



**HAL**  
open science

# Développement de stratégies de test pour les systèmes de communications millimétriques

Matthieu Verdy

► **To cite this version:**

Matthieu Verdy. Développement de stratégies de test pour les systèmes de communications millimétriques. Electronique. Université Grenoble Alpes, 2016. Français. NNT : 2016GREAT078 . tel-01506827

**HAL Id: tel-01506827**

**<https://theses.hal.science/tel-01506827v1>**

Submitted on 12 Apr 2017

**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 DE GRENOBLE

Spécialité : **Automatique**

Arrêté ministériel :

Présentée par

**Matthieu VERDY**

Thèse dirigée par **Suzanne LESECO**  
et codirigée par **Dominique MORCHE**

préparée au sein **LETI/DACLE/LAIR**  
et de **EEATS**

# Stratégies de Test pour les Sys- tèmes de Communications Millime- triques

Thèse soutenue publiquement le **22 septembre 2016**,  
devant le jury composé de :

**Salvador MIR**

Directeur de recherche CNRS, Laboratoire TIMA, HDR, Président

**Thierry TARIS**

Professeur, Université de Bordeaux 1, Rapporteur

**Serge BERNARD**

Chargé de recherche CNRS, Université de Montpellier, Rapporteur

**Patrick LOUMEAU**

Professeur, Télécom ParisTech, département COMELEC, Examineur

**Gildas LEGER**

Docteur, Ingénieur Chercheur, IMSE-CNM-CSIC Institut de Séville, Examineur

**Suzanne LESECO**

Directeur de Recherche CEA, CEA LETI, HDR, Directeur de thèse

**Dominique MORCHE**

Directeur de Recherche CEA, CEA LETI, HDR, Co-Directeur de thèse

**Jean-Pascal MALLET**

Directeur R&D, Presto-Engineering, Invité



À la mémoire de ma tante Mathilde

---

# Remerciements

---

C'est en terminant ce manuscrit que je me rends compte à quel point il est difficile de le commencer. L'effort de la recherche se montre personnel dans les limites qu'il repousse, mais collaboratif dans les rencontres qu'il favorise. Les premières rencontres ont cependant eu lieu bien avant le début des efforts intellectuels, avec Dominique Morche et Suzanne Lesecq. Les travaux présentés dans ce manuscrit se sont construits avec leur soutien, leur patience et leur expérience. Je leur suis infiniment reconnaissant d'avoir dirigé cette thèse.

L'enseignement que je retire de ces travaux a été en partie nourri par la société Presto Engineering, et en particulier par Jean-Pascal Mallet et Cédric Mayor. Ce sont nos discussions qui m'ont fait comprendre les problématiques industrielles du test. Je les remercie de m'avoir transmis des connaissances que je continue d'utiliser dans ma vie professionnelle.

Le laboratoire qui m'a accueilli a été un incomparable incubateur à idées, solutions et rencontres. En effet, le Laboratoire d'Architecture Intégrée Radio-fréquence m'a fait découvrir le monde de la conception de circuit. Je remercie très chaleureusement les directeurs successifs Pierre Vincent et Éric Mercier, ainsi que tous les chercheurs auprès desquels j'ai découvert le monde de l'électronique analogique.

C'est dans ce laboratoire, et aussi dans les laboratoires voisins, que j'ai rencontré bon nombre d'amis, qui ont eu une influence directe sur ces travaux. Je souhaite remercier particulièrement Alin Ratiu, Robert Polster, et Yogadissen Andee, conscrits de thèse avec qui j'ai partagé les moments forts de la vie de thésard. Ces moments, je les ai partagés aussi avec les thésards qui m'ont précédé : Gauthier, Olivier, Clément, Paco, Vincent, mais aussi avec ceux qui suivent : Dominique, Anthony, Jennifer, Zyad, Marguerite, César, Alexandre, David, Florent.

J'ai eu l'occasion de rencontrer des personnes avec qui j'ai pu avoir des

échanges stimulants sur d'autres sujets que celui de la thèse. Je veux parler de mes copains de quadricoptère et autres activités *geek* : Guillaume Waltener, Florent Berthier, André Myko, Maxime Huet, David Rabeau, les musiciens Andreina Liendo, David Coriat, Grégory Vaumourin, Jean-Baptiste David, ainsi que les motards Benjamin Caillat et Alexandra Walrave. Certains ont plusieurs casquettes, ils se reconnaîtront.

Parmi les personnes qui m'ont accompagné, je souhaite remercier mes colloc's grenoblois Maria Moula, Nicolas Kaeffer, Romain De Oliveira, Cécile Blaise, Victor Brochery et Adrien Dulac. J'ai trouvé chez eux un soutien sans faille et une bonne humeur qui m'ont toujours encouragé.

Ces travaux sont aussi le résultat d'un apprentissage qui a commencé bien plus tôt. Je souhaite remercier mes professeurs de l'École Nationale Supérieure d'Ingénieurs Sud-Alsace à Mulhouse, du lycée Carnot de Dijon, du lycée Cournot et du collège Romé de l'Isle de Gray.

Ma famille a été un solide pilier tout au long de ces travaux. Je les remercie tous, en particulier mes parents qui m'ont aidé à être celui que je suis devenu et qui m'ont toujours soutenu dans mes choix. Je remercie aussi mes frères, pour les parties de *Minecraft*, les explications sur les fondamentaux de la physique des matériaux, tous ces bons moments que nous avons eus dans notre jeunesse, et tout ceux que nous aurons encore.

Il est une personne qui partage ma vie depuis maintenant 7 ans et qui mérite bien des remerciements. Elsa m'a supporté, encouragé, et a été à mes côtés pour partager les meilleurs moments comme les moins bons. C'est avec patience qu'elle a relu ce manuscrit et je lui dois quelques jours sacrifiés au beau temps. Merci encore.

---

# Table des matières

---

<b>Table des matières</b>	<b>v</b>
<b>Liste des Acronymes et Symboles</b>	<b>ix</b>
<b>Introduction</b>	<b>xi</b>
Contexte et Problématique . . . . .	xi
Contributions . . . . .	xii
Organisation du document . . . . .	xiv
<b>I Contexte et problématique</b>	<b>1</b>
I.A Origine des fautes dans le cycle de développement d'un circuit .	2
I.A.1 Cycle de développement d'un circuit . . . . .	2
I.A.2 Découpage hiérarchique pour la conception d'un circuit	4
I.A.3 Classification simplifiée des fautes . . . . .	5
I.A.4 Synthèse sur l'origine des fautes considérées . . . . .	7
I.B Réduction des coûts du test . . . . .	8
I.B.1 Évolution des circuits et impact sur le coût du test . . . . .	8
I.B.2 Analyse de méthodes de test . . . . .	10
I.B.3 Précision du test . . . . .	16
I.C Synthèse de l'état de l'Art . . . . .	20
I.D Vision unifiée pour la réduction du coût du test . . . . .	24
<b>II Traitement de l'information pour les stratégies de test</b>	<b>27</b>
II.A Objectifs . . . . .	28
II.B Acquisition d'information . . . . .	31
II.B.1 Information obtenue en simulation . . . . .	32
II.B.2 Informations obtenues <i>via</i> des mesures . . . . .	39
II.B.3 Synthèse sur l'acquisition de données . . . . .	41
II.C Recherche de signaux d'intérêt . . . . .	41
II.C.1 Réduction de dimension . . . . .	43

II.C.2	Sélection de caractéristiques . . . . .	51
II.D	Modélisation . . . . .	59
II.D.1	Arbres de Classification . . . . .	61
II.D.2	Machine à Vecteurs de Support . . . . .	63
II.D.3	Approche k-Plus Proches Voisins . . . . .	65
II.D.4	Méthode MARS . . . . .	67
II.D.5	Méthodes à Réseaux de Neurones . . . . .	69
II.D.6	Conclusion sur les algorithmes de classification . . . . .	70
II.E	Gestion de la Précision . . . . .	72
II.E.1	Précision <i>a priori</i> des algorithmes de classification . . . . .	72
II.E.2	Matrice de Confusion et précision <i>a posteriori</i> . . . . .	76
II.E.3	Conclusion sur la gestion de la précision . . . . .	79
II.F	Conclusion . . . . .	80
<b>III</b>	<b>Méthode de sélection de mesures et classification</b>	<b>83</b>
III.A	Situation de la démarche de sélection dans la Vision Unifiée . . . . .	84
III.B	Sélection de caractéristiques dirigée par le coût : l'algorithme CDC-FOCUS . . . . .	86
III.B.1	Pertinence et coût d'un ensemble de caractéristiques . . . . .	86
III.B.2	Présentation de CDC-FOCUS . . . . .	88
III.B.3	Un exemple d'utilisation de CDC-FOCUS . . . . .	90
III.B.4	Conclusion sur CDC-FOCUS . . . . .	91
III.C	Influence du bruit de mesure sur les métriques de test . . . . .	93
III.C.1	Hypothèse du bruit additif et convolution des PDFs . . . . .	94
III.C.2	Étape 1 : Évaluation de l'impact par déconvolution . . . . .	96
III.C.3	Étape 2 : Analyse du signal reconstruit . . . . .	98
III.C.4	Exemple d'application . . . . .	100
III.C.5	Conclusion sur le recouvrement du signal d'intérêt . . . . .	103
III.D	Induction du modèle . . . . .	105
III.E	Synthèse, avantages et limites de la méthode proposée . . . . .	105
<b>IV</b>	<b>Validation de la méthode</b>	<b>109</b>
IV.A	Acquisition des bases de données . . . . .	110
IV.A.1	Acquisition de bases de données par simulation . . . . .	110
IV.A.2	Acquisition de bases de données par mesure . . . . .	114
IV.B	Validation avec des données de simulation . . . . .	117
IV.B.1	Utilisation de la méthode avec un circuit LNA . . . . .	117
IV.B.2	Convertisseur Analogique Numérique en Temps Continu . . . . .	124
IV.B.3	Conclusion sur l'utilisation de la méthode proposée avec des simulations . . . . .	131
IV.C	Validation de la méthode sur des mesures . . . . .	132
IV.C.1	Présentation de la base de données et procédure d'appli- cation de la méthode . . . . .	132

IV.C.2 Résultats de l'application de la méthode sur la base de données issue de mesures . . . . .	133
IV.C.3 Conclusion de la validation de la méthode proposée sur la base de données de mesures . . . . .	134
IV.D Conclusion sur la validation de la méthode . . . . .	135
<b>V Conclusion et Perspectives</b>	<b>139</b>
V.A Résumé de la Méthodologie complète . . . . .	139
V.B Sélection de caractéristique guidée par le coût . . . . .	141
V.C Validation de la méthode proposée sur des circuits . . . . .	141
V.D Évaluation de l'erreur dans la méthode . . . . .	142
V.E Synthèse des perspectives . . . . .	142
<b>A Nombre d'échantillons en simulation Monte-Carlo</b>	<b>147</b>
<b>B Analyse en Composante Principale</b>	<b>149</b>
<b>C Régression Inverse par Tranche</b>	<b>151</b>
<b>D Application du "Kernel Trick" à l'Analyse en Composante Principale</b>	<b>155</b>
<b>E Exemple d'utilisation d'algorithmes de sélection de caractéristiques</b>	<b>157</b>
E.A Exemple d'utilisation de RELIEF . . . . .	157
E.B Apports de FOCUS-2 . . . . .	158
<b>F Machines à Vecteurs de Support</b>	<b>161</b>
F.A Frontière et marge . . . . .	161
F.B Utilisation des noyaux . . . . .	162
F.C Résolution de la frontière à marge maximale . . . . .	163
<b>G Classification par Arbre de Décision</b>	<b>165</b>
G.A Séparation par mesure d'entropie . . . . .	166
G.B Séparation par mesure de Gini . . . . .	167
<b>H Algorithmes dérivés de FOCUS</b>	<b>169</b>
H.A FOCUS-2 . . . . .	169
H.B CDC-FOCUS . . . . .	170
H.B.1 Cœur de CDC-FOCUS . . . . .	170
H.B.2 Exemple de test de suffisance . . . . .	170
<b>Bibliographie</b>	<b>173</b>
<b>Table des figures</b>	<b>182</b>





---

# Liste des Acronymes et Symboles

---

## Acronymes

ADC	Analog to Digital Converter
BIMBO	Basic Identification Method using Binary Observation
CART	Classification And Regression Tree (Arbre de Classification et de Régression)
CDC-FOCUS	Cost Driven Continuous FOCUS
CT	Classification Tree (Arbre de Classification)
DoE	Design of Experiment
DuT	Device under Test (Dispositif sous Test)
EDR	Effective Dimension Reduction (Réduction de Dimension Effective)
IFFT	Inverse Fast Fourier Transform (Inverse de la Transformée de Fourier Rapide)
IoT	Internet of Things (Internet des Objets)
ITRS	International Roadmap for Semiconductors
kNN	k-Nearest Neighbor (k plus proches voisins)
MARS	Multivariate Adaptive Regression Splines
PAC	Probably Approximately Correct
PCA	Principal Component Analysis (Analyse en Composantes Principales)
PDF	Probability Density Function (Densité de Probabilité)

RNA	Réseau de Neurone Artificiel
SIR	Sliced Inverse Regression (Régression Inverse par Tranches)
SNR	Signal to Noise Ratio (Rapport Signal à Bruit)
SVM	Support Vector Machine (Machine à Vecteur de Support)

### Symboles

$L \in \mathbb{N}$	Dimension de l'espace de redescription
$N \in \mathbb{N}$	Nombre de circuits observés
$O \in \mathbb{N}$	Nombre d'observations par circuit
$P \in \mathbb{N}$	Nombre de performances par circuit
$\mathbb{E}$	Ensemble de représentation de la classification : peut être un ensemble discret, ou un ensemble de labels
$\mathcal{C} \in \mathbb{R}^O$	Espace d'observation de dimension $O$
$\mathcal{L} \in \mathbb{R}^L$	Espace de redescription
$\Theta$	Complexité algorithmique
$\mathbf{c}_i \in \mathbb{R}^O$	Vecteur des observations du $i^{\text{ème}}$ circuit, aussi appelé $i^{\text{ème}}$ instance
$\mathbf{C} \in \mathbb{R}^{O \times N}$	Matrice d'observation représentant toutes les caractéristiques de tous les circuits, représentée par un nuage de points
$\mathbf{p}_i \in \mathbb{R}^P$	Vecteur des valeurs des performances pour le $i^{\text{ème}}$ circuit
$\mathbf{y} \in \mathbb{E}^N$	Vecteur contenant la classification pour tous les circuits

---

# Introduction

---

## Contexte et Problématique

Les systèmes électroniques qui nous entourent sont de plus en plus performants, avec des fonctionnalités de plus en plus intelligentes. Aussi, il est possible d'ajouter plus de composants électroniques par unité de surface et les phénomènes physiques perturbatoires dans les circuits sont de mieux en mieux compris et contrôlés. L'analyse des circuits peut s'articuler autour deux domaines : les circuits analogiques et les circuits numériques. Nos recherches portent sur les premiers, et en particulier dans le cadre des communications sans fil. En effet, dans ce cas, la génération de séquences de test automatique est complexe de par la nature continue des signaux.

Le test des circuits permet de garantir le bon fonctionnement des circuits vendus. Il a donc un objectif économique car de ce fait il limite les retours de circuits défectueux. Cependant, le test des circuits doit aussi être mené pour éviter que des circuits fonctionnels soient considérés comme défectueux, et donc éliminés ou invendus. Le test traditionnel des circuits consiste à mesurer et à analyser un ou plusieurs signaux mesurés sur les circuits, à des fréquences possiblement élevées et selon un protocole parfois complexe.

Les circuits font également l'objet d'une démarche qualité menée avec l'objectif d'un rapport qualité/prix optimal : le coût associé aux tests va donc dépendre de l'importance financière et de la criticité du dispositif. Nous nous inscrivons dans la recherche d'une stratégie de test de systèmes analogiques qui permette de trouver un compromis entre le coût de test et la précision des tests (taux de faux positifs par exemple). En effet, l'intuition pourrait indiquer une relation possible entre le coût du test et la qualité des résultats obtenus : plus le coût de test élevé, meilleurs en sont les métriques (taux de faux positifs ou précision).

Ainsi, la problématique adressée dans nos travaux est celle de l'établissement de tests pour les circuits analogiques permettant de trouver un compromis

entre le coût et la qualité du test. Cette problématique implique l'acquisition d'information sur le circuit, le traitement de cette information en vue du test et l'établissement de modèles de prise de décision quant à l'état fonctionnel ou défectueux des circuits. Ces éléments seront traités dans nos travaux et nous proposons une méthodologie les mettant en œuvre pour sélectionner un ensemble de mesures permettant de diminuer le coût associé au test de circuits. Cette problématique très large laisse place à un grand spectre de réponses et d'études possibles. Dans nos travaux, nous aurons à l'esprit de ne pas focaliser l'étude sur un circuit particulier, ou sur une technologie particulière. La précision des modèles construits est un élément qui doit être également pris en compte dans la méthodologie proposée dans nos travaux.

## Contributions

Dans ce manuscrit, on s'intéresse au problème de la réduction des coûts de test des circuits analogiques. Pour ce faire, nous proposons une méthodologie permettant de répondre à cette problématique, et nous présentons les différents algorithmes utilisés par cette méthodologie.

Nous proposons en particulier un algorithme permettant de traiter la réduction du coût de test pour les circuits analogiques. Cet algorithme, CDC-FOCUS, est un algorithme de sélection de caractéristiques guidée par le coût. Il est issu d'un algorithme existant, à savoir C-FOCUS, pour la sélection de caractéristiques dans le traitement de l'information, que nous modifions pour répondre à notre problématique. La modification proposée est l'introduction d'un coût permettant de guider la sélection des caractéristiques.

Le circuit analogique considéré est représenté par une base de données qui contient toute l'information accessible sur ce circuit. Les outils présentés ici traitent cette base de données pour en extraire des informations pertinentes que nous utilisons par la suite pour réduire le coût du test.

Les contributions répondant à cette problématique sont au nombre de quatre, à savoir :

1. une vision générique du traitement des problèmes de test des circuits analogiques est proposée au chapitre I après une analyse de l'état de l'Art ;
2. un algorithme de sélection de caractéristiques prenant en compte une fonction coût est développé afin de répondre au problème de sélection de caractéristiques guidée par le coût. Cet algorithme, nommé CDC-FOCUS, est introduit au chapitre III ;
3. une étude de l'erreur et du bruit pour la modélisation est menée au chapitre III

4. une validation de la méthode proposée avec des données issues de mesures et de simulations est menée au chapitre IV.

La méthodologie que nous proposons est validée sur trois exemples différents. Deux de ces exemples sont issus de circuits simulés développés dans le laboratoire d'accueil. Le premier est un circuit d'amplification de signaux, le second est un convertisseur analogique numérique en temps continu. Les circuits simulés permettent de maîtriser la génération de données à analyser et sont utilisés pour montrer comment utiliser la méthodologie proposée en phase de développement du circuit. Le troisième circuit consiste en des données provenant de mesures d'un circuits réel. Le circuit est représenté par des mesures "réelles". Cet exemple permet de montrer comment appliquer la méthodologie proposée pendant la phase de test.

Ces travaux ont donné lieu à publication dans des conférences internationales :

1. M. Verdy, D. Morche, S. Lesecq, E. De Foucauld, J.-P. Mallet, C. Mayor, "Weight based Feature Selection Algorithm for Alternative Measurements", STEM Workshop - 19th European Test Symposium, mai 2014.
2. M. Verdy, A. Ratiu, D. Morche, E. De Foucauld, S. Lesecq, J.-P. Mallet, C. Mayor, "Cost-driven statistical analysis for selection of alternative measurements of analog circuits", in 2014 21st IEEE International Conference on Electronics, Circuits and Systems (ICECS), 2014, p. 104-107.
3. M. Verdy, D. Morche, E. De Foucauld, S. Lesecq, J.-P. Mallet, C. Mayor, "Balancing test cost reduction vs. measurements accuracy at test time", in New Circuits and Systems Conference (NEWCAS), 2015 IEEE 13th International, 2015, p. 1-4.

La première publication présente la méthodologie proposée pour répondre au problème de réduction de coût du test des circuits analogiques. Elle correspond à la contribution numéro 1 ci-dessus. Nous y proposons une méthodologie générique, construite en trois étapes fondamentales, à savoir, obtention de données sur le circuit, réduction de la dimension de ces données, construction d'un modèle. Nous y présentons des résultats préliminaires obtenus en utilisant l'algorithme de sélection de caractéristiques spécialement développé, à savoir CDC-FOCUS.

Notre deuxième publication présente plus en détail l'algorithme CDC-FOCUS, spécialement développé pour répondre au problème de réduction de coût du test. Il est dérivé de l'algorithme FOCUS, et plus particulièrement de C-FOCUS. Il est présenté comme une solution à la recherche d'un ensemble de mesures alternatives les plus pertinentes dans un circuit et dont le coût total est le plus faible possible. Elle correspond à la contribution numéro 2.

Notre troisième publication est une étude sur les relations entre la précision d'une mesure et les erreurs de classification induites par le manque de précision. Ces travaux ont pour ambition de traiter de l'analyse de l'erreur de notre méthode globale. Cette publication se rapporte à la contribution numéro 3.

Les travaux ont été réalisés au sein du Laboratoire des Architectures Intégrées Radiofréquences (LAIR) au Département Architecture, Conception et Logiciels Embarqués (DACLE) du Commissariat à l'Énergie Atomique et aux Énergies Alternatives (CEA) de Grenoble, en collaboration avec Presto Engineering. Ils sont donc le fruit d'un environnement à la fois académique et industriel.

## Organisation du document

Ce manuscrit s'organise en cinq chapitres.

Le contexte de nos travaux, une revue succincte du test de circuits analogiques et l'analyse de la problématique sont traités dans le chapitre I. Une vision unifiée de la problématique est donnée. Une méthodologie minimisant le coût du test est ensuite proposée.

Le chapitre II présente des techniques de traitement de l'information issues de l'état de l'Art et potentiellement utilisables dans le cadre de nos travaux. Ce chapitre suit le squelette de la méthode proposée au chapitre I. Il analyse également les avantages et inconvénients des algorithmes envisagés. Le chapitre II met donc en place l'environnement technique de la méthode proposée au chapitre I.

L'algorithme de traitement de données CDC-FOCUS, pierre angulaire de la méthode proposée, fait l'objet du chapitre III. Nous proposons cet algorithme de sélection de caractéristiques dirigée par le coût pour répondre à la problématique de nos travaux. Ce chapitre expose aussi une méthode d'analyse de l'influence du bruit sur les performances des modèles de classification. Il récapitule *in fine* la démarche générique proposée au chapitre I.

Le chapitre IV utilise des exemples issus de simulations et de mesures de circuits réels pour valider la méthode proposée au chapitre III. Les exemples utilisés sont issus de simulations de circuits et de mesures de circuits réels. Cette diversité permet de montrer qu'il est possible d'appliquer la méthode que nous proposons soit en phase de conception du circuit, soit après la phase de conception, pour diminuer le coût du test.

Enfin, le chapitre V résume les avantages de la méthode proposée au fil du manuscrit. Les limites de notre méthode sont également détaillées, et des pistes d'amélioration sont proposées.





## Chapitre I

---

# Contexte et problématique

---

Ne me dites pas que ce problème est difficile. S'il n'était pas difficile, ce ne serait pas un problème.

---

Maréchal FOCH

**L**E TEST des circuits analogiques est une étape nécessaire du processus de développement d'un circuit. Dans ce chapitre, le contexte du test est présenté dans une première section. Le test intervient en fin de cycle de développement pour assurer la vente de circuits fonctionnels et le rejet de circuits défectueux.

Dans la deuxième section, les erreurs affectant les circuits sont introduites. Des erreurs peuvent apparaître tout au long du cycle de développement du circuit, mais ce chapitre s'intéresse en particulier aux erreurs spécifiques à l'étape de production du circuit.

Enfin, le test a un coût qui doit être pris en compte dans le coût global de développement du circuit. Ce chapitre présentera donc rapidement ce lien et introduira deux métriques qui permettent d'apprécier la qualité du test au regard de son impact économique.

Une fois la provenance des erreurs et le caractère économique introduit, ce chapitre étend ce contexte par une discussion de l'état de l'Art de quelques méthodes de test des circuits analogiques permettant de réduire le coût du test. Il sera montré que ces techniques possèdent des points communs qui répondent aux problèmes généraux du test, à savoir l'acquisition de connaissances sur le circuit, la définition de signaux d'intérêt, et la création d'un modèle permettant de lier ces mesures d'intérêt à l'état du circuit.

Ces points communs seront regroupés et une Vision Unifiée de la problématique de réduction du coût du test de circuit analogique sera proposée. Cette Vision Unifiée sera un guide pour la suite de ce manuscrit.

La section I.A situe le test du circuit dans le flot de conception et définit l'origine des fautes. L'évolution du coût du test est présenté dans la section I.B puis différentes techniques de test issues de l'état de l'Art sont résumées. La section I.C analyse les méthodes de réduction de coût de test existantes et propose une analyse de l'état de l'Art des techniques de test présentées. Enfin, la section I.D propose une vision unifiée pour la génération de stratégie de test.

La finalité de ce chapitre est de positionner le contexte de nos travaux sous une vision unifiée permettant par la suite de proposer un flot d'établissement de stratégie de test.

## I.A Origine des fautes dans le cycle de développement d'un circuit

La complexité inhérente d'un circuit analogique RF peut s'illustrer au travers de l'analyse de son cycle de développement. Ce chapitre introduit une vision simplifiée des étapes de conception d'un circuit permettant de positionner le test d'un circuit dans son cycle de développement. Cette vision simplifiée du développement d'un circuit se veut illustrative. Plus d'informations concernant ce point sont disponibles dans [Akkouche, 2011]. Ici, c'est une vision *traitement de données* qui est proposée.

### I.A.1 Cycle de développement d'un circuit

Le cycle classique de développement d'un circuit, présenté en figure I.1, s'organise de la description du cahier des charges au test qui permet d'en valider le respect. Les sources d'erreurs potentielles sont celles provenant de la conception, de la vérification, et de la production.

L'étape de **conception** consiste à générer les schémas réalisant la fonctionnalité décrite dans le cahier des charges et dans les contraintes que ce dernier spécifie. Les erreurs de conception sont évitées par des simulations partielles du circuit qui permettent de valider une partie de son fonctionnement.

Après le dessin des masques, l'étape de **vérification** assure leur cohérence avec les schémas établis en conception. La réussite de cette étape repose sur la bonne compréhension de la technologie utilisée. Elle utilise des modèles établis *a priori* à partir de l'analyse de la technologie (e.g. courant de fuite, capacité

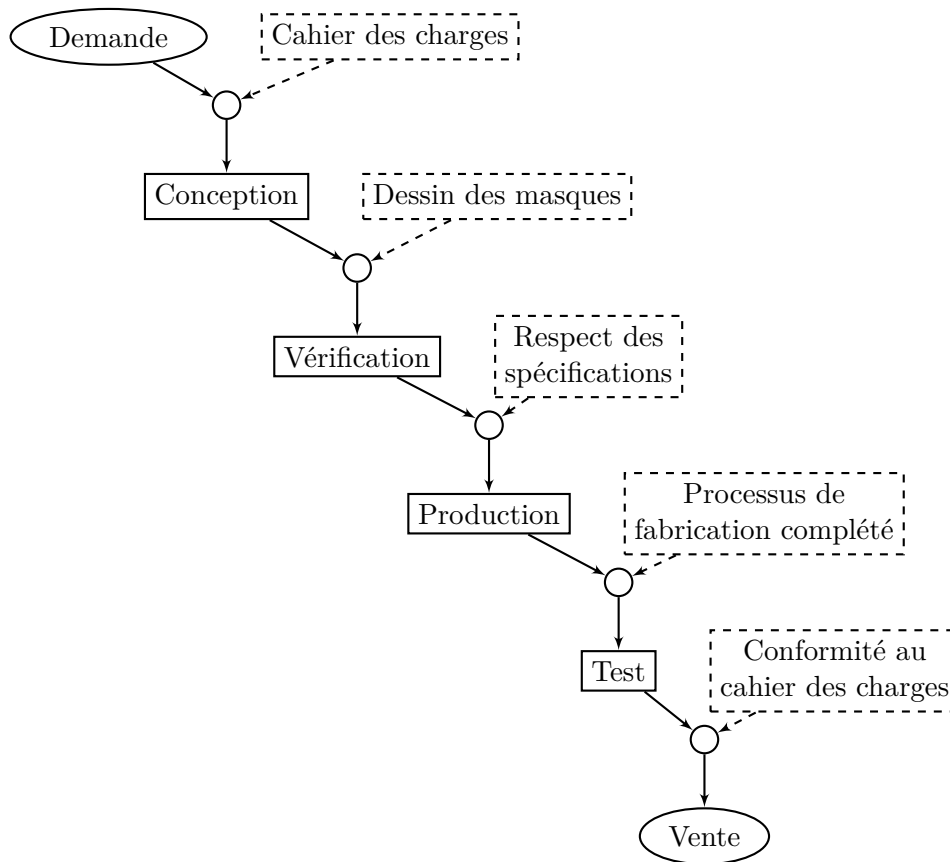


FIGURE I.1 – Schéma de représentation du cycle de développement d'un circuit (reproduit depuis [Dubois, 2011])

entre lignes, *etc.*).

L'étape de **production** utilise les masques vérifiés à l'étape précédente pour fabriquer les circuits. Les erreurs pouvant apparaître à cette étape sont nombreuses. Tout d'abord, le procédé de fabrication est généralement très sensible aux événements extérieurs, par exemple, climatique (pression, température), sismique, calibration des machines, dépôt de poussière, *etc.* Ensuite, les machines et les masques ont une précision limitée ce qui entraîne une légère modification des tailles des composants évaluées à l'étape de conception.

L'étape de **test** est la dernière étape avant la vente. Elle permet de valider les performances du circuit par rapport aux spécifications établies dans le cahier des charges. Les erreurs recherchées à cette étape sont celles de la production. En effet, ces erreurs sont complexes à couvrir. Elles dépendent du circuit et de la technologie, de la variation d'éléments extérieurs qui ne sont pas contrôlables. Des erreurs peuvent apparaître en amont de la production, mais elles sont

supposées détectées par l'étape de vérification.

Pour centrer plus encore la problématique de nos travaux, l'étape de conception de circuit va être maintenant schématisée.

### I.A.2 Découpage hiérarchique pour la conception d'un circuit

La conception d'un circuit repose tout d'abord sur le découpage fonctionnel du circuit. Le principe est de découper le comportement du circuit en plusieurs fonctions réalisant chacune une tâche précise. Par exemple, dans le cas de circuits analogiques de communication sans fils, ces fonctions sont l'amplification de signaux, la synthèse de fréquence, la conversion numérique et les traitements en bande de base.

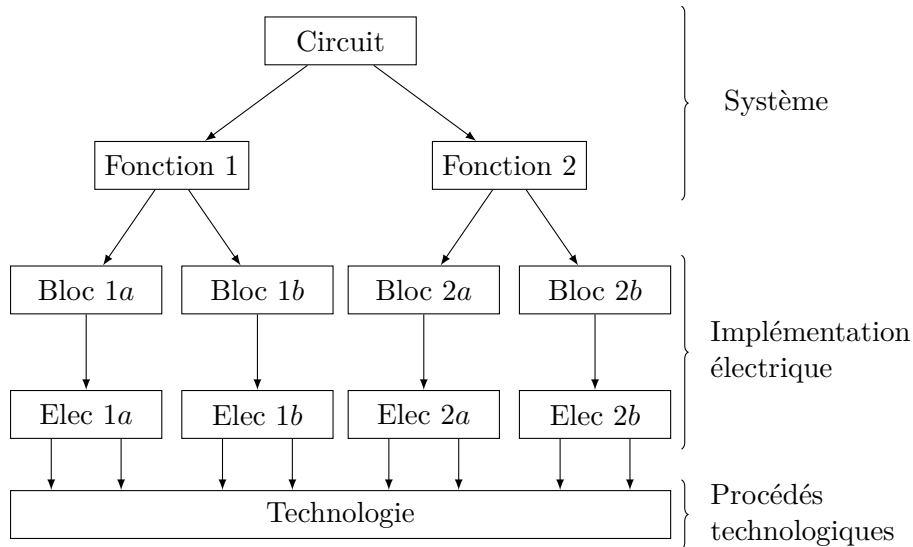


FIGURE I.2 – Schéma de découpage et implémentation d'un circuit

L'intérêt d'un tel découpage est de pouvoir isoler chaque fonction qui peut, dès lors, être développée indépendamment des autres, autorisant le développement en parallèle des fonctions. Chaque fonction est alors spécifiée pour assurer que le circuit complet respecte toujours le cahier des charges.

Les fonctions sont ensuite découpées en blocs électriques dont les spécifications sont héritées des spécifications de la fonction. La répartition de ces spécifications dans le circuit est généralement représentée par des formules analytiques simples, comme par exemple les formules de Friis [Friis, 1944]. Pour illustration, la répartition de la figure de bruit entre les blocs d'un circuit de communication est donnée par :

$$F = F_1 + \frac{F_2 - 1}{G_1} + \frac{F_3 - 1}{G_1 G_2} + \frac{F_4 - 1}{G_1 G_2 G_3} + \dots + \frac{F_n - 1}{G_1 G_2 G_3 \dots G_{n-1}} \quad (\text{I.1})$$

avec :

- $F$  Le facteur de bruit de la fonction
- $F_k$  Le facteur de bruit du bloc  $k$
- $G_k$  Le gain du bloc  $k$

Les architectures des blocs sont ensuite choisies selon la technologie utilisée et les spécifications des blocs. Elles sont alors implémentées électriquement et simulées à l'aide du modèle des composants. Les paramètres électriques de ces modèles sont obtenus par caractérisation du procédé technologique. De la même manière, il y est aussi défini les déviations statistiques de ces paramètres qui ont lieu pendant la fabrication.

Cette méthode dite *top-down*, représentée sur la figure I.2, divise le travail d'implémentation en tâches élémentaires et indépendantes. L'effet néfaste de cette méthode est la sur-spécification induite par la nécessaire spécification indépendante de chacun des blocs. En effet, dans cette approche modulaire, un bloc ne peut pas compenser une performance faible d'un autre bloc. Aussi, la sur-spécification induit une augmentation de la surface silicium occupée et de la consommation du circuit.

Les équations décrivant les transitions d'un niveau vers le suivant sont non linéaires et le nombre de dimensions est élevé. Le tableau I.1 présente les dimensions de chaque niveau de description ainsi que des exemples de noms de paramètres pour ces niveaux dans le cas d'un filtre GmC.

Dans ces conditions, la modélisation des performances à partir des paramètres électriques est un problème complexe, qui souvent ne peut pas être traité par des méthodes simples de traitement des données.

### I.A.3 Classification simplifiée des fautes

L'étape de conception de circuits peut être entachée d'erreurs qui devraient être détectées dans l'étape de vérification. Des erreurs peuvent aussi être introduites dans l'étape de production. Le but ultime du test est de détecter les circuits défectueux compte tenu de ces erreurs.

Les perturbations observables à l'issue de l'étape de production du circuit sont les suivantes [Dubois, 2011] :

<b>Performances circuit (<math>\approx 10</math>)</b>
Gain NF IIP3 <i>etc.</i>
<b>Paramètres électriques (<math>\approx 10000</math>)</b>
R0 I5 V0 <i>etc.</i>
<b>Paramètres technologiques (<math>\approx 400</math>)</b>
Épaisseurs d'oxyde des différents niveaux Concentration de substrat Mobilité des électrons Longueur de grille <i>etc.</i>

TABLE I.1 – Quelques paramètres extraits d'un filtre GmC

- les écarts de paramètres de contrôle du *process* ;
- les écarts induits par l'environnement externe ;
- les erreurs de conception ;
- les défauts de fabrication grossiers.

