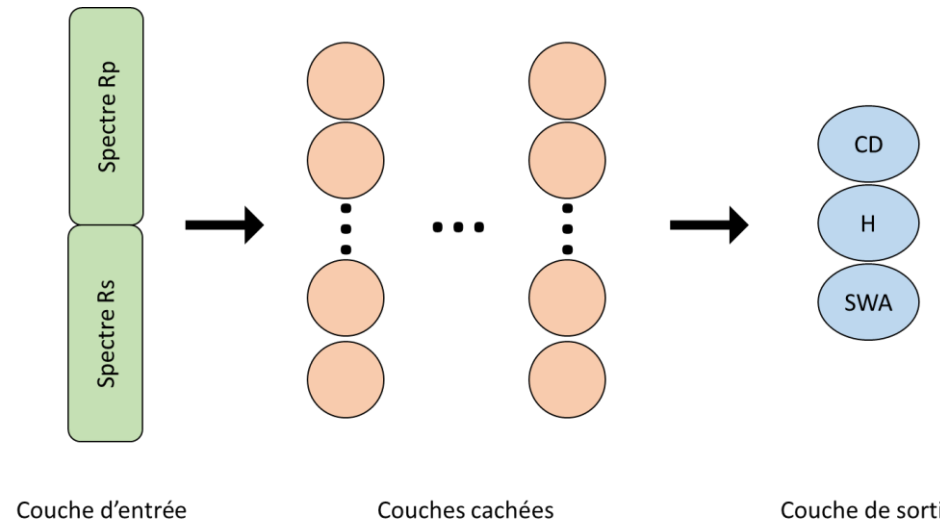

Figure III.20 : Réseau de neurones (MLP) pouvant traiter les données brutes de l'OCD

Comme dans le cas de l'AFM, cette vectorisation impose que tous les spectres soient définis sur un même ensemble de longueurs d'onde. Or, dans la pratique, les mesures présentent souvent de légères variations dans les positions spectrales échantillonnées, ce qui empêche une correspondance directe entre les points. Il est donc nécessaire d'interpoler les points des spectres des données expérimentales afin que les longueurs d'ondes aient les mêmes valeurs, défini au préalable. Cette étape est essentielle pour garantir une cohérence des entrées du réseau de neurones.

III.4.6.4 SAXS

Les données issues des mesures SAXS peuvent se présenter sous différentes formes : images directes des figures de diffraction obtenues à divers angles d'illumination, représentations dans l'espace réciproque, ou encore courbes de diffraction extraite à partir de ces images. Dans cette étude, nous cherchons à minimiser les étapes de traitement afin de préserver au maximum l'information contenue dans les mesures brutes. En effet, chaque opération de transformation ou de réduction de données est susceptible d'entraîner une perte partielle d'information, parfois critique pour les tâches d'apprentissage.

Dans cette optique, l'utilisation des figures de diffraction, extraites à partir d'espace réciproque, est écartée, car trop réductrice. À l'inverse, les représentations dans l'espace réciproque conservent l'essentiel de la richesse structurelle des motifs de diffraction, tout en étant directement exploitables sous forme d'images. Ce format permet de représenter l'information dans un plan 2D, avec une structure interprétable spatialement, ce qui justifie le choix d'une architecture de type réseau de neurones convolutif (CNN) pour leur traitement.

L'usage de CNN est d'autant plus pertinent que ces réseaux sont capables d'extraire automatiquement des motifs visuels complexes à partir d'images, ce qui est particulièrement adapté à la richesse des structures présentes dans les données SAXS. Cette approche est représentée schématiquement dans la Figure III.21.

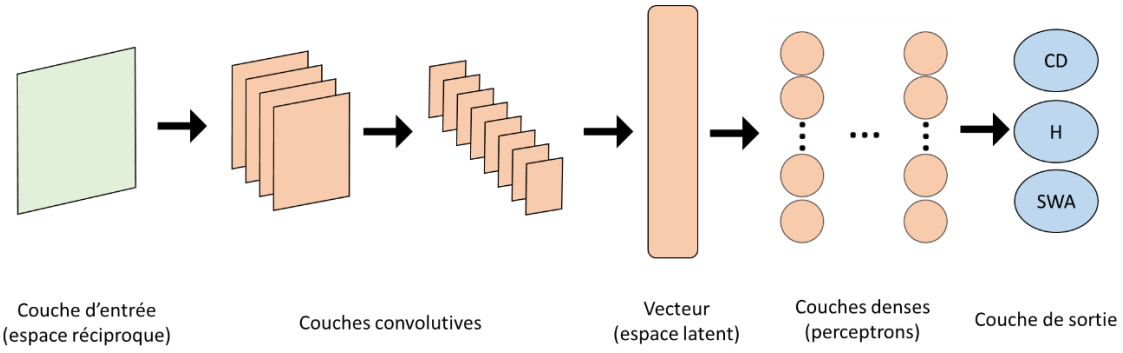

Figure III.21 : Réseau de neurones (CNN) pouvant traiter les données brutes du SAXS

III.4.7 Application des réseaux de neurones à l'extraction de profils depuis les spectres OCD

Des premiers travaux ont été menés sur l'extraction directe de profils à partir de spectres OCD bruts à l'aide de réseaux de neurones [80], [81], [82]. Les données d'entrée correspondent aux spectres Rp et Rs, concaténés en un vecteur unique, et associés aux paramètres dimensionnels de référence (CD, H et SWA) utilisés comme

labels. Les spectres Rp et Rs ont été générés à partir de la modélisation à trois lignes utilisées dans le chapitre précédent. Ces valeurs, naturellement bornées entre 0 et 1, n'ont pas besoin d'être normalisé avant l'apprentissage.

