



HAL
open science

Sciences de données pour la microélectronique : analyse de topographie

Mehdi Kessar

► **To cite this version:**

Mehdi Kessar. Sciences de données pour la microélectronique : analyse de topographie. Modélisation et simulation. Université Grenoble Alpes [2020-..], 2021. Français. NNT : 2021GRALM075 . tel-03702304

HAL Id: tel-03702304

<https://theses.hal.science/tel-03702304v1>

Submitted on 23 Jun 2022

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

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

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ GRENOBLE ALPES

Spécialité : **Mathématiques appliquées**

Arrêté ministériel : 25 mai 2016

Présentée par

Mehdi KESSAR

Thèse dirigée par **Pierre LEMAIRE**, Maître de conférences,
Université Grenoble-Alpes, et
codirigée par **Bertrand LE-GRATIET**, Docteur en sciences,
STMicroelectronics

préparée au sein du **Laboratoire G-SCOP**
dans l'**École Doctorale MSTII**

Sciences de données pour la microélectronique : analyse de topographie

Thèse soutenue publiquement le «**14/06/2021**»,
devant le jury composé de:

M. Bruno AGARD

Professeur, Polytechnique Montréal, Rapporteur

M. Claude YUGMA

Professeur, Mines Saint-Étienne, Rapporteur

M. Jérôme MALICK

Directeur de recherche, Université Grenoble-Alpes, Président

M. Jean-Hervé TORTAI

Chargé de recherche, CNRS, Examineur

Mme. Delphine LE CUNFF

Docteur en sciences, STMicroelectronics, Invitée



Thèse Mehdi KESSAR
Sciences de données pour la micro-électronique:
Analyse de topographie

Mehdi KESSAR

25 Février 2021

Table des matières

Remerciements	3
Introduction	4
1 Des données et des nouveaux métiers	5
1.1 Loi de Moore, volumétrie de données et nouveaux métiers	5
1.1.1 Loi de Moore et conséquence industrielle	5
1.1.2 Flot de fabrication et données volumineuses	7
1.1.3 Des nouveaux métiers	9
1.2 De la métrologie	12
1.2.1 Intérêt de la mesure	12
1.2.2 Définitions	12
1.2.3 Métrologie avancée	14
1.2.4 Métrologie In-Line et cartes de contrôle	18
1.2.4.1 Quels choix techniques pour quels objectifs?	18
1.2.4.2 Capabilité procédé et outil	19
1.2.4.3 Cartes de contrôle	21
2 Des couches minces et de la topographie	22
2.1 Des puces sur un wafer	22
2.2 Micro-électronique et couches minces	24
2.3 Photolithographie	26
2.4 Topographie	30
2.4.1 Origines de la topographie	30
2.4.2 Caractérisation et atténuation de la topographie	31
2.5 Conclusion	32
3 Données PWG : Topographie et usage seul	33
3.1 Présentation de l'outil PWG de KLA pour mesurer la topographie	33
3.1.1 Technique de mesure	33
3.1.2 Volumétrie, précision	34
3.1.3 Traitement de la mesure : Filtre double Gaussien	35
3.1.4 Usages de la mesure	36
3.2 La mesure de topographie en tant que telle	38
3.2.1 Introduction : Enjeux de la description de mesure	38
3.2.2 Des puces sur un wafer	38
3.2.2.1 Méta-données de la mesure	39
3.2.2.2 Puces, champs complets et alignement	40
3.2.2.3 Objets types et objets atypiques	42
3.3 Enrichissement de topographie : Cellules, détection de formes et variabilité	46
3.3.1 Cellules	47

3.3.2	Détection de forme et caractérisation de cellule	47
3.3.3	Gestion de mémoire	49
3.4	Cas d'étude : Choix de slurry pour l'atelier CMP	50
3.5	Conclusion	58
4	Modèle nano-topographie	60
4.1	Présentation du problème	60
4.2	Présentation des données utilisées	61
4.2.1	Données puces : Puces moyennes	61
4.2.2	Données design : Densités périmétriques et surfaciques	61
4.2.3	Cas d'étude	64
4.2.4	Corrélations et distributions entre les données	68
4.3	Méthodologie et choix du modèle	70
4.3.1	Quels modèles et pour quels objectifs?	70
4.3.2	Un premier modèle	72
4.3.3	Évaluation des modèles	74
4.3.4	Modèles avec enrichissement sur les entrées	76
4.3.5	Sélection des vecteurs entrées du modèle	82
4.4	Des modèles pour des puces	85
4.4.1	Confiance et robustesse	85
4.4.2	Modélisation et inférence sur une puce	86
4.4.3	Modélisation et inférence sur trois puces	92
4.5	Critères de confiance a priori et de qualité a posteriori	98
4.5.1	Introduction	99
4.5.2	Distances entre distributions	99
4.5.3	Résultats et discussion	103
4.6	Conclusion	107
5	Détection de défauts : Striations	108
5.1	Des défauts dans la vie d'un produit	108
5.2	Cas d'étude : Striations sur capteurs optiques	109
5.2.1	Présentation du problème	109
5.2.2	Préparation des données	111
5.2.3	Mise en valeur des striations	114
5.2.4	Feature engineering	116
5.3	Conclusion	117
	Conclusion	119
	A Échantillonnage aléatoire	122
	B Détails de calcul pour les transformations	125
	C Échantillonnage pour prédiction	127
	D Classification hiérarchique	128
	Bibliographie	129

Remerciements

Je remercie les membres du jury pour l'évaluation de ce travail.

Merci à l'équipe OPC et Mentor, pour les deux premières années passées à ST. Merci à Charlotte et à Pascal, pour ce que j'ai appris avec vous. Merci à l'équipe métrologie de ST pour leur collaboration avec ce projet de thèse : Victor, Delphine, Thomas et Virginie.

Merci au laboratoire G-SCOP et à STMicroelectronics pour avoir initié et assuré le déroulement de cette thèse.

Enfin, mes remerciements vont à mes deux encadrants, Bertrand et Pierre, pour leur fort soutien lors de cette thèse, particulièrement pour cette troisième année au cours de la rédaction. Vous avez été une base solide d'appui durant ce travail de thèse, et pour cela je vous remercie très sincèrement.

Introduction

Les progrès de la micro-électronique ainsi que des avancées algorithmiques ont largement influencé les domaines de l'informatique dans les dernières décennies. La progression exponentielle de l'industrie de la micro-électronique énoncée par la loi de Moore se fait au prix d'une complexification des procédés de fabrication et d'une augmentation du nombre d'étapes de fabrication.

Cette complexification génère une augmentation du nombre de contrôles et de mesures pour assurer une production fiable, donc du volume des données stockées. Le besoin des compétences en traitement de données est grandissant, et l'intégration de ces compétences est en cours. Il est désormais possible de stocker de gros volumes de données et de les traiter grâce aux progrès de cette même industrie.

Cette thèse s'inscrit dans une dynamique qui vise à intégrer d'avantage de compétences en sciences de données dans des secteurs de la micro-électronique comme les ateliers de gestion des procédés. L'enjeu est d'utiliser les outils de fabrication et de mesure au plus haut de leurs capacités.

Des interactions entre les procédés existent tout au long de la fabrication des puces électroniques. Ces interactions sont mises en évidence par des outils de mesures de plus en plus performants. Lorsqu'un outil de mesure est intégré sur un site de fabrication comme ST-Microelectronics, un travail de fond de traitement de données est nécessaire pour valoriser les mesures produites.

Ce travail consiste d'une part à décomposer l'information contenue dans une mesure, et d'autre part à établir des liens déjà identifiés ou non entre mesures. Établir ces liens entre mesures nécessite des compétences pour manipuler les entités informatiques (les données), et/ou des connaissances sur les entités physiques (les produits).

Ces entités sont différentes dans leurs formes, mais semblables en ce qu'elles représentent. Ce travail d'analyse est d'une part une façon de cristalliser des connaissances physiques déjà établies, et d'autre part de mieux comprendre les procédés en identifiant des liens entre données.

La prise de position de la thèse est double. Elle prendra en compte les enjeux métier d'une industrie qui se doit d'être pérenne, avec des échelles temporelles étendues et subissant des contraintes physiques fortes. Elle prendra également en compte les enjeux académiques des mathématiques appliquées ne subissant que très peu les contraintes physiques, et dont beaucoup de problèmes possèdent déjà des solutions implémentées.

La problématique est la suivante : valoriser des mesures d'un outil acheté par STMicroelectronics, le PWG de KLA Tencor et ses mesures de topographie. Cette thèse fera d'abord état de l'émergence et de la nécessité d'intégrer de nouvelles compétences dans l'industrie de la micro-électronique (Chapitre 1), puis elle présentera les données qui seront utilisées pour répondre à la problématique (Chapitre 2). Dans le Chapitre 3, on valorise les mesures en tant que telles, par des mises en formes, ainsi que des méthodes de description de topographie. Dans le Chapitre 4, on cherchera à faire un modèle prédictif établissant un lien entre le design d'une puce et sa topographie. Dans le Chapitre 5, on s'intéressera à un problème de détection de défauts qui apparaissent sur les mesures : les striations.

Chapitre 1

Des données et des nouveaux métiers

Au cours de ce chapitre, nous abordons les points suivants :

1. L'augmentation de volumétrie de données liée à la Loi de Moore à la fois d'un point de vue de la densité d'information contenue dans une seule mesure comme du nombre de mesures.
2. La métrologie suit les besoins grandissants de l'industrie des semi-conducteurs avec de la métrologie avancée.
3. Le besoin d'inclure des compétences de traitement de données dans l'industrie de la micro-électronique est essentiel, et est au cœur de la motivation de cette thèse.

1.1 Loi de Moore, volumétrie de données et nouveaux métiers

1.1.1 Loi de Moore et conséquence industrielle

La loi de Moore, énoncée en 1975 par Gordon E. Moore [Moore, 2006] nous dit que la densité des transistors sur les semi-conducteurs proposés en entrée de gamme doublerait tous les deux ans. Cette loi fut étonnamment juste de 1971 à aujourd'hui (2020), comme le montre la figure 1.1.

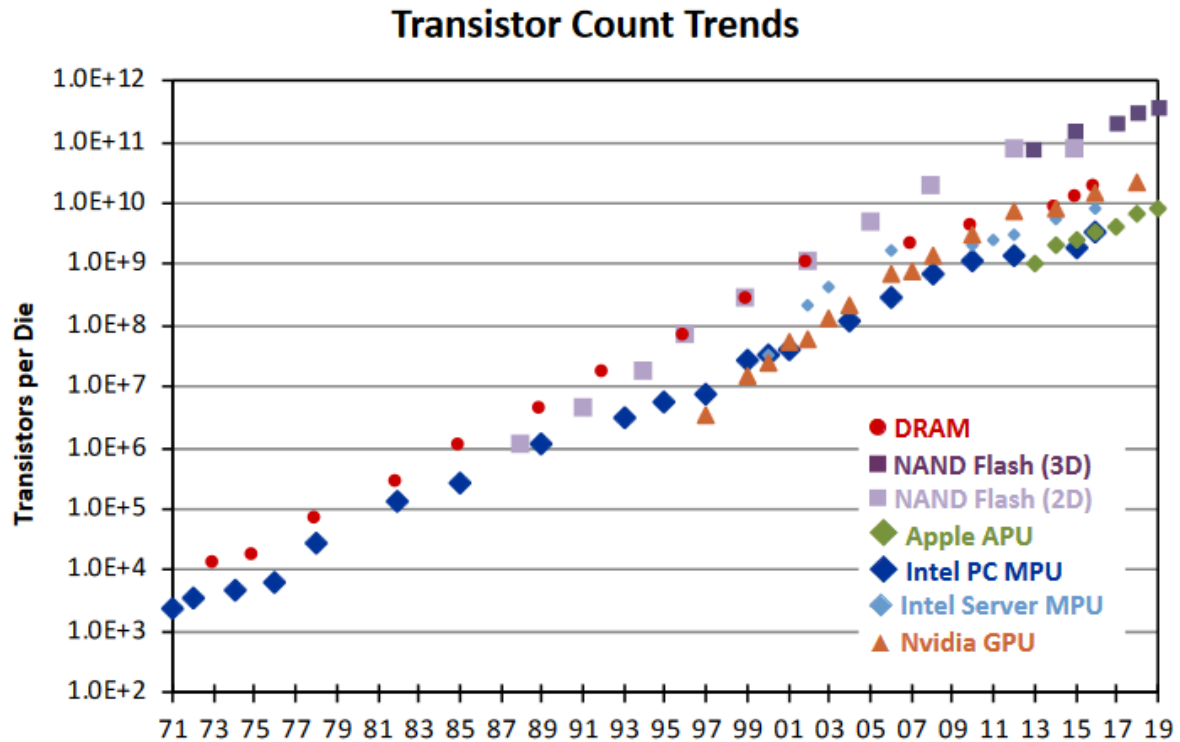


FIGURE 1.1 – Évolution du nombre de transistors sur les semi-conducteurs selon l'année [McClean, 2020]

On voit sur cette figure les différentes technologies poussant la densité des transistors. Ces applications ont des besoins de plus en plus conséquents. Pour le stockage des données, les mémoires doivent être de plus en plus compactes, et étant donné les besoins de calcul sur ces données, on doit également avoir des architectures de processeurs adaptées pour traiter ces volumes de données. Ces besoins imbriqués génèrent une loi de Moore dans plusieurs produits finis de la micro-électronique :

1. DRAM : Dynamic Random Access Memory. Mémoire vive, volatile, basée sur des condensateurs dont l'information ne se conserve pas si la charge des condensateurs n'est pas maintenue pendant quelques millisecondes.
2. NAND Flash : Mémoire flash basée sur la structure électronique NAND. Mémoire vive, mais non volatile, réputée pour sa faible consommation d'énergie et sa possibilité de densité élevée permettant un stockage de masse. On parle de NAND Flash **3D** par rapport à la technique de fabrication consistant à créer une structure en 3 dimensions, mais son utilité logique est la même.
3. APU : Accelerated Protection Unit. Ce sont des processeurs dédiés à certaines opérations. Exemple Accélération 3D, une architecture spécialisée pour faire des calculs sur des points en 3D. L'intérêt est que le CPU attribue ces tâches à ces accélérateurs pour aller plus vite.
4. MPU : Memory Protection Unit. Une protection hardware qui permet de gérer l'accès mémoire.
5. GPU : Graphics Processing Unit. Processeurs initialement conçus pour la gestion de graphisme, ayant une architecture adaptée pour les calculs parallèles. Ces processeurs sont aujourd'hui utilisés dans des domaines où le calcul parallèle est utile, notamment l'entraînement des réseaux de neurones profonds (deep learning).

Cette loi est anecdotique pour les consommateurs, car eux n'ont qu'une visibilité sur les produits finis ayant un nombre de transistors croissant exponentiellement, ce qui implique une

performance croissante de la même façon. Du côté de l'industrie, cela implique une complexité des procédés de fabrication qui est elle aussi exponentielle.

La réduction des dimensions et la multiplication du nombre de transistors sur une puce est un challenge double : d'une part la prouesse technologique de réaliser un transistor de dimension petite, d'autre part la confiance d'être capable de réaliser ces transistors avec une répétabilité constituant une industrie viable. Aujourd'hui, en 2021, l'industrie prépare des transistors de 5nm de largeur, contre 45nm en 2010, et en 2021 les puces comportent plusieurs dizaines de milliards de transistors, contre un peu moins d'un milliard en 2010. Les deux points suivants sensibilisent aux deux enjeux évoqués de complexité et de répétabilité en les reliant à des notions connues de tous :

- Concernant la dimension des transistors récents, il est très facile d'écrire $5 \times 10^{-9}m$ pour exprimer la taille de ceux-ci. En revanche, pour se représenter cette dimension, et à quel point elle est petite, la comparaison avec l'épaisseur d'un cheveu est parlante. L'épaisseur d'un cheveu est comprise entre $50\mu m$ et $100\mu m$. Si l'on considère un cheveu fin de $50\mu m$, soit $5 \times 10^{-5}m$, c'est l'équivalent de 10.000 transistors dernière génération en terme d'épaisseur !
- Concernant les milliards de transistors par puce, cela amène une dimension de nécessité de répétabilité extrêmement forte. En omettant les solutions de redondance mises en place pour éviter qu'un seul défaut ne rende une puce non fonctionnelle [SchneiderElectric, 2020], et en considérant qu'une puce comporte n transistors, avec une probabilité de réussite p pour chacun des transistors, la probabilité qu'une puce ne comporte pas de transistor défaillant est de p^n . Si l'on souhaite avoir une probabilité de succès ps pour une puce, alors la probabilité minimale pour le succès d'un transistor est de $e^{(\ln(ps))/n}$. Le tableau 1.2 ci-dessous donne quelques valeurs pour se représenter en chiffres ces enjeux.

Année	Nombre de transistors	Probabilité minimale p pour des valeurs de ps		
		ps = 0,5	ps = 0,9	ps = 0,99
1975	10^4	p = 0.99993	p = 0.999989	p = 0.9999990
1990	10^6	p = 0.9999993	p = 0.99999989	p = 0.999999990
2020	10^9	p = 0.9999999993	p = 0.99999999989	p = 0.999999999990

FIGURE 1.2 – Conséquence de la loi de Moore sur les besoins de répétabilité

En lisant la figure 1.2, on voit qu'en 1975, pour avoir au moins une chance sur deux de réussir une puce, il faudrait que la probabilité de réussite d'un transistor soit au moins de 0.99993. En 2021, cette probabilité passe à 0.9999999993, soit 100000 fois moins de taux d'échec sur les transistors.

Les solutions pour suivre la loi de Moore ont pris un tournant en 2012 avec l'arrivée des 3D NAND aux nœuds de 22nm et moins [Bunday et al., 2013]. Ces structures, de par leur conception 3D, induisent de nouvelles difficultés physiques qui impliquent des besoins métrologiques bien plus conséquents, en particulier un grand nombre d'étapes de mesures supplémentaires pour garantir le bon déroulement de ces étapes.

1.1.2 Flot de fabrication et données volumineuses

Nous venons de parler de la loi de Moore, de la petitesse d'un transistor et du nombre d'objets mis en jeu impliquant un grand besoin de répétabilité dans l'industrie des semi-conducteurs. Ce besoin de répétabilité implique un grand nombre de mesures et de contrôles.

Dans cette sous-partie, nous faisons le lien entre les deux points de la loi de Moore : le lien entre la petitesse des transistors et la complexité et la taille des données, puis le lien entre la répétabilité et le nombre de contrôles et de mesures.

Aujourd'hui (2021), la fabrication de semi-conducteurs est une succession de plusieurs centaines d'étapes, dont plus de la moitié sont des étapes de mesures, comme le montre la figure 1.3.

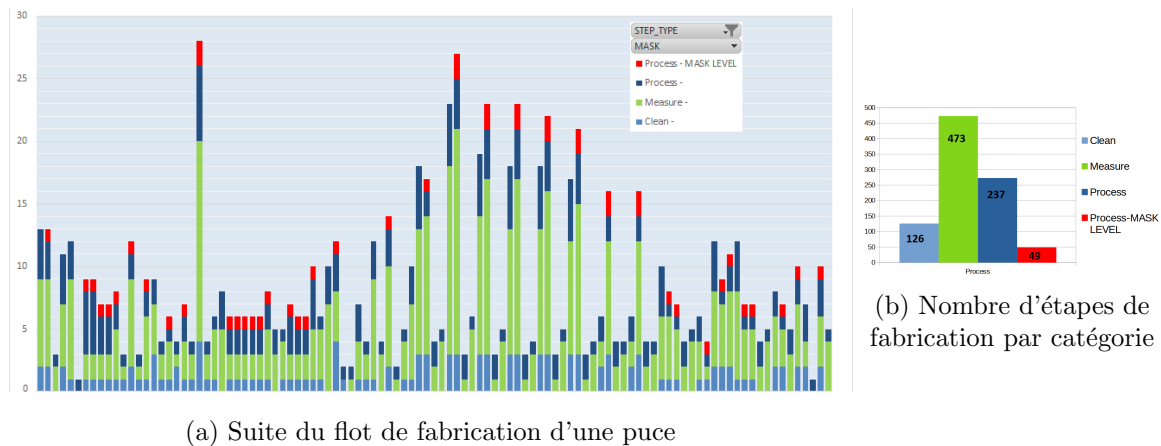


FIGURE 1.3 – Vision du nombre d'étapes de fabrication

Les étapes de mesures supplémentaires génèrent une quantité de données élevée, et cette masse d'information est à la fois une opportunité pour les ingénieurs, et à la fois un challenge pour les exploiter au maximum. Une mesure a vocation à sanctionner/valider une ou plusieurs étapes de fabrication, mais s'il était possible d'en tirer plus de valeur ajoutée, alors il faudrait le faire.

Les produits finis de l'industrie de la micro-électronique sont des puces rectangulaires. Celles-ci sont imprimées sur des wafers, qui sont des tranches de silicium circulaires. Ces tranches de silicium ont des diamètres variables, et dans le cas de ST Crolles 300, elles ont un diamètre de 300mm. Les puces ont des tailles variables, et on peut imprimer plusieurs centaines de puces sur un wafer, parfois dépassant le millier de puces par wafer. Ces proportions amènent d'autant plus à une volumétrie de données élevée.

STMicroelectronics génère 7.6TB de données brutes chaque semaine, soit près de 400TB par an. Notons que ceci n'est qu'une projection, et que ces chiffres sont en croissance permanente, tout comme l'activité de STMicroelectronics.

Valoriser au maximum les données peut signifier dans le contexte d'un produit mature d'en améliorer le rendement. Par exemple, avec ces informations supplémentaires, on peut décider de changer des paramètres dans un flot de fabrication déjà en place. Cela peut également permettre de détecter des puces qui sont déjà défectueuses, et d'arrêter leur production pour ne pas utiliser de ressources inutilement (argent, temps, matière, temps humain...). Dans cette configuration, on dispose de beaucoup de données et de recul sur le produit mature, et on cherchera à gagner des pourcentages de rendement, mêmes infimes.

Dans le contexte de la phase de recherche et développement d'un nouveau produit, on cherche à analyser des mesures sur des plaques dédiées à la R&D, et de définir des spécifications à chaque étape, pour pouvoir calibrer un procédé de fabrication qui n'existe pas encore. Définir un procédé qui n'existe pas encore pour réaliser des puces de façon rentable est une approche différente, car on dispose d'un volume de données plus faible. On doit donner beaucoup de sens à chaque mesure, tandis que dans l'amélioration continue d'un procédé, on donne du sens à une masse de mesures.

Des nouvelles compétences de sciences de données sont nécessaires pour rendre ces données intelligibles, comme nous l'expliquerons dans la partie suivante.

1.1.3 Des nouveaux métiers

Nous venons de parler de la volumétrie des données, du nombre de mesures, qui sanctionnent une ou plusieurs étapes, et de la complexité des mesures. Dans cette partie, nous évoquons plusieurs points :

1. Les données générées par un seul outil sont d'une volumétrie énorme, et gérer la qualité de ces mesures est un métier à part entière dont la responsabilité incombe à la métrologie.
2. Les données générées par un seul outil ont des interactions avec plusieurs procédés, et faire le lien entre une mesure et un procédé est un métier à part entière.
3. Des deux points précédents naît un besoin de "factorisation" de ces compétences. Afin de faire le lien entre les mesures et les ateliers, une personne doit rendre intelligible cette mesure grâce aux sciences de données.

1. Des outils de mesures et des wafers standardisés

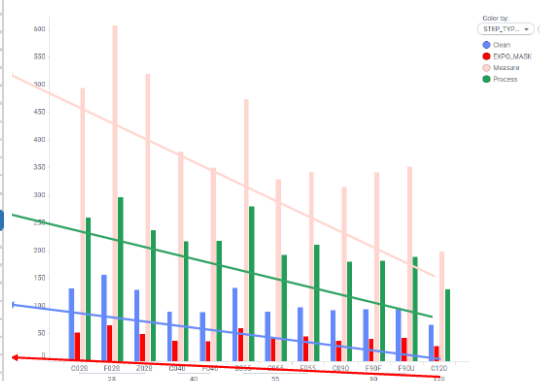
Dans l'industrie des semi-conducteurs, contrairement à d'autres industries, un produit peut voir une même machine plusieurs fois dans son flot de fabrication. Dans la Figure 1.4, on montre la répartition du nombre de chaque étape pour chaque technologie et nœud technologique. Le nœud technologique est la taille du transistor qui est produit dans cette technologie.

Sur ce tableau, on voit par exemple pour la technologie C028, un wafer subira 131 étapes de nettoyage, 52 expositions (donc 52 passages en lithographie), 494 mesures (parfois sur une même machine), et 260 procédés divers, certains étant les mêmes. On note une tendance globale qui confirme les points précédents : plus les nœuds sont petits, plus le nombre d'étapes est élevé.

Node	TECHNO	Clean	EXPO_MASK	Measur	Process	TOTAL_STEP
28	C028	131	52	494	260	937
28	F028	156	65	607	296	1124
28	Z028	129	48	518	237	932
40	C040	90	37	378	216	721
40	F040	89	36	349	217	691
55	B055	132	60	473	279	944
55	C055	90	41	329	192	652
55	F055	97	44	342	210	693
90	C090	92	37	314	179	622
90	F90F	94	40	341	181	656
90	F90U	94	42	351	189	676
120	C120	66	27	198	130	421

Node	TECHNO	Clean	EXPO_MASK	Measur	Process	TOTAL_STEP
28	C028	14%	6%	53%	28%	100%
28	F028	14%	6%	54%	26%	100%
28	Z028	14%	5%	56%	25%	100%
40	C040	12%	5%	52%	30%	100%
40	F040	13%	5%	51%	31%	100%
55	B055	14%	6%	50%	30%	100%
55	C055	14%	6%	50%	29%	100%
55	F055	14%	6%	49%	30%	100%
90	C090	15%	6%	50%	29%	100%
90	F90F	14%	6%	52%	28%	100%
90	F90U	14%	6%	52%	28%	100%
120	C120	16%	6%	47%	31%	100%

(a) Répartition du nombre d'étapes par technologie



(b) Graphique de répartition des étapes par technologie

FIGURE 1.4 – Répartition des procédés par technologie

Ces passages multiples sur une même machine sont dus à la nécessité d'empiler des couches avec des matériaux similaires plusieurs dizaines de fois. Le processus est globalement itératif (photolithographie, gravure, polissage, etc...), et la puce doit nécessairement subir plusieurs fois les mêmes étapes au cours de sa fabrication.

Le fait que des mêmes machines voient plusieurs fois le même produit durant sa fabrication implique qu'elles doivent les voir dans des conditions similaires, car elles sont calibrées pour fonctionner dans un intervalle de conditions. Par exemple, un produit doit être aligné parfaitement avec ses précédents passages dans la machine. Cet alignement s'appelle l'overlay.

Cette nécessité de vérifier les conditions de passage dans la machine implique une voire plusieurs vérifications, et participe à l'augmentation de la volumétrie des données. Ces mesures sanctionnent l'histoire toute entière du produit, et peuvent sanctionner soit une étape, soit une partie de l'histoire du produit.

2. Des mesures à usages multiples

Comme dit précédemment, un produit passe plusieurs fois dans la même machine durant sa fabrication. De la même façon, il est mesuré plusieurs fois par le même outil de mesure. De plus, une mesure produite par un même outil de mesure peut être utilisée pour différents ateliers.

On peut utiliser par exemple une mesure de topographie pour vérifier l'état de planéité après le procédé de polissage visant à aplanir une surface : CMP (Polissage Mécano-chimique). On peut également utiliser cette même mesure pour vérifier si le wafer est en état de passer une photolithographie, un procédé de fabrication développé dans le chapitre 2. Un outil de mesure produit une mesure pouvant interagir avec différents procédés de fabrication. L'exemple donné dans ce paragraphe montre que la mesure de topographie peut servir à l'atelier de polissage mécano-chimique comme à l'atelier photolithographie.

Chez STMicroelectronics, un équipement de mesure est géré par le département métrologie. Certaines mesures ont vocation à servir les intérêts du département métrologie dans la chaîne de fabrication, comme nous l'expliquerons dans la partie suivante. Les mesures peuvent également être demandées par les ateliers de fabrication, soit pour des raisons de contrôle, soit dans le cadre de R&D. Chaque atelier aura des besoins spécifiques, et sa vérification sur la mesure lui sera propre. En revanche, une mesure effectuée peut demeurer la même pour deux ateliers différents.

Comme expliqué précédemment, ces mesures ont une densité d'information conséquente, au vu de la loi de Moore et de son évolution. Elle contient des éléments de réponses aux questions de différents ateliers, et chacun de ces ateliers va répondre à une question qui lui est propre. La charge mentale (ou responsabilité) de la décomposition de l'information, propre à chaque atelier trouve plusieurs acteurs la portant :

1. Le fournisseur de l'outil de mesure. Il peut y avoir un logiciel avec la machine qui mesure, proposant des traitements d'images adaptés aux besoins connus par les fournisseurs.
2. Le responsable de l'outil de mesure peut proposer des scripts de traitement de données sur les mesures qu'il envoie.
3. Les ateliers qui demandent les mesures peuvent faire des scripts spécifiques à leurs besoins.

1. Quand le fournisseur a la charge

Le défaut du mode de fonctionnement où les fournisseurs ont la charge mentale, c'est qu'ils ne subissent pas les problématiques du site de fabrication. Ils ne sont donc pas aptes à proposer seuls une solution, et doivent collaborer avec les ingénieurs de métrologie, et des ateliers, et développer des solutions avec un mode de communication de ping-pong avec le site de fabrication. Ce point n'est pas problématique en lui-même, mais il ralentit grandement la phase de recherche et développement aboutissant à la définition d'un besoin précis.

Les fournisseurs sont externes à l'entreprise, et pour satisfaire un besoin nouveau vont facturer leurs services à l'entreprise subissant la problématique. Cette facturation génère une pression subie par le fournisseur de l'exigence des ingénieurs du site de fabrication. Cette exigence entraîne un mode de fonctionnement où les fournisseurs vont devoir suivre un procédé de gestion de projet pour aboutir à un produit « livrable ». Cela ralentit également la phase de test et de vérification de la fonctionnalité des outils proposés.

2. Quand le responsable de l’outil ou les ateliers ont la charge

Quand la charge mentale est du côté du site de fabrication, le problème vient des compétences. Comme expliqué dans l’introduction, les ateliers de fabrication ainsi que les outils de mesure sont gérés par des personnes de formation physique-chimie. Ils ne sont donc pas formés pour faire du développement, et utilisent leur temps d’expertise métier à faire de la programmation. Au-delà des compétences, la gestion de la production des mesures d’un outil est un métier à temps plein, tout comme gérer les conséquences des mesures.

3. Une solution : des ingénieurs en sciences de données sur le site de fabrication

Un moyen de porter cette charge est de laisser la charge mentale de la R&D au site de fabrication, avec une nouvelle compétence qui sera un métier à part entière : les sciences de données. Cette compétence vise à être le lien entre la mesure et les différents ateliers. En se familiarisant aux problématiques des ateliers, il s’agit de rendre intelligible la mesure pour que les ingénieurs experts en procédés puissent utiliser leur expertise à faire de la physique-chimie, au lieu de faire du développement informatique.

La R&D se fait en interne, et l’exigence d’un développement informatique peut se formuler parfaitement au fournisseur une fois le besoin identifié. Le développement d’une mise à jour sur les logiciels des fournisseurs est plus simple, car le site de fabrication peut en formuler des spécifications exactes. Le site de fabrication peut décider de conserver sa propre librairie de traitement de données également, sans en informer le fournisseur.

Ces compétences sont en train d’être intégrées dans le site de fabrication de ST Crolles, en sensibilisant les personnes aux apports que peuvent avoir des profils en sciences de données. Au-delà de l’intérêt purement pragmatique de l’intelligibilité de la mesure, cela renforce les liens entre les compétences des procédés. Les compétences métier deviennent extrêmement spécifiques, et la manifestation de cette spécificité génère des difficultés à transmettre des connaissances d’un métier à un autre. Elle génère entre autres des fichiers d’une complexité telle que seuls les personnes utilisant ces mesures en ont la compréhension.

Bien que ce ne soit pas le seul moyen, rendre intelligible les fichiers est un pas à faire pour partager les connaissances et les compétences au sein d’une industrie complexe comme celle de la micro-électronique. Le partage de connaissance est utile pour anticiper les problèmes que les autres métiers que le nôtre pourront éventuellement rencontrer. Quand tout le monde comprend les autres métiers et leurs difficultés, l’industrie est de fait bien plus riche.

Ce qui est nouveau, ce sont les techniques de traitement de données ainsi que les besoins grandissants dus à l’augmentation de la volumétrie à la fois d’une mesure, et à la fois du nombre de mesures. Mais rendre intelligible des données n’est pas entièrement nouveau, et est toujours partiellement la responsabilité de l’atelier métrologie, comme nous l’expliquons dans la partie suivante.

1.2 De la métrologie

Cette section parlera de métrologie. La première partie fait état des enjeux de la qualité de la mesure. Ensuite, nous présenterons trois enjeux de métrologie dans l'industrie [KADOK, 2019] : la qualité, la métrologie avancée, et la métrologie in-line mise en place à STmicroelectronics.

1.2.1 Intérêt de la mesure

« *If you cannot measure it, you cannot improve it* »- Lord Kelvin

Toute appréciation, tout adjectif, est intrinsèquement comparatif avec d'autres objets. Lorsqu'on dit que quelque chose est « petit », on le décrit « petit » car il est petit comparativement à des objets semblables qui sont plus grands. Mais deux personnes avec un vécu différent se représenteront une réalité physique différente lorsqu'ils entendront l'adjectif « petit ». Les adjectifs sont sujets à interprétation.

La mesure, c'est l'expression de la volonté se séparer de cette subjectivité pour avoir des grandeurs comparables grâce à des nombres et des entités bien définies aux yeux de tous (mètres, kilogrammes ...). Cela permet de décrire sans qu'il y ait d'interprétation(s) possible(s) quant à la réalité physique décrite.

Au-delà de cette volonté de s'affranchir de la subjectivité, dans l'industrie, on cherche toujours à améliorer les procédés de fabrication déjà en place. Pour cela, on souhaite mesurer au plus précis les grandeurs mises en jeu dans les procédés, et, par la suite, pouvoir comparer des choix industriels. Différents enjeux émergent :

1. Le premier correspond à la volonté de décrire une réalité physique. Par exemple mesurer une distance, en l'exprimant en nombre de mètres, ou mesurer une masse, exprimée en kilogrammes. Cette expression nécessite de savoir et surtout, de définir ce qu'est un kilogramme. C'est de cette volonté qu'émerge le concept d'étalon, expliqué dans la partie 1.2.2.
2. Le second, qui découle du premier, est la volonté d'enrichir les mesures avec des descriptions qui ne sont pas directement liées à la réalité physique. Il s'agit de transformer, d'augmenter des données décrivant une réalité physique pour leur donner une couche d'intelligibilité supplémentaire. Le but étant d'enrichir l'interprétation des mesures devenues massives. Par exemple, on peut mentionner l'extraction de contours, un domaine très en vogue dans l'industrie de la micro-électronique. Ceci sera développé dans la partie « Métrologie avancée ».
3. Le troisième, découlant des deux précédents, est l'aboutissement de la volonté d'améliorer les procédés de fabrication. Pour cela, il est nécessaire de pouvoir comparer des choix industriels, définie par de l'intelligence humaine. On cherche à avoir une mesure de qualité pour faire évoluer les procédés dans la bonne direction. L'utilisation de ce paradigme sera développé dans la partie « Métrologie in-line ».

1.2.2 Définitions

Quelques définitions sont bienvenues pour aborder les différents concepts de la métrologie. Pour commencer, les définitions qui concernent la valeur de la mesure :

1. La grandeur (physique) : Paramètre que l'on souhaite déterminer avec le plus de précision possible.
2. Le mesurage : L'ensemble des opérations qui déterminent la valeur d'une grandeur (physique).
3. La mesure : Évaluation d'une grandeur (physique) par comparaison avec une autre grandeur homogène, souvent un étalon (expliqué ci-dessous)

4. L'incertitude : C'est la demi-largeur de l'intervalle centré sur la mesure donnée contenant la valeur d'une grandeur physique.
5. L'erreur absolue : C'est la différence entre la grandeur physique et la mesure.
6. L'erreur relative : C'est l'erreur absolue divisée par la grandeur physique.

Pour résumer :

On note X la grandeur physique ; x la mesure ; dx l'incertitude ; e l'erreur absolue ; e_r l'erreur relative.

$$x - dx < X < x + dx$$

$$x + e = X$$

$$e_r = e/X$$

Ensuite, les définitions qui concernent les capacités des équipements de mesure :

1. Justesse : aptitude d'un instrument de mesure à donner des indications exemptes d'erreur systématiques.
2. Linéarité : stabilité de l'erreur de justesse sur la gamme d'utilisation de l'instrument de mesure.
3. Étalonnage : Ensemble d'opérations établissant, dans des conditions spécifiées, la relation entre les valeurs de la grandeur indiquées par un appareil de mesure et les valeurs correspondantes de la grandeur réalisées par des étalons.
4. Répétabilité : étroitesse de l'accord entre les résultats des mesures successives du même objet dans les mêmes conditions.
5. Reproductibilité : étroitesse de l'accord entre les résultats des mesures du même objet en faisant varier les conditions de mesure.
6. Résolution : la plus petite différence d'indication d'un système de mesure qui peut être perçue de manière significative.

Un adjectif est relatif à une impression, mais une mesure est relative à un étalon bien déterminé. Cette notion de détermination d'étalon(s) est en réalité un écosystème de certifications hiérarchiques. La figure 1.5 montre les étalons tels qu'ils sont mis en place dans la micro-électronique.

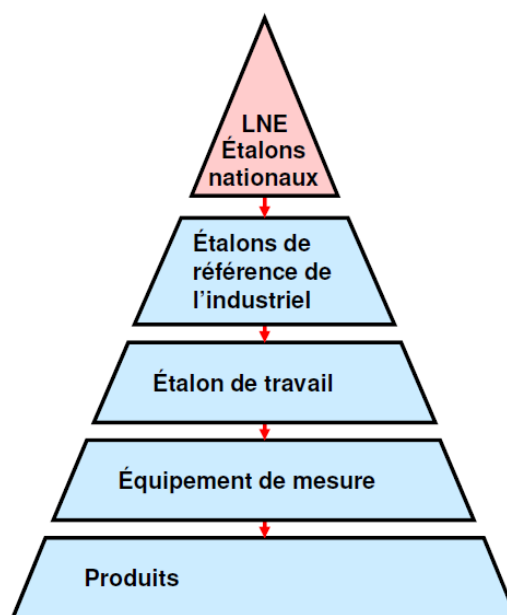


FIGURE 1.5 – Hiérarchie des étalons dans la micro-électronique - *Source interne*

On définit plusieurs types d'étalons :

1. **Étalon primaire (Nationaux)** : Étalon qui est désigné ou largement reconnu comme présentant les plus hautes qualités métrologiques et dont la valeur est établie sans se référer à d'autres étalons de la même grandeur.
2. **Étalon de référence** : Étalon, en général de la plus haute qualité métrologique disponible en un lieu donné ou dans une organisation donnée, dont dérivent les mesurages qui y sont faits.
3. **Étalon de transfert** : Étalon utilisé comme intermédiaire pour comparer les étalons entre eux.
4. **Étalon de travail** : Étalon qui est utilisé couramment pour contrôler des appareils de mesures, ou des matériaux de référence.

Les étalons de transfert n'apparaissent pas dans la hiérarchie de la figure 1.5 car ils sont à usage ponctuel lors de l'installation de la machine.

1.2.3 Métrologie avancée

1. Intérêt de la métrologie avancée

La mesure brute répond à la problématique de la description de la réalité. Dans l'industrie cependant, une mesure coûte des ressources (temps-humain, argent, énergie, temps-machine...), et a vocation à être rentable, comme tout acte industriel. Cette rentabilité se manifeste comme aide à la prise de décision pour les ingénieurs, les techniciens, etc...

Plusieurs points rendent difficile l'utilisation de ces mesures. D'une part, la densité d'information contenue dans une mesure, due à l'augmentation de densité des transistors sur une puce. Cette description de réalité dense nécessiterait un regard approfondi de la part d'un expert métier pour pouvoir prendre les décisions relatives à la mesure effectuée. D'autre part le nombre de mesures effectuées dû au besoin de répétabilité qui est en croissance permanente. Cette quantité de mesure demanderait une attention constante d'un ou plusieurs regards humain(s).

Ce temps passé par un ou des expert(s) métier(s) coûte en réalité très cher. Il coûte à l'entreprise non seulement en tant que salaire, mais il coûte également en tant qu'attention humaine.

De plus, lorsqu'on souhaite prendre une décision en fonction d'une mesure, le biais humain est très fort. Même si l'expertise d'une personne est de valeur, la subjectivité doit disparaître au maximum lorsqu'on souhaite industrialiser en masse une fabrication de puce.

La volonté d'enlever la subjectivité trouve sa solution dans du traitement de données qui enrichit la mesure. Le but est de la rendre compréhensible de façon uniforme par tous. D'une part en transposant l'expertise métier dans un programme logiciel, d'autre part en effectuant un traitement informatique de données aidant les experts métier à mieux comprendre leurs procédés de fabrication.

Dans l'industrie des semi-conducteurs, une approche pour résoudre cette problématique est la métrologie avancée. Il s'agit de transformer une mesure (souvent une image) en une information plus compacte ou plus intelligible.

Des exemples de métrologie avancée à STMicroelectronics Crolles

a. Des images SEM

Le site de ST à Crolles génère plus de 20 millions d'images SEM par an, et pour donner plus de valeur à ces mesures, des travaux de métrologie avancée sont effectués, comme expliqué dans [Le-Gratiet et al., 2019].

Un exemple est l'extraction de contours et la classification sur des images SEM (Microscopie Électronique à Balayage). Dans [Lakcher et al., 2017], des images SEM sont extraites sur la totalité d'un wafer (voir figure 1.6) afin de constituer un groupe d'images. Par la suite, ces images sont classifiées en utilisant un algorithme de classification hiérarchique, puis ces contours sont replacés sur le wafer. Cette cartographie donne une valeur ajoutée aux images SEM massives générées à ST Crolles et permet de constater que les contours qui se ressemblent sont localisés soit au centre du wafer, soit sur la couronne extérieure. Cette information est riche et peut être interprétée par les ingénieurs pour mieux comprendre les procédés de fabrication.

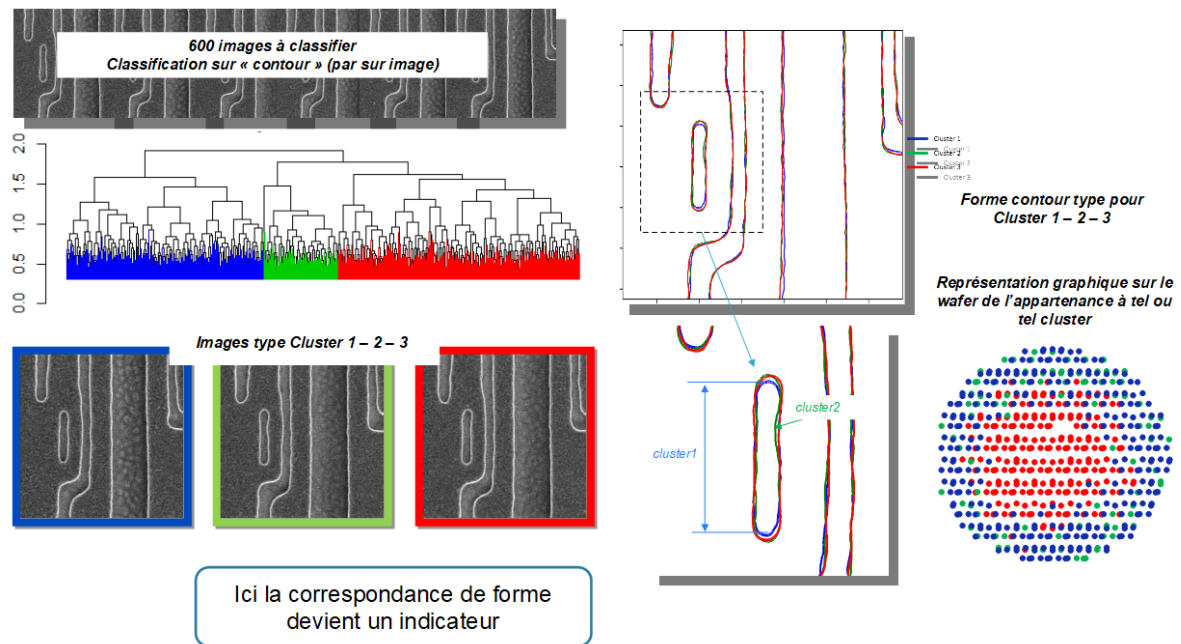


FIGURE 1.6 – Classification d'images SEM - *Source : [Lakcher, 2018]*

Une autre application sur les images SEM est l'extraction de contours, combinée à l'utilisation des fichiers de définition des contours théoriques. Pour identifier des zones à risque, comme montré dans la figure 1.7, on superpose les contours extraits sur les mesures et les contours théoriques. Les différences sont plus visibles que sur une mesure brute, car l'information y est moins dense.

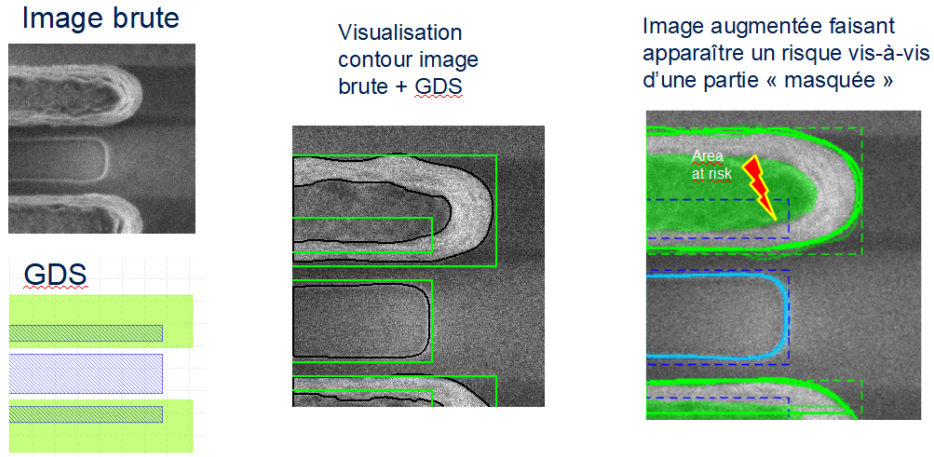


FIGURE 1.7 – Comparaison d’images SEM avec le résultat théorique attendu - *Source : [Le-Gratiet et al., 2020]*

b. Métrologie virtuelle : Overlay

Certaines mesures coûtent cher en micro-électronique. Modéliser ces mesures grâce à des entrées plus faciles d’accès est une solution de choix pour STMicroelectronics. Parmi ces entrées, on peut noter les données logiques, contrôlées et choisies par des personnes, ou des données plus faciles ou moins coûteuses à mesurer. On appelle cette prédiction de la métrologie virtuelle.

La métrologie virtuelle est un usage avancé des mesures, et permet d’économiser de la ressource (temps, argent, énergie...) pour avoir une information suffisamment exploitable.

L’exemple ici est la métrologie virtuelle pour l’overlay. L’overlay est la notion d’alignement entre les étapes de fabrication. Aligner parfaitement un wafer avec son précédent passage est une problématique à part entière étant donné les dimensions mises en jeu. Mesurer un désalignement est coûteux, et utiliser des informations plus faciles d’accès est une solution, comme montré dans [de Caunes et al., 2010] et [van Dijk et al., 2019].

Dans la figure 1.8, on voit la représentation actuelle de la mesure d’overlay entre chaque étape d’implantation ionique : une étape de fabrication. La partie droite de la figure est une perspective mise en place en « shadow », c’est-à-dire en tâche de fond, pour comparer les résultats du déploiement d’une métrologie virtuelle pour la problématique d’overlay.

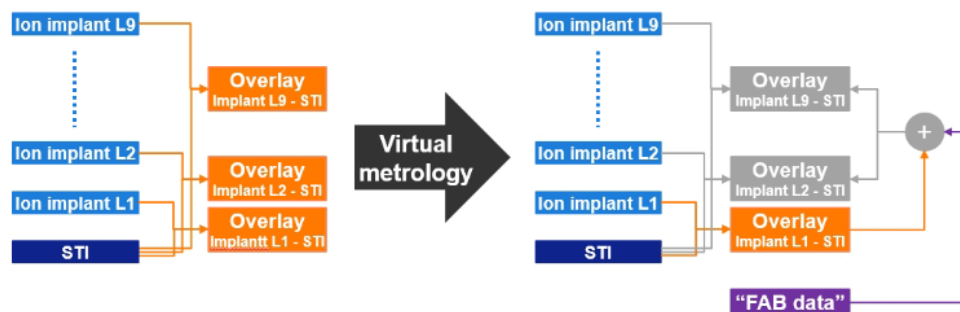


FIGURE 1.8 – Schéma explicatif de la métrologie virtuelle pour la prédiction d’overlay - *Source : [van Dijk et al., 2019]*

Dans 1.8, on veut estimer l’overlay après une opération de maintenance. Les opérations de maintenance étant similaires, on anticipe l’overlay au mieux pour ne pas avoir à le mesurer.

c. Enrichissement de mesures de nano-topographie

Dans un article de Globalfoundries [Lee et al., 2017], on cherche à donner plus de valeur à une mesure de topographie en l'enrichissant via une détection de formes. Cette détection de formes est inspirée de la reconnaissance faciale, en particulier pour détecter quelle émotion ressent une personne en fonction des distributions de topographies sur son visage.

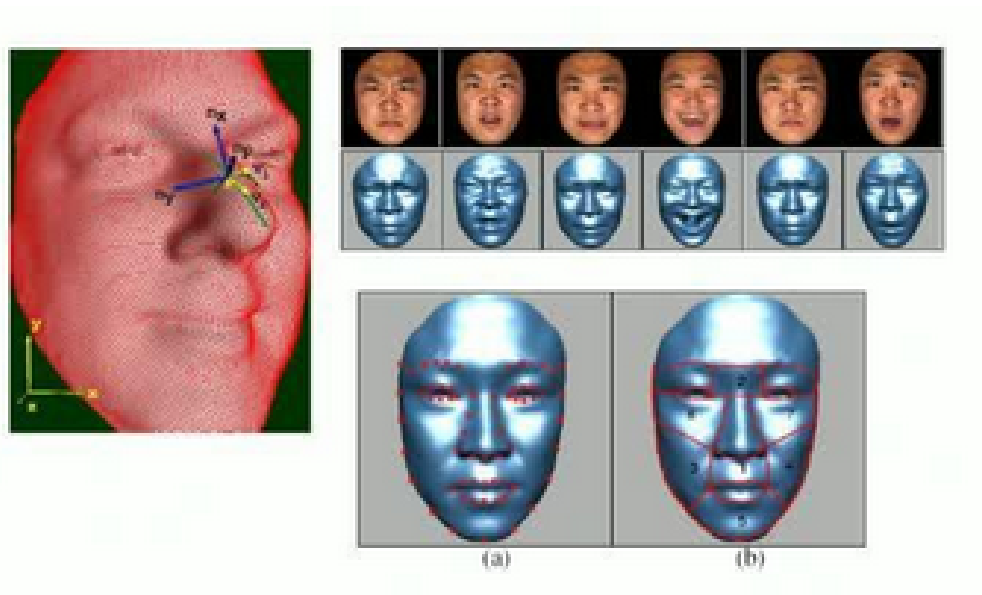


FIGURE 1.9 – Visages avec émotions - *Source : [Wang et al., 2006]*

Dans la figure ci-dessus, grâce à une détection de formes (figure 1.10), on peut déterminer quelle émotion est exprimée sur le visage en utilisant l'histogramme de répartition des formes. Sur chaque zone de 7x7 pixels, on attribue une des 7 formes montrées dans la figure.

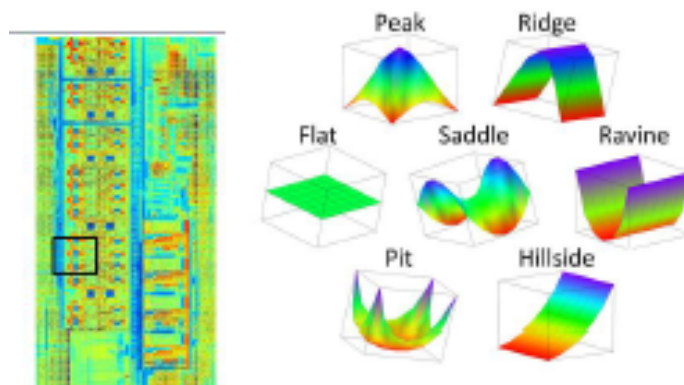


FIGURE 1.10 – Détection de formes sur puce - *Source : [Wang et al., 2006]*

Par la suite, en demandant à des personnes d'avoir ces 6 émotions et de les prendre en photo, une analyse discriminante linéaire permet de classer les émotions en fonction de l'histogramme. La figure 1.11 montre le résultat. Dans les lignes se trouvent les entrées, et dans les colonnes la sortie de la classification linéaire.

Input \ Output	Anger	Disgust	Fear	Happiness	Sadness	Surprise
Anger	80.0%	1.7%	6.3%	0.0%	11.3%	0.8%
Disgust	4.6%	80.4%	4.2%	3.8%	6.7%	0.4%
Fear	0.0%	2.5%	75.0%	12.5%	7.9%	2.1%
Happiness	0.0%	0.8%	3.8%	95.0%	0.4%	0.0%
Sadness	8.3%	2.5%	2.9%	0.0%	80.4%	5.8%
Surprise	1.7%	0.8%	1.2%	0.0%	5.4%	90.8%

FIGURE 1.11 – Pourcentage de détection des émotions - *Source : [Wang et al., 2006]*

Dans cet esprit, Globalfoundries s’est inspiré de cela [Lee et al., 2017] pour évaluer la qualité de la topographie avec des indicateurs objectifs, autre que le regard d’un ingénieur ayant l’expérience de son procédé.

1.2.4 Métrologie In-Line et cartes de contrôle

Nous venons de parler du besoin de nouveaux métiers dû à la volumétrie des données, et du nombre de mesures conséquent dans l’industrie. On sait également qu’on effectue de la métrologie avancée sur ces mesures. Cette partie montre un usage des mesures lié au suivi de procédés de fabrication : la métrologie In-Line. Celle-ci prend en compte les enjeux de calibration des outils de mesure dans le but de faire du suivi de procédés industriels, à travers la mise en place de cartes de contrôle.

1.2.4.1 Quels choix techniques pour quels objectifs ?

Plusieurs enjeux émergent lorsqu’on parle de suivis de procédés. Les enjeux sont détaillés dans l’article [Moyné and Iskandar, 2017]. Cet article fait état de la problématique de « Smart Manufacturing », où les industries souhaitent maximiser l’utilisation de leurs données. En particulier, relier des causes à des effets lors de détections de défauts.

En faisant des prédictions suffisamment fiables, on peut considérer que les déterminants menant à la prédiction d’un défaut sont eux-mêmes en défaut. Si l’on voit par exemple qu’une contamination engendre systématiquement un test électrique défaillant en fin de fabrication du produit, la contamination devient le défaut, et non le test fonctionnel de fin de production. En remontant les causes, on considère les déterminants les plus en amont comme défaillants.

En anticipant, en modélisant et en représentant au maximum les procédés, on converge vers le paradigme du digital twin [Tao et al., 2019], où l’enjeu est de prédire et/ou de stocker de bout en bout les mesures d’un produit, et d’en conserver une version digitale. Dans d’autres contextes, cela peut également aider à la supervision d’aménagements urbains par une modélisation quasi-exhaustive de l’espace 3D de territoires [Siradel, 2021].

Le digital twin est la métrologie virtuelle poussée à son paroxysme, et est une direction de recherche pour aider à anticiper la réalité. Une autre direction du traitement de données, qui lui est complémentaire, est l’analyse et l’interprétation des mesures, pouvant être prédites comme mesurées.

1.2.4.2 Capabilité procédé et outil

Pour superviser un procédé de fabrication, avec de la métrologie in-line, les techniques sont diverses. La situation typique est d'avoir une métrique issue ou non de la métrologie avancée à contrôler, qui doit se trouver dans un certain intervalle pour valider entièrement ou partiellement l'étape. Cette métrique a vocation à estimer l'écart de la réalisation du procédé par rapport à une réalisation cible.

On va mesurer la métrique dont on souhaite détecter la dérive. Cette métrique comporte une incertitude due aux erreurs de justesses des mesures. L'écart-type de cette mesure est notée σ_{OBS} , et on considère que la mesure est issue d'une variable aléatoire suivant une loi normale centrée sur la valeur juste. Cette variabilité observée comporte elle-même deux composantes : la variabilité du procédé notée σ_{proc} , et la variabilité provenant du mesurage, notée : σ_{MS} . On considère ces deux origines comme indépendantes.

$$\text{On a : } \sigma_{OBS}^2 = \sigma_{proc}^2 + \sigma_{MS}^2$$

La variabilité σ_{MS} liée au mesurage comporte plusieurs composantes : $\sigma_{MS}^2 = \sigma_{RPT}^2 + \sigma_{RPD}^2 + \sigma_{IMS}^2$

Où σ_{RPT} est la variabilité due à la répétabilité, σ_{RPD} est la variabilité due à la reproductibilité et σ_{IMS} est la variabilité due aux interactions avec l'environnement.

Pour mesurer chacune de ces variabilités, on doit mettre en place des procédures. La façon dont on mesure toutes ces variabilités n'est pas le sujet de cette thèse. Il est cependant essentiel d'avoir conscience que, lorsqu'on a une mesure, celle-ci comporte des incertitudes aléatoires provenant de sources diverses.

La variabilité de mesure σ_{MS} doit être estimée pour savoir si l'équipement est toujours apte à suivre un procédé. On dispose d'étalons de travail certifiés par une hiérarchie de certifications (Voir figure 1.5). Étant donné un étalon, on connaît le résultat attendu a priori sur l'outil de mesure. Cet étalon sera utilisé pour connaître la variabilité σ_{MS} , et ainsi détecter un équipement qui deviendrait hors de contrôle. La figure 1.12 met en lumière deux aspects dans le contrôle des mesures pour le contrôle des procédés.

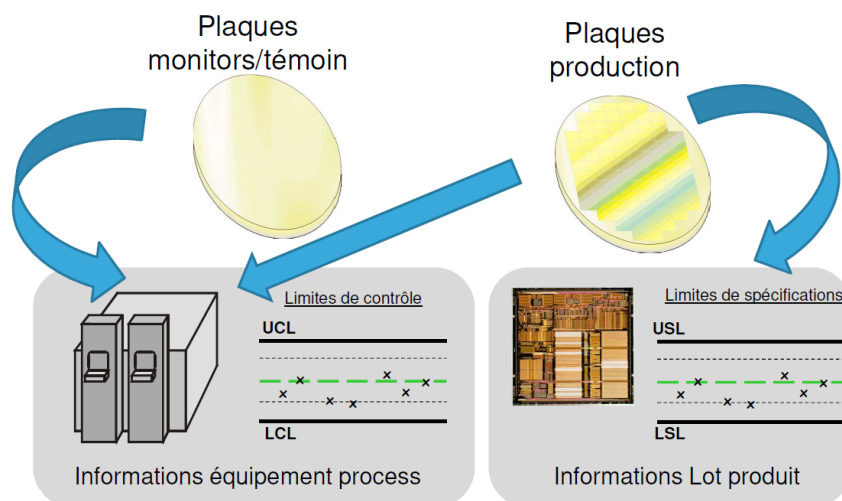


FIGURE 1.12 – Schéma des contrôles de procédés et des équipements de mesure - *Source interne : D. LE CUNFF, Formation Métrologie*

Sur la figure 1.12, UCL (resp. LCL) est la valeur maximum (resp. minimum) limite tolérée pour le contrôle de l'équipement (Upper (resp. Lower) Control Limit). De la même façon, USL et LSL correspondent respectivement aux limites de spécification d'un procédé de fabrication. Dans la pratique, $UCL - LCL < USL - LSL$.