Nous considérons ici que ces perturbations ne sont pas reproductibles : si tel était le cas, alors l'origine de ces perturbations serait identifiée puis corrigée comme, par exemple, le non-respect du protocole de fabrication ou une erreur dans ce dernier.

Les perturbations telles que définies ci-avant engendrent deux types de variations et deux types de défauts :

- les variations locales qui proviennent par exemple d'une variation de la largeur et/ou de la longueur d'un transistor, de sa tension de seuil, entre différents points du circuit ;
- les variations globales qui sont indépendantes de la taille des composants. Ce sont des variations entre différents circuits du *wafers*, ou entre différents *wafers* ;
- le non respect du cahier des charges ;
- l'apparition de courts-circuits et de circuits ouverts non souhaités.

Les variations locales impactent les blocs dont les transistors fonctionnent par appairage. Dans ce cas, les transistors devraient avoir un comportement identique mais une variation locale impacte l'un des transistors et le comportement de la paire de transistors est déséquilibré.

À l'inverse, les variations globales impactent de la même manière deux transistors appairés d'un même circuit, mais différemment deux circuits à des endroits différents du *wafers*. Ainsi, les performances de deux circuits éloignés sur un même *wafers* ou sur deux *wafers* différents seront différentes. L'apparition de courts-circuits et de circuits ouverts sont des défauts qui apparaissent dans le cas d'un dépôt de poussière ou de perturbations importantes pendant la phase de production.

Les variations et les défauts engendrent des fautes, qui peuvent être classées en deux catégories, selon leur importance [Dubois, 2011] :

- **les fautes catastrophiques** traduisent le fait que le circuit est rendu totalement non fonctionnel. Ces fautes catastrophiques sont généralement induites par le dépôt d'un corps étranger sur le *wafers* au moment de la fabrication résultant en un court-circuit ou un circuit ouvert ;
- **les fautes paramétriques** sont définies par le non respect d'au moins une des spécifications du Cahier des Charges. Elles proviennent du fait qu'au moins un des paramètres ne respecte pas le *process* ;

Une troisième catégorie d'erreurs est celle des erreurs de conception, qui sont caractérisées par une production sans erreur d'un circuit dont au moins une performance ne respecte pas les spécifications.

Dans nos travaux nous supposons que les erreurs de conception sont détectées (et corrigées) lors de la phase de vérification. De plus, les erreurs catastrophiques sont supposées facilement détectables. Les travaux présentés ici se focalisent sur les fautes paramétriques issues de variations globales. Ainsi, les défauts adressés proviennent des variations technologiques importantes qui entraînent la violation d'une au moins des spécifications du circuit.

#### I.A.4 Synthèse sur l'origine des fautes considérées

Le cycle de développement d'un circuit a été présenté à la section I.A.1. L'étape de test a été positionnée comme la dernière étape du cycle de développement. Elle valide tout le développement depuis la définition du cahier des charges.

De plus, on a montré qu'un nombre élevé de paramètres (performances, paramètres électriques et technologiques) intervient dans le cycle de dévelop-



pement d'un circuit. Les relations physiques, électriques et système qui relient les différents domaines sont non-linéaires et complexes.

Enfin, une classification des fautes a été proposée pour centrer la problématique de test des circuits spécifiquement sur les fautes paramétriques dues aux variations technologiques qui entraînent une violation des spécifications.

Dans le cadre de la réduction du coût du test, la prochaine section analyse l'évolution du coût du test et présente quelques solutions existantes dans l'état de l'Art.

## I.B Réduction des coûts du test

Cette section relie le coût du test de circuits analogiques à leur complexité, et montre qu'il devient nécessaire de réduire ce coût. Par la suite, les méthodes de l'état de l'Art permettant de réduire ce coût sont analysées de manière à fournir une Vision Unifiée de la problématique de réduction des coûts du test des circuits analogiques.

Par la suite, on proposera notre Vision Unifiée de la problématique de réduction du coût du test des circuits analogiques.

### I.B.1 Évolution des circuits et impact sur le coût du test

La complexité des tests de validation fonctionnelle d'un circuit analogique est liée à sa propre complexité<sup>1</sup>. Les applications des circuits analogiques sont de plus en plus complexes et les performances de plus en plus élevées. Par exemple, dans le domaine des communications sans fil, les applications demandent d'augmenter les bandes passantes et les fréquences de fonctionnement mais aussi d'assurer la compatibilité avec toujours plus de standards.

L'évolution technologique présentée en figure I.3 permet d'intégrer ces fonctionnalités rendues nécessaires par l'évolution du marché des télécommunications, ce qui complexifie les circuits. Cette complexité des circuits se traduit au niveau des performances des transistors, comme le présente la figure I.4 [Hoefflinger, 2012]. La complexité des circuits entraîne aussi une caractérisation plus complexe des circuits.

Pour éviter la chute de rendement des circuits induite par les variations technologiques, une solution consiste à réaliser une conception robuste en prenant des marges sur les performances. Cependant, cette méthode implique

---

1. Un accent particulier est porté sur le concept de fonctionnalités analogiques. En effet, il peut être plus simple de tester un ADC au fonctionnement complexe qu'un amplificateur THz, au fonctionnement plus simple.

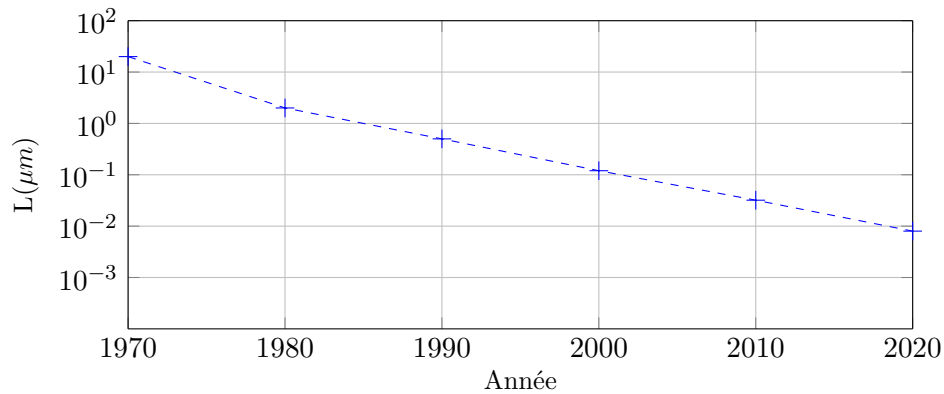


FIGURE I.3 – Évolution de la longueur du canal des transistors MOS de 1970 à 2020 [Hoefflinger, 2012]

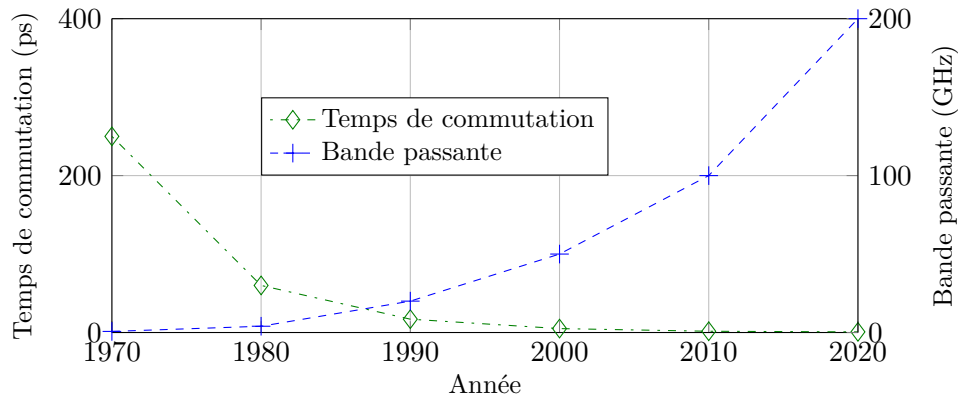


FIGURE I.4 – Évolution de la bande passante et de la vitesse de commutation de transistors NMOS de 1970 à 2020 [Hoefflinger, 2012]

une augmentation de la surface de silicium occupée et de la consommation, ce qui n'est pas acceptable. Une meilleure maîtrise du rendement par des tests et des caractérisations permettrait de concevoir un circuit en diminuant les marges sur les performances, ce qui est possible par une meilleure maîtrise de l'impact de la technologie aussi bien en conception qu'en test. Dans ce cas, des tests et caractérisations supplémentaires s'avèrent nécessaires.

Cependant nous sommes face à un compromis. La réduction du coût du test est possible par l'allègement des contraintes sur le test (nombre de mesures, précision des mesures, *etc.*), ce qui correspond à augmenter l'erreur de classification des circuits testés. Les erreurs possibles sont :

- déclarer un circuit fonctionnel (i.e. conforme aux spécifications) défectueux ;

- déclarer un circuit défectueux fonctionnel.

Ce sont des erreurs de *première* et *deuxième* espèce respectivement, aussi connues sous le nom de **faux négatif** et **faux positif**.

Ainsi, de manière qualitative, accepter de vendre plus de circuits défectueux permet de diminuer les contraintes sur le test, et donc d'en diminuer le coût. Évidemment, les retours clients sont alors plus nombreux et le coût des retours clients est plus élevé. *A contrario*, réaliser un test plus contraignant, donc augmenter son coût, entraîne naturellement une diminution du nombre de circuits défectueux vendus, et donc une diminution des coûts de retours clients.<sup>2</sup>

Il est alors possible de faire apparaître un point d'équilibre entre le coût associé au test et le coût associé aux circuits défectueux vendus, qui sont des retours client. Ce point d'équilibre est représenté qualitativement en figure I.5 et proposé par l'ITRS (*International Roadmap for Semiconductors*). L'augmentation de la performance du test entraîne l'augmentation du coût du test. En effet, pour augmenter la performance du test, il faut réaliser plus de tests, ou mesurer avec plus de précision. Ainsi, [Bottoms, 2007] observe une augmentation exponentielle du coût du test avec la performance du test. Une explication possible de cette allure est que ces contraintes sur le test sont utilisées pour détecter des défaillances dont la probabilité d'apparition est de plus en plus faible.

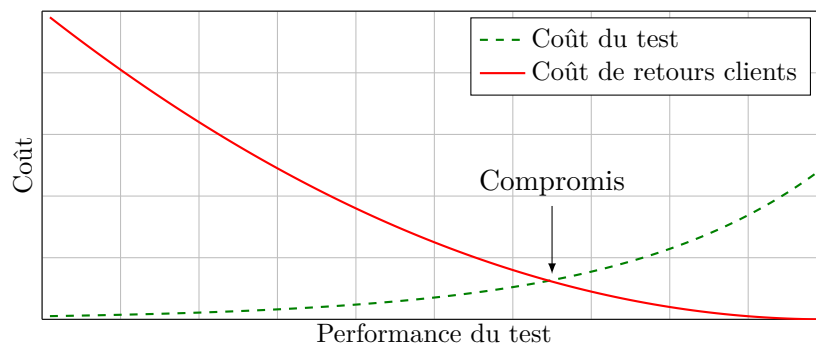


FIGURE I.5 – Illustration qualitative du compromis entre le coût du test et le coût des retours clients (d'après [Bottoms, 2007])

## I.B.2 Analyse de méthodes de test

On propose maintenant une classification et une analyse critique de méthodes issues de l'état de l'Art du test de circuit. Cette analyse se concentre sur les méthodes qui permettent de réduire le coût des tests des circuits analogiques

<sup>2</sup>. Cette assertion dépend évidemment du produit.

soumis aux fautes paramétriques. Elle converge vers une vision unifiée de la problématique, qui permettra *in fine* de proposer une approche de génération de procédure de test en briques élémentaires.

**Le test direct.** Le test d'un circuit dans sa version originelle repose sur la comparaison entre ses performances et ses spécifications comme représenté en figure I.6. Cette représentation distingue le circuit du dispositif testé, appelé *Device under Test* (DuT). En effet, on verra par la suite que le circuit peut contenir, outre le DuT, d'autres blocs utilisés par exemple pour effectuer le test.

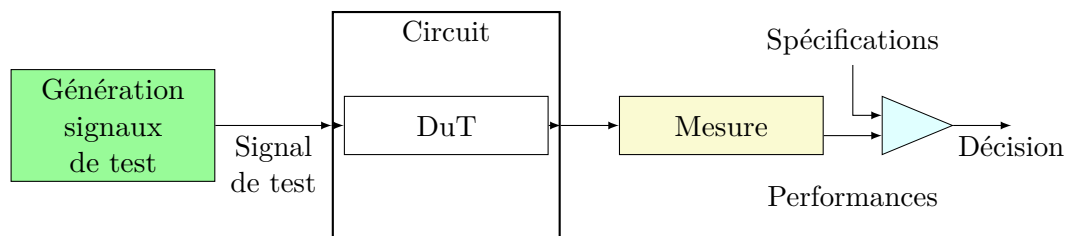


FIGURE I.6 – Schéma simplifié du test direct

Le test direct des circuits utilise trois étapes fondamentales, qui sont la génération des signaux de test, la mesure des signaux du circuit, et la prise de décision quant au bon fonctionnement du circuit. La prise de décision est, dans le cas direct, une comparaison entre les performances mesurées du circuit et les spécifications du circuit établies dans le cahier des charges. La génération des signaux de test, la mesure des signaux de sortie et la prise de décision sont des étapes réalisées à l'extérieur du circuit, par un dispositif que nous appellerons "testeur" dans la suite.

Le coût du test des circuits provient en partie des signaux d'excitation à générer, du nombre et des caractéristiques des signaux à extraire du circuit, des outils de mesure, et, dans une moindre mesure, de la prise de décision. Les méthodes de réduction des coûts du test modifient donc cette représentation, et s'intéressent à au moins l'un des blocs la constituant.

### I.B.2.a Mesures alternatives

Les mesures alternatives consistent en l'ajout de points de mesure à l'intérieur d'un circuit. Notons que des points de mesure non intrusifs (par exemple par mesure de température [Abdallah et al., 2013]) peuvent aussi être utilisés. Ces points de mesure sont utilisés à la place de la mesure des performances pour la prise de décision sur l'état du circuit. L'état de l'Art propose différents

capteurs permettant d'extraire différentes caractéristiques : des analyses spectrales [Negreiros et al., 2002, Negreiros et al., 2004], des mesures de courant [Daubert et al., 1988], des mesures de fenêtres en courant [Velasco-Medina et al., 1999b]. Le test devient alors un **test indirect**, par mesures alternatives. Il est représenté sur la figure I.7.

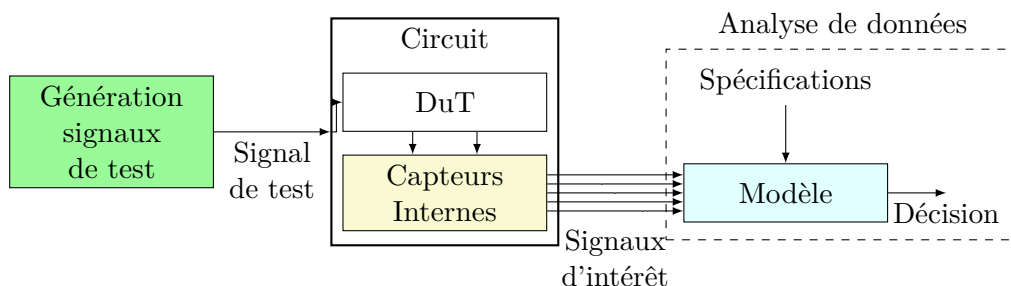


FIGURE I.7 – Schéma simplifié du test indirect

Les mesures alternatives peuvent éliminer les outils de mesure les plus chers de la chaîne de test mais ne mesurent plus directement les performances du circuit. En effet, la mesure (complexe) des performances est remplacée par des mesures plus simples à réaliser de signaux internes, qui sont corrélées aux performances.

Un modèle décrivant le lien entre les performances et les mesures alternatives doit alors être établi *a priori*. Ce modèle est complexe à obtenir en raison de la difficulté à régresser les performances à partir des grandeurs électriques du circuit, comme présenté par le tableau I.1.

Cette démarche, dont la pertinence a été démontrée [Spyronasios et al., 2011, Kupp et al., 2011a], requiert donc une analyse approfondie par un expert du circuit pour établir les points de mesure intéressants, les capteurs à utiliser, et le modèle entre les mesures internes et les performances.

Le test indirect des circuit demande cependant d'extraire des signaux d'intérêt du circuit pour définir les mesures alternatives. Ainsi, bien que ces signaux soient plus simples à mesurer, il est toujours nécessaire d'utiliser un "testeur" qui analyse ces signaux et prend une décision quant à l'état du circuit. Une réponse élégante à cette limitation est apportée par le "BIST".

**BIST.** L'intégration des mesures internes et de la logique de décision est proposée par la méthode du *Built-In Self Test* (BIST) dont la définition est rappelée ci-après.

**Définition 1 (BIST).** Une pratique de conception dans laquelle la logique de test est incorporée dans le circuit d'un élément semiconducteur. Ce circuit peut fournir un test complètement autonome de l'élément (i.e. sans nécessiter de testeur). Il peut être tel qu'un signal d'excitation soit réalisé par un testeur externe. (...) [IEEE, 2005].

Par abus de langage, l'appellation BIST regroupe aussi le recours à la circuiterie de test non nécessairement autonome, c'est-à-dire telle qu'un testeur externe est requis.

Cette définition très large recouvre de nombreuses configurations du test indirect, la plus complète étant représentée par la figure I.8 dans laquelle toutes les fonctionnalités ont été intégrées au BIST : génération des signaux, mesures alternatives et prise de décision. Le plus souvent, le BIST est partiel, et ne réalise qu'une partie de ces fonctionnalités.

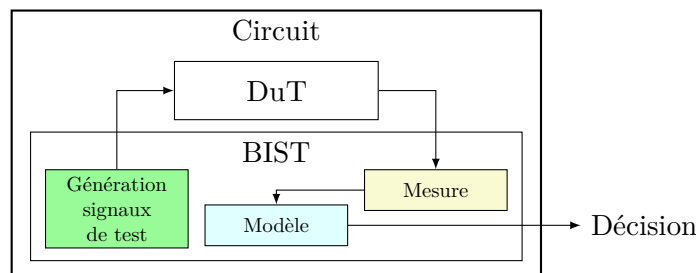


FIGURE I.8 – Schéma simplifié du test indirect par BIST

Le BIST peut réduire l'utilisation d'outils de mesure complexes ou de génération de signaux. Par exemple, [Nassery et al., 2012] développe et utilise un BIST permettant d'extraire l'enveloppe d'un signal pour en déduire certaines spécifications. La mesure de l'enveloppe étant plus simple (car de fréquence moins élevée), le test a été simplifié. Dans ce cas, le BIST implémente un modèle complexe, et la mesure est réalisée par un détecteur d'enveloppe. Dans [Velasco-Medina et al., 1999a], les auteurs proposent d'utiliser des capteurs de courant et des comparateurs de courant. De cette façon, le test consiste à comparer les valeurs en courant mesurées à des références fixes pour réaliser le test.

Le BIST étant ajouté au circuit, la surface de silicium occupé est plus grande et le circuit est plus complexe. En outre, le BIST peut interférer avec le fonctionnement normal du circuit.

De la même manière que pour les mesures internes, il est aussi nécessaire d'identifier les points de mesure pertinents au regard des performances. Ces points de mesure pertinents permettent d'extraire les signaux d'intérêt du

circuit pour en valider son bon fonctionnement.

Enfin, il faut construire un modèle permettant de relier ces signaux d'intérêt aux performances, ou à défaut, à l'état du circuit. Ce modèle doit être intégré dans le circuit, il doit donc être suffisamment simple.

**Loopback.** Le problème de la génération des signaux d'intérêt pour les mesures internes est contourné par l'approche Loopback. Dans le cas particulier des circuits de communication, le procédé de Loopback permet de tester toute la chaîne de communication sans chercher de points de mesure internes, et sans mesurer les performances du bloc d'émission et du bloc de réception indépendamment [Lupea et al., 2003]. Ces deux blocs sont connectés de façon à ce que le bloc d'émission  $T_x$  soit l'équivalent du générateur de signal pour le bloc de réception  $R_x$ . Identiquement, le bloc  $R_x$  est l'équivalent de la mesure pour le bloc  $T_x$ . Les signaux injectés dans le bloc d'émission  $T_x$  sont alors comparés à ceux qui sortent du bloc de réception  $R_x$ , et un modèle permet de déduire les performances du circuit complet (émission et réception) à partir des écarts de signaux observés.

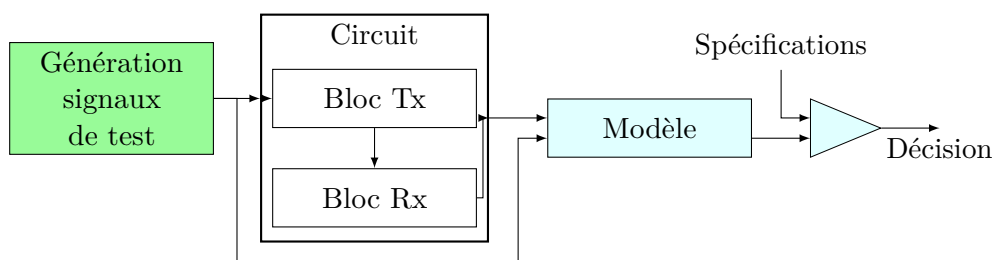


FIGURE I.9 – Schéma simplifié d'un Loopback

Un inconvénient du Loopback est la difficulté du diagnostic. En effet, lorsqu'un test montre que le circuit ne respecte pas les spécifications par Loopback, il est difficile de savoir si l'erreur est dans le bloc  $T_x$  ou  $R_x$ , ou dans les deux. Le défaut est bien détecté, mais pas localisé. [Bhattacharya and Chatterjee, 2004] propose donc d'utiliser de concert le Loopback et la méthode BIST pour résoudre ce problème.

**BIMBO.** La méthode *Basic Identification Method using Binary Observation* (BIMBO) est une réponse élégante au problème d'interaction du BIST avec le DuT, et de limitation du coût du BIST : les mesures sont réalisées par des Convertisseurs Analogique-Numérique (*ADC Analog to Digital Converter*) 1 bit (des comparateurs). Cette mesure 1 bit est utilisée pour identifier un modèle paramétrique du DuT dont la structure est établie *a priori*.

La méthode BIMBO reposant sur une technique d'identification, il est nécessaire de construire un modèle paramétrique *a priori* du DuT. Une connais-

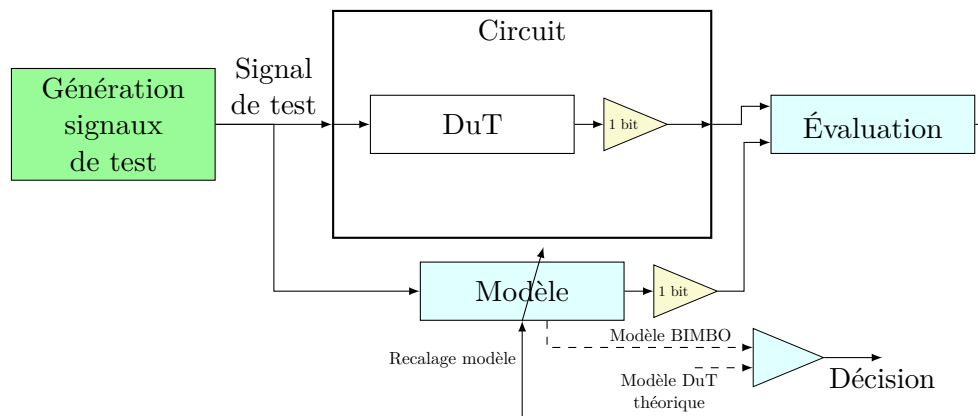


FIGURE I.10 – Schéma simplifié de l'utilisation de la méthode BIMBO

sance profonde du DuT est donc nécessaire. De plus, un générateur de signaux est nécessaire. Il peut être inclus dans le circuit, ou externe, comme représenté sur la figure I.10. De plus, la boucle d'identification des paramètres du modèle peut elle aussi, être externe ou interne comme représenté ici.

### I.B.2.b Organisation du test

Lorsqu'il n'est pas possible de modifier un circuit pour adopter une stratégie de mesures alternatives, ou en complément de cette stratégie, on peut voir la problématique de réduction de coût du test sous l'angle de l'organisation des séquences de tests. Deux approches sont proposées ici, à savoir, l'ordonnancement des séquences de test et la diminution du nombre de tests.

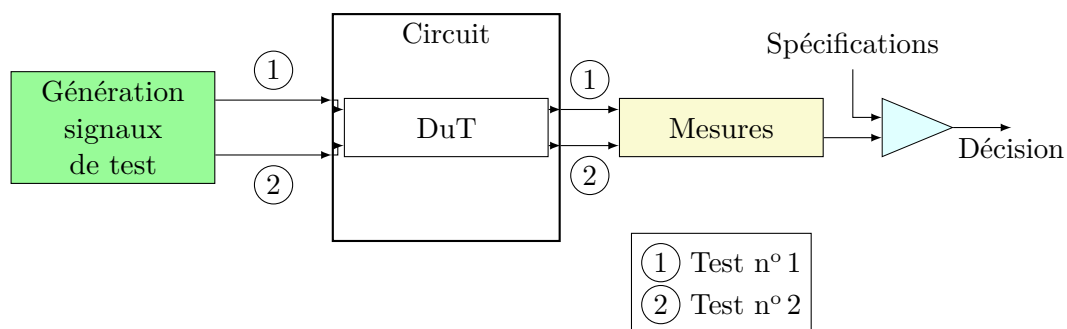


FIGURE I.11 – Schéma simplifié de l'ordonnancement des tests (cas des tests directs)

**Ordonnancement des séquences de test** Cette méthode propose d'ordonner les séquences de test de façon à ce que le temps de test soit diminué. Dans cette configuration, le test d'un circuit est arrêté dès qu'un des tests



du circuit montre que le circuit est défectueux (méthode *Stop on Fail*). Ainsi, réaliser en premier lieu les tests qui échouent le plus souvent (i.e. au rendement le plus faible) permet de diminuer le temps de test moyen des circuits.

Partant de ce constat, [Huss and Gyurcsik, 1991] propose un algorithme permettant d'exécuter le plus tôt possible :

1. les tests qui ont la plus grande probabilité d'échouer ;
2. les tests qui sont exécutés en peu de temps ;
3. les tests qui sont indépendants vis-à-vis des tests antérieurs.

Un avantage important de cette technique est de ne pas diminuer la couverture de test, et de ne pas nécessiter de modélisation *a priori*. Cependant, pour utiliser cet algorithme il est nécessaire de connaître *a priori* les probabilités de détection de faute pour tous les tests ainsi que la probabilité jointe des tests échouant, ce qui peut être obtenu par des simulations Monte-Carlo [Huss and Gyurcsik, 1991, Milor and Sangiovanni-Vincentelli, 1990a].

**Diminution du nombre de tests** Cette méthode est le prolongement de la précédente, dans le sens où les tests qui échouent peu sont éliminés après le tri de la méthode précédente [Milor and Sangiovanni-Vincentelli, 1990b]. Dans ce cas, seuls les tests suffisant à la détection d'une faute paramétrique sont conservés, puis ordonnés. De cette façon, le temps de test est, statistiquement, fortement réduit. Contrairement à l'ordonnancement des séquences de test, la diminution du nombre de tests peut entraîner une diminution de la couverture de rendement.

### I.B.3 Précision du test

La réduction du coût du test s'accompagne naturellement d'une erreur entre le résultat du test à coût réduit et le résultat du test direct complet<sup>3</sup> du circuit. Ainsi, tester des circuits selon une stratégie de test à faible coût peut entraîner une augmentation de la vente de circuits défectueux (faux positifs, erreur de 2<sup>ème</sup> espèce), et un rejet de circuits fonctionnels (faux négatifs, erreur de 1<sup>ère</sup> espèce). Ces deux considérations entraînent d'une part l'insatisfaction des clients et d'autre part, une baisse artificielle du rendement de production.

La définition de ces erreurs est liée à la notion de domaines d'acceptabilité présentés en figure I.12. Ces domaines sont définis dans l'espace des paramètres technologiques, dans l'espace des mesures et dans l'espace des performances. Ils représentent les domaines de variation autorisés dans chacun de ces domaines tels que les performances du circuit respectent les spécifications.

---

3. Par complet, on entend la mesure de toutes les performances et leur comparaison aux spécifications.

La figure I.12 représente schématiquement les régions d'acceptabilité des paramètres technologiques, et des performances. Par ailleurs, l'espace des observations est introduit. Cet espace est une représentation abstraite des observations réalisées pour la réalisation du test. Ce peut être des mesures bruitées, des performances ou encore des mesures alternatives. La zone d'acceptabilité de l'espace d'observation est établie de manière à ce que si une observation y appartient, alors le circuit est déclaré fonctionnel. Il est déclaré défaillant dans le cas contraire.

Puisque la défaillance ou la fonctionnalité d'un circuit est défini par ses performances, la précision du test repose sur la bonne correspondance entre le domaine d'acceptabilité des observations, et celui des performances. Les modèles permettant d'obtenir les performances à partir des paramètres technologiques et les observations à partir des paramètres technologiques existent. Des modèles permettent aussi de relier les performances et les observations aux paramètres technologiques. L'établissement d'une régression entre le domaine de validité des observation et le domaine de validité des performances peut être simple dans le cas où les observations sont des mesures directes des performances, ou complexe si elles sont des mesures alternatives.

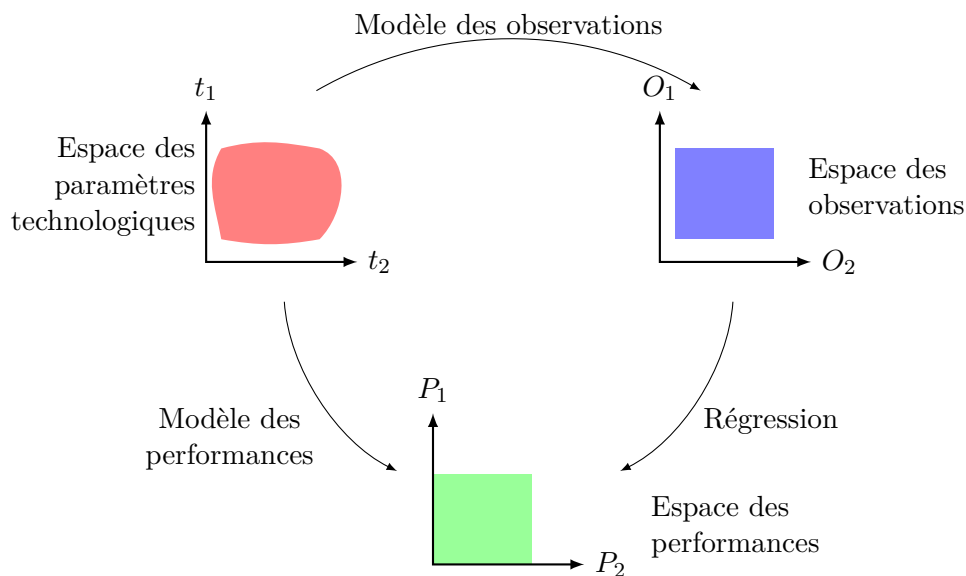


FIGURE I.12 – Représentation des zones d'acceptabilité dans les domaines des paramètres technologiques, mesures et performances

Pour quantifier les erreurs de décision, deux métriques de test sont définies :

- la couverture de faute ;
- la couverture de rendement.

Elles utilisent les domaines de validité tels que présentés ci-avant.

**Couverture de faute.** Pour comprendre la formule adaptée au test des fautes paramétriques, la formule du cas catastrophique sera introduite en premier lieu car la première se déduit de la seconde. La couverture de faute est la proportion de circuits qu'une stratégie de test particulière peut classer défaillants (i.e. qui ne respectent pas les spécifications) par rapport aux circuits qui sont effectivement défaillants. Elle est définie dans le cas des erreurs catastrophiques par :

$$F_{cc} = \frac{\sum_{i=1}^{f_d} p_i}{\sum_{i=1}^f p_i} \quad (\text{I.2})$$

avec :

- $F_{cc}$  Couverture de test des fautes catastrophiques
- $f_d$  Nombre de fautes détectées
- $f$  Nombre de fautes possible
- $p_i$  Probabilité d'apparition de la faute  $i$

[Milor, 1998] adapte cette définition au cas des **fautes paramétriques** par l'utilisation des sommes continues et par l'intégration sur les domaines de validité définis précédemment (les notations sont alors adaptées) :

$$f_{cp} = \frac{\int_{f_d \cap f} P_p(x) dx}{\int_f P_p(x) dx} \quad (\text{I.3})$$

avec :

- $P_p$  Densité de probabilité des paramètres *process*
- $f_{cp}$  Couverture de test des fautes paramétriques
- $f_d$  Ensemble des paramètres technologiques qui génèrent un circuit échouant le test, i.e. hors zone d'acceptabilité des observations
- $f$  Ensemble des paramètres technologiques hors de leur région d'acceptabilité

**Couverture de rendement.** La couverture de rendement représente la proportion de circuits effectivement fonctionnels (i.e. respectant les spécifications) qui réussissent le test. De la même manière que pour la couverture de faute

paramétrique, la couverture de rendement s'écrit [Milor, 1998] :

$$y_c = \frac{\int_{\bar{f} \cap \bar{f}_d} P_p(x) dx}{\int_{\bar{f}} P_p(x) dx} \quad (\text{I.4})$$

avec :

- $\bar{f}$  Ensemble des paramètres technologiques dans leur zone d'acceptabilité
- $\bar{f}_d$  Ensemble des paramètres technologiques qui génèrent un circuit n'échouant pas le test, i.e. dans la zone d'acceptabilité des observations

Les métriques de couverture de faute et de couverture de rendement, qui sont des mesures d'erreur de prise de décision, ont un lien avec l'environnement économique du circuit, à savoir la satisfaction client et la baisse artificielle du rendement de production.

**Mesure des couvertures.** L'évaluation des intégrales des équations (I.3) et (I.4) est en fait réalisée via des expériences Monte-Carlo. La méthode de Monte-Carlo est non déterministe et fournit une approximation de l'intégrale d'une fonction de carré intégrable [Weinzierl, 2000] par l'introduction de l'observateur de Monte-Carlo :

$$I = \int_D f(x) dx \quad (\text{I.5})$$

$$E = \frac{1}{N} \sum_{n=1}^N f(x_n) \quad (\text{I.6})$$

avec :

- $I$  Intégrale à évaluer
- $D \in \mathbb{R}^P$  Domaine d'intégration dont la dimension  $P$  est le nombre de mesures
- $f$  Fonction de carré intégrable
- $E$  Estimateur de Monte-Carlo de  $f$
- $N$  Nombre d'échantillons de  $f$  réalisés
- $x_n \in D$   $n^{\text{ième}}$  échantillon d'évaluation de  $f$

L'évaluation de l'intégrale (I.5) est réalisée par l'estimateur (I.6). En effet, la loi des grands nombres assure la convergence suivante :

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N f(x_n) = I \quad (\text{I.7})$$

Par ailleurs, l'erreur de convergence évolue en  $\frac{1}{\sqrt{N}}$  et ne dépend pas de la dimension  $P$  de  $D$  [Weinzierl, 2000]. Cette indépendance du résultat suivant la dimension du problème rend l'estimateur de Monte-Carlo particulièrement intéressant dans l'évaluation d'intégrales sur des domaines de dimension élevée.

## I.C Synthèse de l'état de l'Art

La problématique adressée dans nos travaux est centrée sur la réduction du coût du test des circuits analogiques, et plus spécifiquement des tests détectant les fautes paramétriques.