Afin de constituer une base d'apprentissage suffisamment large et représentative, des données réalistes supplémentaires ont été générées selon la méthodologie décrite au Chapitre 2, portant la base initiale de 700 à 3700 échantillons. Cette augmentation est essentielle : les données brutes se caractérisent par des vecteurs de grande dimension (plusieurs centaines de points), ce qui accroît la complexité de leur interprétation et requiert des architectures de réseaux plus profondes, donc un volume de données plus important pour garantir une phase d'apprentissage efficace.

Le réseau de neurones utilisé ici est de type perceptron multicouche (MLP), composé de cinq couches cachées dont le nombre de neurones décroît progressivement : 512, 256, 128, 64 et 32 neurones respectivement. Cette architecture est illustrée dans la Figure III.22. Cette architecture, définie dans le cadre de ce travail, a volontairement été surdimensionnée afin de limiter les risques de sous-apprentissage et de permettre au modèle de capturer la complexité des données d'entrée. Ce choix se justifie par la taille suffisante du jeu de données (2000 échantillons d'entraînement, 1000 de validation et 700 de test), qui autorise l'emploi d'un nombre élevé de neurones sans risque majeur de sur-apprentissage.

Les résultats d'apprentissage, présentés en Figure III.23 , montrent une convergence satisfaisante du modèle, sans apparition de sur-apprentissage. En effet, les courbes de perte associées aux données d'apprentissage et de validation restent proches tout au long de l'entraînement, traduisant une bonne capacité de généralisation du réseau sur ces données synthétiques.

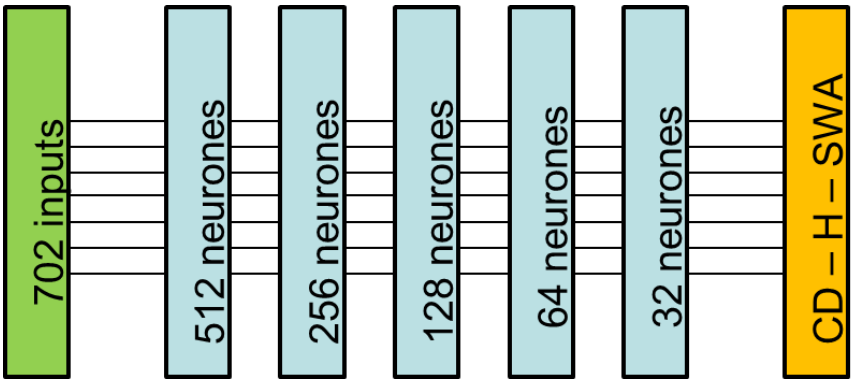

Figure III.22 : Réseau de neurone (MLP 702/512/256/128/64/32/3 avec un learning rate de 0.001 et un batchsize de 100) utilisé pour l'extraction de profils à partir des spectres OCD.

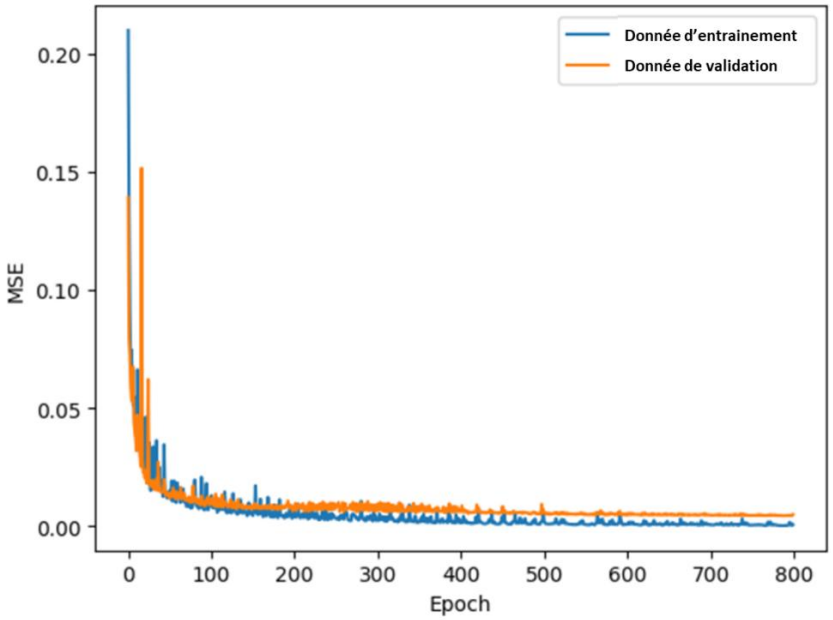

Figure III.23 : Fonctions de pertes du réseau de neurones d'extraction de profils des spectres OCD, 2000 données d'entrainement et 1000 données de validation

À l'issue de l'apprentissage, les erreurs sur les paramètres dimensionnels ont été évaluées. Une comparaison a été réalisée entre les résultats obtenus par extraction via résolution de problème inverse et ceux issus de l'approche par réseau de neurones. Les valeurs associées sont présentées dans le Tableau III.6.

Tout d'abord, on note une disparition des biais dans les prédictions du réseau : les moyennes des erreurs sont nulles, traduisant une prédiction centrée et sans décalage systématique. Ensuite, une réduction significative des écarts-types est observée lors de l'utilisation du réseau de neurones pour les paramètres CD et SWA, indiquant une amélioration notable de l'exactitude sur ces grandeurs. Enfin, on constate une augmentation de l'écart-type pour le paramètre H, qui était auparavant celui présentant les plus faibles incertitudes dans l'approche classique.

Une explication possible de cette augmentation est le déséquilibre déjà évoqué en section III.2, lié à l'entraînement simultané sur plusieurs sorties. La fonction de perte globale, issue de la moyenne des pertes individuelles, peut conduire à privilégier certains paramètres au détriment d'autres, ce qui pourrait expliquer la dégradation observée sur H. Des essais avec une architecture multi-branches permettraient de confirmer cette hypothèse.