Pour des raisons évidentes, une dérive de procédé ne peut pas se détecter si la variabilité liée à l'équipement est plus grande que celle du procédé. De façon générale, ce que l'on souhaite mesurer ou détecter ne doit pas être plus petit que l'incertitude sur la mesure.

On fait l'hypothèse que les mesures sont une somme de lois normales indépendantes, et que les écarts-types de ces lois normales conditionnent la capacité de suivi d'un outil d'une part, et d'un procédé d'autre part. On souhaite évaluer leurs capacités respectives.

On peut alors évaluer la capacité métrologique d'un équipement. On note C_{pm} cette capacité :

$$C_{pm} = \left(\frac{USL - LSL}{6\sigma_{MS}} \right)$$

Plus cette capacité est grande, plus l'équipement est adapté pour suivre le procédé.

$C_{pm} < 3$: Mauvaise capacité

$10 > C_{pm} > 3$: Capacité acceptable

$C_{pm} > 10$: Bonne capacité

Dans une certaine mesure, on considère que la mesure suit une loi normale avec sa propre dispersion. Comme mentionné dans les parties précédentes, le nombre de transistors dépasse 10^9 , et on ne doit pas laisser passer des erreurs. Les besoins métrologiques ont été mis en évidence, et les besoins de répétabilité sont tellement importants que les variabilités ne sont plus acceptées. Cela se traduit par le coefficient 6 devant la variabilité.

D'un point de vue de la loi normale, si l'on considère un outil de bonne capacité avec un $C_{pm} = 10$, alors il y a 60 sigma de largeur contenu dans l'intervalle [LSL, USL], ce qui correspond à presque toutes les valeurs. On peut donc être rassurés quant à l'impact de la variabilité de la mesure sur la mesure de la variabilité du procédé.

On note C_p la capacité procédé, définie comme suit :

$$C_p = \left(\frac{USL - LSL}{6\sigma_{OBS}} \right)$$

Quant à la capacité procédé C_p , il s'agit du nombre de fois que 6 écarts-types sont contenues dans l'intervalle d'acceptation, soit ± 3 écarts-types. Si $C_p = 2$, alors on a 99.99999802% des valeurs contenues dans cet intervalle, soit ± 6 écarts-types.

Jusqu'à maintenant, on ne considérait les performances que d'un seul outil et d'une seule métrique. Dans la pratique, pour certifier la bonne réalisation de procédés, on préférera utiliser un ensemble d'outils de mesure. Cela a l'avantage de pouvoir détecter des dérives d'outils de mesures devenant hors de contrôle sans nécessairement vérifier avec un étalon de travail servant à calibrer l'outil de mesure.

La vision de la métrologie est de réduire au maximum la variabilité des procédés. La difficulté émergeant des possibilités physique de réellement décrire la totalité d'un wafer ou d'un procédé rend impossible une analyse exhaustive sur la totalité du wafer. C'est pour cette raison pragmatique que le choix a été de se focaliser sur la variabilité plutôt que sur une vérification exhaustive à chaque étape. Il n'est pas possible, d'un point de vue du calcul comme d'un point de vue de mesure, de faire cette vérification.

Le raisonnement est le suivant : un procédé qui fonctionne et qui ne varie pas fonctionne toujours. Ainsi, le parti pris est de réduire la variabilité en regardant des descripteurs de variabilité, en supposant qu'ils suffisent à la décrire. En analysant ces descripteurs, nous sommes capables de réduire au mieux la variabilité.

Nous venons de montrer comment la variabilité est gérée pour un procédé de fabrication. Elle peut être mesurée par un seul ou plusieurs outils. Dans la partie suivante, nous parlons des « cartes de contrôle », qui sont un ensemble de contrôles de procédés à effectuer pour mener la fabrication d'un wafer.

1.2.4.3 Cartes de contrôle

Les deux mécanismes suivants sont implantés dans ce que l'on appelle une carte de contrôle :

SPC : Statistical Process Control. Il s'agit de l'observation d'un état. Les mesures vont être utilisées pour déclencher des alarmes et par la suite des actions par les ingénieurs sur les équipements ou étapes de fabrication concernées.

APC : Advanced Process Control. C'est un usage dynamique et automatisé de la mesure pour réguler les procédés de fabrication. Ici la mesure et le décalage par rapport à une valeur cible vont enclencher des changements dans les recettes de fabrication pour corriger l'erreur. Ces systèmes APC aussi appelés Run2Run sont des boucles de régulations plus ou moins avancées qui sont utilisées pour maintenir le produit dans les spécifications et corriger les dérives.

Une carte de contrôle est un ensemble de contrôles effectués sur un wafer. Le but de cette carte est d'assurer la production d'un wafer de bout en bout. La fabrication des wafers a vocation à être automatisée au maximum, et on souhaite définir un processus de fabrication avec le moins d'interventions humaines possible.

Pour ce faire, on automatise des contrôles de validation de procédés : si la variabilité est basse, le wafer peut passer à l'étape suivante. Si elle est haute, l'intervention humaine est nécessaire sur ce wafer. L'expertise métier ou alors un autre contrôle de substitution est effectué afin de vérifier si le wafer doit passer ou non à l'étape suivante. Dans la figure 1.13, on montre ce principe schématiquement.

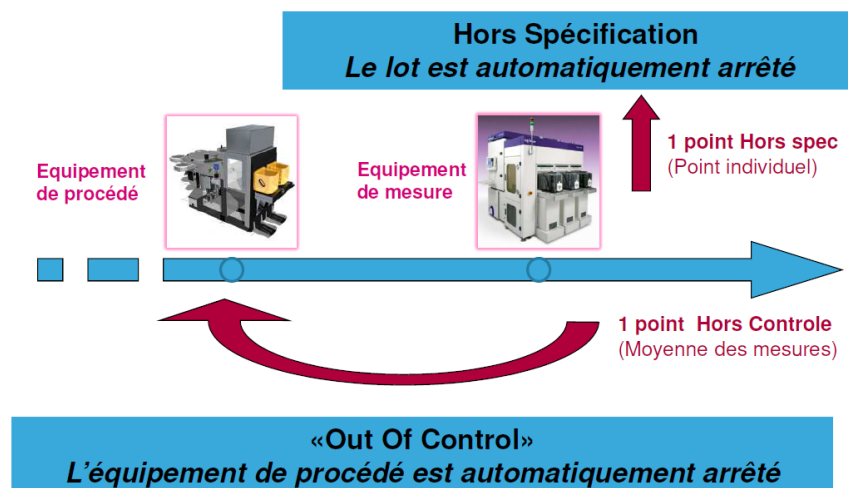


FIGURE 1.13 – Principe d'une étape de carte de contrôles - *Source interne : D. LE CUNFF, Formation métrologie*

Les erreurs peuvent être dues à l'équipement de procédé, ou bien à l'état de la plaque. Dans le premier cas, on arrête l'équipement de procédé, et dans le second cas on sort le « lot » pour le faire analyser plus en détail avec une intervention humaine, qui prendra une décision.

Cette volonté d'automatisation ne souhaite pas remplacer l'expertise métier et la présence d'humains sur le terrain capables de comprendre les procédés. L'enjeu ici est de mettre en place des critères qui minimisent l'intervention humaine dans la plupart des cas. Cette compréhension physique est essentielle, et ne doit pas être écartée des pré-occupations des ingénieurs. C'est le sujet du chapitre suivant, expliquant le contexte industriel dans lequel évoluent les wafers ainsi que les procédés de fabrication qui ont été mis en jeu dans cette thèse.

Chapitre 2

Des couches minces et de la topographie

Ce chapitre se veut relativement indépendant. Malgré quelques références au chapitre précédent, celui-ci peut se comprendre dans sa majorité sans lire le chapitre 1.

En sortant de ce chapitre, le lecteur aura compris les points suivants :

1. La notion de couches minces, et l'importance du rôle de la photolithographie.
2. Une introduction générale à la photolithographie.
3. L'importance devenue critique de la topographie et de son interaction avec d'autres ateliers, en particulier la photolithographie, et du besoin de chercher des indicateurs pour la décrire.

Cette partie présente le contexte industriel dans lequel s'est déroulée cette thèse. Nous vulgarisons l'industrie de la micro-électronique, en centrant le discours sur l'étape de photolithographie et la topographie.

2.1 Des puces sur un wafer

Sur le site de fabrication de Crolles 300, les circuits intégrés ont une forme rectangulaire et sont imprimés sur des plaquettes de silicium circulaires de 300mm de diamètre appelées wafers. Ces wafers arrivent vierges sur le site de fabrication, et sont la base pour la fabrication des puces. Au stade où les wafers sont vierges, on peut décider d'y imprimer n'importe quel produit. Par la suite, ce wafer subira de nombreux traitements qui imprimeront les puces-produits dans un quadrillage maximisant le nombre de puces par plaquette de silicium. La figure 2.1 montre une carte d'exposition d'un produit.

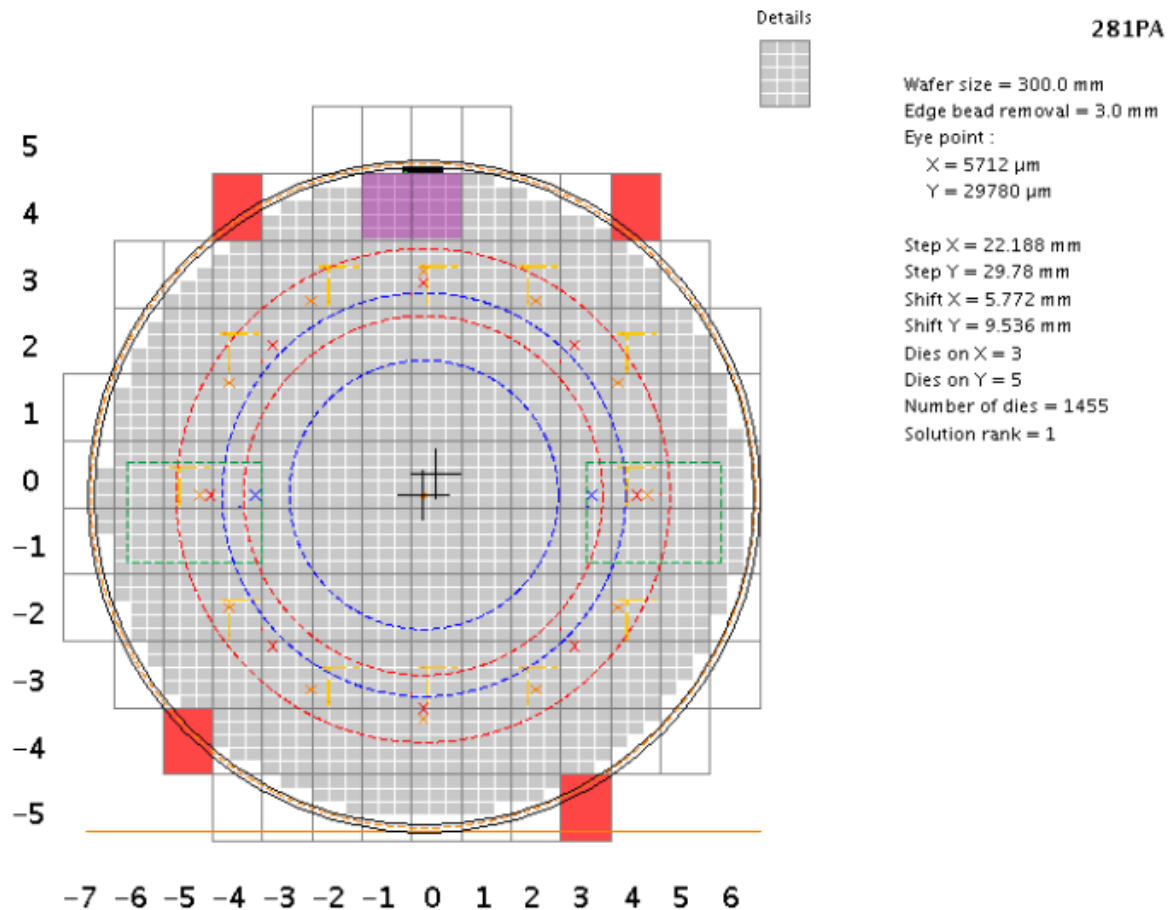


FIGURE 2.1 – Carte d'exposition d'un produit sur un wafer (281PA)

Sur la figure 2.1, chaque cellule grise représente une puce. Le mot carte d'exposition renvoie au procédé de fabrication de la photolithographie, où des ensembles contenant plusieurs puces appelés **champs** sont « exposés » les uns après les autres. Nous reviendrons sur ce procédé dans la partie 2.3.

On voit dans les éléments d'importance de cette figure le StepX (resp. StepY) qui correspond à la dimension du champ d'exposition sur l'axe des abscisses (resp. des ordonnées). On voit le ShiftX (resp ShiftY) qui correspond au dé-centrage entre le centre du wafer circulaire et le centre de la carte d'exposition selon l'axe des abscisses (resp. des ordonnées). Ce décentrage sert à maximiser le nombre de puces entières sur le wafer : il existe une position optimale ajoutant des puces complètes lorsqu'on cherche à insérer un quadrillage dans un cercle. On voit le nombre de puces par champ sur l'axe des abscisses et des ordonnées (Dies on X et Dies on Y). Le nombre total de puces est montré. « Edge bead removal » est une couronne au bord du wafer que l'on retire car elle est dédiée à la manipulation des plaques, pour les saisir et les déplacer par exemple. Ici, il s'agit d'une couronne de 3mm d'épaisseur.

Un wafer verra plusieurs fois une même machine durant son existence. De plus, des wafers différents contenant des produits différents peuvent voir une même machine séparément plusieurs fois. C'est la raison pour laquelle l'industrie standardise sa base de fabrication avec des diamètres fixes pour les wafers : une machine doit pouvoir traiter tout type de produit en étant capable de traiter tout wafer, et doit s'adapter au produit par de la configuration matérielle et/ou logicielle.

Une fois les puces imprimées en quadrillage sur le wafer, elles seront découpées suivant des canaux de découpe entre les puces. Ces canaux ont une largeur inférieure à 100μm. Sur ces canaux de découpe se trouvent des structures de métrologie et de contrôle de procédés (dimension, épaisseur...) permettant par exemple d'aligner le wafer entre ses différents passages

dans différentes machines. La problématique d'alignement est une problématique à part entière, et se nomme « overlay ».

Les structures métrologiques aident la machine à savoir où est situé le wafer. En effet, comme mentionné dans le chapitre précédent, les dimensions sont si petites qu'on a une marge de manœuvre très faible sur un décalage d'un passage à l'autre. Un décalage trop important met en défaut toute une étape, et donc potentiellement tout un procédé de fabrication.

La dimension « horizontale » des puces (axes X et Y) et leur répartition sur la surface du wafer ont été montrées via cette carte d'exposition ainsi que les problématiques d'alignement évoquées. Dans la partie suivante, nous développons la dimension de la profondeur (axe Z) sur le wafer, en parlant de couches minces, élément au cœur de l'industrie de la micro-électronique.

2.2 Micro-électronique et couches minces

Un circuit intégré, qu'on peut appeler puce électronique, est un composant électronique reproduisant une ou plusieurs fonctions électroniques plus ou moins complexes. La fonction désirée est déterminée lors de phases de conception précédant la fabrication. Il s'agit ensuite d'agencer des structures électroniques ayant des fonctions plus ou moins complexes, par exemple des transistors, des balances...

Une fois l'agencement choisi lié à la fonction électronique, le rôle de l'industrie est de bâtir ces interconnexions grâce à un ensemble de briques de fabrication créant des connexions et des isolations. Cela implique de pouvoir placer des matériaux isolants à certains endroits, et conducteurs à d'autres endroits, afin de créer l'agencement électronique propre à la fonction désirée.

Ces contraintes de connexions et d'isolations à satisfaire trouvent leurs solutions dans l'empilement de couches. Le concept d'empilement de couches est fondamental pour comprendre comment l'industrie de la micro-électronique fonctionne, car elle est au cœur des procédés de fabrication. On cherche à empiler des couches les unes sur les autres, qui sont intra-connectées (connectée au sein d'une même couche) et inter-connectées (connectées entre plusieurs couches) à des endroits bien précis.

Pour imprimer la couche mince avec le matériau/la propriété désiré(e), on protège la puce aux endroits où l'impression ne sera pas faite, à l'aide d'une résine, pour appliquer par la suite le traitement sur la totalité de la plaque. Ainsi, la propriété/le matériau sera transféré(e) sur la surface uniquement aux endroits non protégés. Il ne restera plus qu'à retirer la résine pour obtenir le résultat attendu. Schématiquement, on montre le déroulement du procédé pour structurer une couche mince dans la figure 2.2.

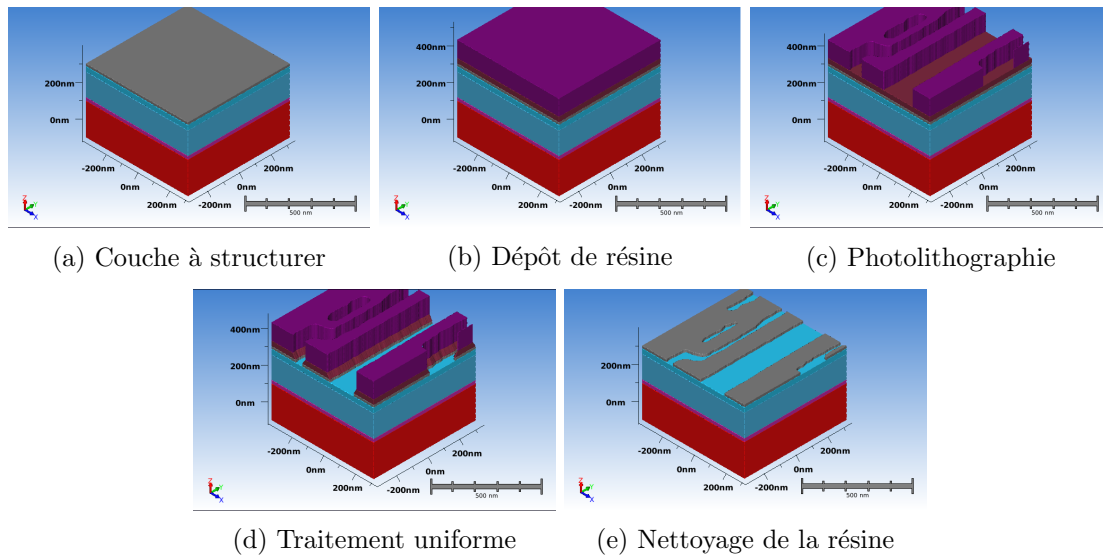


FIGURE 2.2 – Déroulement des étapes pour structurer une couche mince

Dans la figure 2.2, on part d'une couche que l'on souhaite structurer d'une façon déterminée au préalable. Le choix industriel est de protéger la puce grâce à une couche de résine déposée uniformément puis enlevée sur les endroits qu'il ne faut pas protéger grâce au procédé de photolithographie. La suite est de faire le traitement uniforme (gravure, dopage...), et enfin de retirer la résine, pour obtenir le résultat dans la figure 2.2 (e).

Ces couches minces étant omni-présentes, une représentation informatique qui factorise ce concept est nécessaire, et a été proposée par Calma avec le format GDS II, initialement GDS (Graphic Data System) et précédemment RTL (Register Transfer Level) [Kim et al., 2020]. Sur la figure 2.3, ces fichiers correspondent à la deuxième colonne, juste à droite du design (Niveau 1, 2 et 3). Ce format consiste en une liste de polygones étiquetés dont les coordonnées des sommets sont des nombres entiers. Cela rend ce format adaptable et contextuel. On peut par exemple décider arbitrairement de l'unité de la coordonnée. Dans le cas d'une technologie 28nm, la valeur unitaire de la coordonnée est de 0.5nm. Au fil du temps et de la complexification des circuits intégrés, la diminution des dimensions ont fait que ces fichiers pèsent de plus en plus lourds. Ils deviennent de fait de plus en plus difficiles à manipuler et pèsent aujourd'hui (2021) plusieurs giga-octets.

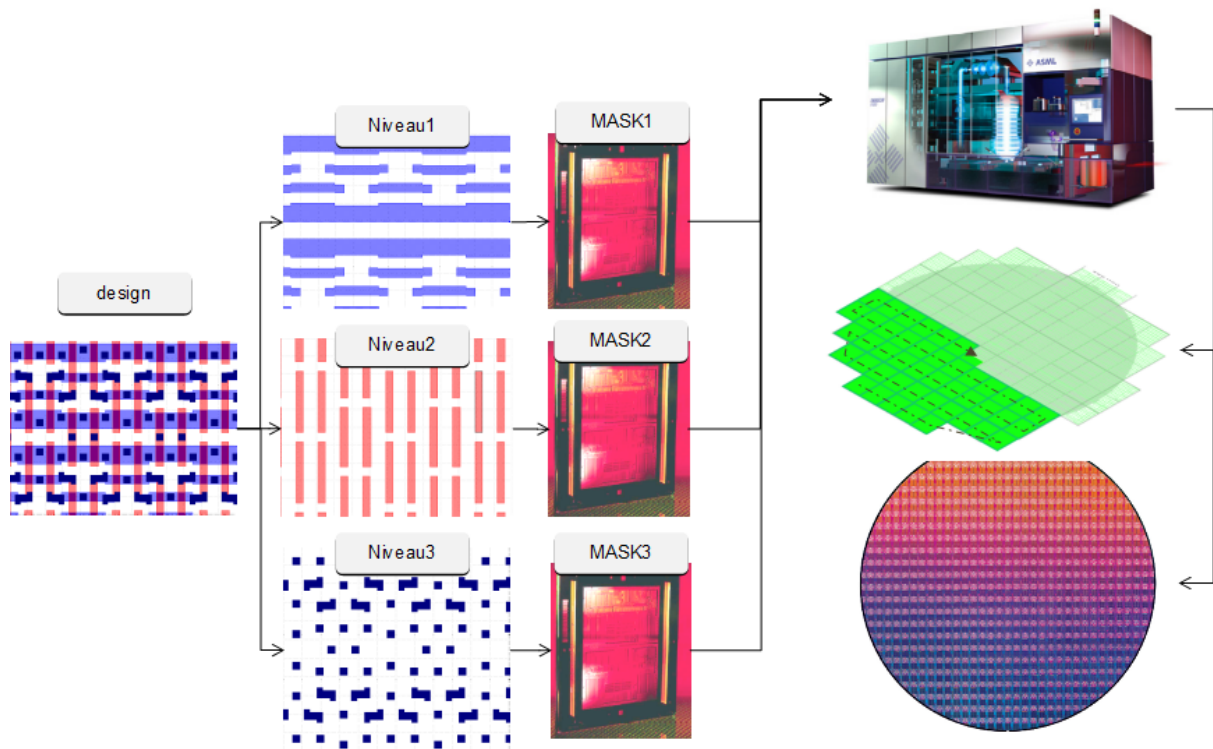


FIGURE 2.3 – Du design au silicium

L’usage des fichiers GDS est d’une telle complexité qu’il n’appartient pas à un humain de regarder un par un ces polygones. Dans les faits, il existe un cahier des charges très précis concernant les fichiers GDS devant représenter des couches minces : il s’agit de trouver, dans l’ensemble de toutes les listes de polygones possibles et imaginables, une liste de polygones qui satisfait des objectifs de connectivité et d’isolation, représentant un design-cible sur la résine.

Ces fichiers servent donc à savoir où doit être déposé un type de matériau, soit conducteur, soit isolant (ou autre), soit à savoir où sera effectué un traitement particulier. Le traitement de la plaque ainsi que les propriétés des matériaux ne sont pas le sujet de cette thèse. Nous nous arrêterons au fait que, dans l’industrie de la micro-électronique, nous souhaitons donner des propriétés aux couches minces en effectuant des dépôts de matière ou des transferts de propriété(s), en suivant un dessin bien précis, contenu dans les fichiers GDS.

Pour dessiner avec précision, on peut prendre un crayon fin, et dessiner. Mais pour dessiner des motifs de façon répétable et rapide, on va préférer l’usage d’un pochoir : le masque de la figure 2.3.

L’étape qui va déterminer la précision avec laquelle les motifs seront imprimés est donc la photolithographie. Ce procédé qui répond au besoin de protéger une puce de façon répétable est développé dans la section suivante.

2.3 Photolithographie

1. Généralités

La lithographie signifie littéralement l’action d’écrire sur la roche (lithos = roche ; graphein = écrire). Il subsiste des abus de langage sur le terme lithographie. Parfois, « lithographie » désigne l’ensemble des étapes visant à imprimer les motifs sur la puce. En d’autres termes, il s’agirait de la totalité des étapes qui visent à produire la puce. Dans d’autres cas, cela désigne uniquement l’étape qui sert à protéger le wafer avec de la résine aux endroits souhaités.

Ici, nous sommes dans la deuxième catégorie mentionnée, où nous expliquerons comment déposer de la résine protectrice en suivant des motifs. Le problème, mentionné précédemment, de dessiner cette protection avec précision demeure.

Aujourd'hui, la technique utilisée à cette fin est la photolithographie (photo = lumière). Les motifs sont imprimés à l'aide d'une résine photosensible dont on recouvre initialement la totalité du wafer (Voir figure 2.2). Les résines peuvent être soit positives, soit négatives. Une résine positive possède la propriété de pouvoir facilement se retirer du wafer une fois exposée à la lumière. Après exposition à la lumière et nettoyage, il ne reste de la résine que sur les endroits qui n'ont pas été exposés à la lumière. À l'inverse, pour une résine négative, il ne reste de la résine que sur les endroits qui ont été exposés à la lumière.

La résine restante sert à protéger le wafer pour l'étape suivante, et la bonne réalisation de la photolithographie conditionne la bonne réalisation du traitement appliqué ensuite (gravure, dopage, autres...). On applique un traitement uniformément sur la totalité de la plaque, et grâce à la résine, seules les zones non protégées sont gravées. On peut ensuite retirer la résine.

2. Principe

La photolithographie met en jeu ces éléments : une source lumineuse, un masque, un système optique et un wafer couvert de résine photosensible. Pour les novices ne connaissant pas ce procédé, un schéma simplifié est proposé dans la figure 2.4.

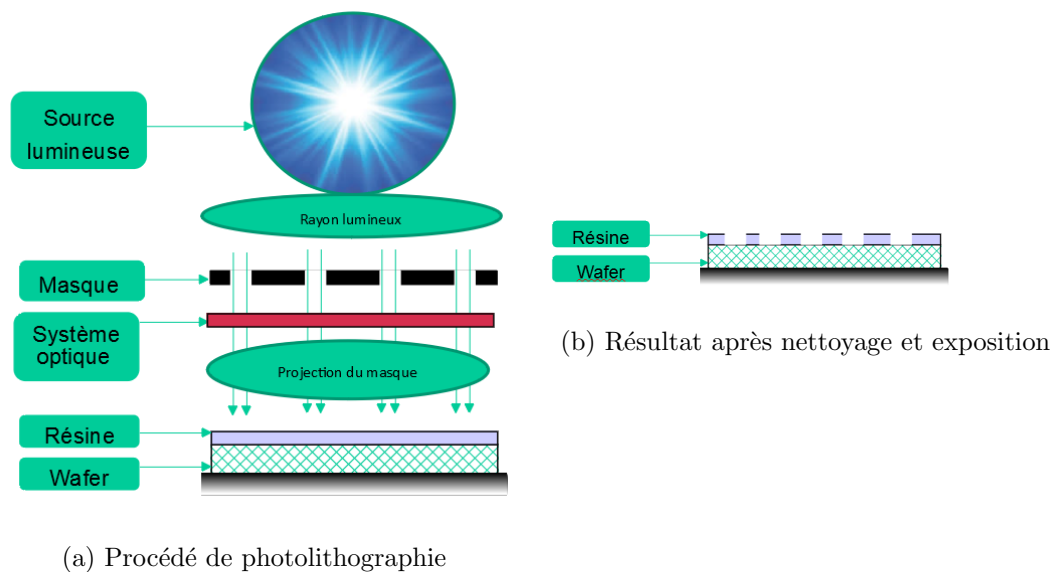


FIGURE 2.4 – Schéma simplifié de la photolithographie

Ce procédé a pour objectif d'imprimer les motifs du masque sur la résine photosensible grâce au système optique : la lentille de projection (voir Figure 2.5). Le mode de fonctionnement du système optique est analogue à celui d'une lentille convergente. Il faut placer le wafer dans le plan focal image du système optique, afin que les motifs du masque soient imprimés sur la résine de façon nette.

La représentation des ombres qui illuminent la résine et transposent les motifs du masque est une simplification pour comprendre le concept de la photolithographie. Le masque possède des parties dites « opaques » de chrome, et des parties « transparentes ». Dans la réalité, et compte-tenu des dimensions, la lumière subit le phénomène physique de diffraction ainsi que d'autres phénomènes qui ne sont pas le sujet de cette thèse.

3. Le masque

Le masque est travaillé grâce à des modèles optiques simulant le trajet de la lumière pour prendre en compte ces phénomènes, grâce à des modèles simulant le profil de la résine en fonction de paramètres tels que la longueur d'onde, la durée d'exposition à la lumière, et des propriétés de la résine. Pour cette thèse, on retiendra que l'information des motifs est contenue dans le masque, et est transmise dans la résine grâce à de la lumière et au système optique. Le format de représentation du masque est également le format GDS précédemment évoqué. Ce masque est un ensemble de polygones opaques, et là où il n'y a pas de polygones, la lumière passe.

Le problème de fabriquer le masque avec précision demeure. Au final, on doit avoir un élément qui satisfait les exigences de précision. Dans la figure 2.5, la lentille de projection a un agrandissement optique en 4X. Cela signifie que les motifs sur le masque sont 4 fois plus grands que le résultat d'impression dans la résine.

La technique de fabrication du masque est plus longue, mais plus précise. Elle s'apparente à l'utilisation d'un crayon, pour continuer sur la métaphore faite précédemment. Il s'agit de lithographie à faisceau d'électrons : une technique permettant de dessiner des motifs avec une résolution descendant jusqu'au nanomètre. Cette technique permet de satisfaire les contraintes liées au masque concernant les difficultés de fabrication du masque.

Le masque doit être **fabricable** en satisfaisant des règles de dessin spécifiques, par exemple, ne pas mettre deux polygones trop proches, car il sera impossible de les fabriquer. Il existe une liste très longue de contraintes pour qu'un masque soit manufacturable, car les dimensions mises en jeu sont tellement petites que des phénomènes physique auparavant négligeables deviennent critiques (diffraction, distorsion, effets quantiques...). Cette liste est contenue dans un DRM (Design Rule Manual). Il s'agit d'une formalisation de la limite de la technique de fabrication.

Deuxièmement, le masque doit **satisfaire ses objectifs** en tant que masque optique, en s'assurant que le résultat sur la résine sera bien celui que l'on souhaite. Pour cela, on dispose d'un modèle optique prenant en entrée la liste de polygones représentant le masque, et en sortie le reste de résine sur le wafer. On possède également l'objectif de résine qu'on souhaite avoir sur le wafer. Avec une comparaison entre l'objectif de résine et la résine simulée, on converge vers une solution, grâce à des algorithmes de recherche présents dans un logiciel spécialisé comme Calibre fourni par Mentor.

4. L'exposition

Le temps d'exposition pour un wafer entier est de l'ordre de 15 à 30 secondes, soit 120 à 240 wafers par heure. On peut ré-utiliser un masque pour exposer plusieurs wafers. La combinaison d'un scanner et d'un masque constitue une étape du procédé de fabrication pour une multitude de wafers d'un même produit.

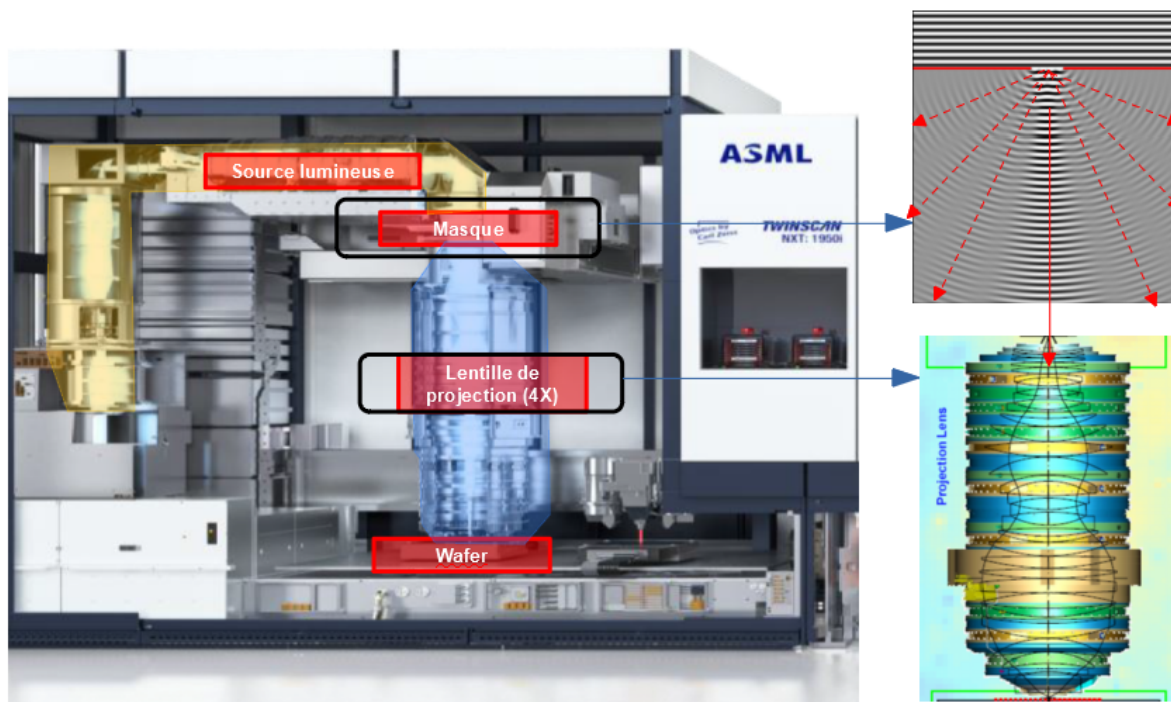


FIGURE 2.5 – Chemin optique dans un scanner ASML TWINSCAN

Sur la figure 2.5, on voit un scanner réel avec tous les éléments précédemment représentés sur le schéma de la figure 2.4. La source lumineuse (d’une longueur d’onde entre 400nm et 13.4nm) envoie de la lumière à travers le masque. Le masque projette le design souhaité dans le plan objet de la lentille de projection où l’on place le wafer. Dans la partie suivante, nous présentons la topographie, élément venant perturber cette distance focale lors de l’exposition.

5. Contrôle du focus

La lentille de projection étant analogue à une lentille convergente, une notion clé pour la réussite de la photolithographie est la notion de distance focale. Il faut placer le wafer à une distance exacte de la lentille pour avoir une image nette.

Le contrôle du focus est la gestion du placement du wafer par rapport au système optique lors de la photolithographie. Le « best focus » est déterminé pendant la phase de recherche et développement, durant laquelle on trouve la meilleure distance d’exposition d’un wafer pour un niveau donné. Pour ce faire, on fait une FEM. Sur cette FEM, on fait varier le focus sur un axe X, et la dose sur un axe Y. La dose est un paramètre qui est le temps d’exposition. Le focus est la distance focale du wafer à la lentille que l’on fera varier. Lors de la calibration de la photolithographie, les limites de l’exposition sont déterminées sur un wafer vierge, par une FEM.

Dans l’espace continu dans lequel nous vivons, rien ne peut être exact. La distance focale à laquelle le wafer sera situé ne fait pas exception, et les marges de manœuvre sur la distance focale sont inférieures à 100nm dans le cas des nœuds technologiques 28nm.

Un élément complique le placement du wafer à une bonne distance focale : la variation de hauteur, appelée topographie. La topographie est compensée en grande partie par le scanner en suivant celle-ci au mieux. Cette topographie était négligée jusqu’à récemment car elle n’était pas critique. Mais depuis peu, celle-ci vient gêner l’exposition et le positionnement du wafer dans le plan focal. L’analyse de la topographie sera au cœur du travail de cette thèse, de par ses interactions devenues critiques avec les procédés de fabrication.

2.4 Topographie

2.4.1 Origines de la topographie

La topographie est une variation de hauteur sur une surface donnée. Celle-ci est toujours relative à une autre surface. Par exemple, dans le cas de la Terre, on se réfère au niveau de la mer pour déterminer les hauteurs alentours. Pour la topographie d'un wafer, la référence à une surface donnée n'est pas toujours la même selon ce que l'on souhaite regarder.

Lorsqu'on effectue un empilement de couches minces les unes à la suite des autres, on génère inévitablement de la topographie. Si il n'y a pas de topographie, c'est qu'il n'y a pas de puce ! Certaines étapes affectent plus la topographie que d'autres. Pour des raisons physiques, lorsqu'on effectue un dépôt de matière ou qu'on retire de la matière, la topographie est altérée. Certaines étapes sont neutres pour la topographie, notamment des étapes métrologiques.

Démêler les composantes de topographie est une façon de relier la topographie à un ensemble de métiers. « Quel atelier doit faire attention à quelle composante ? », « Qui est responsable de cette topographie, et qui sera impacté par cette topographie ? ». Dans la figure 2.6, on montre un exemple de composantes existantes sur un wafer.

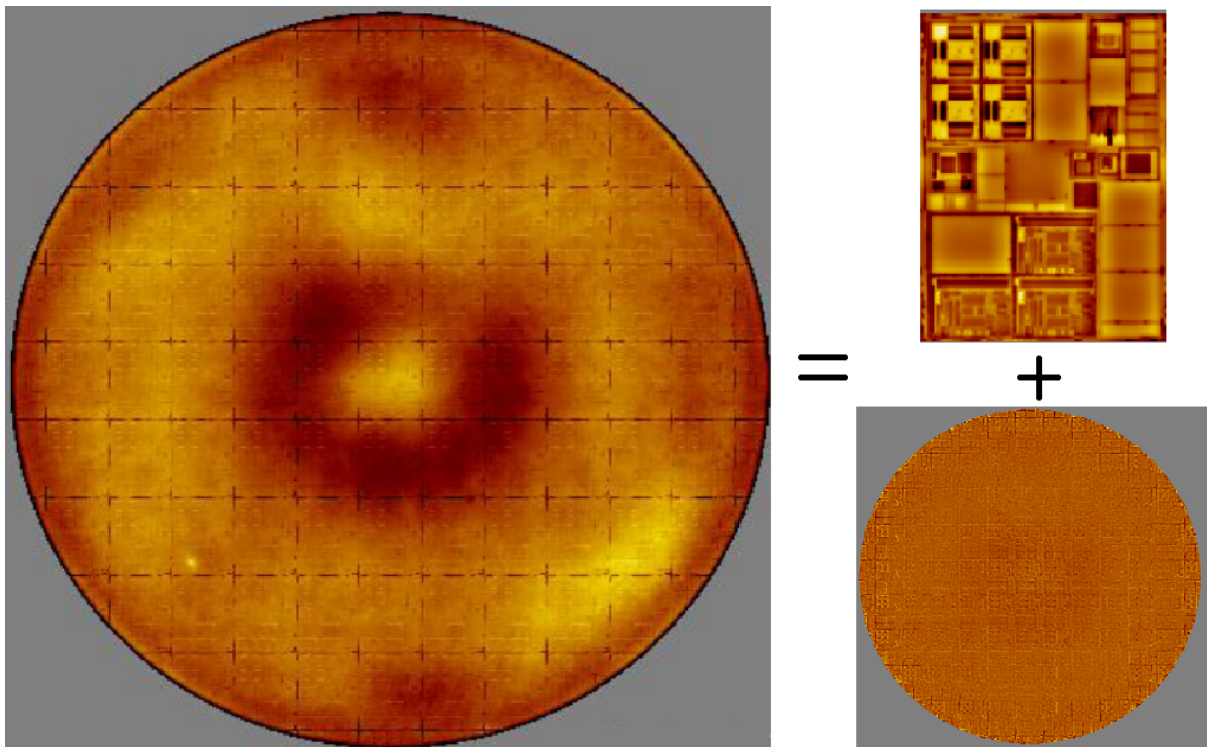


FIGURE 2.6 – Composantes de topographie : topographie totale (gauche), topographie d'une puce (haut droite) et topographie wafer (bas droite)

Sur la figure 2.6, on voit la topographie du wafer entier à gauche. La composante puce apparaît à droite, et en dessous de celle-ci on voit la composante du wafer à laquelle on a retiré la composante puce sur chaque puce du wafer.

Au fur et à mesure que l'industrie se complexifie, des éléments auparavant négligeables ont des interactions qui deviennent critiques. La topographie fait partie de cela, car des variations de hauteurs de quelques dizaines de nanomètres deviennent critiques, alors que ces variations étaient négligeables auparavant.

Dans la sous-partie suivante, nous parlons de la topographie, de sa caractérisation et de son atténuation.

2.4.2 Caractérisation et atténuation de la topographie

1. Rugosité : une mesure de variabilité

On parle de « rugosité » pour mesurer une variation dans une zone d'un wafer. La rugosité est une distance à une référence sur une surface donnée. Cette distance peut être définie de différentes façons, de même que la référence.

Prenant un ensemble de points sur une surface, on peut avoir différentes rugosités : on peut définir la rugosité comme étant la dispersion autour de la valeur moyenne. Dans ce cas, la référence est la valeur moyenne de l'ensemble de points, et la rugosité définit alors la dispersion de ces points. On peut également définir la rugosité comme la différence entre la valeur maximale et la valeur minimale de l'ensemble de points (Peak-to-Valley). La rugosité est une valeur qui quantifie à quel point une surface comporte des variations, et chacune des différentes définitions de rugosités ont un sens différent.

La notion de rugosité traduit une volonté présente dans les contrôles de procédés : mesurer la variabilité. Dans la micro-électronique, la topographie possède plusieurs composantes, chacune ayant des échelles différentes. Elle possède une composante liée au wafer vierge, une composante liée à chaque étape de fabrication, pouvant être à l'échelle de la puce, du champ, ou du wafer selon l'étape. Chacune de ces composantes n'est pas totalement systématique et comporte une variation aléatoire. Par exemple, aucun wafer vierge n'est parfaitement identique à un autre, et ces variations aléatoires ont des échelles différentes.

Le choix de la référence pour mesurer la rugosité, le choix de la surface de référence pour donner une hauteur de topographie sont des choix qui répondent à des questions : Qu'est-ce qui est « normal » pour une topographie ? Que veut-on sur la plaque, et, à quel point est-on écarté de ce que l'on veut ?

2. CMP : Polissage Mécano-Chimique

La topographie est un élément qui peut gêner le passage d'une étape à une autre. Celle-ci doit être maîtrisée, et dans la plupart des cas, on souhaite que les wafers soient les plus plats possibles. L'atelier en charge du contrôle de la topographie est l'atelier CMP (Chemical Mechanical Polishing).

Cet atelier polit la plaque avec un combiné d'action chimique et mécanique. L'action mécanique est un « pad » qui tourne pour polir physiquement le wafer. L'action chimique est l'utilisation d'une « slurry » corrosive et abrasive qui attaque la surface avec laquelle elle est en contact. L'action corrosive agit sur la topographie sans mouvement, et l'action abrasive accentue l'action mécanique du « pad ». La figure 2.7 montre le procédé.

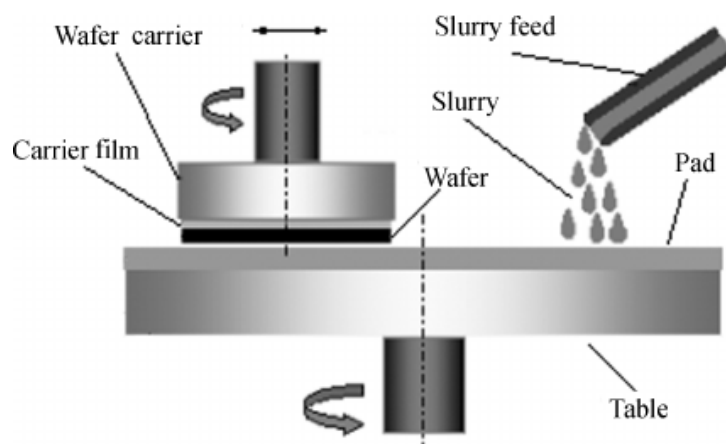


FIGURE 2.7 – Schéma de fonctionnement du polissage mécano-chimique (CMP)

Dans la figure 2.7, le wafer est fixé au « Wafer carrier », et la « slurry » est déposée sur le « pad » en rotation afin de polir la surface du wafer au mieux.

2.5 Conclusion

Dans ce chapitre, nous avons présenté le contexte industriel de STMicroelectronics, en ciblant le propos autour des couches minces, et du procédé de photolithographie permettant cet empilement de couches minces. Ce chapitre 2 fait partie de l'introduction du travail de thèse. Dans le chapitre suivant, nous commencerons à parler du travail réalisé, en spécifiant le cas d'étude de traitement de données de la thèse.

Chapitre 3

Données PWG : Topographie et usage seul

Ici, on considère que le lecteur a lu au moins, soit le chapitre 1, soit le chapitre 2. Nous faisons le lien entre le métrologue cherchant à valoriser sa mesure et les ateliers cherchant à l'exploiter.

Dans ce chapitre, nous abordons les points suivants :

1. Le fonctionnement de l'outil PWG de KLA, de sa volumétrie de données, de sa précision et de rapidité de mesure d'un wafer. Cet outil est récent, et on cherche à bien valoriser ses mesures. Il mesure un wafer qui ne prend pas en compte l'agencement des puces.
2. Du travail effectué qui met en lumière différents objectifs : (1) Lier la mesure au produit et au wafer (2) Enrichir la mesure pour lier les ateliers à cette mesure, en particulier l'atelier CMP.

3.1 Présentation de l'outil PWG de KLA pour mesurer la topographie

Dans cette section, nous présentons l'outil PWG de KLA, capable de mesurer de multiples grandeurs sur un wafer entier. La valeur que nous analysons est la topographie.

Bien que la métrologie ne soit pas le sujet de cette thèse, il est utile de voir comment les façons de mesurer sont complexes dans l'industrie des semi-conducteurs, afin de comprendre les biais qui peuvent être présents dans les mesures qui nous seront fournies par la suite. Les mesures sont issues de connaissances des phénomènes physiques, et également de traitements sur les signaux collectés pour en tirer une grandeur.

3.1.1 Technique de mesure

Dans cette sous-partie, nous décrivons comment l'outil PWG mesure la nano-topographie. La technique de mesure du PWG est l'interférométrie. C'est une technique qui utilise une source lumineuse qui se divise en deux faisceaux cohérents, et qui mesurera les interférences entre une référence (un miroir), et la topographie à mesurer. La figure 3.1 montre schématiquement ce concept.

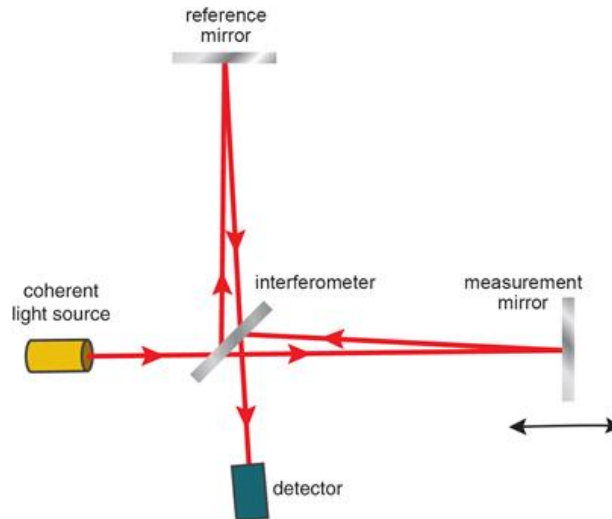


FIGURE 3.1 – Fonctionnement de l'interférométrie - *Source : [Renishaw, 2015]*

Les deux sources se rejoignent et interfèrent, puisqu'elles sont cohérentes et déphasées. Ce déphasage constitue une mesure de la différence de hauteur qui peut se calculer en fonction de la longueur d'onde utilisée, de la distance entre le miroir et le wafer et d'autres paramètres.

La surface obtenue est une surface « modulo », étant donné le caractère ondulatoire de la lumière. Une fois le déphasage obtenu, une deuxième étape du mesurage est la reconstruction de la surface. Cette reconstruction peut être imprécise si la surface que l'on souhaite mesurer ne renvoie pas totalement la lumière. C'est pour cela que sur le schéma, on considère que la surface à mesurer est un miroir.

Le wafer n'est pas une surface totalement réfléchissante. Celle-ci absorbe une partie de la lumière avant d'en renvoyer une partie, donc l'interférométrie n'est pas précise. On souhaite que la surface du wafer possède les mêmes propriétés qu'un miroir. Pour ce faire, on ajoute une couche d'un matériau optiquement dense qui renvoie la quasi-totalité de la lumière : le tantale.

Le retour du bâton de cette technique de dépôt de tantale est son irréversibilité. Une fois le tantale ajouté, on doit détruire le wafer. Cette mesure est donc extrêmement coûteuse.

3.1.2 Volumétrie, précision

Le PWG est capable de mesurer une plaque entière en 60 secondes. La mesure du PWG a une résolution latérale de $100\mu\text{m} \times 100\mu\text{m}$ et une précision au dixième de nanomètre. Il y a environ 7 millions de points par mesure, et chaque mesure pèse environ 180Mo. C'est une mesure rapide pour une plaque entière, comparativement à d'autres techniques de mesures de topographie comme l'AFM [Valade, 2020].

À titre comparatif, l'AFM peut descendre jusqu'à une résolution latérale de l'ordre du nanomètre, et jusqu'à une précision de 0.025nm . L'ordre de grandeur de la durée d'une mesure pour une portion de puce ($150 \times 150 \mu\text{m}$) de l'AFM est de quelques minutes. Pour rappel, un wafer est circulaire de 300mm de diamètre, donc l'outil PWG est très rapide.

Un outil de mesure est un moyen d'assurer des contrôles de procédés, ou bien de calibrer des nouveaux procédés lors de la phase recherche et développement. Mais pour remplir ce rôle, on doit régler des problématiques de système d'information avec notamment la gestion du stockage à court-terme, moyen-terme et long-terme des mesures, sous quel format et pour quels usages. Ces problématiques sont d'autant plus importantes si la mesure a un coût élevé, comme c'est le cas pour cette mesure. Ici, on peut mesurer jusqu'à 60 wafers par minute, ce qui constitue une génération de $60 \times 180 = 10800$ Mo par heure. Dans ces échelles de génération de données, on souhaite à la fois valoriser ces mesures et en assurer un stockage pérenne.

Comme expliqué, cette mesure détruit la plaque. Si l'on souhaite faire du suivi de procédés, on doit avoir une mesure qui ne détruit pas la plaque. La possibilité est alors de ne pas mettre de dépôt réfléchif sur la surface du wafer, quitte à avoir une mesure imprécise. Cette mesure imprécise doit alors être observée avec du traitement a posteriori afin d'y donner le maximum de sens pour prendre les décisions associées à la mesure (arrêt du lot, arrêt du wafer, validation de l'étape...). Nous traitons un tel cas dans le chapitre 5.

Dans ce chapitre, nous nous intéressons à une mesure précise uniquement. L'enjeu est de donner un maximum de valeur à une mesure effectuée par cet outil, en prenant en compte à la fois les besoins des ateliers et des métiers.

3.1.3 Traitement de la mesure : Filtre double Gaussien

Avec l'outil de mesure est fourni un logiciel de traitement de données : PWGtool. Cet outil de traitement de données offre des possibilités.

Le filtre double Gaussien est présenté dans la figure 3.2. Une définition d'un filtre de Gauss peut être trouvée dans [Bergounioux, 2011] (On développe les filtres dans une section plus loin dans la thèse ainsi que leurs définitions : un filtre de Gauss est montré en figure 4.14)

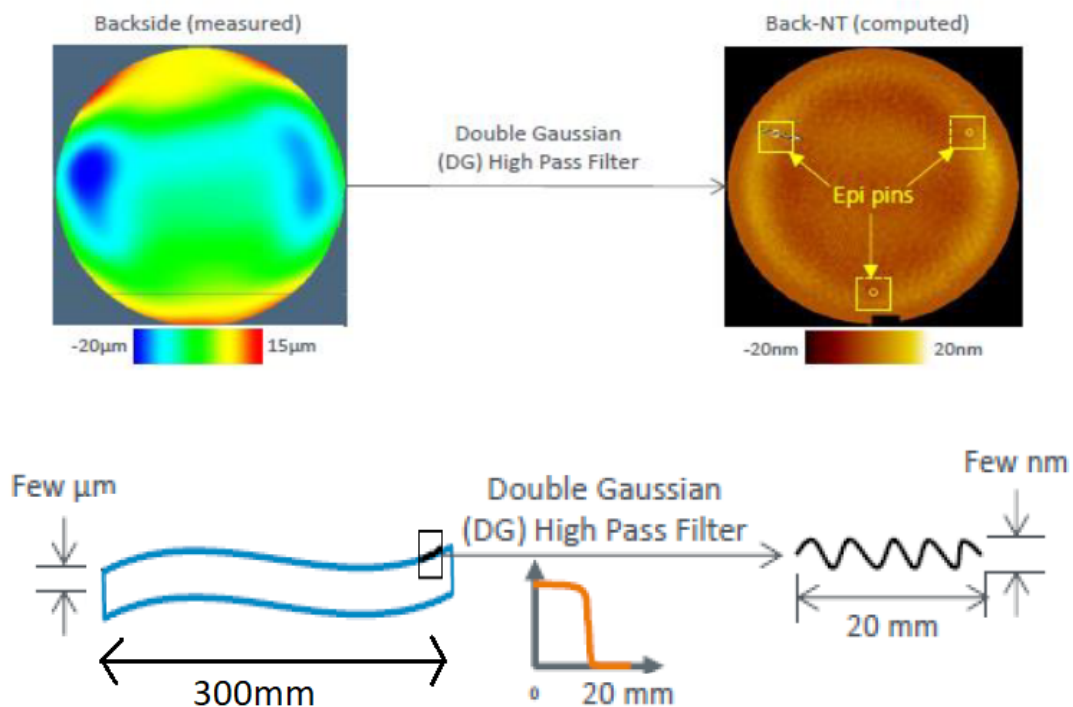


FIGURE 3.2 – Filtre double Gaussien pour le PWG - *Source interne*

Ce filtre est un « leveling ». Il s'agit d'une méthode pour retirer la composante wafer de la puce. Pour ce faire, on effectue deux filtres de Gauss : l'un avec une portée courte, l'autre avec une longue portée. On soustrait le résultat de la longue portée au résultat de la portée courte, et il ne reste ainsi que la composante puce.

Dans la figure 3.2, ce filtre est présenté comme un filtre passe-haut. Les variations hautes fréquences sont conservées (composante puce), et les variations basse fréquences sont retirées (composante wafer). Un tel filtre basé sur une convolution peut s'appliquer dans le domaine fréquentiel comme dans le domaine spatial. La représentation passe-haut est l'implémentation réelle de cet algorithme effectué dans PWGtool, l'outil de traitement proposé par KLA.

On voit que les échelles de variations ne sont pas du tout les mêmes : 35 μ m d'intervalle de variations dans l'image initiale de gauche, et 40nm d'intervalle de variations dans le résultat final de droite.

Le choix des différentes portées des filtres Gaussiens permettent de mettre en avant différents éléments. Dans la figure 3.3, on montre qu'un leveling effectué avec une grande portée de 2mm et une petite portée de 0.1mm (DG 20 :1) fait un filtre passe-haut avec une fréquence de coupure moins haute que sur une grande portée de 0.5mm et une petite portée de 0.1mm (DG 5 :1). On appelle signature d'une entité une trace visible provenant de cette entité. On peut faire ressortir des signatures provenant de différents ateliers.

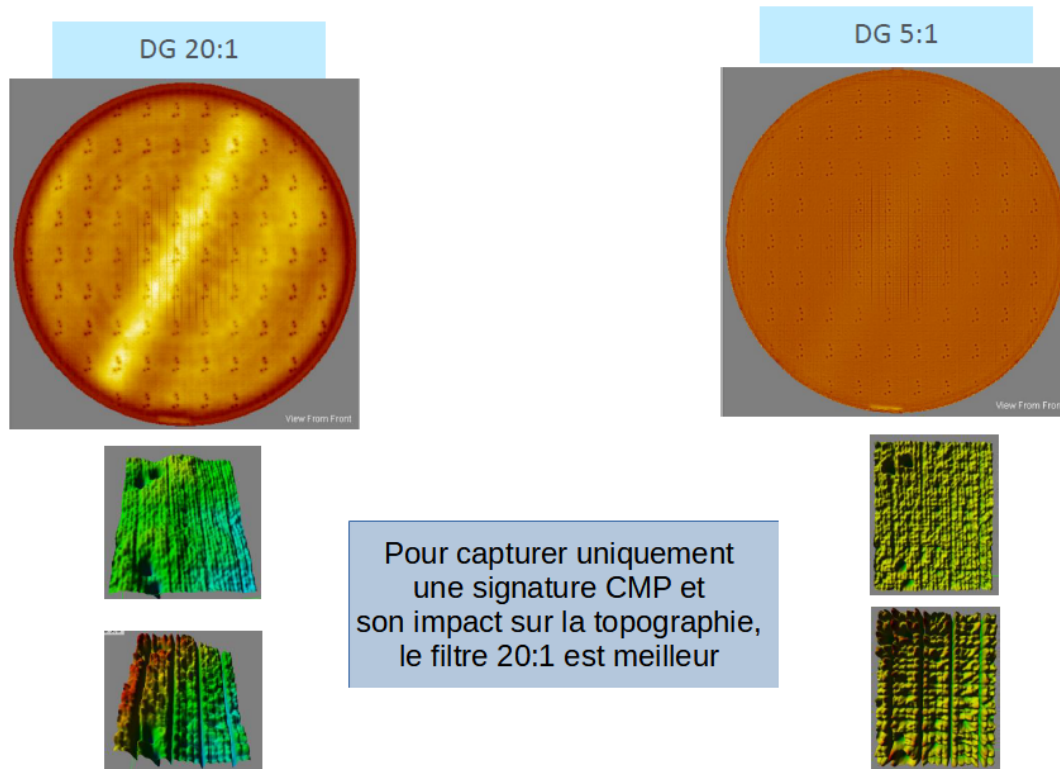


FIGURE 3.3 – Choix des portées différentes de filtres double Gaussien pour faire ressortir des signatures CMP - *Source : KLA Tencor*

Comme expliqué dans la figure 3.3, cela aide des ateliers avec des problématiques spatiales plus hautes fréquences comme l'atelier CMP de faire ressortir ces composantes.

3.1.4 Usages de la mesure

Dans le chapitre précédent, l'importance de la variation est mise en avant. Dans l'idée de détecter la variation, un indicateur disponible dans le traitement est le « Peak-to-Valley ». Pouvant être considéré comme l'une des nombreuses expressions de rugosité, le Peak-to-Valley est la différence entre la valeur la plus haute et la valeur la plus basse d'un ensemble de points.