Les méthodes de l'état de l'Art présentées dans ce chapitre ont montré que la réalisation du test nécessite une étape d'élaboration de méthodes de test *a priori*. Cette étape permet, entre autres, d'établir les mesures alternatives à réaliser, de construire les modèles de régression, ou encore d'organiser les séquences de test.

Dans le cas du test indirect, des mesures alternatives sont établies par la mesure de signaux d'intérêts. Les signaux d'intérêt sont mesurés aux points de mesure pertinents, qui sont généralement établis heuristiquement ou par une connaissance experte.

Des modèles sont construits pour relier les mesures indirectes aux performances ou à l'état du circuit. Selon la méthode mise en œuvre, des modèles de nature différente sont utilisés. Ainsi, BIMBO utilise un modèle paramétrique dont la structure est connue *a priori*; les mesures alternatives utilisent des modèles de régression, et le BIST peut utiliser des modèles de classification. Quelque soit la méthode mise en œuvre, une étape indispensable à ces approches est donc la construction d'un modèle.

Dans le cas de l'organisation des séquences de test, la connaissance du rendement de chaque test et les probabilités jointes des échecs de chaque test sont utilisées pour ordonner la procédure de test. Se pose alors le problème de l'acquisition de ces informations, en simulation ou en mesure.

Ainsi, la phase d'établissement d'une stratégie de test regroupe différents problèmes :

- Quels signaux mesurer dans le circuit ?
- Comment définir un signal d'intérêt ?
- Comment mesurer les signaux d'intérêt ?
- Est-il possible d'éliminer une mesure ?
- Quels capteur placer pour mesurer un signal d'intérêt, et où le placer ?
- Quel modèle choisir pour lier les mesures alternatives aux performances ou à l'état du DuT ?
- Comment prendre une décision sur l'état du circuit ?

**Présentation de la Vision Unifiée.** Les problèmes auxquels les méthodes de l'état de l'Art présentées répondent peuvent être organisés selon trois catégories, à savoir, l'acquisition d'information sur le circuit, la recherche et la mesure de signaux d'intérêt, et enfin la modélisation et la prise de décision quant à l'état du circuit. Ces trois catégories sont explicitées ci-après.

- **l'acquisition d'information** sur le circuit. Les simulations, mesures ou connaissances expertes sont utilisées pour définir les points de mesure d'intérêt, ou orienter la structure d'un modèle. Une grande quantité d'informations peut être extraite et seules les informations pertinentes sont utilisables pour la recherche des points de mesure d'intérêt et la construction d'un modèle. Le problème de l'acquisition de l'information est donc à la base de la démarche de réduction du coût du test des circuits. Cette étape de mise en place de la stratégie est sous-jacente dans l'état de l'Art présenté dans ce chapitre ;
- **la définition de points de mesure d'intérêt** à partir de ces informations. Cette brique permet l'extraction des informations pertinentes, si elles sont présentes dans les informations disponibles précédemment. Il est alors nécessaire de définir la notion de pertinence ;
- **la construction d'un modèle** qui relie les points de mesure d'intérêt aux performances ou à l'état du circuit. Le modèle issu de cette brique est utilisé pour prendre une décision quant à l'état de chaque circuit testé. La qualité de la prise de décision est évaluée par les couvertures de faute et de rendement présentées en section I.B.3, et elle dépend de la qualité du modèle construit.

**Une Vision Unifiée adaptée à différents problèmes.** Nous allons voir à travers quelques exemples très différents que la résolution de problèmes passe systématiquement par l'approche de la Vision Unifiée proposée.

Ces briques élémentaires peuvent être mises en place et utilisées pour plusieurs tâches. Par exemple, dans le cas de test de systèmes de communication sans-fils à plusieurs fréquences, il est possible que les mesures de performances

sur plusieurs fréquences soient fortement corrélées. Dans ce cas, il est envisageable d'éliminer les mesures sur les fréquences hautes pour ne conserver que les mesures sur les fréquences basses. Ce problème trouve une solution dans la figure I.13. Ici, l'acquisition des données est réalisée par des mesures, l'analyse des corrélations permet de détecter les mesures corrélées entre elles, et enfin, un modèle est construit pour régresser les valeurs des mesures éliminées à partir des mesures conservées.

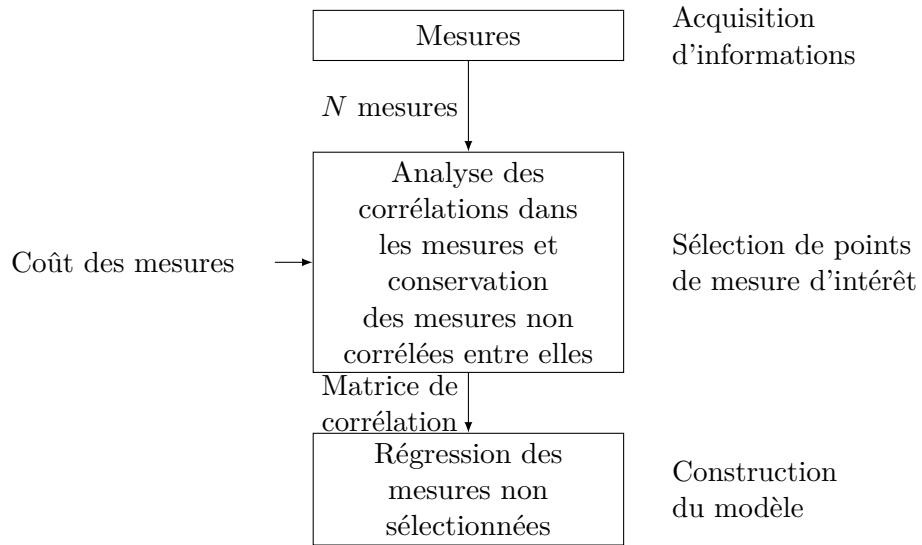


FIGURE I.13 – Élimination de mesure corrélées

L'exemple suivant est celui de la sélection de mesures alternatives. Dans cet exemple, un circuit peut être simulé par Monte-Carlo pour extraire les performances, ainsi que des mesures alternatives. Ensuite, une analyse de corrélations entre les mesures alternatives et les performances permet de sélectionner les mesures alternatives les mieux corrélées aux performances. Enfin, un modèle est construit pour régresser les mesures alternatives sélectionnées aux performances. Cette configuration est représentée sur la figure I.14.

Ces applications sont limitées en raison de l'utilisation de corrélations linéaires, qui ne permet pas de conserver des mesures dont l'impact sur les performances est plus complexe. D'autres outils sont alors proposés, en particulier dans [Stratigopoulos and Makris, 2008] où il est fait usage d'un algorithme à base de réseaux de neurones pour classer les circuits fonctionnels/défaillants, ainsi que pour évaluer la précision de cette classification comme représenté en figure I.15.

L'assemblage des trois briques d'acquisition d'information, de définition

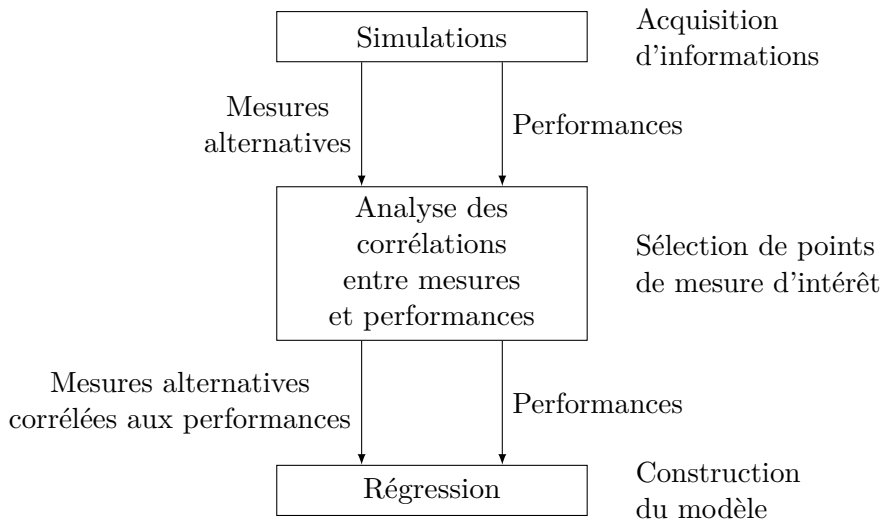


FIGURE I.14 – Conservation des mesures corrélées aux performances

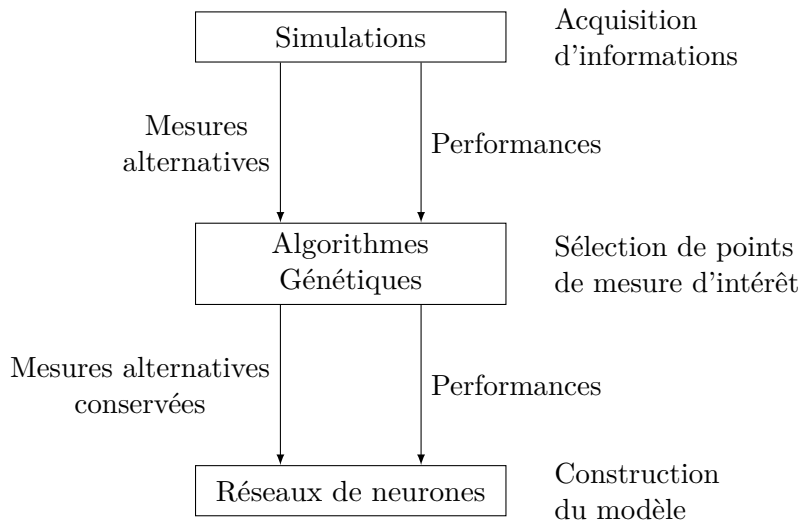


FIGURE I.15 – Sélection des points de mesure d'intérêt et construction d'un modèle de classification [Stratigopoulos and Makris, 2008]

des points de mesure pertinents et de construction du modèle sont à la base de la vision unifiée de réduction des coûts de test des circuits analogiques qui est proposée dans la section suivante.



## I.D Vision unifiée pour la réduction du coût du test

Ce chapitre a présenté le contexte du test des circuits analogiques pour la détection des fautes paramétriques. L'état de l'Art, analysé dans la section précédente, a permis d'exprimer des problèmes variés auxquels il apporte des réponses similaires, et de regrouper les solutions proposées dans la Vision Unifiée. Dans cette section, la Vision Unifiée est étendue pour intégrer la notion de précision du test.

En effet, l'acquisition d'information extrait un nombre fini de d'échantillons. De ces échantillons seront extraits les informations pertinentes puis un modèle sera construit à partir de ces dernières. Il y a ainsi trois sources d'erreurs identifiées : l'ensemble d'échantillons, la sélection des "bonnes" informations et la construction des modèles "adaptés". Il est donc nécessaire de conserver une connaissance des erreurs tout au long du flot.

Cette erreur est à relier au nombre de faux positifs et faux négatifs présentés en section I.B.3. Ces grandeurs permettent d'apprécier directement la faisabilité d'une méthode.

Ces briques sont organisées et présentées dans la Vision Unifiée de la Problématique sur la figure I.16. Il faut maintenant comprendre les méthodes et algorithmes qui peuvent réaliser les objectifs de chacune de ces briques, ainsi que les organiser dans un flot d'établissement de la stratégie de test.

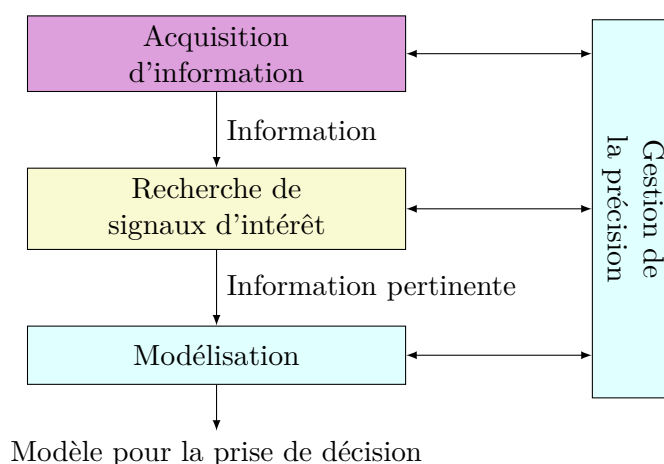


FIGURE I.16 – Vision Unifiée de la Problématique de construction de la procédure de test

La problématique telle qu'exposée dans la Vision Unifiée relève du domaine

du traitement de l'information, et des algorithmes de ce domaine seront utilisés pour répondre au problème de réduction du coût du test.

En particulier, pour la recherche de signaux d'intérêt, des algorithmes de réduction de dimension pourront être mis en œuvre. Il s'agit par exemple de l'Analyse en Composante Principale utilisée dans [Stratigopoulos et al., 2009a] pour extraire d'une simulation les combinaisons de mesures qui permettent d'expliquer la dispersion des performances.

D'autres algorithmes sont aussi utilisés pour répondre à la vision unifiée proposée : [Stratigopoulos and Makris, 2008] propose un algorithme à base de réseau de neurones pour construire un modèle de classification, et [Kupp et al., 2011b] propose un algorithme de sélection de caractéristiques dans le cadre de l'estimation de métriques de test. Concernant la modélisation, des algorithmes de classification comme le Support Vector Machine (SVM) ont été utilisés pour établir le modèle qui sera utilisé dans le diagnostic de défauts [Stratigopoulos and Mir, 2010].

Certains de ces algorithmes, ainsi que d'autres issus du domaine du traitement de l'information sont présentés dans le chapitre II. L'objectif de ce chapitre est de présenter de manière générale les méthodes utilisables pour résoudre des problèmes associés à la réduction du coût du test. Le choix des méthodes présentées est guidé par l'objectif de nos travaux, à savoir "concevoir un flot de conception d'une stratégie de test prenant en compte la réduction du coût". On y retrouve des méthodes utilisables dans les briques de la Vision Unifiée.



## Chapitre II

---

# Traitement de l'information pour les stratégies de test

---

Je ne perds jamais. Soit je gagne,  
soit j'apprends

---

Nelson MANDELA

Le contexte présenté au chapitre précédent a proposé une Vision Unifiée de la Problématique de réduction du coût du test des circuits analogiques. Cette Vision fait intervenir quatre briques fondamentales. Ces briques sont l'*acquisition d'information*, la *recherche de signaux d'intérêt*, la *construction d'un modèle* entre les signaux d'intérêt et les performances ou l'état du circuit, et enfin la *gestion de la précision*. Ce chapitre propose une discussion sur les méthodes et algorithmes existant et pouvant être mis en œuvre dans la Vision Unifiée.

Comme signalé en conclusion du chapitre précédent, des réponses au problème de réduction des coût de test existent et ont été proposées dans le cadre du test de circuits analogiques. Elles sont organisées selon la démarche présentée par la Vision Unifiée figure I.16. De plus, leurs inconvénients et limites ont aussi été discutés, en particulier celui de la sélection de mesures alternatives et de la construction de modèles.

La première section de ce chapitre en précise les objectifs. Ensuite, les outils de l'état de l'Art pouvant être mis en œuvre dans chacune de ces briques sont résumés et discutés. Cette discussion est organisée pour suivre la Vision Unifiée.

## II.A Objectifs

Les techniques de traitement de l'information sont des candidates naturelles pour les briques de la Vision Unifiée présentée précédemment. En effet, l'objectif des techniques de traitement de l'information est d'extraire de l'information utile de données brutes. Cet objectif concorde avec les objectifs de recherche de signaux d'intérêt et de modélisation dans la Vision Unifiée, rappelée sur la figure II.1.

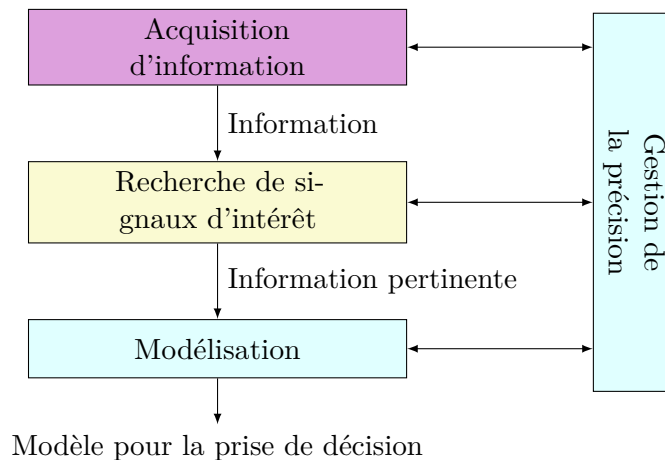


FIGURE II.1 – Vision Unifiée de la Problématique de construction de la procédure de test

Les méthodes de traitement de données peuvent être ordonnées, dans un premier temps, selon deux catégories : tout d'abord, les méthodes pouvant être utilisées dans la recherche des signaux d'intérêt, et ensuite celles liées à la modélisation. Pour le traitement de données, les premières sont dites d'extraction d'information, et les secondes d'induction de modèle. L'extraction d'information consiste à définir et extraire l'information utile contenue dans des données brutes. L'induction du modèle fournit un modèle de prédiction, permettant d'évaluer des grandeurs ou de réaliser une classification à partir de données brutes. Dans le cadre des présents travaux, les grandeurs sont les performances du circuit, la classification est l'indicateur de son respect des spécifications, et les données brutes sont les mesures.

L'organisation des méthodes de traitement de données en méthodes d'extraction d'information et en méthodes d'induction peut être affinée en introduisant deux sous-catégories, qui sont les méthodes supervisées et les méthodes non-supervisées. Les méthodes supervisées utilisent une connaissance supplémentaire aux données brutes qui, dans le cas de nos travaux, est les performances des circuits ou leur classification (entre *défectueux* et *fonctionnels*). Ainsi, leur

	Extraction d'information	Induction de modèle
Supervisé	SIR[Saracco et al., 1999]	kNN[Peterson, 2009]
Non-supervisé	ACP[Cho et al., 2008]	Clustering[Jain et al., 1999]

TABLE II.1 – Une première classification des méthodes de traitement de données, avec quelques exemples <sup>1</sup>

objectif est d'extraire des informations utiles des données brutes, en prenant en compte cette connaissance.

A l'inverse, les méthodes non supervisées n'utilisent pas de connaissance supplémentaire mais recherchent une structure particulière dans les données brutes. Elles permettent, par exemple, de mettre en avant des corrélations ou de faire apparaître des groupes d'échantillons. Le tableau II.1 récapitule cette première classification des méthodes de traitement de données et en propose des exemples.

Compte tenu du contexte du problème que nous adressons dans ces travaux, à savoir la mise en place d'une méthode permettant de réduire le coût du test de circuits analogiques, des méthodes non supervisées permettant d'extraire des structures, en particulier de corrélation et de dispersion, peuvent être mises en place. De plus, comme les performances sont aussi connues, des méthodes supervisées sont utilisables. Ce chapitre va donc étudier des méthodes supervisées et non supervisées pouvant être utilisées pour les briques de recherche des signaux d'intérêt et de modélisation.

Comme l'a conclu le chapitre précédent, la gestion de l'erreur dans la Vision Unifiée joue un rôle important. Elle permet en effet d'anticiper les métriques de test, et d'aider au paramétrage des différents algorithmes utilisés. Ainsi, ce chapitre étudiera la provenance des erreurs et approximations des algorithmes proposés.

## Organisation du chapitre

Le chapitre est organisé pour suivre l'architecture de la Vision Unifiée rappelé en figure II.1.

La section II.B propose une discussion sur les méthodes d'acquisitions d'information. Les méthodes d'acquisition d'informations à partir de mesures ou de simulations sont discutées. En particulier, les méthodes d'acquisition en simulation à base de Monte Carlo sont développées.

---

1. SIR : *Sliced Inverse Regression*, Régression Inverse par Tranches. ACP : Analyse en Composantes Principales. kNN : *k-Nearest Neighbors*, k-Plus Proches Voisins.

Dans la section II.C, les méthodes d'extraction d'information sont présentées selon deux aspects. Tout d'abord, les méthodes de *réduction de dimension* sont introduites. Elles utilisent des projections et leur objectif est de réduire le nombre de dimensions nécessaires à l'observation d'un nuage de points. Ensuite, des méthodes de *sélection de caractéristiques* sont présentées. Leur objectif est de sélectionner les dimensions de l'espace d'origine qui correspondent à une définition donnée de "pertinence".

Ensuite, la section II.D présente plusieurs méthodes permettant de construire des modèles de classification. Ces méthodes sont construites selon différents principes qui sont :

- la régression ;
- le calcul d'une séparation ;
- la recherche de ressemblance ;
- l'analyse de la structure des données disponibles.

Comme précisé précédemment, le problème de précision des résultats doit être abordé aux différentes briques de la Vision Unifiée. Ce sujet sera traité dans la section II.E.

On donne ci-après des notations et définitions utilisées dans le chapitre.

**Notations et définitions.** Soient les notations suivantes :

$$\mathbf{c}_i = [c_{1,i}, c_{2,i}, \dots, c_{O,i}]^T \in \mathcal{C} \quad (\text{II.1})$$

$$\mathbf{C} = [\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_N] \quad (\text{II.2})$$

$$\mathbf{p}_i = [p_{1,i}, p_{2,i}, \dots, p_{P,i}] \quad (\text{II.3})$$

$$\mathbf{y} = [y_1, y_2, \dots, y_N] \quad (\text{II.4})$$

avec :

- $N$  Le nombre de circuits
- $O$  Le nombre d'observations par circuit
- $P$  Le nombre de performances par circuit
- $\mathbf{c}_i$  Le vecteur des  $O$  observations du circuit  $i$
- $\mathbf{p}_i$  Le vecteur des  $P$  valeurs prises par les performances du circuit  $i$
- $\mathbf{C}$  La matrice contenant toutes les observations de tous les circuits
- $\mathbf{y}$  Les vecteurs de la classification des  $N$  circuits
- $\mathcal{C}$  L'espace d'observation

**Définition 2** (Espace d'observation). *La représentation de la matrice  $\mathbf{C}$  dans l'espace d'observation  $\mathcal{C}$  est appelée nuage de points. Cette notion de nuage de points s'étend lorsque  $\mathbf{C}$  est représenté dans les différents espaces de redescription et de projection définis dans la section II.C.1.*

Par soucis de clarté, les notations suivantes sont introduites

- $c(i, j)$  désigne la valeur de la  $j^{\text{ème}}$  observation du  $i^{\text{ème}}$  circuit ;
- $c(i, :)$  désigne toutes les observations du  $i^{\text{ème}}$  circuit ;
- $c(:, j)$  désigne les valeurs prises par tous les circuits pour l'observation  $j$ .

## II.B Acquisition d'information

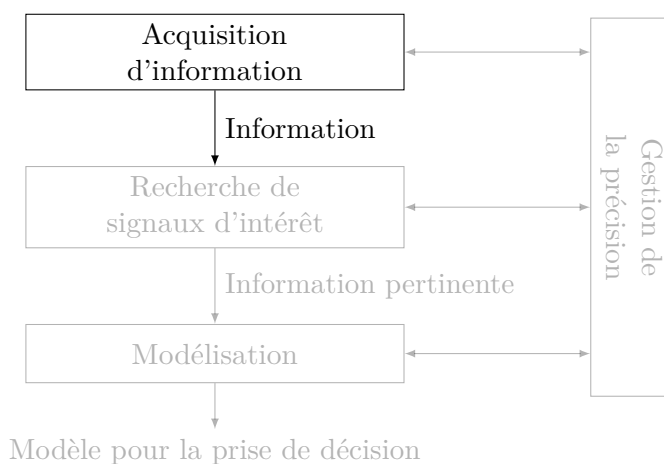


FIGURE II.2 – Acquisition d'information dans la Vision Unifiée

Dans la Vision Unifiée, la brique d'acquisition d'information prend place en début de méthode. Elle permet d'extraire des informations sur le circuit. Ces informations peuvent être des mesures de performances ou des mesures alternatives, mais aussi des données issues de simulations. La capacité à obtenir des informations de simulations ou de mesures a été introduite dans les exemples d'utilisation de la section I.C.

Les problématiques de l'acquisition de données dans les deux cas sont de disposer d'informations représentant "correctement" la dispersion réelle des circuits. Cependant, différents objectifs nécessitent d'acquérir des informations sur le circuit. Par exemple, le cas de la mesure des métriques de test demande d'utiliser des simulations dont la dispersion est proche de celle de la réalité ; la construction de modèle de régression nécessite un parcours aussi complet que possible de l'espace des performances, *etc.*



Dans cette section, différentes méthodes d'acquisition d'information sont présentées. Tout d'abord, celles se rapportant à la simulation, où un contrôle fort est possible sur la génération des échantillons. Ainsi, la méthode de Monte-Carlo est présentée en premier lieu, puis des méthodes utilisant le contrôle de la génération d'échantillons sont introduites. Le contrôle des simulations est utile pour construire des modèles qui soient valables sur un domaine défini de paramètres technologiques. En particulier pour la construction de modèles couvrant régulièrement l'espace des paramètres technologiques, ou de modèles autour de la frontière entre circuits défectueux et fonctionnels.

Les contraintes associées à l'acquisition d'information par mesures sont introduites en section II.B.2.

### II.B.1 Information obtenue en simulation

Les simulations de circuits analogiques utilisent des modèles prenant en compte les dispersions technologiques et leurs impacts sur le circuit. Traditionnellement, des simulations statistiques de Monte-Carlo sont utilisées pour évaluer le rendement de fabrication (la réalisation d'une expérience de Monte-Carlo permet le calcul d'une intégrale de rendement, voir la section II.B.1.a). L'avantage majeur de l'acquisition de données en simulation est la possibilité d'obtenir un grand nombre de données et de contrôler la génération de ces données. En d'autres termes, la simulation peut ne pas suivre la dispersion statistique de la technologie.

Les questions qui se posent alors sont liées à la bonne couverture des "cas possibles" :

- y a-t-il assez d'échantillons pour couvrir l'espace des possibilités ?
- Ces échantillons sont-ils produits de manière à respecter la dispersion réelle des observations ?

La réponse à ces questions diffère selon l'objectif souhaité : construction de modèle, évaluation de rendement, *etc.* Ainsi, différentes méthodes sont présentées pour répondre aux différents objectifs. En premier lieu, la méthode de Monte-Carlo est présentée. Elle permet d'évaluer les métriques liées à la statistique des circuits : calcul des rendements, des faux-positifs, *etc.* Ensuite, une méthode adaptée à la construction de modèles est présentée : le quasi Monte-Carlo. Cette méthode permet un parcours régulier de l'espace des paramètres technologiques et non pas guidé par la dispersion technologiques. La construction d'un modèle autour d'une zone d'intérêt nécessite des échantillons dans cette zone, ce que les méthodes de Monte-Carlo et Quasi Monte-Carlo ne permettent pas. Ainsi, la méthode de Statistical Blockade est introduite dans cette section. Enfin, la méthode de plan d'expériences est présentée.

Elle répond à la modélisation des relations entre performances et paramètres technologiques en limitant le nombre de simulations nécessaires.

### II.B.1.a Nombre d'échantillons générés par Monte-Carlo

L'intégration de Monte-Carlo, présentée en section I.B.3, calcule l'intégrale d'une expression de manière non déterministe. Cette méthode est particulièrement utile pour le calcul d'intégrales en dimension élevée. [Brockman and Director, 1989] l'utilise pour relier la précision attendue sur la mesure du rendement avec le nombre de simulations réalisées.

Pour se faire, il faut rappeler que l'observateur de Monte-Carlo, qui correspond à l'évaluation du rendement, est donné par :

$$\hat{Y} = \frac{n_g}{N} \quad (\text{II.5})$$

avec :

- $\hat{Y}$  L'observateur de Monte-Carlo
- $N$  Le nombre total de circuits dont les performances sont mesurées
- $n_g$  Le nombre de circuits observés dont les performances respectent les spécifications

Pour caractériser l'erreur commise lors de l'évaluation du rendement, [Brockman and Director, 1989] propose de relier le rendement, le nombre d'échantillons disponibles, l'erreur sur le rendement et la confiance dans l'erreur par la relation suivante :

$$N = \frac{(1 - Y) Y}{\epsilon^2} \left[ \phi^{-1} \left( \frac{1 + \rho}{2} \right) \right]^2 \quad (\text{II.6})$$

avec :

- $N$  Le nombre d'échantillons minimal nécessaire pour assurer la précision requise sur le calcul du rendement
- $Y$  La vraie valeur du rendement ou, à défaut, une évaluation de la valeur minimale du vrai rendement
- $\epsilon$  L'erreur entre le rendement réel et le rendement observé
- $\rho$  La confiance dans l'erreur sur le rendement
- $\phi^{-1}$  La fonction inverse de la fonction de distribution normale standard

La justification de cette formule établie par [Brockman and Director, 1989] est rappelée en annexe A. Elle exprime le nombre d'échantillons à observer pour respecter une erreur déterminée avec un intervalle de confiance donné. Dans le cadre de nos travaux, elle permet de savoir *a priori* si la base de données peut respecter les contraintes d'erreur exigées.

Un avantage de la méthode de Monte-Carlo est que l'erreur ne dépend pas de la dimension des échantillons. Cependant, la génération des échantillons respecte la dispersion naturelle de la technologie, ce qui ne permet pas de parcourir identiquement tout l'espace des paramètres. Ce faisant, la construction de modèles de régression ou de classification peut être biaisée par un grand nombre d'échantillons dans une région de l'espace, et l'absence d'échantillons dans une autre région. Ce déséquilibre est caractérisé par la *discrédance* qui est présentée dans la section suivante.

### II.B.1.b Génération à discrédance minimale

La construction d'un modèle entre les performances et les paramètres technologiques permet de comprendre quels sont les paramètres technologiques qui sont susceptibles de dégrader les performances d'un circuit. L'obtention de modèles précis sur tout l'espace d'observation demande de le parcourir régulièrement. Or la dispersion des paramètres technologiques ne conduit pas forcément à une répartition régulière des échantillons dans l'espace des paramètres technologiques. En effet, lors de la génération pseudo-aléatoire des points, des zones de l'espace d'observation peuvent être couvertes par plus de points que d'autres. La figure II.3 montre visuellement deux générations pseudo-aléatoires possibles. La figure II.3a utilise une génération uniforme pseudo-aléatoire. Cette génération laisse apparaître des zones où la densité de points est plus faible. À l'inverse, la figure II.3b est plus homogène. Les points utilisés pour cette dernière sont issus d'une génération de Sobol[Niederreiter, 1988].

La *densité de points* est définie par la discrédance. [Singhee et al., 2008] indique qu'il existe plusieurs définitions de la discrédance. La plus usitée est la *Star Discrepancy*. Cette grandeur caractérise la non-uniformité d'une séquence de points. Elle est donnée par la définition 3.

**Définition 3** (Star Discrepancy). *Soit une séquence  $\mathbf{x} = \{x_0, \dots, x_{N-1}\}$  de  $N$  points appartenant à un cube  $\bar{\mathbf{I}}^s = [0, 1]^s$  de dimension  $s$ . La Star Discrepancy  $D_N^*(\mathbf{x})$  est définie par [Thiérmard, 2001] :*

$$D_N^*(\mathbf{x}) = \sup_{P \in \mathcal{F}^*} \left| \frac{A(P, \mathbf{x})}{N} - \text{vol}(P) \right| \quad (\text{II.7})$$

avec :

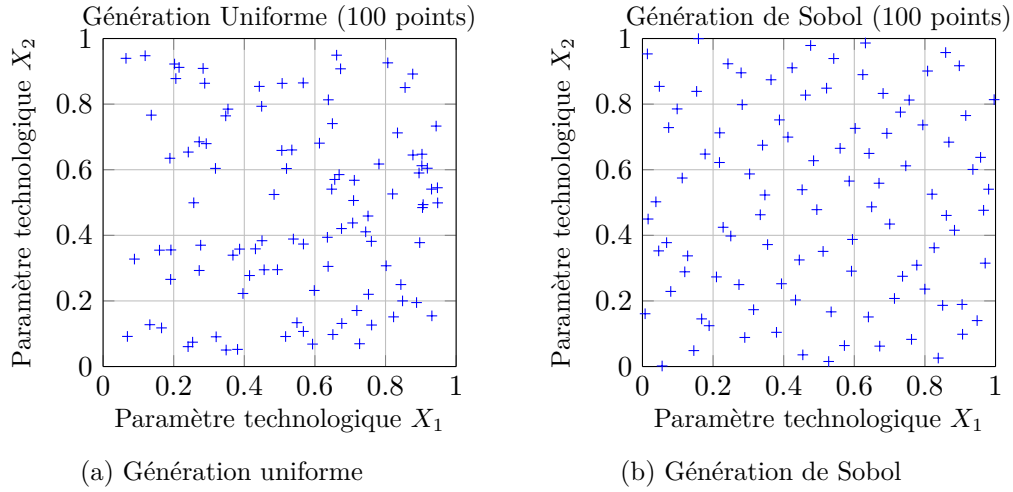


FIGURE II.3 – Deux générations différentes d'échantillons

- $\mathcal{F}^*$       *La famille de tous les intervalles de  $\mathbf{I}^s = [0, 1]^s$*
- $\text{vol}(P)$     *Le volume de l'intervalle  $P$*
- $A(P, \mathbf{x})$    *Le nombre de points de  $\mathbf{x}$  appartenant à l'intervalle  $P$*

Dans le cadre du test de circuit, la discrédance quantifie la répartition irrégulière des échantillons dans l'espace d'observation. Plus elle est faible, plus les échantillons sont régulièrement répartis dans l'espace. Une telle génération est intéressante pour la construction de modèles dont la précision est alors la même dans tous l'espace d'observation. Typiquement, l'utilisation d'échantillons générés par une simulation de Monte-Carlo ne permet pas de décrire précisément le comportement du circuit autour des spécifications par manque d'échantillons. Une génération à discrédance minimale des échantillons permet de décrire régulièrement les circuits sur tout l'espace d'observation.

Le calcul de cette grandeur est complexe pour  $s \geq 2$ , et [Thiémond, 2001] propose un algorithme permettant de calculer les bornes supérieures et inférieures de  $D_n^*(\mathbf{x})$ . Cet algorithme repose sur une décomposition de  $\bar{\mathbf{I}}^s = [0, 1]^s$  paramétrée par l'erreur souhaitée entre les bornes inférieures et supérieures de  $D_n^*(\mathbf{x})$ . L'algorithme a l'avantage de ne pas requérir l'enregistrement de toute la décomposition. Cependant, le nombre de partitions générées par la décomposition croît rapidement avec la dimension des échantillons. Cette quantité est donc un indicateur de la qualité de la couverture de l'ensemble d'observation.

Dans le cas de la simulation, il est possible d'assurer la génération d'une

séquence de points à discrédance minimale. Dans ce cas, la simulation est dite de *Quasi Monte-Carlo*.

Plusieurs algorithmes existent pour générer des séquences qui minimisent la discrédance. La figure II.3b utilise l'algorithme de Sobol pour générer une séquence de points dans deux dimensions à discrédance minimale.

Le Quasi Monte Carlo assure le parcourt régulier de l'espace d'observation, mais la précision des modèles dépend du nombre de point couvrant l'espace d'observation. Pour modéliser une zone restreinte de l'espace d'observation, par exemple la limite de spécification des circuit, il n'est pas utile de modéliser tout l'espace, qui demande de simuler un grand nombre de point. La méthode du statistical blockade permet de ne simuler qu'une zone particulière de l'espace d'observation, diminuant ainsi le nombre de points nécessaire pour l'établissement d'un modèle local.

### II.B.1.c Statistical Blockade

Le *Statistical Blockade* est un outil de contrôle de la simulation qui permet de ne simuler que les circuits dont les performances sont à la limite des spécifications [Singhee and Rutenbar, 2009]. En effet, les déviations statistiques font que les circuits dont les performances sont proches des spécifications peuvent avoir une faible probabilité d'apparition, comme représenté en figure II.4.