Globalement, ces résultats indiquent qu'une extraction directe des paramètres dimensionnels à partir des spectres OCD par réseau de neurones est non seulement réalisable, mais également avantageuse. L'approche permet une réduction de certaines incertitudes, tout en éliminant les biais introduits par la résolution inverse. Dans l'article de Liu et al [82], les auteurs soulignent toutefois que l'apprentissage à partir

de données purement simulées limite la généralisation du modèle aux spectres expérimentaux, en raison du bruit et des imperfections réelles. Dans notre cas, la génération de spectres simulés mais réalistes vise précisément à réduire cet écart entre simulation et expérience. Bien que l'on observe une légère dégradation des performances sur la hauteur, celle-ci reste contenue, et le gain global en temps de calcul et en exactitude rend cette méthode particulièrement prometteuse.

Erreur en fonction de la méthode d'extraction de profil		Résolution de problème inverse	Réseau de neurones
CD	**Ecart-type (σ)**	**2,05 nm**	**0,50 nm**
	Moyenne (μ)	-0,41 nm	-0,02 nm
H	**Ecart-type (σ)**	**0,92 nm**	**1,16 nm**
	Moyenne (μ)	0,13 nm	-0,03 nm
SWA	**Ecart-type (σ)**	**3,49 °**	**0,46°**
	Moyenne (μ)	-1,32 °	-0,03 °

Tableau III.6 : Comparaison des erreurs lors de l'extraction de profils (résolution de problème inverse et réseau de neurones) sur 700 données de test

III.5 Vers une métrologie hybride fondée sur l'apprentissage automatique

Ce chapitre a été consacré à l'étude de l'utilisation des réseaux de neurones pour l'exploitation conjointe de données métrologiques issues de différentes techniques, dans le but d'améliorer l'estimation des paramètres dimensionnels. Dans un premier temps, l'architecture multi-branches a été mise en place et comparé à une architecture dite classique (MLP). Cette nouvelle architecture a été identifiée comme la plus adaptée à ce contexte. Sa structure modulaire permet de traiter chaque grandeur cible de manière indépendante, limitant ainsi les interférences entre les sorties et facilitant l'optimisation de l'apprentissage. Ce choix architectural a constitué la base des développements menés par la suite.

Les travaux menés sur données virtuelles ont permis de déterminer un ordre de grandeur du nombre d'échantillons nécessaires. Cette estimation est essentielle pour garantir un apprentissage stable et pertinent. Elle a ensuite servi de référence pour

guider la génération de données simulées réalistes. Celles-ci intègrent des incertitudes et des bruits représentatifs des conditions expérimentales.

L'hybridation de ces données simulées, combinant plusieurs techniques métrologiques, a permis de démontrer l'intérêt de cette approche. L'architecture proposée s'est révélée capable de produire des prédictions plus précises, sans biais systématique, et avec une réduction des incertitudes sur plusieurs paramètres. Cette validation à partir de données simulées permet de démontrer la faisabilité et la pertinence méthodologique de l'approche proposée. Dans une optique applicative, l'exploitation de jeux de données expérimentales représenterait une étape complémentaire. La combinant d'un pré-entraînement sur des données simulées avec un affinement (fine-tuning), c'est-à-dire un second apprentissage, sur données expérimentales, pourrait accroître la robustesse et la capacité de généralisation aux mesures réelles.

Concernant l'hybridation des données brutes, les travaux réalisés dans ce chapitre se sont limités à une réflexion approfondie sur la manière d'intégrer ces données dans un cadre de fusion via un réseau de neurones. Cette réflexion s'est traduite par la proposition d'une architecture dédiée, combinant des sous-réseaux spécialisés par technique, suivis d'un réseau de fusion avec apprentissage en deux parties. Bien que cette approche n'ait pas été implémentée dans sa globalité, une étude spécifique menée sur les spectres OCD a permis de valider la pertinence de cette voie. Les résultats obtenus montrent que l'extraction directe des paramètres dimensionnels à partir des spectres, sans résolution de problème inverse, est non seulement réalisable, mais aussi plus précise et robuste pour certaines grandeurs.

Ces travaux posent les bases d'une nouvelle approche pour la métrologie dimensionnelle, fondée sur l'apprentissage automatique et l'hybridation des données. À plus long terme, la généralisation de ces méthodes à d'autres types de mesures, ainsi que leur extension à des configurations expérimentales variées, pourrait ouvrir la voie à des chaînes d'analyse métrologique plus flexibles, plus rapides et mieux adaptées aux exigences industrielles. L'intégration progressive des données expérimentales réelles, dans une logique de pré-entraînement simulé suivi d'ajustement fin, constitue une étape clé à envisager dans la continuité de ces travaux.

Conclusions et perspectives

Comme nous l'avons évoqué, la microélectronique moderne se caractérise par la complexité croissante des structures fabriquées, la diversité des matériaux employés et l'exigence accrue en matière d'exactitude, de rapidité et de fiabilité des mesures. Dans un tel contexte, il est acté qu'aucune technique métrologique, prise de manière isolée, ne peut satisfaire l'ensemble des besoins. La problématique posée en introduction, comment améliorer l'exactitude et la robustesse des mesures dans un contexte où les structures deviennent de plus en plus complexes et où les données expérimentales sont limitées, a guidé l'ensemble des travaux présentés dans ce mémoire.

Pour y répondre, une double approche a été mise en œuvre durant ces travaux de thèse :

- **Développement d'une méthodologie de métrologie hybride** reposant sur l'utilisation de réseaux de neurones artificiels multi-branches pour fusionner les résultats de plusieurs techniques de mesure.
- **Mise en place d'une chaîne de génération de données simulées réalistes** à partir de simulateurs adaptés à chaque technique (AFM, CDSEM, OCD et SAXS), intégrant des modèles d'incertitudes spécifiques, afin de pallier le manque de données expérimentales.