Dans l'industrie de la micro-électronique, on cherche à savoir si une puce peut passer à l'étape suivante. Ainsi, on va regarder dans ce cas le Peak-to-Valley d'une puce, c'est-à-dire la variation à l'échelle de la puce. Cette variation est une vérification qui « confirme » si une puce peut passer à l'étape suivante de par sa définition. Un Peak-to-Valley bas implique une puce très homogène, donc une puce pouvant passer à l'étape suivante. Cependant, un Peak-to-Valley élevé ne peut être suffisant pour conclure sur l'état de la puce. Cet indicateur permet donc de donner des conclusions « pessimistes » sur le nombre de puces qui pourront parfaitement passer à l'étape suivante.

Dans la figure 3.4, on voit trois cartographies. Celle à gauche est la nano-topographie obtenue après application d'un filtre double gaussien 20 :1 (2mm de portée longue, et 0.1mm de portée courte : voir précédemment). Celle du milieu est une cartographie des Peak-to-Valley obtenus sur chacune des puces de la carte de gauche. Celle de droite est la même cartographie précédente, mais avec comme nano-topographie d'origine l'application d'un filtre 5 :1 au lieu de 20 :1.

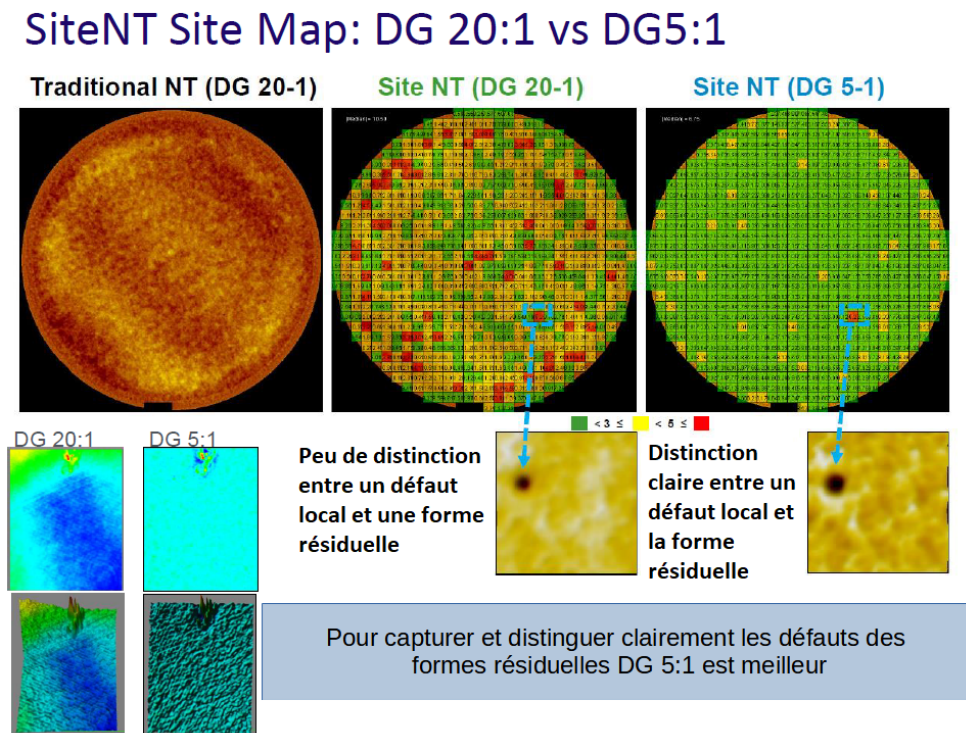


FIGURE 3.4 – Choix des portées différentes de filtres double Gaussien pour faire ressortir des défauts - *Source : KLA Tencor*

Cette figure a vocation à montrer l'importance du choix du filtre double gaussien lorsqu'on veut détecter des petits défauts, petit voulant dire d'une taille inférieure à celle d'une puce. Dans la partie précédente, on montrait que le filtre 20 :1 faisait ressortir les traces macroscopiques, et dans ce cas, on montre que les défauts petits sont mieux mis en lumière avec un filtre 5 :1.

Sur la cartographie de droite, seule une puce est rouge et on voit en regardant la vraie topographie que celle-ci comporte un défaut. Ce défaut impacte significativement le Peak-to-Valley.

Sur la cartographie du milieu, on voit qu'il y en a beaucoup plus. La portée des filtres fait partie des choix que la métrologie doivent faire pour répondre aux besoins des ateliers.

Ces choix sont dépendants des traitements disponibles sur le logiciel de traitement informatique fourni avec l'outil de mesure. Le Peak-to-Valley est une mesure utilisée pour le contrôle de procédés comme le CMP. On souhaite qu'à l'issue de ce procédé, la puce soit « la plus plate possible », dans le but de passer à l'étape suivante. Dans ce but, on utilise le Peak-to-Valley chez STmicroelectronics comme critère de validation.

Dans la figure 3.5, on montre un suivi de différents lots par évaluation du Peak-to-Valley (PV). Si celui-ci est suffisamment bas, on peut considérer que la puce est très homogène. Si celui-ci est haut, un ingénieur regardera ce qu'il en est en détail.

Front Site NT PV: Flagging Sites with Defects, Capturing Unique CMP signatures.

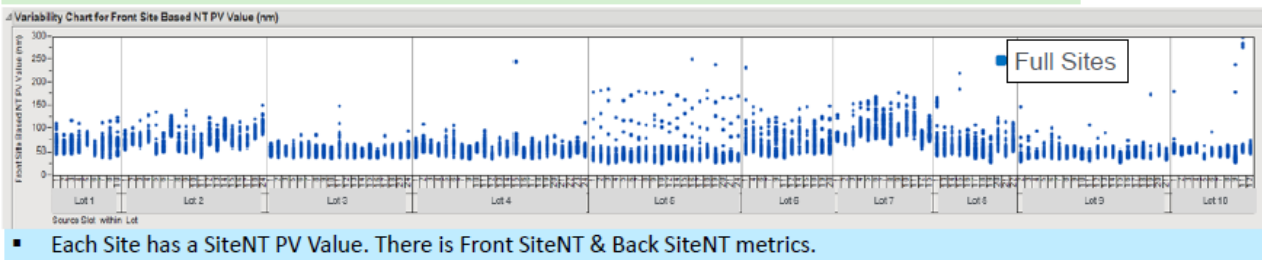


FIGURE 3.5 – Suivi de procédé par Pic-to-Valley pour chaque puce (site) - *Source : KLA Tencor*

On voit sur cette figure que les lots 1 et 2 n'ont pas de variations inquiétantes. On voit sur le lot 3 que certains sites montent plus haut que le groupe de puces de leur lot. Dans le lot 4, une grosse variation est présente sur l'une des puces, et nécessitera le regard d'une personne. L'apport de ce critère est d'éliminer un maximum de puces qui auront besoin d'être regardées en détail par un ingénieur. On voit sur le lot 6 qu'il y a beaucoup de points que l'on devra regarder, et qu'il faut tracer l'historique de ce lot pour expliquer cela.

La création de métriques pour le suivi de procédé est nécessaire dans l'industrie pour minimiser l'intervention humaine, sans pour autant remplacer les personnes comprenant les procédés. À mesure que les besoins métrologiques évoluent, non seulement les techniques de mesure évoluent, mais le travail de fond concernant les métriques doit également évoluer pour satisfaire ces besoins. Dans la suite du chapitre, nous présentons les travaux effectués qui vont dans le sens de la création d'indicateurs participant à l'aide à la décision des métiers.

3.2 La mesure de topographie en tant que telle

A picture is worth a thousand words - W. Churchill. Mais si une image vaut 1000 mots, comment extraire 1000 mots d'une image ?

Dans les chapitres et parties précédentes, nous avons fait état du contexte industriel incluant les données, le contexte métier, et de l'outil dont nous allons traiter les données : le PWG. À partir d'ici, le travail de thèse de réflexion et de traitement de données commence.

3.2.1 Introduction : Enjeux de la description de mesure

Dans ce chapitre, nous répondons à plusieurs questions successives visant à rendre intelligible la mesure de nano-topographie fournie par le PWG de KLA. On veut être capable de voir le maximum d'information contenue dans cette mesure dense pour aider à prendre des décisions et à améliorer la compréhension des procédés de fabrication.

L'aspect de traitement de données qui sera mis en lumière et développé dans la partie 3.2 est la séparation des composantes du wafer. C'est l'objet de la partie 3.2.2, où des composantes sont démêlées pour enrichir la compréhension de la mesure. Dans la partie 3.3, nous nous intéressons à la description de la mesure en tant que mesure de topographie.

3.2.2 Des puces sur un wafer

Cette sous-partie fait état de la démarche effectuée pour passer d'une mesure d'un wafer entier à des composantes faisant sens pour différents objets : puces, champs et wafers. Démêler ces composantes permet d'attribuer chacune de celles-ci à un ensemble d'ateliers gérant des procédés à l'échelle de ladite composante.

3.2.2.1 Méta-données de la mesure

La mesure de l’outil PWG est une mesure d’un wafer, et n’est pas liée à la carte d’exposition du produit contenu sur le wafer. La mesure fournit une liste de points dont la dimension latérale est de $100\mu\text{m}\times 100\mu\text{m}$, et la hauteur est donnée en nanomètres.

Comme le montre la figure 3.6, on observe une photo-répétition à l’échelle de la puce sur la mesure fournie par l’atelier métrologie. Démêler les composantes permet de relier des composantes à des ateliers, afin d’avoir des liens pertinents entre la mesure et les métiers déjà présents à ST.

Dans un premier temps on souhaite en faire une liste d’objets similaires : une liste de puces, une liste de champs. Dans un second temps, en faire la moyenne spatiale, puis retirer la moyenne de l’unité de photo-répétition sur chaque unité pour faire ressortir la composante globale : ici la composante wafer.

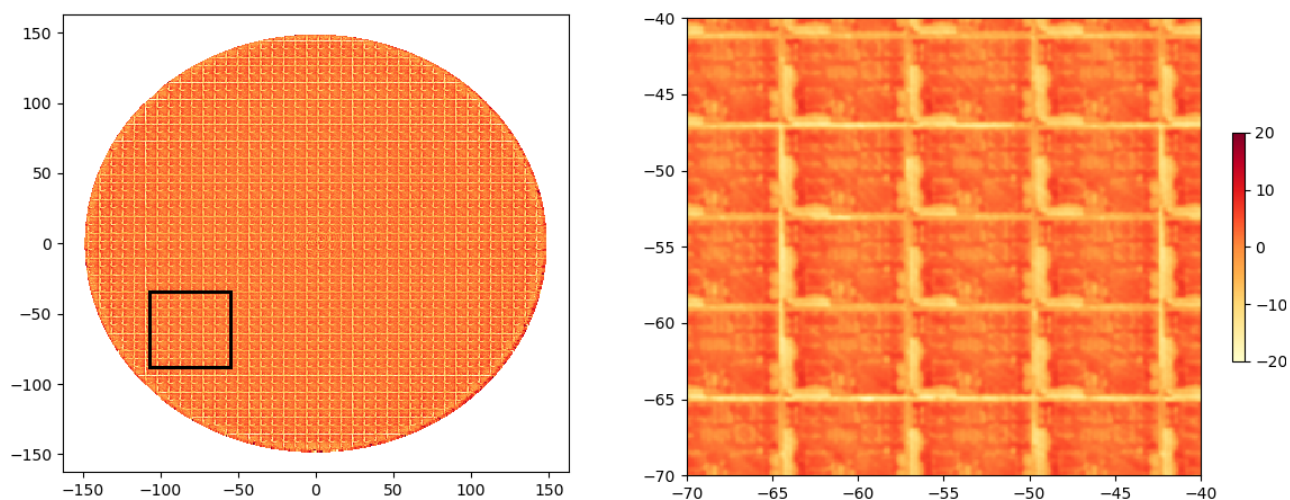


FIGURE 3.6 – Mesure brute PWG (gauche) et zoom (droite)

Nous disposons de plusieurs informations dans l’en-tête de la mesure produites, en particulier « l’offset » de mesure, qui correspond à la coordonnées du (0, 0) de la carte d’exposition (voir figure 2.1 pour la carte d’exposition de ce même produit) dans les coordonnées de la mesure PWG. En d’autres mots, c’est la coordonnée du point en bas à gauche de la puce située la plus au centre du wafer. Cet offset est essentiel, car comme mentionné, la mesure n’est pas reliée à la carte d’exposition, et l’outil PWG ne fait que mesurer un wafer. Cet offset permet de se repérer puis de s’aligner avec la carte d’exposition du produit situé sur le wafer, aidant ainsi à interpréter les données mesurées.

Une autre information contenue dans la carte d’exposition ainsi que dans l’en-tête de la mesure, est la dimension en X et en Y d’un champ (pour rappel, un champ est un ensemble rectangulaire contenant plusieurs puces qui sont exposées en même temps lors de la photolithographie), et le nombre de puces en X et en Y dans chaque champ. Ces dimensions ne sont pas nécessairement des multiples de $100\mu\text{m}$, et causent un dés-alignement avec la grille de mesure de l’outil PWG.

Une donnée non contenue dans la mesure, mais connue du métier, est la dimension exacte des canaux de coupes situés entre les puces (appelés « scribelines »), sur lesquels se trouvent également des structures métrologiques [Li et al., 2009]. Ces canaux sont des lignes qui serviront à découper les puces. Ils ont une dimension de $64\mu\text{m}$ de largeur dans notre cas d’étude. Il y a donc un espace de $64\mu\text{m}$ entre chaque puce dans les 4 directions. Un dés-alignement entre les bords de puces et les bords des pixels de la mesure PWG est engendré, car comme pour la

dimension des puces, cette valeur n'est pas un multiple de $100\mu\text{m}$.

Les informations précédemment mentionnées, à savoir l'offset, la dimension en X et en Y d'un champ, l'espace situé entre chaque puce, ainsi que le nombre de puces en X et en Y dans chaque champ permettent l'enrichissement de la mesure, à commencer par l'extraction des puces et champs.

3.2.2.2 Puces, champs complets et alignement

Une mesure PWG est une grille constituée de carrés de $100\mu\text{m}$ et placée sur la totalité du wafer. Chaque carré de la grille aura une valeur de nano-topographie. Si l'on considère le point le plus en bas à gauche d'une puce d'une dimension de $100\mu\text{m}$, il n'est pas forcément totalement aligné avec cette grille. D'une puce à l'autre, le décalage de $64\mu\text{m}$ dû aux canaux de découpe ainsi que la dimension de la puce désaligne les puces avec cette grille quoi qu'il en soit. La figure 3.7 illustre ce dés-alignement.

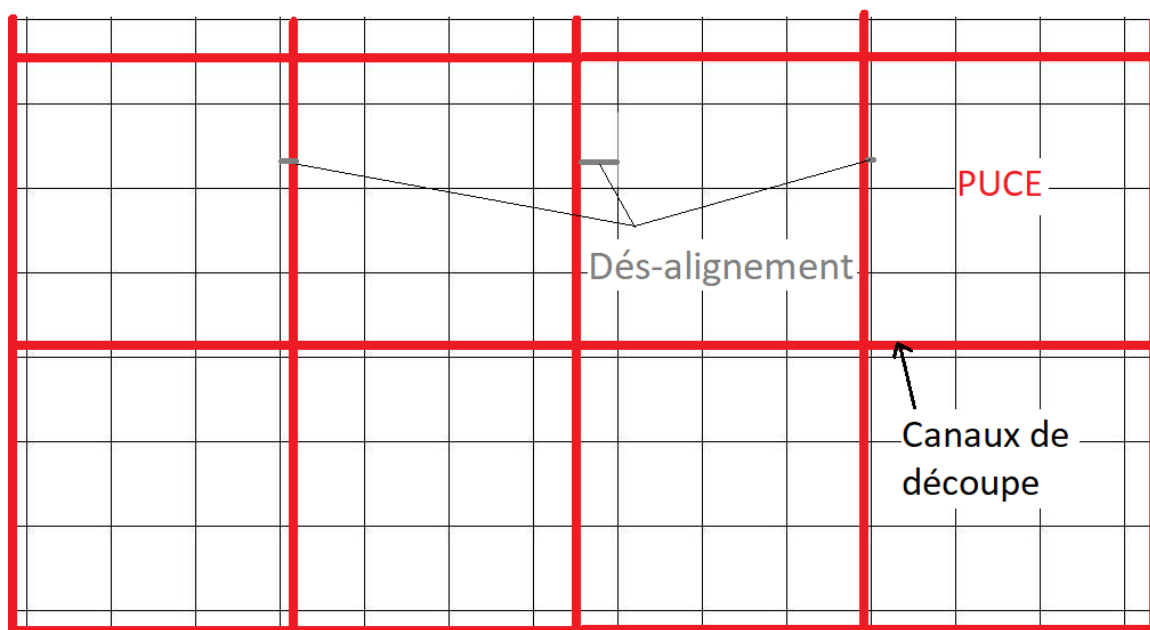


FIGURE 3.7 – Illustration du dés-alignement

Sur cette figure, la grille du fond en noir représente la grille de mesure PWG. Les traits rouges représentent les canaux de découpe, et chaque puce est contenue dans un carré rouge à l'intérieur. La largeur des canaux est non-nulle. On voit que le dés-alignement peut être soit négatif (gauche), soit positif (milieu et droite).

La solution retenue pour gérer ce dés-alignement est d'attribuer le pixel le plus proche du bas gauche de la grille rouge comme étant le pixel de la puce tout en bas à gauche. Si la puce a une dimension qui n'est pas un multiple de $100\mu\text{m}$, on tronque sa valeur comme étant un multiple de $100\mu\text{m}$, et on considère que le reste s'ajoute avec canal de découpe pour trouver la nouvelle puce. On montre le résultat sur la figure 3.8.

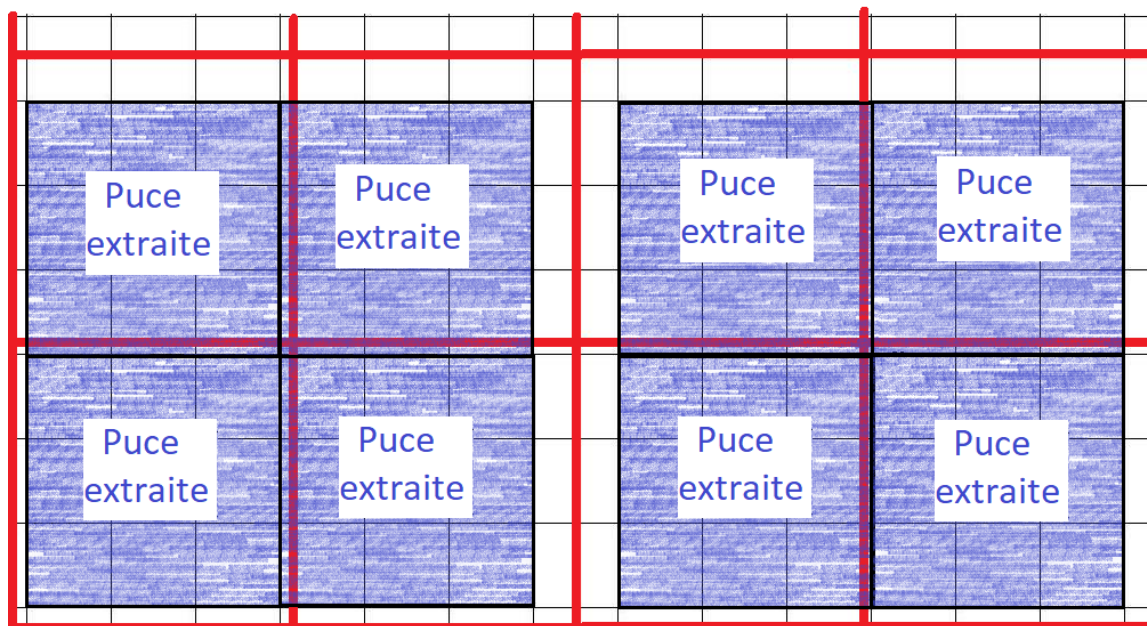


FIGURE 3.8 – Illustration de la solution retenue pour l’extraction de puces

On voit sur la figure 3.8 que la dimension de la puce réelle est entre 3 et 4 pixels sur les deux axes. On se retrouve donc en toute logique avec des puces extraites de dimension 3x3 pixels. Dans la réalité, ces puces ont beaucoup plus de pixels. Elles en comportent $72 \times 57 = 4104$ sur l’exemple 3.9, ainsi la troncature crée une perte d’information d’une seule ligne de pixels par côté au maximum (donc une perte de $72 + 57 = 129$ pixels au maximum), ce qui est une proportion faible du nombre de points (3%). Au-delà de cela, l’avantage de tronquer les puces de cette façon est qu’elles ont toutes les mêmes dimensions, ce qui permet des opérations de moyennes, médianes, écart-types. Nous explorons ce point dans la suite.

Ce choix d’alignement se justifie par sa simplicité de mise en œuvre. D’autres choix possibles auraient été de conserver la grille rouge réelle, d’y dessiner des pixels et d’y attribuer des valeurs pondérées par la surface en contact avec les pixels avec des méthodes d’interpolation (interpolation bicubique [Mitchell and Netravali, 1988] par exemple).

Dans notre cas, le défaut des méthodes basées sur les filtres est une dénaturation de la mesure d’origine. En choisissant la méthode de filtrage, on appliquerait une transformation des pixels donnés par la mesure, et en un sens, cela constitue un jugement biaisé sur la mesure. La façon dont la mesure est effectuée par le PWG dépasse le rôle des sciences de données. On créerait des nouveaux pixels avec des nouvelles valeurs, alors que le PWG a déjà effectué des opérations pour attribuer des valeurs à une grille de pixels. En appliquant la méthode choisie précédemment, on extrait des pixels qui existent déjà dans la mesure initiale, ce qui nous place dans une situation de confiance quant aux valeurs.

La décomposition en puces complètes, ainsi qu’en champs complets, avec la cartographie associée permet d’observer des objets mis en jeu dans la micro-électronique. Comme dit précédemment, la mesure de l’outil PWG ne tient pas compte de l’alignement des champs ou des puces, ni du produit : c’est une mesure d’un wafer sous la forme d’une liste de plus de 7 millions de points.

Transformer 7 millions de points en une liste d’objets compréhensibles tels qu’une liste de puces et de champs est une couche d’intelligibilité. La figure 3.9 montre ces cartographies.

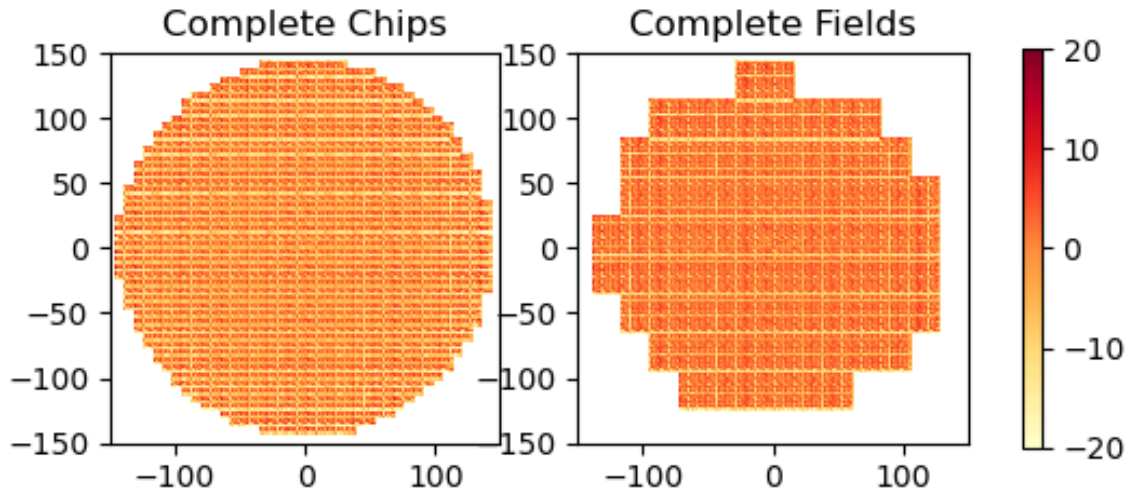


FIGURE 3.9 – Puces complètes (gauche) et champs complets (droite)

Sur le wafer de la figure 3.9 qui est mesuré sur la figure 3.6, on peut compter 1475 puces complètes, et 81 champs complets contenant chacun 15 puces. Générer un ensemble d'objets semblables dans la forme permet d'en faire des analyses statistiques.

3.2.2.3 Objets types et objets atypiques

Dans cette partie de la démarche, on part de la liste d'objets et on aborde différentes questions statistiques.

Quel est l'objet-type pour cette liste d'objets? La moyenne, la médiane, sont des réponses possibles à cette question. Par « moyen », il s'agit d'une moyenne pixel à pixel : on souhaite créer un objet de la même dimension de chaque objet. On remplit donc chaque pixel de coordonnée (i, j) avec la moyenne (resp. médiane) des pixels de coordonnées (i, j) de tous les autres objets. On procède de manière analogue pour la puce médiane : c'est l'objet dont chaque pixel est la médiane des pixels de toutes les puces. La puce moyenne et le champ moyen apparaissent sur la figure 3.10. On les montre également en 3D dans la figure 3.11.

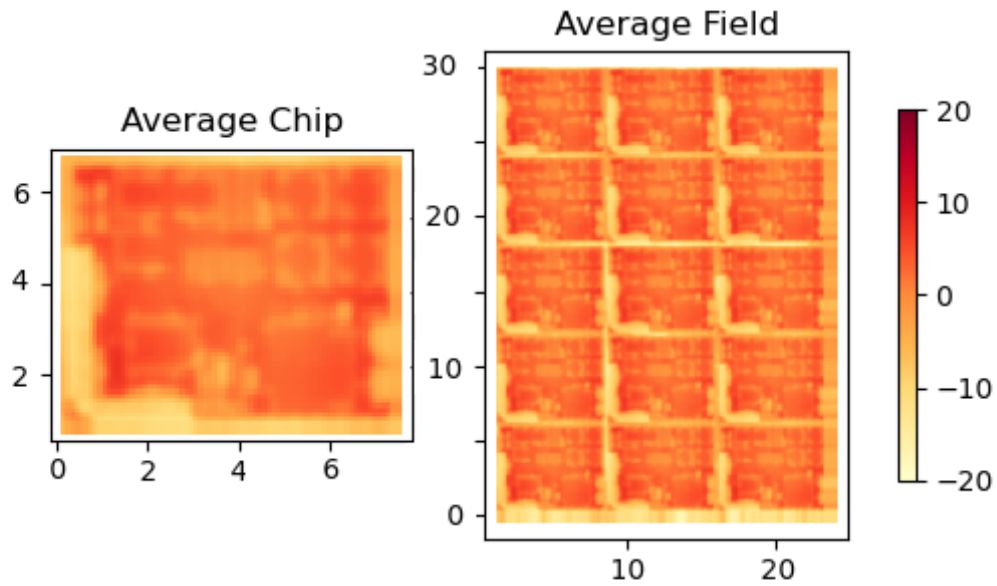


FIGURE 3.10 – Puce moyenne (gauche) et champ moyen (droite)

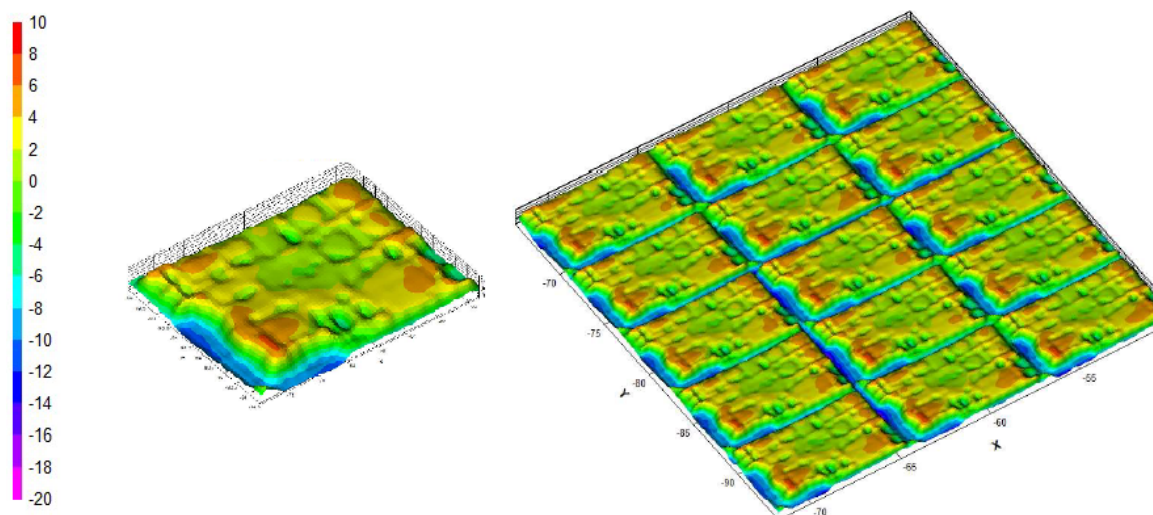


FIGURE 3.11 – Puce moyenne (gauche) et champ moyen (droite)

En observant le champ moyen et la puce moyenne, on voit que le champ moyen est une photo-répétition de la puce moyenne (Figure 3.10 et 3.11).

La conjecture que l'on peut faire est qu'il existe un lien entre les procédés à l'échelle de la

puce et cette nano-topographie. Ce qui est à l'échelle de la puce, ce sont les designs des couches minces empilées. Cela correspond à une information logique, car il n'existe pas de procédés que l'on applique littéralement « puce à puce ». Ce lien sera développé dans le Chapitre 4, où l'on réalisera un modèle de nano-topographie, en prenant comme objectif de prédiction la nano-topographie de la puce moyenne, et en entrée les designs à l'échelle de la puce des couches précédentes.

On souhaite discriminer les puces normales des puces anormales. On considère qu'une puce est normale si elle est proche de la puce moyenne. Enlever la puce moyenne à chacune des puces du wafer permet d'avoir une cartographie résiduelle pour commencer la réponse à cette question. Cette cartographie est montrée dans la figure 3.12, en plus de la même démarche réalisée sur les champs, en image de droite. Ces cartographies sont en réalité des composantes à l'échelle du wafer, car les composantes liées au produit ont été enlevées. On enlève les pixels n'appartenant pas à des puces ou des champs complets de la cartographie.

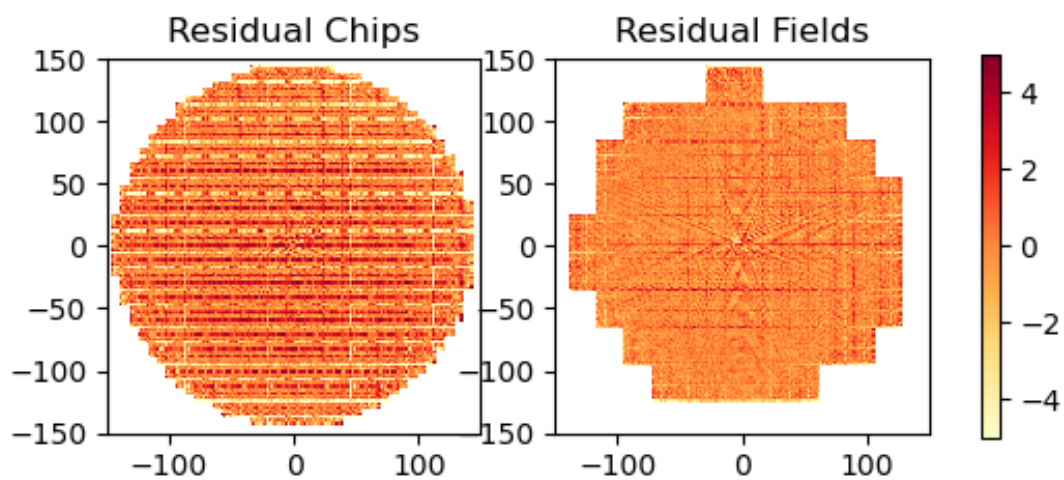
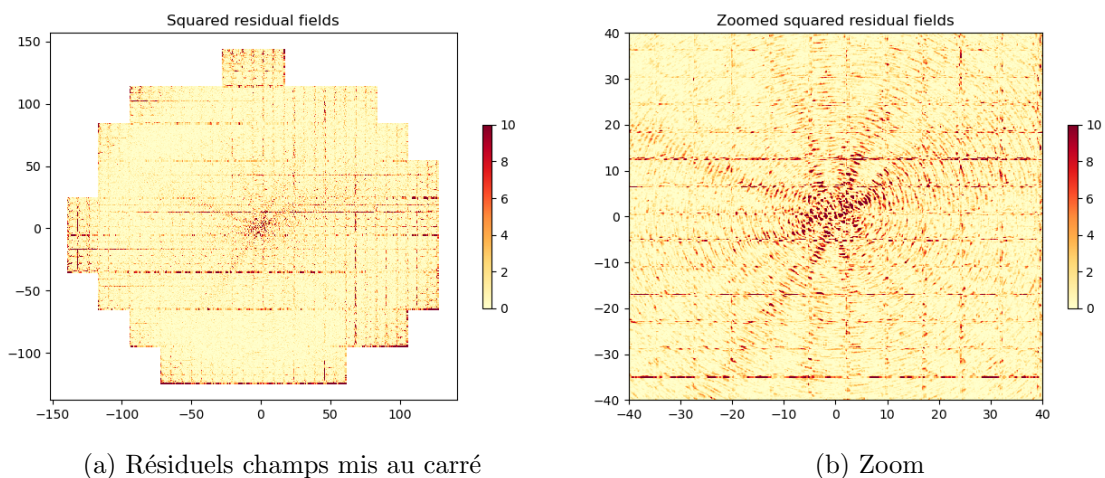


FIGURE 3.12 – Résiduels puces (gauche) et résiduels champs (droite)

Afin de faire ressortir les variations visuellement, on met au carré les valeurs. Ainsi, on augmente la lisibilité des variations à l'échelle du wafer présentes dans la mesure. Cette cartographie est montrée dans la figure 3.13.



(a) Résiduels champs mis au carré

(b) Zoom

FIGURE 3.13 – Composante inter-champ au carré

On peut faire la même manipulation pour les résidus puces et observer une cartographie similaire, avec la même trace, comme montré dans la figure 3.14.

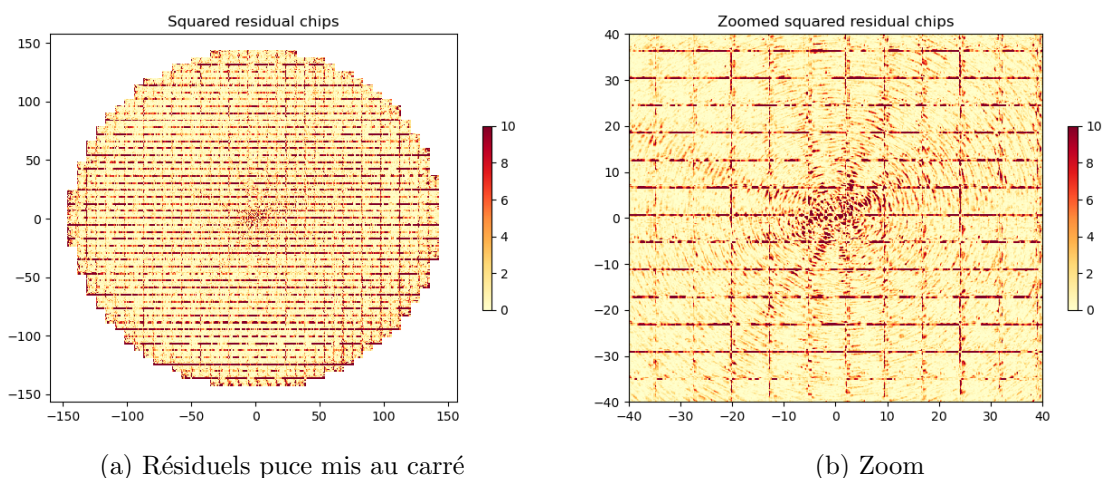


FIGURE 3.14 – Composante inter-puce au carré

Le résiduel et la variation font apparaître la variabilité wafer (Figures 3.12, 3.13 et 3.14). Dans ce cas, on observe une trace circulaire concentrique. Le métrologue aura intérêt à montrer ce phénomène lorsqu’il produit cette mesure, pour le montrer aux ateliers qui la demandent.

Cette trace circulaire est due au procédé CMP, où la plaque est polie par une solution abrasive, ainsi qu’un disque polissant la plaque. Faire ressortir cette trace est une porte d’entrée pour mettre en place des critères d’évaluation pour le procédé CMP.

Comme expliqué dans le chapitre 1, un objectif de la métrologie est de contrôler la bonne réalisation des étapes de fabrication. En analysant ces mesures de topographie, extraire l’information « hors-contexte » permet aux ateliers demandant la mesure de pouvoir mieux comprendre leurs procédés de fabrication, et ainsi de mieux les évaluer.

On veut un critère d’évaluation pour pouvoir contrôler la qualité de réalisation du procédé, soit pour un suivi de production, soit pour des choix de recherche et développement.

Dans la partie suivante, nous montrerons en quoi aller plus loin dans la transformation de données permet d’évaluer des choix de procédés dans un contexte de recherche et développement d’un produit comme de mieux comprendre la mesure pour assurer un suivi de production.

On souhaite mettre une métrique d’écart à la puce moyenne, pour tenter d’identifier les configurations atypiques. Pour chaque puce, on prend la moyenne des pixels de la cartographie résiduelle montrée sur la figure 3.14. Le résultat est montrée dans la figure 3.15a. Ceci étant observé, un métrologue produisant la mesure pourra attirer l’attention sur l’objet moyen à l’échelle de la puce, du champ, ou bien du wafer selon l’atelier concerné. En ce qui concerne les variations wafer, nous abordons le sujet dans le point suivant.

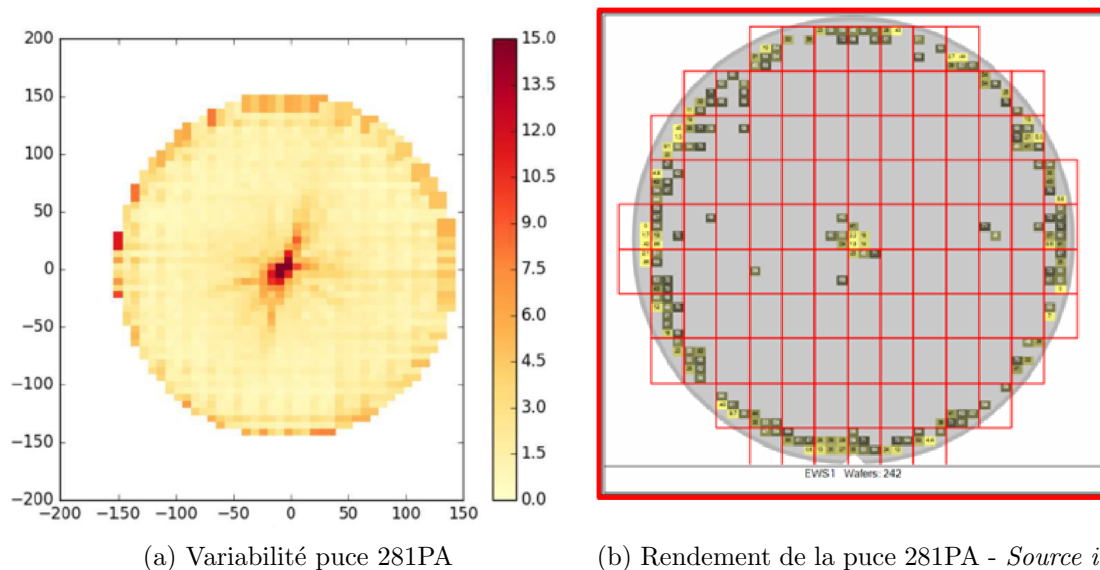


FIGURE 3.15 – Mise en regard entre la variabilité et le rendement du produit 281PA

Ces cartographies permettent, à travers des analyses simples de démantèlement des composantes, de rendre la mesure plus lisible pour l'œil en générant une image pour chaque composante. Bien que cela ne constitue pas un énorme travail technique de développement, il a une valeur ajoutée pour l'entreprise. Premièrement ce type de script reste largement à déployer dans une entreprise comme STmicroelectronics, où les compétences de traitement de données ne sont pas présentes dans tous les ateliers. Quant à la figure 3.15a, on la met en regard d'une cartographie de rendement sur la figure 3.15b.

Le rendement sur le produit 281PA est le plus faible au centre et en bord de plaque, comme le montre la figure 3.15b, où seules les puces ayant un rendement inférieur à 75% sont dans une autre couleur que le gris, avec des rendements d'autant plus faibles que la couleur se rapproche du jaune. Cependant, pouvoir montrer que la topographie est la plus atypique au centre et en bord de plaque permet de supposer une corrélation entre le rendement et la topographie atypique.

C'est là tout l'intérêt de faire de l'enrichissement de données. STmicroelectronics souhaite en permanence contrôler ses rendements, et surtout les augmenter. Pour ce faire, trouver des corrélations entre une mesure et un rendement est un résultat riche. Dans l'optique de faire de l'amélioration continue, extraire le maximum d'information d'une mesure permet d'expliquer les procédés, et le lien entre une mesure et un rendement, comme c'est le cas ici. Ainsi, [Simiz et al., 2015] ont démontré l'impact de la topographie sur le contrôle du focus, ainsi que les conséquences sur le rendement.

Dans la suite de ce chapitre, on veut donner un sens à ces composantes via un enrichissement plus technique de la mesure en tant que telle. En particulier, répondre à des questions de caractérisation de topographie.

3.3 Enrichissement de topographie : Cellules, détection de formes et variabilité

L'enjeu dans cette partie est de décrire la topographie. On veut pouvoir caractériser les objets précédemment extraits (puces ou champs) aussi bien qu'un wafer entier. Pour cela, on décide, comme dans tout ce chapitre, de regarder la mesure en tant que telle. Cette mesure représente une topographie, une différence de hauteur.

La topographie est une grandeur qui apparaît dans d'autres domaines que la micro-

électronique, en particulier pour analyser les terrains, pour diverses applications : bâtiment, géologie, sports de glisse... Ainsi, nous prendrons le parti dans cette partie de faire une bibliographie des techniques d'analyse de terrains dans divers domaines, et de les appliquer à notre description de topographie.

3.3.1 Cellules

La notion de base qui sera utilisée est la notion de « Cellule ». Une cellule est une surface d'une taille (fixe ou non) inférieure à la taille de l'image analysée. Cette cellule sera l'unité d'analyse à laquelle on souhaite attribuer une forme ou une caractéristique. Chaque algorithme ou type d'analyse aura comme finalité une caractéristique.

Dans la suite, nous fixerons la taille de la cellule, et elle sera de taille carrée. En définissant une taille de côté, on fixe l'analyse de fréquence spatiale : si la cellule est petite, l'analyse aura une haute fréquence spatiale. Si la cellule est grande, la fréquence spatiale est basse, et on attribuera des caractéristiques sur de larges surfaces.

La suite de l'algorithme est d'attribuer une caractéristique à une cellule. Il y a diverses finalités, divers objectifs qui peuvent être satisfaits. On peut attribuer soit une métrique à une cellule, soit une classe dans le cas d'une classification. La figure 3.16 résume le concept de cellule ainsi que le concept d'attribution d'une caractéristique à une cellule.

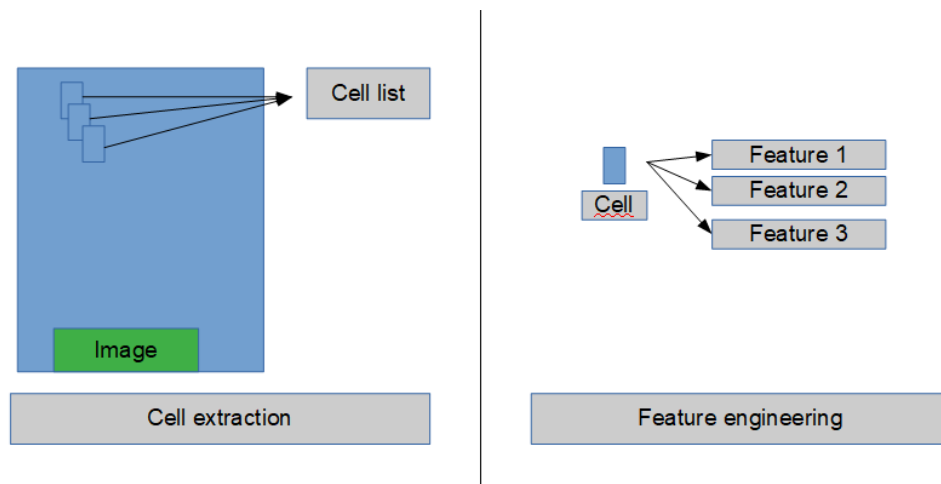


FIGURE 3.16 – Algorithme générique de description de topographie

Sur cette figure, on appelle « feature » une caractéristique, une forme... Dans tous les cas, une valeur attribuée à une cellule. Le feature engineering est utilisé dans les méthodes d'apprentissage, en particulier dans le deep learning, où une couche de sens supplémentaire est donnée afin de faciliter l'apprentissage du réseau de neurones [Reid Turner et al., 1999].

3.3.2 Détection de forme et caractérisation de cellule

La détection de formes est un thème abordé dans la littérature de la description de topographie [Mokarram et al., 2015] [Jasiewicz and Stepinski, 2013]. Il y a différentes façons de faire de la détection de forme. On peut chercher à reconnaître ou non la présence ou non d'une forme spécifique dans une image, en attribuant une valeur pour déterminer la proximité de la cellule à cette forme. On peut faire de la classification, c'est-à-dire étant donné une liste de formes (ou classes), attribuer une forme (ou classe) pour chaque cellule. Un cours sur la reconnaissance de formes peut être trouvé dans [Guesmia, 2017], et explique ces deux différences. On abordera ces deux façons de faire dans les deux sous-parties suivantes.

1. Détection d'une forme particulière ou d'un objet particulier

Dans le cas d'une détection ciblée, on cherche à savoir si, étant donné une forme F , celle-ci se trouve dans une image. La littérature de l'apprentissage automatique, traite de sujets de classification d'images, notamment en deep learning [Kaur and Gill, 2019]. Via des réseaux de neurones, on cherche à savoir si un chat se trouve sur l'image ou non. Ce type de détections n'est pas le sujet dont nous souhaitons parler. Dans notre cas, on souhaite donner du sens aux images, et pouvoir interpréter au maximum la démarche aussi bien que le résultat. Ainsi, utiliser une boîte noire ira à l'encontre de la compréhension.

De plus, dans le cas d'une classification d'images contenant ou non des chats, il peut être dans un coin de l'image comme en plein centre et plus gros, de dos, de face, etc... Il y a là des enjeux dont nous ne souhaitons pas parler comme des changements d'échelles, des rotations, et en ce sens, il ne s'agit pas d'une détection de forme définie en 2D, mais de la détection d'un objet ayant plusieurs définitions en 2D. Afin de mettre en regard le travail qui sera effectué par rapport à celui qui est fait en deep learning, on montre sur la figure 3.17 que des « features » (ou attributs) sont attribuées automatiquement dans les couches intermédiaires.

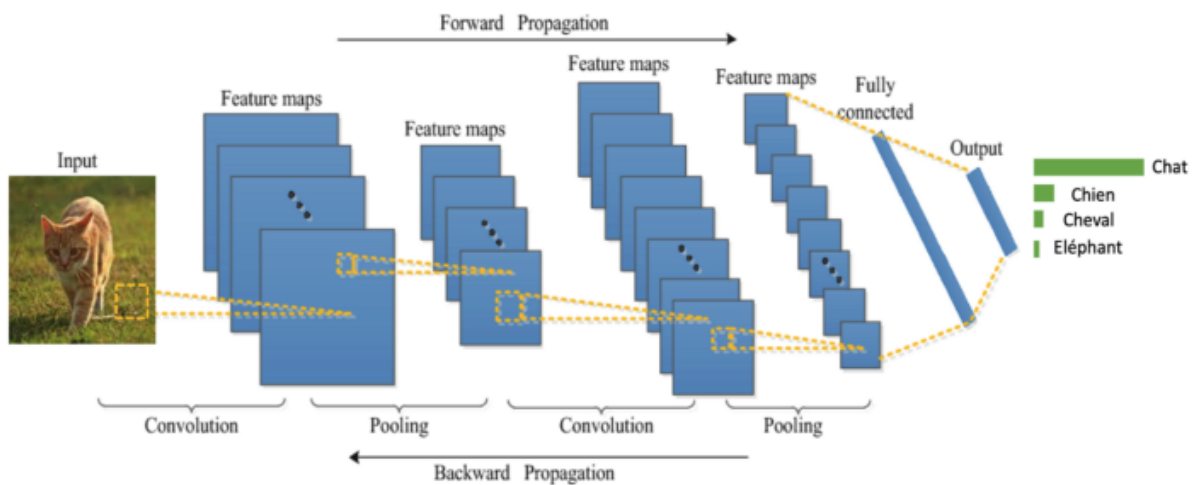


FIGURE 3.17 – Paradigme de deep learning

En observant la figure 3.17, le résultat final « Chat ; Chien ; Cheval ; Éléphant » est un attribut également. Le deep learning est une succession de feature engineering pour arriver à un attribut final. En ce sens, choisir humainement des attributs pour aider le réseau à évoluer vers une interprétation pertinente équivaut à une couche « humaine » d'un réseau.

Ici, on fixe les formes, et on souhaite en attribuer une à chaque cellule. Dans les faits, si l'on veut détecter la forme correspondante, on peut mettre un seuil : si la cellule est suffisamment proche de cette forme, alors on peut considérer que la cellule est de cette forme. Pour ce faire, toute fonction de distance entre une cellule et la forme en question est pertinente. On peut mettre la forme à l'échelle de la cellule si cela est possible, et ainsi toutes les fonctions de distance dans un espace vectoriel de dimension n fonctionneront, n étant le nombre de points de la cellule.

Une autre façon de détecter non pas une forme mais un objet est d'en définir une liste de caractéristiques, et pour chaque cellule de les vérifier. Toute cellule vérifiant la liste de caractéristiques sera considérée comme étant cet objet. Dans le cas d'une forme, comme nous le verrons par la suite, on peut vérifier par exemple des valeurs de dérivées premières et secondes afin de savoir comment varie la surface autour d'un point, et s'en servir pour en déterminer la forme. Cet algorithme est expliqué dans [Watson et al., 1985].

2. Classification de cellules parmi un ensemble de formes

L'objectif ici est, étant donné une liste de formes, d'attribuer une unique forme à chaque cellule. On peut mettre une case « forme indéfinie » pour les cellules ne correspondant à aucun cas afin de rendre l'algorithme complet (chaque cellule a une forme attribuée).

Dans le même esprit que dans la détection de forme, on peut estimer la proximité d'une cellule avec chacune des formes, et considérer que la cellule est de la forme qui lui est la plus proche. Si aucune forme n'est suffisamment proche, en définissant un seuil par exemple, on peut considérer que la cellule est de forme indéfinie.

Des articles utilisant la géométrie différentielle utilisent les signes des dérivées premières et secondes pour déterminer la forme autour d'un pixel d'une image [Wang et al., 2006] [Watson et al., 1985].

3. Caractérisation de cellules

Un autre type de transformation est d'attribuer une valeur numérique à une cellule, comme par exemple le gradient ou le TPI (Topographic Position Index) [Mokarram et al., 2015], comme nous le montrons dans la suite.

3.3.3 Gestion de mémoire

Dans l'implémentation d'un algorithme de caractérisation de cellules, des éléments d'attention concernant l'implémentation sont relevés concernant la gestion de la mémoire. Dans notre cas, toutes les opérations sont parallélisables. Chaque cellule peut être extraite et traitée séparément, et également chaque attribut peut être calculé séparément sur une même cellule. Au-delà de vouloir optimiser le temps de calcul en parallélisant au maximum, on veut éviter de stocker un trop grand nombre de cellules dans la mémoire vive pour ne pas « swapper » et considérablement augmenter le temps de calcul.

Lorsqu'on travaille sur une petite surface, petite voulant dire : pouvant entièrement se contenir dans la mémoire vive, il n'y a aucun problème. Cependant, lors de l'extraction de cellules sur une surface plus conséquente comme un wafer entier de 7 millions de points, il faut, d'une part, gérer le stockage de chaque cellule sans dupliquer l'image pour ne pas surcharger la mémoire temporaire, et d'autre part ne pas stocker les variables de calcul en nettoyant régulièrement la mémoire utilisée.

On présente ici les cas extrêmes résultant des erreurs que l'on peut commettre dans l'implémentation, afin de clarifier le choix de la solution choisie pour d'une part optimiser la parallélisation et d'autre part limiter les sollicitations des disque durs liées aux transferts de données.

Cas extrême 1 : Traiter les cellules une par une, puis nettoyer la cellule une fois sa caractéristique ou sa forme attribuée. Le défaut de cette méthode est le nombre de fois que l'on sollicitera le disque dur. Le transfert du disque vers une mémoire vive est de l'ordre de quelques millisecondes (quelques dizaines de millisecondes dans le cas d'une architecture distante). Ainsi, il faut solliciter le disque le moins souvent possible. Dans ce cas extrême, on sollicitera le disque dur autant de fois qu'il existe de cellules : autrement dit, 7 millions d'appels disque, soit plus de 2h passées à solliciter la mémoire pour un wafer.

Cas extrême 2 : Charger le wafer en entier dans la mémoire. Ce n'est pas possible, car on cherchera à attribuer plusieurs caractéristiques par cellule avec des algorithmes différents. Aussi, il faudra garder en mémoire ces écritures, et le programme devra alors « swapper », ce qui aura pour conséquence de grandement ralentir le temps de calcul. Dans ce cas, encore une fois, le disque dur (distant ou non) sera sollicité un grand nombre de fois, ce qui aboutira à un temps de calcul extrêmement long.

Solution : On maîtrisera donc l'usage de la mémoire, en faisant en sorte de traiter un bloc de cellules pouvant se stocker dans la mémoire. Plus le nombre de cellules par portion est grand,

moins le nombre d'appels disque pour stocker les points du wafer est élevé. Moins le nombre de cellule par bloc est élevé, moins on s'approche du fait que la mémoire surcharge et se mette à « swapper ». Il faut donc écrire par blocs de taille la plus grande possible selon l'architecture. Le choix qui a été fait ici a été de faire un bloc par « puce ». Dans le cas de la mesure précédente, il y aura donc 1475 appels disque, contenant 4104 cellules chacun. Il n'y a pas de swap, et le nombre d'appels est limité.

La problématique d'implémentation étant réglée pour éviter tout problème lié à la mémoire et profiter d'une parallélisation efficace, on peut commencer l'implémentation d'algorithmes traitant une cellule en attribuant une caractéristique. Ces problèmes ont été rencontrés dans d'autres cas d'analyse de topographie. Une gestion attentive de la mémoire est mentionnée dans [Jasiewicz and Stepinski, 2013], où l'on attribue des valeurs à des cellules comme dans notre cas.

3.4 Cas d'étude : Choix de slurry pour l'atelier CMP

L'enrichissement de données se rentabilise au travers des prises de décisions de la part des ingénieurs. Dans le cas d'étude de cette sous-partie, il s'agit de faire un choix industriel pour un procédé dans la phase de Recherche et Développement d'un produit. Il s'agit du choix de slurry pour l'atelier CMP. Nous avons expliqué ce qu'était l'atelier CMP dans le chapitre 2, ainsi que ce qu'est une slurry. Les slurry ont des propriétés différentes, et des choix techniques de compositions chimiques ont des impacts sur les procédés. Dans ce cas, on en essaie 3, on regarde le résultat, et l'idée est de choisir la plus efficace selon des critères de planéité. Mais comment évaluer la planéité d'une plaque ? Le choix est de comparer les champs moyens, et de choisir celui qui a la meilleure configuration. On montre ces champs moyens dans la figure 3.18.

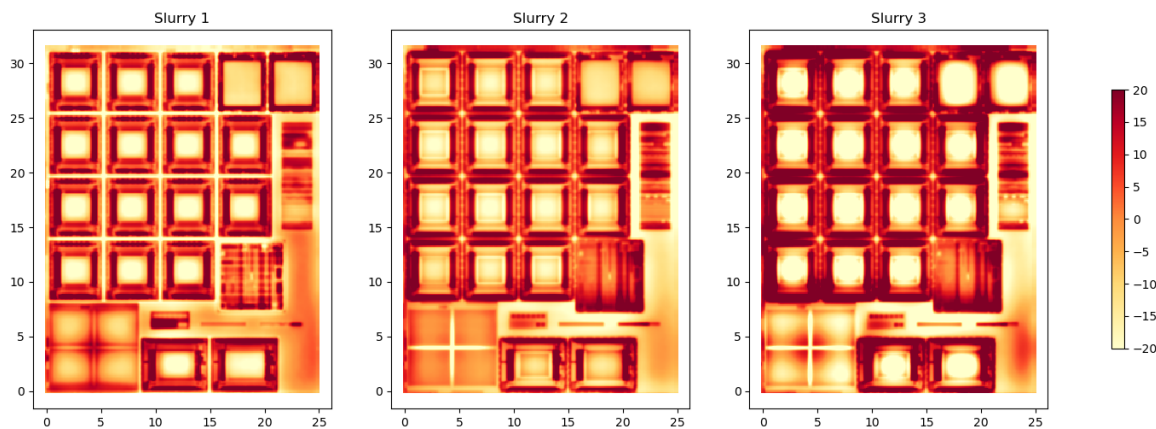


FIGURE 3.18 – Topographie des slurry mesurées

L'atelier CMP a donc 3 choix, qu'ils évaluent en les essayant et en mesurant les différentes plaques ayant subi les différents choix. Ces mesures sont montrées en Figure 3.18. A l'œil nu, ces images se ressemblent beaucoup. On a cependant besoin de pouvoir prendre une décision, qui influencera le rendement de production lors de l'industrialisation du procédé de polissage.

Le parti qui a été pris ici, et c'est celui de ce chapitre 3, est que ce n'est pas aux sciences de données de décider. On respecte l'expertise métier des ingénieurs en physique-chimie travaillant à l'atelier CMP, en ne faisant qu'une chose : montrer le maximum d'information pertinente contenue dans les mesures, et offrir des critères de comparaison objectifs pour les ingénieurs. Indépendamment de la cause technique de savoir « pour quelles raisons » un choix est meilleur qu'un autre, on cherche alors des potentiels critères de description de la mesure de topographie.

La détection de formes, les différentes transformations d'images telles que les cartographies de gradient, sont des indicateurs. Les variations de topographies sont intrinsèques à la mesure elle-même. On fait ressortir la variation en montrant des cartographies de gradient, comme dans la Figure 3.19.