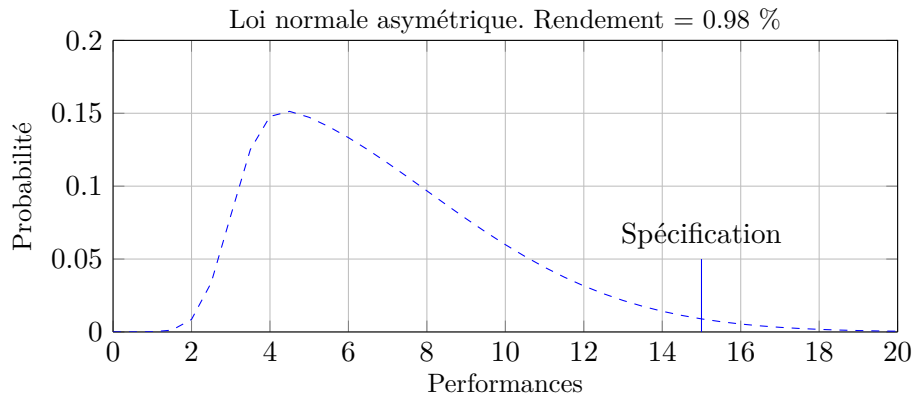


FIGURE II.4 – Exemple de répartition de performances

De ce fait, beaucoup de points seront nécessaires pour avoir suffisamment d'échantillons autour de la spécification. Le *Statistical Blockade* suppose qu'il est plus rapide de générer un ensemble de valeurs de paramètres technologiques que de simuler un circuit pour en obtenir les performances. Ainsi, [Singhee and Rutenbar, 2009] propose de réaliser des simulations sur l'espace complet des

paramètres avec peu d'échantillons, puis d'entraîner un classifieur pour ne sélectionner par la suite que les échantillons susceptibles d'avoir des performances proches des spécifications. L'organigramme de cet algorithme est reproduit en figure II.5, d'après [Stratigopoulos and Sunter, 2014]. Un modèle doit donc être construit et entraîné pour détecter les paramètres technologiques qui génèrent des performances proches des spécifications.

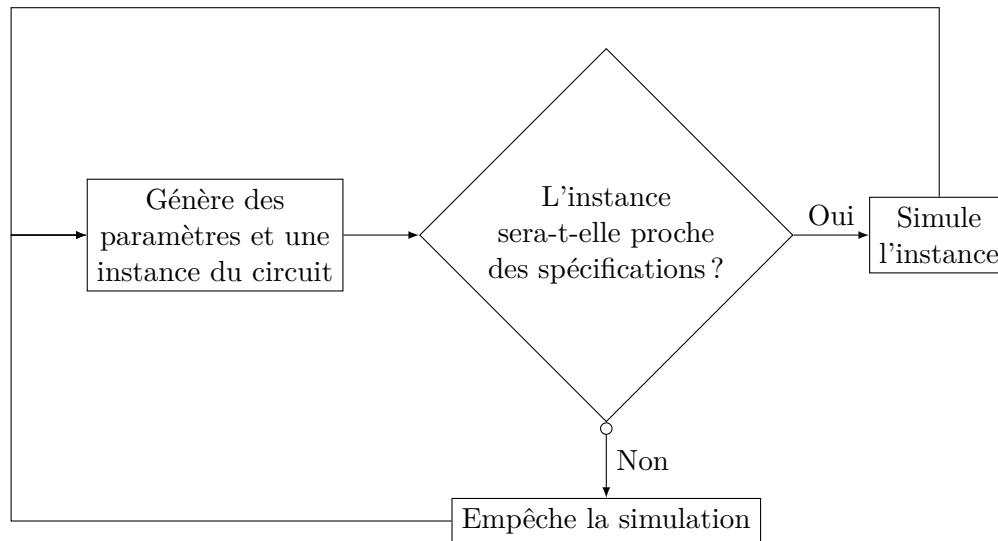


FIGURE II.5 – Organigramme du *Statistical Blockade*, d'après [Stratigopoulos and Sunter, 2014]

Dans l'exemple de la figure II.6, un modèle est simulé en utilisant la méthode de *Statistical Blockade*. Un classifieur de Machine à Vecteurs de Support (voir section II.D.2) est entraîné sur la distribution des paramètres technologiques avec 50 échantillons. Ensuite, seuls des échantillons proches des spécifications sont entraînés. La sélection des échantillons à simuler est réalisée par le classifieur.

La répartition des performances autour de la spécification ainsi obtenue est présentée sur la figure II.7. Cette répartition permet d'avoir plus de points autour de la spécification, pour construire un modèle autour de cette dernière.

La réponse à la question de la proximité de l'instance avec les spécifications est apportée par le classifieur préalablement entraîné. Cette méthode a été mise en place, par exemple, dans [Spyronasios et al., 2011] pour évaluer les couvertures de faute et de rendement, dans le cadre de test indirect. Pour nos travaux, cette génération autour des spécification peut être utilisée pour construire des modèles précis sans nécessiter la simulation de nombreux circuits qui, en suivant la dispersion des paramètres technologiques, seront en majorité

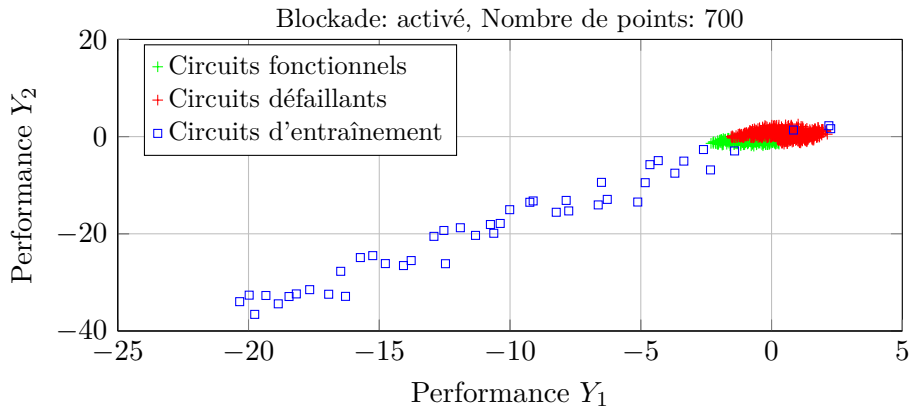


FIGURE II.6 – Représentation de simulations utilisant le Statistical Blockade

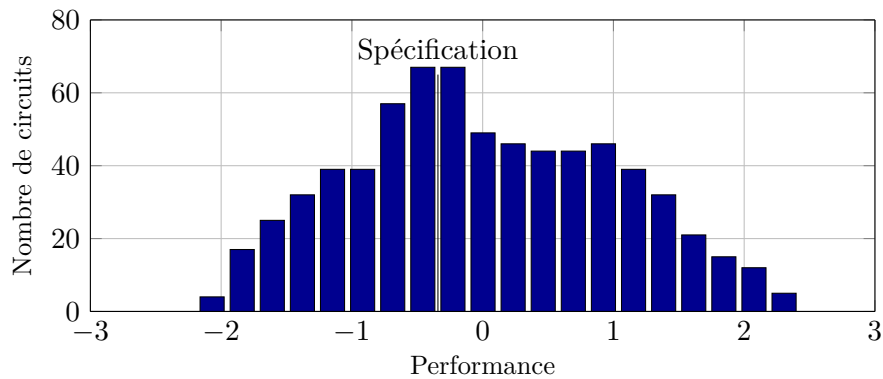


FIGURE II.7 – Histogramme d'une performance obtenue par *Statistical Blockade*

éloignés des spécifications (voir figure II.4).

#### II.B.1.d Plan d'expériences

Cette approche de contrôle de la génération des échantillons, aussi connue sous le nom de DOE - *Design of Experiment*, permet de construire des modèles de régression entre les paramètres technologiques et les performances. Son intérêt réside dans le fait de limiter le nombre d'échantillons à simuler pour construire des modèles de structure connue *a priori*.

Le nombre de simulations requises pour parcourir toutes les variations de tous les paramètres technologiques peut être très important. Les plans d'expérience permettent de modifier des paramètres technologiques d'une simulation à la suivante pour extraire l'impact sur les performances de ces paramètres technologiques. L'objectif est alors de trouver un plan d'expériences qui diminue le nombre de simulations requises. En notant  $P_t$  le nombre de paramètres tech-

nologiques et  $k$  la quantification de chaque variable technologique, le nombre d'expériences nécessaires pour couvrir tout le domaine technologique s'élève au maximum à  $k^{P_t}$  [Zhang and Styblinski, 1995] (plan d'expériences factoriel complet). L'utilisation d'un plan d'expériences différent, qui réduit le nombre d'expériences à mener, peut cacher une influence croisée des paramètres technologiques. Le choix d'un plan d'expériences demande donc une connaissance *a priori* de la complexité du circuit.

Pour répondre à cette problématique de sélection du plan d'expériences, [Rémond and Necessian, 2011] propose une méthode permettant de construire un plan d'expériences utilisé pour la modélisation des performances d'un LNA à partir des variables technologiques. Pour nos travaux, cette méthode fournit un modèle reliant les performances aux paramètres technologiques. Bien qu'elle limite le nombre de simulations requises pour l'établir, elle utilise une classe de modèles définie *a priori*.

#### II.B.1.e Limites de la simulation

Des problèmes inhérents à la simulation sont les contraintes en temps et en espace mémoire des simulateurs. Ces limitations amènent à réduire le nombre d'échantillons à simuler pour établir un modèle. C'est à cette question que répond la méthode de Quasi Monte Carlo. Lorsque qu'un modèle est connu *a priori*, l'utilisation de plan d'expérience permet de ne simuler que les échantillons utiles au paramétrage dudit modèle.

Les simulations utilisent des modèles décrivant le comportement physique et électrique des circuits. Ces modèles peuvent être incomplets ou incorrectement représenter le comportement des circuits observé en réalité. Dans ces travaux, nous considérons ces modèles comme étant suffisamment précis pour expliquer la réalité.

#### II.B.2 Informations obtenues *via* des mesures

L'acquisition de données par mesures se différencie de celle par simulation par plusieurs aspects. Tout d'abord, elle n'est pas contrôlable comme le sont certaines approches par simulation (DOE ou Quasi Monte-Carlo par exemple). Ensuite la différence entre simulations et mesures touche aux questions de bruits de mesure. Enfin d'autres aspects liés à l'acquisition de données interviennent. Ils sont maintenant développés.

**Tests de continuité.** La mesure des circuits pour le test se déroule schématiquement en deux phases. La première phase est celle dite de *test de continuité* qui permet d'assurer que le circuit est testable sans endommager les appareils



de mesure. Cette première phase de test est critique et ne peut être ni évitée, ni modifiée. Si le circuit testé réussit les tests de continuité, alors le test peut continuer. Dans le cas contraire les tests ne sont pas poursuivis. Il s'agit d'un test grossier qui élimine les circuits ouvertement défectueux.

**Stop on Fail.** La deuxième phase est celle des mesures de spécifications et de signaux internes. Afin de diminuer le temps de test, l'approche *Stop on Fail* est utilisée : le test est interrompu dès qu'une mesure indique que le circuit est en dehors des spécifications. Cette approche permet de réduire le temps de test de façon à ne pas mesurer des circuits connus défectueux. Cependant, il ne permet pas la création d'une base de données complète. Une base de données complète est importante pour nos travaux car la recherche des signaux d'intérêt et la modélisation considèrent identiquement les circuits défectueux et les circuits fonctionnels. La figure II.8 représente le comportement *Stop on Fail*. Les tests des circuits sont représentés dans une image, où les pixels bleus sont les tests réussis, et les pixels blancs sont les tests échoués ou non réalisés.

Remarquons que la construction d'une base de données complète peut être obtenue par une campagne de mesure dite de *caractérisation* qui permet aussi de valider les modèles de déviation statistique des circuits.

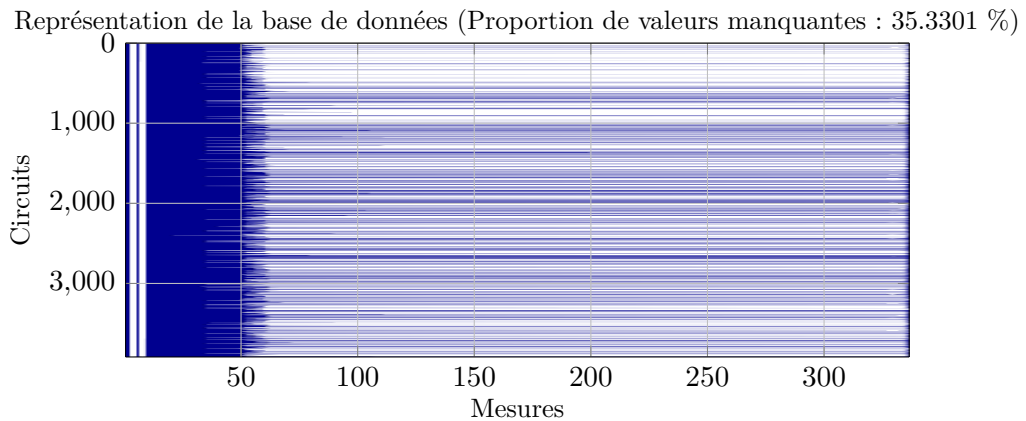


FIGURE II.8 – Illustration du comportement *Stop on Fail* en mesure

**Confidentialité des données.** Contrairement au cas de la simulation, il n'est pas toujours possible d'obtenir des informations sur les grandeurs mesurées. La réalisation de mesures complémentaires est généralement difficile car coûteuse. De plus, la confidentialité des données impose de ne pas avoir accès à la conception du circuit testé, ni à des connaissances d'expert sur le circuit. Il est donc difficile d'accéder à la signification des données mesurées ce qui pose problème lors de la validation des algorithmes et de l'interprétation

des résultats.

### II.B.3 Synthèse sur l'acquisition de données

Cette section a présenté les méthodes d'acquisition de données dans les cas de simulations et de mesures. Une première question qui se pose lors de l'acquisition de données est celle du nombre d'échantillons requis. La réponse apportée par [Brockman and Director, 1989] est celle de la confiance et de l'erreur commise sur le calcul du rendement. Certaines des méthodes proposées montrent qu'il est possible d'utiliser les simulations pour construire des modèles sans recourir à la simulation d'un grand nombre d'instances du circuit.

L'assurance de correctement couvrir l'espace des paramètres technologiques est décrit par deux grandeurs. Tout d'abord par l'équation (II.6) qui décrit la précision de la mesure de rendement dans le cas de simulation Monte-Carlo, puis par les notions de discrédance, dont une définition est fournie en définition 3.

Le cas de l'acquisition de données par mesures est particulier dans le sens où il n'offre pas le contrôle que permet la simulation. De plus, les erreurs qui atteignent les circuits mesurés ne sont pas nécessairement induites par des fautes paramétriques, mais par d'autres erreurs en-dehors du sujet de ces travaux : fautes catastrophiques ou fautes de conception par exemple. Cependant, ces mesures sont utiles pour valider le comportement statistique des circuits (rendement, *etc.*) et pour augmenter la précision des modèles.

De même, l'étude statistique des mesures impose qu'un nombre minimal de circuits aient été mesurés. La mesure de ces circuits peut représenter un coût important qui peut être évité par l'utilisation des simulations.

## II.C Recherche de signaux d'intérêt

L'objectif de la recherche de signaux d'intérêt est de diminuer le nombre de signaux à mesurer, pour réduire la complexité du test de circuit. Comme présenté en section I.A.2, un grand nombre de paramètres peuvent expliquer la valeur des performances observées, ce qui rend complexe la régression des performances à partir des mesures électriques.

Ainsi, la recherche de signaux d'intérêt prend place après l'acquisition d'information sur le circuit, voir la figure II.9. Elle permet d'extraire uniquement l'information "pertinente" de cette acquisition.

Deux approches sont possibles. La première dite de *réduction de dimension* analyse les données brutes et les transforme dans un espace de dimension

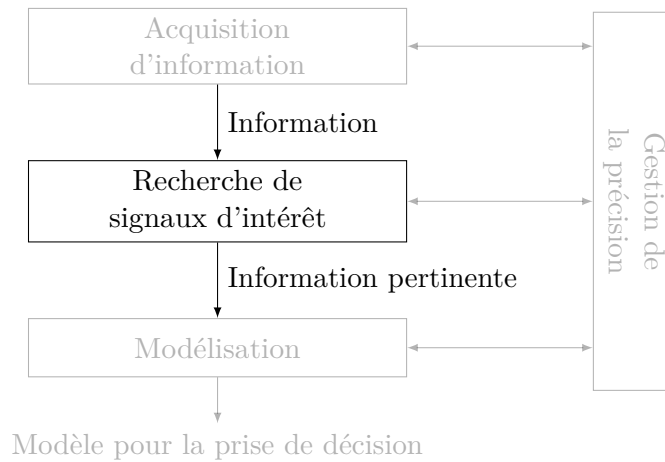


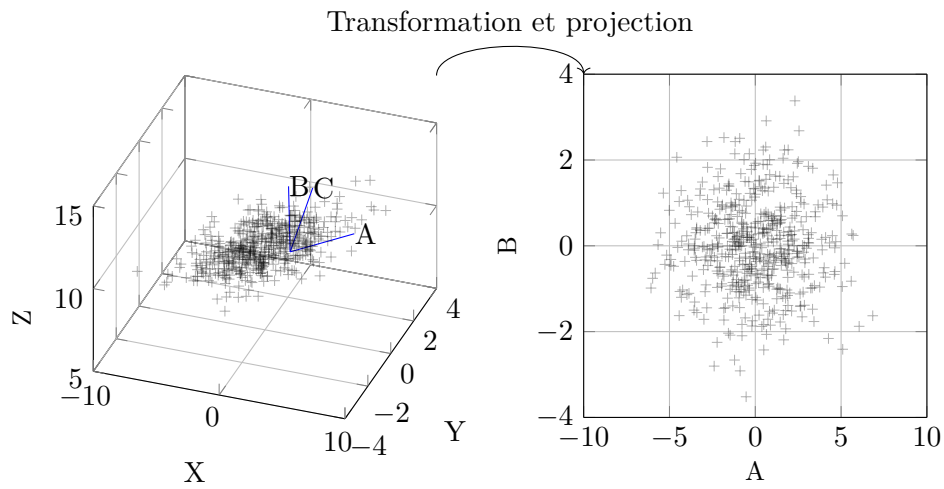
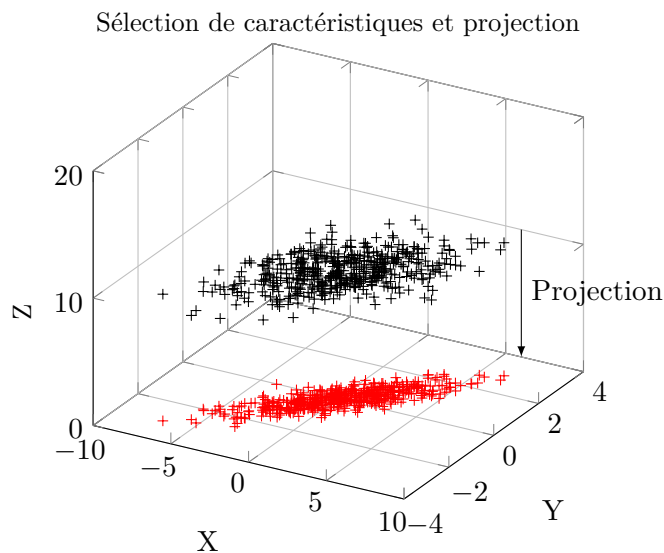
FIGURE II.9 – Recherche de signaux d'intérêt dans la Vision Unifiée

plus faible. Cette projection est représenté sur la figure II.10. Dans cette figure, un nuage de points est représenté dans deux espaces. Le premier est l'espace d'observation  $\{X, Y, Z\}$ . Dans cet espace le nuage de point a une représentation complexe dans le sens où trois coordonnées doivent être utilisées pour décrire le nuage. Le second espace utilisé est l'espace de projection  $\{A, B, C\}$  dans lequel le nuage peut être représenté "plus simplement", c'est-à-dire en utilisant moins de coordonnées pour décrire chacun des points. Dans l'exemple proposé ici, le critère de transformation choisi est l'explication de la variance du nuage. Ainsi, la direction  $A$  supporte la plus grande proportion de la variance du nuage, la direction  $B$  en supporte moins, et la direction  $C$  supporte la plus faible part de variance<sup>2</sup>. Ainsi, selon l'application visée, il est possible de négliger l'observation suivant l'axe  $C$  tout en conservant la majeure partie de la variance du nuage d'origine. La dimension est alors réduite.

La deuxième approche est celle de *sélection de caractéristiques*. Dans ce cas, des composantes du nuage de points sont éliminées selon un critère établi *a priori*. Cette approche est similaire à une projection représentée en figure II.11. Dans cet exemple, la dimension  $Z$  a été éliminée, ce qui permet de ne conserver que les dimensions  $X$  et  $Y$ . Dans ce cas, les dimensions conservées sont dites *pertinentes* relativement à une performance ou à une classification. Les algorithmes réalisant la sélection de caractéristiques recherchent donc à conserver les dimensions pertinentes.

On détaille maintenant ces deux approches issues de l'état de l'Art.

2. Cet exemple est un cas simple de l'Analyse en Composante Principale qui sera développée ci-après.

FIGURE II.10 – La *réduction de dimension* est une transformationFIGURE II.11 – La *sélection de caractéristiques* est une projection

### II.C.1 Réduction de dimension

Les techniques de réduction de dimensions permettent de traiter un problème, exprimé initialement dans un espace de dimension élevée, dans un espace de dimension plus faible. Ces méthodes reposent sur la transformation des données d'origine dans un espace de dimension plus faible, en conservant “le mieux possible” les informations qu’elles contiennent. Les informations d’intérêt sont, dans les techniques présentées ici, la variance du nuage de points.

Dans cette section, on présente l'Analyse en Composantes Principales (PCA) et la Régression Inverse par Couche (SIR). Les vecteurs directeurs de l'espace de projection expliquent la dispersion statistique.

On verra ensuite que les problèmes non linéaires peuvent aussi être traités par les méthodes d'Analyse en Composantes Principales et de Régression Inverse par Couche *via* l'introduction de fonctions *noyaux*. Cette approche, connue sous le nom de *Kernel Trick*, permet de remplacer une projection par un calcul de produit vectoriel.

### II.C.1.a Analyse en Composantes Principales

L'objectif de l'Analyse en Composantes Principales (PCA) est de représenter un nuage de points selon les directions de variance maximale grâce au calcul des vecteurs propres de la matrice de covariance.

On note  $\mathbf{C}$  la matrice contenant les observations des  $N$  circuits considérés, de dimension  $O \times N$ . Les observations peuvent être des performances ou des mesures alternatives. L'Analyse en Composantes Principales permet d'obtenir la matrice  $\mathbf{L}$ , qui est la projection de  $\mathbf{C}$  dans l'espace  $\mathcal{L}$  de dimension  $L$  inférieure à la dimension  $O$  de l'espace d'observation  $\mathcal{C}$ . Les vecteurs de la base de  $\mathcal{L}$  sont tels que [Janecek et al., 2008] :

1. ils forment une base orthonormée ;
2. ils sont des combinaisons linéaires des vecteurs de la base de l'espace d'observation  $\mathcal{C}$  ;
3. ils décrivent la plus grande variation du nuage de points  $\mathbf{C}$ .

La transformation de  $\mathcal{C}$  dans  $\mathcal{L}$  est définie par :

$$\mathbf{L} = \mathbf{Q}^T \mathbf{C} \tag{II.8}$$

avec :

- $\mathbf{L}$  La matrice projetée de  $\mathbf{C}$  dans  $\mathcal{L}$
- $\mathbf{Q}^T$  La matrice de transformation de  $\mathcal{C}$  vers  $\mathcal{L}$ , de taille  $[L \times O]$

Pour construire la matrice de projection  $\mathbf{Q}$ , il suffit de rechercher les  $L$  plus grandes valeurs propres de la matrice de covariance  $\mathbf{C}_C$  de  $\mathbf{C}$ . Les vecteurs propres associés, notés  $[\mathbf{u}_1, \dots, \mathbf{u}_L]$ , sont alors les vecteurs de changement de base :

$$\mathbf{Q}^T = [\mathbf{u}_1, \dots, \mathbf{u}_L] \tag{II.9}$$

La démonstration de ce résultat est reportée en annexe B.

Cette décomposition spectrale permet de fixer la dimension de l'espace de projection à  $L$ . Les  $\mathbf{u}_L$  sont les  $L$  premiers vecteurs propres de la décomposition spectrale de  $\mathbf{C}_C$ , correspondant au  $L$  plus grandes valeurs propres.

La matrice projetée  $\mathbf{L}$  contient alors les projections des observations des  $N$  circuits dans l'espace de dimension  $L$ . Ces dimensions sont alors décrites par des combinaisons des dimensions d'origines, qui n'ont pas forcément de sens physique.

Les directions à variances maximales sont maintenant explicitées par la matrice  $\mathbf{Q}$ . Il reste cependant à déterminer le nombre  $L$  de ces directions à conserver pour projeter le nuage de points.

Une méthode simple de calcul de  $L$  consiste à décider à l'avance de la quantité  $K$  de dispersion totale à conserver :

$$\frac{\sum_{i=1}^{L_m} \lambda_i}{\sum_{i=1}^O \lambda_i} \geq K \quad (\text{II.10})$$

avec :

- $L_m$  La plus petite valeur de  $L$  telle que l'équation (II.10) soit vraie
- $O$  Le nombre total de valeurs propres, qui correspond au nombre d'observations par circuits, issu de l'analyse spectrale de  $C_C \in \mathbb{R}^{[O \times O]}$

Cette méthode permet d'obtenir une représentation de la dispersion du nuage de points dans un espace de faible dimension. Elle est utilisée par [Stratigopoulos et al., 2009a] pour conserver  $K = 99.9\%$  de la dispersion du nuage de points. Dans [Peres-Neto et al., 2005], les auteurs proposent d'autres méthodes pour choisir le nombre  $L$  de dimensions à conserver.

**Remarques.** Dans le domaine du test de circuits, cette méthode permet de représenter les circuits dans les directions de dispersion maximale. Par exemple, [Stratigopoulos et al., 2009a] propose d'analyser la fonction de densité de probabilité à plusieurs variables (ou loi jointe) des mesures alternatives après avoir simplifié la représentation des données par la méthode PCA.

En fournissant des combinaisons des composantes du nuage de points qui expliquent sa dispersion, on perd le sens physique des variables qui seront par la suite manipulées.

En outre, la méthode PCA ne peut pas utiliser les performances ou la classifications des circuits d'entraînements. Ces données peuvent être établies en simulation mais ne peuvent pas être exploitées par cette méthode non supervisée. Ainsi, cette limite de la méthode PCA est adressée par la méthode

de Régression Inverse par Tranches, qui réalise une analyse spectrale exploitant la valeur des performances.

### II.C.1.b Régression Inverse par Tranches

L'objectif de la Régression Inverse par Tranches (SIR) est, comme pour l'Analyse en Composantes Principales, de construire un espace de dimension plus faible que l'espace d'observation. Dans le cas de la méthode SIR, l'espace de projection est appelé espace EDR (*Effective Dimension-Reduction*). Contrairement à l'Analyse en Composantes Principales, la méthode SIR est utilisée pour une régression : elle fait donc intervenir les performances des circuits.

La tâche de régression non paramétrique consiste à évaluer la relation  $f$  entre une performance  $p_i \in \mathbb{R}$  et des observations  $\mathbf{c}_i \in \mathbb{R}^O$  telles que [Saracco et al., 1999] :

$$p_i = f(\mathbf{c}_i) + \epsilon \quad (\text{II.11})$$

où  $\epsilon$  représente l'erreur de régression à minimiser, indépendante de  $\mathbf{c}_i$ .

Lorsque la dimension  $O$  de  $\mathbf{c}_i$  est grande, la régression de  $f$  requiert des échantillons de grande taille. Une solution à ce problème est apportée par [Li, 2012] qui propose de préalablement projeter les  $\mathbf{c}_i$  de dimension  $O$  sur un espace de dimension  $K < O$ , tel que le problème de régression soit exprimé par

$$p_i = f\left(\beta_1^T \mathbf{c}_i, \beta_2^T \mathbf{c}_i, \dots, \beta_K^T \mathbf{c}_i, \epsilon\right) \quad (\text{II.12})$$

avec :

- $\beta_i$  Le  $i^{\text{ème}}$  vecteur colonne inconnu défini sur  $\mathbb{R}^O$ ,  $i \in \{1, \dots, K\}$
- $f$  La fonction de régression définie sur  $\mathbb{R}^{K+1}$
- $\epsilon$  Le terme d'erreur indépendant de  $\mathbf{c}_i$

La méthode SIR introduit alors une transformation monotone  $T$  permettant d'écrire équation (II.12) sous la forme :

$$T(p_i) = (T \circ f)\left(\beta_1^T \mathbf{c}_i, \dots, \beta_K^T \mathbf{c}_i, \epsilon\right) \quad (\text{II.13})$$

ce qui simplifie le calcul de la matrice de covariance. En effet, la transformation à l'origine de la méthode SIR est la transformation par tranches où  $T$  introduit une discrétisation des valeurs de  $y$ .

Régresser  $f$  dans le cas de l'équation (II.13) est alors simplifié par la réduction de la dimension de l'ensemble d'observation en raison de la diminution

de la dimension introduite par les  $\{\beta_k\}_{k \in \{1, \dots, K\}}$ . Notons que la méthode SIR s'intéresse à l'évaluation des vecteurs  $\{\beta_k\}_{k \in \{1, \dots, K\}}$ , mais pas à la régression de  $f$ .

Cette évaluation des vecteurs  $\beta_k, k \in \{1, \dots, K\}$  est réalisée par l'introduction de la courbe de régression inverse :

$$p_i \mapsto E[\mathbf{c}_i | T(p_i)] \quad (\text{II.14})$$

$$\Sigma_{\mathbf{c}_i \mathbf{c}_i} = \text{var}(E[\mathbf{c}_i | T(p_i)]) \quad (\text{II.15})$$

où  $E$  représente l'espérance statistique.

Lorsque  $T$  est une discrétisation de  $p_i$  en  $H$  tranches  $s_1, \dots, s_H$ , alors la matrice de covariance  $\Sigma_{\mathbf{c}_i \mathbf{c}_i}$  est exprimée par :

$$\Sigma_{\mathbf{c}_i \mathbf{c}_i} = \sum_{h=1}^H R_h \mathbf{m}_h \mathbf{m}_h' \quad (\text{II.16})$$

avec :

- $\Sigma_{\mathbf{c}_i \mathbf{c}_i}$  La matrice de covariance de  $E[p_i | \mathbf{c}_i]$
- $H$  Le nombre de tranches définies par la transformation  $T$
- $R_h$  La proportion des valeurs de  $p_i$  appartenant à la tranche  $s_h$
- $\mathbf{m}_h$  La moyenne des valeurs de  $\mathbf{c}_i$  tel que  $p_i \in s_h$

L'hypothèse ci-après est fondamentale pour obtenir la base de l'espace EDR.

**Hypothèse 1.** Soit  $\mathbf{b} \in \mathbb{R}^O$ , alors l'espérance  $E[\mathbf{b}^T \mathbf{c}_i | \beta_1^T \mathbf{c}_i, \dots, \beta_K^T \mathbf{c}_i]$  est linéaire en  $\beta_1^T \mathbf{c}_i, \dots, \beta_K^T \mathbf{c}_i : \exists \alpha_k \in \mathbb{R}, k \in \{1, \dots, K\}$  tels que

$$E[\mathbf{b}^T \mathbf{c}_i | \beta_1^T \mathbf{c}_i, \dots, \beta_K^T \mathbf{c}_i] = \sum_{k=1}^K \alpha_k \beta_k^T \mathbf{c}_i \quad (\text{II.17})$$

Les vecteurs de l'espace EDR sont alors les  $K$  vecteurs propres associés aux  $K$  plus grandes valeurs propres de  $\Sigma_{\mathbf{c}_i \mathbf{c}_i}$ . La démonstration est développée en annexe C.

Le choix du nombre  $H$  de tranches et la définition des tranches sont traités dans [Saracco et al., 1999, Li, 2012]. En particulier, il est nécessaire d'avoir  $H > K$  pour retrouver au moins les  $K$  vecteurs EDR. Quant à la définition des tranches, il est préférable de les définir non pas comme un découpage régulier de l'ensemble des valeurs prises par les  $p_i$ , mais comme un découpage permettant d'avoir "suffisamment" d'individus dans chaque tranche [Saracco et al., 1999]. Par exemple, un découpage en quantiles assure d'avoir approximativement la même quantité d'individus dans chaque tranche.



**Remarques.** La méthode SIR propose de calculer une projection du nuage de points d'origine dans un espace EDR qui explique la dispersion du nuage de points au regard de la variation de  $p_i$ , sans évaluer une fonction de régression  $f$  de  $\mathbf{c}_i$  vers  $p_i$ . La méthode d'Analyse en Composantes Principales n'extrait que les directions de dispersion maximale du nuage de points de manière non supervisée. La méthode SIR est *a contrario* supervisée puisqu'elle est basée sur l'analyse de  $\text{var}(E[\mathbf{c}_i|p_i])$ .

Cependant, les dimensions de  $\mathbf{c}_i$  qui ont une influence importante sur la valeur de  $p_i$  ne sont pas nécessairement celles qui répondent au problème de réduction du coût de la procédure de test. Par ailleurs, les dimensions de l'espace EDR restent une combinaison linéaire des dimensions d'observation, et ces combinaisons n'ont pas de sens physique.

L'hypothèse 1 est une hypothèse forte qui n'est pas facilement vérifiable. Les auteurs de [Kuentz and Saracco, 2010, Saracco et al., 1999, Li, 2012] signalent qu'il est possible de la remplacer par l'hypothèse de distribution elliptique symétrique des mesures. Cette hypothèse est plus forte que l'hypothèse 1, mais plus simple à vérifier en pratique.

Les méthodes SIR et PCA possèdent une limitation intrinsèque due à la projection des données d'origine dans un espace. Cette projection est adaptée au cas où les relations entre observations peuvent être décrites par des relations linéaires. Le cas non linéaire nécessite l'utilisation de la méthode à noyaux présentée dans la section suivante.

### II.C.1.c Cas non linéaires Méthodes à Noyaux

Les solutions PCA et SIR ne sont pas adaptées lorsque le nuage de points d'origine n'est pas décrit dans un espace, mais dans une variété (*manifold*), comme représenté en figure II.12.

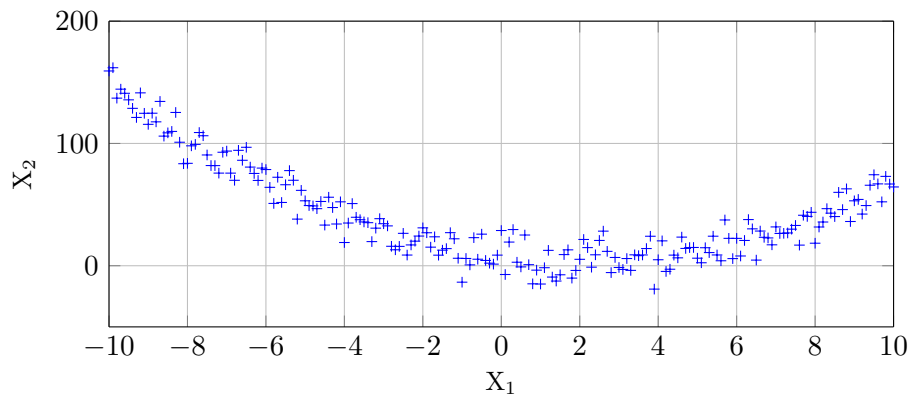


FIGURE II.12 – Exemple de représentation d'un nuage de points projetable sur une variété

Dans ce cas, la méthode à noyaux consiste à projeter le nuage de points dans un espace de dimension supérieure à l'espace d'observation, appelé *espace de redescription*, pour le représenter linéairement. Il est ensuite possible d'appliquer les méthodes de réduction de dimension précédentes (Analyse en Composantes Principales ou Régression Inverse par Tranches).