Conclusions

La méthodologie globale a reposé sur un travail préalable de modélisation des échantillons et d'identification des sources d'erreurs propres à chaque technique. Ces sources ont ensuite été intégrées aux simulateurs, permettant de générer un volume important de données d'entraînement représentatives des conditions réelles de mesure. Les réseaux de neurones, entraînés sur ces données, ont permis de réaliser l'hybridation des mesures issues de différentes techniques, dans le but de réduire les incertitudes et d'accroître la robustesse des résultats.

Les résultats obtenus sont significatifs :

- **Critical Dimension (CD)** : écart-type réduit à 0,35 nm après hybridation, contre 0,54 nm pour la technique la plus précise prise isolément (SAXS).
- **Side Wall Angle (SWA)** : écart-type réduit à 0,43°, contre 0,67° pour la meilleure technique seule (SAXS).
- **Hauteur (H)** : l'écart-type obtenu est de 0,47 nm. Cette valeur est légèrement meilleure que celle fournie par la technique la plus précise prise isolément (SAXS), dont l'écart-type le plus élevé est de 0,49 nm. Il est à noter que le comportement du SAXS est atypique, avec une double distribution

normale des résultats, ce qui peut influencer la comparaison directe des performances.

Ces gains démontrent que l'hybridation par réseaux de neurones multi-branches permet non seulement d'atteindre mais parfois de dépasser l'exactitude des meilleures techniques individuelles, tout en réduisant drastiquement les biais systématiques identifiés.

Malgré ces résultats encourageants, plusieurs limites doivent être soulignées. La première concerne la **modélisation des incertitudes**, exercice particulièrement complexe en raison de la multiplicité et de la complexité des phénomènes physiques en jeu. Certaines hypothèses simplificatrices ont été introduites lors de la modélisation des échantillons et des incertitudes. Par exemple, la symétrie des profils (SWA gauche et droite identiques) a été imposée. Ces choix, motivés par la nécessité de limiter la complexité des modèles, peuvent engendrer des écarts entre les résultats simulés et les observations expérimentales.

La deuxième limite est liée aux **simulateurs eux-mêmes**, qui bien qu'améliorés pour intégrer les sources d'erreur identifiées, reposent sur des hypothèses et modèles physiques dont la validité peut être restreinte à un domaine spécifique. Une transposition directe à d'autres procédés ou architectures de dispositifs pourrait nécessiter une adaptation importante.

Enfin, la validation expérimentale a principalement été menée sur des données traitées issue de simulateurs (paramètres dimensionnels). Les travaux préliminaires sur l'hybridation à partir de données brutes, qui visent à supprimer l'étape de traitement numérique avant l'hybridation, montrent déjà des résultats prometteurs, comme on a pu le voir avec les données OCD et dans la littérature.

En conclusion, ce travail démontre qu'il est possible d'améliorer significativement l'exactitude et la robustesse des mesures métrologiques en microélectronique grâce à l'hybridation par réseaux de neurones, même dans un contexte de données expérimentales limitées. L'utilisation de simulateurs enrichis par la modélisation des incertitudes permet de constituer des jeux de données réalistes et suffisamment riches pour entraîner des modèles performants.

L'intelligence artificielle n'apparaît pas ici comme un substitut à la métrologie physique, mais comme un complément puissant, capable d'exploiter et de combiner efficacement les informations issues de techniques variées. Ce travail contribue ainsi à ouvrir la voie à une nouvelle génération de méthodes métrologiques, plus flexibles, plus adaptatives et mieux intégrées aux contraintes industrielles, fondées sur la collaboration entre les savoirs physiques et les capacités d'apprentissage des machines.

Perspectives

Les résultats obtenus dans le cadre de cette thèse ouvrent la voie à de nombreuses pistes de recherche et de développement, tant sur le plan scientifique que technologique et industriel.

Sur le plan scientifique, un premier axe de travail concerne l'amélioration de la modélisation des sources d'incertitudes. Les effets complexes, souvent non linéaires, qui interviennent lors des mesures, ne peuvent pas tous être représentés de manière satisfaisante. Il faudrait des modèles physiques plus complets et des approches numériques plus fines, par exemple en utilisant des modèles à plus grand nombre de lignes. L'intégration de ces phénomènes dans les simulateurs existants permettrait de générer des données synthétiques encore plus représentatives des conditions réelles, renforçant ainsi la fiabilité des modèles d'hybridation.

Un second axe scientifique viserait à approfondir la méthodologie développée en l'appliquant à des cas d'étude plus variés au sein des principales techniques de métrologie dimensionnelle nanométrique (AFM, CDSEM, OCD et SAXS). L'objectif serait de vérifier la capacité de l'approche à rester performante pour des structures plus complexes, des matériaux hétérogènes ou des conditions de mesure particulières. Dans ces derniers cas, l'apport de techniques complémentaires comme la spectroscopie photo-électronique (XPS) permettrait d'enrichir l'analyse en reliant les propriétés dimensionnelles à la composition chimique. Une telle démarche contribuerait également à améliorer la combinaison entre différentes techniques.

De plus, l'hybridation à partir de données brutes constitue une troisième piste particulièrement prometteuse. Les nombreux travaux montrent que l'exploitation directe des signaux de mesure, sans passer par les étapes traditionnelles de traitement comme d'extraction de paramètres dimensionnels par résolution de problème inverse, permet d'obtenir une exactitude accrue. Poursuivre cette voie pourrait non seulement améliorer les performances des modèles, mais aussi réduire le temps total de traitement, un avantage considérable dans un contexte industriel à haute cadence.