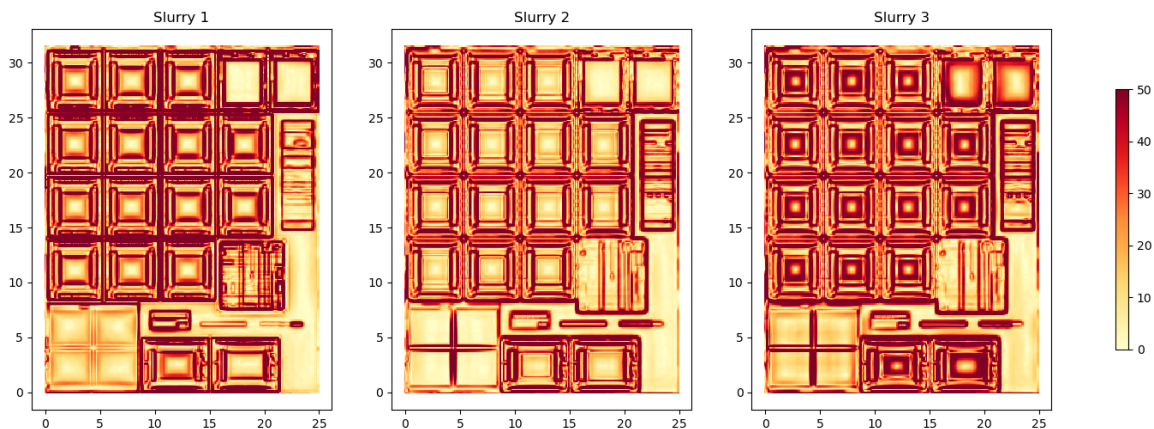


FIGURE 3.19 – Gradient des slurry mesurées

Cette cartographie offre un regard différent, mais ne permet pas vraiment de mieux décider que la mesure de topographie. Si l'objectif est de minimiser les variations, alors il faudrait en extraire des critères comparables. Pour le gradient, on propose la distribution de gradient, comme montré sur la Figure 3.20.

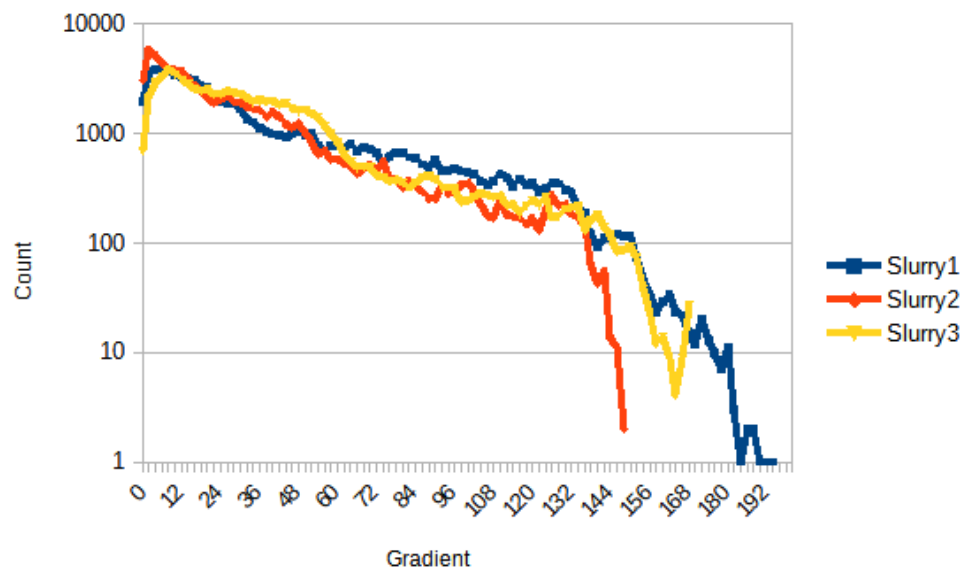


FIGURE 3.20 – Distribution de gradient des slurries

Cette distribution offre une possibilité de comparaison sur des domaines différents de gradients. On peut accorder de l'importance à la présence de gradients de magnitude élevée, si ceux-ci sont problématiques pour le procédé, tout comme on peut plutôt s'intéresser au gradient moyen. Dans la figure 3.21, on montre des critères comparables sur ces distributions de magnitudes de gradient.

Slurry 1		Slurry 2		Slurry 3	
Premier quartile	11,52	Premier quartile	8,72	Premier quartile	14,21
Médiane	25,84	Médiane	21,21	Médiane	30,29
Troisième quartile	60,45	Troisième quartile	43,22	Troisième quartile	51,02
Moyenne	40,01	Moyenne	31,34	Moyenne	38,34
Peak-to-Valley	194,07	Peak-to-Valley	149,73	Peak-to-Valley	169,34

FIGURE 3.21 – Statistiques des magnitudes de gradient des slurries

On souhaite que les magnitudes des gradients soient basses. Les magnitudes sont toujours positives. Dans ce tableau et dans le contexte d’une variable positive, chacun de ces indicateurs implique certaines propriétés allant dans le sens de gradients bas. La moyenne, la médiane, le premier et le troisième quartiles, sont autant d’indicateurs nous aidant à nous représenter la distribution de l’ensemble de ces points. Dans cet exemple, on voit que sur **tous** les indicateurs, la Slurry 2 est plus basse que les deux autres. Il est donc raisonnable de conclure que celle-ci est une meilleure slurry que les autres lorsqu’on regarde ces valeurs.

Dans le cas d’étude présent, sur le graphique de distribution des gradients (figure 3.20), on voit que pour les gradients faibles, la courbe orange (Slurry 2) est plus haute pour les gradients plus faible, plus basse pour les gradients hauts, et plutôt en-dessous pour les gradients moyens. On en déduit qu’elle a des gradients plus faibles que les autres slurries. Elle est donc le meilleur choix selon ce critère. L’atelier CMP avait choisi cette slurry comme étant la meilleure au final. On peut donc proposer ces indicateurs comme description pertinente et comparable.

Dans la pratique, notamment pour la photolithographie, tous les points n’ont pas la même sensibilité au procédé. Certains points peuvent avoir des variations hautes de topographie sans que cela impacte le rendement car la topographie n’est pas critique, mais d’autres ne pourront pas dépasser une certaine variation de topographie sans impacter le rendement de la puce. C’est en cela que la **cartographie** de gradient prend son sens métier. L’enjeu est toujours de fournir une information qui puisse être interprétée par un expert métier pour prendre une décision.

Une autre description possible est la détection de formes. Dans l’esprit de détecter des configurations problématiques pour un procédé, un algorithme de détection de formes prend son sens, en cherchant à détecter des formes problématiques dans l’état de la puce. On se réfère à un article de géométrie différentielle pour cela [Watson et al., 1985]. On peut voir la cartographie des formes dans la figure 3.24.

L’algorithme procède en plusieurs étapes pour la description de formes :

1. Extraction de cellule.
2. Interpolation polynomiale d’ordre 2.
3. Sur cette interpolation, calculer la matrice Hessienne (« H » sur la figure 3.22).
4. Calculer les valeurs propres $L1, L2$ et les vecteurs propres $u1, u2$ de la matrice Hessienne :
5. Calculer les dérivées premières suivant les directions respectives des vecteurs propres $V1, V2$, ainsi que la magnitude du gradient M .
6. Selon les signes de ces 5 valeurs, déterminer la forme grâce à la table 3.23.

Les 5 premières étapes sont détaillées sur la figure 3.22 pour clarification. La magnitude du gradient est calculée grâce à $V1$ et $V2$ avec pour valeur $M = \sqrt{V1^2 + V2^2}$.

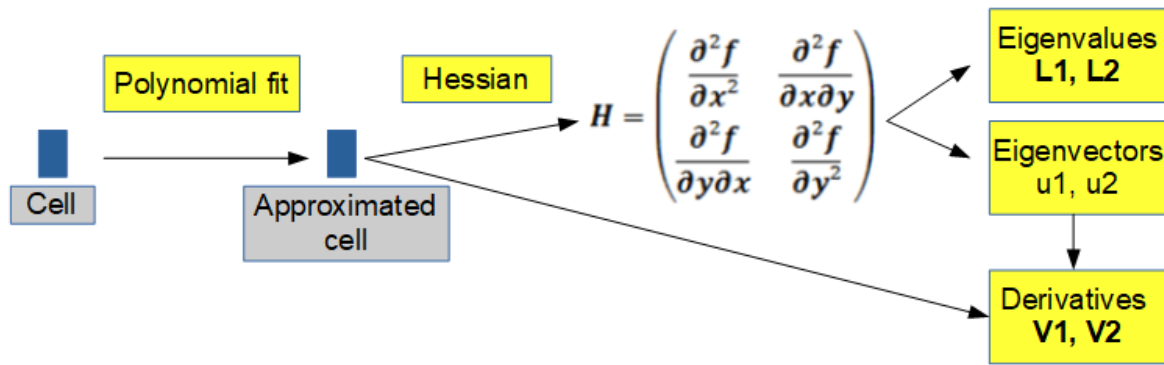


FIGURE 3.22 – Algorithme d'obtention des valeurs

M	L1	L2	V1	V2	Label
0	-	-	0	0	Peak
0	-	0	0	0	Ridge
0	-	+	0	0	Saddle
0	0	0	0	0	Flat
0	+	-	0	0	Saddle
0	+	0	0	0	Ravine
0	+	+	0	0	Pit
+	-	-	-, +	-, +	Hillside
+	-	*	0	*	Ridge
+	*	-	*	0	Ridge
+	-	0	-, +	*	Hillside
+	-	+	-, +	-, +	Hillside
+	0	0	*	*	Hillside
+	+	-	-, +	-, +	Hillside
+	+	0	-, +	*	Hillside
+	+	*	0	*	Ravine
+	*	+	*	0	Ravine
+	+	+	-, +	-, +	Hillside
+	*	*	0	0	Cannot Occur

FIGURE 3.23 – Tableau de répartition des formes en fonction des valeurs de dérivées - *Source : [Watson et al., 1985]*

Sur ce tableau, M est la magnitude du gradient, $L1$ et $L2$ sont les valeurs propres de la matrice Hessienne, $V1$ et $V2$ sont les dérivées secondes dans la direction des vecteurs propres de la matrice Hessienne.

Un point d'attention concernant le « signe » des dérivées : hors-contexte, le signe d'une dérivée semble évident à déterminer. Cependant, savoir si une dérivée est considérée comme positive, nulle, ou négative dépend du contexte. Techniquement, on doit trouver un intervalle qui permet de dire que la dérivée en question peut être considérée comme nulle. Le choix de cet intervalle est basé sur des critères pouvant être liés aux distributions des valeurs des données, ou bien liés à des facteurs métiers.

Dans le cas d'étude courant, sachant que la résolution latérale est en μm et la résolution de la hauteur est en nm , l'unité d'une dérivée première est en $\text{nm}/\mu\text{m}$. Une fois qu'on a déterminé un intervalle de nullité pour une dérivée première (centré sur zéro dans notre cas), que l'on note $[-u, u]$, on traduit ce choix a priori mathématique/informatique (car on sait qu'on ne tombe jamais sur zéro exactement) par la phrase suivante : « Toute variation inférieure à $u \text{ nm}/\mu\text{m}$ en

valeur absolue est considérée comme nulle ».

Ce choix technique a un sens métier, et dans la mesure où l'on souhaite utiliser le signe de la dérivée pour détecter des formes pour un atelier, il est pertinent d'avoir la démarche de se demander : « Étant donné le contexte métier auquel cette mesure est destinée, quel est l'intervalle dans lequel la variation de hauteur peut être considérée comme nulle ? ». En répondant à cette question, on peut détecter les formes avec pertinence pour l'atelier.

Dans les faits, on souhaitait initialement valoriser une mesure, et en voulant y donner du sens, on trouve un besoin d'écho chez les personnes du métier. On veut appliquer un algorithme, a priori sensé, et générer des questionnements pour appliquer correctement l'algorithme. Dans l'idéal, les personnes du métier ont les réponses. Au-delà de l'application de l'algorithme, la question précédente n'a jamais été posée au métier. Cette réflexion est de valeur pour mieux comprendre les procédés et savoir ce qui importe dans une mesure. Dans notre exemple, considérer une dérivée comme nulle dépend du contexte dans lequel elle est utilisée.

C'est là tout l'enjeu de déployer des compétences en mathématiques et informatique dans un contexte industriel. Décrire une mesure pousse à trouver des objets mathématiques donc provenant d'un domaine différent, et les interactions entre les métiers montrent que ces objets sont en fait similaires. La nullité d'une dérivée est ici l'absence d'impact sur le procédé métier.

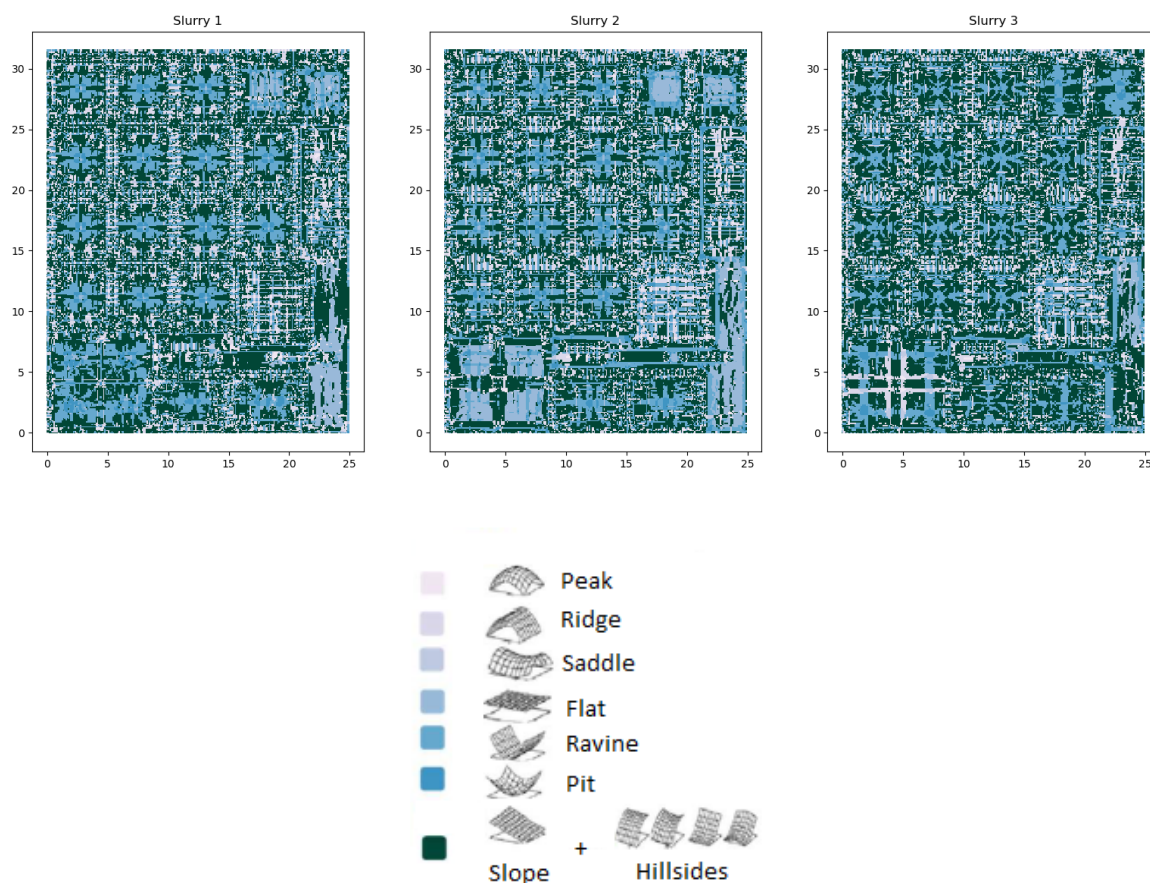


FIGURE 3.24 – Cartographie des formes des slurries

Cette cartographie des formes (3.24) est difficile à lire. On peut cependant montrer un histogramme de répartition des formes qui lui est plus lisible (figure 3.25).

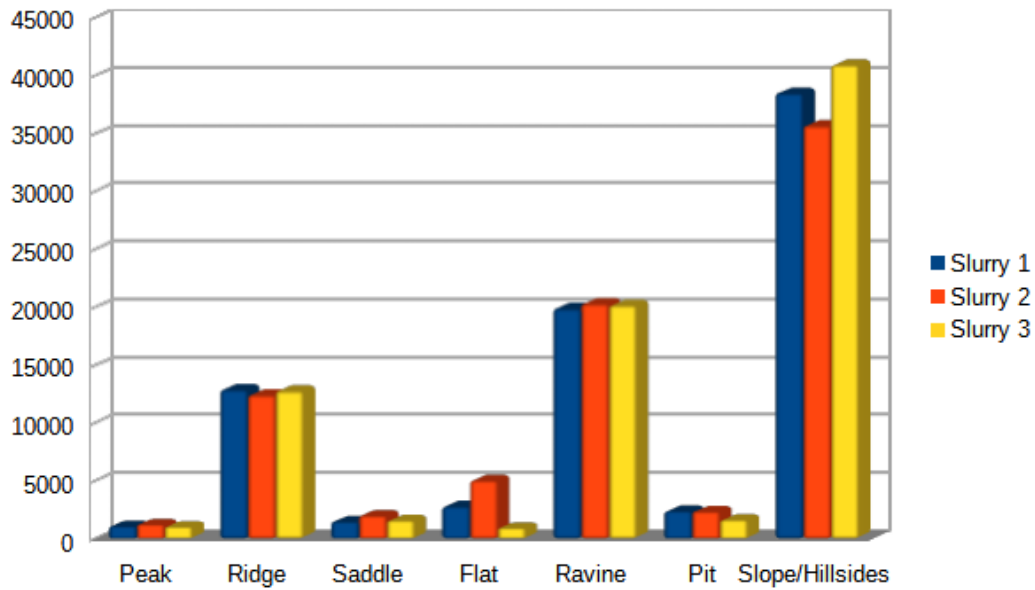


FIGURE 3.25 – Histogramme des formes des slurries en pourcentage

Bien qu'il soit plus lisible, l'histogramme ne permet pas une comparaison complète pour prendre une décision concernant le meilleur des trois procédés. En effet, on est obligés d'accorder plus ou moins d'importance à la présence de certaines formes, mais cela demande un travail de fond qui peut manquer de sens. Par exemple, comment décider s'il est pire d'avoir des pics ou des vallées pour un procédé, en sachant qu'une vallée peut être très abrupte et un pic peu profond, tout comme un pic peut être très profond et une vallée plutôt douce. Si l'information peut être la même, la résultante sur les procédés de fabrication peut ne pas être la même.

Pour résoudre cette difficulté liée à la multiplicité des formes, on peut décider de changer l'algorithme précédemment évoqué, et faire une interpolation polynomiale d'ordre 1, à savoir un plan pour chaque cellule. Dans ce cas, les formes possibles sont uniquement les pentes et les plats. Nous sommes donc dans une situation où il y a un pourcentage de plat et un pourcentage de pentes. Cela permet d'avoir un critère de décision peu biaisé pour les ingénieurs. Bien sûr, il peut toujours subsister des situations où un pic profond est approximé par un plan, mais celui-ci sera vraisemblablement entouré de pentes. On montre le résultat de cette cartographie dans la figure 3.26.

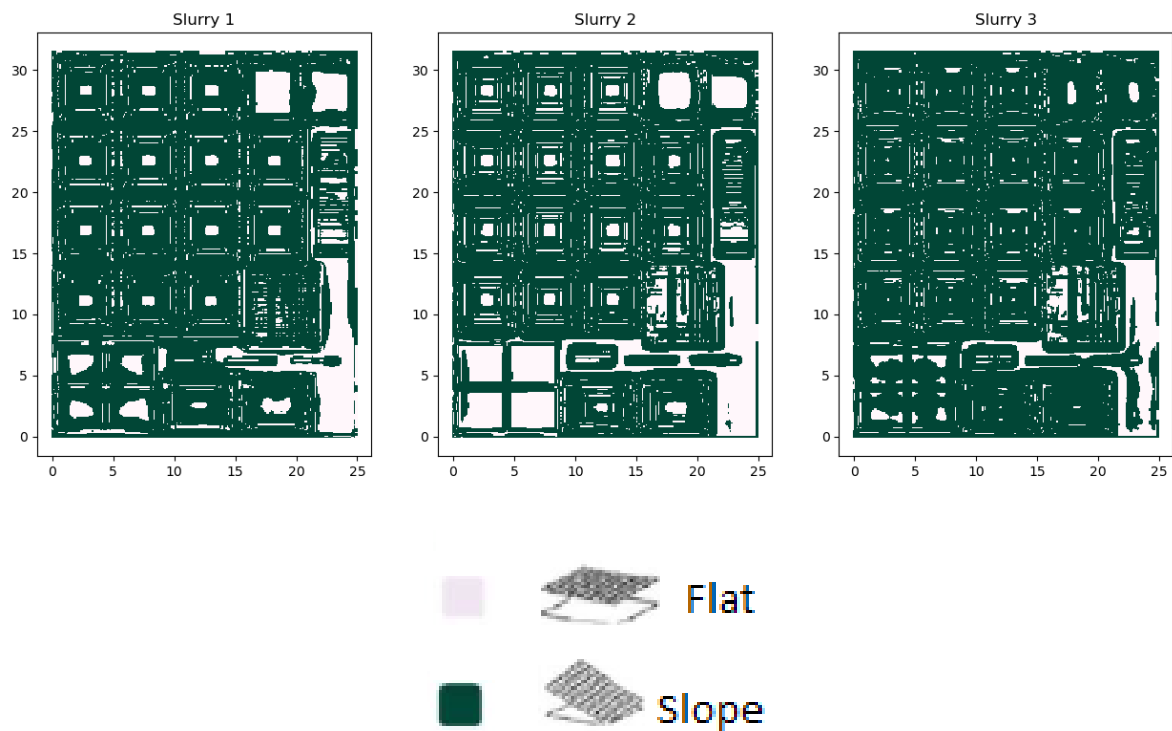


FIGURE 3.26 – Cartographie des formes des slurries à l'ordre 1

Cette cartographie est plus lisible que la précédente, et contient bien moins d'information à interpréter. On a une information binaire, qui peut être synthétisée par l'histogramme montré en figure 3.27.

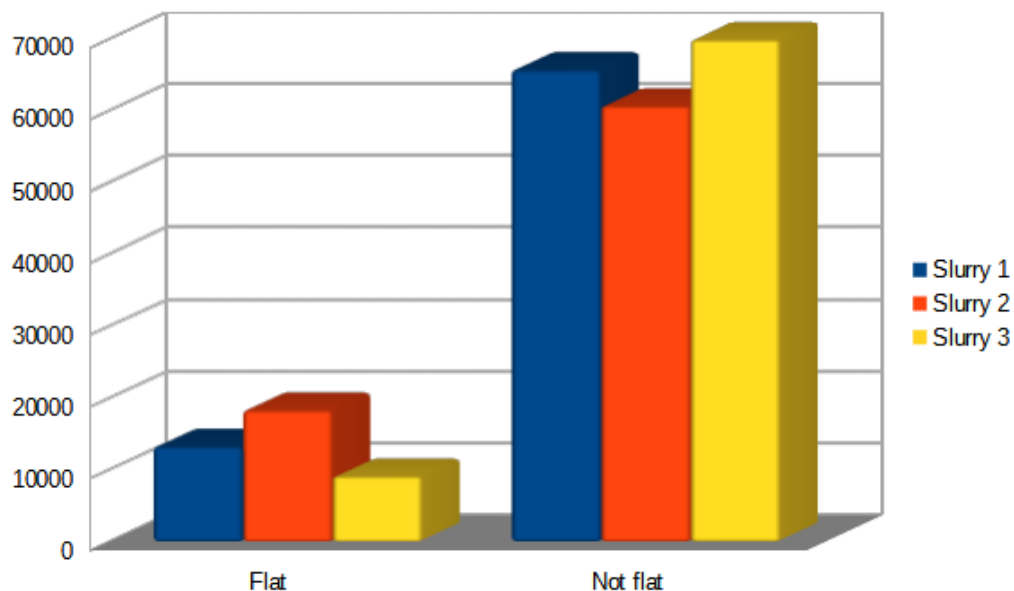


FIGURE 3.27 – Histogramme des formes d'ordre 1 des slurries en pourcentage

Encore une fois, la Slurry 2 est la meilleure selon cet indicateur. En multipliant les indicateurs qui confirment un choix, on est de plus en plus confiants pour confirmer une décision.

Une autre description de topographie est le TPI (Topographic Position Index), utilisé pour

décrire des paysages naturels [Mokarram et al., 2015]. Le TPI est un calcul effectué sur chaque pixel et son entourage. Il s'agit de déterminer l'écart entre le pixel concerné et son entourage **proche** comparativement à son entourage **éloigné**. La valeur du TPI est la différence entre la valeur moyenne de l'entourage proche et celle de l'entourage éloigné. Les paramètres prennent donc en entrée deux zones d'entourages définies. Dans la figure 3.28, on montre le résultat du calcul d'un TPI sur des cellules de taille 3x3. L'entourage proche est le singleton du pixel central, et l'entourage éloigné est la totalité de la cellule.

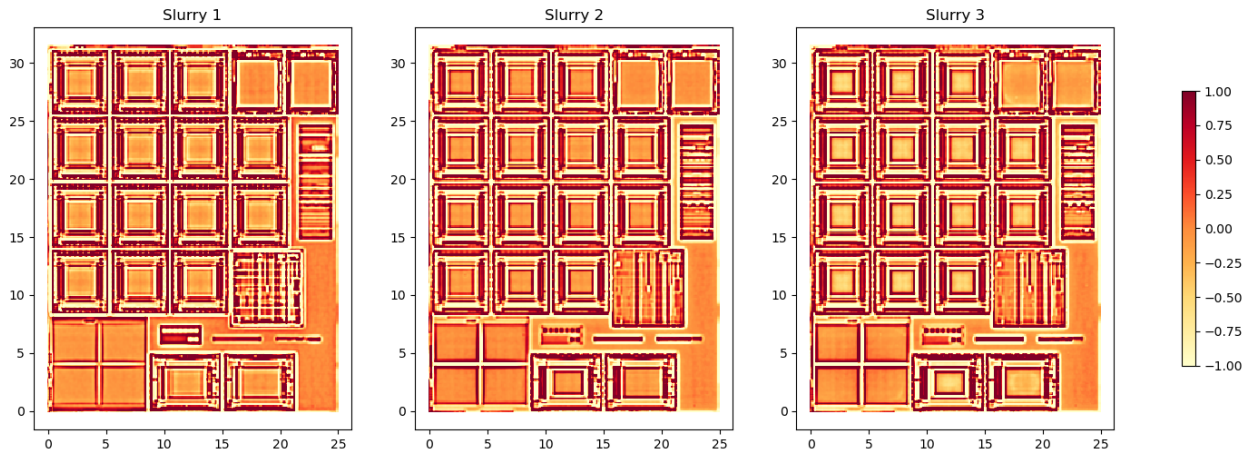


FIGURE 3.28 – Cartographie des TPI des slurries

Cette cartographie apporte encore un nouveau regard, mais cela ne va pas dans le sens de la création de métrique pour évaluer un procédé. Tout comme le gradient, on donne une courbe de comparaison des slurries sur 3.29 ainsi que des indicateurs sur la figure 3.30. Pour le TPI, c'est un TPI proche de zéro qui est la conséquence d'une zone plane.

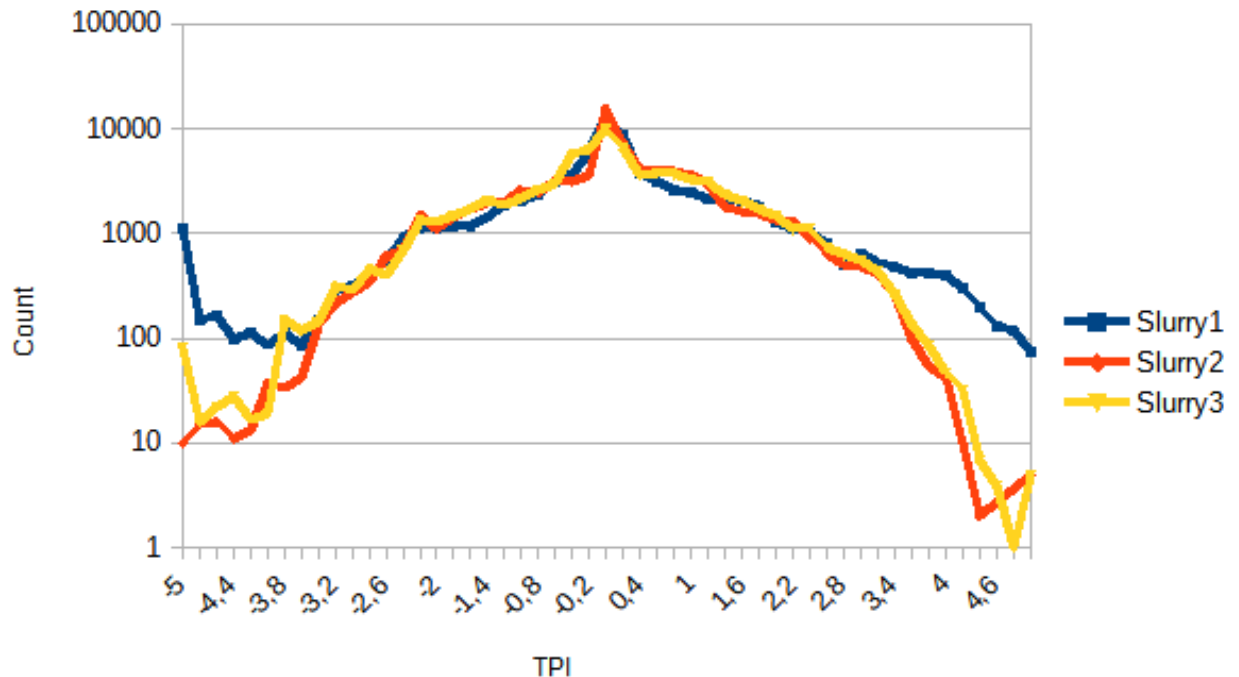


FIGURE 3.29 – Distribution de TPI des slurries

Sur cette courbe, on discrimine difficilement les slurries entre elles. On peut voir que la slurry 1 n'est pas la meilleure, car elle dépasse visiblement sur les extrémités de la courbe de répartition. On ne peut cependant pas distinguer la meilleure des deux slurries entre les 2 et 3. Cette courbe donne une information, qui permet de discriminer la slurry 1, mais pas les autres. Elle n'est pas entièrement satisfaisante, mais va dans le sens de la recherche d'indicateurs pour prendre des décisions.

Afin d'avoir des indicateurs plus comparables, comme pour le gradient, on donne le tableau suivant sur la figure 3.30. L'avantage de ce tableau, est la présence d'un nombre raisonnable de valeurs comparables.

Slurry 1		Slurry 2		Slurry 3	
Premier quartile	-0,60	Premier quartile	-0,63	Premier quartile	-0,63
Médiane	-0,04	Médiane	-0,02	Médiane	-0,05
Troisième quartile	0,73	Troisième quartile	0,68	Troisième quartile	0,76
Moyenne	0,00	Moyenne	-0,00	Moyenne	-0,00
Peak-to-Valley	16,00	Peak-to-Valley	9,45	Peak-to-Valley	10,62

FIGURE 3.30 – Statistiques de TPI des slurries

Ces indicateurs ne montrent pas de différences significatives les uns avec les autres. Le Peak-to-Valley éliminerait encore une fois la slurry 1 en tant que meilleur choix, mais pour départager les deux autres, la différence est trop minime pour trancher. L'indicateur du TPI ne semble pas approprié dans ce cas-ci pour caractériser la dominance d'une seule et unique slurry.

3.5 Conclusion

La volonté de ce chapitre est de s'approprier une mesure afin de l'observer dans ses différentes composantes, mais aussi pour trouver des indicateurs qui caractérisent la mesure pour prendre des décisions. Le cas d'étude sur les slurry a montré que le gradient ainsi que ses statistiques permettaient de se conformer à la décision prise par l'atelier CMP. L'histogramme de formes Flat/Slope est également un critère qui va dans le sens du choix effectué par l'atelier CMP.

La figure 3.31 a été présentée lors de la conférence SPIE 2019, et constitue une proposition d'indicateurs résultant de transformations d'images résumant partiellement le travail effectué sur la transformation de données. Il ne s'agit pas du même produit précédent mais d'un autre produit mis en production : le produit 281PA.

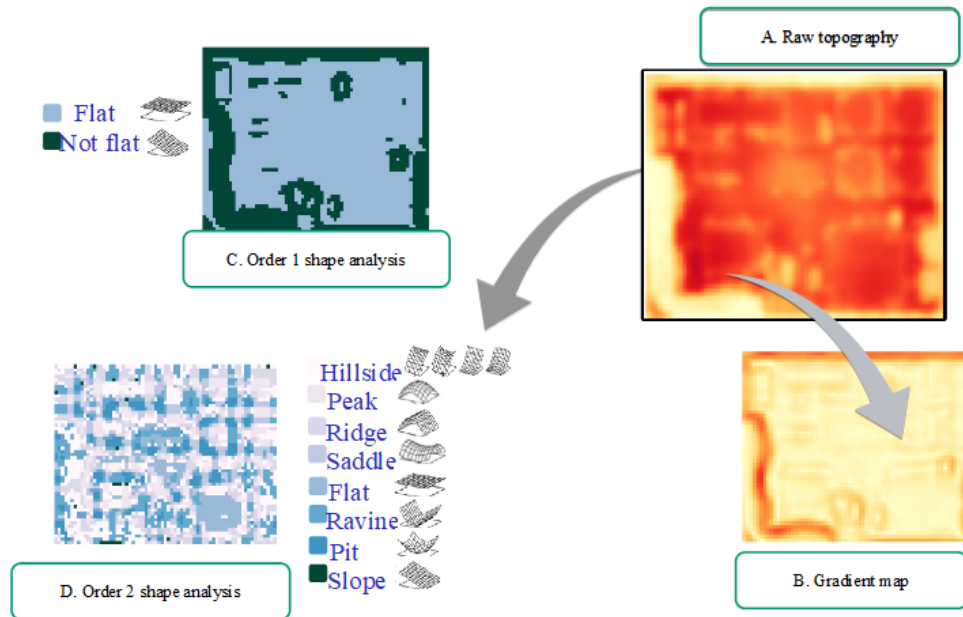


FIGURE 3.31 – Transformations d’images - *Source : [Kessar et al., 2019]*

Choisir des indicateurs, c’est aussi exclure tous les autres possibles. Le TPI n’a pas été ici pertinent pour le choix des slurry, mais il serait prématuré d’en éliminer l’utilité. Il en est de même pour la détection de formes. La recherche, c’est aussi explorer, présenter, et donner de nouvelles idées pour de nouvelles applications. En l’occurrence, le TPI, le gradient, et la détection de formes ont permis une re-considération de l’image initiale pour faire du « feature engineering », comme nous le présentons dans le Chapitre 5.

Chapitre 4

Modèle nano-topographie

Dans ce chapitre nous présentons le travail principal effectué lors de la thèse. Il s’agit de faire un modèle de topographie dont les enjeux sont explicités au cours du chapitre. Dans l’ordre, nous présentons :

1. La motivation du modèle, à la fois généralement dans une industrie, puis plus spécifiquement dans notre cas d’étude en parlant de l’importance de la topographie.
2. Les données disponibles et utilisées pour ce projet de modélisation. En particulier les densités et la nano-topographie des produits de notre cas d’étude.
3. La construction d’un modèle linéaire est justifiée, les choix de transformation des entrées sont expliqués, et les enjeux de simplification et robustesse sont mis en évidence.
4. Des protocoles de simplifications de modèles sont présentés ainsi que leurs résultats, qui amènent le besoin de quantifier la robustesse d’un modèle.
5. Une proposition de métrique de robustesse est présentée.

4.1 Présentation du problème

Dans l’industrie, il est essentiel de pouvoir anticiper les problèmes le plus en amont possible. Parfois, la conception-même d’un produit peut induire de mauvaises dispositions à être fabriqué dans des rendements optimaux. Dans l’optique de détecter des conceptions induisant des problèmes de rendement, on met en place des modèles prenant en entrée des choix de conceptions, et en sortie les mesures d’intérêt. Quand des choix sont possibles, les modèles permettent de prendre les meilleures décisions pour avoir les meilleurs rendements possibles sur le long terme.

Deux enjeux complémentaires sont présents dans ce contexte. Le premier est celui de la modélisation et de la prédiction (dont l’extrapolation est le « digital twin » [Tao et al., 2019]), c’est-à-dire un suivi complet de la production simulée. Le second est dans l’optique de détecter des défauts en analysant l’état d’une puce via la valeur de mesures.

La topographie est devenue critique pour certaines étapes, notamment lors de l’exposition à cause du contrôle du focus, comme présenté dans le Chapitre 2 et dans [Simiz, 2016]. Cette topographie présente différentes composantes : composante wafer, composante champ, composante puce, composante aléatoire. Ces composantes sont mises en évidence dans le chapitre 3 sur la mesure PWG, et dans le chapitre 2 sur les origines de la topographie.

Le cœur du travail de ce chapitre sera de relier la mesure de nano-topographie du PWG au design, étant à l’origine de la variation de topographie à la plus haute fréquence sur un wafer (la composante puce). Il s’agit donc de construire un modèle de prédiction liant la composante puce à ses déterminants contenus dans le design. Ceci a pour but de prendre en compte cette variation haute fréquence devenue problématique pour des procédés durant la phase de conception.

Modéliser cette composante est utile pour le procédé de photolithographie, car la composante wafer est majoritairement compensée par le système de leveling du scanner [Simiz, 2016]. Ce système a vocation à suivre la topographie pour exposer le wafer avec une distance focale image constante. Il est capable de compenser les variations de topographie basses fréquences (variations douces). Cependant, il ne peut pas compenser les variations hautes fréquences. Le résiduel de topographie non compensé par le scanner correspond en grande partie à cette composante puce, comme expliqué dans [Simiz, 2016].

Comme présenté dans le chapitre 3, la nano-topographie est photo-répétée à l'échelle de la puce. Pour deux puces ayant le même procédé de fabrication, cette nano-topographie diffère, mais la photo-répétition au sein d'un même wafer est toujours présente. Elle peut donc vraisemblablement se modéliser avec des informations concernant la puce et son procédé de fabrication. L'enjeu de ce chapitre est de faire un modèle à partir d'une mesure d'un ou plusieurs produits pour prédire la nano-topographie d'un autre produit qui a subi les mêmes étapes de fabrication.

En premier lieu, la préparation des données est détaillée, ce qui nous placera dans un problème académique de modéliser une valeur numérique à partir d'un ensemble de descripteurs. Les choix du modèle, la transformation et la sélection des descripteurs, le choix des critères d'évaluations sont justifiés, et mis en regard d'autres choix. En particulier, l'enjeu pour un usage métier est un compromis entre l'intelligibilité du modèle et sa performance. Le choix du critère d'évaluation sera déterminant pour savoir à quelle question nous répondons. Des critères de confiance a priori sont présentés dans la dernière partie.

4.2 Présentation des données utilisées

Dans cette section, nous introduisons les données utilisées pour la réalisation du modèle. Dans un premier temps, nous présentons les données de sortie : la nano-topographie de la puce, que l'on souhaite prédire. Dans un second temps, les données d'entrée concernant le design : les densités périmétriques et surfaciques.

4.2.1 Données puces : Puces moyennes

La puce moyenne calculée à partir d'une mesure PWG ne comporte plus que la composante puce. C'est cette nano-topographie que l'on souhaite modéliser, et prédire par la suite.

Cette mesure ne comporte plus que la composante nano-topographie car le PWG possède un algorithme de leveling qui filtre majoritairement la composante wafer avec un filtre double Gaussien, comme présenté dans le chapitre 3.

4.2.2 Données design : Densités périmétriques et surfaciques

1. Design d'une couche : format GDS II

Comme dit dans le Chapitre 1, une puce voit plusieurs fois une même machine durant sa fabrication. De la même façon, différentes puces voient la même machine. Certaines puces possèdent le même flot de fabrication, mais avec des designs ou paramétrages différents. La puce s'imprime par empilement de couches minces structurées. Chacune de ces couches a vocation à créer des motifs sur la puce, à des fins différentes (gravure, isolation...). L'ordre dans lequel ces couches sont empilées est le même pour certains produits, mais l'agencement horizontal des couches n'est pas le même. Cet agencement horizontal est connu et représenté par le format GDS II précédemment présenté. Dans la suite de l'explication, on considère que l'intérieur des polygones vaut 1, et l'extérieur des polygones vaut 0 pour chaque couche.

Dans ce chapitre, on souhaite relier le design d'une puce, représenté par un ensemble de fichiers .gds, à cette composante puce. Leur manipulation est donc difficile, et se fait grâce à des

logiciels spécifiques fournis par des entreprises spécialisées dans ce domaine, comme par exemple Mentor avec son logiciel Calibre.

Grâce au logiciel Calibre, et à son langage de programmation associé, le langage SVRF, on peut extraire un format condensé du fichier GDS : les densités périmétriques et surfaciques.

2. Densité surfacique

La densité surfacique d'un fichier GDS est une image, donc une liste de pixels. La valeur associée à chaque pixel est le pourcentage de présence de polygone(s) au sein d'un pixel. Elle représente le pourcentage de matière ou le pourcentage de traitement qui sera ajouté dans la surface que représente le pixel. C'est une bonne estimation de la quantité horizontale de matière.

Ce fichier s'obtient en prenant divers paramètres : une taille de pixel T_p , et une taille de pas d'avancement T_a , ainsi que le fichier .gds. Le fichier de sortie sera alors une liste de pixels de dimension T_p , se superposant (ou en laissant un espace) sur une distance en abscisse et ordonnée de $T_p - T_a$.

Dans le flot de fabrication actuel, la densité surfacique est générée sur les fichiers .gds avec une taille de pixel $T_p = 50 \mu\text{m}$ et un pas de $T_a = 25 \mu\text{m}$, soit une superposition de $25 \mu\text{m}$. Une fois cette compression effectuée, elle est stockée sous la forme d'une image pour que les ingénieurs disposent d'une représentation synthétique de la couche en question.

Au-delà d'une diminution de la taille du fichier, on a surtout une correspondance spatiale avec la mesure de nano-topographie, sous la forme d'une grille (image classique) et non sous la forme d'une liste de polygones comme les fichiers GDS.

Cette expression provoque une perte d'information. On peut trouver des configurations très différentes de polygones où la densité surfacique sera la même. On donne des exemples dans la figure 4.1. Pour compenser partiellement cette perte d'information, une autre compression d'information existe : la densité périmétrique. La combinaison de ces deux vecteurs offre une plus riche expression.

3. Densité périmétrique

La densité périmétrique est une autre forme d'expression d'un fichier .gds. Elle a vocation à générer une image de densité, comme la densité surfacique. Chaque pixel renverra une valeur qui sera plus grande si il y a la présence de périmètres dans ce pixel. Cette valeur a vocation à représenter la quantité de matière « verticale » de traitement qui sera effectué.

L'algorithme de calcul est le suivant :

1. On sauvegarde la liste de polygones initiale
2. On génère la liste de polygones suivante : chaque polygone précédent est « gonflé » de 1nm
3. On calcule la densité du « ou exclusif » des deux listes de polygones. Autrement dit, la différence entre les polygones gonflés et les polygones initiaux.

De cette façon, on voit que la densité périmétrique est d'autant plus grande que des périmètres sont présents au sein d'un pixel. C'est une information complémentaire de la densité surfacique qui donne une idée de la quantité de polygones contenus dans un pixel.

4. Des densités complémentaires

Les densités surfaciques et périmétriques sont pertinentes, car lorsqu'un traitement est effectué sur ces polygones, il y a dépôt de couche mince, ou alors retrait de couche(s) mince(s), même infimes. Dans le cas de la densité surfacique, on a alors une grossière idée de la « surface horizontale » qui sera traitée, et dans le cas de densité périmétrique, cela donne une idée de la quantité de « surface verticale » qui sera en contact avec le traitement.

Sur la figure 4.1, des exemples montrant que les densités périmétriques et surfaciques apportent des informations différentes. On propose trois exemples simples ayant des configurations

spécifiques. L'unité de base de la densité périmétrique est arbitraire. Compte-tenu de l'algorithme, il s'agit d'une projection de la surface verticale sur la surface horizontale. L'information qui est contenue dans la densité périmétrique est la notion de proportionnalité entre elles. Si une densité périmétrique est deux fois plus grande qu'une autre, on sait qu'elle comporte deux fois plus de surface verticale, mais la valeur exacte dépendra de la taille choisie du pixel ainsi que du « gonflement » choisi, qui sont tous deux arbitraires. Pour ces raisons, on ne donne pas de valeur exacte dans les exemples ci-dessous, car ceux-ci rendraient confus le message porté. On notera a la valeur de densité périmétrique d'une « bande » périmétrique complète.

Contours (GDS)	Périmètres	Densité surfacique	Densité périmétrique
		DS = 0,5	DP = a
		DS = 0,5	DP = 5a
		DS = 0,25	DP = a
		DS = 0,5	DP = 2a
		DS = 0,5	DP = 2a

FIGURE 4.1 – Exemples de densités

Sur la figure 4.1 : on voit que dans les exemples 1 et 2, les densités surfaciques sont les mêmes, mais que la densité périmétrique de l'exemple 2 est 5 fois plus grande que dans l'exemple 1. On voit également que dans les exemples 1 et 3, la densité périmétrique est la même, mais la densité surfacique est différente de moitié dans l'exemple 3.

À travers ces exemples simples, on comprend que les deux densités montrent des variables différentes. Ces densités ayant un sens physique et étant à la fois synthétiques, on utilise ces densités pour les entrées en modèle.

On peut également trouver des configurations différentes ayant exactement le même couple (densité surfacique, densité périmétrique), avec un agencement de polygones différent, comme on le montre dans la figure 4.1 avec les exemples 4 et 5. Cela prouve qu'il y a une perte d'information lorsqu'on met l'information sous cette forme.

4.2.3 Cas d'étude

1. Les données du problème

Les topographies moyennes de la figure 4.2 et les densités montrées sur les figures 4.3, 4.4 et 4.5 sont celles des produits auxquels nous nous intéressons. Les fichiers sont générés dans le flot normal de la fabrication des produits. Ainsi, le travail qui sera effectué pourra s'intégrer dans le flot de fabrication sans générer de nouveaux fichiers. On enrichit l'usage de ces données existantes avec un nouvel usage.

Dans la pratique, on souhaite minimiser la génération de fichiers afin de ne pas surcharger à la fois le temps de calcul pour créer ces fichiers, et également ne pas alourdir le besoin de stockage de l'entreprise. Il s'agit d'un enrichissement strict de la donnée : on fait un lien entre des données existantes en en faisant un modèle. Il n'y a pas de coût supplémentaire en terme de temps de calcul pour générer de nouvelles données.

Les problématiques de stockage et d'accès aux données ne sont pas à négliger dans l'industrie. Ici, les densités sont générées dans le processus de fabrication, mais celles-ci ne sont stockées que sous la forme d'images, donc avec un usage limité. Leur usage est informatif et une représentation visuelle simple suffisait au procédé pour les ingénieurs car aucun traitement n'est effectué à ce jour sur ces fichiers. Ainsi, utiliser ces données pour le modèle demandera à intégrer à long terme leur stockage et on pourra le faire au moment de leur première génération, qui fait déjà partie du processus.

Dans l'histoire du déroulement de la thèse, on disposait des mesures d'un seul produit (figure 4.3), puis on a pu disposer des mesures des deux autres produits ainsi que de leurs densités (figure 4.4 et 4.5). Ces trois puces ont un flot de fabrication similaire. Elles ont subi les mêmes étapes de fabrication, et on souhaite modéliser leur nano-topographie à un stade où celle-ci est critique. Il s'agit de la couche CNT (contact), où un lien entre des couches minces (lien vertical) est effectué : c'est une étape critique, et l'exposition doit se faire à une distance focale précise, donc la topographie impacte directement la bonne réalisation de cette étape.

Les produits 281PA et 281RA sont très similaires en termes de conception. Ils sont à la demande du même client de ST, et ont une fonctionnalité similaire. En revanche, le produit 282FA n'est pas du même client, et a une conception différente. Leurs procédés de fabrication suivent cependant les mêmes étapes.

On peut observer dans les figures 4.2, 4.3, 4.4, 4.5 que les produits 281PA et 281RA se ressemblent (présence de la forme en « L » dans les angles), et que le produit 282FA est différent. On peut également observer une similitude à l'œil nu entre les densités et la topographie. On remarque que les structures de design apparaissent également sur la topographie de la puce.

Les densités périmétriques de la couche NOSO ne sont pas générées dans le flot standard de fabrication, car son utilité n'est pas avérée jusqu'à maintenant. Une piste serait de demander à la générer si le besoin était présent dans la suite.

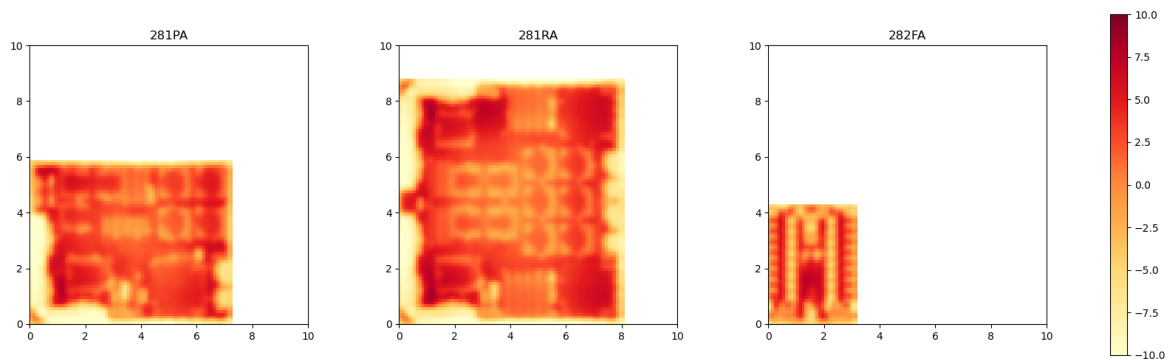


FIGURE 4.2 – Topographies moyennes à l'échelle des produits 281PA, 281RA et 282FA

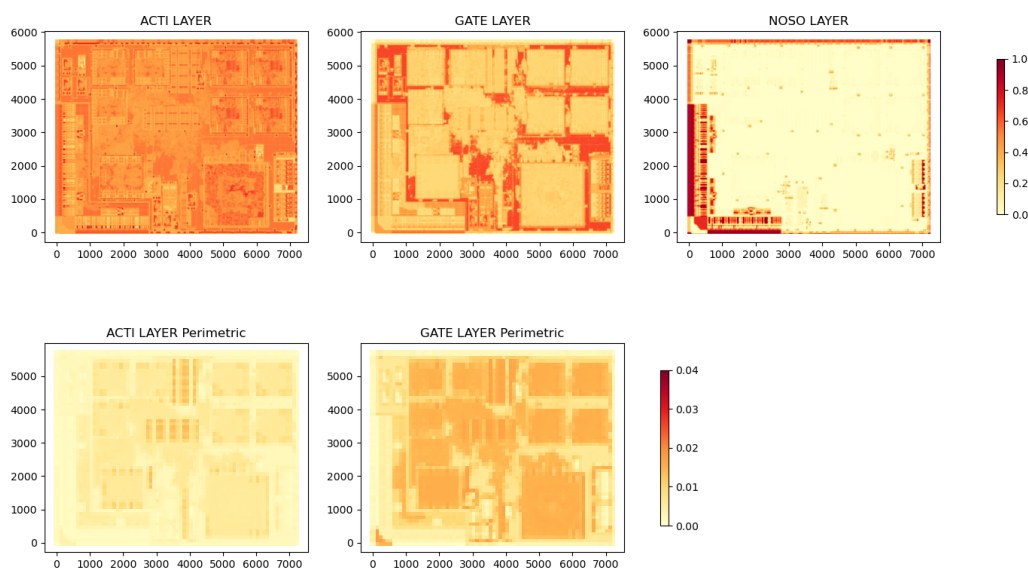


FIGURE 4.3 – Densités surfaciques et périmétriques du produit 281PA

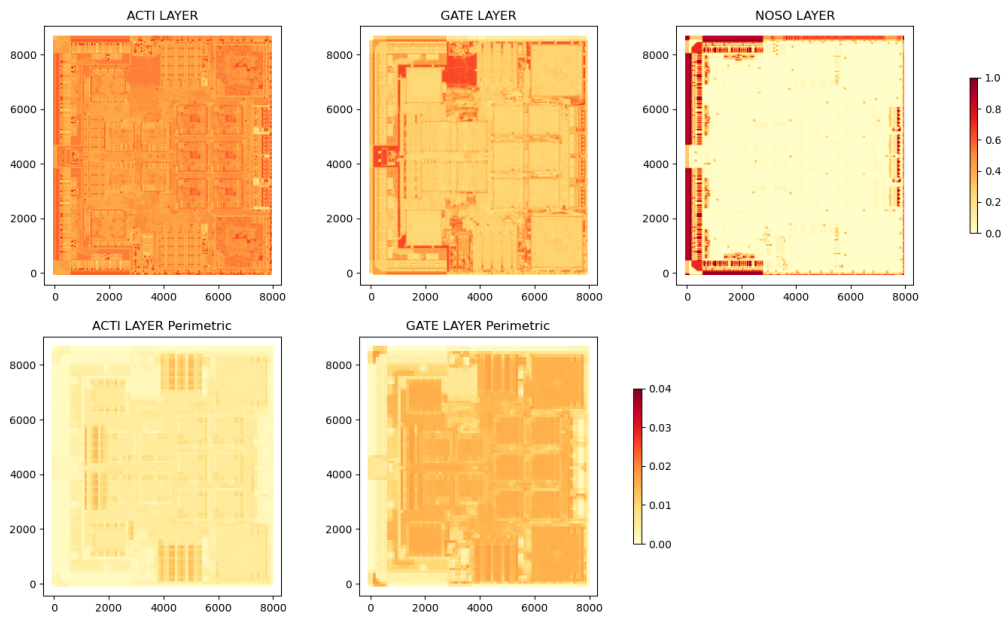


FIGURE 4.4 – Densités surfaciques et périmétriques du produit 281RA

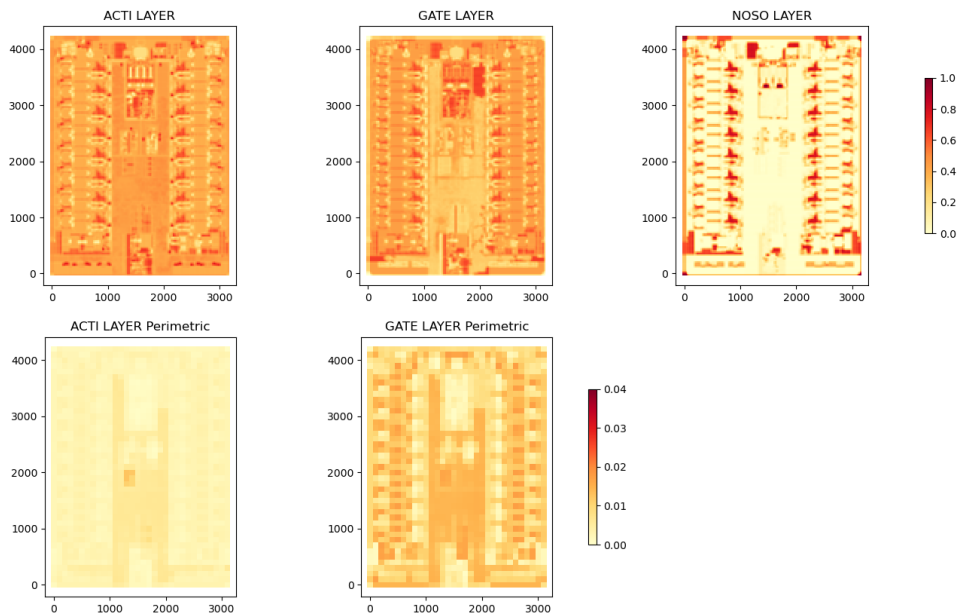


FIGURE 4.5 – Densités surfaciques et périmétriques du produit 282FA

2. Mise à l'échelle des données

On dispose des données de 3 couches pré-sélectionnées pour chacune des trois puces. On sait pour des raisons métier que seules celles-ci impactent la topographie de la puce. On dispose

de ces 3 couches pour les 3 produits. Les densités ont été générées avec des pixels d'une taille différente de la puce moyenne. On décide de commencer par mettre ces points aux mêmes dimensions, pour avoir le même nombre de points sur chaque couche et sur la puce moyenne associée. Ensuite, on fera correspondre les points spatialement en réglant l'alignement.

Le choix qui a été fait a été de mettre tous les fichiers à la résolution la plus haute : 100 μm , à savoir la résolution de la mesure de nano-topographie. On ne transformera pas la mesure moyenne de nano-topographie, mais celles des densités uniquement. Il existerait plusieurs façons de changer la résolution d'une image (interpolation bicubique [Mitchell and Netravali, 1988], Glasner [Glasner et al., 2009], réseaux de neurones (EnhanceNet [Sajjadi et al., 2017])...).

Le choix qui a été fait est lié au sens des données. On a des données qui sont des densités données sur une surface. On souhaite re-mettre à une échelle plus grande, et la dimension de la nouvelle grille est un multiple de la grille précédente. En ce sens, chaque pixel de la nouvelle grille est exactement rempli par 4 pixels ayant une valeur de densité. Si l'on recalculait la densité avec les données brutes, on aurait alors la nouvelle densité qui est exactement le barycentre des points qui sont situés sur la grille. Nous obtenons donc en fichier la densité surfacique et périmétrique exacte qu'on aurait pu calculer directement sur les fichiers GDS mais avec une grille plus grande avec la méthode utilisée.

Il aurait pu être demandé de re-calculer cette densité en modifiant le programme d'origine (langage SVRF), mais comme dit précédemment, on souhaite minimiser l'impact sur la chaîne de fabrication déjà existante. Ainsi, re-générer une densité à partir de celle qui a été calculée précédemment est une meilleure solution pour ne pas alourdir la charge de travail des autres métiers.

Le problème reste de savoir si les points du design (information logique) et les points de la mesure (information mesurée) sont alignés. On détaille cela dans le point suivant.

3. Alignement des données

Dans le chapitre 3, nous mentionnions que la grille de la mesure n'était pas alignée avec les puces présentes sur le wafer. Ici, il faut aligner les densités avec la puce moyenne. Les dimensions sont les mêmes, et on fait l'hypothèse que les points les plus en bas à gauche de la puce correspondent exactement aux points les plus en bas à gauche du design.

Cette hypothèse est raisonnable, car les problèmes d'alignements qui étaient présents dans le chapitre 3 sont « lissés » par la moyenne calculée. Le dés-alignement est tantôt positif, tantôt négatif, et borné par un intervalle de +/- 50 μm par arrondi. En ayant une infinité de puces sur un wafer de rayon infini, on convergerait vers une puce moyenne dont les bords sont parfaitement alignés avec le design.

On donne la figure 4.6 ci-dessous pour montrer comment les pixels sont regroupés.