La projection  $\phi$  est définie par :

$$\begin{aligned} \phi : \mathbb{R}^O &\longrightarrow \mathbb{R}^L \\ \mathbf{c} &\longmapsto \phi(\mathbf{c}) \end{aligned} \quad (\text{II.18})$$

avec :

- $\phi$  Une projection du nuage de points
- $O$  La dimension de l'espace d'observation,  $O < \infty$
- $L$  La dimension de l'espace de redescription,  $L \in \mathbb{N}$

En général,  $L \gg O$ , voire  $L$  peut être infini. Dans ce cas, on ne peut pas appliquer les méthodes PCA ou SIR.

Pour résoudre le problème de dimension de l'espace de redescription, l'astuce du noyau (appelée *Kernel Trick*) représente la projection comme un produit scalaire par l'introduction d'un *Noyau* [Hofmann et al., 2008] tel que défini par :

$$\begin{aligned} KT : \mathbb{R}^O \times \mathbb{R}^O &\longrightarrow \mathbb{R} \\ (\mathbf{x}_i, \mathbf{x}_j) &\longmapsto k(x_i, x_j) = \langle \phi(x_i), \phi(x_j) \rangle \end{aligned} \quad (\text{II.19})$$

Ce noyau autorise alors l'utilisation de PCA ou SIR, y compris dans les cas où l'espace de redescription est infini. L'astuce permet de ne pas décrire le nuage de points dans l'espace de redescription, mais d'utiliser à la place le noyau tel que défini en II.19 dans l'expression des méthodes PCA et SIR.

Un exemple d'illustration apporté en annexe D propose une application du *Kernel Trick* à une Analyse en Composante Principale.

**Remarques.** L'utilisation du *kernel trick* permet l'application des méthodes de réduction de dimension PCA et SIR à des nuages de points qui ne sont pas linéairement séparables. Cette technique demande la projection du nuage de points dans un espace de redescription de grande dimension dans lequel les méthodes PCA ou SIR s'appliquent. L'avantage est que le calcul de la projection n'est pas explicite. Il est remplacé par le calcul d'un produit scalaire, bien plus facile à évaluer.

Pour que l'utilisation du *kernel trick* soit possible, il faut cependant avoir une connaissance *a priori* de la variété sur laquelle projeter le nuage de points. Dans le cas de l'Analyse en Composantes Principales, [Saerens et al., 2004] applique la méthode à noyaux dans l'analyse de graphes. [WU, 2007] applique la méthode à noyaux pour l'utilisation de la Régression Inverse par Tranches.

Pour nos travaux, la méthode à noyau est utilisée pour les appliquer des méthodes linéaires a des cas non-linéaires pour lesquels il est possible de connaître un noyau *a priori*. Dans notre étude, ce ne sont pas seulement PCA et SIR qui utilisent le noyau, mais aussi le classifieur SVM présenté en section II.D.2.

### II.C.1.d Conclusion sur la réduction de dimensions

Les méthodes de réduction de dimensions permettent d'extraire les axes sur lesquels la variance du nuage de points est maximale. La méthode PCA calcule ces directions indépendamment des performances des circuits. Ainsi, elle extraira les même directions quelque soit la performance d'intérêt. *A contrario*, la méthode SIR généralise la démarche PCA en incluant la performance d'intérêt par la régression inverse. Ainsi les dimensions extraites dépendent de la performance considérée.

Dans le cadre du test des circuits analogiques, ces méthodes retournent comme résultats des combinaisons des observations, qui n'ont plus de signification physique. Cependant, les grandeurs extraites sont celles qui expliquent la dispersion des circuits dans l'espace des mesures. Par ailleurs, PCA et SIR supposent que les dimensions les plus intéressantes pour le test sont celles sur lesquelles le nuage de points a la plus grande variance.

Ces deux méthodes proposent des transformations "linéaires". Le *Kernel Trick* permet d'utiliser ces méthodes pour des transformations sur des variétés. Cependant, il est nécessaire de trouver un noyau qui décrive correctement la projection par une analyse en amont de leur application.

Les techniques de réduction de dimension ne répondent donc que partiellement au problème de réduction des coûts du test : elles permettent de réduire la dimension du problème en sélectionnant un ensemble de combinaisons de mesures qui explique la dispersion du nuage de points. Les mesures qui justifient la dispersion du nuage de points ne sont pas nécessairement celles qui expliquent la classification du circuit, ou le respect d'une performance en particulier. De plus, la notion de coût des mesures est absente de cette réduction de dimension.

Dans la section suivante, on propose une autre approche de recherche des signaux d'intérêt qui tienne compte de la relation entre les observations et la

classification des circuits.

### II.C.2 Sélection de caractéristiques

La sélection de caractéristiques, au contraire de la réduction de dimension (section II.C.1), permet d'extraire les dimensions d'intérêt dans l'espace d'observation directement, sans chercher à projeter le nuage de points dans un espace de dimension plus faible.

Dans cette section, l'appellation *caractéristique* est utilisée pour désigner les *observations* des circuits (mesures ou simulations). En effet, le terme *caractéristique* est le terme d'usage dans le contexte du traitement de données qui nous intéresse ici.

On formalise maintenant la notion de pertinence d'une caractéristique qui est un critère traduisant l'explication qu'apporte une caractéristique à classification observée d'un ensemble de points. Par la suite, deux algorithmes d'intérêt pour la sélection de caractéristiques sont présentés, à savoir Relief et Focus. Le choix de ces algorithmes, détaillé en section II.C.2.b, s'appuie sur l'analyse rapportée dans [Liu and Yu, 2005] qui introduit des critères de comparaison des algorithmes dédiés à la sélection de caractéristiques. Par la suite, les algorithmes Focus et Relief sont développés et discutés en sections II.C.2.c et II.C.2.d.

Dans un premier temps, les observations sont considérées binaires  $\mathbf{c}_i \in \{0, 1\}^O, \forall i \in \{1, \dots, N\}$ . De même,  $y_i \in \{0, 1\} \forall i \in \{1, \dots, N\}$ . Ces simplifications permettent d'introduire les algorithmes de sélection de caractéristiques. Le caractère continu des mesure est adressé par la suite. On désigne par  $C_j$  la  $j^{\text{ème}}$  caractéristique.

#### II.C.2.a Notion de Pertinence d'une caractéristique

Les algorithmes de sélection de caractéristiques reposent sur la définition de *caractéristique pertinente*. [John et al., 1994] rappelle plusieurs définitions de la pertinence d'une caractéristique. Il propose deux nouvelles notions qui sont la *pertinence forte* et la *pertinence faible*.

Pour ces définitions, [John et al., 1994] introduit une notation supplémentaire à celles de la section II.A :  $\mathbf{S}_i = \{C_j, j \in \{1, \dots, P\} \setminus \{i\}\}$  représente l'ensemble des  $O$  caractéristiques d'un circuit privé de la caractéristique  $C_i$ .

**Définition 4** (Pertinence forte [John et al., 1994]). *Une caractéristique  $C_i$  est pertinente si et seulement s'il existe  $x, y, \mathbf{s}$  pour lesquels  $\Pr(C_i = x, \mathbf{S}_i = \mathbf{s}) > 0$*

tels que<sup>3</sup>

$$\Pr(Y_i = y|C_i = x, \mathbf{S}_i = \mathbf{s}) \neq \Pr(Y_i = y|\mathbf{S}_i = \mathbf{s}) \quad (\text{II.20})$$

où  $Y_i$  représente la classification.

Cette définition signifie qu'une caractéristique fortement pertinente est absolument nécessaire et ne peut pas être éliminée sans une perte d'information sur la classification du circuit. Une caractéristique fortement pertinente ne peut donc pas être remplacée par une fonction des autres caractéristiques.

Dans le cadre de nos travaux, une observation fortement pertinente est une observation nécessaire pour connaître l'état du circuit, et il n'est pas possible de la remplacer par d'autres observations potentiellement moins chères.

**Définition 5** (Pertinence faible [John et al., 1994]). *Une caractéristique  $C_i$  est faiblement pertinente si elle n'est pas fortement pertinente et s'il existe un sous-ensemble  $\mathbf{S}'_i \subseteq \mathbf{S}_i$  pour lequel il existe  $x, y, \mathbf{s}$  avec  $\Pr(p_i = x, \mathbf{S}'_i = \mathbf{s}) > 0$  tels que :*

$$\Pr(Y_i = y|p_i = x, \mathbf{S}'_i = \mathbf{s}) \neq \Pr(Y_i = y|\mathbf{S}'_i = \mathbf{s}) \quad (\text{II.21})$$

Une caractéristique faiblement pertinente est donc une caractéristique qui peut être remplacée par une combinaison d'autres caractéristiques.

La suite de cette section expose une organisation des méthodes de sélection de caractéristiques, et justifie le choix de deux algorithmes pouvant répondre à la problématique de réduction du coût de test, à savoir FOCUS et RELIEF. Un premier inconvénient notable des définitions ci-avant sont qu'elles ne manipulent que des grandeurs binaires. Nous verrons que les algorithmes présentés s'étendent aux grandeurs continues.

### II.C.2.b Différents algorithmes de sélection de caractéristiques

Les algorithmes de sélection de caractéristiques utilisés dans ces travaux prennent place en amont de la construction d'un modèle de classification, comme indiqué dans la Vision Unifiée. Les informations accessibles à ce stade incluent les observations (mesures ou simulations) des échantillons d'entraînement et leur classification (ou leur performances qui peuvent être comparées aux spécifications connues). Ce sont donc des algorithmes supervisés, qui seront sélectionnés.

Les algorithmes de sélection de caractéristiques peuvent être organisés selon trois catégories : *filtre*, *wrapper* et *hybride* [Liu and Yu, 2005].

---

3. La notation  $\Pr(X = x)$  est valable car  $X$  est une variable aléatoire discrète

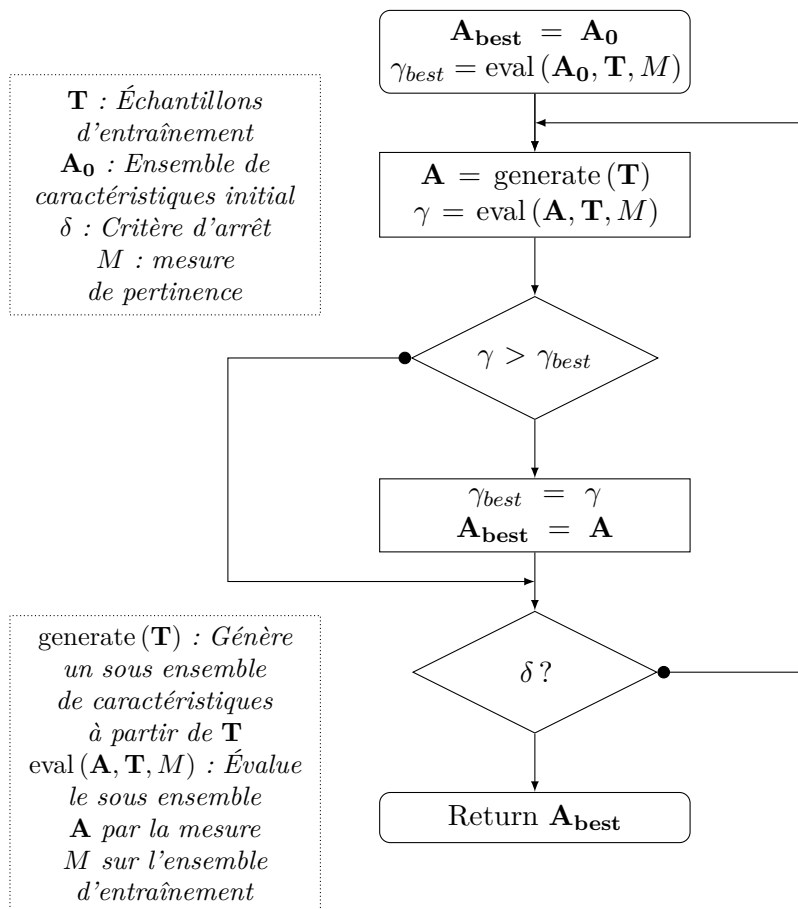


FIGURE II.13 – Algorithme *filtre* pour la sélection de caractéristiques (reproduit d'après [Liu and Yu, 2005]). **A<sub>best</sub>** est le sous ensemble de caractéristiques le plus pertinent selon la mesure de pertinence **M**

Les algorithmes de la catégorie *filtre* (voir figure II.13) ont la particularité de ne pas dépendre d'autres algorithmes. Ainsi, le résultat n'est basé que sur la définition de la pertinence au travers d'une mesure de pertinence **M** et des données d'entraînement. *A contrario*, les algorithmes *wrappers* (figure II.14) dépendent d'un algorithme tiers de classification. Les algorithmes de type *hybride* utilisent la mesure de pertinence **M** pour sélectionner différents ensembles de caractéristiques potentiellement solution et utilisent aussi un algorithme tiers pour extraire l'ensemble le plus pertinent selon la mesure de pertinence **M** parmi ceux-là.

[Liu and Yu, 2005] compare différents algorithmes de type *filtre*, *wrapper* ou *hybride*, selon les critères de stratégie de recherche (fonction generate(.)), d'évaluation (fonction eval(.)) et de l'objectif de la sélection (classification ou

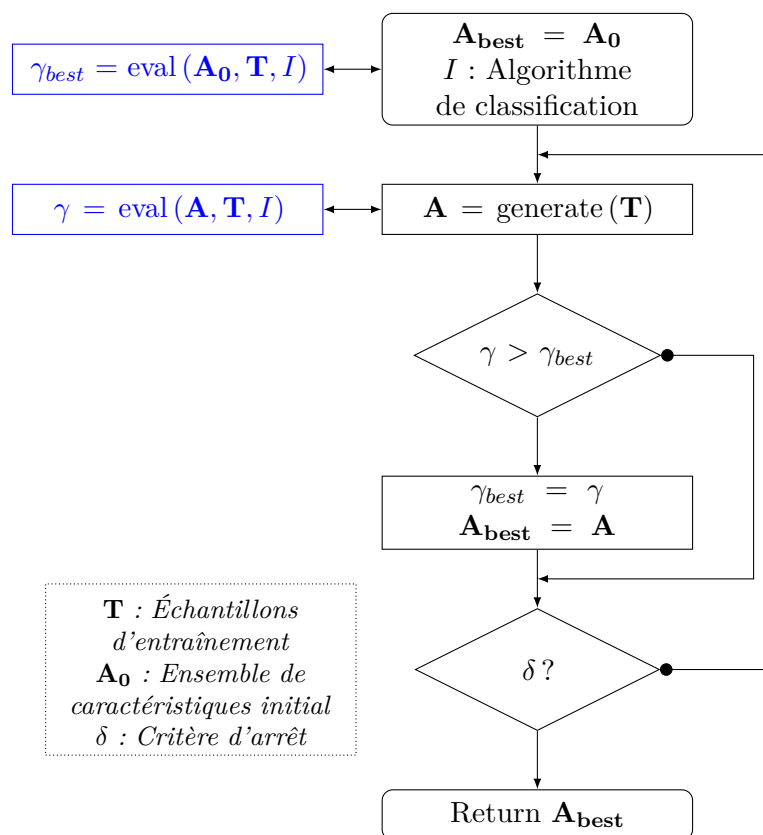


FIGURE II.14 – Algorithme *wrapper* de sélection de caractéristiques (reproduit d’après [Liu and Yu, 2005]). Dans ce cas, l’évaluation des sous ensembles est réalisée par un algorithme de classification externe.

partitionnement).

Dans [Liu and Yu, 2005], les auteurs proposent de les organiser selon trois catégories : la stratégie de recherche (notée *generate* dans les figures II.13 et II.14), le critère d’évaluation (noté *eval*), et leur objectif (classification ou partitionnement). Différentes stratégies de recherche (fonction *generate* (.) dans la figure II.14) sont possibles. Dans nos travaux, la volonté de conserver des résultats déterministes permet d’éliminer les stratégies de recherche de type aléatoire. De plus, nous adressons des problèmes de classification, et non pas de partitionnement. Enfin, les structure *filter* seront préférées car elles ne nécessitent pas l’utilisation d’un classifieur lors de la recherche des caractéristiques.

Ainsi, les algorithmes retenus pour nos travaux sont FOCUS, qui a une stratégie de recherche “complète”, et RELIEF, qui a une stratégie de recherche “séquentielle”.

FOCUS, de par sa stratégie de recherche complète, parcourt donc potentiellement la totalité des ensembles de caractéristiques possibles. Ce comportement, avec le défaut d'être de complexité élevée ( $\Theta(2^N)$ ), a l'avantage d'assurer l'extraction de l'ensemble de caractéristiques le plus pertinent. Cependant, il extrait le sous-ensemble de caractéristiques de cardinal le plus faible, mais pas de coût le plus faible. Il ne répond donc pas à l'exigence de diminution du coût adressé dans nos travaux.

L'algorithme RELIEF a une stratégie de recherche séquentielle. Ainsi, il ajoute (ou retire) des caractéristiques à chaque itération de `generate(.)`. Ce comportement a l'avantage d'être plus rapide que la recherche complète ( $\sim \Theta(N^2)$ ) mais il risque de fournir comme résultat un ensemble de caractéristiques qui n'est pas optimal, c'est-à-dire qui n'est pas l'ensemble de caractéristiques pertinent au coût le plus faible.

L'algorithme FOCUS utilise comme critère d'évaluation une mesure de cohérence qui assure que les caractéristiques extraites suffisent à justifier la classification observée dans l'ensemble d'entraînement. RELIEF évalue les ensembles de caractéristiques par leur capacité à discriminer les classes de l'ensemble d'entraînement, à partir d'une mesure de distance entre les échantillons.

Dans la suite, ces deux algorithmes seront développés et leur utilisation détaillée. Ce développement permettra de présenter une modification de FOCUS qui réponde au problème de sa complexité algorithmique.

### II.C.2.c Algorithme de sélection de caractéristiques à recherche séquentielle : RELIEF

RELIEF est un algorithme de sélection de caractéristiques qui associe un poids  $W$  à chaque caractéristique. Ce poids est calculé de manière itérative et les caractéristiques au poids le plus élevé sont désignées comme les plus pertinentes [Kira and Rendell, 1992]. L'algorithme fonctionne sur une analyse statistique des données. Soit un nuage de  $N$  points  $\mathbf{C}$  où chaque instance  $\mathbf{c}_i, i \in \{1, \dots, N\}$  est classifiée binairesment. L'algorithme sélectionne  $M$  instances aléatoirement et, pour chaque instance  $\mathbf{c}_s$  sélectionnée, il recherche l'instance  $\mathbf{c}_h$  la plus proche de  $\mathbf{c}_s$  de même classe que  $\mathbf{c}_s$  et l'instance  $\mathbf{c}_m$  la plus proche de classe différente de  $\mathbf{c}_s$ <sup>4</sup>. Le poids de chaque caractéristique est alors mis à jour par la formule suivante :

$$W_{new}(C_i) = W_{old}(C_i) - \frac{\text{diff}(S, H, C_i)}{m} + \frac{\text{diff}(S, M, C_i)}{m} \quad (\text{II.22})$$

avec :

---

4.  $\_s$  :selected,  $\_h$  : hit,  $\_m$  : miss



$W_{new}(C_i)$	Le nouveaux poids de la caractéristique $C_i$
$W_{old}(C_i)$	L'ancien poids de la caractéristique $C_i$
$\text{diff}(X, Y, C_i)$	La distance entre un échantillon $X$ et un échantillon $Y$ projeté sur la caractéristique $C_i$
$S$	Une des $M$ instances sélectionnées aléatoirement
$H$	L'instance la plus proche de $S$ qui est de la même classe que $S$
$M$	L'instance la plus proche de $S$ de classe différente
$m$	Le numéro de l'itération $m \in \{1, \dots, M\}$

Comme l'indiquent les auteurs de l'algorithme dans [Kira and Rendell, 1992], la complexité de RELIEF est  $\Theta(OMN)$ , qui se ramène à  $\Theta(ON)$  puisque  $M$  est une constante fixée arbitrairement, ce qui rend la sélection des caractéristiques pertinentes linéaire selon le nombre total de caractéristiques  $O$  et le nombre d'instances d'entraînement  $N$ .

L'algorithme RELIEF extrait les caractéristiques fortement pertinentes, mais aussi des caractéristiques pouvant être faiblement pertinentes. En effet, comme l'explique [Kira and Rendell, 1992], les poids des caractéristiques représentent l'impacte qu'elles ont pour la classification, indépendamment les unes des autres. Ainsi, deux caractéristiques corrélées qui interviennent dans la classification auront un poids élevé par RELIEF, et sont faiblement pertinentes par définition. En d'autres termes, RELIEF extrait des caractéristiques redondantes pour la classification.

L'annexe E.A propose un exemple simple de mise en œuvre de l'algorithme RELIEF.

**Conclusion et discussion sur RELIEF.** Cet algorithme est originellement construit pour répondre à la sélection de caractéristiques sur des données binaires. Son objectif est de détecter les caractéristiques faiblement et fortement pertinentes. Son temps d'exécution et sa précision dépendent de  $M$  qui fixe le nombre d'itérations de la boucle principale de l'algorithme.

Le paramétrage de l'algorithme exige aussi de fixer la limite  $\tau$  qui représente le seuil des poids à partir duquel une caractéristique peut être définie "pertinente". Dans l'exemple fourni en annexe E.A,  $\tau$  est calculé par l'équation (E.1). Cette définition est motivée par l'allure des poids  $W$  et la séparation claire des caractéristiques pertinentes et non pertinentes. [Kira and Rendell, 1992] propose une analyse statistique permettant de fixer la valeur de  $\tau$  selon l'erreur

de type  $I$ ,  $\alpha$ , *via* la formule  $\tau = \frac{1}{\sqrt{\alpha * M}}$ .

RELIEF a donné lieu au développement de nombreux algorithmes dérivés qui permettent de résoudre des problèmes plus complexes : classification en plusieurs classes, données manquantes, *etc.* Une comparaison des différents algorithmes issus de RELIEF est proposée dans [Kononenko, 1994].

L'algorithme RELIEF présenté dans cette section ne répond que partiellement au problème de réduction du coût du test. Il permet d'ordonner les mesures selon leur pertinence dans la classification des circuits d'entraînement, mais il est possible que l'algorithme n'élimine pas des mesures qui fournissent la même information. En d'autres termes, il n'est pas possible de l'utiliser en l'état pour déterminer les mesures à faible coût pouvant remplacer une mesure plus coûteuse.

#### II.C.2.d Algorithme de sélection de caractéristiques à recherche complète : FOCUS

À la différence de RELIEF, FOCUS est un algorithme dont l'objectif est d'extraire le sous-ensemble de cardinal le plus faible qui permette d'expliquer la classification dans un nuage de points [Almuallim and Dietterich, 1991]. Cet algorithme fonctionne pour des caractéristiques et une classification binaire. L'approche proposée par FOCUS est de tester tout les sous-ensembles de l'ensemble des caractéristiques, dans l'ordre des cardinaux croissants, jusqu'à trouver une partie qui ne présente pas de contradiction dans la classification. L'algorithme 1 résume le fonctionnement de FOCUS [John et al., 1994]. Il ne sélectionne que les caractéristiques nécessaires à l'explication de la classification observée dans l'ensemble d'entraînement.

---

**Algorithme 1** : L'algorithme FOCUS [Almuallim and Dietterich, 1991]

---

**input** :  $\mathbf{C}$  : Les caractéristiques des échantillons,  $\mathbf{y}$  : La classification des échantillons

**output** :  $A$  : Le plus petit ensemble pertinent

```

1 for  $i = 0, 1, \dots, O$  do
2   for  $A \subseteq \{c_1, \dots, c_O\}$  of size  $i$  do
3     if Test( $A, \mathbf{y}$ ) then
4       return  $A$ ;

```

---

L'instruction 3 de l'algorithme 1 est exécutée pour tout les sous-ensembles de l'ensemble des caractéristiques  $\{\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_O\}$ , jusqu'à en extraire une partie  $A$  qui satisfasse l'instruction. La complexité algorithmique est [Almuallim

and Dietterich, 1991] :

$$\Theta \left( O^{\log_2(s-O)} \right) \quad (\text{II.23})$$

avec :

- $O$  Le nombre de caractéristiques pour chaque circuit
- $s$  Le nombre de caractéristiques minimales qui expliquent la classification, inconnu avant l'exécution de l'algorithme

Cette équation indique que le temps de recherche des caractéristiques pertinentes est fortement impacté par  $s$ . Pour pallier le problème de complexité, [Almuallim and Dietterich, 1994] propose d'améliorer l'algorithme original par FOCUS-2 représenté en algorithme 4 en annexe H. Cette amélioration est suggérée par la remarque que chaque couple d'échantillons de classe différente doit avoir au moins une caractéristique qui diffère. Cette connaissance n'est pas prise en compte dans FOCUS et FOCUS-2 met en place une queue FIFO<sup>5</sup>  $Q$  pour ne pas tester toutes les parties de l'ensemble des caractéristiques. L'exemple fourni en annexe E.B illustre cet apport.

Par ailleurs, les auteurs introduisent la notation  $M_{U,V}$  qui désigne tout les sous ensembles de caractéristiques qui contiennent toutes les caractéristiques de l'ensemble  $U$ , et aucune de  $V$ .

Par la suite, [Arauzo Azofra et al., 2003] propose une nouvelle amélioration de FOCUS-2, nommée C-FOCUS, qui permet de l'appliquer à des grandeurs continues, et non plus binaires. Pour ce faire, les auteurs proposent de normaliser le nuage de points sur  $[0, 1]^O$  puis de considérer les *conflits* définis ci-après.

**Définition 6** (Conflits). *Les conflits sont l'ensemble de tous les couples d'instances  $c_x$  dont la classification  $y_x$  est différente :*

$$\mathbf{G} = \{(\mathbf{c}_i, \mathbf{c}_j) \mid y_i \neq y_j\} \quad (\text{II.24})$$

C-FOCUS établit donc l'ensemble des conflits puis cherche à les résoudre. La résolution d'un conflit est définie ci-après.

**Définition 7** (Résolution). *Un ensemble de caractéristiques  $A$  résout un conflit lorsqu'il est possible de distinguer les deux instances du conflit par l'observation de  $A$  uniquement.*

L'objectif de C-FOCUS est de trouver l'ensemble des caractéristiques  $A$  de cardinal le plus faible résolvant tous les conflits de  $\mathbf{G}$ .

---

5. First In, First Out

Pour distinguer deux valeurs d'une même caractéristique, les auteurs introduisent un seuil  $\Theta$  et normalisent chaque caractéristique entre 0 et 1. Si pour une caractéristique  $i$ , la différence est supérieure à  $\Theta$ , alors les caractéristiques pour ces instances sont considérées comme différentes. Dans le cas contraire, elles sont considérées comme identiques.

### II.C.2.e Synthèse sur les algorithmes de sélection de caractéristiques

A l'inverse des techniques présentées à la section II.C.1, la sélection de caractéristiques ne projette pas le nuage de points dans un nouvel espace mais elle sélectionne les caractéristiques les plus pertinentes dans ce nuage de points.

Deux algorithmes de l'état de l'Art ont été résumés. Ces algorithmes ont un parcours des caractéristiques déterministe et non pas aléatoire<sup>6</sup>. Ils se distinguent par leur stratégie de parcours des solutions possibles : RELIEF a un parcours séquentiel qui en diminue la complexité algorithmique, alors que FOCUS réalise un parcours complet, et a donc une complexité algorithmique élevée.

Les algorithmes RELIEF et FOCUS présentés ne sont pas applicables en l'état pour la réduction du coût des tests de circuits analogiques. De plus, pour RELIEF, les caractéristiques sélectionnées peuvent être redondantes. FOCUS évite par nature cette redondance car il assure de sélectionner le plus petit ensemble de caractéristiques qui permet d'assurer la classification des circuits.

En effet, l'objectif de la sélection de caractéristiques dans notre étude n'est pas de sélectionner le nombre minimal de caractéristiques, mais les moins coûteuses à mesurer qui permettent d'assurer la couverture de test.

## II.D Modélisation

Les techniques de réduction de dimension ou de sélection de caractéristiques permettent d'extraire l'information pertinente d'un nuage de points. Cependant, elles ne permettent pas de construire un modèle pour la classification des circuits. Or, l'objectif du test est, *in fine*, de prendre une décision quant au respect des spécifications par les circuits testés. Pour ce faire, la Vision Unifiée propose une étape de modélisation permettant d'obtenir un modèle pour la classification, à partir seulement des caractéristiques extraites à l'étape de recherche des signaux d'intérêt (voir figure II.15).

---

6. RELIEF laisse apparaître une sélection aléatoire des échantillons d'entraînement, mais pas des caractéristiques à évaluer. La recherche des caractéristiques reste donc déterministe.

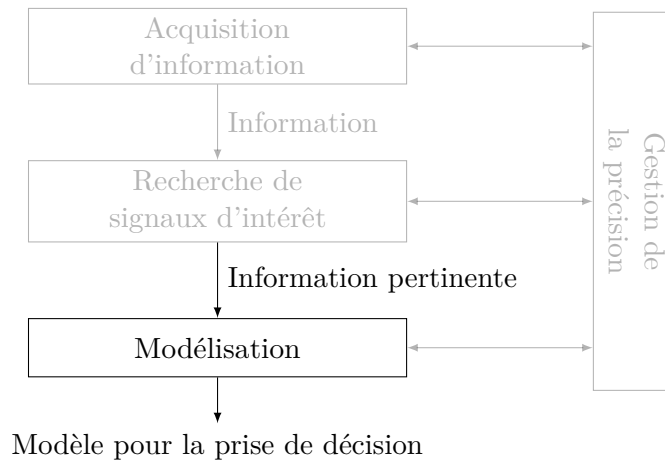


FIGURE II.15 – La modélisation dans la Vision Unifiée

La construction d'un modèle de classification permet de construire les modèles qui sont utilisés pour prendre une décision quant à l'état du circuit (fonctionnel ou défectueux). Cette étape est rendue nécessaire car ce ne sont plus nécessairement les performances des circuits qui sont manipulées, mais des observations de ce circuit, comme par exemple des mesures alternatives. Ici, le domaine des observations est réduit aux observations contenue dans l'information pertinente extraite lors de l'étape de recherche de signaux d'intérêt.

Le modèle construit à partir de l'information pertinente permettra de classer les circuits dans les états "fonctionnel" ou "défectueux", compte tenu des spécifications données dans l'espace des performances.

Pour nos travaux, des algorithmes supervisés sont retenus en raison de la nature des échantillons d'entraînement disponibles. En effet, les simulations et les mesures directes des performances permettent d'avoir une classification de l'ensemble d'entraînement. Cette information doit être utilisée pour discriminer les circuits entre ceux respectant les spécifications et ceux ne les respectant pas. À l'opposé, des méthodes à apprentissage non supervisé n'utilisent pas la classification *a priori* des échantillons d'apprentissage, ce qui n'assure pas de construire un modèle de classification selon le respect des spécifications.

Les algorithmes présentés dans cette section sont les suivants :

- les **Arbres de Décision** construisent un modèle de classification en recherchant une structure sous forme d'arbre parmi les échantillons d'entraînement. Dans le cadre de nos travaux, les arbres de classification ont l'avantage de pouvoir manipuler des données bruitées (issues de mesures par exemple) et des ensembles avec des données manquantes

(cas *Stop on Fail* des mesures) [Rokach and Maimon, 2010] ;

- les **Machines à Vecteurs de Support** établissent une frontière permettant de séparer les échantillons d’entraînement de classe différente ;
- l’approche  **$k$ -Plus Proches Voisin** dont le modèle de classification est une “simple” recherche de similitude. Cette méthode est simple à mettre en place, mais repose sur une définition *a priori* de la similitude ;
- les **Réseaux de Neurones** peuvent prendre en compte des données d’entraînement de natures diverses, mais ont la particularité de fournir un modèle de classification dont l’interprétation n’est pas simplement reliée aux données d’entraînement.

Les algorithmes discutés précédemment vont être développés dans cette section. Leurs avantages et inconvénients vis-à-vis du test de circuits y seront exposés.

L’étape de modélisation étant précédée de l’étape de recherche de signaux d’intérêt, le nuage de points vu par la modélisation n’est plus  $\mathbf{C}$ . Le nuage de points vu par la brique de modélisation est donc noté  $\mathbf{D}$  dans la suite. La dimension de cet espace est notée  $D$ .

### II.D.1 Arbres de Classification

La particularité de ce modèle de classification est d’être représentable par un arbre dont les feuilles sont les classes, ou des probabilités d’appartenance à une classe. Les nœuds de l’arbre sont des entrées d’évaluation d’une caractéristique des échantillons d’apprentissage, et les liens sont issus des nœuds représentant les valeurs possibles de l’évaluation du nœud précédent.

Les algorithmes d’induction d’arbres de classification sont une sous-classe d’une appellation plus globale : *Classification And Regression Tree (CART)* originellement introduite par [Breiman et al., 1984].

L’objectif de ces algorithmes est de construire l’arbre de classification à partir d’un ensemble d’échantillons d’entraînement. L’idée est de détecter la caractéristique de discrimination la meilleure, puis de séparer les échantillons d’entraînement selon un critère choisi sur cette caractéristique, et d’itérer jusqu’à ce que les séparations ne contiennent que des échantillons de même classe, cf. l’algorithme 2.

Comme le montre le pseudo algorithme 2, la méthode traditionnelle de construction de l’arbre de classification est une méthode *top-down* récursive.

Par ailleurs, les algorithmes d’induction d’arbres de classification se distinguent en partie par leur stratégie de sélection de la caractéristique de séparation. D’après [Donald, 1994], les deux stratégies majeures sont la mesure

---

**Algorithme 2 :** Pseudo algorithme d'induction d'arbres de classification

---

```

input   :  $\mathbf{D}, \mathbf{y}$ 
output :  $T$  : Le model d'arbre de classification

1 Function split( $\mathbf{D}$  : Set of samples,  $\mathbf{y}$  : classes of every  $\mathbf{d}$ ) is
   | /* Extract the best splitting feature  $F$  with limit  $f$  */
   | 2 ( $F, f$ )  $\leftarrow$  bestSplitCriterion ( $\mathbf{d}, \mathbf{y}$ ) ;
   | 3  $T$ .addNode ( $F$ );
   |
   | /* Split  $\mathbf{D}$  into two groups */
   | 4 ( $\mathbf{D}^+, \mathbf{y}^+$ )  $\leftarrow$  { $\mathbf{D} | \mathbf{D}(F) \geq f$ };
   | 5 ( $\mathbf{D}^-, \mathbf{y}^-$ )  $\leftarrow$  { $\mathbf{D} | \mathbf{D}(F) < f$ };
   |
   | 6 if unique( $\mathbf{y}^+$ ) > 1 then
   |   |  $\lfloor$  split ( $\mathbf{D}^+, \mathbf{y}^+$ )
   | 7
   | 8 else
   |   |  $\lfloor$   $T$ .addLeaf(unique( $\mathbf{y}^+$ ))
   | 9
  10 if unique( $\mathbf{y}^-$ ) > 1 then
  11   |  $\lfloor$  split ( $\mathbf{D}^-, \mathbf{y}^-$ )
  12 else
  13   |  $\lfloor$   $T$ .addLeaf(unique( $\mathbf{y}^-$ ))
  14 return  $T$ ;
  15  $T =$  split ( $\mathbf{D}, \mathbf{y}$ );

```

---

d'entropie et l'évaluation de la fonction de Gini.