D'un point de vue technologique, l'intégration des modèles de machine learning au sein de chaînes de mesure en temps réel représente un défi technologique majeur, notamment dans la perspective des environnements de production dits d'usine 4.0. Une telle mise en œuvre nécessite de repenser les flux de données, d'optimiser les temps de calcul et d'assurer une robustesse suffisante pour un fonctionnement continu. Une telle intégration permettrait de s'affranchir des phases de post-traitement, souvent longues et coûteuses, pour fournir directement aux opérateurs ou aux systèmes de contrôle des résultats hybrides exploitables. Du point de vue industriel, la validation de la méthodologie dans des environnements de production à forte cadence constitue une étape indispensable pour son adoption. Tester les

performances du système dans des conditions réelles, en présence de contraintes logistiques, environnementales et économiques, permettrait de confirmer sa robustesse et son efficacité à grande échelle.

Enfin, cette approche s'inscrit dans une vision plus large de la « métrologie augmentée », où les données issues de différentes sources (physiques ou simulées) sont intégrées de manière intelligente dans les systèmes de contrôle et de pilotage des procédés. En exploitant au mieux les forces et la complémentarité des différentes techniques de mesure, il deviendrait possible d'optimiser leur utilisation au sein des lignes de production, en réduisant le nombre de mesures redondantes tout en maintenant, voire en améliorant, l'exactitude globale. Une telle évolution transformerait la métrologie, la faisant passer d'un rôle principalement vérificateur à un rôle prédictif et proactif, contribuant ainsi à la maîtrise et à l'optimisation des procédés de fabrication en microélectronique.

Bibliographies

[1] R. J. C. Brown, « Measuring measurement – What is metrology and why does it matter? », *Measurement*, vol. 168, p. 108408, janv. 2021, doi: 10.1016/j.measurement.2020.108408.

[2] Bureau International des Poids et Mesures, « International vocabulary of metrology – Basic and general concepts and associated terms. », BIPM. Consulté le: 6 mars 2025. [En ligne]. Disponible sur: https://www.bipm.org/en/committees/jc/jcgm/publications

[3] 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 », 1983, Consulté le: 6 mars 2025. [En ligne]. Disponible sur: https://research.wur.nl/en/publications/precision-of-test-methods-determination-of-repeatability-and-repr

[4] « IRDS™ 2022: Metrology - IEEE IRDS™ ». Consulté le: 6 mars 2025. [En ligne]. Disponible sur: https://irds.ieee.org/editions/2022/irds%E2%84%A2-2022-metrology

[5] G. Binnig, « Atomic Force Microscope », *Phys. Rev. Lett.*, vol. 56, n° 9, p. 930-933, 1986, doi: 10.1103/PhysRevLett.56.930.

[6] R. Bowen et N. Hilal, *Atomic Force Microscopy in Process Engineering: An Introduction to AFM for Improved Processes and Products*. Butterworth-Heinemann, 2009.

[7] A. Trache et G. A. Meininger, « Atomic Force Microscopy (AFM) », *Curr. Protoc. Microbiol.*, vol. 8, n° 1, févr. 2008, doi: 10.1002/9780471729259.mc02c02s8.

[8] R. Dixson, J. Fu, N. Orji, W. Guthrie, R. Allen, et M. Cresswell, « CD-AFM reference metrology at NIST and SEMATECH », présenté à Microlithography 2005, R. M. Silver, Éd., San Jose, CA, mai 2005, p. 324. doi: 10.1117/12.601972.

[9] T. A. Germer, H. J. Patrick, R. M. Silver, et B. Bunday, « Developing an uncertainty analysis for optical scatterometry », in *Metrology, Inspection, and Process Control for Microlithography XXIII*, SPIE, mars 2009, p. 255-265. doi: 10.1117/12.814835.

[10] J. Reche, « Nouvelle méthodologie hybride pour la mesure de rugosités sub-nanométriques », p. 157.

[11] K. D. Vernon-Parry, « Scanning electron microscopy: an introduction », *III-Vs Rev.*, vol. 13, n° 4, p. 40-44, juill. 2000, doi: 10.1016/S0961-1290(00)80006-X.

[12] J. S. Villarrubia, A. E. Vladar, B. D. Bunday, et M. Bishop, « Dimensional metrology of resist lines using a SEM model-based library approach », présenté à Microlithography 2004, R. M. Silver, Éd., Santa Clara, CA, mai 2004, p. 199. doi: 10.1117/12.536871.

[13] A. J. Cepler, B. Bunday, B. L. Thiel, et J. S. Villarrubia, « Scanning electron microscopy imaging of ultra-high aspect ratio hole features », in *Metrology, Inspection, and Process Control for Microlithography XXVI*, SPIE, avr. 2012, p. 562-575. doi: 10.1117/12.916552.

[14] Ya. M. Blanter et M. Büttiker, « Shot noise in mesoscopic conductors », *Phys. Rep.*, vol. 336, n° 1, p. 1-166, sept. 2000, doi: 10.1016/S0370-1573(99)00123-4.

[15] C. J. Wu, W.-S. Huang, K. R. Chen, C. N. Archie, et M. E. Lagus, « Investigation on the mechanism of the 193-nm resist linewidth reduction during the SEM measurement », in *Advances in Resist Technology and Processing XVIII*, SPIE, août 2001, p. 190-199. doi: 10.1117/12.436848.