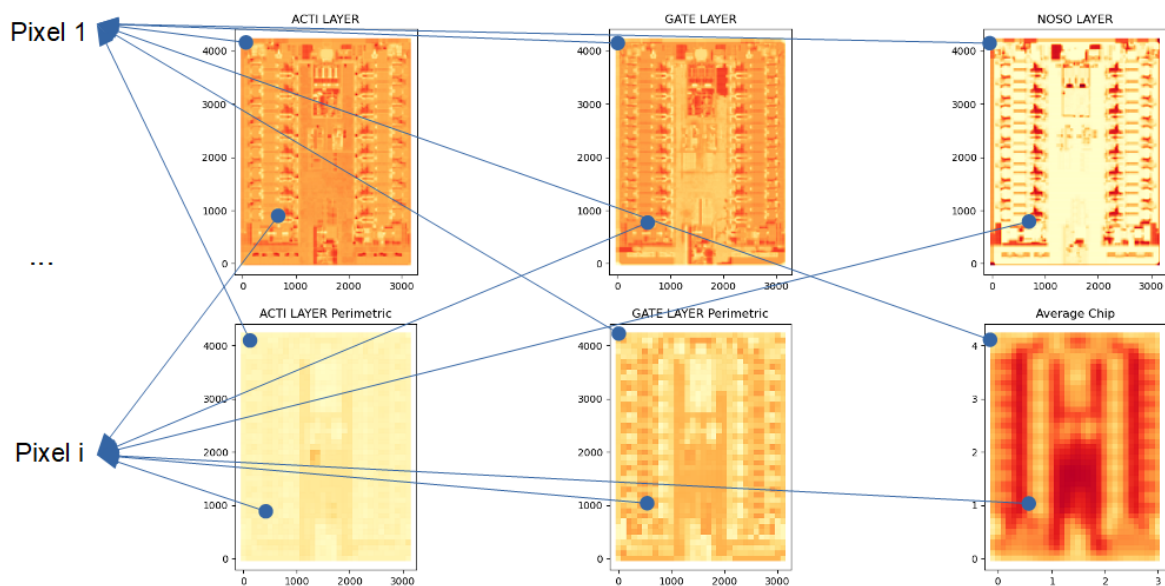


FIGURE 4.6 – Schéma explicatif pour la mise en forme des données (282FA)

On génère une liste de valeurs où chaque ligne représentera un pixel de taille 100 μ m et dont les valeurs associées seront :

1. la localisation du point en X et en Y sur la puce
2. la valeur de la nano-topographie moyenne correspondant à ce point
3. les valeurs des densités de ce même point

On propose une description de certaines caractéristiques des données dans la sous-section suivante.

4.2.4 Corrélations et distributions entre les données

Dans l'optique de se familiariser avec les données ainsi que d'observer d'éventuels liens entre eux, on montre les figures 4.7 et 4.8. Les corrélations sur la figure 4.7 mettent en évidence qu'il existe un lien entre la composante à prédire (Z) et la plupart des couches. Ce graphique montre que la composante Acti_surfacique n'est pas beaucoup en corrélation avec la composante à prédire.

D'un produit à l'autre, il existe une différence de coefficients sur les matrices de corrélation de la figure 4.7. En particulier celui concernant la corrélation entre Acti_surfacique et Acti_perimetrique. Il y a un changement de signe sur la matrice du produit 282FA. Cette corrélation s'interprète comme cela : la couche Acti ne comporte pas beaucoup de périmètres (ou de contours) sur les produits 281PA et 281RA, car plus il y a de surface, et moins il y a de périmètres. Cette inversion de corrélation sur le produit 282FA implique que la couche Acti de ce produit contient plus de contours à mesure que la valeur Acti augmente : les structures sont plus petites sur ce produit.

La corrélation entre Noso_surfacique, d'une part, et Acti_perimetrique et Gate_perimetrique d'autre part est également bien plus prononcée sur le produit 282FA que sur les deux autres. L'interprétation est qu'il y aura d'autant plus de couches Acti et Gate sur la couche Noso sur le produit 282FA, alors que ce n'est pas le cas sur les deux autres produits, sur lesquels la corrélation est proche de 0.

Sur l'axe des abscisses de la figure 4.8, les échelles sont représentées : les densités périmétriques ne dépassent pas la valeur 0.02, tandis que les densités surfaciques ont des valeurs

plus élevées, en particulier dans le cas de la couche NOSO (allant jusqu'à 0.94). On remarque également que la topographie descend jusqu'à -6nm pour le produit 282FA tandis qu'elle descend jusqu'à -10nm dans les deux autres produits.

Ces deux figures montrent que le produit 282FA se distingue des deux autres. Cette différence est présente dans les distributions des produits, mais également sur la matrice des corrélations. Cette différence a priori peut se retrouver sur la modélisation si la puce 282FA n'est pas représentative des deux autres, et vice-versa. Nous explorons, discutons et montrons que cette singularité est retrouvée dans la modélisation dans les parties suivantes du chapitre.

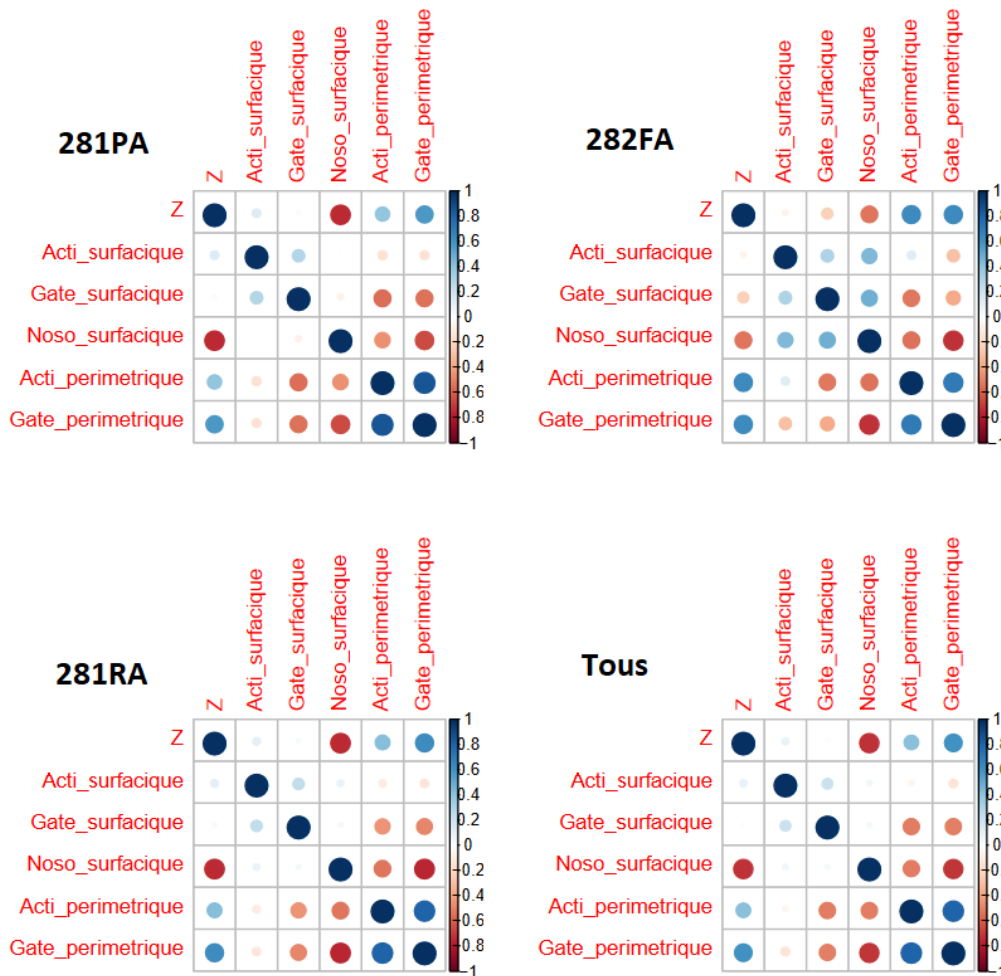


FIGURE 4.7 – Corrélations entre les vecteurs d'entrée

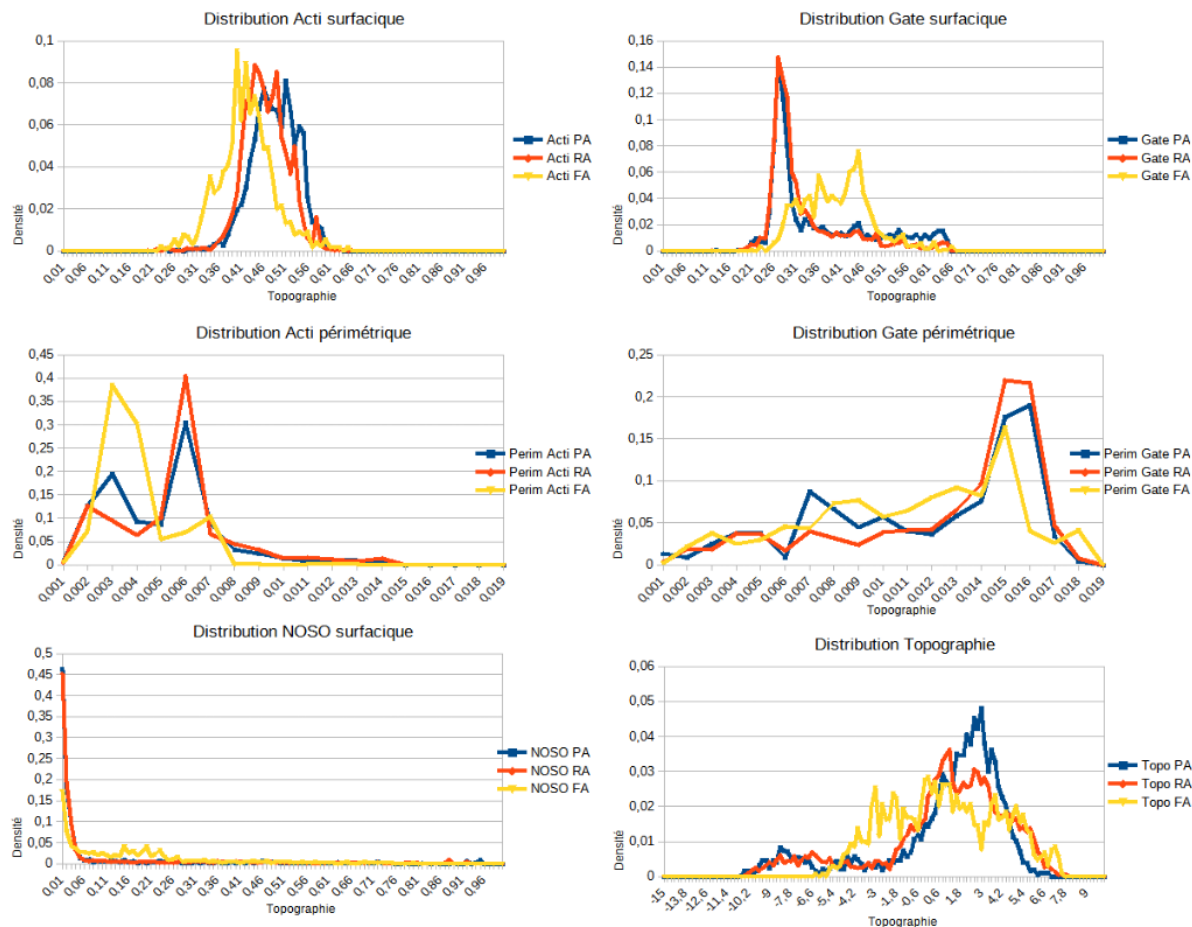


FIGURE 4.8 – Distributions des valeurs des couches et de la topographie

4.3 Méthodologie et choix du modèle

4.3.1 Quels modèles et pour quels objectifs ?

On souhaite faire un modèle prédictif de la colonne « topographie » avec toutes les autres colonnes en entrée. On commence par se positionner dans la littérature, puis on décrit les choix qui seront faits pour la modélisation dans cette thèse.

1. Précédent travaux de modélisation

Un précédent travail effectué lors d'une thèse [Simiz, 2016] avait montré des résultats encourageants avec des modèles linéaires. Celui-ci avait pris pour cas d'étude un autre produit, et des mesures provenant d'un d'autre outil de mesure de nano-topographie : le Wyko. L'outil Wyko est un outil fournissant une mesure de nano-topographie avec une résolution horizontale variable allant de 50 μ m à 3 μ m. C'est également par interférométrie qu'il effectue sa mesure.

Dans ce travail, le choix avait été de mettre beaucoup de vecteurs afin de constater un lien. La figure 4.9 montre la façon dont le modèle est calculé.

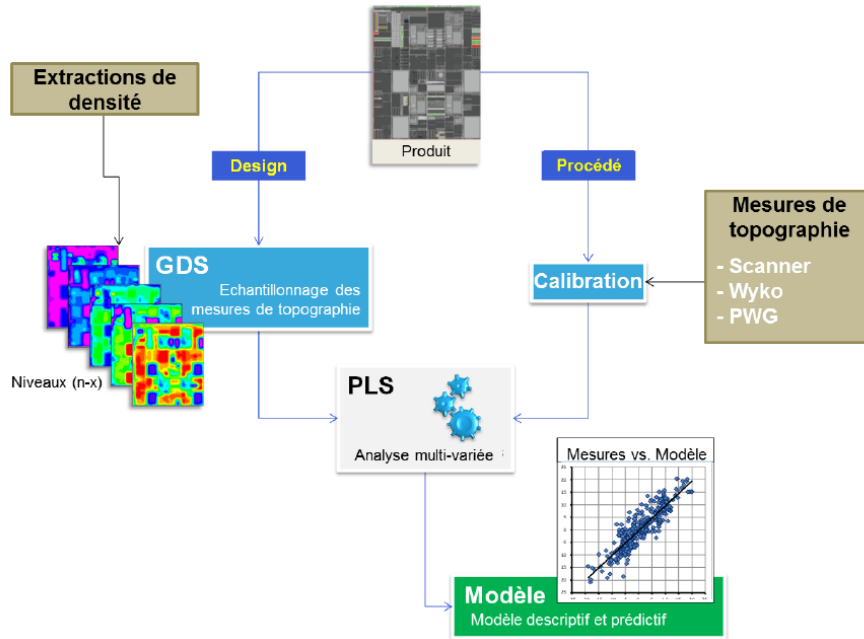


FIGURE 4.9 – Figure explicative de la méthode de prédiction - *Source : [Simiz, 2016]*

Dans cette figure, on voit le paradigme évoqué par l’auteur : prendre des mesures de topographie pour la calibration en sortie, et des échantillons de densité en entrée pour construire un modèle linéaire en utilisant la technique PLS (Partial Least Squares) [Parodi, 1965], une technique de régression linéaire. Dans la figure 4.10, les résultats de modélisation sont présentés.

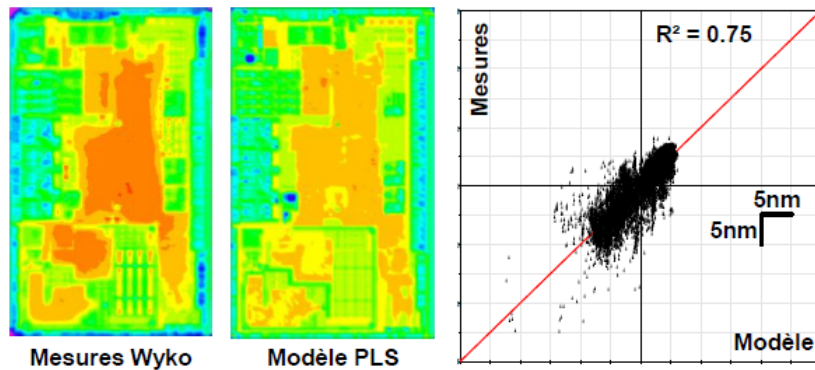


FIGURE 4.10 – Résultats d’un modèle calibré sur les mesures Wyko - *Source : [Simiz, 2016]*

Pour donner du contexte à ce travail, il a été effectué par un ingénieur en micro-électronique. Notre enjeu est de reprendre ses idées, avec un regard sciences de données et métier, accentué en sciences de données. À notre connaissance, il n’existe pas d’autres travaux concernant la modélisation de topographie utilisant le design en entrée en micro-électronique. En revanche, il existe des modèles de topographie simulant le résultat de procédés dont la finalité est la topographie, en particulier des entreprises spécialisées comme Coventor et Synopsys.

2. Notre choix de modèle

Il existe diverses techniques de modélisation : modèles linéaires, réseaux de neurones, arbres de décisions [Rakotomalala, 2005]... Aujourd’hui, les réseaux de neurones suscitent beaucoup d’intérêt et d’engouement, en particulier le deep learning [LeCun et al., 2015]. Il y a une volonté d’utiliser ces techniques de prédiction gourmandes en calcul désormais exploitables, juste-

ment grâce aux progrès de la micro-électronique, qui permettent des parallélisations de calculs extrêmement efficaces, notamment grâce au calcul GPU [Owens et al., 2008]. Mais ces réseaux sont complexes et leurs contenus échappent à l'utilisateur, surtout dans le cas de réseaux profonds, où seul un résultat est donné en sortie. Libre à l'utilisateur d'aller regarder en détail le contenu des coefficients dans ces réseaux, mais donner du sens est fastidieux, voire impossible.

La coopération entre les sciences de données et les métiers est au cœur de difficultés à la fois techniques et humaines. En utilisant un modèle compréhensible par tous les métiers, le langage commun permet de mieux répondre aux besoins de l'entreprise. En particulier, on cherche à relier des conséquences à des causes, comme dans [Lam et al., 2015], où l'on cherche à déterminer les facteurs causant l'overlay.

Dans notre cas d'étude, nous réalisons des modèles linéaires, car la modélisation de topographie n'est qu'à ses débuts. Il est donc sensé de commencer par des modèles simples. Ce choix se justifie également par une volonté d'avoir un modèle toujours compréhensible. Retirer la complexité de l'élément « modèle » de l'ensemble du système de prédiction simplifie grandement la compréhension.

La compréhension est un élément important dans les sciences, et non uniquement les prédictions. Dans une régression linéaire, on peut donner du sens à chacun des coefficients, et ce sens que l'on peut donner est riche pour faire des raisonnements plus complexes en les reliant aux métiers associés. Un choix qui aurait été possible est un modèle de complexité modérée, mais le parti pris est de commencer par un modèle extrêmement simple et de faire fonctionner une prédiction. Si un modèle linéaire est suffisamment efficace, alors on devrait l'utiliser.

Nous allons donc, dans la réalisation des modèles, nous focaliser sur les données d'entrées, leurs transformations, leur sélection ainsi que leur échantillonnage. Dans les deux sous-parties suivantes, nous réalisons un modèle sur une seule puce, afin d'expliquer la méthodologie de réalisation du modèle, et également de déterminer les limites de prédictions d'un modèle.

La formalisation du problème est la suivante :

- On dispose de P vecteurs colonnes (v_1, v_2, \dots, v_p) ayant N points chacun qui sont ici les entrées, les densités surfaciques et périmétriques précédemment évoquées. On dispose également d'une colonne qui est l'objectif de prédiction : la topographie précédemment mentionnée, que l'on note B ayant N points également.

- On cherche à obtenir un modèle linéaire approximant B à partir des vecteurs d'entrées. Autrement dit, un ensemble de P coefficients $X = (c_1, c_2, \dots, c_p)$ qui, lorsqu'on effectue la combinaison linéaire $c_1.v_1 + c_2.v_2 + \dots + c_p.v_p$ s'approche le plus possible de B .

En notation matricielle, cela donne : trouver X qui minimise $\|AX - B\|$. La technique qui sera utilisée est la pseudo-inverse de Moore-Penrose. Cette technique est celle qui donne le modèle linéaire minimisant la norme L_2 le plus efficace sur les données d'entraînement. C'est une régression linéaire classique.

4.3.2 Un premier modèle

On dispose de trois puces. Pour commencer, on effectue un modèle sur une seule de ces puces à la fois. Cela permet de vérifier les limites du lien entre les densités et la nano-topographie. En calibrant le modèle sur la totalité de la puce, on obtient une borne optimiste de la performance d'un modèle prédictif linéaire.

Dans la figure 4.11, on voit les résultats de modélisation individuelle sans transformer les entrées. Le coefficient de détermination R^2 , est un indicateur de proximité de la valeur de la nano-topographie des puces. Ce résultat constitue une évaluation optimiste pour le critère du R^2 pour toute prédiction sur cette puce avec ce modèle linéaire. Cette évaluation est de 0.60 pour le produit 281PA, de 0.63 pour le produit 281RA et de 0.51 pour le produit 282FA.

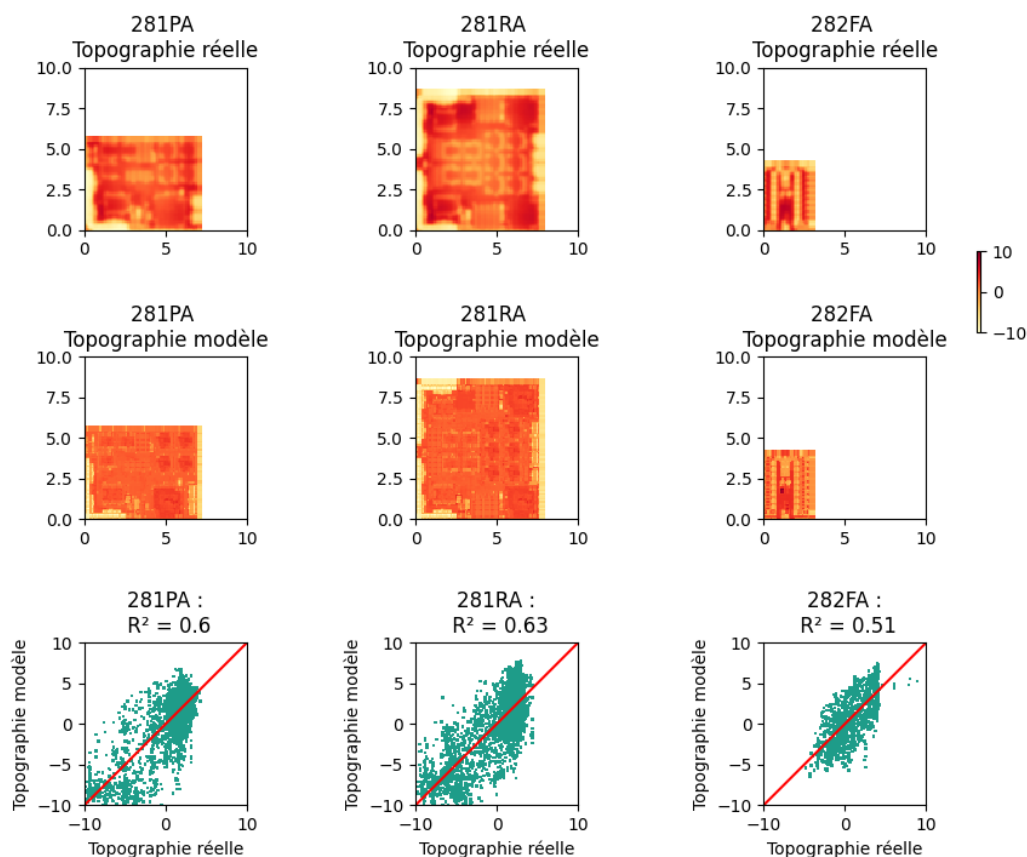


FIGURE 4.11 – Modélisation séparée des puces avec densités brutes

L'avantage de ce modèle est sa simplicité à la fois dans le modèle qui est linéaire, et à la fois dans les données qui ne sont pas modifiées. On laisse les densités périmétriques et surfaciques telles qu'elles ont été générées par le processus de fabrication, et ainsi on simplifie le déploiement du modèle sans générer de transformations supplémentaires.

Comme évoqué, l'autre avantage d'un modèle linéaire est d'avoir la possibilité d'interpréter le modèle grâce aux coefficients. Cela reste possible tant que le nombre de colonnes n'est pas trop élevé. En effet, plus le nombre de coefficients est élevé, plus l'interprétation est difficile. Dans ce cas, il y a 5 coefficients associés aux entrées et un coefficient associé à une constante. La figure 4.12 montre ces coefficients pour interprétation.

	281PA	281RA	282FA
Acti surf.	11.22	11.97	-2.22
Gate surf.	3.09	8.16	8.54
Noso surf.	-10.53	-9.56	-4.09
Acti perim	-140.43	-168.87	1034.07
Gate perim	328.67	430.19	154.39
Constante	-7.83	-10.80	-6.25

FIGURE 4.12 – Coefficients du modèle sans filtre

Sur la figure 4.12, tous les coefficients sont différents de 0 et jouent un rôle dans la modélisation (p-valeur du test de Student inférieure à 0.001) à l’exception du vecteur Acti surf pour le produit 282FA qui n’est pas significatif.

Dans les trois modèles, on obtient un signe négatif pour la couche Noso. Cela fait sens, car le procédé associé à cette couche est de « creuser » dans la puce. Les coefficients associés à la couche Gate sont tous positifs, qu’ils soient surfaciques ou périmétriques et la composante Gate correspond à un ajout de matière.

Les différences dans les valeurs absolues des coefficients sont liées aux échelles des données mises en jeu : les densités périmétriques sont significativement inférieures aux densités surfaciques, et les coefficients sont significativement plus élevés en valeur absolue sur les couches périmétriques. On a montré grâce aux distributions les ordres de grandeurs des densités sur la figure 4.8.

Les produits 281PA et 281RA étant similaires dans leurs conceptions et dans leurs données, il n’est pas surprenant de voir deux modèles relativement similaires avec des coefficients proches sur toutes les colonnes. Cette similitude apparaît avec la ressemblance forte entre les corrélations (Figure 4.7) et les distributions (Figure 4.8). Dans le produit 282FA, qui est très différent, on voit des changements de valeurs dans la plupart des coefficients des colonnes.

Une perspective serait d’expliquer le changement de signe des couches Acti surfacique et Acti périmétrique sur le modèle. Une piste d’exploration est de vérifier si les informations dans la couche Acti surf et Acti perim sont redondantes, et si, selon les données, le modèle ne va pas privilégier l’un ou l’autre des vecteurs pour optimiser le R^2 , quitte à inverser les signes dans l’un ou l’autre cas. La figure de corrélation 4.7 donne un élément de réponse : la corrélation entre les deux couches change de signe : elle est négative sur les produits 281PA et 281RA, et positive sur le produit 282FA.

Lorsqu’on regarde les formes des puces, et qu’on les compare avec leurs réelles nano-topographies sur la figure 4.11, elles ne sont pas si éloignées des puces réelles, en tout cas à l’œil nu. Les graphiques montrent également que la topographie est « attrapée » dans sa tendance, et avec une marge d’erreur constante. Si l’on compare à la courbe $Y = X$, on voit que les valeurs sont contenues dans un tube de largeur $+/- 5nm$.

Le R^2 n’est peut-être pas le seul indicateur qui devrait être utilisé pour évaluer la performance d’un modèle puisque la forme est bien capturée pour les puces. En s’intéressant aux usages de la nano-topographie, on peut vérifier que des usages sont les mêmes, ou bien utiliser d’autres critères de ressemblance.

4.3.3 Évaluation des modèles

Dans cette sous-section, on fait état des différentes façons d’évaluer un modèle de la nano-topographie d’une puce et la nano-topographie réelle. On peut répondre avec deux points de vues : le point de vue sciences de données, et le point de vue métier. Ces deux points de vue ne s’opposent pas, mais ont vocation à répondre à des questions différentes. Le premier, dans l’état de l’art de la modélisation, répond plutôt à la question de la proximité entre deux vecteurs ou deux images, et certains de ces indicateurs prennent également en compte la complexité du modèle. Le second, plutôt lié au contexte, a vocation à vérifier si l’usage de la mesure est le même que celui du modèle, par exemple en vérifiant si les Peak-to-Valley sont proches.

1. Indicateurs de sciences de données

D’un point de vue sciences de données, évaluer la proximité entre deux vecteurs laisse le champ à une infinité de choix. Il s’agit de toutes les distances possibles entre deux vecteurs. Dans [Witten et al., 2011], on mentionne diverses normes classiques. En notant $P = (P_1, P_2, \dots)$ le vecteur des valeurs prédites et $R = (R_1, R_2, \dots)$ le vecteur des valeurs réelles, on a :

$$L_p = \sqrt[p]{\sum_{i=1}^n |P_i - R_i|^p}$$

Ces normes, assez semblables dans leurs calculs, ont l'avantage de se calculer en des temps proches, mais ont pour différence que plus p est grand, et plus « sanctionnés » seront les gros écarts. En faisant tendre p vers l'infini, on converge vers l'indicateur de l'écart maximum divisé par le nombre de points, qui est lui-même une norme : L_∞ . Ce dernier indicateur ne sanctionne pas un vecteur en entier mais sanctionne uniquement le pire des points.

$$L_\infty = \lim_{p \rightarrow \infty} \sqrt[p]{\sum_{i=1}^n |P_i - R_i|^p} = \max_i |P_i - R_i|$$

Plutôt que de mesurer la distance entre les vecteurs, on considère souvent leur adéquation ou qualité d'ajustement, couramment mesuré par le coefficient de détermination R^2 .

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i^{mesure} - y_i^{prédit})^2}{\sum_{i=1}^n (y_i^{mesure} - y^{moyen})^2}$$

La fraction $\frac{\sum_{i=1}^n (y_i^{mesure} - y_i^{prédit})^2}{\sum_{i=1}^n (y_i^{mesure} - y^{moyen})^2}$ est une distance entre la variabilité prédite et la variabilité réelle, allant de 0 dans la meilleure des prédictions, en passant par 1 dans le cas particulier de la valeur moyenne en prédiction, et pouvant aller jusqu'à $+\infty$ dans des très mauvaises prédictions. Cela donne respectivement dans ces trois cas mentionnés un R^2 de 1 (prédiction parfaite), 0 (prédiction de qualité équivalente à la valeur moyenne) et $-\infty$ (très mauvaise prédiction).

Les normes ainsi que le critère du R^2 montrent une proximité entre deux objets, mais ne font pas état de la complexité d'un modèle. Des indicateurs prennent en compte la complexité d'un modèle pour apprécier à la fois la performance du résultat tout comme la parcimonie du modèle. Parmi ces indicateurs, on peut citer le R^2 ajusté [Witten et al., 2011] ou encore l'AIC [Hastie and Pregibon, 1992].

On souhaite avoir le modèle le plus simple possible pour décrire un phénomène. Ce principe, connu sous le nom de rasoir d'Ockham, dit qu'il ne faut pas utiliser plus d'éléments que nécessaires pour décrire quelque chose. Bien qu'il soit attribué au philosophe Guillaume d'Ockham vivant au XIVème siècle, ce concept avait déjà été évoqué par Aristote dans son premier livre de Physique : « Il vaut mieux prendre des principes moins nombreux et de nombre limité ». Autrement appelé longueur de description minimale, il s'agit de privilégier les descriptions les plus compactes, les plus simples, dans les sciences.

Avec ce principe, dans notre cas on limite le nombre de colonnes pour éviter de complexifier le modèle. Le deuxième effet bénéfique de cela est que plus un modèle est simple, plus hautes sont les chances d'avoir attrapé le phénomène sous-jacent de la modélisation car le sur-apprentissage est moins probable. On a ainsi une capacité de prédiction plus forte. Lorsqu'un modèle est trop complexe, on augmente les chances de sur-apprentissage sur les données, et on passe à côté du phénomène qui induit le lien entre les entrées et les sorties.

2. Indicateurs métier

D'un point de vue métier, on veut s'assurer que les usages de la nano-topographie restent inchangés, c'est-à-dire que les décisions prises seront les mêmes avec la prédiction qu'avec la mesure. Dans le chapitre 3, nous avons listé différentes métriques, telles que le Peak-to-Valley, le gradient, les distributions de gradient, les métriques de la distribution de gradients, le TPI... On pourrait par exemple vérifier que même si la nano-topographie n'est pas exactement la même selon un critère de distance, ces usages restent les mêmes.

On va ici se demander si la topographie modélisée possède les mêmes usages que la topographie réelle. Dans ce cas, on peut considérer la modélisation comme pragmatique, sans nécessairement satisfaire des critères de proximité mathématique comme évoqué précédemment.

On utilisera le Peak-to-Valley en tant que critère d'usage, car il est celui qui est utilisé actuellement à ST pour caractériser la topographie. Pour avoir un critère de performance, le critère sera donc la différence entre le Peak-to-Valley réel, et le Peak-to-Valley de la modélisation.

3. Indicateurs retenus

Pour évaluer les modèles dans la suite du manuscrit, on utilisera le R^2 , le Peak-to-Valley et le nombre de vecteurs. On montre sur la figure 4.13 un exemple d'évaluation de prédictions avec un code couleur associé.

Produits de calibration	Produit d'inférence	Modèle 1		
		R^2	Nb vecteurs	PV diff
281PA	281PA	0,6	5	-2,39
281RA	281RA	0,63	5	-2,61
282FA	282FA	0,51	5	1,52

Nb Vecteurs	R^2	PV diff abs.
5	< 0.2	< 2
5 à 15	0.2-0.35	2 – 5
15 à 25	0.35-0.5	5 – 10
25 à 40	0.5-0.7	10 – 15
> 40	> 0,7	> 15

	Peak-to-Valley (nm)
281PA	17,39
281RA	18,28
282FA	13,97

FIGURE 4.13 – Évaluation du premier modèle avec les critères retenus

Ce tableau est une forme synthétique de l'évaluation d'un modèle contenant les indicateurs mentionnés, avec un code couleur. Le code couleur aide à la lecture, et les valeurs choisies pour le définir sont plutôt arbitraires. Lorsqu'un R^2 est supérieur à 0.5, on a une prédiction satisfaisante dans notre contexte. Pour le Peak-to-Valley, une différence inférieure à 5nm est satisfaisante. Pour le nombre de vecteurs, on peut donner du sens à un modèle si le nombre de coefficients ne dépasse pas un trop grand nombre, dépendant du nombre de couches. On considère que moins de 3 coefficients par couche est très intelligible.

4.3.4 Modèles avec enrichissement sur les entrées

On cherche à améliorer les capacités de prédiction du modèle tout en conservant un modèle linéaire. Dans le modèle précédent, la correspondance se fait point à point, et ne prend pas en compte les points voisins des densités lors de la modélisation. Dans la réalité physique des procédés, la plupart des procédés affectant la topographie ont des effets courte portée et longue portée, variant selon le procédé, mais se situant entre 0 et quelques millimètres d'impact.

Afin de représenter ce phénomène dans le modèle, on transforme les vecteurs d'entrée pour que chaque nouveau point prenne en compte son entourage. On applique donc un filtre convolutif pour représenter cela. Un filtre convolutif est l'application d'un noyau sur chaque pixel d'une image. Ce noyau est une matrice de coefficients qui multipliera chaque pixel de l'entourage du pixel central par la valeur associée sur la matrice selon son positionnement. La figure 4.14 montre un exemple de résultat.

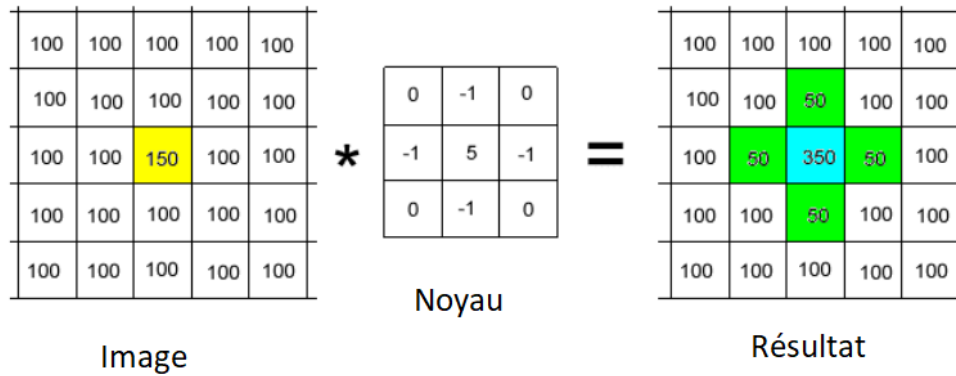


FIGURE 4.14 – Exemples de convolution

Différents types de filtres convolutifs existent. La figure 4.15 montre des exemples-type de noyaux de convolution utilisés dans le traitement d'images. Un cours complet fait un état de l'art des filtres classiquement utilisés [GIMP, 2021].

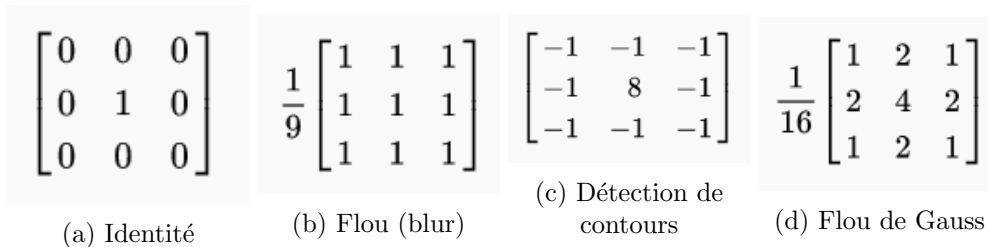


FIGURE 4.15 – Exemples de noyaux de convolution 3x3

Dans la mesure où la modélisation de topographie est un domaine peu exploré, on commence par un noyau simple : le flou (figure 4.15b). Ce noyau peut s'étendre avec la taille de la matrice, si l'on divise par le nombre de cases au final, en laissant des 1 dans les coefficients de la matrice.

Dans la figure 4.16, on montre le résultat des filtrages effectués. Dans la pratique, nous avons généré 20 images de noyaux flou ayant un côté de matrice de valeur 1, 3, 5, ..., 39. On décide de s'arrêter à 39 car cela correspond à une portée de 2mm, ce qui correspond à la limite spatiale d'action des procédés mis en jeu. De plus, l'image devient presque totalement uniforme à la portée 19, comme on le voit sur la figure 4.16. Le nombre de colonnes d'entrée du modèle passe de 5 à 100.

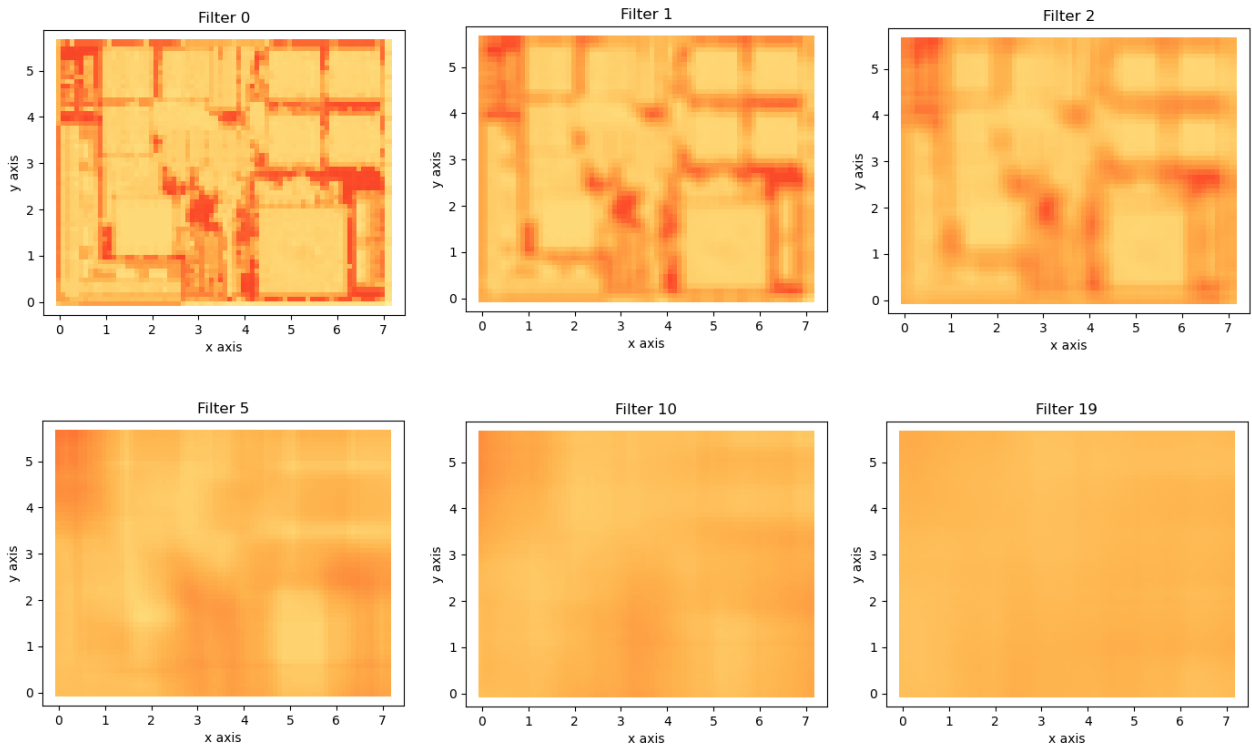


FIGURE 4.16 – Exemples de filtrages blur avec un noyau de matrice carrée : 0, 1, 2, 5, 10, 19

Un produit est manufacturé par centaines sur un wafer circulaire, comme on l'explique dans le chapitre 2. Certains procédés ont lieu à l'échelle de la puce, mais certains se déroulent à des échelles plus grandes, et ainsi il existe des interactions entre puces voisines. On montre une carte d'exposition d'un des produits dans la figure 4.17. Les bords de puces interfèrent les uns avec les autres au niveau topographique. Chaque puce est séparée d'un canal de découpe, mais des interactions existent quand même. Afin de prendre en compte celles-ci, on transforme chaque couche de design en ajoutant la notion de puce voisine en les juxtaposant, comme montré dans la figure 4.18.

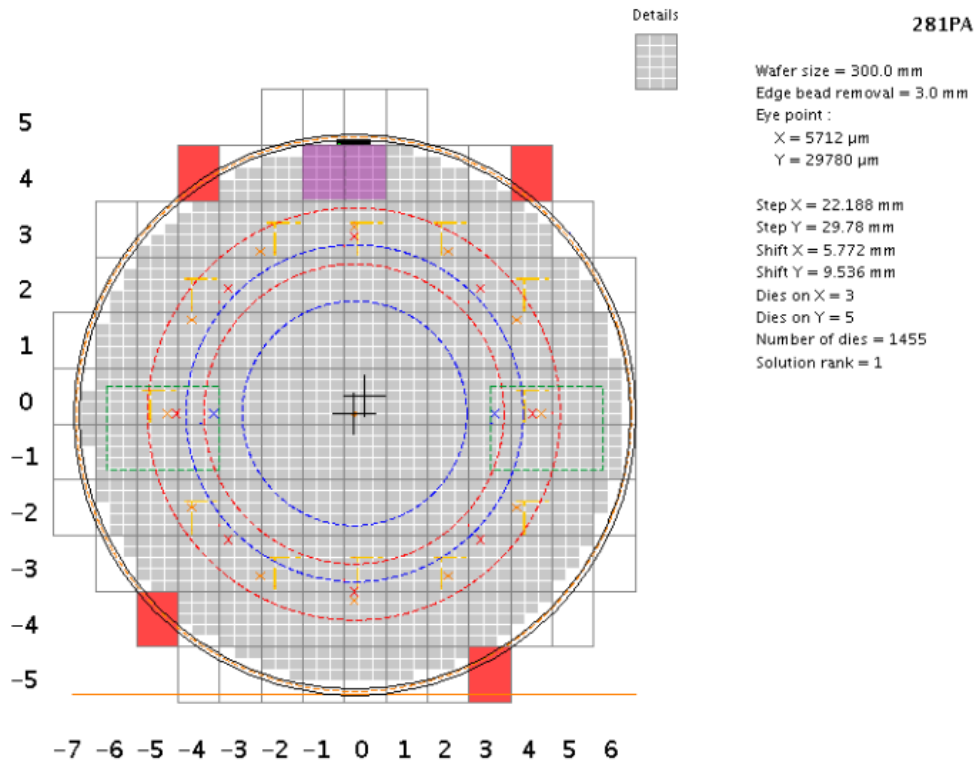


FIGURE 4.17 – Carte d'exposition du produit 281PA

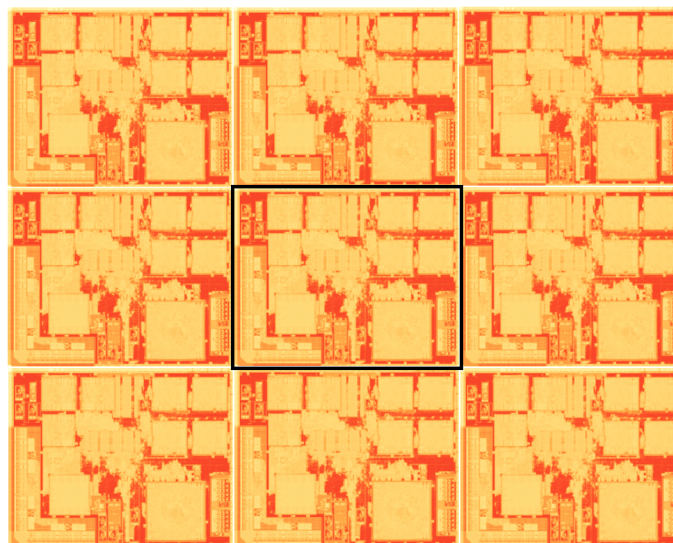


FIGURE 4.18 – Juxtaposition des puces

Une fois cette juxtaposition effectuée, on filtre cette nouvelle image générée comme précédemment avec des filtres « blur », puis on sélectionne uniquement la partie centrale encadrée en noir dans l'image précédente pour obtenir le résultat montré en figure 4.19.

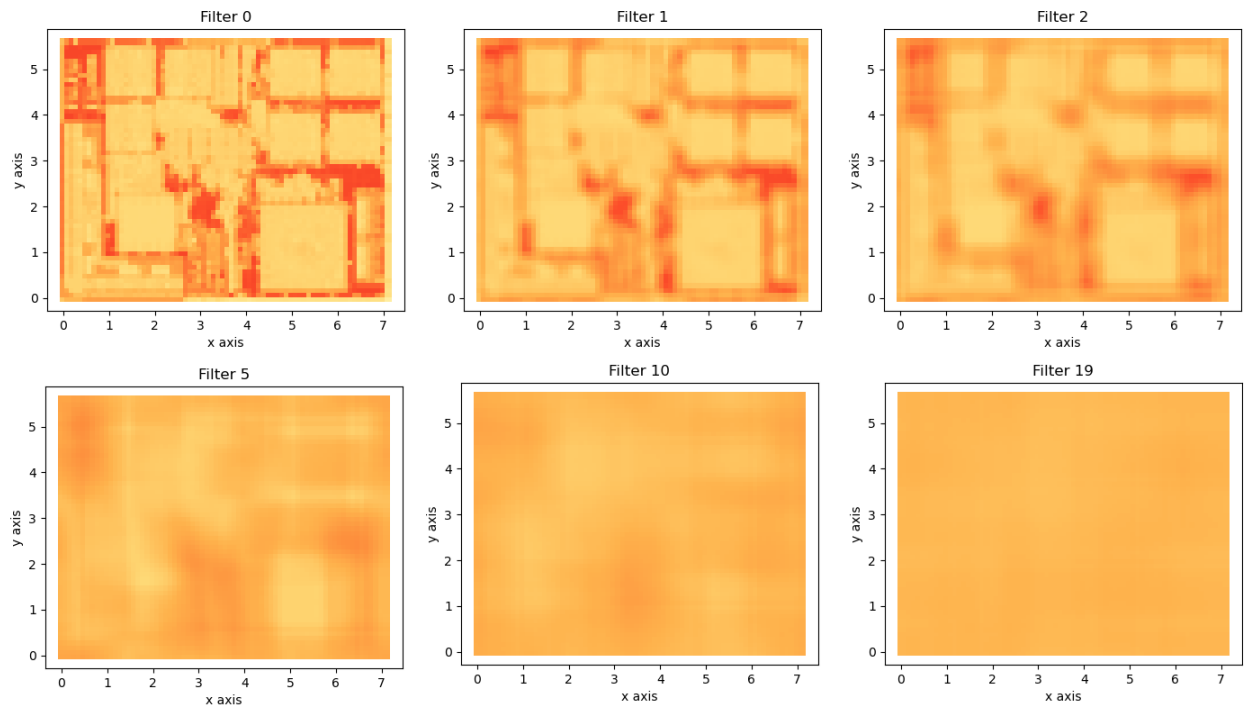


FIGURE 4.19 – Exemples de filtrages blur avec voisinage et un noyau de matrice carrée : 0, 1, 2, 5, 10, 19

Les modèles obtenus en mettant toutes les colonnes générées sont montrés en figure 4.20. Sur cette figure, chacun des 3 modèles est calibré sur la puce en question. Il s'agit donc ici d'une borne optimiste de la capacité de modélisation de cet ensemble de colonnes. Cette capacité est haute, étant donné les R^2 , mais le modèle est complexe car il possède 195 colonnes.

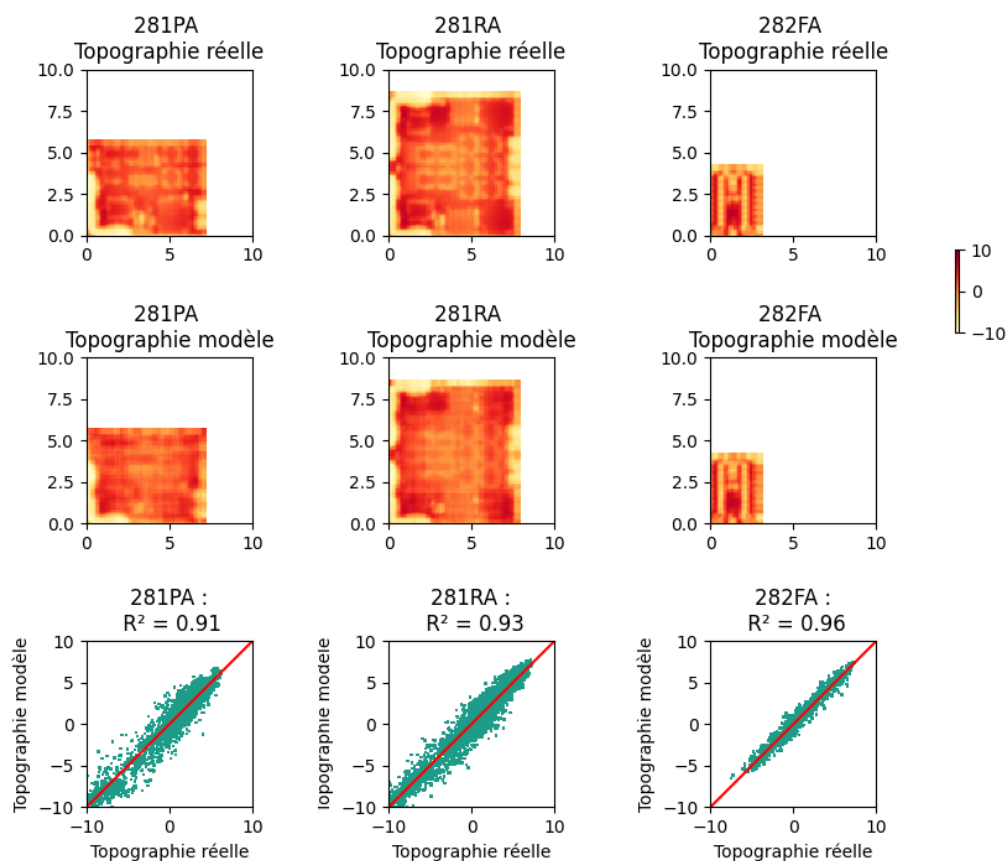


FIGURE 4.20 – Modèles avec les 195 colonnes en entrée

Produits de calibration	Produit d'inférence	Modèle plein		
		R ²	Nb vecteurs	PV diff
281PA	281PA	0,91	195	1,31
281RA	281RA	0,93	195	0,22
282FA	282FA	0,96	195	0,98

Nb Vecteurs	R ²	PV diff abs.
5	< 0.2	< 2
5 à 15	0.2-0.35	2 – 5
15 à 25	0.35-0.5	5 – 10
25 à 40	0.5-0.7	10 – 15
> 40	> 0,7	> 15

	Peak-to-Valley (nm)
281PA	17,39
281RA	18,28
282FA	13,97

FIGURE 4.21 – Tableau récapitulatif du modèle calibré sur chacune des puces

Ce modèle est extrêmement précis dans la modélisation : le R^2 est au-dessus de 0.9 pour les trois produits, et la différence entre Peak-to-Valley est inférieure à 2nm. En revanche, on peut difficilement interpréter le modèle car il y a 195 coefficients, et certains des vecteurs sont très ressemblants, ce qui génère du sur-apprentissage. On montre sur la figure 4.22 quelques coefficients associés.

Couche	Acti surf 0	Acti surf 1	Acti surf 2	Acti surf 3	Acti surf 4	Acti surf 5	Acti surf 6	Acti surf 7	Acti surf 8	Acti surf 9
Coefficient	-8,44	-15,94	-57,43	-70,76	57,85	41,95	-87,14	82,16	181,76	-111,22
Couche	Acti surf 10	Acti surf 11	Acti surf 12	Acti surf 13	Acti surf 14	Acti surf 15	Acti surf 16	Acti surf 17	Acti surf 18	Acti surf 19
Coefficient	44,41	147,67	-82,43	-28,37	126,19	59,92	-165,22	-3,01	10,32	-12,13

FIGURE 4.22 – Coefficients des portées de la couche Acti surfacique sans voisinage sur le modèle avec 195 colonnes

Les coefficients ont des valeurs qui varient énormément, même pour des couches avec des portées assez proches : en particulier les couches 8 et 9 où le coefficient est à 181.76 pour la portée 8 et -111.22 pour la portée 9. La plus haute valeur pour cette couche est 181.76, et la plus basse valeur est -165.22. On ne montre pas les 195 coefficients du modèle, mais ce qu'il se passe sur cette couche est parlant. On donne difficilement du sens à ces coefficients.

Ces variations sont la conséquence de trop grands degrés de liberté, et le modèle va chercher à tout prix à se rapprocher des données. Il s'agit d'un cas de sur-apprentissage linéaire, où des vecteurs proches vont avoir des valeurs totalement différentes pour créer des augmentations minimales du R^2 . On parle alors de multi-colinéarité.

Dans la prochaine partie, on cherche à sélectionner les colonnes pour à la fois éviter le sur-apprentissage, comme simplifier le modèle. L'idée est de trouver un compromis entre le premier modèle qui comporte 5 colonnes et celui qui en comporte 195.

4.3.5 Sélection des vecteurs entrées du modèle

Avec 195 vecteurs en entrée, la capacité de modélisation est forte. Mais un modèle possédant trop de degrés de liberté a un risque de sur-apprentissage élevé et une interprétation difficile. La figure 4.23 montre le sur-apprentissage dû à la complexité du modèle, avec une faible capacité de prédiction sur le produit 282FA en ayant une calibration sur le produit 281PA. On remarque que la prédiction du produit 281RA est satisfaisante, car les produits 281RA et 281PA ont les mêmes concepteurs et se ressemblent beaucoup.

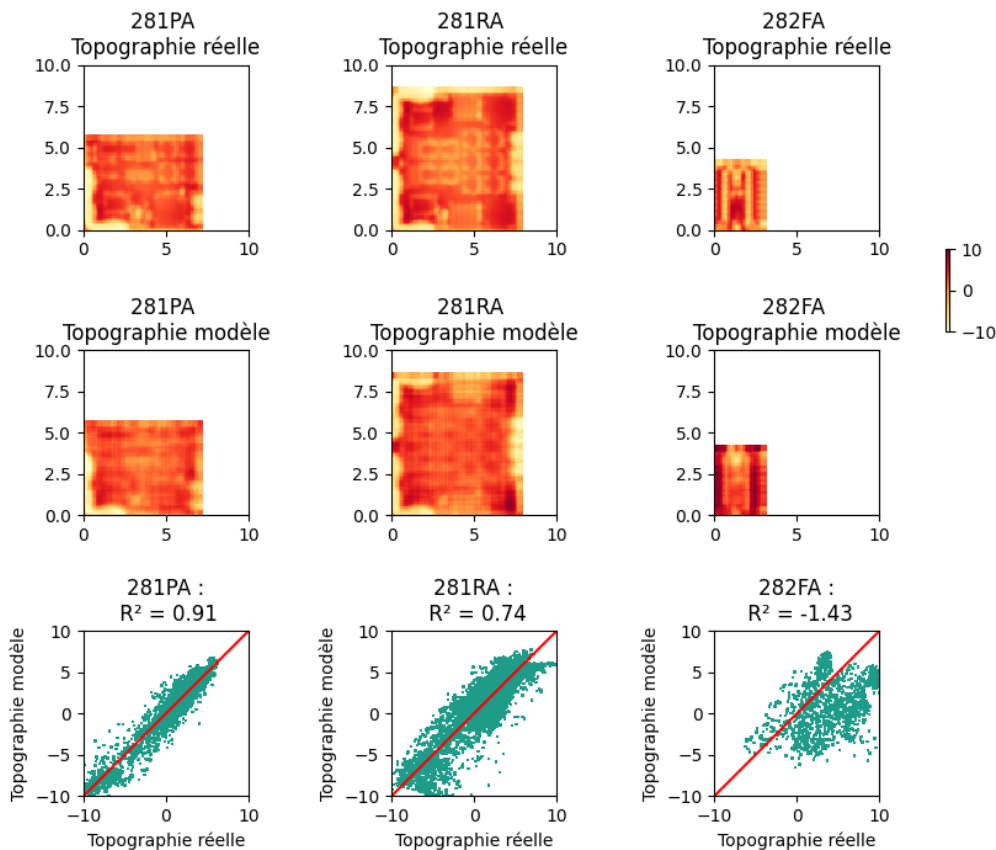


FIGURE 4.23 – Modèle avec les 195 colonnes en entrée calibré sur le produit 281PA

Produits de calibration	Produit d'inférence	Modèle plein		
		R ²	Nb vecteurs	PV diff
281PA	281PA	0,91	195	1,31
	281RA	0,74	195	4,09
	282FA	-1,43	195	7,6

Nb Vecteurs	R ²	PV diff abs.
5	< 0.2	< 2
5 à 15	0.2-0.35	2 – 5
15 à 25	0.35-0.5	5 – 10
25 à 40	0.5-0.7	10 – 15
> 40	> 0.7	> 15

	Peak-to-Valley (nm)
281PA	17,39
281RA	18,28
282FA	13,97

FIGURE 4.24 – Tableau récapitulatif du modèle calibré sur la puce 281PA

Afin d'améliorer la capacité de prédiction du modèle et pour le simplifier, on souhaite diminuer le nombre de colonnes en entrée. Pour ce faire, on peut utiliser diverses techniques. Dans les techniques automatiques, on peut par exemple filtrer les colonnes qui sont trop corrélées entre elles. On peut également utiliser un algorithme itératif qui enlève et rajoute des colonnes qui améliorent le plus le modèle. Autrement, on sélectionne les colonnes à la main pour des raisons métier.

Méthode a. Pour sélectionner les colonnes suivant leur corrélation, on souhaite qu'il n'existe pas de paires de deux colonnes dont la corrélation est supérieure à un seuil s :

1. On calcule la matrice de corrélation des colonnes : $M_{i,j} = Corr(v_i, v_j)$.
2. Pour tout i allant de 1 à n : Pour tout j allant de $i+1$ à n : Si $M_{i,j} > s$, alors : on supprime v_j des vecteurs d'entrée.

À titre d'illustration, on montre dans la figure 4.25 la matrice de corrélation des vecteurs avec le résultat de filtrage. Les cases en rouge sont celles concernées par le filtrage. Dans ce cas, l'une des deux valeurs est supprimée du modèle. Cette matrice est symétrique, car c'est une matrice de corrélation. La coloration est nécessairement symétrique, elle aussi.