La mesure fondée sur l'entropie cherche à conserver l'équilibre entre l'arbre de classification et la classification de l'ensemble des échantillons d'entraînement. La mesure de Gini, elle, isole les classes les plus représentées dans l'ensemble des échantillons d'entraînement. Ces stratégies ne sont pas développées ici, mais le lecteur peut se reporter à l'annexe G pour plus de détails sur l'implémentation de la mesure d'entropie et de la fonction de Gini.

Dans le cadre du test de circuits analogiques, les arbres de classification ont l'avantage d'être simples à interpréter. Ainsi, ils fournissent un modèle en *boite blanche*, dont le résultat est obtenu par l'analyse directe des grandeurs observées, comparées à des seuils sur ces observations. De plus, cette méthode ne présuppose pas de distribution des échantillons d'origine. Les arbres de décision sont mis en œuvre dans [Stratigopoulos and Sunter, 2014] pour construire une frontière de simulation dans le cadre de métriques de test.

Les algorithmes d'induction d'arbres de classification ont l'inconvénient

de générer des arbres complexes et redondant lorsque les relations entre la classification observée dans les échantillons d'entraînement et les mesures alternatives sont complexes [Rokach and Maimon, 2010]. Par ailleurs, ces algorithmes sont sensibles au bruit de mesure et aux caractéristiques non pertinentes. En d'autres termes, dans le cadre de la Vision Unifiée, un algorithme de modélisation par arbre de décision sera sensible aux erreurs potentielles de la sélection des signaux d'intérêt. Dans le cadre de nos travaux, les arbres de classifications seront mis en œuvre en particulier lorsque les données sont issues de simulations, en raison de la sensibilité au bruit de cette méthode.

### II.D.2 Machine à Vecteurs de Support

La Machine à Vecteurs de Support (SVM) est un algorithme de classification qui recherche la frontière (si elle existe) qui sépare les échantillons de classes différentes avec la distance la plus grande entre la frontière calculée et les points qui lui sont le plus proche [Boser et al., 1992]. Cette méthode utilise la notion de séparation, définie par l'hyperplan :

$$H_b = \left\{ \mathbf{x} \mid D(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + b = 0 \right\}, \forall \mathbf{x} \in \mathbb{R}^D \quad (\text{II.25})$$

avec :

- $H_b$  Un hyperplan de l'espace d'observation définissant la limite de séparation des échantillons de classes différentes
- $D(.)$  Une fonction distance qui définit la distance entre un point de l'espace d'observation et l'hyperplan de séparation
- $\mathbf{w}$  Un vecteur orthogonal à  $H_b$ , à calculer
- $b$  Un scalaire à calculer
- $D$  La dimension de l'espace de l'information pertinente  $\mathbf{D}$

La méthode SVM calcule  $\mathbf{w}$  et  $b$  qui définissent l'hyperplan  $H_b$  de séparation des échantillons de classe différente. Pour cela, elle introduit la notion de vecteurs de support qui sont quelques échantillons d'observation permettant la définition de la marge la plus grande. La complexité de l'algorithme est évaluée par [Bousquet, 2001] comme étant  $O(DN^3)$ . La figure II.16 propose un exemple simple d'application de la méthode SVM qui fait apparaître les vecteurs de support.

L'exposé de cette méthode est reporté en annexe F.



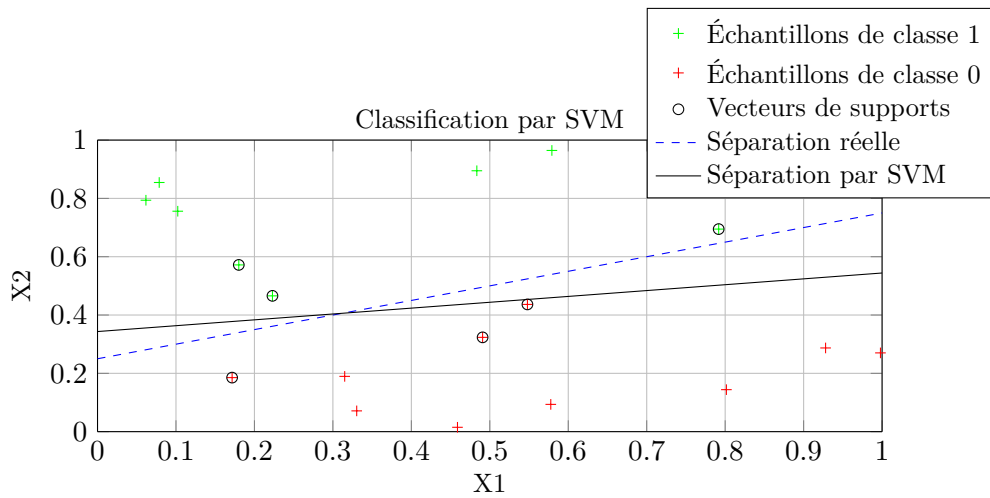


FIGURE II.16 – Exemple de classification par SVM

### Conclusion sur la méthode SVM.

La méthode SVM est originellement construite sur un séparateur linéaire mais elle s'étend aux frontières non linéaires par l'introduction de fonctions à noyaux.

La classification par méthode SVM conduit à un simple test de signe au moment de la classification, ce qui en fait un candidat idéal pour le test de circuits analogiques en production. Cette simplicité au moment de la classification est contrebalancée par la détermination de la frontière de séparation. Cette frontière est établie à partir des échantillons de l'ensemble d'entraînement et demande l'utilisation de méthodes à noyaux lorsque la frontière ne peut pas être linéaire. La méthode SVM a donc les inconvénients des méthodes à noyaux, en particulier la nécessité de connaître *a priori* le noyau adapté au nuage de points.

Dans l'état de l'Art du test analogique, la méthode SVM est utilisé pour des algorithmes de sélection de mesures alternatives [Sun et al., 2014], ou encore pour du diagnostic [Stratigopoulos and Mir, 2010]. Dans le cadre de nos travaux, cette méthode permet de déterminer une frontière, si elle existe, exprimée dans l'espace des caractéristiques sélectionnées, entre les circuits défectueux et les circuits fonctionnels. Son utilisation au moment du test est donc très simple puisqu'il suffit de calculer l'écart entre le circuit mesuré et l'hyperplan  $H_b$ .

### II.D.3 Approche k-Plus Proches Voisins

La méthode de classification des  $k$  plus proches voisins (kNN *k-Nearest Neighbors*) est une méthode de classification supervisée dont l'objectif est de classer un échantillon selon ses  $k$  voisins les plus proches dans l'ensemble d'entraînement [Peterson, 2009]. Elle ne requiert pas de noyau comme la méthode SVM, mais pose deux questions : *quel  $k$  choisir ? Et quelle distance choisir ?*

De plus, cette méthode utilisant la notion de distance, elle est soumise au problème d'échelle. Il est donc recommandé de normaliser les données pour utiliser l'approche kNN.

#### II.D.3.a Choix de $k$

Le choix du nombre de plus proches voisins  $k$  est critique car il peut, à lui seul, modifier la classification d'échantillons. L'exemple de la figure II.17 montre comment le choix de  $k$  influe sur la classification.

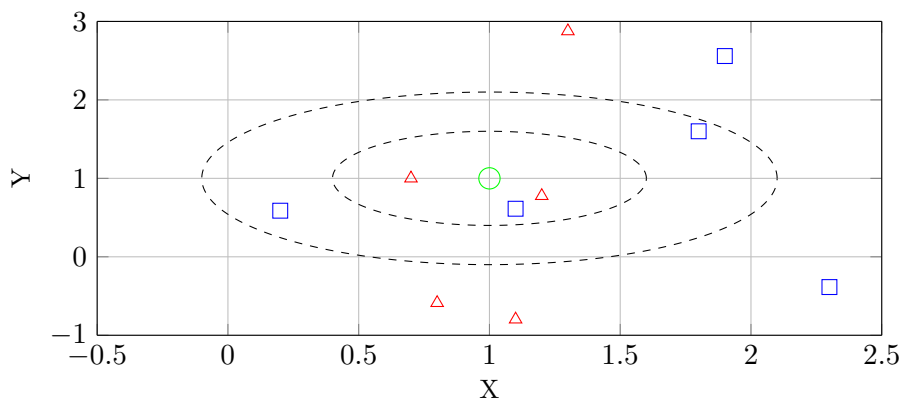


FIGURE II.17 – Exemple illustrant l'impact du choix de  $k$ . L'échantillon de test est le cercle vert.

Dans ce cas simple, si  $k = 3$  est retenu, alors l'échantillon de test vert sera classé préférentiellement en triangle rouge. Dans le cas  $k = 5$ , il sera classé préférentiellement en carré bleu.

Si à l'extrême  $k = 1$ , la classification fournie par l'algorithme est constante par morceau. L'espace des échantillons est alors découpé en morceaux dont les frontières sont perpendiculaires à la direction définie par le couple d'échantillons de classes différentes les plus proches. Dans le cas  $k = T$  (nombre d'échantillons de l'ensemble d'entraînement), la classification est constante, et correspond à la classe la plus représentée dans l'ensemble d'entraînement.  $k$  a donc un rôle de lissage de la classification.

Le choix de  $k$  est donc critique. Pour opérer efficacement, [Hall et al., 2008] justifie l'utilisation d'une méthode heuristique pour l'évaluation de  $k_{opt}$ , le nombre de plus proches voisins qui minimise l'erreur de classification. Les auteurs proposent de construire un *bootstrap*<sup>7</sup> de l'ensemble d'entraînement et de réaliser plusieurs tentatives de classification par kNN avec différentes valeurs de  $k$  pour sélectionner la valeur de  $\widehat{k}_{opt}$  donnant la meilleure classification.

Ensuite,  $k_{opt}$  est calculé à partir de  $\widehat{k}_{opt}$  par la formule  $k_{opt} = r^{\frac{-4}{D+4}} \widehat{k}_{opt}$  où  $D$  est la dimension des échantillons et  $r$  le ratio utilisé pour la construction du *bootstrap*.

Cette méthode est justifiée par [Hall et al., 2008] lorsque les classes peuvent être approximées par des distributions de Poisson, ou des distributions binomiales.

### II.D.3.b Réduction de l'ensemble d'entraînement

Un problème majeur associé à l'utilisation de la classification kNN est la gestion de la quantité de données dans l'ensemble d'entraînement. En effet, plus il y a d'échantillons dans l'ensemble d'entraînement, plus la recherche des plus proches voisins est longue, et plus l'espace mémoire occupé est grand.

Par ailleurs, les effets néfastes de la taille de l'ensemble d'entraînement ne sont pas seulement induits par le nombre d'échantillons, mais aussi par la dimension des échantillons. Comme rappelé en section I.A.2, le nombre d'observations peut être élevé. Ainsi, [Zhou et al., 2015] propose une réduction de dimension *a priori*. Les auteurs étudient l'utilisation de l'Analyse en Composantes Principales et proposent une technique de projection aléatoire. Selon les auteurs, cette projection permet de conserver localement les distances entre échantillons et est utile dans les cas de faible nombre d'échantillons d'entraînement en dimension élevée, ce qui peut être le cas dans nos travaux.

### Avantages et inconvénients de l'approche des $k$ plus proches voisins.

Le calcul des  $k$  plus proches voisins d'un point nécessite d'évaluer la distance entre tous les points de l'ensemble d'entraînement à ce point. La complexité du tri de l'ensemble d'entraînement selon la distance dépend de l'algorithme choisi. L'algorithme de recherche parmi tous les échantillons d'entraînement est à complexité  $\Theta(D \cdot N)$  où  $D$  est la dimension des échantillons, et  $N$  le nombre d'échantillons.

La classification par kNN fournit la classe d'un échantillon de test à partir des  $k$  plus proches voisins de ce point dans l'ensemble d'entraînement. Ce comportement permet de calculer la probabilité de correctement classifier

---

7. C'est un sous-ensemble d'échantillons qui est représentatif de l'ensemble d'échantillons initial, i.e. le ratio d'échantillons par classe est respecté.

l'échantillon de test en utilisant les distributions empiriques des échantillons d'entraînement dans le voisinage de l'échantillon de test. Ainsi, dans le cas où le voisinage de l'échantillon de test ne contient pas un nombre significatif d'échantillons d'une classe, il est possible de considérer cet échantillon de test comme incertain, plutôt que de le classer.

Dans le cadre du test analogique, la classification par kNN est simple d'implémentation mais repose sur une notion de distance. Au vu de son fonctionnement, il est nécessaire d'acquérir suffisamment d'échantillons d'entraînement pour couvrir correctement l'espace des caractéristiques pertinentes.

Par ailleurs, l'algorithme est, comme la méthode par arbre de décision, sensible aux caractéristiques non pertinentes ou bruitées. Pour ces raisons, cette méthode est peu représentée dans l'état de l'Art du test de circuits analogiques. [Zhou et al., 2015] utilise la modélisation kNN pour détecter les fautes lors de la production de circuits. Les auteurs expliquent le besoin de réduire la dimension de l'espace d'observation pour utiliser kNN et proposent une projection aléatoire qui préserve les distances entre espace d'observation et espace de projection.

#### II.D.4 Méthode MARS

MARS n'est pas à proprement parler un modèle de classification, mais c'est un modèle de régression. Dans le cadre de nos travaux, les modèles de régression peuvent être utilisés pour établir une formule reliant les valeurs des mesures alternatives retenue dans l'information pertinente aux performances. Puis, les performances régressées sont comparées aux spécifications connues pour prendre une décision sur le bon fonctionnement du circuit.

Le modèle de la méthode MARS est construit comme une succession de fonctions linéaires par morceaux définies de la façon suivante [Hastie et al., 2001] :

$$M_t : \mathbb{R} \longrightarrow \mathbb{R} \\ x \longmapsto \begin{cases} x - t, & \text{si } x > t \\ 0 & \text{sinon} \end{cases} \quad (\text{II.26})$$

ou

$$M_{-t} : \mathbb{R} \longrightarrow \mathbb{R} \\ x \longmapsto \begin{cases} t - x, & \text{si } x < t \\ 0 & \text{sinon} \end{cases} \quad (\text{II.27})$$

où  $x$  représente la valeur d'une mesure alternative.

Les fonctions  $M_t$  et  $M_{-t}$  sont appelées des *splines*, avec un nœud en  $t$ . L'objectif de MARS est de construire un modèle de régression de la forme

$$f(\mathbf{x}) = \beta_0 + \sum_{m=1}^M \beta_m h_m(\mathbf{x}) \quad (\text{II.28})$$

avec :

$\beta_i, i \in \{0, \dots, M\}$	Les coefficients à calculer de façon à réduire l'erreur quadratique
$h_m$	Un produit de splines appartenant à la collection $\mathcal{C} = \{M_t, M_{-t}   t \in \mathbf{C}_{i,j}, j \in \text{SelectedFeatures}\}$
$M$	Le nombre de produits à utiliser dans la régression
$\mathbf{x} \in \mathbb{R}^D$	Le vecteur des informations pertinentes pour un circuit

L'induction du modèle n'est pas développée ici, les détails sont fournis dans [Friedman, 1991]. L'induction se déroule généralement en deux étapes : une première étape *forward* qui recherche de manière itérative les  $\beta_m$  et  $t$  qui permettent de minimiser l'erreur quadratique de régression, puis une étape *backward* d'élimination de certains produits. En effet, le modèle issu de l'étape *forward* contient de nombreux produits ce qui peut conduire au risque de sur-apprentissage [Hastie et al., 2001].

La simplicité du modèle utilisé par MARS et ses performances en font un candidat sérieux dans le cadre du test de circuit. Il est en particulier utilisé pour régresser les performances de circuits à partir de mesures internes [Bhattacharya and Chatterjee, 2004], [Srinivasan et al., 2006], [Halder et al., 2005], [Haider et al., 2003] et [Nassery et al., 2012]. MARS ne requiert pas l'utilisation d'un noyau, mais requiert une connaissance des relations pouvant relier les observations (ie. mesures ou simulations désignées pertinentes) aux performances, afin de paramétrer l'algorithme, en particulier le nombre de splines et le nombre de produits  $M$  dans la régression.

Dans le cadre de nos travaux, la méthode MARS sera utilisée pour régresser les valeurs des performances à partir des observations pertinentes. Ensuite, les performances régressées sont comparées aux spécifications, ce qui permet de classer les circuits entre fonctionnels et défectueux à partir des observations pertinentes seulement.

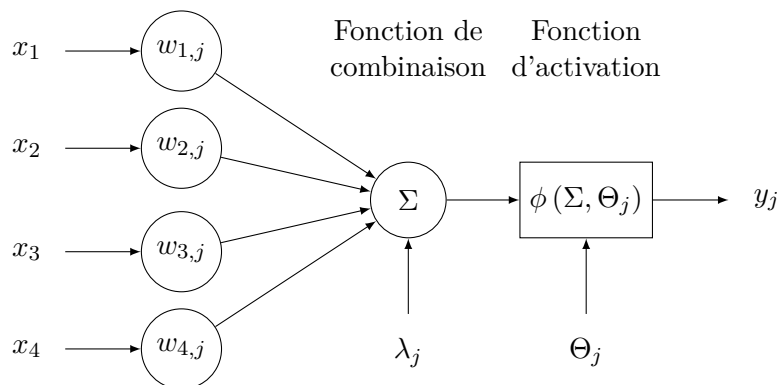


FIGURE II.18 – Représentation d'un neurone [Guyon, 1991, Vapnik, 1995b]

### II.D.5 Méthodes à Réseaux de Neurones

Les Réseaux de Neurones Artificiels (RNA) constituent une partie non négligeable de l'état de l'Art du traitement de données, en particulier de la classification et de la régression. Les réseaux de neurones sont des systèmes mimant l'apprentissage (au sens biologique). Leur apparition provient d'une analyse biologique [Lettvin et al., 1959] qui montre que quatre opérations sont réalisées biologiquement par des neurones pour représenter une image. Cette observation a permis de développer des modèles de neurones dont l'association peut réaliser des opérations mathématiques complexes [Hopfield, 1982]. L'apprentissage par RNA est une méthode d'induction permettant d'obtenir un réseau de neurones à partir d'une base de données servant d'entraînement. La figure II.18 représente un modèle de neurone. Le fonctionnement d'un neurone artificiel consiste en la sommation des valeurs d'entrées  $x_i$  pondérées par des poids  $w_{i,j}$ . Puis une fonction non linéaire  $\phi$  est appliquée à cette somme ainsi qu'à une valeur d'offset  $\lambda_j$ . Enfin, le résultat de cette fonction non linéaire est comparé à un seuil  $\Theta_j$  qui détermine la sortie du neurone : activé ou désactivé.

La construction d'un RNA consiste en l'assemblage de neurones artificiels entre eux, où les sorties de neurones sont les entrées des autres. Le choix de la structure d'assemblage du réseau et des fonctions non linéaires  $\phi$  permet au réseau de résoudre des classes de problèmes différentes. En particulier, [Arinton and Korbicz, 2004, Korbicz and Janczak, 1996] proposent des méthodes de recherche automatique de structure de réseau de neurones. La figure II.19 présente un exemple de RNA à une couche d'entrée (en vert sur le schéma), une couche cachée (rouge) et une couche de sortie (bleue). Le nombre de couches cachées est aussi déterminant pour la sélection des problèmes solvable par un RNA : [Cybenko, 1989] a démontré qu'un RNA à une couche cachée peut approximer toute fonction linéaire à un nombre fini de variables. L'apprentissage d'un RNA consiste à modifier les valeurs des poids  $w_{i,j}$  pour satisfaire un ensemble d'entraînement.

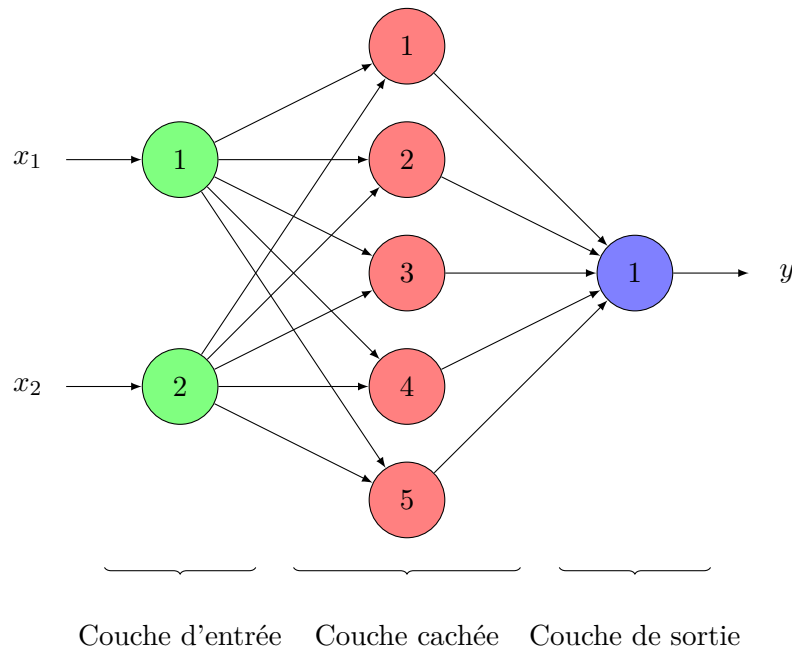


FIGURE II.19 – Exemple de RNA

Les RNA sont utilisés dans le domaine du test essentiellement pour la prédiction du respect des spécifications d'un circuit [Stratigopoulos and Makris, 2008, Starzyk and El-Gamal, 1990].

Les réseaux de neurones souffrent de plusieurs inconvénients. Tout d'abord, leur comportement *boîte noire* empêche d'associer une description physique au modèle de classification construit. En outre, la structure du réseau et la fonction d'activation sont des choix critiques au regard de la fonction à régresser qui impactent le modèle obtenu et la qualité de la classification. Dans ces travaux, les réseaux de neurones ne seront pas utilisés.

### II.D.6 Conclusion sur les algorithmes de classification

Plusieurs algorithmes de classification de l'état de l'Art ont été présentés dans cette section. Seul des algorithmes supervisés sont retenus en raison de la nature des informations disponibles dans le cadre de l'établissement de stratégies de test. En effet, les mesures et simulations peuvent toujours s'accompagner d'une classification en comparant les performances relevées aux spécifications.

Une comparaison des performances des inducteurs présentés ci-avant est

donnée sur la figure II.20 pour un cas particulier et trois performances. Cette comparaison permet de quantifier les résultats obtenus. En effet, ces grandeurs sont liées aux considérations économiques du test, et permettent de représenter les performances des classifieurs.

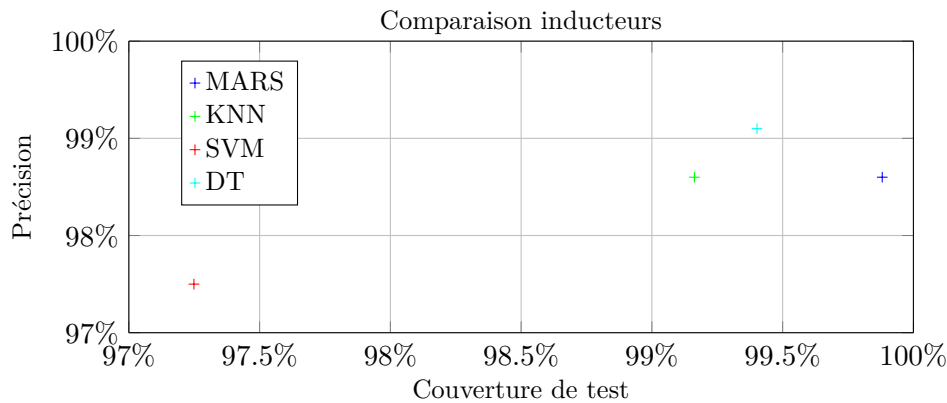


FIGURE II.20 – Comparaison des erreurs de classifieurs

Les distinctions entre les différentes méthodes de classification peuvent être résumées selon plusieurs critères :

- la dimension originale ;
- la complexité de la structure des données ;
- La complexité algorithmique.

Dans le cadre du test de circuits, la méthode MARS régresse les valeurs de chacune des performances, qui sont alors comparées aux spécifications. MARS est performant lorsque les relations entre les performances et les observations ne sont pas fortement non linéaires. De la même manière, SVM recherche une séparation entre deux classes sous la forme d'un hyperplan. Cependant, l'utilisation de méthodes à noyaux permet d'évaluer une séparation qui n'est pas un hyperplan. La performance de cette séparation repose donc en partie sur le choix du noyau.

La méthode kNN utilise la notion de distance pour estimer la classe d'un échantillon. La méthode mise en place ici réalise une transformation de l'espace d'observation avant la construction d'un modèle. Cette transformation ne conserve pas nécessairement la distance, ce qui peut engendrer des erreurs de classification par la méthode kNN. Dans nos travaux, nous supposons que l'espace dans lequel sont projetés les échantillons avant la classification est celui dans lequel la classification fait sens.



## II.E Gestion de la Précision

Dans le cadre de la Vision Unifiée, différentes méthodes de traitement des données sont utilisées pour extraire les signaux d'intérêt d'une base de données et construire des modèles de classification. Le modèle résultant de ces étapes doit permettre de classer les circuits uniquement à partir des mesures des signaux désignés comme d'intérêt. Cependant, les erreurs peuvent entraîner une mauvaise classification des circuits. Ainsi apparaissent des faux positifs, circuits défectueux qui sont classés fonctionnels, et des faux négatifs, circuits fonctionnels classés défectueux. Ces erreurs sont les grandeurs à minimiser lors de l'établissement de la stratégie de test.

Cette section propose d'analyser les erreurs commises au moment de la classification. On présente aussi des outils théoriques permettant d'anticiper et d'interpréter ces erreurs. Ainsi, la section II.E.1 propose une analyse théorique, appelée PAC-learning, permettant d'établir la précision de la classification en fonction, entre autre, des échantillons disponibles. Puis la section II.E.2 explique l'analyse de la précision par matrices de confusion. Enfin, la section II.E.3 conclut cette partie sur la gestion de la précision.

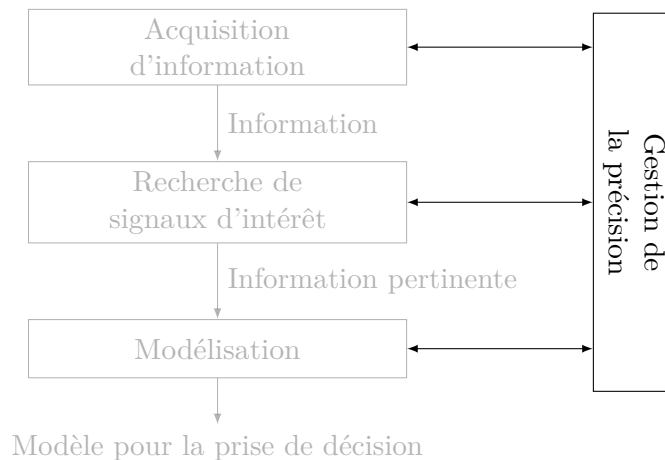


FIGURE II.21 – Gestion de la précision dans la Vision Unifiée

### II.E.1 Précision *a priori* des algorithmes de classification

Le principe de construction d'un modèle est rappelé en figure II.22. Les données sont découpées en données d'entraînement destinées à la construction d'un modèle par un *inducteur*, et en données de validation sur lesquelles le modèle construit est appliqué pour sa validation. Cette étape de validation compare alors les résultats issus du modèle à ceux effectivement observés et en déduit l'erreur introduite par l'utilisation de ce modèle. La description

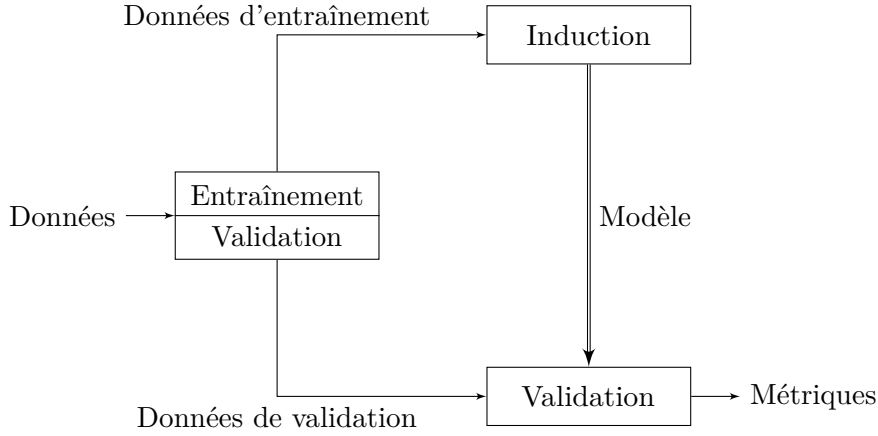


FIGURE II.22 – Construction et validation du modèle

mathématique de l'opération d'induction, et d'apprentissage en général, est donnée par la théorie PAC-learning (*Probably Approximately Correct*) [Valiant, 1984]. Cette dernière décrit un cadre de travail pour l'analyse des algorithmes d'apprentissages, et notamment le calcul *a priori* des erreurs qu'ils commettent.

Dans cette section, on décrit ces erreurs. Pour ce faire, PAC introduit un *concept* comme la relation *réelle* à apprendre et le définit comme suit.

**Définition 8** (Concept). *Soient le domaine  $\mathbf{D} = \{0, 1\}^O$ , un échantillon  $d_i = (d_{i,1}, \dots, d_{i,O}) \in \mathbf{D}$ , un concept est défini par [Dhagat and Hellerstein, 1994] :*

$$\begin{aligned} Do : \mathbf{D} &\longrightarrow \mathbb{E} = \{0, 1\} \\ d_i &\longmapsto P_i = Do(d_i) \end{aligned} \quad (\text{II.29})$$

où  $P_i$  est la classe de l'échantillon  $D_i$ .

On appelle *concept* une procédure qui associe un label à chaque échantillon, ce dernier étant alors représenté par le couple  $(\mathbf{d}_i, P_i)$ . L'objectif de l'algorithme d'induction est de trouver une fonction  $h$ , appelée hypothèse, "aussi proche que possible" du concept  $Do$ , à partir de l'observation des  $(\mathbf{d}_i, P_i)$ ,  $i \in \{1, \dots, T\}$ , où  $T$  représente le nombre d'échantillons d'entraînement. L'hypothèse est appelée  $h \in H$ , où  $H$  est une classe de concept.

Une hypothèse supplémentaire est que les échantillons utilisés pour l'entraînement sont générés selon une distribution  $D$  inconnue. L'erreur d'entraînement  $\epsilon_T$  sur une hypothèse  $h$  est définie comme la probabilité que, pour un échantillon  $\mathbf{d}_i$  choisi selon  $D$  (noté  $\mathbf{d}_i \sim D$ ), on ait  $Do(\mathbf{d}_i) \neq h(\mathbf{d}_i)$ .

$$\epsilon_T(h) = \Pr_{\mathbf{d}_i \sim D} \{Do(\mathbf{d}_i) \neq h(\mathbf{d}_i)\} \quad (\text{II.30})$$

Si l'algorithme d'induction génère l'hypothèse  $h$  pour tout concept  $Do \in H$ , pour toute distribution  $D$ , pour tout  $0 < \epsilon < \frac{1}{2}$  et pour tout  $0 < \delta < \frac{1}{2}$ , avec une probabilité  $1 - \delta$  que l'erreur soit plus faible que  $\epsilon$ , alors  $H$  est dit *PAC apprenable*.

La théorie PAC établit, à partir de l'équation (II.30), une relation entre l'erreur d'entraînement  $\epsilon$ , la confiance dans l'erreur  $\delta$  et le nombre d'échantillon d'entraînement  $T$ .

Si la génération est consistante, c'est-à-dire que l'algorithme recherche des hypothèses qui sont valides pour tout les échantillons de l'ensemble d'entraînement, alors la probabilité qu'une de ces hypothèses ait une erreur inférieure à  $\epsilon$ , et soit consistante avec un échantillon d'entraînement est au plus  $(1 - \epsilon)$  [Pichuka et al., 2007]. Par la suite, la probabilité que cette hypothèse ait une erreur supérieure à  $\epsilon$  et soit consistante avec  $m$  échantillons d'entraînement est alors  $(1 - \epsilon)^m$ . Enfin, la probabilité que  $k$  hypothèses aient une erreur supérieure à  $\epsilon$  et soient consistantes avec  $m$  échantillons est  $k * (1 - \epsilon)^m$ . En notant  $|H|$  le nombre maximal d'hypothèses possibles :

$$\Pr \{ \epsilon_T(h) > \epsilon \} \leq |H| * (1 - \epsilon)^m \leq \delta \quad (\text{II.31})$$

$$\leq |H| * e^{-m\epsilon} \leq \delta \quad (\text{II.32})$$

d'où il vient [Pichuka et al., 2007] :

$$m \geq \frac{1}{\epsilon} \left( \ln(|H|) + \ln\left(\frac{1}{\delta}\right) \right) \quad (\text{II.33})$$

L'équation (II.33) fait apparaître la limite minimale d'échantillons requis pour un inducteur, en fonction de l'erreur  $\epsilon$  et de la confiance dans l'erreur  $\delta$  souhaitée. Il y apparaît aussi une dépendance en  $|H|$  qui est le nombre d'hypothèses que l'inducteur peut parcourir. Ce résultat est établi pour un inducteur consistant, c'est-à-dire qui recherche des solutions qui satisfassent tous les échantillons de l'ensemble d'entraînement. Il est possible de considérer un inducteur agnostique, qui autorise une solution ne satisfaisant pas tous les échantillons d'entraînement. Dans ce cas, l'équation (II.33) devient [Pichuka et al., 2007] :

$$m \geq \frac{1}{2\epsilon^2} \left( \ln(|H|) + \ln\left(\frac{1}{\delta}\right) \right) \quad (\text{II.34})$$

Dans les équations (II.33) et (II.34), le terme en  $|H|$  doit être évalué. [Guestrin, 2007] propose d'évaluer cette grandeur dans le cadre des arbres de classification binaires, selon deux approches : à savoir, en fonction de la profondeur de l'arbre, ou du nombre de feuilles de l'arbre. Dans les deux cas, la technique de résolution est récursive, et donne les résultats suivants :

$$|H|_{k\text{-depth}} = \ln(2) \left( (2^k - 1) (1 + \log_2(n)) + 1 \right) \quad (\text{II.35})$$

$$|H|_{k\text{-leaves}} = n^{k-1} (k + 1)^{2k-1} \quad (\text{II.36})$$

avec :

- $k$  La profondeur de l'arbre pour l'équation (II.35) ou nombre de feuilles pour l'équation (II.36)
- $n$  Le nombre de dimensions dans l'espace des échantillons

L'équation (II.33) est utilisable dans les cas où  $|H|$  est fini. Dans les cas où  $|H|$  est infini, ce qui est courant avec des modèles de classification à base de séparation par hyperplan (e.g. SVM), la taille de l'espace des hypothèses est substituée par la *Dimension VC* (pour Vapnik Chervonenkis) qui est la mesure de la complexité d'un algorithme de classification. Elle est définie pour un ensemble de fonctions comme le cardinal du plus grand ensemble de points classifiables sans erreur d'entraînement par ces fonctions, quelque soit la répartition et la classification des points de cet ensemble [Vapnik, 1995a]. Ainsi, en notant  $h$  la dimension *VC* d'un modèle de classification, l'équation (II.33) devient [Blumer et al., 1989] :

$$m \geq \frac{1}{\epsilon} \left( 4 \log_2 \left( \frac{2}{\delta} \right) + 8h \log_2 \left( \frac{13}{\epsilon} \right) \right) \quad (\text{II.37})$$

La théorie de la dimension *VC* n'est pas développée ici, les détails sont disponibles dans [Vapnik, 1995c]. La dimension *VC* est développée par [Srihari, 2004] dans le cas des réseaux de neurones. On obtient alors :

$$m \geq \frac{1}{\epsilon} \left( 4 \log_2 \left( \frac{2}{\delta} \right) + 16 (r + 1) s \log (es) \log \left( \frac{13}{\epsilon} \right) \right) \quad (\text{II.38})$$

avec :

- $r$  Le nombre d'entrée par perceptrons
- $s$  Le nombre de perceptrons dans le réseau
- $e$  Le nombre d'euler

La théorie du PAC-learning et de la dimension *VC* sont utilisées pour estimer l'erreur *a priori* d'un modèle de classification, en fonction du nombre d'échantillons d'entraînement et de la confiance souhaitée pour l'erreur. Les relations établies pour caractériser la précision des modèles fait toujours apparaître le nombre d'échantillons d'apprentissage, l'erreur et la confiance, ainsi que la dimension *VC* ou le nombre d'hypothèses qui caractérise l'inducteur utilisé. Ainsi, ces relations construisent un compromis entre la précision et le nombre d'échantillons nécessaires. Dans le cadre du test de circuits, ces relations permettent de paramétrer, *a priori*, les algorithmes d'induction.