[16] « Mechanistic Studies on the CD Degradation of 193nm Resists during SEM Inspection ». Consulté le: 6 mars 2025. [En ligne]. Disponible sur: https://www.jstage.jst.go.jp/article/photopolymer1988/14/3/14_3_407/_article
[17] « Radiation damage in the TEM and SEM », *Micron*, vol. 35, nº 6, p. 399-409, août 2004, doi: 10.1016/j.micron.2004.02.003.
[18] K. Levenberg, « A Method for the Solution of Certain Non-Linear Problems in Least Squares », *Q. Appl. Math.*, vol. 2, nº 2, p. 164-168, 1944.
[19] D. W. Marquardt, « An Algorithm for Least-Squares Estimation of Nonlinear Parameters », *J. Soc. Ind. Appl. Math.*, vol. 11, nº 2, p. 431-441, 1963.
[20] J. A. Nelder et R. Mead, « A Simplex Method for Function Minimization », *Comput. J.*, vol. 7, nº 4, p. 308-313, janv. 1965, doi: 10.1093/comjnl/7.4.308.
[21] T. Choisnet *et al.*, « Critical dimension measurement: from synchrotron small angle x-ray scattering to industrial optical scatterometry techniques », in *Metrology, Inspection, and Process Control XXXVII*, SPIE, avr. 2023, p. 428-437. doi: 10.1117/12.2657661.
[22] R. L. Jones *et al.*, « Small angle x-ray scattering for sub-100 nm pattern characterization », *Appl. Phys. Lett.*, vol. 83, nº 19, p. 4059-4061, nov. 2003, doi: 10.1063/1.1622793.
[23] T. Hu *et al.*, « Small angle x-ray scattering metrology for sidewall angle and cross section of nanometer scale line gratings », *J. Appl. Phys.*, vol. 96, nº 4, p. 1983-1987, août 2004, doi: 10.1063/1.1773376.
[24] « Auger catalog calculated transition energies listed by energy and element », *At. Data Nucl. Data Tables*, vol. 5, nº 4, p. 317-469, janv. 1973, doi: 10.1016/S0092-640X(73)80005-1.
[25] A. H. Compton, « A Quantum Theory of the Scattering of X-rays by Light Elements », *Phys. Rev.*, vol. 21, nº 5, p. 483-502, mai 1923, doi: 10.1103/PhysRev.21.483.
[26] L. Boldon, F. Laliberte, et L. Liu, « Review of the fundamental theories behind small angle X-ray scattering, molecular dynamics simulations, and relevant integrated application », *Nano Rev.*, vol. 6, p. 10.3402/nano.v6.25661, févr. 2015, doi: 10.3402/nano.v6.25661.
[27] G. Freychet, « Analyses morphologiques et dimensionnelles de nanostructures organisées par diffusion centrale des rayons X ».
[28] V. Ukraintsev, « Review of reference metrology for nanotechnology: significance, challenges, and solutions », *J. MicroNanolithography MEMS MOEMS*, vol. 11, nº 1, p. 011010, févr. 2012, doi: 10.1117/1.JMM.11.1.011010.
[29] G. A. Dahlen *et al.*, « TEM validation of CD AFM image reconstruction », présenté à Advanced Lithography, C. N. Archie, Éd., San Jose, CA, mars 2007, p. 651818. doi: 10.1117/12.711943.
[30] J. J. Hwu, T. J. Pham, S. Dulay, A. Lopez, et P. Wilkens, « The estimation of total measurement uncertainty in a multiple metrology tool environment », présenté à Microlithography 2004, R. M. Silver, Éd., Santa Clara, CA, mai 2004, p. 413. doi: 10.1117/12.535526.
[31] R. M. Silver *et al.*, « Improving optical measurement accuracy using multi-technique nested uncertainties », présenté à SPIE Advanced Lithography, J. A. Allgair et C. J. Raymond, Éd., San Jose, California, USA, mars 2009, p. 727202. doi: 10.1117/12.816569.

[32] H. Yang, Z. Wu, K. Zhang, D. Wang, et H. Yu, « Uncertainty quantification on small angle x-ray scattering measurement using Bayesian deep learning », *J. Appl. Phys.*, vol. 136, nº 14, p. 143101, oct. 2024, doi: 10.1063/5.0226735.

[33] S.-B. Yoo *et al.*, « Automated measurement and analysis of sidewall roughness using three-dimensional atomic force microscopy », *Appl. Microsc.*, vol. 52, p. 1, mars 2022, doi: 10.1186/s42649-022-00070-5.

[34] J. J. Faria-Briceno, V. Sasidharan, A. Neumann, S. Singhal, S. V. Sreenivasan, et S. R. J. Brueck, « High-speed in-line optical angular scatterometer for high-throughput roll-to-roll nanofabrication », in *Novel Patterning Technologies 2021*, E. M. Panning et J. A. Liddle, Éd., Online Only, United States: SPIE, févr. 2021, p. 44. doi: 10.1117/12.2593730.

[35] M. Wormington *et al.*, « X-ray critical dimension metrology solution for high aspect ratio semiconductor structures », in *Metrology, Inspection, and Process Control for Semiconductor Manufacturing XXXV*, SPIE, févr. 2021, p. 150-159. doi: 10.1117/12.2583966.

[36] K. Ueda, T. Mizuno, et K. Setoguchi, « High accuracy CD matching monitor for CD-SEM beyond 20nm process », présenté à SPIE Advanced Lithography, A. Starikov et J. P. Cain, Éd., San Jose, California, USA, avr. 2013, p. 868135. doi: 10.1117/12.2011389.

[37] « Transmission Electron Microscopy (TEM) », in *Membrane Characterization*, Elsevier, 2017, p. 145-159. doi: 10.1016/B978-0-444-63776-5.00008-5.

[38] D. B. Williams et C. B. Carter, « Imaging in the TEM », in *Transmission Electron Microscopy: A Textbook for Materials Science*, D. B. Williams et C. B. Carter, Éd., Boston, MA: Springer US, 1996, p. 349-366. doi: 10.1007/978-1-4757-2519-3_22.