	ACTISURF_filtered_0	ACTISURF_filtered_1	ACTISURF_filtered_2	ACTISURF_filtered_3	ACTISURF_filtered_4	ACTISURF_filtered_5	ACTISURF_filtered_6	ACTISURF_filtered_7	ACTISURF_filtered_8	ACTISURF_filtered_9	ACTISURF_filtered_10	ACTISURF_filtered_11
ACTISURF_filtered_0	1	0.81	0.69	0.56	0.45	0.39	0.36	0.36	0.37	0.37	0.37	0.35
ACTISURF_filtered_1	0.81	1	0.91	0.77	0.64	0.56	0.52	0.51	0.52	0.52	0.51	0.49
ACTISURF_filtered_2	0.69	0.91	1	0.93	0.81	0.68	0.72	0.68	0.67	0.65	0.64	0.62
ACTISURF_filtered_3	0.56	0.77	0.93	1	0.95	0.88	0.83	0.8	0.78	0.76	0.75	0.73
ACTISURF_filtered_4	0.45	0.64	0.81	0.95	1	0.93	0.97	0.93	0.89	0.88	0.84	0.82
ACTISURF_filtered_5	0.39	0.56	0.72	0.88	0.97	1	0.98	0.94	0.91	0.88	0.86	0.85
ACTISURF_filtered_6	0.36	0.52	0.68	0.83	0.93	0.98	1	0.98	0.95	0.92	0.9	0.89
ACTISURF_filtered_7	0.36	0.51	0.67	0.8	0.89	0.94	0.98	1	0.99	0.96	0.94	0.92
ACTISURF_filtered_8	0.37	0.52	0.66	0.78	0.85	0.91	0.95	0.99	1	0.99	0.97	0.95
ACTISURF_filtered_9	0.37	0.52	0.65	0.76	0.84	0.88	0.92	0.96	0.99	1	0.99	0.97
ACTISURF_filtered_10	0.37	0.51	0.64	0.75	0.82	0.86	0.9	0.94	0.97	0.99	1	0.99
ACTISURF_filtered_11	0.35	0.49	0.62	0.73	0.8	0.85	0.89	0.92	0.95	0.97	0.99	1
ACTISURF_filtered_12	0.34	0.48	0.61	0.72	0.79	0.83	0.87	0.9	0.93	0.95	0.98	0.99
ACTISURF_filtered_13	0.33	0.46	0.6	0.7	0.78	0.82	0.86	0.89	0.91	0.94	0.96	0.98
ACTISURF_filtered_14	0.32	0.46	0.58	0.69	0.76	0.8	0.84	0.87	0.9	0.92	0.94	0.97
ACTISURF_filtered_15	0.32	0.45	0.57	0.68	0.74	0.79	0.82	0.86	0.89	0.91	0.93	0.95
ACTISURF_filtered_16	0.31	0.44	0.56	0.66	0.73	0.77	0.81	0.84	0.87	0.89	0.91	0.93
ACTISURF_filtered_17	0.3	0.42	0.55	0.65	0.71	0.75	0.79	0.82	0.85	0.87	0.9	0.92
ACTISURF_filtered_18	0.29	0.41	0.53	0.63	0.69	0.73	0.77	0.81	0.83	0.86	0.88	0.9
ACTISURF_filtered_19	0.28	0.4	0.52	0.61	0.67	0.71	0.75	0.79	0.81	0.84	0.87	0.89
GATESURF_filtered_0	0.3	0.29	0.2	0.11	0.05	0.02	0	0	0.01	0.02	0.01	0.01
GATESURF_filtered_1	0.27	0.3	0.23	0.15	0.08	0.04	0.02	0.02	0.03	0.03	0.03	0.03
GATESURF_filtered_2	0.19	0.23	0.22	0.17	0.1	0.06	0.04	0.03	0.03	0.03	0.03	0.04
GATESURF_filtered_3	0.1	0.14	0.17	0.16	0.11	0.08	0.05	0.04	0.03	0.03	0.04	0.05
GATESURF_filtered_4	0.06	0.08	0.11	0.13	0.12	0.09	0.07	0.06	0.04	0.04	0.04	0.06

FIGURE 4.25 – Partie haut-gauche de la matrice de corrélation avec le résultat de filtrage pour un seuil de 0.8 sur le produit 281PA

À la suite de l'algorithme avec le seuil de corrélation de 0.8 sur le produit 281PA, il reste alors 22 colonnes en entrée. Chacune de ces colonnes ont des corrélations mutuelles inférieures à 0.8. Sur le produit 281RA, il reste 17 colonnes avec cette heuristique, et il en reste 46 sur le produit 282FA. Cette heuristique est dépendante des données sur lesquelles on effectue le filtrage.

Cette méthode évite de la redondance dans le modèle en enlevant des colonnes. Étant donné que le modèle est linéaire, il est pertinent d'enlever les vecteurs qui ont des corrélations linéaires trop hautes entre eux, puisque la majorité de l'information de l'un est contenue dans l'autre.

Méthode b. Pour simplifier la compréhension et rendre le modèle indépendant du jeu de données, on peut sélectionner à la main les portées de filtres. Le modèle sera plus compréhensible, car on connaîtra exactement les vecteurs qui seront dans celui-ci par choix.

Chacune des solutions **a.** **b.** possède des avantages et inconvénients. Dans le cas **a.**, on a une confiance en la non-redondance linéaire des vecteurs présents dans le modèle, mais ceux-ci dépendent de l'ensemble de données sur lequel on calibre le modèle. Dans la pratique, on souhaite toujours avoir un modèle compréhensible. Par compréhensible, on entend qu'il ne doit pas y avoir un trop grand nombre de colonnes, mais aussi qu'on sait ce que l'on utilise pour la modélisation.

Dans la méthode **b.**, on a une connaissance des vecteurs présents, et ceux-ci sont indépendants du modèle. A priori, la méthode **b** est la plus simple et pertinente à mettre en œuvre, si les critères évoqués dans la partie concernant l'évaluation la valident.

On montre sur la figure 4.26 différents modèles proposés pour chacun des trois produits calibrés sur eux-mêmes. On présente ces cinq modèles :

- Le modèle 1 est le modèle sans transformer les données d'entrées. Le modèle plein est le modèle contenant les 195 colonnes générées précédemment.
- Les modèles 2 et 3 sont des modèles basés sur le filtrage des colonnes par corrélation. Le modèle 2 filtre les colonnes ayant une corrélation supérieure à 0.8, et le modèle 3 avec une corrélation supérieure à 0.9.
- Les modèles 4 et 5 sont basés sur des choix métiers traduisant la volonté de regarder différentes longueurs de portées : 0, 1, 5, 15. Le modèle 4 **ne prend pas** en compte les puces voisines, et le modèle 5 prend en compte **uniquement** les puces voisines.

Sur cette figure 4.26, il s'agit en quelque sorte d'une borne optimiste de ces modèles pour chacun des produits. Pour chaque modèle, on calibre celui-ci sur les données complètes de chaque puce. En ce sens, le critère du R^2 est maximisé pour chacun des modèles.

Produits de calibration	Produit d'inférence	Modèle 1			Modèle plein		
		R ²	Nb vecteurs	PV diff	R ²	Nb vecteurs	PV diff
281PA	281PA	0,6	5	-2,39	0,91	195	1,31
281RA	281RA	0,63	5	-2,61	0,93	195	0,22
282FA	282FA	0,51	5	1,52	0,96	195	0,98

Produits de calibration	Produit d'inférence	Modèle 2			Modèle 3		
		R ²	Nb vecteurs	PV diff	R ²	Nb vecteurs	PV diff
281PA	281PA	0,8	22	-0,2	0,86	36	0,65
281RA	281RA	0,87	17	1,84	0,9	30	-0,25
282FA	282FA	0,89	46	1,01	0,92	78	1,03

Produits de calibration	Produit d'inférence	Modèle 4			Modèle 5		
		R ²	Nb vecteurs	PV diff	R ²	Nb vecteurs	PV diff
281PA	281PA	0,78	20	-0,34	0,79	20	-0,43
281RA	281RA	0,84	20	0,9	0,88	20	-0,22
282FA	282FA	0,79	20	4,87	0,8	20	4,72

Modèle 1	Modèle linéaire sans filtre	Nb Vecteurs	R ²	PV diff abs.	Peak-to-Valley (nm)	
Modèle 2	Modèle linéaire avec filtre seuil de corrélation 0.8	5	< 0.2	< 2	281PA	17,39
Modèle 3	Modèle linéaire avec filtre seuil de corrélation 0.9	5 à 15	0.2-0.35	2 - 5	281RA	18,28
Modèle 4	Modèle linéaire avec filtres sans puces voisines de portés [0,1,5,15]	15 à 25	0.35-0.5	5 - 10	282FA	13,97
Modèle 5	Modèle linéaire avec filtres avec puces voisines de portés [0,1,5,15]	25 à 40	0.5-0.7	10 - 15		
Modèle plein	Modèle linéaire avec la totalité des 195 vecteurs	> 40	> 0.7	> 15		

FIGURE 4.26 – Tableau récapitulatif des modèles sur chacun des trois produits

Un code couleur a été donné sur ce tableau pour en simplifier la lisibilité. Ce code couleur est détaillé en bas à droite de la figure. Les colonnes « PV diff » correspondent à la différence entre le Peak-to-Valley du modèle et le Peak-to-Valley de la topographie réelle. Plus celui-ci est proche de zéro, mieux cela est. Les coloris sont en fonction de la valeur absolue de cette différence.

Sur ce tableau, les R^2 respectifs des puces sont tous supérieurs à 0.5, et montent jusqu'à 0.9. Comparativement au modèle plein où l'on a 195 vecteurs, on diminue fortement la taille du modèle sans perdre énormément en capacité de modélisation dans chacune des méthodes de filtrage. Les meilleures capacités de modélisation sont obtenues sur les modèles basés sur les filtrages de corrélation. Cela s'explique par l'enlèvement de redondance linéaire grâce à la méthode de sélection. Cette méthode est particulièrement adaptée pour un modèle linéaire.

En ce qui concerne les méthodes 4 et 5, celles-ci rendent compte d'un plus gros écart de Peak-to-Valley avec la vraie valeur pour le produit 282FA, malgré un R^2 proche des autres produits.

La question restante est de savoir si ces modèles s'appliquent ou non d'une puce sur l'autre, et dans quelles conditions cela est ou n'est pas le cas. On souhaite avoir confiance en le modèle, et des critères a priori sont nécessaire pour avoir cette confiance. C'est l'objet de la section suivante.

4.4 Des modèles pour des puces

4.4.1 Confiance et robustesse

Lorsqu'on fait un modèle, on sépare habituellement les données en deux ensembles pour faire de la validation croisée : les données d'entraînement et les données de validation. On définit un ensemble pour l'entraînement du modèle, et un ensemble distinct pour la validation. Séparer ces ensembles permet d'évaluer la capacité de prédiction sur des nouvelles données. On définit la robustesse comme la stabilité de performance d'un système. Ici, la performance est la qualité de prédiction suivant les indicateurs choisis : le R^2 et le Peak-to-Valley.

Il y a plusieurs enjeux. D'une part les enjeux d'un modèle prédictif efficace, avec des méthodes classiques de validation croisée, et d'autre part les enjeux métiers d'avoir un indicateur de confiance sur les prédictions concernant de nouveaux produits arrivants. Nous disposons des données de trois produits différents. On peut partitionner les données pour des enjeux métier, où chaque produit est un ensemble de données d'entraînement ou un ensemble de données de validation.

Ou bien, on peut considérer ces trois produits comme un unique ensemble de points, et utiliser des méthodes de validation et d'entraînement classiques, par exemple en séparant les données de façon aléatoire.

Ces deux façons de séparer les données font sens pour des enjeux différents. Les enjeux qui nous intéressent ici sont des enjeux de confiance concernant la prédiction des nouvelles puces. Nous considérons majoritairement la première façon de séparer les données dans cette section.

Cette section commence par vérifier qu'un modèle linéaire est réalisable pour une puce en satisfaisant les exigences de validation croisée ainsi que des enjeux métiers avec des puces virtuelles. Ensuite, on vérifie que des modèles calibrés sur une ou deux puces permettent de prédire la topographie d'autres puces. Identifier des critères a priori pour évaluer la confiance sur la réalisation d'un modèle est un enjeu clé, qui reste à explorer dans notre contexte.

4.4.2 Modélisation et inférence sur une puce

On se place ici dans le contexte où l'on ne dispose que d'une puce, et de voir si calibrer un modèle sur une portion de la puce permet d'avoir un modèle efficace pour le reste de la puce. L'enjeu est d'évaluer la pertinence de la modélisation avec des données limitées, à savoir en ne possédant que les données d'une seule puce.

Historiquement, dans le déroulement de la thèse, on ne disposait que des données de la puce 281PA. Les problématiques d'accès de l'existence des données doivent être prises en compte lorsqu'on génère un modèle. On se pose des questions quant à la capacité de prédiction du modèle choisi. Comme évoqué précédemment, on peut faire de la validation croisée afin de voir si la modélisation est possible. Cependant, on ne peut pas être certain de la prédiction sur une autre puce, en particulier si celle-ci a une configuration totalement différente, comme nous le montrons plus tard dans la thèse.

On peut prendre des portions spatiales de la puce, en la découpant en quarts. Cela permet de considérer chaque quart comme étant une puce fictive, et ainsi vérifier la capacité de prédiction du modèle sur les quarts restants. La figure 4.27 montre des modèles, avec la zone en surbrillance noire comme zone de calibration, sans donner des critères d'évaluation. Ces critères sont développés dans le tableau 4.30.

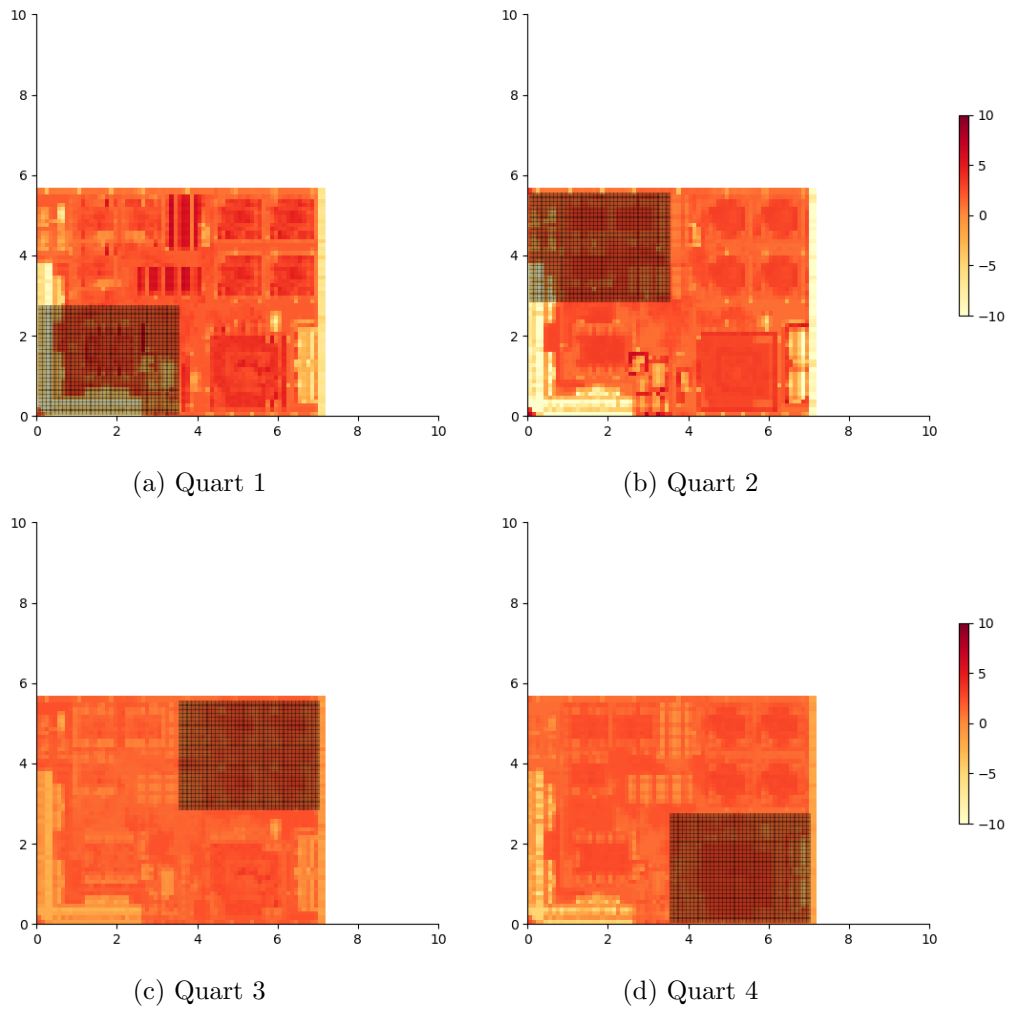


FIGURE 4.27 – Prédications des modèles linéaires, sans transformation des données, calibrés sur les quarts de puces (en noir)

Dans le même esprit, on prend un échantillonnage prenant en calibration 25% de la puce aléatoirement. On montre cela sur la figure 4.28. L'échantillonnage des puces avec différents pourcentages est développé en Annexe C.

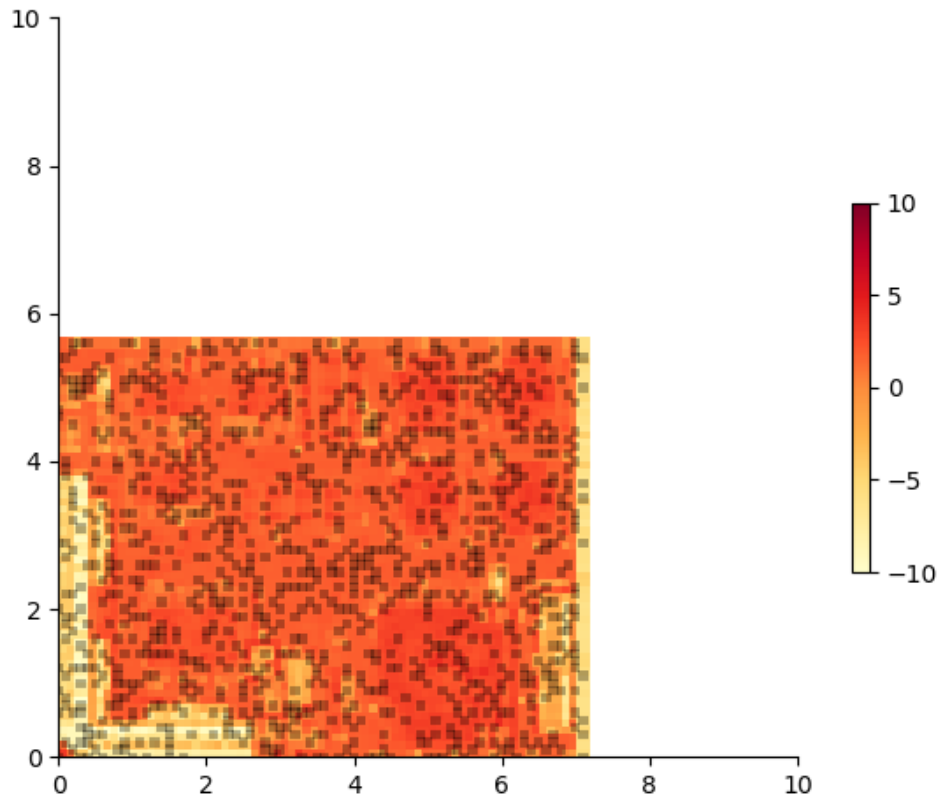


FIGURE 4.28 – Modèle linéaire, sans transformation des données, calibré sur un échantillon aléatoire de 25%

Ces mêmes sélections prises, on peut effectuer la même modélisation avec les types de modèle avec filtre spatial sur les entrées. On conserve les modes de sélection évoqués précédemment, et on les nommera « Quart 1 », « Quart 2 », « Quart 3 », « Quart 4 », « Aléatoire ». Le tableau récapitulatif suivant présente les résultats des prédictions en fonction de la zone de calibration sur chacun des trois produits. On montre, pour rappel, les topographies des trois puces avec leur Peak-to-Valley respectif sur la figure 4.29.

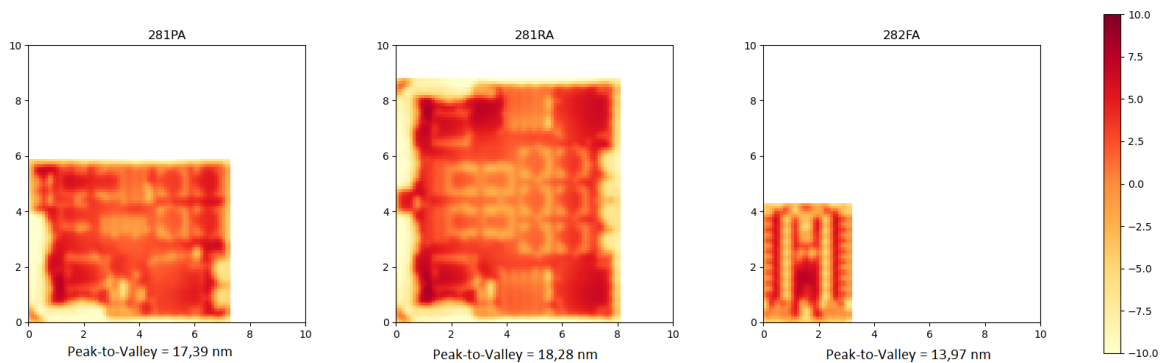


FIGURE 4.29 – Topographie des puces moyennes avec Peak-to-Valley

	Produits de calibration	Produit d'inférence	Modèle 1			Modèle plein		
			R ²	Nb vecteurs	PV diff	R ²	Nb vecteurs	PV diff
	281PA Q1	281PA	0,5	5	3,12	-23,87	195	95,5
	281PA Q2	281PA	0,41	5	6,74	-11,5	195	62,04
	281PA Q3	281PA	0,33	5	-6,16	-6,53	195	63,38
	281PA Q4	281PA	0,47	5	-6,86	-0,9	195	27,73
	281RA Q1	281RA	0,56	5	-1,62	-0,59	195	14,95
	281RA Q2	281RA	0,47	5	2,03	-0,37	195	13,87
	281RA Q3	281RA	0,54	5	-3,06	-2,91	195	31,19
	281RA Q4	281RA	0,5	5	-4,38	0,02	195	16,83
	282FA Q1	282FA	0,42	5	0,02	-92,13	195	145,7
	282FA Q2	282FA	0,38	5	-4,17	-44,48	195	91,23
	282FA Q3	282FA	0,42	5	-2,61	-42,78	195	122,25
	282FA Q4	282FA	0,42	5	4,17	-109,55	195	99,4

	Produits de calibration	Produit d'inférence	Modèle 2			Modèle 3		
			R ²	Nb vecteurs	PV diff	R ²	Nb vecteurs	PV diff
	281PA Q1	281PA	0,53	17	2,29	0,53	27	1,61
	281PA Q2	281PA	0,59	13	1,75	0,66	22	3,04
	281PA Q3	281PA	0,29	12	-6,61	0,7	22	1,31
	281PA Q4	281PA	0,66	12	0,87	0,37	21	5,2
	281RA Q1	281RA	0,56	12	0,26	0,36	23	2,84
	281RA Q2	281RA	0,52	11	-1,45	0,74	19	2,73
	281RA Q3	281RA	0,03	9	-8,2	0,68	17	3,04
	281RA Q4	281RA	0,08	9	-6,49	0,69	16	2,59
	282FA Q1	282FA	-1,59	34	10,84	-11,8	65	32,06
	282FA Q2	282FA	0,53	25	4,03	-0,63	62	10,41
	282FA Q3	282FA	0,52	25	3,5	-8,45	60	18,59
	282FA Q4	282FA	0,62	24	1,71	-7,28	59	24,89

	Produits de calibration	Produit d'inférence	Modèle 4			Modèle 5		
			R ²	Nb vecteurs	PV diff	R ²	Nb vecteurs	PV diff
	281PA Q1	281PA	-0,5	20	3,84	0,64	20	-0,25
	281PA Q2	281PA	0,38	20	9,24	0,42	20	4,95
	281PA Q3	281PA	-2,3	20	8,97	0,59	20	-2,11
	281PA Q4	281PA	0,41	20	13,97	0,65	20	5,1
	281RA Q1	281RA	0,47	20	2,49	0,73	20	-0,5
	281RA Q2	281RA	0,73	20	0,83	0,84	20	-0,46
	281RA Q3	281RA	0,71	20	11,81	0,79	20	7,72
	281RA Q4	281RA	-0,04	20	16,86	0,75	20	9,5
	282FA Q1	282FA	0,3	20	4,15	0,39	20	3,5
	282FA Q2	282FA	0,44	20	0,88	-0,04	20	4,39
	282FA Q3	282FA	0,54	20	2,51	0,17	20	2,65
	282FA Q4	282FA	0,64	20	4,89	0,1	20	6,33

Produits de calibration	Produit d'inférence	Modèle 1			Modèle plein		
		R ²	Nb vecteurs	PV diff	R ²	Nb vecteurs	PV diff
281PA Quart Aléatoire	281PA	0,6	5	-2,48	0,89	195	2,56
281RA Quart Aléatoire	281RA	0,58	5	-2,73	0,92	195	1,4
282FA Quart Aléatoire	282FA	0,46	5	2,58	0,82	195	5,92

Produits de calibration	Produit d'inférence	Modèle 2			Modèle 3		
		R ²	Nb vecteurs	PV diff	R ²	Nb vecteurs	PV diff
281PA Quart Aléatoire	281PA	0,8	22	-0,14	0,85	37	0,53
281RA Quart Aléatoire	281RA	0,87	17	1,9	0,85	32	-0,54
282FA Quart Aléatoire	282FA	0,87	46	0,96	0,92	76	1,14

Produits de calibration	Produit d'inférence	Modèle 4			Modèle 5		
		R ²	Nb vecteurs	PV diff	R ²	Nb vecteurs	PV diff
281PA Quart Aléatoire	281PA	0,77	20	-0,21	0,78	20	-0,54
281RA Quart Aléatoire	281RA	0,84	20	0,55	0,87	20	-0,17
282FA Quart Aléatoire	282FA	0,78	20	4,77	0,78	20	5,5

Modèle 1	Modèle linéaire sans filtre
Modèle 2	Modèle linéaire avec filtre seuil de corrélation 0.8
Modèle 3	Modèle linéaire avec filtre seuil de corrélation 0.9
Modèle 4	Modèle linéaire avec filtres sans puces voisines de portés [0,1,5,15]
Modèle 5	Modèle linéaire avec filtres avec puces voisines de portés [0,1,5,15]
Modèle plein	Modèle linéaire avec la totalité des 195 vecteurs

Nb Vecteurs	R ²	PV diff abs.
5	< 0,2	< 2
5 à 15	0,2-0,35	2 - 5
15 à 25	0,35-0,5	5 - 10
25 à 40	0,5-0,7	10 - 15
> 40	> 0,7	> 15

Peak-to-Valley (nm)	
281PA	17,39
281RA	18,28
282FA	13,97

FIGURE 4.30 – Tableau récapitulatif des modèles calibrés sur des quarts de puces

1. Lecture du tableau

Le but de séparer les ensembles de données est de questionner la robustesse des modèles. La séparation en quarts sert à créer des puces artificielles dont la cohérence spatiale est maintenue. La séparation aléatoire correspond à la vision « classique » de la validation croisée, où l'on sélectionne un échantillon aléatoire de la population pour calibrer un modèle.

2. Interprétation des sélections

Le modèle 1 est celui qui présente le plus de robustesse. Ses résultats sont homogènes selon le critère du R^2 , et les différences absolues de Peak-to-Valley ne dépassent jamais 7, et sont globalement en dessous de 5. Il est parallèlement celui qui va le moins haut en terme de R^2 , ainsi que lui qui descend le moins bas. De même pour le Peak-to-Valley, il est celui qui monte le moins haut de tous les modèles en valeur absolue. En ce sens, il est celui dont la robustesse est la plus forte, tout en étant celui qui peut être le moins précis. Cela s'explique par le peu de degrés de liberté, ce qui limite sa capacité de modélisation. Ce modèle ne comporte que 5 colonnes. Il sera aisé d'implémenter un tel modèle et d'en interpréter le contenu, en plus d'en obtenir une information robuste. On peut vraisemblablement conclure qu'une nouvelle puce sera modélisée avec un R^2 situé entre 0.4 et 0.6 grâce à ce tableau. Cette hypothèse sera étudiée dans la partie suivante, lorsqu'on utilisera les trois puces pour créer des modèles prédictifs.

Au contraire, le modèle plein possède le plus de colonnes donc de degrés de liberté et est celui avec les moins bons résultats. Les résultats du modèle plein montrent un sur-apprentissage sur le quart et une mauvaise inférence sur la totalité de la puce. Sa capacité de prédiction est pauvre, comme nous l'avons déjà montré précédemment sur la figure 4.24. Cette observation, combinée avec le paragraphe précédent, nous donne confiance dans la démarche d'éliminer des vecteurs du modèle plein pour trouver un compromis robuste et le plus performant possible avec un nombre de vecteurs le plus bas possible.

Les modèles 2, 3, 4 et 5 ont des résultats variables, dépendant du quart sur lequel a été calibré le modèle, mais aussi du produit concerné. On discute certaines lignes permettent cependant d'identifier des limites de ces quatre modèles.

Sur les quarts du produit 281PA, on voit que le Modèle 5 est le plus satisfaisant sur les critères du R^2 et du Peak-to-Valley, et que le Modèle 4 l'est beaucoup moins sur ces critères. On ne peut conclure quant au fait qu'un des deux modèles soit meilleur de façon générale, car on observe l'inverse sur le produit 282FA.

De même, sur les quarts du produit 281PA, le modèle 3 donne de meilleurs résultats que le modèle 2, on observe encore une fois l'inverse sur le produit 282FA. L'avantage du modèle 2 est d'avoir bien moins de vecteurs que le modèle 3.

Sur les quarts du modèle 281RA, les résultats sont également imprévisibles en fonction du quart concerné. On peut penser grâce aux R^2 du modèle 2 que les quarts 1 et 2 sont des bons représentants de la puce dans sa totalité comparativement aux quarts 3 et 4, mais avec les R^2 du modèle 3, on montre que ce n'est pas le cas pour le quart 1. Les R^2 du modèle 4 semblent indiquer que les quarts 2 et 3 sont des bons représentants, et ceux du modèle 5 montrent que les quatre quarts représentent la puce. Le Peak-to-Valley est cependant consistant dans sa démonstration que les quarts 1 et 2 sont de bons représentants de la puce complète dans les 6 modèles (même dans le modèle plein, les résultats sont moins catastrophiques). En regardant la topographie de la puce, on voit que les quarts 1 et 2 sont ceux qui comportent des zones basses et hautes, et couvrent donc le plus grand intervalle de nano-topographie.

En ce qui concerne la comparaison des modèles, il est difficile d'en extraire un meilleur que les autres, ou même d'en éliminer un de la liste. On voit clairement qu'une sélection aléatoire rend de bien meilleurs R^2 de modélisation, avec des résultats très similaires à une sélection sur la totalité de la puce (figure 4.30). Cela se vérifie sur tous les produits. Cela s'explique par le fait qu'un échantillon est le mieux représenté dans la totalité de ses variables lorsqu'on ne

considère aucune variable de sélection a priori. Aussi, la capacité de modélisation plus haute que le modèle 1 s'explique par le nombre de colonnes qui augmente. On montre en annexe C une analyse quantifiant l'évolution du R^2 en fonction de la proportion de sélection de l'échantillon.

La capacité de prédiction de ces modèles n'est pas garantie, mais ils possèdent une excellente capacité de modélisation sur des échantillons aléatoires. Ce qui est mis en lumière dans ces tableaux est le besoin de données représentatives de ce que l'on souhaite modéliser afin de profiter de cette forte capacité de modélisation pour en faire un modèle prédictif efficace. On tire des conclusions sur la pertinence des modèles lorsqu'on utilise plus de données, ou en tout cas les « bonnes » données adaptées à la prédiction que l'on souhaite faire.

3. Mise en regard des indicateurs

Le graphique 4.31 montre le Peak-to-Valley en fonction du R^2 pour tous les modèles. On regroupe les R^2 inférieurs à -1 pour leur donner la valeur -1 car on considère qu'un R^2 à -1 est déjà une très mauvaise valeur. Cela permet de rendre le graphique plus lisible et éviter de montrer un graphique avec les trois quarts remplis de vide, nous faisons ce choix.

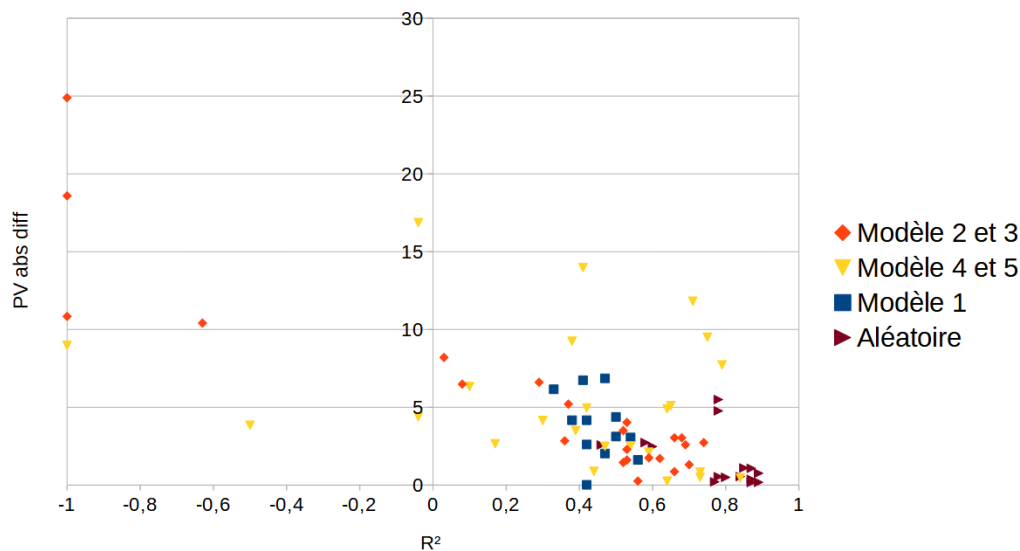


FIGURE 4.31 – Peak-to-Valley en fonction du R^2 pour le tableau 4.30

On a décidé de considérer les échantillons aléatoires comme à part, car ils ne représentent pas de « puces » à proprement parler étant donné que les points sont éparses, comme montré dans la figure 4.28. Les points des échantillons Aléatoires sont situés le plus en bas à droite du graphique, soit les meilleurs résultats.

En ce qui concerne les points associés aux quarts, on observe que les modèles 2 et 3 sont ceux ayant les variations les plus hautes. Cela s'explique par le fait que le mode de sélection des colonnes fait du sur-apprentissage, et l'inférence sur le reste de la puce ne fonctionne pas bien.

La tendance générale est que plus le R^2 est élevé, plus le Peak-to-Valley a globalement des valeurs basses. Cela n'est pas toujours le cas, ce qui rend ces indicateurs complémentaires. Il est rassurant d'y voir une cohérence globale, et utile de conserver ces deux indicateurs en complément.

4. Conclusion

Le questionnement traité dans cette sous-partie est de déterminer la confiance à avoir en la capacité prédictive d'un modèle en ne possédant les données que d'une seule puce. Ce qui a été montré, est que les données de calibration du modèle jouent un rôle significatif pour la performance de la prédiction. On a montré qu'une sélection aléatoire représentait bien la puce,

et générant un modèle proche du modèle calibré sur la puce entière. On a montré qu'en changeant les zones de sélections, on obtenait des résultats rarement satisfaisants, et souvent mauvais. On a montré que le modèle 1 prenant en entrée les données brutes était robuste, et constituait un choix de modélisation de confiance, bien qu'imprécis.

Les données de calibration étant identifiées comme le point de travail le plus important pour ces modèles, la sous-section suivante aura pour but de répondre aux mêmes questions que cette partie, mais avec des enjeux métier concrets. On répond aux questions des capacités de prédiction des 6 modèles choisis en disposant de plus de données, à savoir les données de 3 puces. La question est de savoir si on peut prédire la topographie d'une puce en utilisant les données des autres puces.

4.4.3 Modélisation et inférence sur trois puces

Dans cette sous-partie, on dispose des données de trois puces, et on s'autorise à utiliser les données de toutes les puces dans un même problème. On pose des questions de construction de modèles, de capacité de prédiction, et d'indices de confiance sur la prédiction a priori.

1. Modèle complet

Pour commencer, et afin d'avoir un ordre d'idée de ce qu'il serait possible d'atteindre, on se propose d'utiliser la totalité des données en tant que données de calibration. On montre les trois puces dans la figure 4.32, mises en sur-brillance noire pour montrer que celles-ci ont été prises comme données de calibration. Les résultats sont montrés dans le tableau 4.33.

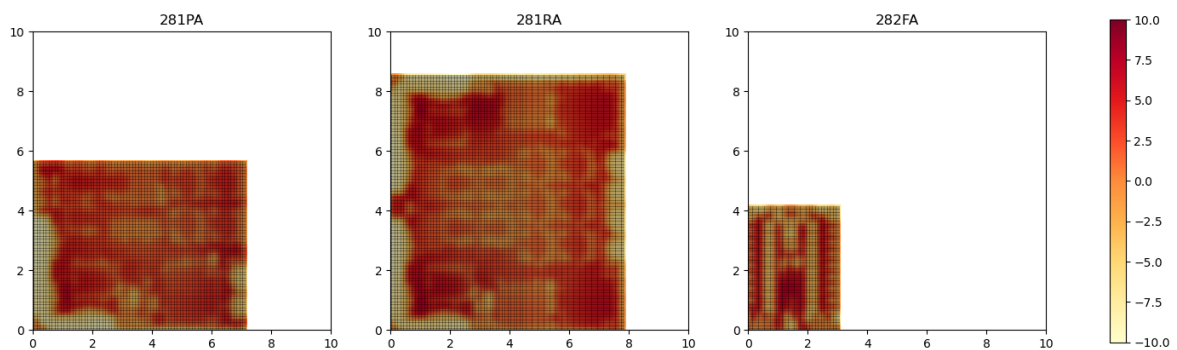


FIGURE 4.32 – Sélection sur les 3 produits

Produits de calibration	Produit d'inférence	Modèle 1			Modèle plein		
		R ²	Nb vecteurs	PV diff	R ²	Nb vecteurs	PV diff
281PA,281RA,282FA	281PA	0,59	5	-3	0,87	195	2,3
	281RA	0,63	5	-3,39	0,91	195	-0,11
	282FA	0,33	5	-2,74	0,77	195	-1,61

Produits de calibration	Produit d'inférence	Modèle 2			Modèle 3		
		R ²	Nb vecteurs	PV diff	R ²	Nb vecteurs	PV diff
281PA,281RA,282FA	281PA	0,78	16	1,04	0,81	29	1,03
	281RA	0,86	16	0,24	0,88	29	-0,4
	282FA	0,58	16	-5,84	0,66	29	-4,33

Produits de calibration	Produit d'inférence	Modèle 4			Modèle 5		
		R ²	Nb vecteurs	PV diff	R ²	Nb vecteurs	PV diff
281PA,281RA,282FA	281PA	0,73	20	1,14	0,78	20	0,58
	281RA	0,82	20	0,56	0,86	20	0,57
	282FA	0,63	20	-1,37	0,66	20	-1,53

Modèle 1	Modèle linéaire sans filtre	Nb Vecteurs	R ²	PV diff abs.
Modèle 2	Modèle linéaire avec filtre seuil de corrélation 0.8	5	< 0.2	< 2
Modèle 3	Modèle linéaire avec filtre seuil de corrélation 0.9	5 à 15	0.2-0.35	2 – 5
Modèle 4	Modèle linéaire avec filtres sans puces voisines de portés [0,1,5,15]	15 à 25	0.35-0.5	5 – 10
Modèle 5	Modèle linéaire avec filtres avec puces voisines de portés [0,1,5,15]	25 à 40	0.5-0.7	10 – 15
Modèle plein	Modèle linéaire avec la totalité des 195 vecteurs	> 40	> 0.7	> 15

	Peak-to-Valley (nm)
281PA	17,39
281RA	18,28
282FA	13,97

FIGURE 4.33 – Tableau récapitulatif des modèles calibrés sur les trois produits

Ce tableau est une synthèse de la modélisation qui prend en calibration tous les points. On constate que la plupart des R^2 sont supérieurs ou égaux à 0.6, ce qui signifie qu'il existe des modèles efficaces sur les trois puces. Chaque colonne représente un même modèle.

On constate que les modèles sont moins performants sur le produit 282FA que sur les deux autres produits. Cela peut être pour différentes raisons. La première serait le nombre de points. En effet, comme on peut le constater sur la figure 4.32, le produit 282FA est bien plus petit, et par conséquent pèse bien moins lourd dans la calibration du modèle linéaire. La contribution des deux autres produits est significativement plus importante, et ils sont donc privilégiés pour la modélisation. Pour vérifier cette hypothèse, on échantillonne les produits 281PA et 281RA pour faire en sorte que la modélisation se calibre sur le même nombre de points que sur le produit 282FA. On a donc une calibration sur la concaténation de 1302 points de chacun des trois produits. La sélection des points des produits 281PA et 282RA est aléatoire. On montre le résultat sur le tableau récapitulatif 4.34.

Produits de calibration	Produit d'inférence	Modèle 1			Modèle plein		
		R ²	Nb vecteurs	PV diff	R ²	Nb vecteurs	PV diff
281PA,281RA,282FA	281PA	0,58	5	-3,76	0,85	195	2,81
	281RA	0,61	5	-4,25	0,88	195	0,29
	282FA	0,39	5	-3,3	0,84	195	-0,47

Produits de calibration	Produit d'inférence	Modèle 2			Modèle 3		
		R ²	Nb vecteurs	PV diff	R ²	Nb vecteurs	PV diff
281PA,281RA,282FA	281PA	0,79	18	0,52	0,81	30	0,89
	281RA	0,87	18	-0,99	0,88	30	-0,89
	282FA	0,7	18	-4,4	0,73	30	-1,96

Produits de calibration	Produit d'inférence	Modèle 4			Modèle 5		
		R ²	Nb vecteurs	PV diff	R ²	Nb vecteurs	PV diff
281PA,281RA,282FA	281PA	0,73	20	2,27	0,77	20	0,94
	281RA	0,8	20	3,02	0,85	20	0,52
	282FA	0,7	20	-0,12	0,7	20	-0,41

Modèle 1	Modèle linéaire sans filtre	Nb Vecteurs	R ²	PV diff abs.
Modèle 2	Modèle linéaire avec filtre seuil de corrélation 0.8	5	< 0.2	< 2
Modèle 3	Modèle linéaire avec filtre seuil de corrélation 0.9	5 à 15	0.2-0.35	2 – 5
Modèle 4	Modèle linéaire avec filtres sans puces voisines de portés [0,1,5,15]	15 à 25	0.35-0.5	5 – 10
Modèle 5	Modèle linéaire avec filtres avec puces voisines de portés [0,1,5,15]	25 à 40	0.5-0.7	10 – 15
Modèle plein	Modèle linéaire avec la totalité des 195 vecteurs	> 40	> 0.7	> 15

	Peak-to-Valley (nm)
281PA	17,39
281RA	18,28
282FA	13,97

FIGURE 4.34 – Tableau récapitulatif des modèles calibrés sur les trois produits avec le même nombre de points sur chaque

La performance est améliorée sur le produit 282FA, bien qu'on constate toujours une

différence de performance entre les produits. Le R^2 est amélioré sur toutes les prédictions du produit 282FA, et la différence du Peak-to-Valley est diminuée sur toutes les prédictions, sauf sur le premier modèle.

Parallèlement, on constate une augmentation de la différence du Peak-to-Valley sur les deux autres produits, ainsi qu'une diminution mineure du R^2 . La différence du nombre de points engendre en effet un biais sur le modèle en faveur des puces les plus grosses, car la régression linéaire les privilégie de par sa nature. Ce biais est cependant faible, comme nous venons de le montrer, et la différence entre les produits n'est pas uniquement liée à cela.

Une autre raison de la différence est, comme évoqué précédemment, une différence de conception et d'agencement des structures. Cette différence est une difficulté pour la prédiction lorsqu'on cherche de la certitude, mais une opportunité lorsqu'on cherche à mettre en défaut une méthode de modélisation, et à en identifier les limites de fonctionnement.

Comment être certain de la robustesse de prédiction du modèle lorsqu'un produit différent arrive en production ?

2. Modèle puce sur puce

Maintenant, on calibre le modèle sur toutes les combinaisons possibles des puces. Le tableau récapitulatif des puces est montré figure 4.35. On veut vérifier si les modèles proposés peuvent servir de prédiction, ou bien uniquement de modélisation.

Pour ce faire, on met sur la colonne de gauche les combinaisons possibles des produits, qui seront les données de calibration des modèles. On vérifie ensuite l'inférence des 6 modèles sur chacun des produits. Cela permettra de savoir s'il y a suffisamment de données pour que ces modèles soient utilisés en tant que modèles prédictifs.

Produits de calibration	Produit d'inférence	Modèle 1			Modèle plein		
		R ²	Nb vecteurs	PV diff	R ²	Nb vecteurs	PV diff
281PA,281RA,282FA	281PA	0,59	5	-3	0,85	195	2,81
	281RA	0,63	5	-3,39	0,88	195	0,29
	282FA	0,33	5	-2,74	0,84	195	-0,47
281PA	281PA	0,6	5	-2,39	0,91	195	1,31
	281RA	0,62	5	-2,86	0,74	195	4,09
	282FA	0,22	5	-3,59	-1,43	195	7,6
281RA	281PA	0,58	5	-2,16	0,58	195	6,77
	281RA	0,63	5	-2,61	0,93	195	0,22
	282FA	0,31	5	-2,21	-16,75	195	-2,51
282FA	281PA	0,21	5	1,99	-126,11	195	115,45
	281RA	0,12	5	1,51	-111,35	195	133,32
	282FA	0,5	5	1,53	0,96	195	0,98
281PA,281RA	281PA	0,6	5	-2,36	0,88	195	2,06
	281RA	0,63	5	-2,84	0,92	195	0,15
	282FA	0,28	5	-2,8	-2	195	-1,15
281PA,282FA	281PA	0,59	5	-3,9	0,87	195	1,15
	281RA	0,61	5	-4,37	0,79	195	2,13
	282FA	0,37	5	-3,47	0,82	195	-0,17
281RA,282FA	281PA	0,57	5	-2,61	0,75	195	5,66
	281RA	0,63	5	-2,99	0,92	195	-0,51
	282FA	0,35	5	-2,43	0,8	195	-1,5

Produits de calibration	Produit d'inférence	Modèle 2			Modèle 3		
		R ²	Nb vecteurs	PV diff	R ²	Nb vecteurs	PV diff
281PA,281RA,282FA	281PA	0,78	16	1,04	0,81	29	1,03
	281RA	0,86	16	0,24	0,88	29	-0,4
	282FA	0,58	16	-5,84	0,66	29	-4,33
281PA	281PA	0,81	22	-0,2	0,86	36	0,66
	281RA	0,78	22	-1,08	0,8	36	-1,15
	282FA	-0,34	22	-5,9	0,31	36	-3,65
281RA	281PA	0,74	17	2,41	0,74	30	1,15
	281RA	0,87	17	1,84	0,9	30	-0,25
	282FA	-0,1	17	0,44	-3,96	30	-0,89
282FA	281PA	-2,71	46	24,14	-12,45	78	48,55
	281RA	-7,57	46	29,09	-26,64	78	54,89
	282FA	0,89	46	1,02	0,92	78	1,03
281PA,281RA	281PA	0,8	16	0,89	0,84	32	0,92
	281RA	0,87	16	0,4	0,89	32	-1,18
	282FA	-0,27	16	-5,88	-0,96	32	-2,15
281PA,282FA	281PA	0,81	21	-0,14	0,85	35	1,31
	281RA	0,82	21	-1,12	0,84	35	0,28
	282FA	0,66	21	-3,35	0,71	35	-1,93
281RA,282FA	281PA	0,73	15	1,44	0,75	26	1,4
	281RA	0,88	15	1,08	0,88	26	0,57
	282FA	0,69	15	0,71	0,68	26	-0,23

Produits de calibration	Produit d'inférence	Modèle 4			Modèle 5		
		R ²	Nb vecteurs	PV diff	R ²	Nb vecteurs	PV diff
281PA,281RA,282FA	281PA	0,73	20	1,14	0,78	20	0,58
	281RA	0,82	20	0,56	0,86	20	0,57
	282FA	0,63	20	-1,37	0,66	20	-1,53
281PA	281PA	0,78	20	-0,34	0,79	20	-0,43
	281RA	0,7	20	-0,89	0,86	20	-0,17
	282FA	-2,25	20	-3,35	0,4	20	-2,64
281RA	281PA	0,49	20	2,49	0,73	20	0,36
	281RA	0,84	20	0,91	0,88	20	-0,22
	282FA	0,28	20	-1,36	-5,49	20	-2,01
282FA	281PA	-0,21	20	13,14	-3,64	20	21,72
	281RA	-0,35	20	11,91	-14,92	20	24,08
	282FA	0,79	20	4,88	0,8	20	4,72
281PA,281RA	281PA	0,74	20	1,28	0,78	20	0,36
	281RA	0,82	20	0,04	0,88	20	0,08
	282FA	0,43	20	-2,22	-2,02	20	-2,26
281PA,282FA	281PA	0,76	20	-0,65	0,78	20	0,05
	281RA	0,63	20	-0,31	0,85	20	0,14
	282FA	0,67	20	-0,93	0,68	20	-0,96
281RA,282FA	281PA	0,51	20	2,46	0,74	20	1,06
	281RA	0,84	20	2,33	0,87	20	0,43
	282FA	0,65	20	-0,5	0,67	20	-1,01

Modèle 1	Modèle linéaire sans filtre
Modèle 2	Modèle linéaire avec filtre seuil de corrélation 0.8
Modèle 3	Modèle linéaire avec filtre seuil de corrélation 0.9
Modèle 4	Modèle linéaire avec filtres sans puces voisines de portés [0,1,5,15]
Modèle 5	Modèle linéaire avec filtres avec puces voisines de portés [0,1,5,15]
Modèle plein	Modèle linéaire avec la totalité des 195 vecteurs

Nb Vecteurs	R ²	PV diff abs.
5	< 0,2	< 2
5 à 15	0,2-0,35	2 – 5
15 à 25	0,35-0,5	5 – 10
25 à 40	0,5-0,7	10 – 15
> 40	> 0,7	> 15

	Peak-to-Valley (nm)
281PA	17,39
281RA	18,28
282FA	13,97

FIGURE 4.35 – Tableau récapitulatif des modèles

Totalité des données

Les trois premières lignes des trois tableaux contiennent les mêmes informations que le tableau 4.33. Les commentaires précédemment faits s'appliquent. Il est théoriquement possible de trouver des modèles aussi généralisables que ceux des trois premières lignes. Il reste à déterminer des raisons pour lesquelles ce n'est pas toujours le cas sur les autres lignes.

Un modèle 1 toujours robuste

La robustesse du modèle 1 est encore une fois démontrée. C'est le modèle avec le R^2 descendant le moins bas, et également celui avec le R^2 montant le moins haut sur la totalité des produits. Le Peak-to-Valley suit cette conclusion : le modèle 1 est celui avec les différences absolues de Peak-to-Valley ayant le plus petit intervalle. La différence entre le produit 282FA et les deux autres produits est évidente ici. On remarque que lorsque le produit 282FA est le modèle de calibration seul, les prédictions n'ont des R^2 que de 0.21 et 0.12. Les différences de Peak-to-Valley demeurent satisfaisante dans ce cas-ci cependant. En terme de ce dernier critère, l'écart est correct sur toutes les calibrations et inférences possibles ($< 5nm$). La différence entre le produit 282FA et les deux autres produits doit être gardée en tête lorsqu'on lit le reste du tableau, pour comprendre les autres résultats.

Le produit 282FA d'un côté, 281PA et 281RA de l'autre

Dans les trois lignes suivantes, la calibration s'effectue sur le produit 281PA. Les modèles sur les produits 281RA sont de couleur verte clair dans le modèle 1, et verte foncée dans tous les autres cas. En revanche, le modèle n'est pas performant pour le produit 282FA, sauf pour le modèle 5. Le modèle plein sur-apprend tout de même légèrement sur le produit 281PA, avec une différence de R^2 (0.91 pour 281PA et 0.74 pour 281RA), et une différence de Peak-to-Valley de 1.31 pour 281PA contre 4.09 pour 281RA.

Dans les trois lignes suivantes, le produit est calibré sur le produit 281RA. Les conclusions sont approchantes du paragraphe précédent. Les prédictions sont peu satisfaisantes sur le produit 282FA, et satisfaisantes sur le produit 281PA. On note des exceptions cependant. Sur le modèle plein, le sur-apprentissage est encore plus marqué, avec une différence de R^2 plus grande encore (0.58 pour 281PA et 0.93 pour 281RA), et une différence de Peak-to-Valley de 6.77 pour 281PA contre 0.22 pour 281RA. Sur le modèle 4, ce sur-apprentissage est marqué sur le critère du R^2 , et peu marqué sur le critère du Peak-to-Valley. Sur chacun des modèles 2, 3 et 5, on observe une différence sur les R^2 en faveur du produit 281RA et en défaveur du produit 281PA.

Ces observations rendent compte de l'asymétrie de la calibration et de l'inférence. On conclut que le produit 281PA est un meilleur représentant du produit 281RA que 281RA n'est un représentant de 281PA. On re-démontre que le choix des données de calibration sont clé dans notre cas.

Dans les lignes où l'on calibre le modèle sur le produit 282FA, aucun des 6 modèles ne donne de résultats pleinement satisfaisants en inférence sur les produits 281PA et 281RA. On conclut que 282FA n'est pas un bon représentant des deux autres produits, faisant encore une fois ressortir sa différence. L'exception à relever est le critère de la différence du Peak-to-Valley sur le modèle 1 qui est inférieur à 2 nm.

Dans les lignes où l'on calibre les données sur le produit 281PA et 281RA, la performance est satisfaisante sur les critères du R^2 et du Peak-to-Valley pour les produits 281PA et 281RA. Par contre, sur les critères du R^2 pour le produit 282FA dans les modèles plein, 2, 3 et 5, on observe une faible capacité de prédiction. En revanche, pour les modèles plein, 3, 4 et 5, le Peak-to-Valley a une différence faible pour les trois produits, malgré un R^2 faible. Cela permet de rendre compte que l'usage métier peut être maintenu malgré un écart à la réalité élevé.

Des sélections plus représentatives

Pour terminer de commenter la figure, on rassemble pour chaque tableau les 6 dernières lignes qui concernent les données de calibration (281PA, 282FA) et (281RA, 282FA), très semblables dans leurs résultats ainsi que dans leurs conclusions. Il semblerait que l'ensemble des données soit bien représenté dans ce cas, puisque les résultats ressemblent beaucoup aux trois premières lignes du tableau, avec quelques différences que nous notons dans le paragraphe suivant.

Il y a une augmentation de performance pour le produit 282FA, et un léger contrecoup pour le produit n'appartenant pas aux données de calibration : 281RA dans la sélection 281PA, 282FA et 281PA dans la sélection 281RA, 282FA. Le contrecoup est le plus significatif dans le modèle plein, en particulier sur le critère du Peak-to-Valley. On passe à une différence de Peak-to-Valley de 0.29 à 2.13 pour le produit 281RA, et de 2.81 à 5.66 pour le produit 281PA. Le R^2 quant à lui passe de 0.85 à 0.75 pour le produit 281PA, et de 0.88 à 0.82 pour le produit 281RA.

Le contrecoup est aussi observé pour le produit 282FA, alors qu'il fait partie des deux ensembles de calibration. Cela est contre-intuitif, car on pourrait penser qu'il « pèse » moins dans la calibration lorsqu'il y a 3 produits, comparativement à 2 produits. Pourtant, les résultats sont moins bons sur le critère du R^2 dans les deux sélections à 2 produits que dans la sélection à 3 produits. Cette observation appuie la thèse que la représentativité de l'échantillon est essentielle pour une modélisation performante.

Sur les 6 dernières lignes, les résultats des modèles 2 et 3 sont très satisfaisants, avec des R^2 supérieurs à 0.66 dans tous les cas, et des Peak-to-Valley avec des écarts inférieurs à 2, sauf dans un seul cas. Le bémol est le nombre de vecteurs, en particulier dans le modèle 3 : 35 vecteurs ou 26 ce qui est élevé, comparativement aux modèles 4 et 5.

Les résultats des modèles 4 et 5 montrent une différence. On remarque ici que prendre les puces voisines dans le lissage des vecteurs d'entrée pour la prédiction fait une différence car le modèle 5 est plus performant suivant les deux critères choisis. Le modèle 4 montre que les données de calibration jouent un rôle important : le R^2 est plus élevé sur le produit qui a servi à calibrer le modèle : dans les trois dernières lignes, il est plus élevé sur le produit 281RA, et dans les trois lignes précédentes, il est plus élevé sur le produit 281PA. Les différences de Peak-to-Valley sont bien plus satisfaisantes dans les lignes de calibration 281PA et 282FA. Le modèle 5 montre une confiance élevée sur les 6 dernières lignes, avec des performances élevées sur les deux critères d'évaluation, et un nombre de vecteurs de 20.

R^2 et Peak-to-Valley

Comme dans le tableau précédent où l'on sélectionnait les quarts, on montre dans la figure 4.36 le lien entre le R^2 et le Peak-to-Valley. On clampe les R^2 inférieurs à -1 pour leur donner la valeur -1 également, pour la même raison que précédemment.

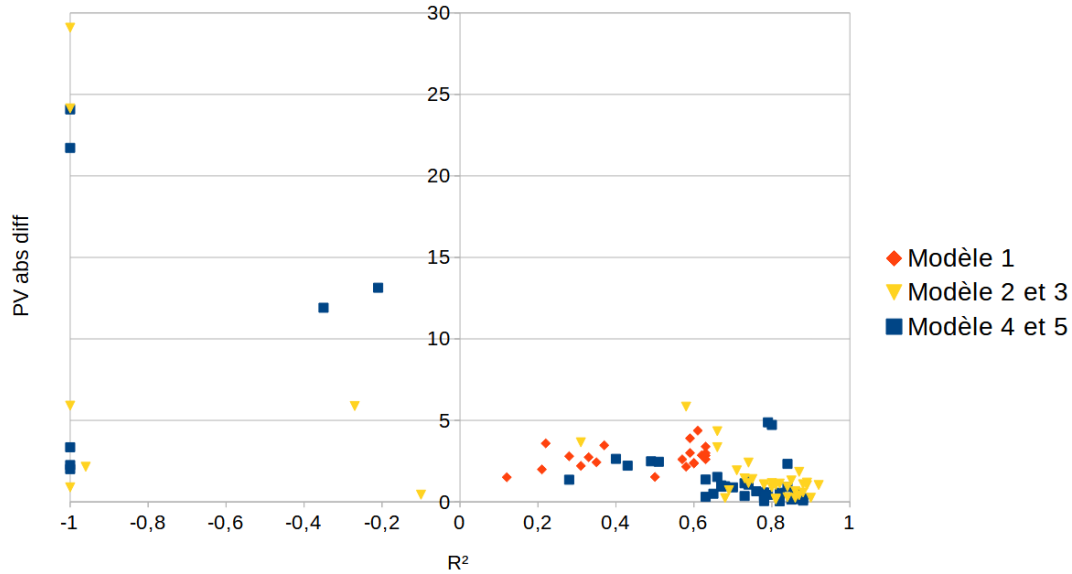


FIGURE 4.36 – Lien entre différence absolue du Peak-to-Valley et R^2 sur les sélections et inférences de tous les modèles

On observe une cohérence globale entre les deux critères, où les points sont inférieurs à 5 lorsque le R^2 est positif. Cette cohérence n'est pas totale cependant, ce qui montre que les deux indicateurs sont complémentaires, et doivent être conservés.

Complexités des modèles : nombre de colonnes

Le critère choisi pour évaluer la complexité est le nombre de colonnes du modèle. Le modèle 1 est le plus simple avec 5 colonnes et une absence totale de transformation des données, et le modèle plein est le plus complexe avec 195 colonnes.