Pour nos travaux, le PAC-learning est utilisé pour savoir, *a priori*, s'il est possible d'atteindre une classification avec l'erreur demandée, compte tenu du nombre d'échantillons disponibles et du modèle utilisé.

L'approche par matrice de confusion permet d'obtenir d'autres informations sur l'erreur commise, par l'analyse *a posteriori* des résultats de classification. Cette approche permet d'obtenir des indicateurs d'erreur proches des indicateurs économiques présentés en section I.B.3.

### II.E.2 Matrice de Confusion et précision *a posteriori*

Une Matrice de Confusion est un tableau formaté de façon à visualiser la performance d'un inducteur [Stehman, 1997]. Cette représentation permet de définir une métrique de test qui est, en règle générale, représentative de la probabilité d'obtenir une erreur de classification d'un certain type.

Ici, deux classes sont considérées pour les circuits : la classe *Fonctionnel* et la classe *Défectueux*. Par ailleurs, un classifieur évalue l'appartenance de  $N$  échantillons à une de ces classes. Les résultats de cette classification sont alors consignés dans la matrice de confusion donnée sur la tableau II.2. Les grandeurs VP (Vrai Positif) et VN (Vrai Négatif) représentent respectivement le nombre d'échantillons classés positifs (i.e. fonctionnels) et qui le sont effectivement, et le nombre d'échantillons classés négatifs (i.e. défectueux) et qui le soient effectivement. FN (Faux Négatif) représente le nombre d'échantillons classés négatifs mais qui sont en réalité fonctionnels. FP (Faux Positif) représente, à l'inverse, le nombre d'échantillons classés positifs mais qui en réalité sont défectueux. Les quantités FN et FP sont les erreurs de classification, aussi appelés respectivement erreur de type *II* et erreur de type *I*.

		Classification	
		Fonctionnel	Défectueux
Réel	Fonctionnel	VP	FN
	Défectueux	FP	VN

TABLE II.2 – Exemple de matrice de confusion à deux classes

Ce dénombrement peut être réalisé pendant la phase de validation de l'algorithme de classification. La matrice de confusion servira alors de base pour définir les métriques de test, et comparer les performances de différents algorithmes. La suite de cette section introduit des métriques de classifieurs, calculées à partir de la matrice de confusion, qui servent de point de comparaison des algorithmes de classification.

Parmi ceux-ci, les couvertures de test et les pertes de rendement sont très utilisés car très proches des métriques de test présentées en section I.B.3. Cependant, ces métriques sont atteinte par le biais du déséquilibre des classes rencontré dans le test de circuit. Ainsi, les métriques  $\kappa$  et  $F_\beta$ -Mesure seront introduites.

### II.E.2.a Couverture et perte de rendement

La couverture de rendement  $Y_C$  représente le taux de circuits fonctionnels qui passent le test. En utilisant les notations introduites dans le tableau II.2, cette métrique peut être évaluée par :

$$Y_C = \frac{VP}{VP + FN} \quad (\text{II.39})$$

La métrique complémentaire est la perte de rendement  $Y_L$  définie par :

$$Y_L = 1 - Y_C = \frac{FN}{VP + FN} \quad (\text{II.40})$$

La perte de rendement représente le nombre de circuits fonctionnels rejetés par le test. Économiquement, elle représente la quantité de circuits fonctionnels qui ne sont pas vendus. On cherchera donc naturellement à la minimiser.

### II.E.2.b Taux de défauts et valeur prédictive positive

Le taux de défauts est la quantité d'échantillons "défectueux" qui sont classés "fonctionnels". Elle est donnée par :

$$T_E = \frac{FP}{VP + FP} \quad (\text{II.41})$$

Le taux de défaut est important dans le test de circuits car il est liée au taux de retour client, c'est-à-dire la probabilité qu'a un client d'acheter un circuit défectueux alors qu'il est classé comme fonctionnel. Cet aspect économique est souvent à équilibrer avec la perte de rendement. En effet, selon le produit visé, il peut être préférable d'avoir une stratégie de test qui ait un taux de défaut faible, quitte à avoir une perte de rendement plus élevée, ou l'inverse.

La valeur prédictive positive  $S_S$  est la grandeur complémentaire :

$$S_S = 1 - T_E = \frac{VP}{VP + FP} \quad (\text{II.42})$$

Dans nos travaux, le taux de défaut est une grandeur qui sera observé lors des résultats en raison de son aspect économique.

### II.E.2.c Précision

La précision représente la proportion d'échantillons correctement classés parmi tous les échantillons :

$$A_{cc} = \frac{VP + VN}{N} \quad (\text{II.43})$$

### II.E.2.d Limite des métriques de test

Les métriques de test présentées ci-avant sont influencées par la proportion initiale d'échantillons fonctionnels ou défectueux. En effet, l'exemple suivant montre qu'il est dangereux de ne considérer qu'une seule métrique de test.

Considérons un *process* ayant pour rendement 95% et un algorithme de classification qui ait un comportement tel qu'il classe tous les circuits comme fonctionnels. Cet algorithme aura une précision de 95%, ce qui pourrait laisser croire que l'algorithme est relativement précis, alors que la naïveté de sa classification est cachée par le haut rendement du *process*.

Afin de pallier cet effet, d'autres métriques sont considérées pour prendre en compte l'impact d'autres paramètres comme l'équilibre des classes. Quoiqu'il en soit, toutes ces mesures restent une observation particulière, et incomplète, de la matrice de confusion, mais permettent d'apprécier un caractère particulier de cette dernière. On présente maintenant deux de ces métriques.

**Prise en compte de la précision et des erreurs d'indicateur.** La métrique  $\kappa$  prend en compte la précision et les erreurs dans la matrice de confusion. Il est défini par :

$$\kappa = \frac{A_{cc} - \sum_{k=1}^q p_{k+} p_{+k}}{1 - \sum_{k=1}^q p_{k+} p_{+k}} \quad (\text{II.44})$$

où

$$p_{k+} = \sum_{j=1}^q p_{kj} \quad (\text{II.45})$$

$$p_{+k} = \sum_{i=1}^q p_{ik} \quad (\text{II.46})$$

avec :

- $A_{cc}$  La mesure de la précision telle que définie dans le section II.E.2.c
- $q$  Le nombre de classes
- $p_{ij}$  La probabilité de la cellule  $(i, j)$  de la matrice de confusion

D'autres définitions de  $\kappa$  existent. Elles permettent de mettre en évidence d'autres aspects de la matrice de confusion [Stehman, 1997]. Les métriques basées  $\kappa$  sont elles aussi partielles puisque cette mesure n'est qu'une observation particulière de la matrice de confusion. Il faut donc être prudent avant d'utiliser cette mesure pour comparer les résultats de différents algorithmes, ou les résultats obtenus sur différentes populations.

### Prise en compte de l'importance des faux positifs et faux négatifs.

La mesure  $F_\beta$  permet de prendre en compte un déséquilibre d'importance entre la couverture de rendement et la valeur prédictive positive. En d'autres termes, elle s'interprète comme le rapport entre le nombre de vrais positifs parmi les classés positifs et les vrais positifs parmi les réellement positifs. Cette mesure utilise un poids  $\beta$  qui introduit un déséquilibre entre la couverture de rendement et la valeur prédictive positive :

$$\begin{aligned} F_\beta &= \frac{(1 + \beta^2) \frac{S_S * Y_C}{(\beta^2 * S_S) + Y_C}}{(1 + \beta^2) \frac{VP}{(\beta^2 + 1) VP + \beta^2 FN + FP}} \quad (\text{II.47}) \\ &= \end{aligned}$$

La  $F_\beta$ -Mesure prend en compte la couverture de rendement ainsi que la valeur prédictive positive. Ainsi, elle contient à la fois les erreurs de type *I* et les erreurs de type *II*. Cette métrique est utilisable à la condition de définir la valeur du poids  $\beta$ . Dans notre application, nous ne visons pas d'application particulière, et n'avons donc pas de valeur  $\beta$  déterminée. Nous préférons donc utiliser directement les métriques de taux de défaut et perte de rendement.

### II.E.3 Conclusion sur la gestion de la précision

La gestion de la précision est une composante essentielle dans la mise en place du flot proposé. Elle permet de garantir les résultats fournis. Dans le cadre du test, les métriques importantes sont celles de la couverture de faute et de la couverture de rendement, présentées en section I.B.3. On a donné ici des méthodes d'analyse des résultats fournis permettant d'évaluer ces métriques *a priori*.

Une mesure de précision de l'acquisition de données a été fournie en section II.B.1.a. Elle se fonde sur l'évaluation de l'erreur sur le rendement d'un observateur de Monte-Carlo.

Concernant l'erreur commise en phase de modélisation, deux approches ont été présentées dans cette section. La première est celle de la matrice de confusion. Simple à mettre en place, elle est particulièrement adaptée aux algorithmes de classification et permet d'anticiper les métriques de test. La seconde approche est celle portée de l'analyse PAC-learning. Ce cadre de travail



relie le nombre d'échantillons d'entraînement, l'erreur d'entraînement et la confiance dans l'erreur. Associée à la théorie de Vapnik Chervonenkis, elle permet d'établir cette relation pour les classifieurs proposés à partir de leur *dimension VC*.

La gestion de la précision vise essentiellement à prédire les taux de faux positifs et faux négatifs qu'il est possible d'atteindre. Ces grandeurs sont décrites par la répartition des circuits défectueux ou fonctionnels, et par une distribution décrivant l'incertitude qui atteint les échantillons : du bruit de mesure par exemple. Une analyse statistique peut alors être conduite pour obtenir *a priori* les métriques de test à partir de la connaissance d'une distribution de bruit et de la distribution des échantillons défectueux ou fonctionnels. Une telle analyse sera conduite au prochain chapitre pour exprimer l'impact du bruit de mesure sur les métriques de test.

## II.F Conclusion

Les méthodes de traitement de données sont particulièrement adaptées à la problématique de réduction du coût du test. Cette problématique, présentée au chapitre I, a permis de proposer une Vision Unifiée de la réduction du coût du test. Cette dernière est décomposée en trois étapes, qui sont l'acquisition de données, la sélection de données pertinentes, et la construction de modèle. Ces trois étapes peuvent être réalisées par les méthodes de traitement de données présentées dans ce chapitre.

La réduction de dimension est une étape clé dans l'établissement d'une stratégie de test. En effet, la dimension des observations d'un circuit est élevée. Deux approches sont proposées, la réduction de dimension d'une part, et la sélection de caractéristiques d'autre part. La réduction de dimension projette les données dans un espace de dimension plus faible que l'espace d'observation, après avoir réalisé une transformation de l'espace d'observation initial. La sélection de caractéristiques ne construit pas d'espace de projection, mais sélectionne les dimensions de l'espace d'observation les plus "pertinentes". La notion de pertinence doit alors être définie *a priori*.

Les méthodes de construction d'un modèle, aussi appelée induction, sont aussi couvertes par des méthodes de traitement de données. Plusieurs approches ont été présentées dans ce chapitre. Elles permettent de construire des modèles fondés soit sur la recherche des différences entre classes (SVM), sur les similitudes au sein des classes (kNN). Deux autres approches ont aussi été décrites dans ce chapitre, l'une construisant une structure en arbre (arbre de décision) et la dernière une structure neuronale.

Les résultats que proposent ces méthodes dans le cadre de nos travaux sont soumis à des erreurs. L'impact de ces erreurs peut s'apprécier au moment de la classification, où un circuit défaillant peut être classé fonctionnel et inversement. Ces erreurs, présentées à la section I.B.3, impactent la performance de la stratégie de test. Il devient alors intéressant d'analyser cette erreur au fil du flot de conception de la stratégie de test pour assurer que les erreurs de classification seront "acceptables". Par exemple, la quantité d'observations disponibles est liée à la précision atteignable.

Le chapitre suivant propose une solution au problème de la sélection de caractéristiques qui est l'algorithme CDC-FOCUS. Cet algorithme exploite en particulier C-FOCUS présenté dans l'état de l'Art, et l'améliore pour répondre à la réduction d'une fonction coût.

Ensuite, le problème de l'impact du bruit de mesure sur les métriques de test est adressé. L'étude statistique qui y est menée permet d'établir *a priori* le niveau maximal de bruit de mesure acceptable pour respecter des contraintes de faux positifs ou faux négatifs.



## Chapitre III

---

# Méthode de sélection de mesures et classification

---

[...] Il n'y a point de méthode facile pour apprendre les choses difficiles ; l'unique méthode est de fermer sa porte, de faire dire qu'on n'y est pas et de travailler.

---

Joseph DE MAISTRE

LA DÉMARCHE de sélection proposée dans ce chapitre répond à l'analyse du chapitre II. Elle s'inscrit dans la Vision Unifiée présentée à la section I.D et est conçue pour être utilisée dans deux cas différents : pendant la conception du circuit pour déterminer les mesures utiles au test, et pendant le test pour éliminer les mesures redondantes. Nous avons montré d'un côté qu'il existe la possibilité de mesures alternatives, et de l'autre que la notion de pertinence proposée par l'état de l'Art à la section II.C.2.a permet de sélectionner les mesures alternatives d'intérêt pour le test du circuit. Or, cette approche ne prend pas en compte l'objectif de réduction du coût du test : c'est dans ce but que l'on propose ici une définition améliorée de la pertinence. L'algorithme FOCUS présenté en section II.C.2.d est alors modifié pour implémenter cette nouvelle définition.

Ce chapitre replace en section III.A les algorithmes étudiés dans la Vision Unifiée proposée suite à l'analyse de la problématique en section I.D. Ensuite, l'algorithme CDC-FOCUS qui est le résultat de la modification de la définition de la pertinence, est proposé en section III.B. Il est mis en pratique dans l'étape de sélection de dimensions, et présente la capacité de réduire le coût du test. Le comportement de CDC-FOCUS est validé dans ce chapitre sur un exemple de formes géométriques utilisé dans l'état de l'Art. Nous verrons

que la convergence de cet algorithme dépend d'une étape de paramétrage préalable non encore définie. Nous nous proposons alors en section III.C de définir ce paramètre en fonction d'une étude de bruit. Enfin, la validation des mesures sélectionnées par l'algorithme CDC-FOCUS passe par la construction d'un modèle permettant de relier ces mesures à la classification des circuits en fonctionnels ou défectueux. Dans nos travaux nous ne développons pas de nouveaux algorithmes de classification, mais utilisons les algorithmes existant de l'état de l'Art comme le détaille la section III.D.

Ainsi, ce chapitre présente l'algorithme CDC-FOCUS appliqué à l'étape de sélection de dimensions de la méthode pour aboutir à un modèle validant la réduction du coût du test. Il s'agit de la contribution algorithmique de nos travaux.

### III.A Situation de la démarche de sélection dans la Vision Unifiée

La figure III.1 représentant la Vision Unifiée permet de situer l'étape de sélection de dimensions sur laquelle s'applique l'algorithme CDC-FOCUS. Pour rappel, la Vision Unifiée est une réponse proposée à plusieurs problématiques exposées dans l'état de l'Art : élimination des redondances et détection des mesures alternatives corrélées aux performances.

Dans ce cadre, la sélection de dimensions est précédée de l'acquisition de données et suivie par la modélisation.

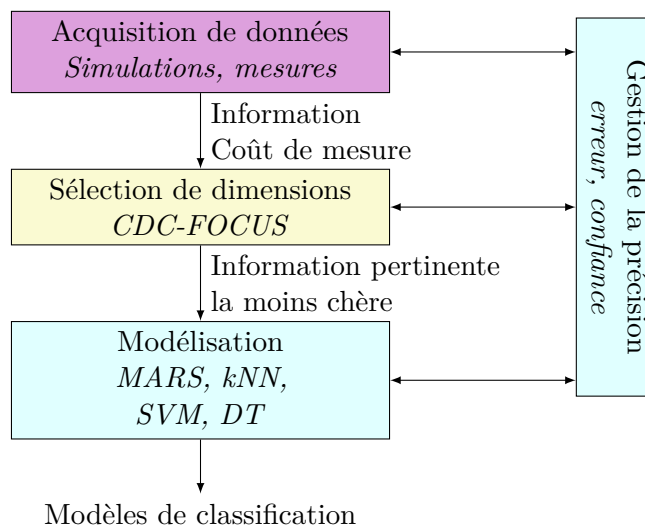


FIGURE III.1 – Les différents blocs de la Vision Unifiée

La structure de la vision proposée ici permet une implémentation par “blocs” et un contrôle sur chacun de ces blocs. En particulier, l’étape de modélisation n’interfère pas avec l’étape de sélection de dimensions. Cette approche relève des méthodes dites “filtre”. Si l’étape de sélection de dimensions est couplée à l’étape de modélisation, dans ce cas l’approche est dite “*wrapper*”. L’intérêt est alors de sélectionner les dimensions en fonction de l’algorithme de modélisation. L’approche *hybride* combine les deux approches précédentes en exploitant la mesure de pertinence pour sélectionner un ensemble de caractéristiques potentiellement pertinentes puis un classifieur pour sélectionner un sous-ensemble parmi ces dernières. [Kohavi and John, 1997] propose une discussion des avantages et inconvénients des méthodes *filtre* et *wrapper*. Il apparaît que les méthodes à *wrapper* peuvent obtenir une bonne précision de classification, les auteurs expliquant ces résultats par la définition de la pertinence relativement à l’algorithme de classification. Or, ce qui nous intéresse est d’avoir la meilleure sélection de dimensions indépendamment de la modélisation. Dans nos travaux, les dimensions doivent en effet être sélectionnées selon la définition des caractéristiques pertinentes, et non selon l’algorithme de classification utilisé. C’est la raison pour laquelle on préfère donc utiliser l’architecture filtre représentée en figure III.1.

Dans ce contexte, l’état de l’Art présenté dans le chapitre II a permis d’identifier deux catégories de méthodes de recherche de signaux d’intérêt : la réduction de dimensions et la sélection de caractéristiques. La première regroupe des méthodes de transformation et de projection qui permettent de représenter un nuage de points selon les directions de variances maximales. Cependant, ces méthodes se prêtent mal aux problèmes fortement non linéaires, comme discuté en section II.C.1. Par ailleurs, les directions de fortes variances ne sont pas nécessairement celles d’intérêt. Enfin, les méthodes de réduction de dimensions proposent comme résultat une combinaison des dimensions d’origine, qui n’a pas de signification physique évidente. Les méthodes de sélection de caractéristiques sont par essence plus adaptées au problème adressé dans ce chapitre puisqu’elles permettent de détecter des dimensions d’intérêt correspondant à une définition particulière de la pertinence. Une telle définition est apportée par l’état de l’Art [John et al., 1994].

On propose dans la suite de modifier la définition de la pertinence afin d’y intégrer la notion de coût. L’algorithme CDC-FOCUS, dérivé de C-FOCUS est développé dans ce but : il permet d’extraire les caractéristiques correspondant à la nouvelle définition de la pertinence.

### III.B Sélection de caractéristiques dirigée par le coût : l'algorithme CDC-FOCUS

Dans cette section, la notion de coût est développée pour modifier la notion de pertinence de l'état de l'Art. Ensuite, l'algorithme de l'état de l'Art C-FOCUS est modifié en conséquence. De cette manière, l'extraction de caractéristiques assure d'obtenir les caractéristiques pertinentes (au sens de l'état de l'Art) et au coût le plus faible.

Cette section précisera tout d'abord les définitions de coût. On présentera ensuite l'algorithme CDC-FOCUS (Cost Driven Continuous FOCUS).

#### III.B.1 Pertinence et coût d'un ensemble de caractéristiques

Les définitions introduites en section II.C.2.a définissent la pertinence d'une caractéristique. Dans le cadre de la réduction du coût du test, plusieurs définitions intégrant la notion de coût sont proposées ci-après. La première définition propose de simplement calculer le coût de mesure d'un ensemble de caractéristiques comme la somme des coûts unitaires de chacune des caractéristiques. Le coût associé à la mesure de la caractéristique  $c_i$  est alors noté  $f_c(\text{Mes}(c_i))$ .

**Définition 9** (Coût de la mesure d'un ensemble de caractéristiques). *Le coût de la mesure d'un ensemble de caractéristiques est la somme des coûts de mesure de chacune des caractéristiques le composant :*

$$C_S(\mathcal{C}) = \sum_{c_i \in \mathcal{C}} f_c(\text{Mes}(c_i)) \quad (\text{III.1})$$

avec :

$C_S(\mathcal{C})$	Le coût de l'ensemble de caractéristiques $\mathcal{C}$
$\mathcal{C}$	L'ensemble de caractéristiques
$\text{Mes}(c_i)$	La mesure d'une caractéristique $c_i \in \mathcal{C}$

Cependant, cette définition n'intègre pas les coûts associés aux changements d'état du testeur et du circuit entre deux mesures. Pour le prendre en compte, nous proposons de distinguer un *ensemble* de caractéristiques et une *séquence* de caractéristiques, cette dernière étant une suite ordonnée de caractéristiques. Le coût d'une séquence de caractéristique prends alors en compte ce changement d'état, et nous proposons alors la définition suivante :

**Définition 10** (Coût d'une séquence de caractéristiques). *Le coût d'une séquence de caractéristiques est donné par la somme des coûts de mesure de*

chaque caractéristique de la séquence et des coûts associés aux changements d'état du circuit et du testeur :

$$C_O(\mathcal{S}_n) = C_S(\{c|c \in \mathcal{S}_n\}) + \sum_{i=2}^n f_s(c_{i-1}, c_i) \quad (\text{III.2})$$

avec :

- $C_O(\mathcal{S}_n)$     *Le coût de la séquence de caractéristiques  $\mathcal{S}_n$*
- $\mathcal{S}_n$             *La séquence de  $n$  caractéristiques*
- $f_s(c_a, c_b)$     *Le coût du passage de la mesure de la caractéristique  $c_a$  à la mesure de la caractéristique  $c_b$*

Une hypothèse simplificatrice forte considère  $f_s$  comme une fonction nulle, ce qui porte le coût d'une séquence de caractéristiques au coût de l'ensemble des caractéristiques la composant. De cette façon, le coût de mesure d'une séquence de caractéristiques est le coût de l'ensemble des mesures la composant. Cette hypothèse n'est généralement pas vérifiée au moment du test, car pour faire une mesure, comme précisé avant, le circuit et le testeur doivent être placés dans un état particulier, adapté à la nouvelle mesure. Ainsi, la séquence de mesures  $\{c_a, c_b\}$  n'aura pas le même coût que la séquence  $\{c_b, c_c\}$ . Un algorithme présenté par [Huss and Gyurcsik, 1991] adresse le problème d'organisation de la séquence de tests. Il peut être utilisé pour contourner le problème de l'ordonnancement d'une séquence de mesures de caractéristiques après la sélection des caractéristiques à tester. Cet algorithme organise la séquence de tests de façon à minimiser la quantité  $\sum_{i=2}^n f_s(c_{i-1}, c_i)$ . Dans le cadre de nos travaux, cette organisation n'impacte pas la construction du modèle. Ainsi, l'algorithme de [Huss and Gyurcsik, 1991] peut être utilisé en aval de la méthode proposée.

Les caractéristiques recherchées dans le cadre de la méthode proposée sont dérivées de l'algorithme C-FOCUS (voir section II.C.2.d). Ce dernier extrait l'ensemble des caractéristiques dont le cardinal est le plus faible et qui explique le mieux la classification du nuage de points considéré. Ici il s'agit d'extraire l'ensemble des caractéristiques de coût le plus faible, expliquant le mieux la classification. C'est alors les séquences correspondant à cette nouvelle définition que l'algorithme CDC-FOCUS doit extraire :

**Définition 11** (Séquence de caractéristiques pertinentes de plus faible coût). *La séquence recherchée  $\hat{\mathcal{S}}$  est telle qu'elle contienne toutes les caractéristiques à pertinence forte et les caractéristiques à pertinence faible suffisantes pour expliquer la classification du nuage de points considéré, et de coût  $C_S(\hat{\mathcal{S}})$*



*minimal.*

Comme indiqué précédemment, ce ne sera pas la grandeur  $C_O(\hat{\mathcal{S}})$  qui sera minimisée, mais la grandeur  $C_S(\hat{\mathcal{S}})$ . Ainsi, lorsque  $\hat{\mathcal{S}}$  aura été obtenu par minimisation de  $C_S$ , il sera possible de minimiser la grandeur  $C_O$  par l'utilisation d'un algorithme tel que celui présenté dans [Huss and Gyurcsik, 1991]. Les deux minimums ne convergeront pas vers la même valeur (car dans ce cas, il existerait une séquence  $\hat{\mathcal{S}}_n$  telle que  $\sum_{i=2}^n f_s(\hat{c}_{i-1}, \hat{c}_i) = 0$ , ce qui est faux dans la plupart des cas), mais sous réserve que  $C_S(\{c|c \in \hat{\mathcal{S}}_n\}) \gg \sum_{i=2}^n f_s(\hat{c}_{i-1}, \hat{c}_i)$ , alors  $\hat{\mathcal{S}}_n$  minimise  $C_O(\mathcal{S}_n)$ .

### III.B.2 Présentation de CDC-FOCUS

L'algorithme CDC-FOCUS donné par l'algorithme 5 en annexe est dérivé de C-FOCUS, ce dernier exploitant les connaissances des conflits déjà analysés par l'utilisation d'une queue de calcul  $Q$ , apportée par FOCUS2, et s'appliquant à des grandeurs continues.

La notion de coût est introduite par l'évaluation systématique des coûts de chaque ensemble de caractéristiques satisfaisant potentiellement les conditions de pertinence. Par ailleurs, il est possible de savoir *a priori* les coûts minimum et maximum des ensembles de caractéristiques en fonction de leur cardinal. Cette connaissance permet à l'algorithme de savoir s'il est utile de poursuivre la recherche d'un ensemble de caractéristiques pertinentes moins cher.

L'algorithme CDC-FOCUS utilise un test de suffisance "isSufficient" dont le but est de vérifier, pour un ensemble de caractéristiques donné, qu'il n'y a pas de contradictions dans l'ensemble d'entraînement. L'algorithme "isSufficient" proposé dans ce travail est décrit par l'algorithme 6 en annexe. Sa particularité est d'itérer sur les échantillons. Notons qu'un autre algorithme de test de suffisance par récursivité est proposé par [Almuallim and Dietterich, 1994]. Ce dernier itère sur les caractéristiques à tester. Le choix de l'implémentation par itération sur les échantillons est justifié par l'absence de récursivité, ce qui simplifie l'implémentation Matlab.

L'innovation de CDC-FOCUS est de déterminer un *horizon* de calcul. Cet horizon est un tableau contenant les coûts minimaux et maximaux des cardinaux respectifs de toutes les parties de l'ensemble des caractéristiques. Pour obtenir le coût minimal et le coût maximal du cardinal  $N$ , les coûts sont triés par ordre croissant, et la somme des  $N$  plus petits coûts est le coût minimal tandis que la somme des  $N$  plus grands est le coût maximal. L'horizon est considéré comme un outil permettant de répondre à la question : *est-il préfé-*

table de réaliser une mesure complexe et coûteuse ou un ensemble de mesures simples et peu chères ?

Par exemple, soit un problème de dimension 3 tel que  $f_C(\text{Mes}(c_1)) = 2$ ,  $f_C(\text{Mes}(c_2)) = 5$ ,  $f_C(\text{Mes}(c_3)) = 1$ . L'horizon est représenté dans le tableau III.1.

Card.	Coût min.	Coût max.
1	1	5
2	3	7
3	8	8

TABLE III.1 – Exemple d'horizon déterminé par CDC-FOCUS

L'intérêt d'utiliser l'horizon est de relier le cardinal d'une partie au coût atteignable en poursuivant l'exécution de l'algorithme. Pour l'exemple précédent, si l'algorithme détecte la séquence de caractéristique  $\{c_2\}$ , de coût de mesure 5 comme étant pertinente, alors l'horizon indique qu'il est possible de poursuivre l'algorithme car le coût minimal des parties de cardinal 2 est 3 (à savoir la mesure de  $\{c_1, c_3\}$ ). De la même façon, si l'algorithme détecte la séquence de caractéristiques  $\{c_2, c_3\}$  de coût 6 comme étant pertinente, l'horizon lui indique qu'il est possible de poursuivre la recherche d'une partie de cardinal 2 car le coût minimal en est de 3, mais qu'il est inutile de chercher une séquence de cardinal 3 car le coût minimal est de 8.

Pour ce qui est de l'initialisation de l'algorithme CDC-FOCUS, elle est similaire à celle de FOCUS-2, à ceci près que les nouvelles variables `baseCost` et `baseResult` sont initialisées à 0 et  $\emptyset$  (ensemble vide) respectivement. La découverte de la première séquence pertinente entraîne son enregistrement dans la variable `baseResult` et celui de son coût dans la variable `baseCost`. Le coût `baseCost` est comparé à l'horizon, ce qui permet de savoir jusqu'à quel cardinal l'algorithme doit être exécuté. Par la suite, seules les séquences dont le coût est inférieur à `baseCost` sont testées par le test de suffisance "isSufficient".

L'implémentation de l'algorithme est critique pour en préserver les performances. Ainsi, l'essentiel de l'algorithme est codé en MATLAB, et les fonctionnalités les plus critiques en occupation mémoire et temps de calcul sont implémentées en C. Ces fonctionnalités critiques sont le test de suffisance `isSufficient` et la construction de la matrice des conflits `ConflictMatrix`.

### III.B.3 Un exemple d'utilisation de CDC-FOCUS

L'exemple introduit dans cette section est celui des figures géométriques, utilisé par [Arauzo Azofra et al., 2003] et qui implique quatre classes de figures géométriques, à savoir, triangle isocèle, triangle équilatéral, carré et rectangle. Les grandeurs mesurées sont :

- nombre de cotés ;
- plus grand coté  $LS$  ;
- plus petit coté  $SS$  ;
- périmètre ;
- surface ;
- rapport du plus petit et du plus grand coté  $SS/LS$  ;

Ces mesures sont détaillées dans le tableau III.2.  $s_1$  et  $s_2$  représentent deux mesures des figures géométriques.

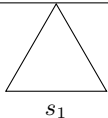
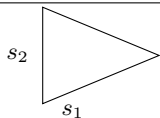
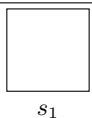

Classes	 $s_1$	 $s_2$ $s_1$	 $s_1$	 $s_2$ $s_1$
Nombre de côtés	3	3	4	4
Plus petit côté SS	$s_1$	$\min(s_1, s_2)$	$s_1$	$\min(s_1, s_2)$
Plus grand côté LS	$s_1$	$\max(s_1, s_2)$	$s_1$	$\max(s_1, s_2)$
Périmètre	$3 \times s_1$	$2 \times s_1 + s_2$	$4 \times s_1$	$2 \times s_1 + 2 \times s_2$
Surface	$\sqrt{\frac{3}{4} \times s_1^2}$	$\frac{s_2 \times \sqrt{4 \times s_1^2 - s_2^2}}{4}$	$s_1^2$	$s_1 \times s_2$
SS/LS	1	SS/LS	1	SS/LS

TABLE III.2 – Les 4 classes et les caractéristiques de l'exemple

L'objectif de cet exemple est d'extraire les caractéristiques les moins chères qui permettent de distinguer un élément appartenant à l'une des quatre classes. Dans le cadre de nos travaux, l'algorithme CDC-FOCUS doit obtenir des résultats identiques à ceux établis par [Arauzo Azofra et al., 2003] lorsqu'aucun coût n'est associé aux caractéristiques. À savoir, l'algorithme doit détecter l'ensemble de caractéristiques *Nombre de côtés* et *SS/LS* comme étant pertinent, et de cardinal le plus faible. Et lorsqu'un coût est associé, il doit extraire les caractéristiques les moins chères qui permettent de classer un échantillon parmi les quatre classes proposées. Pour obtenir les résultats des performances de CDC-FOCUS, des échantillons de figures géométriques sont générés et CDC-FOCUS est exécuté avec différents paramètres. Ces paramètres sont le nombre d'échantillons utilisés, le nombre de caractéristiques non pertinentes

ajoutées aux caractéristiques des figures géométriques, et le seuil introduit par C-FOCUS. Pour notre exemple, les différents nombres d'échantillons utilisés sont 50, 100, 250 et 500, les différentes valeurs de caractéristiques non pertinentes sont 1,10 et 25, et les différentes valeurs de seuil sont 0.025, 0.050, 0.100 et 0.200.

La génération des échantillons d'entraînement est similaire à la méthode proposée par [Arauzo Azofra et al., 2003]. En particulier la génération assure que les échantillons des classes Triangle Isocèle et Rectangle observent un écart d'au moins 5% entre  $s_1$  et  $s_2$ . Cet écart permet de marquer la distinction entre un triangle isocèle et un triangle équilatéral, et entre un rectangle et un carré. Dans ces conditions, lorsque CDC-FOCUS est utilisé sans fonction coût (i.e. toutes les caractéristiques ont le même coût), les résultats obtenus sont effectivement similaires à C-FOCUS présentés par [Arauzo Azofra et al., 2003].

Pour exploiter la capacité de CDC-FOCUS à gérer les coûts de test, on attribue maintenant aux caractéristiques un coût de 1, sauf à la caractéristique SS/LS qui porte un coût de 3. Le résultat attendu est le remplacement de cette caractéristique par les caractéristiques SS et LS dont dépend directement SS/LS, mais dont le coût est plus faible. Les résultats, pour différents seuils, sont présentés dans le tableau III.3 page suivante. Ils montrent que le remplacement de la caractéristique SS/LS par les caractéristiques SS et LS a bien été effectué pour un seuil strictement inférieur à 0.2. Pour ce seuil, le remplacement n'a pas été effectué car le test de suffisance a échoué. En effet, il existe deux échantillons de classes différentes dont la distance est inférieure à 0.2. Il est à noter aussi qu'aucune caractéristique non pertinente n'a été sélectionnée par l'algorithme. Cet exemple valide donc le comportement de CDC-FOCUS, et montre que le seuil de paramétrage influence les résultats. Cet exemple reste cependant simple et éloigné des applications sur circuit. Nous le mettrons en œuvre sur des bases de données issues de circuits dans le chapitre IV.