[39] N. G. Schuch, « Hybrid Metrology Applied to dimensional Control in Lithography », phdthesis, Université Grenoble Alpes, 2017. Consulté le: 6 mars 2025. [En ligne]. Disponible sur: https://theses.hal.science/tel-01725224

[40] M.-A. Henn *et al.*, « Optimizing hybrid metrology: rigorous implementation of Bayesian and combined regression », *J. MicroNanolithography MEMS MOEMS*, vol. 14, nº 4, p. 044001, nov. 2015, doi: 10.1117/1.JMM.14.4.044001.

[41] J. Hazart *et al.*, « Data fusion for CD metrology: heterogeneous hybridization of scatterometry, CDSEM, and AFM data », in *Metrology, Inspection, and Process Control for Microlithography XXVIII*, SPIE, avr. 2014, p. 688-701. doi: 10.1117/12.2046484.

[42] N. F. Zhang, R. M. Silver, H. Zhou, et B. M. Barnes, « Improving optical measurement uncertainty with combined multitool metrology using a Bayesian approach », *Appl. Opt.*, vol. 51, nº 25, p. 6196-6206, sept. 2012, doi: 10.1364/AO.51.006196.

[43] F. Castanedo, « A Review of Data Fusion Techniques », *Sci. World J.*, vol. 2013, nº 1, p. 704504, janv. 2013, doi: 10.1155/2013/704504.

[44] P. Digraci, M. Besacier, P. Gergaud, G. Rademaker, et J. Rêche, « Multi-branch neural network for hybrid metrology improvement », in *Metrology, Inspection, and Process Control XXXVI*, J. C. Robinson et M. J. Sendelbach, Éd., San Jose, United States: SPIE, mai 2022, p. 46. doi: 10.1117/12.2612798.

[45] M. E. Kodadi, « Développement des techniques de scatterométrie en temps réel pour le suivi des procédés de gravure plasma », phdthesis, Université de Grenoble, 2010. Consulté le: 8 juin 2023. [En ligne]. Disponible sur: https://theses.hal.science/tel-00670095

[46] R. Wagner, R. Moon, J. Pratt, G. Shaw, et A. Raman, « Uncertainty quantification in nanomechanical measurements using the atomic force microscope », *Nanotechnology*, vol. 22, n° 45, p. 455703, nov. 2011, doi: 10.1088/0957-4484/22/45/455703.
[47] H. Gross, M.-A. Henn, S. Heidenreich, A. Rathsfeld, et M. Bär, « Modeling of line roughness and its impact on the diffraction intensities and the reconstructed critical dimensions in scatterometry », *Appl. Opt.*, vol. 51, n° 30, p. 7384, oct. 2012, doi: 10.1364/AO.51.007384.
[48] « OpenCV ». [En ligne]. Disponible sur: https://docs.opencv.org/4.10.0/
[49] « correlate2d — SciPy v1.15.2 Manual ». Consulté le: 18 mars 2025. [En ligne]. Disponible sur: https://docs.scipy.org/doc/scipy/reference/generated/scipy.signal.correlate2d.html
[50] N. Metropolis et S. Ulam, « The Monte Carlo Method », *J. Am. Stat. Assoc.*, vol. 44, n° 247, p. 335-341, sept. 1949, doi: 10.1080/01621459.1949.10483310.
[51] D. Wales et J. Doye, « Global Optimization by Basin-Hopping and the Lowest Energy Structures of Lennard-Jones Clusters Containing up to 110 Atoms », *J. Phys. Chem. A*, vol. 101, n° 28, p. 5111-5116, juill. 1997, doi: 10.1021/jp970984n.
[52] M. Iwamatsu et Y. Okabe, « Basin hopping with occasional jumping », *Chem. Phys. Lett.*, vol. 399, n° 4-6, p. 396-400, déc. 2004, doi: 10.1016/j.cplett.2004.10.032.
[53] Y. Zhao, X. Chen, et J. Li, « TGMin: A global-minimum structure search program based on a constrained basin-hopping algorithm », *Nano Res.*, vol. 10, n° 10, p. 3407-3420, oct. 2017, doi: 10.1007/s12274-017-1553-z.
[54] F. Glover, « Tabu Search: A Tutorial », *Interfaces*, vol. 20, n° 4, p. 74-94, août 1990, doi: 10.1287/inte.20.4.74.
[55] A. Hertz, E. Taillard, et D. de Werra, « 5. Tabu search », in *Local Search in Combinatorial Optimization*, E. Aarts et J. K. Lenstra, Éd., Princeton University Press, 2018, p. 121-136. doi: 10.1515/9780691187563-008.
[56] M. Armatte, « Robert Gibrat et la loi de l'effet proportionnel », *Mathématiques Inform. Sci. Hum.*, vol. 129, p. 5-35, 1995.
[57] « SciPy API — SciPy v1.15.1 Manual ». Consulté le: 21 janvier 2025. [En ligne]. Disponible sur: https://docs.scipy.org/doc/scipy/reference/index.html
[58] *Nebula - Github*. (12 novembre 2024). C++. Nebula-simulator. Consulté le: 18 novembre 2024. [En ligne]. Disponible sur: https://github.com/Nebula-simulator/Nebula
[59] L. Van Kessel et C. W. Hagen, « Nebula: Monte Carlo simulator of electron–matter interaction », *SoftwareX*, vol. 12, p. 100605, juill. 2020, doi: 10.1016/j.softx.2020.100605.
[60] W. Schottky, « Über spontane Stromschwankungen in verschiedenen Elektrizitätsleitern », *Ann. Phys.*, vol. 362, n° 23, p. 541-567, janv. 1918, doi: 10.1002/andp.19183622304.
[61] W. Schottky, « On spontaneous current fluctuations in various electrical conductors », *J. MicroNanolithography MEMS MOEMS*, vol. 17, n° 4, p. 041001, oct. 2018, doi: 10.1117/1.JMM.17.4.041001.
[62] É. Sezestre *et al.*, « Model-based contour extraction: an enabler for very low-frame SEM images metrology », in *Metrology, Inspection, and Process Control XXXVI*, SPIE, mai 2022, p. 471-486. doi: 10.1117/12.2616527.