Les modèles 4 et 5 sont ensuite les moins complexes, de façon équivalente : 20 colonnes, et des transformations qui font sens d'un point de vue métier.

Les modèles 2 et 3 ont un algorithme de sélection qui ajoute une complexité quant à la compréhension des modèles, mais par cette étape, cela assure qu'on s'adapte bien au jeu de données qui correspond, avec des colonnes qui ne comportent pas ou peu de redondance linéaire. Ils sont les plus complexes en terme de conception, et majoritairement aussi en terme de colonnes, mais pas toujours. Il semblerait que le modèle 2 soit plus pertinent que le modèle 3 car il comporte moins de vecteurs, et ce pour une performance approchant au modèle 3.

Conclusions du tableau

La thèse appuyée par les observations sur le tableau est que le modèle 1 est très robuste, et utilisable avec peu de données disponibles, bien que ne pouvant pas aller très haut en terme de performance. La seconde thèse appuyée par les observations est que les modèles plus complexes prennent plus de risques, avec des performances pouvant monter très haut, mais avec une incertitude d'autant plus grande. La difficulté est ici d'avoir de la robustesse, et la recherche de critères de confiance a priori pour la qualité de prédiction est un enjeu mis en lumière ici, à la fois pour le métier et pour les sciences de données.

4.5 Critères de confiance a priori et de qualité a posteriori

Les enjeux de cette section sont de connaître a priori la qualité de prédiction d'une nouvelle puce dont on ne dispose pas de mesure de topographie. La sous-partie précédente montre que

les données de calibration sont clés pour la prédiction. Il s'agit donc d'explorer des variables descriptives permettant de savoir en avance la qualité de la prédiction.

Pour ce faire, on cherchera à vérifier si les données de calibration représentent les données à prédire. On s'intéressera en particulier à la variable « distribution » des ensembles de données, en vérifiant si la similarité entre des distributions implique ou non une confiance sur la prédiction.

4.5.1 Introduction

Dans [Gerville-Réache and Couallier, 2011], des problématiques de représentativité d'échantillons sont expliquées. Dans notre travail, on ne parle pas d'échantillons d'un grand ensemble de données mais de deux ensembles de données distincts dont on souhaite qualifier la représentativité mutuelle pour la modélisation.

Pour qu'une prédiction soit efficace, on veut que les données de calibration soient représentatives des données de prédiction. Dans [Gerville-Réache and Couallier, 2011], il est stipulé que la représentativité n'a de sens que par rapport à un contexte défini, et que le concept est largement intuitif. Lorsqu'on parle de représentativité, on parle de représentativité suivant des variables.

Les variables doivent être calculées à partir des vecteurs d'entrée, puisqu'on veut des critères a priori. Ici, ce sont les densités périmétriques et surfaciques brutes. On s'intéresse à une variable intuitive concernant la représentativité d'un échantillon : la distribution. Dans notre cas, les distributions des 5 vecteurs d'entrée sont considérées.

La régression linéaire modélise un comportement moyen. Ainsi, si en moyenne les données de calibration et les données de l'ensemble à prédire sont proches selon les variables qui mesurent cette proximité, on peut raisonnablement supposer que les résultats de prédiction seront proches.

Dans le cas où les données suivent une loi parfaitement linéaire et sans variance, il n'y a pas ou peu de problèmes de représentativité. En revanche, comme c'est le cas dans les mesures, les données ont une variance liée à plusieurs facteurs (précision de mesure, variabilité de procédé, facteurs non identifiés...).

Le problème pouvant se produire est lié à l'impact de certains cas sous-représentés face à une masse de cas sur-représentés. La variance des cas sur-représentés sera alors modélisée et pourra prendre le dessus sur un ensemble de points moins nombreux des données dont la composante moyenne est le centre d'intérêt du modèle. C'est un cas de sur-apprentissage dans le cas de la régression linéaire.

On souhaite donc mesurer la distance des densités périmétriques et surfaciques entre les produits. Le procédé final visé est le suivant :

1. On dispose de p produits mesurés avec la topographie moyenne, ainsi que les données périmétriques et surfaciques.
2. Un nouveau produit arrive en production. On dispose de son design, mais pas de mesures de topographie car il n'a pas encore été mis en production. On compare ses distributions de densités surfaciques et périmétriques aux ensembles des p produits.
3. La proximité nous donne un indicateur de confiance ou de méfiance envers la prédiction. Si la distance est faible, on peut se fier à la prédiction, sinon on procède avec parcimonie, et on a conscience que la modélisation peut être fautive. Ce qui n'empêche pas de voir la prédiction pour information.

4.5.2 Distances entre distributions

1. Tests statistiques et distances

On se pose la question de savoir si deux ensembles de données sont issus ou non d'une même distribution. Les tests de Kruskal [Sawilowsky and Fahoome, 2014] et de Wilcoxon

[Fagerland and Sandvik, 2009], très semblables, vont vérifier que deux ensembles de points sont approximativement ordonnés de la même façon si on agrège ces ensembles de points. Ces tests ne s'intéressent pas aux valeurs des points, mais à leur ordonnancement lors de l'agrégation, aussi ils ne seront pas présentés car peu concluants.

Un test permettant de vérifier une égalité entre distribution est le test de Kolmogorov-Smirnov [Hodges, 1958]. Celui-ci donne en résultat une p-valeur d'acceptation de l'hypothèse « les deux échantillons sont issus d'une même distribution ». Dans notre cas, le test de Kolmogorov-Smirnov rejette systématiquement l'hypothèse d'égalité.

L'explication est que les ensembles de données ont un grand nombre de points, et le test aura avec certitude la réponse que **non**, les distributions ne sont pas **exactement** égales, et ce sur chacun des vecteurs. Le grand nombre de points confirmant avec certitude l'existence d'une différence.

De façon générale, le problème des tests statistiques est la binarité de leurs résultats. On cherche ici plutôt une métrique intuitive et indicative, car comme expliqué dans [Gerville-Réache and Couallier, 2011], le concept de représentativité est intuitif. Au lieu de considérer le test, on se tourne vers des mesures de distances entre fonctions.

2. Distances entre distributions de densités

Pour évaluer les distances entre les distributions des densités, on commencera par transformer la liste de points en une fonction de répartition. On évaluera ensuite la proximité des fonctions de répartition des densités périmétriques et surfaciques en utilisant différentes distances. On montre les fonctions de répartition des produits dans la figure 4.37.

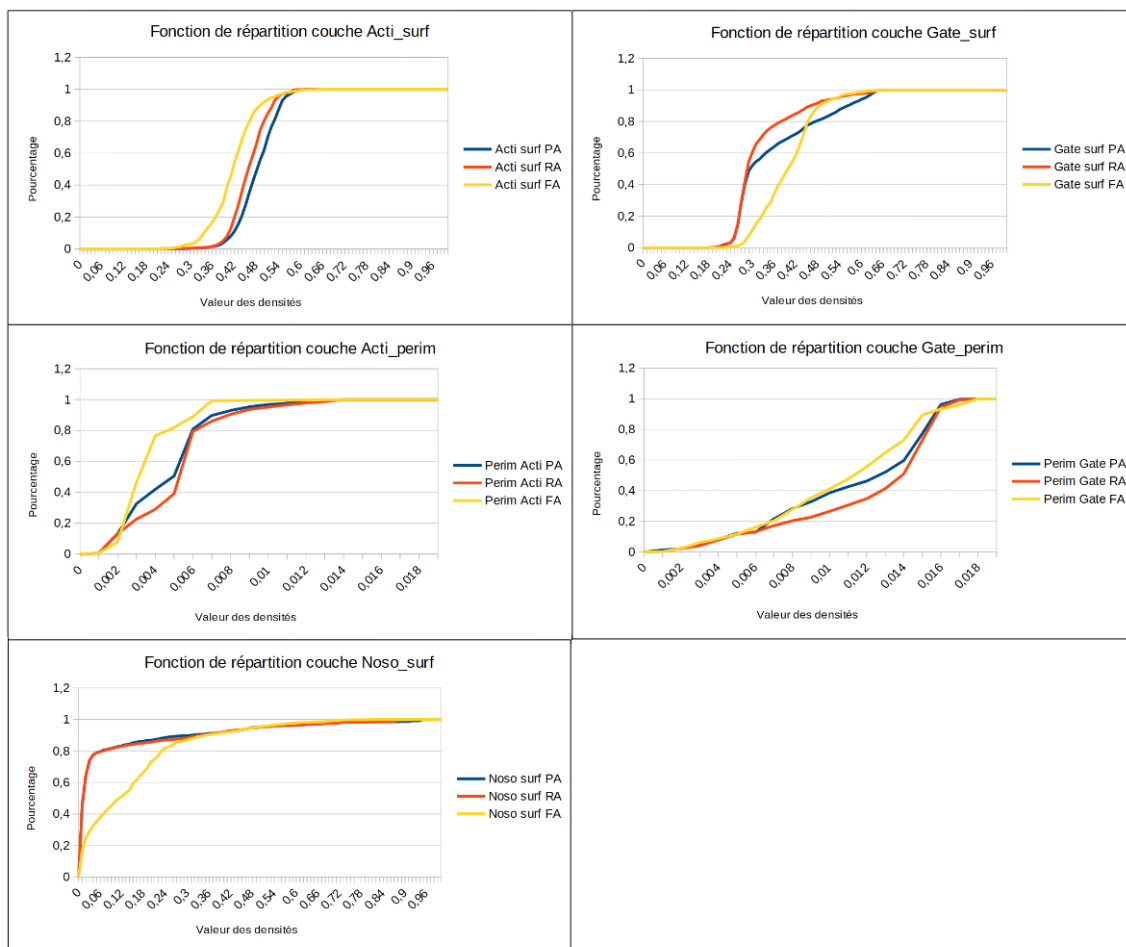


FIGURE 4.37 – Fonctions de répartition des densités surfaciques et périmétriques

Dans le cas où l'on dispose de plusieurs produits pour calibrer les données de calibration (comme c'est le cas ici), on devra également considérer les fonction de répartition d'ensembles de données agrégés de plusieurs produits. Ces ensembles agrégés peuvent servir de données de calibration, et donc doivent être évalués à l'arrivée d'un nouveau produit. On ne doit pas nécessairement choisir la totalité des produits disponibles pour calibrer le modèle, mais choisir l'ensemble de données le plus représentatif du nouveau produit suivant les variables.

Nous présentons la distance L2 entre les fonctions de répartition, expliquée dans [Deza and Deza, 2009]. Nous présentons aussi la mesure de Kolmogorov, utilisée par le test du même nom, elle aussi mentionnée dans [Deza and Deza, 2009]. Elles sont intuitivement simples à comprendre et également rapides à calculer. Ainsi, si un lien entre ces distances et la qualité de prédiction est trouvé, on aura un processus composé d'opérations simples pour construire un modèle, et à la fois obtenir un indicateur de robustesse pour savoir si on aura une prédiction fiable sur un nouveau produit.

La première distance que nous considérons pour évaluer les distances entre ces courbes est la distances L2. On montre dans la figure 4.38 la distance entre les ensembles de données et chacun des trois produits. Pour aider à la lecture, on donne un code couleur selon la distance à gauche de la figure. Ces distances seront mises en regard avec les résultats des modèles sur les tableaux de la figure 4.35 pour voir s'il existe une cohérence pour en extraire une variable de robustesse.

	Données 1	Données 2	Acti_surf	Gate_surf	Noso_surf	Acti_perim	Gate_perim	Total L2
	281PA	281RA	2,23	2,579	0,293	1,647	1,976	8,725
	281PA	282FA	6,725	5,161	5,046	4,334	1,548	22,814
	281RA	282FA	4,682	7,115	4,95	5,98	3,338	26,065
	281RA	281PA	2,23	2,579	0,293	1,647	1,976	8,725
	282FA	281PA	6,725	5,161	5,046	4,334	1,548	22,814
	282FA	281RA	4,682	7,115	4,95	5,98	3,338	26,065
	281PA, 281RA	281PA	1,515	1,619	0,26	1,021	1,249	5,664
	281PA, 281RA	281RA	0,873	0,971	0,198	0,635	0,744	3,421
	281PA, 281RA	282FA	5,369	6,335	4,983	5,338	2,668	24,693
	281PA, 282FA	281PA	1,627	1,173	1,399	0,989	0,367	5,555
	281PA, 282FA	281RA	1,281	3,396	1,353	2,595	2,288	10,913
	281PA, 282FA	282FA	5,012	3,953	3,662	3,32	1,195	17,142
	281RA, 282FA	281PA	3,027	1,797	1,013	0,794	1,469	8,1
	281RA, 282FA	281RA	0,74	1,16	0,916	0,938	0,564	4,318
	281RA, 282FA	282FA	3,928	5,362	4,067	4,993	2,806	21,156
	281PA, 281RA, 282FA	281PA	2,009	1,207	0,628	0,537	0,99	5,371
	281PA, 281RA, 282FA	281RA	0,601	1,529	0,56	1,163	1,011	4,864
	281PA, 281RA, 282FA	282FA	4,794	5,599	4,396	4,773	2,392	21,954

L2

- 0 à 1
- 1 à 2
- 2 à 3
- 3 à 4
- > 4

L2 total

- 0 à 5
- 5 à 10
- 10 à 15
- 15 à 20
- > 20

FIGURE 4.38 – Distances L2 entre les fonctions de répartition des densités

La première observation sur ce tableau, est que le produit 282FA seul n'est jamais proche (une seule fois vert clair, jamais vert foncé) d'aucune des distributions des autres ensembles de données. Il n'est donc pas bien représenté par les ensembles de données sur la variable « fonction de répartition des densités ». L'ensemble qui s'en sort le mieux à le représenter est l'ensemble « 281PA, 282FA ». On remarque un manque de robustesse global pour la prédiction du produit 282FA sur les tableaux 4.35, mais une bonne prédiction sur le modèle 5 pour la calibration « 281PA, 282FA ».

La seconde observation est que le produit 281RA seul est bien représenté lorsque celui-ci fait partie des ensembles de gauche (vert clair ou vert foncé uniquement). On peut ensuite noter que le produit 281PA le représente partiellement bien sur la distance L2 : deux cases vertes claires (Acti_perim et Gate_perim), une case verte foncée (Noso_surf) et deux cases jaunes (Acti_surf et Gate_surf). Pour l'ensemble « 281PA, 282FA », la distance est moyennement élevée, avec des cases autour de la moyenne des distances : entre 1.2 et 3.4 pour la distance L2. On peut alors raisonnablement supposer que les ensembles « 281PA, 282FA » et « 281PA » seront des bons ensembles de données de calibration pour un modèle prédictif du produit 281RA. Dans les tableaux 4.35, c'est effectivement le cas : les résultats des modèles calibrés sur ces ensembles ont de très bons résultats pour prédire la topographie du produit 281RA.

Le produit 281PA est très bien représenté lorsque celui-ci fait partie des ensembles de données, et ce sur les deux normes choisies. Il est également représenté avec 3 cases vertes claires, une cases verte foncée sur l'ensemble « 281RA, 282FA » pour la distance L2, et avec 3 jaunes, 1 verte claire et une verte foncée sur l'aire entre courbes. De même, on observe une robustesse et une précision élevées sur ces ensembles pour le modèle 5 sur les tableaux 4.35.

Une autre mesure de distance entre fonctions de répartition est la mesure de Kolmogorov, définie comme le plus grand écart en valeur absolue entre les fonctions de répartition. On montre sur la figure 4.39 les mesures de Kolmogorov des ensembles de données précédents.

Données 1	Données 2	Acti_surf	Gate_surf	Noso_surf	Acti_perim	Gate_perim	Total
281PA	281RA	0,186	0,138	0,093	0,128	0,121	0,666
281PA	282FA	0,471	0,383	0,492	0,35	0,133	1,829
281RA	282FA	0,342	0,507	0,497	0,475	0,232	2,053
281RA	281PA	0,186	0,138	0,093	0,128	0,121	0,666
282FA	281PA	0,471	0,383	0,492	0,35	0,133	1,829
282FA	281RA	0,342	0,507	0,497	0,475	0,232	2,053
281PA, 281RA	281PA	0,117	0,087	0,117	0,08	0,077	0,478
281PA, 281RA	281RA	0,069	0,053	0,086	0,05	0,046	0,304
281PA, 281RA	282FA	0,384	0,461	0,498	0,429	0,197	1,969
281PA, 282FA	281PA	0,118	0,091	0,192	0,082	0,032	0,515
281PA, 282FA	281RA	0,106	0,221	0,188	0,21	0,14	0,865
281PA, 282FA	282FA	0,354	0,296	0,362	0,269	0,105	1,386
281RA, 282FA	281PA	0,216	0,108	0,169	0,063	0,095	0,651
281RA, 282FA	281RA	0,062	0,093	0,159	0,076	0,039	0,429
281RA, 282FA	282FA	0,288	0,422	0,41	0,4	0,196	1,716
281PA, 281RA, 282FA	281PA	0,145	0,072	0,116	0,042	0,063	0,438
281PA, 281RA, 282FA	281RA	0,051	0,1	0,103	0,093	0,063	0,41
281PA, 281RA, 282FA	282FA	0,343	0,41	0,441	0,383	0,178	1,755

Mesures de Kolmogorov

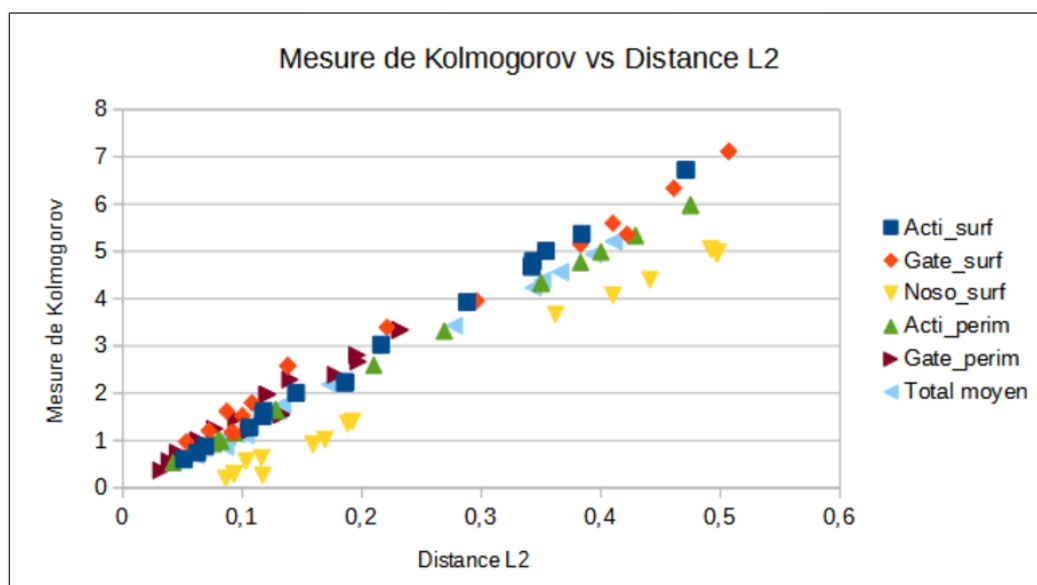


FIGURE 4.39 – Mesure de Kolmogorov et comparaison avec distance L2

La tendance est que les deux mesures sont largement proportionnelles. Les commentaires sur le tableau des mesures sont très similaires. On retrouve des commentaires semblables pour les produits 281PA et 281RA, et on remarque que le produit 282FA n'est jamais bien représenté seul par un ensemble de données.

La conjecture de cette étude sur ces trois mesures, est que l'on peut utiliser de façon relativement équivalente les normes choisies pour évaluer les distances entre distributions. On peut

en revanche trouver des cas où la mesure de Kolmogorov sera très délocalisée de la courbe de proportionnalité avec la distance L2 montrée sur le graphe de la figure 4.39.

L'exemple serait deux ensembles de données de la forme suivante : (a, a, a, \dots, a) et (b, b, b, \dots, b) . Les fonctions de répartition de ces deux ensembles de données sont deux fonctions en escalier, valant 0 respectivement sur $[0, a[$ et $[0, b[$ et 1 respectivement sur $[a, \infty[$ et $[b, \infty[$. La mesure de Kolmogorov entre ces deux ensembles de données vaut toujours 1 lorsque a est différent de b . La distance L2 dépend de l'écart entre a et b . En faisant tendre a vers b , la distance L2 tend vers 0, mais la distance de Kolmogorov vaut 1. En faisant tendre b vers l'infini, la distance L2 tend vers l'infini, mais la distance de Kolmogorov vaut toujours 1.

On choisit de n'utiliser que la distance L2 dans la suite, car elle prend en compte la totalité des valeurs, et ne fait pas état d'une différence localisée uniquement. Nous étudions cette distance par la suite.

4.5.3 Résultats et discussion

Avec la distance L2 comme outil de comparaison entre les ensembles de données des produits, on montre le lien entre le tableau 4.35. On prend chaque ligne des tableaux dont les données de calibration sont différentes des données d'inférence, et on place les résultats sur un graphique de la figure 4.40 et 4.41 en fonction du vecteur concerné (voir titre du graphique). On ajoute le graphique « Total » qui, au lieu de séparer les vecteurs prend en compte la somme des distances des vecteurs.

Les figures 4.40 et 4.41 montrent **strictement** les mêmes points graphique à graphique. Cependant, on les montre tous les deux, pour mettre en lumière les différences entre les modèles pour la figure 4.40 et les différences entre les produits pour le graphique 4.41.

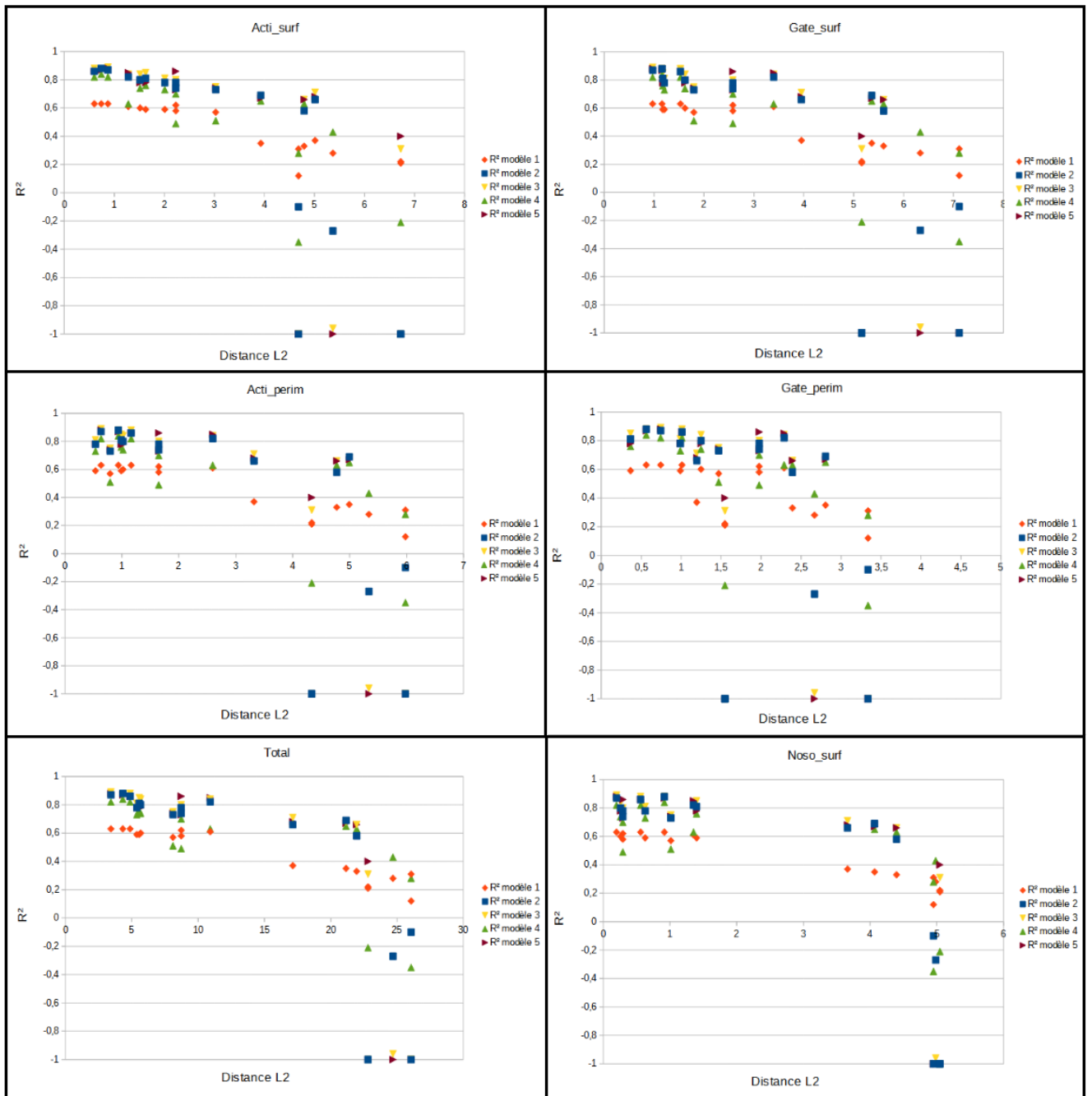


FIGURE 4.40 – Graphiques de points plaçant le R^2 des différents **modèles** en fonction de la distance L2 entre chaque vecteur

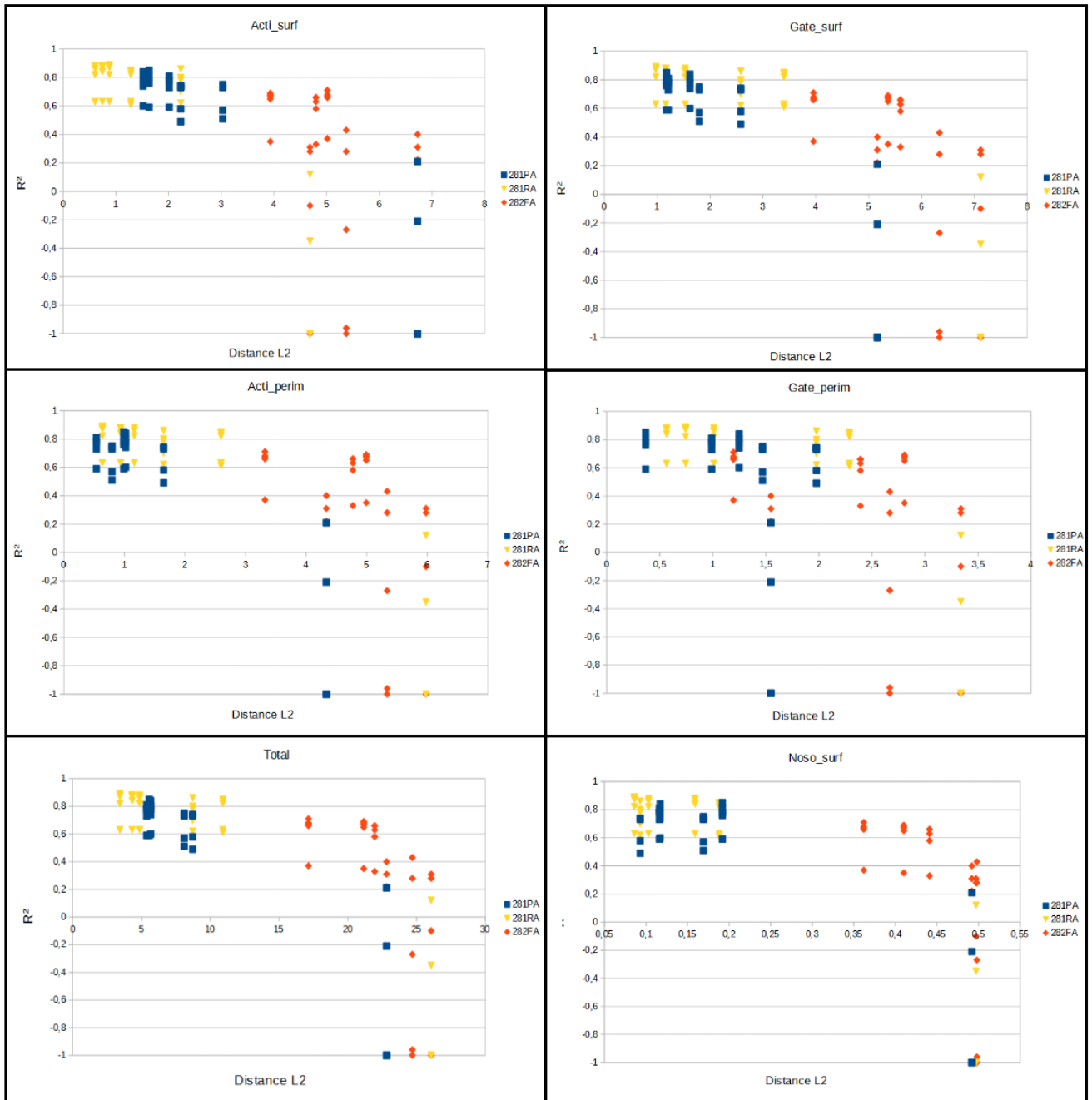


FIGURE 4.41 – Graphiques de points plaçant le R^2 des différents **produits** en fonction de la distance L2 entre chaque vecteur

1. Position des points

Pour commencer, on commente la position des points. On peut donc regarder indifféremment l'une ou l'autre des deux figures 4.40 ou 4.41. Chacun des vecteurs peut servir de critère a priori. Aucun des vecteurs ne semble discriminer plus que l'autre. Aussi, ces critères sont complémentaires car chacun des vecteurs ordonne les distances différemment pour des mêmes modèles.

Ce que l'on peut conclure sur ces graphiques dans leur globalité est que lorsqu'on se situe avec une distance faible, les R^2 sont au-dessus de 0.5. Lorsque les distances dépassent un seuil, les résultats varient et sont peu fiables. Ils peuvent être bons comme très mauvais. La robustesse

n'est pas assurée pour des distances élevées.

Pour simplifier la lecture, et afin de limiter les informations à consulter pour évaluer la proximité entre les distributions, on propose la métrique « Total » qui est la somme des métriques. Pour cette métrique, on remarque que lorsque la métrique est inférieure de 15, on a une prédiction avec un R^2 supérieur à 0.5. On peut aller un peu plus loin et dire que lorsque la métrique est inférieure à 22, le R^2 est supérieur à 0.3.

2. Interprétation des produits

Ici, on retrouve le fait que le produit 282FA se distingue des deux autres, de par ses distances et de par ses résultats de prédiction. Dans le cas où nous aurions disposé uniquement des produits 281PA et 281RA, on aurait eu un indicateur a priori quant à la précision de la prédiction, et de la méfiance qu'on peut accorder à ces modélisations.

On aurait pu observer que le produit 281RA est suffisamment bien représenté par les produits 281PA et 282FA, particulièrement par 281PA, et qu'on aurait pu être confiant sur ses prédictions. De façon symétrique, le produit 281PA est bien représenté par le produit 281RA.

3. Interprétation des modèles

Comme nous l'avons dit lors du commentaire du tableau 4.35, on observe sur la figure 4.40 que le modèle 1 est celui qui comporte le plus de robustesse, et qui ne descend jamais en-dessous de 0. S'il y avait un modèle à choisir pour un produit qui a une distance élevée, ce serait celui-ci.

Le modèle 5 est celui qui a la plus grande robustesse pour le critère choisi. Les performances des modèles ont un R^2 au-dessus de 0.6 pour des distances totales inférieures à 22. On a ensuite des résultats variables en lesquels on a peu de robustesse.

Les modèles 2, 3 et 4 ont une robustesse plus faible, et une performance équivalente, voire inférieure. Les modèles 2 et 3 ont le désavantage de rendre imprévisibles les colonnes choisies. Aussi la compréhension et l'interprétation du modèle sera plus difficile lorsqu'on regarde les coefficients.

4. Conclusion

Le choix final de ce chapitre se porte sur les modèles 1 et 5. En conclusion, si la métrique de distance totale est basse, on peut avoir confiance en un modèle qui prend des risques : le modèle 5. Si la métrique de distance totale est haute, alors on utilisera le modèle 1 qui prend moins de risque, et on aura conscience de la limite de capacité de prédiction du modèle. Avec les données dont on a disposé pour effectuer ce travail, et sur cette technologie, on place la limite à 20 pour la métrique totale.

Dans des conditions où on dispose d'un nombre différent de 5 couches, on préconise comme limite de $4 \times$ nombre de couches. Cette hypothèse reste largement à vérifier dans les conditions concernées. L'usage de distances entre les fonctions de répartition est à conserver, à la fois en tant que métrique pour anticiper des similarités dans le comportement des modèles de nano-topographie.

Le protocole est le suivant dans notre cas :

1. Un nouveau produit P_0 arrive en production. On dispose de ses densités périmétriques et surfaciques, mais pas de sa nano-topographie. On souhaite la modéliser, et on dispose également de P produits avec leurs densités périmétriques et surfaciques, mais également leurs nano-topographies.
2. On cherche, parmi l'ensemble des tuples des P produits, la combinaison C_0 dont la distance L_2 totale est la plus basse.
3. Si cette distance est inférieure à 20, alors on utilise le modèle 5 calibré sur C_0 pour modéliser le produit P_0 . Si cette distance est supérieure à 20, alors on utilise le modèle 1 calibré sur C_0 pour modéliser le produit P_0 , en sachant que ce modèle est imprécis.

La similarité des comportements plus généraux d'un nouveau produit arrivant dans l'industrie est également un axe ouvert à explorer, aidant les ingénieurs à orienter leurs décisions pour industrialiser un nouveau produit.

4.6 Conclusion

Le parti-pris de ce chapitre a été de réaliser un modèle simple et compréhensible. Cela s'est traduit par l'établissement d'un modèle linéaire avec des entrées transformées. Le modèle linéaire représente l'empilement de couches minces, et les entrées transformées représentent les différentes portées des procédés.

Des critères de robustesse des modèles ont été présentés, en établissant un protocole comparant le produit à prédire aux produits déjà disponibles. Cette démarche est à retenir, et créer des indicateurs de confiance quant à la robustesse d'un modèle est important lors de la création d'une modélisation.

L'enjeu double de pouvoir fournir une explication métier du modèle et d'avoir une performance élevée et fiable est au cœur de ce travail. La modélisation de topographie résiduelle liée au design n'en est qu'à ses débuts, et sera sans aucun doute l'un des points limitant lors de la production de masse des puces des prochaines générations. La topographie est déjà en cause de baisses de rendement, et celle-ci aura un impact d'autant plus grand, à mesure que les dimensions des transistors continuent à diminuer. Un protocole complet a été proposé, et des portes restent largement ouvertes pour poursuivre ce travail dans les directions mises en lumière.

À terme, utiliser un modèle de topographie doit faire partie du processus de conception du design. Le but est d'identifier des configurations problématiques pour la production dans un rendement optimal. Cette perspective va dans le sens du « digital twin », où une version digitale de chaque étape de la puce est disponible, et ce de bout en bout.

Dans le contexte de la modélisation de topographie, il n'y aura au final pas un, mais des modèles : un pour chaque étape de fabrication. Il faudra dans un premier temps modéliser la topographie aux étapes où la topographie est la plus critique, puis continuer dans la direction préconisée par l'anticipation des problèmes et modéliser la topographie à chacune des étapes.

L'enjeu ici est en fait triple. Ce travail sur la modélisation de topographie se poursuivra, et l'existence de la présentation d'un tel travail compréhensible par tous les métiers est essentiel pour prendre conscience de tous les enjeux de la modélisation de topographie (simplicité, performance, évaluation, compréhension, robustesse). Prendre connaissance de ce travail de thèse permet à un profil métier ou informatique de se familiariser avec ces enjeux et d'assurer la pérennité des choix techniques à faire pour l'industrie de la micro-électronique.

Chapitre 5

Détection de défauts : Striations

Dans ce chapitre, nous présentons un cas d'étude de détection de défaut, en abordant les points suivants :

1. L'importance de la détection de défauts et des sur-coûts engendrés lors d'une non-détection dans l'industrie.
2. Le problème des striations et de ses enjeux.
3. La façon dont on souhaite traiter le problème : réseau de neurones avec entrées augmentées.

5.1 Des défauts dans la vie d'un produit

La détection de défauts dans une industrie fait partie intégrante des cartes de contrôles mentionnées dans le chapitre 1. Pour rappel, une carte de contrôle est un ensemble de contrôles visant à détecter des défauts ou des dérives durant la fabrication d'un produit.

Pour détecter la dérive d'une mesure, on doit connaître son comportement habituel, ou standard. Pour détecter un défaut, on doit connaître le résultat attendu, pouvant soit être la mesure habituelle ou standard, ou bien le résultat d'une modélisation. On peut réaliser des tests fonctionnels, ou bien des tests de durabilité par exemple.

On cherche à détecter les défauts ou des dérives le plus en amont possible, afin d'éviter des sur-coûts qui auraient pu être évités si l'on avait détecté le défaut ou la dérive en amont. Ces sur-coûts prennent différentes formes de gravité, selon le stade auquel les défauts sont détectés.

Pour commencer à présenter les sur-coûts, on se place dans le cas le plus extrême à éviter dans une industrie : livrer un produit non fonctionnel. Ce sur-coût est le plus grave pour une industrie, car il faut rappeler les produits de chez le client, puis rembourser et/ou re-livrer avec une pénalité. Le préjudice causé est également l'image de l'entreprise et la confiance que le client accorde à celle-ci. La gravité de rappeler des produits une fois en usage peut également aller jusqu'à la vie humaine, comme dans le cas d'avions ou de voitures.

On peut détecter un défaut sur un produit à la fin de la chaîne de fabrication. Celui-ci peut être fonctionnel mais par exemple ne pas passer des tests de durabilité. Il est de fait non commercialisable, et la totalité du coût de la fabrication du produit devient une perte pour l'entreprise. Entre autres, la perte correspond à la totalité des ressources utilisées pour la production, ainsi que les délais engendrés par une re-fabrication.

Les défauts peuvent également être détectés au cours de la chaîne de fabrication du produit. Dans le cas où le défaut ne peut être corrigé, c'est la totalité des procédés précédemment effectués qui sont un sur-coût. Comparativement au cas précédent, le gain apporté par la détection est de ne pas engendrer les coûts de la suite de la chaîne de fabrication qui aboutiraient à un produit non fonctionnel. Dans certains cas, le défaut est corrigible, en refaisant une brique de fabrication, par exemple. Ainsi, le sur-coût sera la brique à refaire entièrement, ainsi que d'éventuelles étapes

pour « annuler » cette brique. Comparativement au cas précédent, le gain est la totalité des étapes précédant la brique en question qui ne sont pas à refaire.

5.2 Cas d'étude : Striations sur capteurs optiques

Dans cette section, nous présentons les striations : un défaut lié à la topographie intervenant lors d'une étape de couchage de résine. Le couchage est une étape réversible : on peut enlever la résine et refaire l'étape. Il est très intéressant de pouvoir détecter un défaut à ce stade de fabrication des puces. Le gain est fort car on ne recommence pas le procédé du début, mais juste la brique précédente.

Nous commençons par présenter le contexte physique du problème, en expliquant l'intérêt de la résine ainsi que la conséquence de ce défaut. Ensuite, nous détaillons les données utilisées pour ce problème. Enfin, nous montrons l'apport du travail effectué qui a amélioré la détection de ce défaut en complétant un autre travail de thèse.

5.2.1 Présentation du problème

1. Imageur et capteur optique : fonctionnement

Un imageur est un capteur visant à reconstituer une image. Ceux fabriqués à ST [Demésy, 2009] sont une matrice de pixels, tous séparés d'une distance de quelques μm , comme montré dans la figure 5.1. Sur chacun des pixels se trouve une micro-lentille, et en-dessous de chacune d'elle une pile de couches de résines, et d'autres couches dont nous ne parlerons pas. L'ordre dans lequel sont étalées les résines a pour but de construire des motifs.

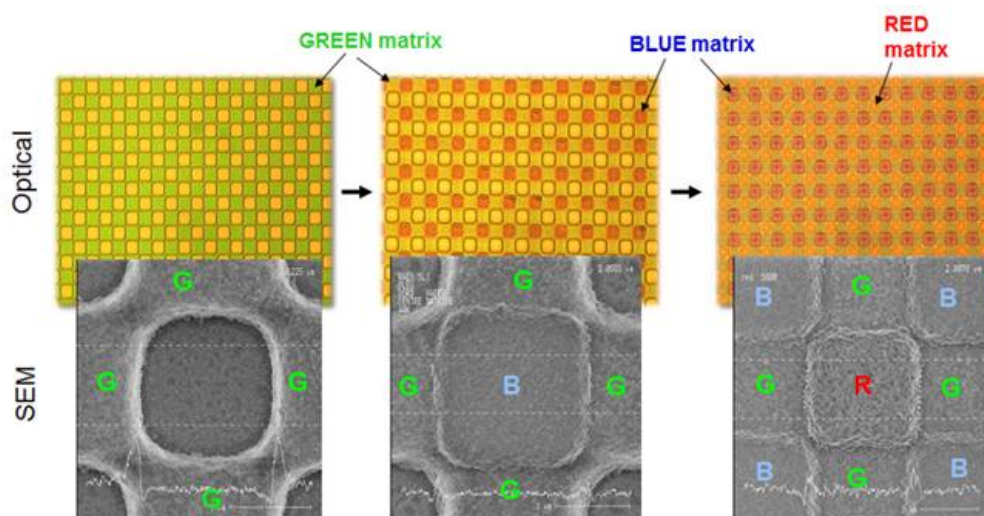


FIGURE 5.1 – Matrices de pixels et ordre d'étalement des résines colorées - *Source interne*

Ces couches de résines doivent former un agencement précis, pour que le dispositif capte la quantité de lumière rouge, bleue et verte, afin de former un pixel. Sur chaque micro-lentille, il y a une pile d'éléments ayant chacun une fonction. Les deux éléments qui nous intéressent le plus sont le capteur de lumière (photo-diode), et la résine colorée. L'idée est, sur chaque pile, de filtrer une lumière avec la résine colorée, et de capter la quantité de lumière restante. En faisant fi des autres éléments, on comprend qu'en ne laissant passer que la lumière verte, par exemple, et en mesurant la quantité de lumière du signal restant, on a mesuré la quantité de lumière verte présente. On montre une coupe longitudinale sur la figure 5.2.

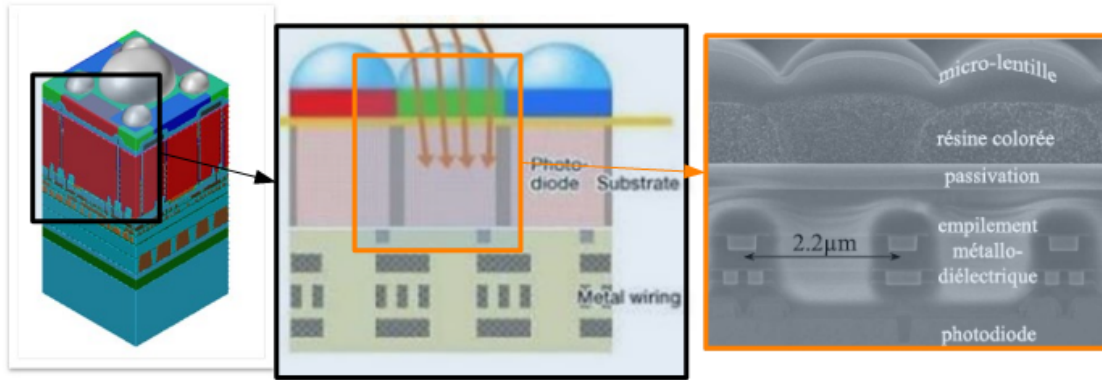


FIGURE 5.2 – Schéma de coupes longitudinales - *Source interne et [Demésy, 2009]*

Sur la coupe de la figure 5.2, on voit la pile de couches pour chacune des micro-lentilles. Sur la partie gauche, on voit le motif final théorique.

2. Couchage de résine colorée et striations

Dans le point précédent, nous expliquons l'importance de la résine dans le fonctionnement du capteur. L'étape de procédé qui permet de mettre ces couches de résines colorées est ce que l'on appelle un couchage. Cela consiste en l'étalement de résine sur la totalité du wafer, sur lequel se trouvent les capteurs optiques. On souhaite que la résine soit d'épaisseur égale sur la totalité d'un capteur afin d'éviter de fausser la quantité de lumière captée. Le résultat de l'opération sur 9 capteurs est montré sur la figure 5.3.

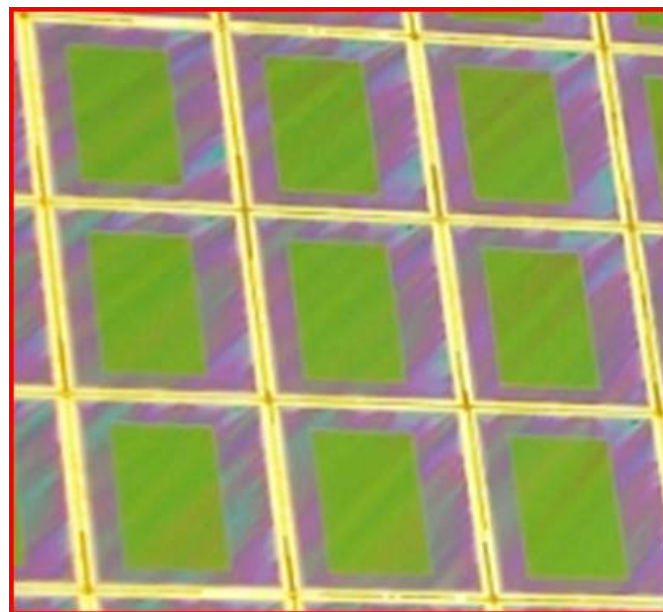


FIGURE 5.3 – Striations sur capteurs - *Source interne*

Sur la figure 5.3, la forme attendue à l'issue du procédé est un rectangle vert uniforme. On peut constater une irrégularité d'étalement prenant la forme de « vaguelettes ». Cela est dû à une différence d'épaisseur, et ces « vaguelettes » s'appellent des **striations**. Ce défaut est présent dans les zones du wafer « Nord-Ouest », « Nord-Est », « Sud-Ouest » et « Sud-Est ». Cette variation d'épaisseur peut ou non causer un dysfonctionnement, qui rend la puce non commercialisable dans certains cas.

Ce sont sur ces rectangles verts que l'intérêt est porté. Sur le reste du capteur, les striations n'ont pas d'impact sur le rendement. Ce rectangle identifié comme la zone d'intérêt est appelé « Care Area ». C'est dans cette Care Area que les striations doivent être détectées, et décrites comme un défaut tueur ou non.

3. Métriques et actions correctives

La question à laquelle STMicroelectronics souhaite répondre est : dans quelle mesure le rendement est affecté par la présence de striations ?

Le problème des striations est un problème qui ne résulte pas d'une erreur, car celles-ci sont inhérentes au procédé de couchage et toujours présentes. En revanche, on veut savoir si celles-ci sont trop importantes pour avoir un produit commercialisable ou non. Pour ce faire, le besoin à STMicroelectronics est d'avoir une métrique pouvant quantifier l'ampleur des striations.

Avoir une métrique permet d'y mettre un seuil de gravité, avec pour règle : si la métrique dépasse le seuil, alors le produit n'est pas commercialisable. Au-delà de cet intérêt pour commercialiser, cela a pour utilité de mettre en place des actions correctives lors de la phase « Recherche et Développement » d'un produit, et de vérifier la pertinence de ces actions, en regardant si celle-ci s'est améliorée, détériorée, ou bien si l'action n'a eu aucun impact. Par action on peut faire référence à un changement de paramétrage sur une machine, par exemple.

Dans notre cas spécifique, le fait de pouvoir refaire l'étape génère un grand intérêt économique pour la détection de ce défaut. Si les striations sont trop importantes, alors on refait la brique concernée. Le couchage est une étape à la fin de la production, comme on peut le voir sur la figure 5.2, la résine colorée est en haut du produit donc l'une des dernières étapes de fabrication.

Avoir une métrique pour quantifier la gravité des striations et chercher un seuil pour savoir si un produit est commercialisable est un axe de recherche. Un autre axe de recherche est la prédiction des puces commercialisables en utilisant un réseau de neurones (ou autre méthode). Un travail de thèse à STMicroelectronics fait une approche par deep learning et a donné lieu à une publication d'article [Alcaire et al., 2021].

Le travail effectué dans ce manuscrit est de rendre visibles les striations, pour en faciliter l'observation, et également pour se diriger vers l'obtention d'une métrique. Nous montrons également que ce qui a été fait est complémentaire avec le travail de thèse en cours à STMicroelectronics en tant que pré-traitement de données.

5.2.2 Préparation des données

L'outil PWG possède la bonne échelle dimensionnelle pour mettre en évidence ce défaut. Les striations font environ $300\mu\text{m}$ de largeur pour une longueur d'environ 1mm , ce qui correspond plutôt bien à la dimension des pixels du PWG ($100\mu\text{m}$ de pixels de large). Il a également une rapidité de mesure intéressante pour un suivi de procédé (60 secondes pour mesurer un wafer entier de 1500 puces).

Dans l'usage classique, la mesure de l'outil PWG est destructive car on y applique un dépôt réfléchissant. Si l'on souhaite faire de la détection de défaut comme c'est le cas, on ne peut utiliser ces mesures destructives. L'idée est de ne pas appliquer le dépôt réfléchissant sur la plaque pour exploiter la forte capacité métrologique du PWG. La conséquence de ce choix est une mesure imprécise, nécessitant des traitements pour la rendre exploitable au maximum.

L'enjeu difficile du travail, est de pouvoir valoriser une mesure avec beaucoup d'imprécision, comme montré sur la figure 5.5 qui comporte des aberrations optiques de reconstruction de surface. Ces erreurs sont observées dans les mesures qui fonctionnent par interférométrie, une technique de métrologie comportant beaucoup de traitement du signal, et donc extrêmement sensible à des changements sur la collecte de signaux [Su et al., 2020].

On montre une carte d'exposition décrivant le produit ainsi que le layout du produit sur la

figure 5.4. Sur la carte d'exposition on constate que le nombre de puces est de 925 réparties sur des champs de 12 puces (3 en X par 4 en Y). On montre sur le layout l'endroit où se situe la matrice du capteur précédemment évoquée. Une ambiguïté existe : « pixel », « puce » et « capteur » sont différents noms faisant référence à un seul et même objet dans ce contexte. Sur la figure 5.4, il s'agit de la puce.

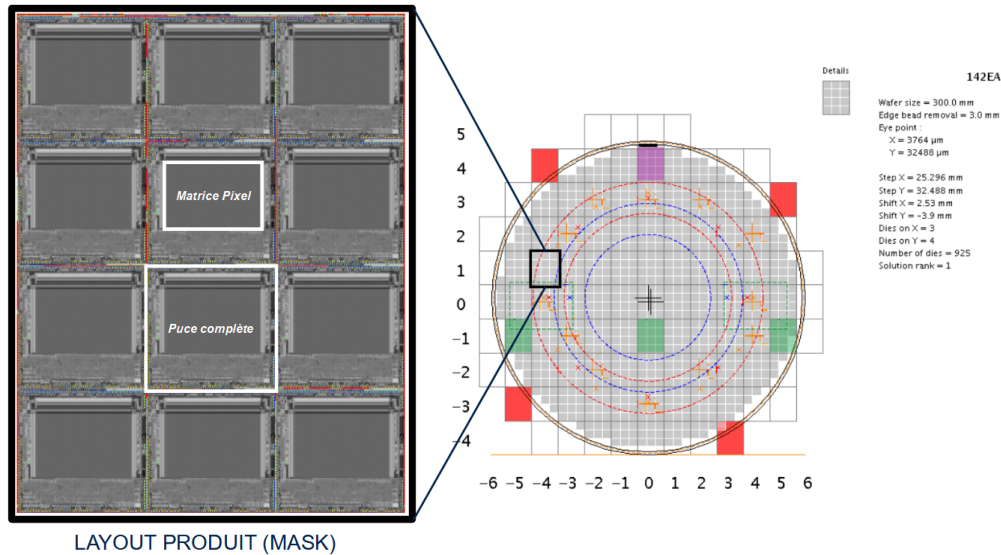


FIGURE 5.4 – Carte d'exposition et layout du produit 142EA

La figure 5.5 montre une mesure imprécise sans revêtement optique. On constate sur cette mesure qu'il y a des zones comportant beaucoup de distorsions, en particulier sur la partie moyenne-haute du wafer où l'on voit deux bandes horizontales. Le résultat théorique attendu sur une mesure précise est un capteur uniforme sur la partie centrale. Sur une mesure imprécise, le résultat théorique attendu n'est pas connu.

Le zoom montre des capteurs comportant des striations, ce qui impacte le fonctionnement des puces. Ces striations sont en revanche difficile à voir à l'œil nu, rendant difficile la détection. Au cours de ce travail, on utilise uniquement la mesure PWG en tant que telle. On se focalise sur la mise en évidence du défaut, mais son lien avec le rendement n'est pas l'objet de ce travail.

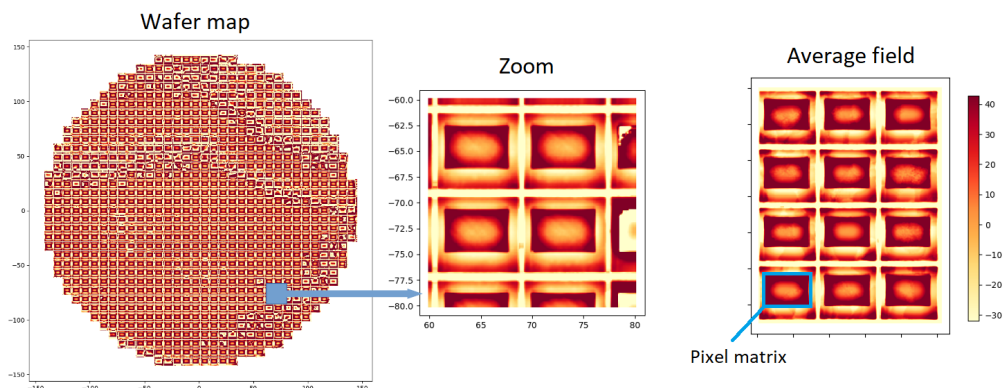


FIGURE 5.5 – Mesure brute sans revêtement avec un zoom là où il y a des striations - *Source interne*

Sur la figure 5.5, on montre également le champ moyen comportant 12 capteurs. On remarque sur cette figure que le champ moyen est une photo-répétition, malgré les erreurs de reconstruction

de surface sur la partie haute du wafer. Ces erreurs sont compensées par le calcul du champ moyen. On montre aussi une matrice de pixel encadrée en bleu, sur laquelle la présence de striations peut impacter le rendement. On met en regard cela avec la figure 5.3, où l'on voit la zone verte précédente comme apparaissant ici dans un rectangle au centre du capteur. Cette zone est la care area à isoler pour le traitement de données.

Pour détecter les striations sur les care areas, on extrait une cartographie du wafer ne comportant que ces zones. L'enjeu est toujours de rendre de plus en plus lisible la mesure de départ de la figure 5.5. On montre la cartographie des care areas sur la figure 5.6, avec un zoom sur une zone censée comporter des striations. À l'œil nu, on distingue difficilement ces striations.

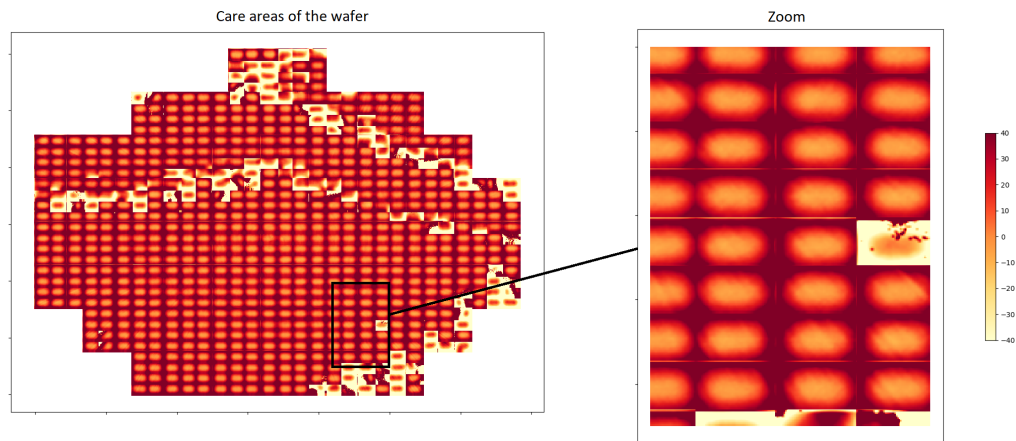


FIGURE 5.6 – Cartographie des Care areas sur les champs complets

Dans notre cas d'étude, on cherche à détecter un défaut. En enlevant la composante moyenne à chacune des care areas, on rendra plus visibles les signaux qui sortent de la moyenne, comme nous l'avons expliqué dans le chapitre 3. On montre le résultat sur la figure 5.7.

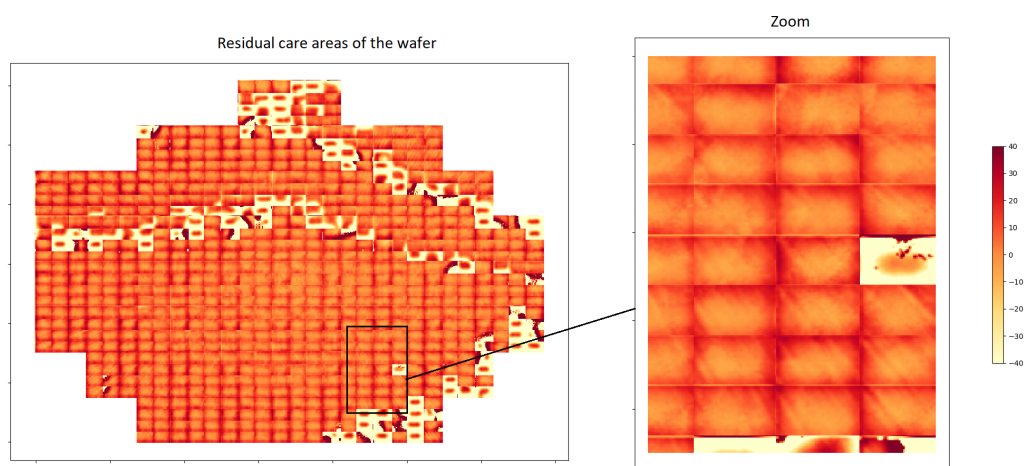


FIGURE 5.7 – Cartographie des Care areas résiduelles sur les champs complets

Sur cette figure 5.7, on distingue bien mieux les striations à l'œil nu. L'enjeu est de rendre visible ce défaut pour le regard humain. Le travail de mise en forme et de visualisation jusqu'à maintenant avait pour finalité de retirer de l'information superflue dans le contexte de la détection de striation. Ces opérations de nettoyage des données sont mentionnées dans les frameworks de traitement de données, notamment dans [Chapman et al., 2000]. Une fois les

données nettoyées au maximum, on entre dans une phase d'exploitation du signal, où l'on va traiter l'information extraite pour rendre visible le défaut.

5.2.3 Mise en valeur des striations

1. Transformations d'images

On cherche à mettre en valeur les striations en transformant les mesures de nanotopographie. Sur la figure 5.8, on montre le résultat de deux algorithmes développés durant la thèse.

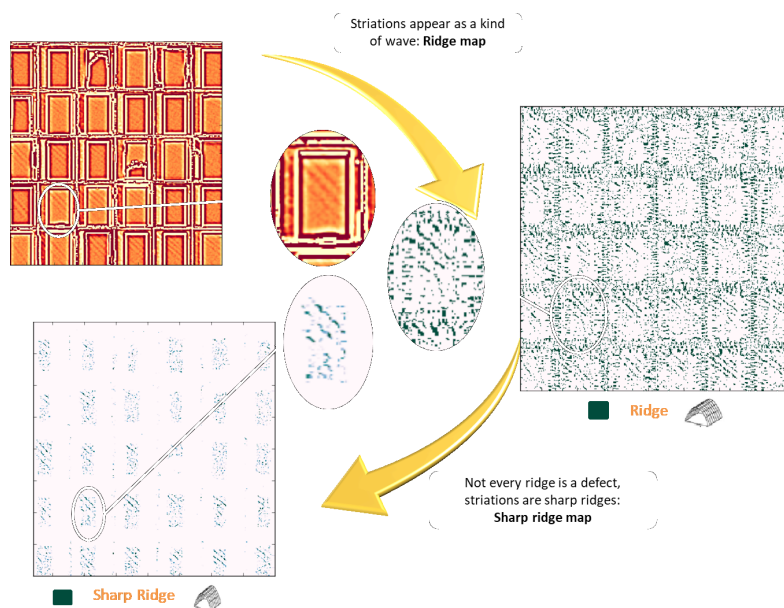


FIGURE 5.8 – Travail présenté à SPIE 2019 - *Source : [Kessar et al., 2019]*

Sur la figure 5.8, il y a 3 images : la première en haut à gauche est la topographie brute sur un autre cas d'étude que la mesure de la figure 5.5, la seconde à droite est la cartographie des crêtes (= Ridges). Cette cartographie est issue de l'algorithme de détection de formes présenté dans le chapitre 3. Dans la mesure où ce que l'on cherche à détecter est une vague, de la forme d'une crête, cette cartographie est pertinente. Elle montre les points au sommet des crêtes sur l'image, donc potentiellement des striations. La cartographie fait ressortir des bandes vertes sur le sommet des striations, et rend l'information plus visible. L'information montrée est binaire : il s'agit d'une classification rangeant les pixels et leur entourage dans la classe « Sommet d'une crête » ou dans une classe « Autre ».

Cette information binaire est dépendante d'un paramétrage de l'algorithme de détection de forme. Pour rappel, on montre sur la table 3.23 l'attribution des formes en fonction de la valeur des dérivées premières et secondes dans les directions des vecteurs propres Hessiens (voir Chapitre 3.4 pour l'algorithme détaillé). On montre le résultat des transformations pour le calcul des valeurs Gradient, L1, L2, V1, V2 qui servent à la détection de formes d'un autre cas d'étude sur la figure 5.9.

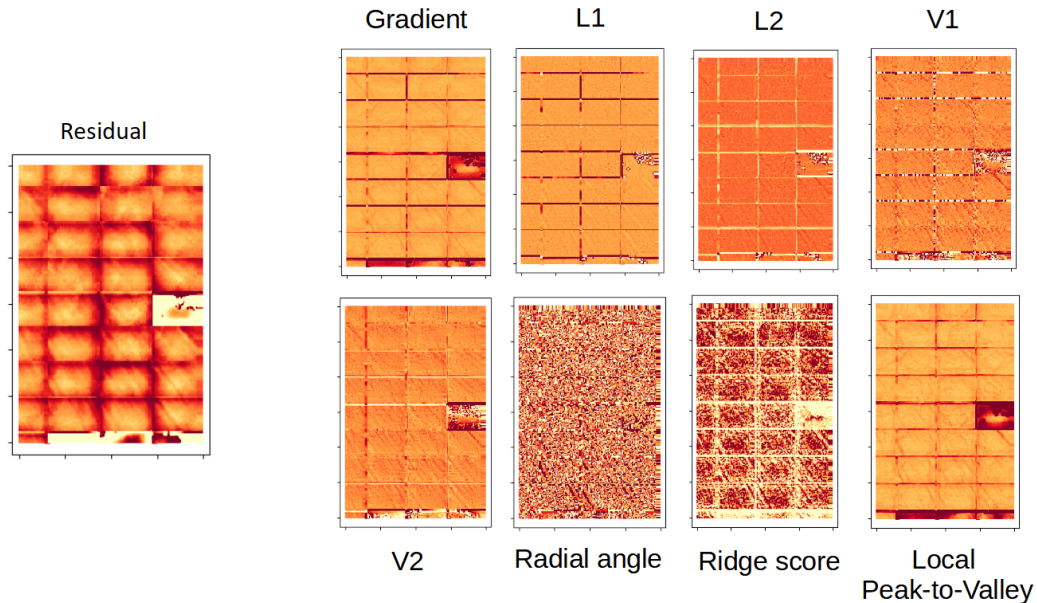


FIGURE 5.9 – Transformations des images

Ces 5 valeurs sont celles qui sont utilisées pour fournir une information binaire quant au fait d’être ou non placé au sommet d’une striation dans l’image de droite de la figure 5.8. Elles sont également utilisées pour le calcul du « Ridge Score » montré sur cette même figure en bas à gauche, qui est une métrique largement intuitive ayant pour but de quantifier à quel point un ensemble de points a la forme d’une crête, et à quel point elle est pentue. Avec le « Ridge score », on observe les striations par une alternance de jaune et noir autour d’une striation sur l’image de la figure 5.9.

La détection de striations a fait l’objet de premiers travaux qui ont été présentés à la conférence SPIE 2019 [Kessar et al., 2019], résumés par la figure 5.8 contenue dans le poster présenté. Ce poster faisait état d’un premier prototype, et les autres transformations de la figure 5.9 n’avaient pas encore été effectuées à ce moment-là.

Le « Radial angle » met en valeur l’angle formé par le vecteur propre de la matrice Hessienne et le vecteur radial du point concerné. Il a été créé pour mettre en avant le fait que les striations sont un phénomène radial. En ce sens, les striations sont parallèles au vecteur radial, et donc le vecteur propre du Hessien aussi. Sur l’image concernée de la figure 5.9, on peut distinguer des lignes quasi-uniformes d’angles sur les sommets des striations, comparativement aux autres points qui sont plutôt anarchiques dans leurs valeurs.

Le « Local Peak-to-Valley » est une métrique mettant en valeur une variation locale. Il s’agit de prendre l’ensemble des points de la cellule et de renvoyer en valeur la différence entre la valeur maximum et la valeur minimum. Une striation est une vague, comportant une variation élevée si l’ensemble de points contient la largeur de la striation. On observe que les valeurs sont effectivement plus élevées, donc les points plus foncés au sommet des striations sur l’image associée.

On montre en Annexe D un travail de classification hiérarchique effectué, afin d’identifier les points au sommet des striations. Ce travail n’est pas inclus dans la rédaction principale du manuscrit car la partie suivante montre un travail plus concluant et prometteur.

Ce travail de transformation d’images est une porte d’entrée pour faire du « Feature engineering », comme nous l’expliquons dans la sous-partie suivante.

5.2.4 Feature engineering

Un travail de thèse à STMicroelectronics sur la détection de striations en utilisant les mesures PWG imprécises est en cours [Alcaire et al., 2021]. Ce travail explore des techniques de deep learning pour détecter les striations, en prenant en données d'entrée les care areas brutes (figure 5.6), et en sortie des données de tests électrique à la fin de fabrication du produit, montrant si oui ou non la puce est commercialisable. Cette correspondance permet d'avoir des données pour entraîner un réseau de neurones profond qui permet de prédire l'impact des striations sur le rendement d'un wafer. La figure 5.10 explique schématiquement le fonctionnement du deep learning mis en place par le travail de la thèse en cours.

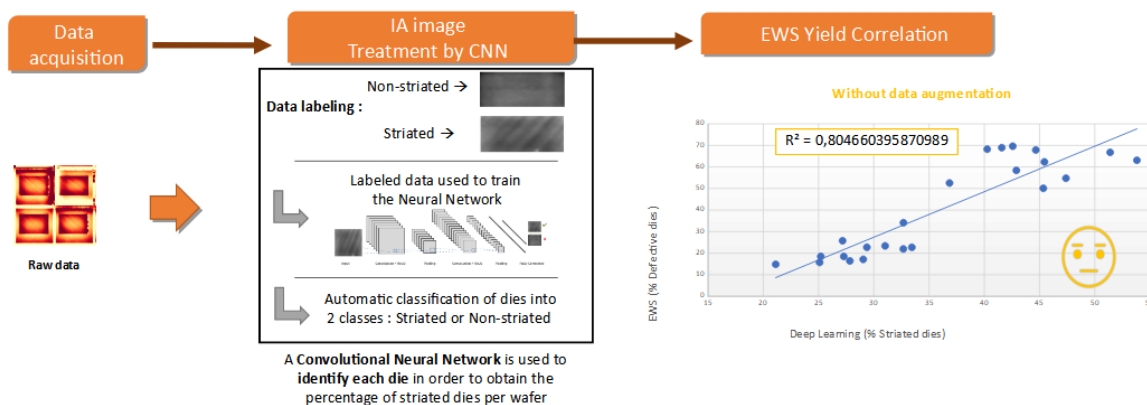


FIGURE 5.10 – Deep learning pour corrélation entre classification de striations et rendement optique. Sources internes : Victor GREDY (Acquisition des données), Thomas ALCAIRE PhD metrology (CNN)

Sur cette figure, les données EWS sont les données permettant de valider le test fonctionnel final du capteur. Il s'agit de tests électriques. Suite à ce test, on obtient l'information pour savoir si oui ou non le capteur peut être commercialisé. Ces données sont utilisées pour étiqueter les images des capteurs précédemment mesurés par l'outil PWG. Ces images PWG étiquetées sont utilisées pour entraîner le réseau de neurones. Le réseau est basé sur ResNet [Szegedy et al., 2017].

Le réseau de deep learning est entraîné par « transfert learning ». La pratique est de prendre un réseau déjà entraîné sur des cas d'usages, et de conserver les premières couches du réseau. Ces couches, au travers de leur similarité dans les cas d'usages, ont vocation à exprimer de différentes façons une image. En un sens, les couches intermédiaires extraient des « features » automatiques. On montre sur la figure 5.11 le paradigme du transfert learning.

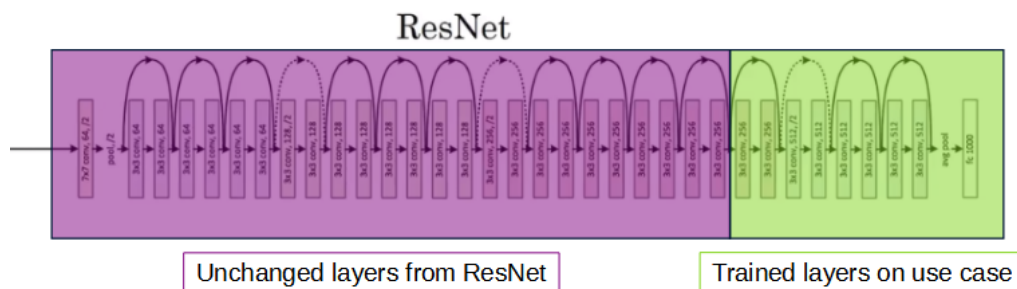


FIGURE 5.11 – Schéma explicatif du « transfert learning »- Source interne : L. Bidault

Sur la figure 5.11, les couches en mauves sont les couches pré-entraînées extraites du réseau de neurones ResNet, qui ne seront pas changées. L'entraînement sur les données PWG étiquetées ne change que les coefficients des quatre dernières couches montrées en vert.

Dans notre cas d'étude, on souhaite transformer des care areas, afin de rendre plus visibles les striations, pour les prendre en entrée dans le réseau de neurones profond pour améliorer les résultats de prédiction. On voit sur la figure 5.10 le résultat sans transformation des entrées (« Without data augmentation »), où la prédiction du nombre de puces prédites correctement (striée ou non striée) atteint un R^2 à 0.80.

Le paradigme de transformation de données pour rendre visible un défaut est complémentaire avec ce type de travail, en tant que « Feature engineering ». Il s'agit, dans la figure 5.10, de ne pas mettre en entrée les données brutes, mais des données transformées. L'idée est de remplacer l'image en entrée du réseau de neurones profond par une couche intelligible utilisant une des transformations contrôlées, au lieu d'un réseau entièrement basé sur de l'apprentissage automatique. On montre sur la figure 5.12 le paradigme utilisant les données transformées.

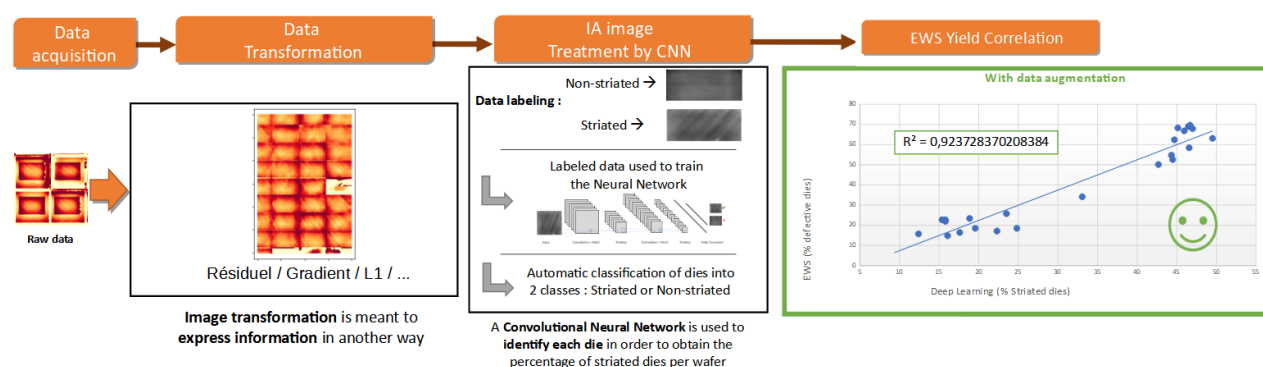


FIGURE 5.12 – Deep learning pour corrélation entre classification de striations et rendement optique avec pré-traitement sur les données. Sources internes : Victor GREDY (Acquisition des données), Mehdi KESSAR (pré-traitement), Thomas ALCAIRE PhD Metrology (CNN)

Les transformations proposées pour aider le réseau à améliorer la lisibilité des données de calibration sont exactement ceux présentées dans la figure 5.9. Le calcul de ces vecteurs est détaillé dans l'annexe B. Le résultat de cette modification a montré une amélioration (voir figure 5.12), en passant d'un R^2 à 0.8 (sans transformation) à un R^2 à 0.92 pour le Résiduel, et à 0.91 pour la valeur L1.

5.3 Conclusion

Les travaux de transformations réalisés sont complémentaires avec l'utilisation d'un réseau de neurones. Ce travail a montré la pertinence du choix technique du feature engineering conjoint à l'utilisation du deep learning déjà très performant (R^2 à 0.80), en augmentant significativement la détection des puces striées (R^2 à 0.92).

Les couches pré-entraînées sur d'autres cas d'usages du réseau ResNet ont pour but de décomposer une image d'une façon « universelle », ou en tout cas s'approchant d'une telle décomposition au travers diverses couches intermédiaires.

Une perspective dans le cas d'étude des striations serait d'expliquer l'augmentation aussi importante des résultats par ces transformations, car c'est en théorie le rôle des couches intermédiaires pré-entraînées d'exprimer l'information contenue dans une image. On pourrait par exemple regarder le contenu de ces couches pré-entraînées en détail pour évaluer la signification de celles-ci.

Une autre perspective, qui peut s'appliquer de façon générique à d'autres cas d'études, est d'intégrer la totalité des transformations proposées dans ce chapitre pour en faire une couche

supplémentaire du réseau de neurones. Le but est d'utiliser leurs apports de façon conjointe. La figure 5.13 montre cette perspective.

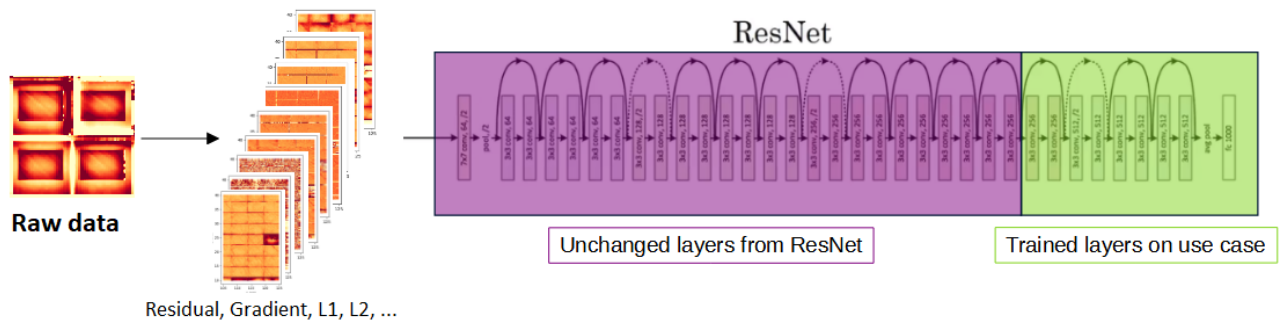


FIGURE 5.13 – Schéma explicatif du « transfert learning » combiné à une couche supplémentaire greffée

Le feature engineering est à privilégier à l'utilisation de réseaux de neurones purs. Cela permet un compromis intéressant entre performance et intelligibilité, où le réseau de neurones fait office de « boîte noire » de performance, et les features choisies font office de couche « compréhensible ». Ces enjeux font référence à ceux mis en avant dans le chapitre 4, où l'on souhaite concilier robustesse, performance, simplicité et compréhension. En choisissant des features qui ont du sens pour le contexte (comme le résiduel pour détecter), on rend simple et compréhensible l'information. En utilisant le réseau de neurones, on rend performant et robuste le résultat.

Conclusion

1. Problématique et résumé du plan

La problématique de cette thèse est d'insérer la mesure de topographie de l'outil PWG de KLA dans l'écosystème des données déjà existantes de STMicroelectronics. Insérer veut dire valoriser au maximum ce type de mesure, d'une part en se familiarisant aux métiers présents sur le site de fabrication, et d'autre part en ayant un point de vue informatique sur ces données disponibles pour sensibiliser les métiers aux traitements possibles.

En présentant la volumétrie des données (Chapitre 1), puis le contexte industriel (Chapitre 2), le besoin de nouvelles compétences et de nouveaux métiers en traitement de données a été mis en lumière.

Les axes de travail développés dans ce manuscrit sont thématiques, en commençant par le démêlage des composantes (Chapitre 3), puis en reliant une composante de la mesure à ses déterminants en réalisant un modèle (Chapitre 4) et enfin en liant l'une des composantes de la mesure à un problème de détection de défauts déjà en cours de résolution à STMicroelectronics (Chapitre 5).

Chacun des trois axes de travail a vocation à aider à la décision en rendant intelligible l'information contenue dans les données, et cela valorise également l'expertise métier grâce à la capacité à relier cette information aux procédés de fabrication.

2. Travail effectué

Beaucoup de points techniques développés dans cette thèse sont basés sur des transformations d'images (Filtrage, extraction des composantes...). La figure récapitulative 5.14 montre le bilan des travaux effectués. La légende de cette figure est la suivante :

1. Les encadrés rouges sont les données extérieures au travail de cette thèse : en particulier, la mesure de topographie du PWG telle qu'elle est fournie, les densités périmétriques et surfaciques brutes, la cartographie de rendement, et le travail de thèse (T. Alcaire) utilisant le deep learning pour résoudre la détection des striations.
2. Les encadrés verts générés par des flèches vertes sont le premier travail de cette thèse, avec des transformations des données. En particulier, la décomposition de la mesure brute du PWG en champs moyens, puces moyennes, résiduels, variabilités... L'encadré isolé dans la droite est le travail de description de topographie visant à caractériser la mesure pour les procédés de fabrication.
3. Les doubles flèches bleues correspondent à la cristallisation de ces manipulations, et sont les étapes de valorisation les plus importantes. Il s'agit des liens entre les manipulations et les données existantes à STMicroelectronics. C'est à ce moment-là que le travail prend tout son sens. En particulier, la double flèche bleue en bas à gauche de la figure correspond au travail du chapitre 4, où des modèles de nano-topographie ont été réalisés, et où les enjeux d'une prédiction efficace ont été mis en lumière. La mise en regard entre la cartographie de rendement et les variations de topographie est montré par la double flèche bleue en haut à droite. L'utilisation des transformations de données pour les connecter au réseau

de neurones et améliorer les résultats de prédiction (R^2 à 0.92 au lieu de 0.80) est montré par la double flèche bleue en bas à droite.

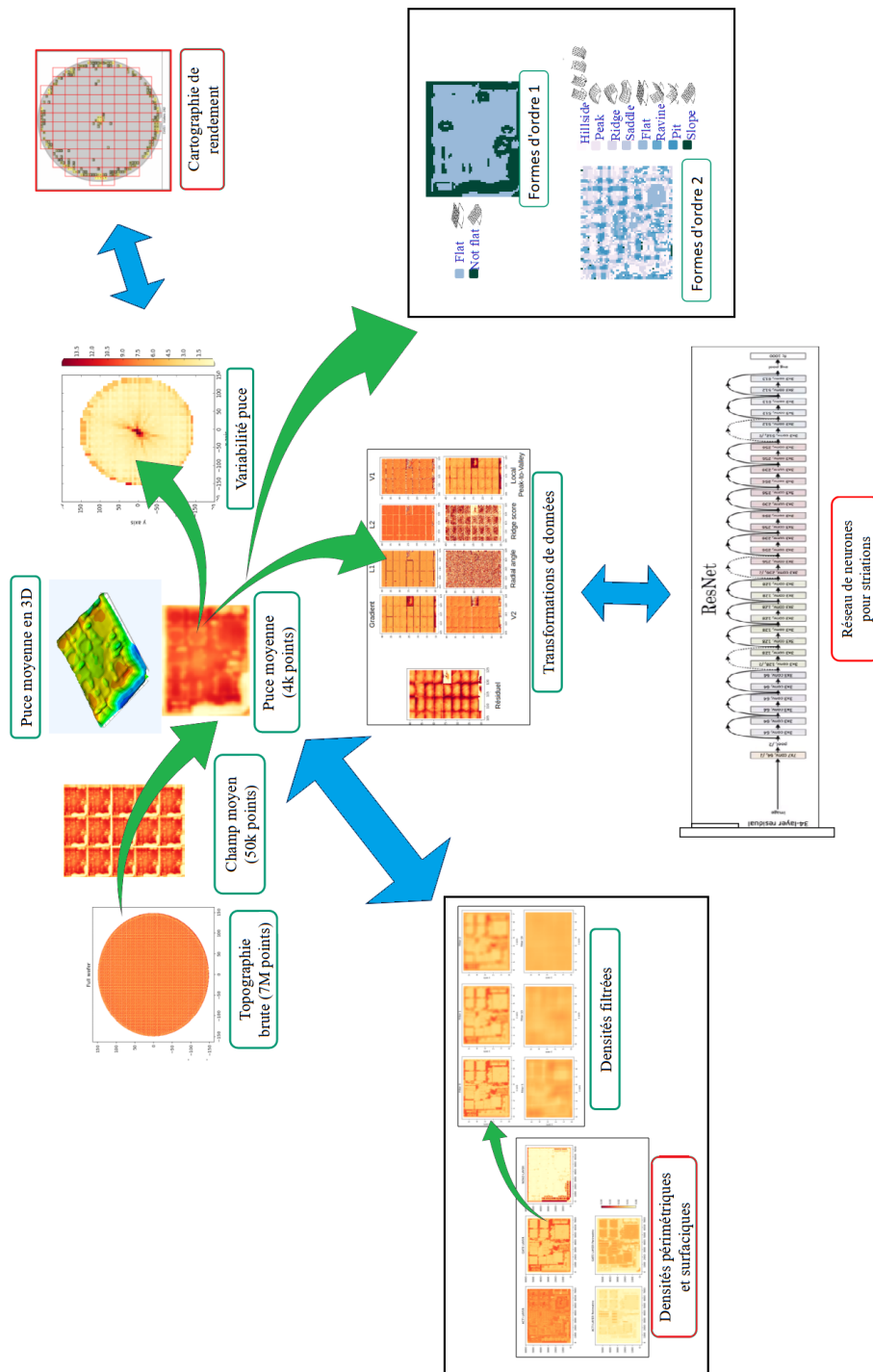


FIGURE 5.14 – Figure récapitulative du travail effectué

Ces liens sont à établir dans les sciences de données. De nombreux objets physiques sont représentés par des objets informatiques via les données (données logiques, et mesures physiques). En identifiant les liens entre ces entités, on rend évidents les liens des entités physiques, qui, au bout du compte, sont les plus essentielles dans une industrie. Au final, l'objectif reste de

produire des puces avec un rendement optimal, en comprenant les procédés de l'industrie grâce aux données et aux connaissances physiques.

Comprendre ces liens avec un point de vue physique puis les mettre en œuvre avec du développement informatique, ou bien la démarche inverse de décomposer une mesure informatique pour les lier à d'autres mesures permet de mieux comprendre l'industrie, et d'aider à prendre de meilleures décisions.

3. Bilan

Cette thèse est à cheval entre plusieurs disciplines : l'informatique, les mathématiques et les procédés de fabrication. Le profil du doctorant choisi (moi-même) était purement mathématiques et informatiques, sans connaissance de la micro-électronique. Cette perspective d'intégration de profils informatiques dans de nouvelles branches de la micro-électronique, à savoir le côté gestion d'un procédé, est nouveau pour cette industrie.

Cette nouveauté est au cœur de difficultés humaines, de compréhension et (parfois) de sens. La technicité du point de vue du développement informatique n'est pas le facteur limitant dans la réalisation de ce type de travail, mais bien des difficultés liées aux différences de profils et de connaissances.

Cette thèse s'adresse à toute personne souhaitant faire du traitement de données dans le domaine de la micro-électronique, qu'elle soit issue d'une formation physique-chimie comme mathématiques ou informatique, abordant de nombreux aspects du traitement de données. Cette thèse aborde de nombreux aspects du traitement de données, en détaillant les enjeux de chaque aspect sans aller dans une technicité profonde trop rapidement. Les éléments abordés permettent d'effectuer des choix techniques assurant une industrie pérenne : compréhension, performance et robustesse.

4. Perspectives

La poursuite du travail spécifique d'analyse des mesures de topographie contient plusieurs points. Le premier est de continuer à expliciter et caractériser les composantes de la topographie d'une puce, d'un champ, ou bien d'un wafer, pour faire de l'aide à la décision sur des mesures brutes, aussi bien pour du suivi de procédés comme pour de la R & D. Le second point est sur le modèle de topographie d'une puce, où des éléments d'importance ont été identifiés : simplicité du modèle, performance du modèle, et anticipation de la robustesse. Sur le dernier point concernant le feature engineering, les transformations peuvent être essayées sur d'autres cas d'études, ou bien intégrées au réseau de neurones dans leur intégralité comme couche supplémentaire. On peut également envisager de chercher d'autres transformations efficaces pour le cas d'étude des striations comme pour d'autres cas d'études.

Au-delà du travail spécifique réalisé qui concerne la topographie, cette thèse a tenté d'apporter un regard sur des problèmes d'intégration de mesures dans l'écosystème des données du site de fabrication, en souhaitant qu'elle permettra à des scientifiques de tous métiers de la micro-électronique de se familiariser aux problématiques de traitement de données dans cette même industrie, en ayant tous les points d'importance abordés en tête.

Annexe A

Échantillonnage aléatoire

Au cours du chapitre 4, nous avons sélectionné des quarts de puces pour évaluer la représentativité d'un échantillonnage sur cette même puce. Nous avons également sélectionné des quart aléatoires pour voir leur représentativité au sein d'une puce. Les résultats étaient que les quarts aléatoires étaient de très bons représentants de chacune des puces.

La démarche dans cette annexe est de vérifier à partir de quand chacune des sélections devient représentative de la puce elle-même. Pour ce faire, on prend un ensemble de proportions variant entre 0 et 1. Pour chacune de ces proportions, on réalise 10 sélections aléatoires différentes et on calibre les modèles 1 et plein sur chacune des sélections.

Pour représenter cet ensemble de 10 R^2 sur un graphique, la fonction boxplot est utilisée [Chambers J. M. and A., 1983]. Autrement appelée « boîte à moustache » due à sa forme, c'est un moyen synthétique de représenter un ensemble de données sur un axe, détaillé sur la figure A.1. L'ensemble de données représenté sur chaque abscisse correspond ici à l'ensemble des 10 R^2 de la sélection concernée.

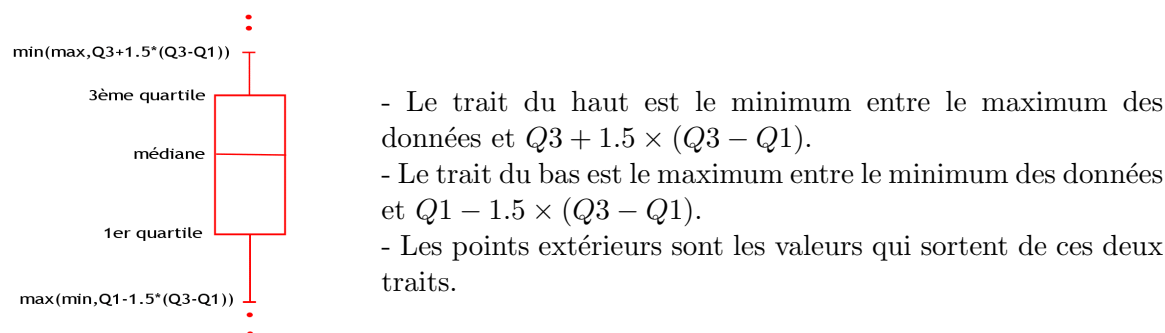
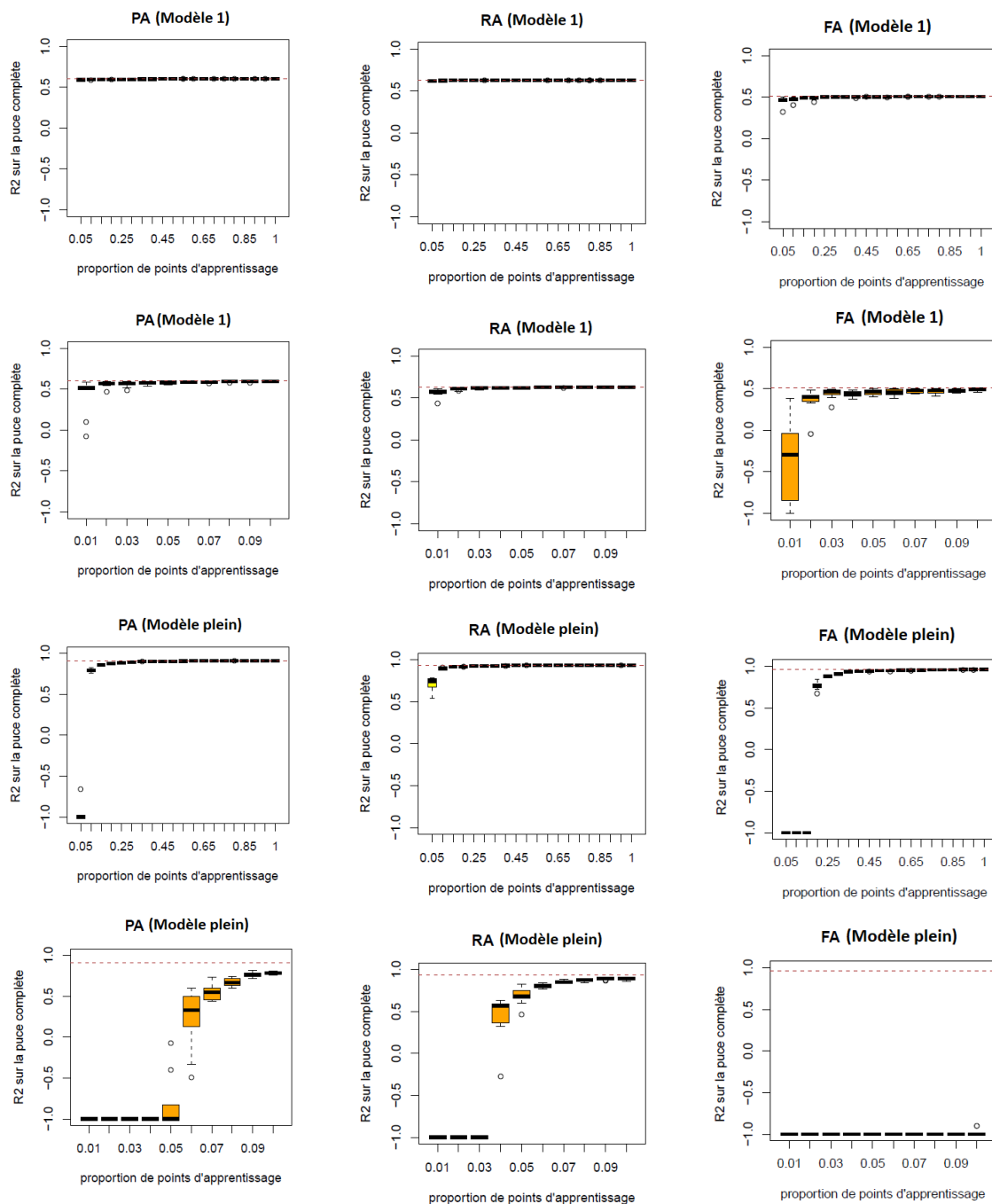


FIGURE A.1 – Boxplot

Sur la figure A.2, on souhaite montrer à partir de quelle proportion de sélection de la puce celle-ci atteint une représentativité approchant le meilleur des cas, à savoir une calibration sur la puce complète. On montre 4 graphiques sur chacune des trois puces (281PA à gauche, 281RA au milieu, 282FA à droite). Deux de ces graphiques concernent le modèle 1 (au-dessus), et les deux autres concernent le modèle plein (en-dessous). Sur chacun des modèles, le graphique du haut est un échantillonnage allant de 0.05 à 0.95, et le graphique du haut a vocation à être plus précis, en parcourant un intervalle allant de 0.01 à 0.1.



(a) 281PA

(b) 281RA

(c) 282FA

FIGURE A.2 – Modélisations calibrées sur des échantillons aléatoires des puces sur les modèles 1 et plein

Sur ces graphiques, on représente en ordonnée la boîte à moustaches des 10 R^2 de la proportion concernée (on a limité la valeur minimale du R^2 à -1 pour plus de clarté), et en abscisse la proportion de la sélection aléatoire de la puce.

On observe qu'à partir de 0.1, les puces 281PA et 281RA sont bien représentées par leurs sélections sur le modèle plein, et ce sans beaucoup de variations. La représentativité de la puce sur le modèle 1 est en revanche atteinte à partir de 0.04 et ce sans variations sur les R^2 également.

La puce 282FA se distingue encore une fois ici. Il est représentatif de lui-même sur le modèle

plein sans beaucoup de variations à partir de proportions de 0.3. Il est représentatif sur le modèle 1 sans beaucoup de variation à partir de 0.1. Cependant, on observe tout de même des variations faibles, mais visibles pour le modèle 1 même pour des proportions dépassant 0.5.

Annexe B

Détails de calcul pour les transformations

On souhaite ici donner les détails du calcul de ces différentes features, en particulier du Radial angle, du Ridge Score, et du Local Peak-to-Valley. Nous sommes dans le cas déjà défini, où l'on extrait une cellule prenant un pixel et son entourage. Dans cette partie, on détaille la transformation entre cette cellule et l'une de ces features.

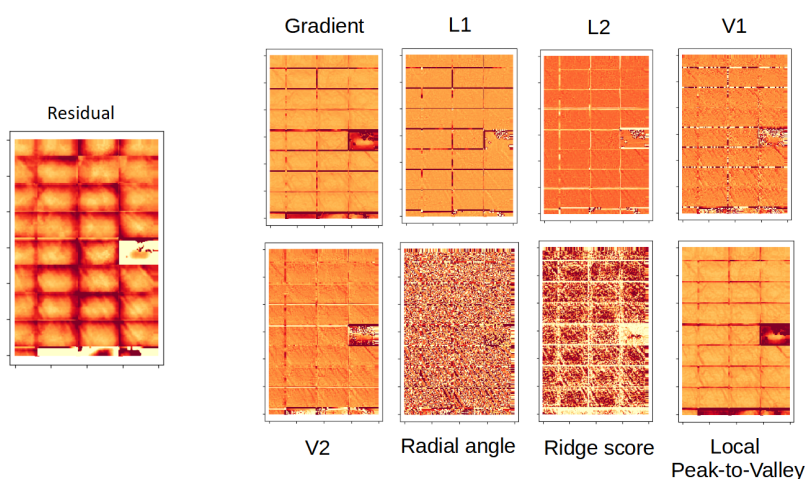


FIGURE B.1 – Transformations des images (répétition de la figure 5.9)

Les calculs des vecteurs gradient, L1, L2, V1, V2 sont déjà expliqués dans la partie 3.4, aussi nous expliquons uniquement les 3 restants ici. On ré-utilise les notations précédentes également pour les vecteurs propres : u_1, u_2 .

1. Radial angle

On nomme (X, Y) les coordonnées du pixel au centre de la cellule. On calcule l'angle formé entre le vecteur (X, Y) et le vecteur u_1 pour obtenir le Radial angle.

Au sommet d'une striation, le vecteur u_1 est orienté en longeant la crête de la striation. La striation étant un phénomène radial, la valeur du Radial angle est proche de 0 pour un pixel au sommet d'une striation. Cette transformation a été pensée pour faire ressortir ce phénomène radial.

2. Ridge score

Cette valeur se base sur l'une des définitions du Ridge suivant le papier [Watson et al., 1985] :
($M = 0, L1 < 0, L2 = 0, V1 = 0, V2 = 0$).

$$M_{coef} = \min\left(1, \frac{1}{1 + \frac{|M|}{M_{seuil}}}\right)$$

$$L1_{coef} = \max\left(0, 1 - e^{\frac{L1}{L_{seuil}}}\right)$$

$$L2_{coef} = \min\left(1, \frac{1}{1 + \frac{|L2|}{L_{seuil}}}\right)$$

$$V1_{coef} = \min\left(1, \frac{1}{1 + \frac{|V1|}{V_{seuil}}}\right)$$

$$V2_{coef} = \min\left(1, \frac{1}{1 + \frac{|V2|}{V_{seuil}}}\right)$$

L'idée ici est de générer des valeurs pour chacun des vecteurs comprises entre 0 et 1, avec pour propriété : plus le critère associé au vecteur est validé, plus on est proche de 1.

Le choix des valeurs « seuils » est dépendant du contexte, et plus ces valeurs sont élevées, plus l'intervalle au sein duquel les valeurs peuvent être considérées comme nulles est grand. C'est un choix technique qui est le même que lorsqu'on définit un intervalle de nullité dans la version « discrète » de la qualification d'une crête.

$$RidgeScore = M_{coef} + L1_{coef} + L2_{coef} + V1_{coef} + V2_{coef}$$

Une critique de cette métrique est qu'un point avec un entourage très plat ($M = 0, L1 = 0, L2 = 0, V1 = 0, V2 = 0$) aura un Ridge Score approchant 4, tandis qu'il n'est pas proche d'une striation. Un point avec ce profil ($M > 0, L1 < 0, L2 = 0, V1 < 0, V2 = 0$) aura un Ridge Score approchant 3, alors qu'il est plus à même d'être au sommet d'une striation avec ces valeurs.

3. Local Peak-to-Valley

Ici, on renvoie la valeur du Peak-to-Valley de la cellule : la valeur maximale de la cellule moins sa valeur minimale.

Annexe C

Échantillonnage pour prédiction

Un paradigme dans la continuité du protocole établi au chapitre 4 est de sélectionner non pas des ensembles de puces, mais de « créer » une puce fictive. Lorsqu'un nouveau produit arrive, on crée une puce fictive de même dimension que le nouveau produit dont chaque point est le meilleur représentant parmi l'ensemble des puces dont on dispose la nano-topographie.

L'algorithme heuristique est le suivant :

1. Un nouveau produit P_0 arrive, ayant N pixels. On dispose de ses densités périmétriques et surfaciques, mais pas de sa nano-topographie. On dispose également de P produits.
2. Pour chacun des N pixels de P_0 , on cherche dans l'ensemble des points des P produits celui qui est le plus « proche » selon une distance choisie dans l'espace des densités. On peut par exemple choisir la distance L2. Ce point est le « représentant » du pixel de P_0 . On obtient une puce fictive de N pixels avec le meilleur représentant pour chacun des points.
3. On calibre le modèle 5 sur cette puce, et on prend la prédiction de P_0 .

On montre le résultat de cette heuristique sur la figure C.1. On se place dans les 3 cas où un seul des trois produits dont on dispose est inconnu, et les deux autres sont connus. Les autres cas ne sont pas pertinents pour notre étude. La distance choisie pour évaluer les distances entre les points est la distance L2, sur laquelle on a normalisé chacun des vecteurs d'entrée.

Produits de calibration	Produit d'inférence	Modèle 1			Modèle 5		
		R ²	Nb vecteurs	PV diff	R ²	Nb vecteurs	PV diff
281PA, 281RA	282FA	0,18	5	-3,51	0,62	20	-1,91
281PA, 282FA	281RA	0,61	5	-3,65	0,84	20	0,69
281RA, 282FA	281PA	0,5	5	-3,05	0,71	20	-0,27

Peak-to-Valley (nm)	
281PA	17,39
281RA	18,28
282FA	13,97

Modèle 1		Modèle linéaire sans filtre	
Modèle 5		Modèle linéaire avec filtres avec puces voisines de portés [0,1,5,15]	

Nb Vecteurs	R ²	PV diff abs.
5	≤ 0,2	< 2
5 à 15	0,2-0,35	2 – 5
15 à 25	0,35-0,5	5 – 10
25 à 40	0,5-0,7	10 – 15
> 40	> 0,7	> 15

FIGURE C.1 – Tableau récapitulatif de l'heuristique présentée

Si l'on ne normalise pas les vecteurs d'entrée, on rend « insignifiants » les vecteurs de densités périmétriques, car les échelles ne sont pas du tout les mêmes.

Cette heuristique est l'une des nombreuses possibilités pour exploiter une masse de points représentant la puce du produit P_0 entrant en production. Les résultats sont peu concluants, aussi cette heuristique n'a pas été intégrée au manuscrit.

En montrant cette heuristique simple, le message est qu'on peut chercher des algorithmes pour assurer robustesse grâce aux distances dans l'espace métrique des densités, et performance dans la modélisation en utilisant ce genre de technique. Cette heuristique s'apparente à un « plus proche voisin », mais n'en est pas un car on ne renvoie pas le résultat de la topographie du plus proche voisin.

Annexe D

Classification hiérarchique

L'utilisation des transformations de données du chapitre 5 crée un espace métrique de dimension $N = 8$, où N est le nombre de caractéristiques par cellule. Dans cet espace, il était supposé au travers des algorithmes de détections de formes que nous pouvions identifier les points au sommet des striations. Ce raisonnement a été présenté par la détection de crêtes du papier [Watson et al., 1985] dans le manuscrit.

Une tentative a été effectuée, en faisant de la classification hiérarchique [Pedregosa et al., 2011], dont on montre le résultat sur la figure D.1. Dans notre cas particulier, on s'est placé dans l'espace métrique $(M, L1, L2, V1, V2, RadialAngle)$ dont les éléments sont mentionnés dans le chapitre 5, car ces valeurs ont été les plus pertinentes parmi les transformations effectuées.

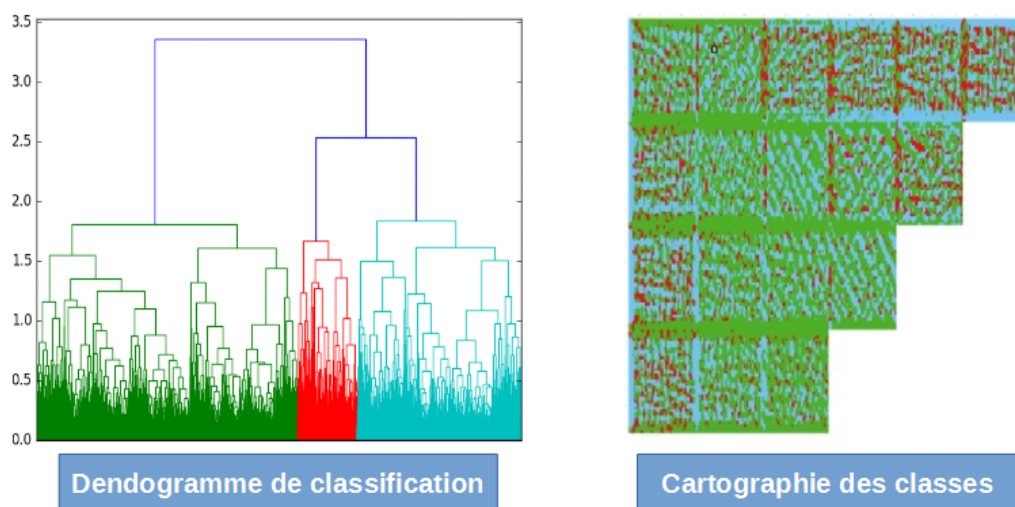


FIGURE D.1 – Classification hiérarchique sur un coin de wafer Sud-Est

Ce qui s'observe sur cette classification, est que la classe bleue ciel correspond à peu près aux sommets des striations. Cette piste n'a pas été poursuivie, car le feature engineering combiné à un réseau de neurones a donné des résultats plus prometteurs, en plus d'avoir une rapidité de calcul plus intéressante. La classification de 7 millions de points est extrêmement longue, en particulier si l'on utilise une distance de type « moyenne » pour la distance inter-classe. Le temps de calcul diminue lors de l'utilisation de la distance inter-classe « maximum » ou « minimum ».

Bibliographie

- [Alcaire et al., 2021] Alcaire, T., Cunff, D. L., Tortai, J.-H., Soulan, S., Greedy, V., Templier, M., Kessar, M., Bianchini, R., and Berthoud, A. (2021). Detection and correlation of yield loss induced by color resist deposition deviation with a deep learning approach applied to optical acquisitions. In Adan, O. and Robinson, J. C., editors, *Metrology, Inspection, and Process Control for Semiconductor Manufacturing XXXV*, volume 11611, pages 467 – 475. International Society for Optics and Photonics, SPIE.
- [Bergounioux, 2011] Bergounioux, M. (2011). Quelques méthodes de filtrage en Traitement d’Image. Cours donné dans le cadre d’une école CIMPA - en attente de publication dans les actes.
- [Bunday et al., 2013] Bunday, B., Germer, T. A., Vartanian, V., Cordes, A., Cepler, A., and Settens, C. (2013). Gaps analysis for CD metrology beyond the 22nm node. In Starikov, A. and Cain, J. P., editors, *Metrology, Inspection, and Process Control for Microlithography XXVII*, volume 8681, pages 870 – 898. International Society for Optics and Photonics, SPIE.
- [Chambers J. M. and A., 1983] Chambers J. M., Cleveland W. S., K. B. and A., T. P. (1983). *Graphical methods for data analysis*. Belmont, Calif : Wadsworth International Group.
- [Chapman et al., 2000] Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., and Wirth, R. (2000). CRISP-DM 1.0 Step-by-step data mining guide. Technical report, The CRISP-DM consortium.
- [de Caunes et al., 2010] de Caunes, J., Ferreira Porto Rosa, A., Kernaflen, C., Lam, A., Oussar, Y., Pages, G., and Pasqualini, F. (2010). Virtual metrology modeling for predicting overlay of photolithography process after a maintenance action. *Conference : ASMC 2010*.
- [Demésy, 2009] Demésy, G. (2009). *Modélisation électromagnétique tri-dimensionnelle de réseaux complexes. Application au filtrage spectral dans les imageurs CMOS*. Theses, Université Paul Cézanne - Aix-Marseille III.
- [Deza and Deza, 2009] Deza, M. and Deza, E. (2009). *Encyclopedia of Distances*. Encyclopedia of Distances. Springer Berlin Heidelberg.
- [Fagerland and Sandvik, 2009] Fagerland, M. W. and Sandvik, L. (2009). The wilcoxon–mann–whitney test under scrutiny. *Statistics in medicine*, 28(10) :1487–1497.
- [Gerville-Réache and Couallier, 2011] Gerville-Réache, L. and Couallier, V. (2011). Échantillon représentatif d’une population finie : Définition statistique et propriétés. Échantillon représentatif, Sondage, Quotas, Probabilités d’inclusion.
- [GIMP, 2021] GIMP (2021). Chapter 17 : Filters. <https://docs.gimp.org/2.10/fr/filters.html>.
- [Glasner et al., 2009] Glasner, D., Bagon, S., and Irani, M. (2009). Super-resolution from a single image. In *IEEE International Conference on Computer Vision*, pages 349 – 356.
- [Guesmia, 2017] Guesmia, K. Y. (2017). La reconnaissance de forme dans le traitement d’images. *University of Eloued*.
- [Hastie and Pregibon, 1992] Hastie, T. J. and Pregibon, D. (1992). *Generalized linear models*. Boca Raton.

- [Hodges, 1958] Hodges, J. L. (1958). The significance probability of the smirnov two-sample test. *Ark. Mat.*, 3(5) :469–486.
- [Jasiewicz and Stepinski, 2013] Jasiewicz, J. and Stepinski, T. F. (2013). Geomorphons — a pattern recognition approach to classification and mapping of landforms. *Geomorphology*, 182 :147 – 156.
- [KADOK, 2019] KADOK, P. (2019). Métrologie industrielle et scientifique. *Bureau luxembourgeois de métrologie*. <https://portail-qualite.public.lu/fr/metrologie/industrielle-scientifique.html>.
- [Kaur and Gill, 2019] Kaur, J. and Gill, N. (2019). *Artificial Intelligence and Deep Learning for Decision Makers : A Growth Hacker’s Guide to Cutting Edge Technologies*. BPB Publications.
- [Kessar et al., 2019] Kessar, M., Le-Gratiet, B., Lemaire, P., Brouzet, V., Cunff, D. L., and Gredy, V. (2019). Analysis and modelling of patterned wafer nano-topography using multiple linear regression on design GDS and silicon PWG data. In Ukraintsev, V. A. and Adan, O., editors, *Metrology, Inspection, and Process Control for Microlithography XXXIII*, volume 10959, pages 581 – 588. International Society for Optics and Photonics, SPIE.
- [Kim et al., 2020] Kim, J., Murali, G., Vanna-Iampikul, P., Lee, E., Kim, D., Chaudhuri, A., Banerjee, S., Chakrabarty, K., Mukhopadhyay, S., and Lim, S. K. (2020). Rtl-to-gds design tools for monolithic 3d ics. In *2020 IEEE/ACM International Conference On Computer Aided Design (ICCAD)*, pages 1–8.
- [Lakcher, 2018] Lakcher, A. (2018). *Nouvelles perspectives de métrologie dimensionnelle par imagerie de microscope électronique pour le contrôle de la variabilité des procédés de fabrication des circuits intégrés*. PhD thesis, Université Grenoble Alpes (ComUE) 2018.
- [Lakcher et al., 2017] Lakcher, A., Le-Gratiet, B., Ducoté, J., Fanton, P., Kiers, T., Gemmink, J.-W., Hunsche, S., Prentice, C., and Besacier, M. (2017). Robust 2D patterns process variability assessment using CD-SEM contour extraction offline metrology. In Sanchez, M. I., editor, *Metrology, Inspection, and Process Control for Microlithography XXXI*, volume 10145, pages 389 – 402. International Society for Optics and Photonics, SPIE.
- [Lam et al., 2015] Lam, A., Ypma, A., Gatefait, M., Deckers, D., Koopman, A., van Haren, R., and Beltman, J. (2015). Pattern recognition and data mining techniques to identify factors in wafer processing and control determining overlay error. In Cain, J. P. and Sanchez, M. I., editors, *Metrology, Inspection, and Process Control for Microlithography XXIX*, volume 9424, pages 545 – 554. International Society for Optics and Photonics, SPIE.
- [Le-Gratiet et al., 2019] Le-Gratiet, B., Bouyssou, R., Ducoté, J., Dezaudier, C., Ostrovsky, A., Beylier, C., Gardin, C., Petroni, P., Milléquant, M., Chagoya-Garzon, A., and Schiavone, P. (2019). Contour based metrology : getting more from a SEM image. In Ukraintsev, V. A. and Adan, O., editors, *Metrology, Inspection, and Process Control for Microlithography XXXIII*, volume 10959, pages 332 – 340. International Society for Optics and Photonics, SPIE.
- [Le-Gratiet et al., 2020] Le-Gratiet, B., Bouyssou, R., Ducoté, J., Ostrovsky, A., Beylier, C., Gardin, C., Schuch, N., Annezo, V., Schneider, L., Millequant, M., Petroni, P., Figueiro, T., and Schiavone, P. (2020). Contour based metrology : “make measurable what is not so”. In Adan, O. and Robinson, J. C., editors, *Metrology, Inspection, and Process Control for Microlithography XXXIV*, volume 11325, pages 9 – 19. International Society for Optics and Photonics, SPIE.
- [LeCun et al., 2015] LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *nature*, 521(7553) :436–444.
- [Lee et al., 2017] Lee, J., Bello, A. F., Kakita, S., Pieniazek, N., and Johnson, T. A. (2017). Surface topography analysis and performance on post-CMP images (Conference Presentation). In Sanchez, M. I., editor, *Metrology, Inspection, and Process Control for Microlithography XXXI*, volume 10145, pages 303 – 303. International Society for Optics and Photonics, SPIE.

- [Li et al., 2009] Li, X., Rutenbar, R., and Blanton, R. (2009). Virtual probe : A statistically optimal framework for minimum-cost silicon characterization of nanoscale integrated circuits. In *ACM International Conference on Computer-Aided Design*, pages 433–440.
- [McClellan, 2020] McClellan, B. (2020). Transistor count trends continue to track with Moore’s law. *IC Insight*.
- [Mitchell and Netravali, 1988] Mitchell, D. P. and Netravali, A. N. (1988). Reconstruction filters in computer-graphics. *SIGGRAPH Comput. Graph.*, 22(4) :221–228.
- [Mokarram et al., 2015] Mokarram, M., Roshan, G., and Negahban, S. (2015). Landform classification using topography position index (case study : salt dome of Korsia-Darab plain, Iran). *Modeling Earth Systems and Environment*, 1.
- [Moore, 2006] Moore, G. E. (2006). Progress in digital integrated electronics [technical literature, copyright 1975 ieee. reprinted, with permission. technical digest. international electron devices meeting, ieee, 1975, pp. 11-13.]. *IEEE Solid-State Circuits Society Newsletter*, 11(3) :36–37.
- [Moyne and Iskandar, 2017] Moyne, J. and Iskandar, J. (2017). Big data analytics for smart manufacturing : Case studies in semiconductor manufacturing. *Processes*, 5.
- [Owens et al., 2008] Owens, J. D., Houston, M., Luebke, D., Green, S., Stone, J. E., and Phillips, J. C. (2008). GPU computing. *Proceedings of the IEEE*, 96(5) :879–899.
- [Parodi, 1965] Parodi, M. (1965). *Mathématiques appliquées à l’art de l’ingénieur*. Number vol. 3 in *Mathématiques appliquées à l’art de l’ingénieur*. Société d’édition d’enseignement supérieur.
- [Pedregosa et al., 2011] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. (2011). Scikit-learn : Machine learning in Python. *Journal of Machine Learning Research*, 12 :2825–2830.
- [Rakotomalala, 2005] Rakotomalala, R. (2005). Arbres de décision. *Revue Modulad*, 33 :163–187.
- [Reid Turner et al., 1999] Reid Turner, C., Fuggetta, A., Lavazza, L., and Wolf, A. L. (1999). A conceptual basis for feature engineering. *Journal of Systems and Software*, 49(1) :3 – 15.
- [Renishaw, 2015] Renishaw (2015). explication de l’interférométrie : les fondements. <https://www.renishaw.fr/fr/explication-de-linterferometrie-7854>.
- [Sajjadi et al., 2017] Sajjadi, M. S. M., Schölkopf, B., and Hirsch, M. (2017). Enhancenet : Single image super-resolution through automated texture synthesis.
- [Sawilowsky and Fahoome, 2014] Sawilowsky, S. and Fahoome, G. (2014). Kruskal-wallis test : Basic. *Wiley StatsRef : Statistics Reference Online*.
- [SchneiderElectric, 2020] SchneiderElectric (2020). Qu’est-ce que la redondance ? <https://www.se.com/fr/fr/faqs/FA17854/>.
- [Simiz, 2016] Simiz, J.-G. (2016). *Approche holistique du contrôle du focus en photolithographie 193nm immersion pour les niveaux critiques en 28nm et 14nm FD-SOI*. PhD thesis, Université Jean Monnet, Saint-Etienne. Thèse de doctorat dirigée par Jourlin, Yves Microélectronique Lyon 2016.
- [Simiz et al., 2015] Simiz, J.-G., Hasan, T., Staals, F., Le-Gratiet, B., Gilgenkrantz, P., Villaret, A., Pasqualini, F., Tel, W. T., Prentice, C., and Tishchenko, A. (2015). Predictability and impact of product layout induced topology on across-field focus control. In *SPIE Advanced Lithography 2015*, volume 9424 of *Proceedings of the SPIE*, San José, CA, United States.
- [Siradel, 2021] Siradel (2021). Digital twins and data aggregation. <https://www.siradel.com/solutions/3d-gis-data-and-digital-twin/digital-twins-and-data-aggregation/>.

- [Su et al., 2020] Su, R., Thomas, M., Liu, M., Drs, J., Bellouard, Y., Pruss, C., Coupland, J., and Leach, R. (2020). Lens aberration compensation in interference microscopy. *Optics and Lasers in Engineering*, 128 :106015.
- [Szegedy et al., 2017] Szegedy, C., Ioffe, S., Vanhoucke, V., and Alemi, A. (2017). Inception-v4, inception-resnet and the impact of residual connections on learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 31.
- [Tao et al., 2019] Tao, F., Sui, F., Liu, A., Qi, Q., Zhang, M., Song, B., Guo, Z., Lu, S. C.-Y., and Nee, A. Y. C. (2019). Digital twin-driven product design framework. *International Journal of Production Research*, 57(12) :3935–3953.
- [Valade, 2020] Valade, C. (2020). *Développement d’une méthodologie adaptée à l’industrie microélectronique pour la reconstruction topographique par imagerie SEM à faisceau inclinable*. Theses, Université Grenoble Alpes.
- [van Dijk et al., 2019] van Dijk, L., Hasibi, F., Larrañaga, M., Pastol, A., Lam, A., and van Haren, R. (2019). Smart implant-layer overlay metrology to enable fab cycle time reduction. In Ukraintsev, V. A. and Adan, O., editors, *Metrology, Inspection, and Process Control for Microlithography XXXIII*, volume 10959, pages 298 – 310. International Society for Optics and Photonics, SPIE.
- [Wang et al., 2006] Wang, J., Yin, L., Wei, X., and Sun, Y. (2006). 3d facial expression recognition based on primitive surface feature distribution. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2 :1399– 1406.
- [Watson et al., 1985] Watson, L., Laffey, T., and Haralick, R. (1985). Topographic classification of digital image intensity surfaces using generalized splines and the discrete cosine transformation. *Comput. Vis. Graph. Image Process.*, 29 :143–167.
- [Witten et al., 2011] Witten, I. H., Frank, E., and Hall, M. A. (2011). *Data Mining : Practical Machine Learning Tools and Techniques*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 3rd edition.