[63] D. Fuard, N. Troscompt, I. El Kalyoubi, S. Soulan, et M. Besacier, « S-Genius, a universal software platform with versatile inverse problem resolution for scatterometry », présenté à SPIE Optical Metrology 2013, B. Bodermann, K. Frenner, et R. M. Silver, Éd., Munich, Germany, mai 2013, p. 878919. doi: 10.1117/12.2020674.

[64] R. Smaali, M. Besacier, et P. Schiavone, « Three-dimensional rigorous simulation of EUV defective masks using modal method by Fourier expansion », présenté à SPIE 31st International Symposium on Advanced Lithography, M. J. Lercel, Éd., San Jose, CA, mars 2006, p. 615124. doi: 10.1117/12.655495.

[65] S. Essig et K. Busch, « Generation of adaptive coordinates and their use in the Fourier Modal Method », *Opt. Express*, vol. 18, n° 22, p. 23258, oct. 2010, doi: 10.1364/OE.18.023258.

[66] J. J. Hench et Z. K. Strakoš, « THE RCWA METHOD - A CASE STUDY WITH OPEN QUESTIONS AND PERSPECTIVES OF ALGEBRAIC COMPUTATIONS », 2008.

[67] Y. Wang, « Gauss–Newton method », *WIREs Comput. Stat.*, vol. 4, n° 4, p. 415-420, 2012, doi: 10.1002/wics.1202.

[68] S. M. Sedlak, L. K. Bruetzel, et J. Lipfert, « Quantitative evaluation of statistical errors in small-angle X-ray scattering measurements », *J. Appl. Crystallogr.*, vol. 50, n° 2, p. 621-630, avr. 2017, doi: 10.1107/S1600576717003077.

[69] J. W. Tukey, « The Future of Data Analysis », in *Breakthroughs in Statistics: Methodology and Distribution*, S. Kotz et N. L. Johnson, Éd., New York, NY: Springer, 1992, p. 408-452. doi: 10.1007/978-1-4612-4380-9_31.

[70] X. Glorot, A. Bordes, et Y. Bengio, « Deep Sparse Rectifier Neural Networks », in *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, JMLR Workshop and Conference Proceedings, juin 2011, p. 315-323. Consulté le: 31 août 2025. [En ligne]. Disponible sur: https://proceedings.mlr.press/v15/glorot11a.html

[71] Y. LeCun, Y. Bengio, et G. Hinton, « Deep learning », *Nature*, vol. 521, n° 7553, p. 436-444, mai 2015, doi: 10.1038/nature14539.

[72] L. Penlap Woguia, J. Rêche, M. Besacier, P. Gergaud, et G. Rademaker, « Data fusion by artificial neural network for hybrid metrology development », in *Metrology, Inspection, and Process Control for Semiconductor Manufacturing XXXV*, O. Adan et J. C. Robinson, Éd., Online Only, United States: SPIE, févr. 2021, p. 84. doi: 10.1117/12.2583590.

[73] H. Zhang, J. Shao, et R. Salakhutdinov, « Deep Neural Networks with Multi-Branch Architectures Are Less Non-Convex », *ArXiv180601845 Cs Stat*, juin 2018, Consulté le: 18 février 2022. [En ligne]. Disponible sur: http://arxiv.org/abs/1806.01845

[74] D. E. Knuth, *Art of Computer Programming, Volumes 3*, 1st éd. Addison-Wesley Professional., 2022. Consulté le: 31 août 2025. [En ligne]. Disponible sur: https://www.informit.com/store/art-of-computer-programming-volumes-1-4b-boxed-set-9780137935109

[75] A. Nowak-Vila, D. Folqué, et J. Bruna, « Divide and Conquer Networks », 14 octobre 2018, *arXiv*: arXiv:1611.02401. doi: 10.48550/arXiv.1611.02401.

[76] S. Xie, R. Girshick, P. Dollar, Z. Tu, et K. He, « Aggregated Residual Transformations for Deep Neural Networks », in *2017 IEEE Conference on Computer Vision and Pattern*

Recognition (CVPR), Honolulu, HI: IEEE, juill. 2017, p. 5987-5995. doi: 10.1109/CVPR.2017.634.

[77] F. Chollet, « Xception: Deep Learning With Depthwise Separable Convolutions », présenté à Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, p. 1251-1258. Consulté le: 26 mai 2025. [En ligne]. Disponible sur: https://openaccess.thecvf.com/content_cvpr_2017/html/Chollet_Xception_Deep_Learning_CVPR_2017_paper.html

[78] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, et K. Keutzer, « SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5MB model size », 4 novembre 2016, *arXiv*: arXiv:1602.07360. doi: 10.48550/arXiv.1602.07360.

[79] « Fiddle CNRS - CNN ». Consulté le: 16 juin 2025. [En ligne]. Disponible sur: https://www.fidle.cnrs.fr/w3/archives/07-CNN.html

[80] T. Alcaire, « Approches métrospection pour la détection de dérives de procédés », phdthesis, Université Grenoble Alpes, 2023. Consulté le: 16 juin 2025. [En ligne]. Disponible sur: https://hal.science/tel-04992193

[81] B. M. Barnes et M.-A. Henn, « Contrasting conventional and machine learning approaches to optical critical dimension measurements », in *Metrology, Inspection, and Process Control for Microlithography XXXIV*, O. Adan et J. C. Robinson, Éd., San Jose, United States: SPIE, mars 2020, p. 40. doi: 10.1117/12.2551504.

[82] S. Liu *et al.*, « Machine learning aided solution to the inverse problem in optical scatterometry », *Measurement*, vol. 191, p. 110811, mars 2022, doi: 10.1016/j.measurement.2022.110811.