#### III.B.4 Conclusion sur CDC-FOCUS

L'algorithme présenté dans cette section, CDC-FOCUS, est construit pour tirer profit du comportement de C-FOCUS tout en garantissant la recherche complète des caractéristiques pertinentes. Ainsi, il assure l'extraction des caractéristiques pertinentes, si elles existent. De plus, son comportement est modifié de façon à poursuivre les recherches de caractéristiques s'il est possible d'extraire des caractéristiques pertinentes moins chères.

La complexité algorithmique de CDC-FOCUS rend le temps d'exécution et l'espace mémoire occupé sensible aux choix d'implémentation. Dans le cadre de la méthode mise en place, la recherche se déroule hors-ligne, c'est-à-dire en

Seuil : 0.025		Nombre de caractéristiques non pertinentes		
Nombre d'échantillons	1	10	25	
50	SS, Perimeter	Perimeter, SS	SS, Perimeter	
100	NSides, LS, SS	NSides, LS, SS	LS, SS, NSides	
250	NSides, LS, SS	LS, NSides, SS	NSides, LS, SS	
500	NSides, LS, SS	NSides, LS, SS	NSides, LS, SS	

Seuil : 0.050		Nombre de caractéristiques non pertinentes		
Nombre d'échantillons	1	10	25	
50	NSides, LS, SS	NSides, LS, SS	NSides, LS, SS	
100	NSides, LS, SS	NSides, LS, SS	NSides, SS, LS	
250	NSides, LS, SS	LS, NSides, SS	LS, NSides, SS	
500	NSides, LS, SS	NSides, LS, SS	NSides, LS, SS	

Seuil : 0.100		Nombre de caractéristiques non pertinentes		
Nombre d'échantillons	1	10	25	
50	NSides, SS, LS	NSides, LS, SS	LS, NSides, SS	
100	NSides, LS, SS	NSides, LS, SS	LS, NSides, SS	
250	NSides, LS, SS	NSides, LS, SS	LS, NSides, SS	
500	NSides, SS, LS	NSides, SS, LS	NSides, LS, SS	

Seuil : 0.200		Nombre de caractéristiques non pertinentes		
Nombre d'échantillons	1	10	25	
50	NSides, SS/LS	SS/LS, NSides	NSides, SS/LS	
100	NSides, SS/LS	NSides, SS/LS	NSides, SS/LS	
250	NSides, SS/LS	NSides, SS/LS	NSides, SS/LS	
500	NSides, SS/LS	SS/LS, NSides	SS/LS, NSides	

TABLE III.3 – Sélection de caractéristiques pertinentes avec l'algorithme CDC-FOCUS, avec un coût différent pour la mesure des caractéristiques

amont du test effectif des circuits. Ainsi, le temps d'exécution de cet algorithme n'est pas inclus dans le temps de test des circuits.

Par ailleurs, un inconvénient provient de la manipulation de grandeurs continues. Cette capacité a été apportée par C-FOCUS qui a introduit un seuil  $\delta$  permettant de caractériser la proximité de deux valeurs de caractéristiques. Le choix de ce seuil *a priori* est délicat. Une méthode heuristique pour le définir est proposée par [Arauzo Azofra et al., 2003]. Les auteurs réalisent plusieurs essais avec un seuil de plus en plus faible, jusqu'à ce que l'algorithme extraie des caractéristiques qui résolvent tous les conflits. Ils laissent cependant ouverte la question de l'étude de ce seuil, et du choix de sa valeur.

Dans la suite, nous proposons une analyse de bruit pour apporter des éléments de réponse à ce problème du choix du seuil. Cette analyse est suggérée par la nature du seuil qui est utilisé dans C-FOCUS et dans CDC-FOCUS, comme un paramètre de sensibilité. D'un côté, l'intuition indique que le seuil se comporte comme un réjecteur de bruit de mesure. De l'autre côté, le bruit de mesure impacte les métriques de test (précision, taux de faux positifs) et la compréhension de l'impact du bruit de mesure sur les métriques de test permettrait d'évaluer le seuil de CDC-FOCUS en fonction des métriques de test souhaitées.

### III.C Influence du bruit de mesure sur les métriques de test

Le problème de la dégradation des métriques de test a été étudié par [Stratigopoulos et al., 2009b]. Les auteurs proposent une analyse statistique du test de circuit qui permet de déduire les métriques de test obtenus par Monte-Carlo. En particulier, ils analysent l'erreur commise sur l'évaluation d'une métrique de test par l'inégalité de Chebyshev :

$$Pr \{ |T^* - T| < \epsilon \} \geq \delta = 1 - \frac{\sigma^2}{\epsilon^2} \quad (\text{III.3})$$

avec :

$T^*$	L'évaluation de la valeur de la métrique de test $T$ par simulation Monte-Carlo
$T$	La vraie valeur de la métrique de test
$\delta$	La confiance dans le respect de l'erreur $\epsilon$
$\epsilon$	L'erreur entre la vraie valeur et la valeur obtenue par simulation Monte-Carlo
$\sigma$	La variance de l'observateur de Monte-Carlo

L'étude de l'impact du bruit sur les métriques de test permet de fixer une limite qui assure un certain taux de succès au test (taux de faux positifs et taux de faux négatifs). Les résultats de cette analyse ont pour intérêt d'anticiper l'impact et de permettre la définition d'un seuil pour l'utilisation de CDC-FOCUS. On souhaite donc relier les métriques de test à un modèle de bruit statistique. En plus d'aider à fixer le seuil  $\delta$  utilisé dans CDC-FOCUS, cette étude est directement utile pour l'objectif de réduction du coût du test de circuit. En effet, le coût du test peut être réduit par l'utilisation d'outils de mesure moins précis, mais le manque de précision de mesure s'accompagne d'une dégradation de la qualité de la prise de décision quant à l'état du circuit testé. On peut alors mener une étude de l'impact du bruit sur les métriques de test [Verdy et al., 2015].

Dans cette section, une analyse statistique est conduite pour relier l'incertitude de mesure aux métriques de test. Pour ce faire, nous posons une hypothèse sur le modèle de bruit en premier lieu, puis opérons une déconvolution. Les métriques de test sont ensuite exprimées sous forme de probabilités [Stratigopoulos et al., 2009b]. Pour simplifier, nous supposons ici que les spécifications ne sont exprimées que sur une seule dimension.

### III.C.1 Hypothèse du bruit additif et convolution des PDFs

Le signal considéré ici est une mesure de performance sur un ensemble de circuits. Il n'y a pas de modèle de bruit *a priori* défini. Le bruit est supposé additif, indépendant du signal, cf. la figure III.2a :

$$\mathcal{M} = \mathcal{X} + \mathcal{N} \tag{III.4}$$

avec :

$\mathcal{M}$	La variable aléatoire caractérisant la mesure
$\mathcal{X}$	La variable aléatoire caractérisant la vraie valeur à mesurer
$\mathcal{N}$	La variable aléatoire caractérisant le bruit sur la valeur à mesurer

Ce bruit peut être, par exemple, issu de l'incertitude induite par la réduction du temps de mesure. De cette façon, la densité de probabilité du signal bruité est la convolution de la densité de probabilité du signal avec celle du bruit :

$$f_{\mathcal{M}}(z) = \int_{-\infty}^{+\infty} f_{\mathcal{X}}(x) f_{\mathcal{N}}(z - x) dx \quad (\text{III.5})$$

avec :

- $f_{\mathcal{M}}$  La densité de probabilité de  $\mathcal{M}$
- $f_{\mathcal{X}}$  La densité de probabilité de  $\mathcal{X}$
- $f_{\mathcal{N}}$  La densité de probabilité de  $\mathcal{N}$

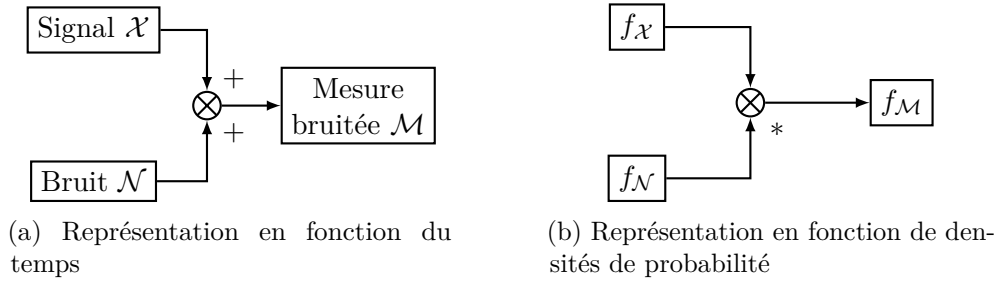


FIGURE III.2 – La densité de probabilité d'un signal affecté par un bruit additif non corrélé est la convolution de leurs densités de probabilité respectives

Dans le cas de la figure III.2, qui correspond à l'équation (III.4), l'expression du rendement à partir des variables aléatoires caractérisant le signal d'intérêt et le bruit est explicitée par :

$$Y_{imp} = F_{\mathcal{M}}(S) = \int_{-\infty}^S \int_{-\infty}^{+\infty} f_{\mathcal{X}}(x) f_{\mathcal{N}}(z - x) dx dz \quad (\text{III.6})$$

avec :

- $Y_{imp}$  le rendement théorique observable après application du bruit
- $F_{\mathcal{M}}$  la fonction de répartition de  $\mathcal{M}$
- $S$  la spécification

La relation entre les densités de probabilité apparaissant sous forme de convolution, il est possible d'analyser les informations bruitées mesurées dans le domaine de Fourier, où la convolution sera une multiplication, plus simple à manipuler. Les informations extraites de cette analyse ne concernent que les



densités de probabilité, et pas les signaux eux-mêmes. Ce choix est pertinent puisque les métriques de test sont définies à partir des densités de probabilité des performances.

On propose maintenant une méthode en deux étapes pour évaluer l'impact du bruit sur les métriques de test. La première étape permet de déconvoluer les mesures pour en extraire, à partir des relations précédemment établies, les densités de probabilité des signaux d'intérêt. Ensuite, les métriques de test (faux positif et faux négatif) sont explicitées à l'aide de PDFs dans la deuxième étape. Ces deux étapes sont représentées en figure III.3.

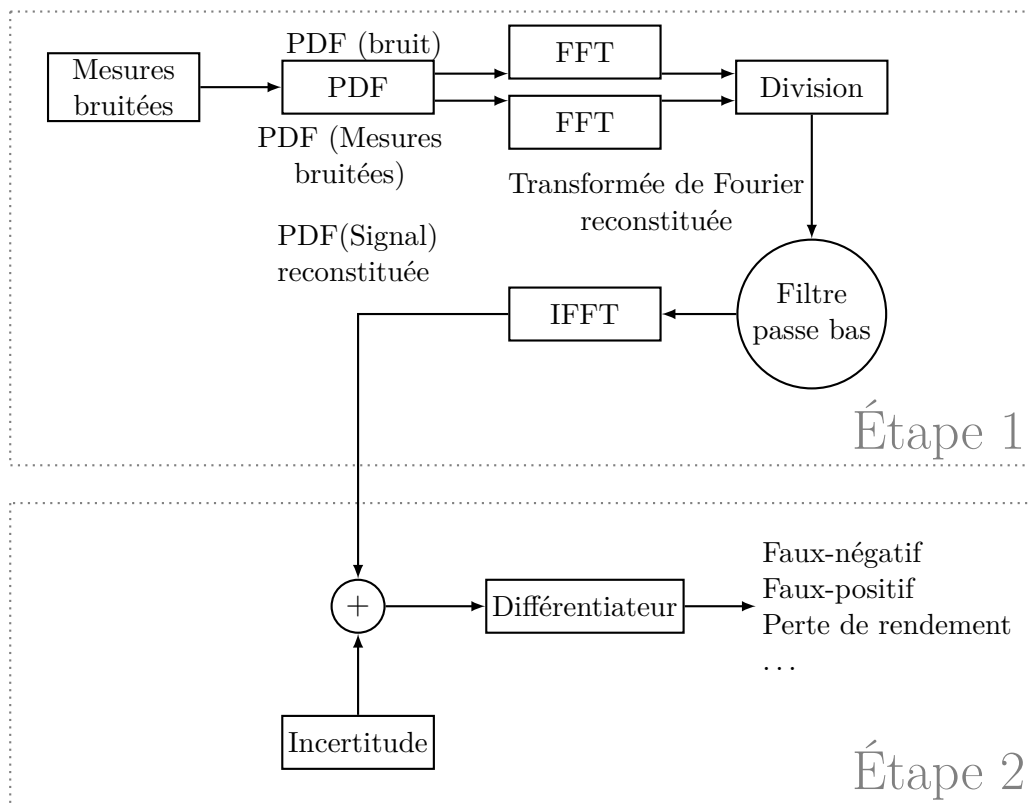


FIGURE III.3 – Méthode d'évaluation de l'impact des incertitudes de mesure sur les métriques de test. *IFFT* signifie *Inverse Fast Fourier Transform* (inverse de la transformée de Fourier rapide) et *PDF* signifie *Probability Density Function* (Densité de probabilité).

### III.C.2 Étape 1 : Évaluation de l'impact par déconvolution

Pour analyser l'impact du bruit sur la mesure, les densités de probabilité des performances sont évaluées par l'intermédiaire des signaux mesurés. Une première étape consiste à évaluer le bruit subit par les mesures. D'après les

hypothèses sur le modèle de bruit, il est possible de l'extraire par la mesure d'un unique échantillon à plusieurs reprises. La densité de probabilité du bruit est alors déduite de ces mesures.

Après obtention de la densité de probabilité du bruit, il est possible *d'éliminer* son impact. En effet, la densité de probabilité des mesures bruitées étant la convolution du signal d'intérêt avec le bruit, la Transformée de Fourier de la densité de probabilité est la multiplication des Transformées de Fourier des densités de probabilité du signal d'intérêt et du bruit. Dans le domaine de Fourier, il est alors possible de retrouver la densité de probabilité du signal d'intérêt. La méthode mise en place est une déconvolution dans le domaine de Fourier. Elle peut être résumée comme suit.

Soient  $a, b, c$  trois signaux tels que  $(a, b, c) \in (L^1(\mathbb{R}))^3$ , liés entre eux par la relation de convolution :

$$a * b = c \quad (\text{III.7})$$

La transformée de Fourier de  $c$  s'exprime alors comme la multiplication des transformées de Fourier de  $a$  et  $b$  :

$$\mathcal{F}(a) \times \mathcal{F}(b) = \mathcal{F}(c) \quad (\text{III.8})$$

Ensuite, sous réserve que  $b$  soit non nul sur son domaine de définition, il vient :

$$\mathcal{F}(a) = \frac{\mathcal{F}(c)}{\mathcal{F}(b)} \quad (\text{III.9})$$

Et enfin, si  $b$  et  $c$  sont de carré intégrable, alors il vient :

$$\mathcal{F}^{-1}(\mathcal{F}(a)) = \hat{a} = \mathcal{F}^{-1}\left(\frac{\mathcal{F}(c)}{\mathcal{F}(b)}\right) \quad (\text{III.10})$$

où  $\mathcal{F}^{-1}$  est la transformée de Fourier inverse.

Cette méthode de déconvolution permet alors d'estimer la densité de probabilité du signal d'intérêt à partir des mesures bruitées et d'un modèle du bruit supposé additif :

$$\widehat{f_{\mathcal{X}}} = \mathcal{F}^{-1}\left(\frac{\mathcal{F}(f_{\mathcal{M}})}{\mathcal{F}(f_{\mathcal{N}})}\right) \quad (\text{III.11})$$

avec :

- $\widehat{f_{\mathcal{X}}}$  La densité de probabilité estimée du signal d'intérêt  $\mathcal{X}$
- $f_{\mathcal{N}}$  La densité de probabilité du bruit  $\mathcal{N}$

Notons que le problème de déconvolution est connu comme étant mal-posé

(au sens de Hadamard). En effet, l'existence d'une solution au problème de déconvolution n'est pas assurée. Par ailleurs, si une solution existe, elle peut être très sensible aux variations des données d'entrée [Jacquelin et al., 2003]. La reconstruction de la division de l'équation (III.11) ne donne alors pas le résultat attendu en raison des approximations de calcul de la division et du dénominateur approchant 0. Ces phénomènes s'observent dans le domaine fréquentiel à partir d'une certaine fréquence. Pour résoudre ce problème de sensibilité, une solution simple envisagée est d'utiliser un filtre passe-bas pour ne reconstruire la densité de probabilité que sur les basses fréquences. La fréquence de coupure de ce filtre est calculée en tenant compte de la connaissance de la composition fréquentielle de la densité de probabilité du signal mesuré.

### III.C.3 Étape 2 : Analyse du signal reconstruit

La déconvolution précédente a permis de reconstruire la densité de probabilité du signal d'intérêt  $\mathcal{S}$  qui peut alors être comparée à celle du signal mesuré  $\mathcal{M}$ . Cette comparaison permet de retrouver les métriques de perte de rendement, mais aussi de faux positifs et de faux négatifs.

Le rendement peut être obtenu directement par l'intégration de la fonction de probabilité sur les spécifications, voir l'équation (III.6). Il est alors possible, par l'intégration des densités de probabilité du signal reconstruit et du signal mesuré, d'évaluer la perte de rendement induite par la mesure, et ce pour différents niveaux de bruit.

L'estimation du nombre de faux positifs demande une analyse plus délicate des densités de probabilité, présentée ci-après. La probabilité d'apparition FP d'un faux positif est donnée par :

$$\begin{aligned} \text{FP} &= \Pr \left\{ \bar{R} \cap M \right\} \\ &= \Pr \{M\} \times \Pr \left\{ \bar{R} | M \right\} \end{aligned} \quad (\text{III.12})$$

avec :

$R$	L'évènement <i>un circuit respecte réellement les spécifications</i>
$M$	L'évènement <i>un circuit mesuré respecte les spécifications</i>
$\bar{X}$	Tous les évènements complémentaires de $X$
$\Pr \{X\}$	La probabilité d'apparition de $X$

Ensuite, le calcul de la probabilité de l'évènement  $M$  est direct :

$$\Pr \{M\} = \int_{-\infty}^{\text{Spec}} f_{\mathcal{M}}(t) dt$$

Le calcul de la probabilité conditionnelle consiste à calculer, à l'aide du modèle de bruit, la probabilité qu'une valeur mesurée soit en réalité en dehors des spécifications. Puis à intégrer cette probabilité sur le support  $\Gamma_P$  défini par la répartition des valeurs de mesures validant les spécifications :

$$\Pr \{ \bar{R} | M \} = \int_{-\infty}^{\text{Spec}} \left( \int_{\text{Spec}}^{\infty} f_{\mathcal{N}}(t) \times \Gamma_{\mathcal{M}}(m) dt \right) dm \quad (\text{III.13})$$

$$\Gamma_{\mathcal{M}}(m) = \frac{f_{\mathcal{M}}(m)}{Y_P}$$

avec :

- Spec La spécification
- $\Gamma_{\mathcal{M}}$  Le support d'intégration des circuits mesurés qui respectent les spécifications
- $Y_P$  Le rendement calculé sur la dispersion reconstruite

Le calcul de la probabilité d'apparition d'un faux négatif FN est similaire :

$$\begin{aligned} \text{FN} &= \Pr \{ R \cap \bar{M} \} \\ &= \Pr \{ \bar{M} \} \times \Pr \{ R | \bar{M} \} \end{aligned} \quad (\text{III.14})$$

avec

$$\Pr \{ \bar{M} \} = \int_{\text{Spec}}^{\infty} P_{\mathcal{M}}(t) dt$$

$$\Pr \{ R | \bar{M} \} = \int_{\text{Spec}}^{\infty} \left( \int_{-\infty}^{\text{Spec}} P_{\mathcal{N}}(t) \times \Gamma_N(m) dt \right) dm$$

et

$$\Gamma_N(m) = \frac{P_M(m)}{1 - Y_P}$$

où  $\Gamma_N$  est le support d'intégration des circuits mesurés qui ne respectent pas les spécifications.

Les équations (III.12) et (III.14) permettent donc d'estimer les métriques de test. L'hypothèse de bruit additif, quant à elle, permet d'utiliser la convolution des distributions du signal d'intérêt et du bruit, et donc la déconvolution dans le domaine de Fourier.

Dans le cadre de nos travaux, ces résultats permettent d'anticiper la valeur maximale de bruit autorisée sur une mesure pour respecter les métriques de test connues *a priori*. Comme indiqué en introduction de cette section, ces résultats sont établis pour des spécifications exprimées sur une seule dimension. Le cas

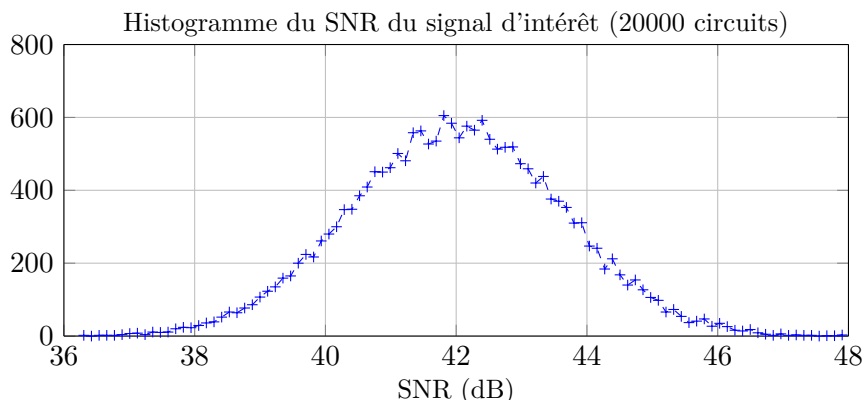


FIGURE III.4 – Distribution théorique du SNR

de spécifications multidimensionnelles est résolu en remplaçant les intégrales simples par des intégrales multidimensionnelles et les densités de probabilité par des densités de probabilité conjointes. Cependant, la résolution numérique de ces intégrales multidimensionnelles peut être un problème.

### III.C.4 Exemple d'application

L'objectif de cet exemple est de mesurer le rapport signal-à-bruit (SNR) d'un ensemble de circuits afin de valider leur bon comportement. Cette expérience n'est réalisée qu'en simulation. La seule perturbation considérée sur le signal d'intérêt est induite par la variation technologique. Dans le cadre de cet exemple, la simulation est construite de manière à ce que le SNR du Signal soit de 42dB. Les variations technologiques considérées modifient ce signal, ce qui conduit à un Signal d'Intérêt dont le SNR dévie autour de 42dB (voir figure III.4). C'est ce signal d'intérêt qui permet d'évaluer le bon fonctionnement du circuit, i.e. le respect de la spécification de SNR.

Cependant, la mesure du SNR nécessite une transformée de Fourier et un détecteur de puissance. Le calcul de la transformée de Fourier est réalisé avec un nombre limité d'échantillons du signal d'intérêt, voir la figure III.5. On se propose d'analyser l'impact de la limitation du nombre d'échantillons mesurés sur la qualité du test fondé sur cette mesure. En effet, la limitation du nombre de points mesurés engendre une erreur sur le calcul de la transformée de Fourier du signal d'intérêt, qui induit une erreur dans la mesure du SNR.

La figure III.6 présente l'impact du nombre de points conservés pour le calcul de la transformée de Fourier. Il s'agit de mesures répétées sur un même circuit, avec différents nombres d'échantillons. Cette figure représente *in fine* le modèle de bruit associé à différentes acquisitions de mesure. Ce bruit est considéré comme additif, indépendant du signal d'intérêt.

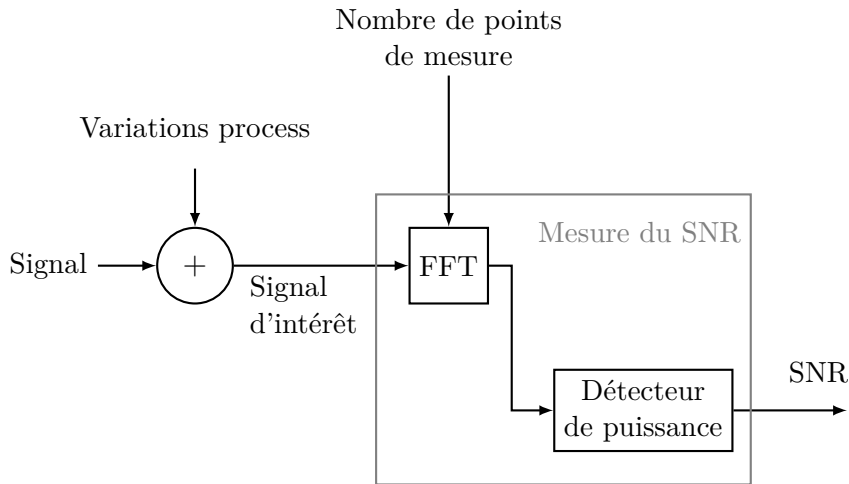


FIGURE III.5 – Schéma de l'évaluation du SNR. Les variations technologiques influent sur la puissance du bruit en sortie du dispositif. Le nombre de points de mesure utilisés pour le calcul de la FFT est source d'incertitude dans l'évaluation du SNR.

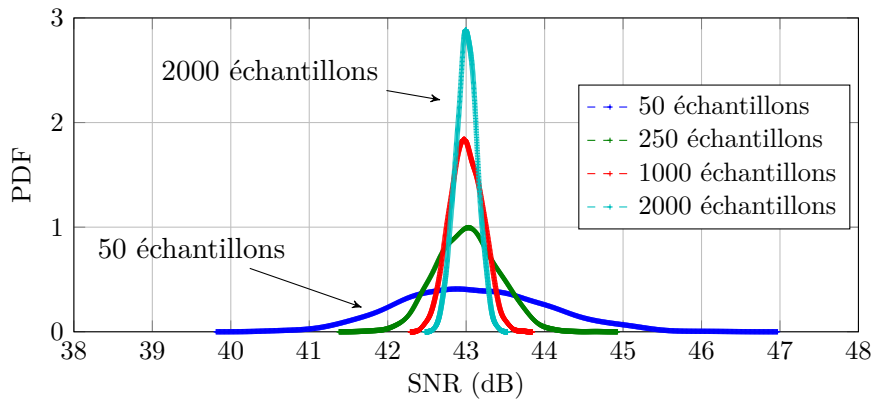


FIGURE III.6 – Dispersion des mesures sur un unique circuit, lorsque différentes quantités d'échantillons sont utilisées pour le calcul de la Transformée de Fourier

On souhaite évaluer les métriques de test en fonction de différentes valeurs de bruit, induite par la réduction du nombre d'échantillons conservés pour la mesure de la transformée de Fourier. Comme le présente la figure III.3, une première étape consiste à reconstruire la distribution du signal d'intérêt à partir des modèles de bruit et d'une mesure du SNR de tous les circuits. Le recouvrement dans le domaine de Fourier est donné sur la figure III.7, la fréquence de coupure du filtre passe-bas  $y$  est également représentée. Cette fréquence de coupure est calculée à partir de la transformée de Fourier du signal

bruitée où l'on recherche à conserver seulement les fréquences pour lesquelles la transformée de Fourier du signal bruité est non-nulle.

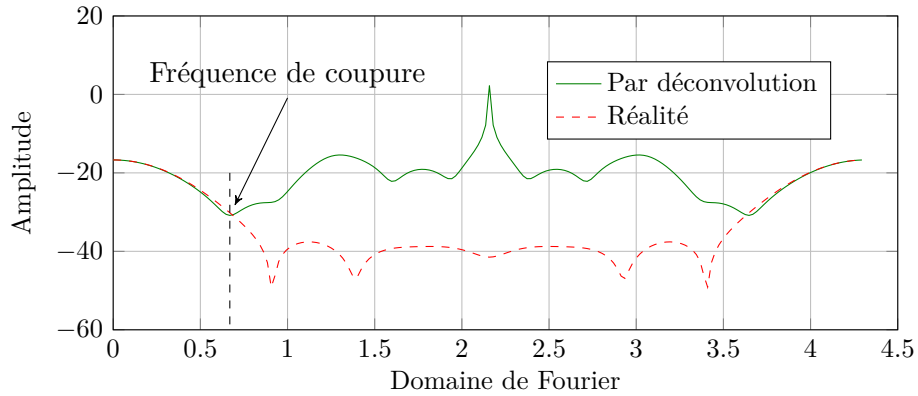


FIGURE III.7 – Transformée de Fourier de la distribution du signal reconstruit par déconvolution, avec sa représentation réelle

La transformée de Fourier inverse sur les basses fréquences conservées permet de retrouver la densité de probabilité du signal d'intérêt, représenté sur la figure III.8. Par la suite, la comparaison de la distribution recouverte et de la distribution mesurée bruitée permet d'évaluer les métriques de test.

La figure III.9 représente la baisse de rendement en fonction du nombre d'échantillons utilisés dans le calcul de la transformée de Fourier du SNR. De la même façon, les équations (III.12) et (III.14) permettent d'anticiper la probabilité de faux positifs et de faux négatifs, représentés sur la figure III.10 et la figure III.11.

Les courbes de faux positifs et faux négatifs montrent que l'évaluation des métriques réelles est entachée d'erreur. Cette erreur est induite par le nombre limité de circuits utilisés pour le calcul des métriques de test. L'évaluation de

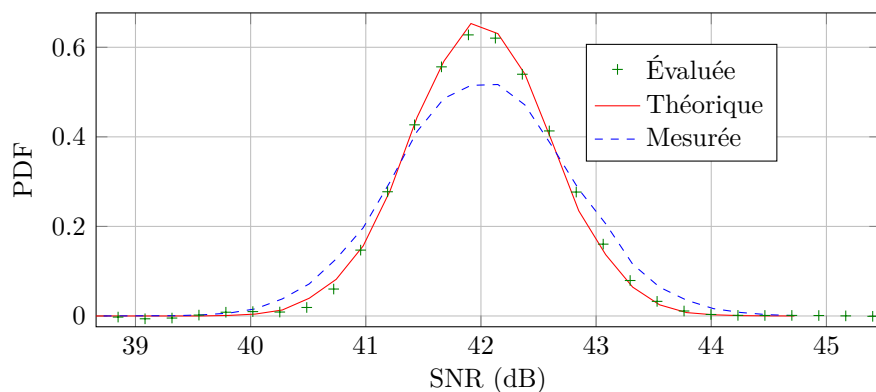


FIGURE III.8 – Comparaison des signaux reconstruit, d'origine et bruité

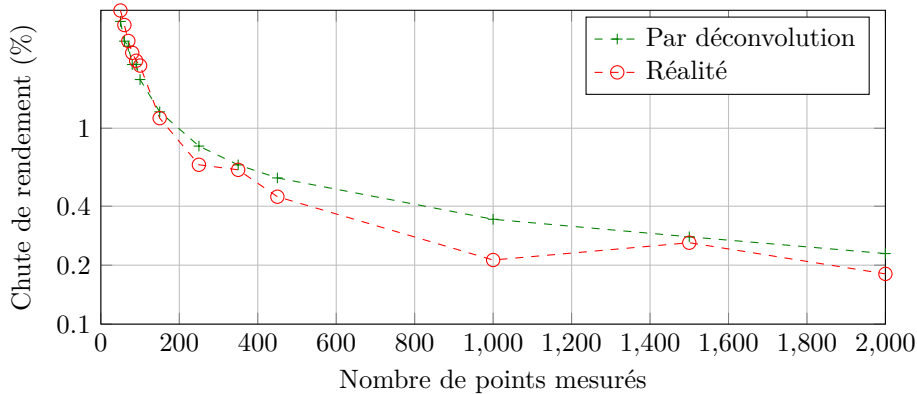


FIGURE III.9 – Écart de rendement en fonction du nombre de points conservés

cette erreur est dérivée de [Brockman and Director, 1989]. Elle a été présentée à la section II.B.1.a.

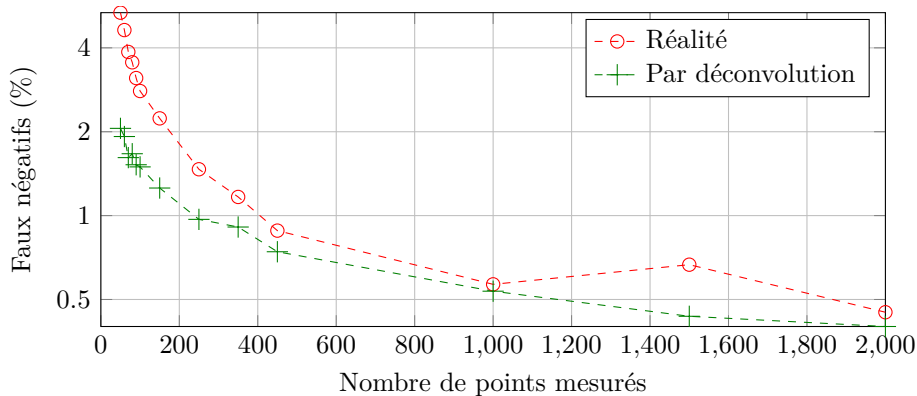


FIGURE III.10 – Évolution du nombre de faux négatifs suivant le modèle d'incertitude (III.14)

Les résultats obtenus pour cet exemple montrent qu'il est possible de diminuer le nombre d'échantillons utilisés pour la mesure de SNR, sans détériorer les métriques de test. Cette diminution du nombre de points de mesure permet une réduction du temps de test par circuit et donc une réduction du coût global du test. Pour ce faire, les modèles de bruit ont été obtenus par mesures sur un seul circuit. Ensuite, le SNR de tous les circuits n'a été mesuré qu'une seule fois.

### III.C.5 Conclusion sur le recouvrement du signal d'intérêt

L'intérêt de recouvrir la distribution des signaux est de pouvoir en déduire les métriques de test. En effet, ces métriques sont définies à partir de la distri-



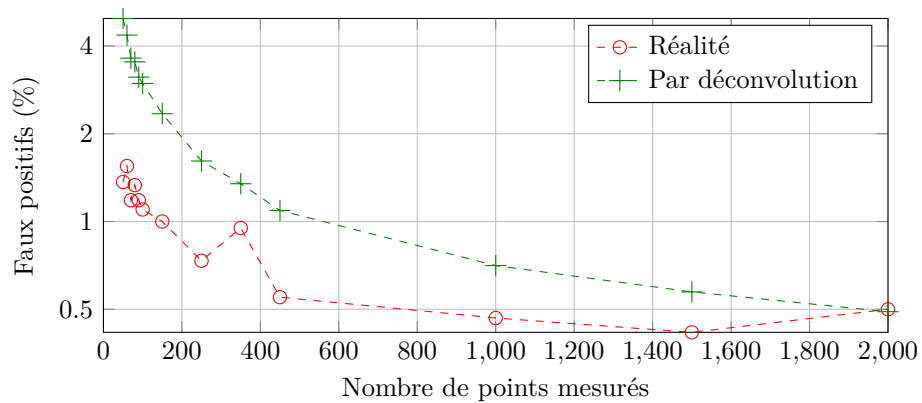


FIGURE III.11 – Évolution du nombre de faux positifs suivant le modèle d’incertitude (III.12)

bution des mesures et des spécifications.

La méthode de recouvrement de la distribution d’un signal bruité repose sur une simple déconvolution dans le domaine de Fourier. Cette déconvolution impose que le bruit soit additif et indépendant du signal d’intérêt. Par ailleurs, elle repose sur une division des représentations fréquentielles des densités de probabilité des mesures et du modèle de bruit. Un filtre passe-bas doit être utilisé pour reconstruire la densité de probabilité du signal d’intérêt, et ainsi contourner le problème de la sensibilité de la division dans le domaine de Fourier. Notons que pour éviter l’utilisation d’un filtre et le calcul de sa fréquence de coupure, il est possible d’utiliser une déconvolution régularisée [Neelamani et al., 2004].

Pour l’exemple proposé, nous réalisons une analyse de mesure de SNR où le bruit considéré est induit par la limitation du nombre de points pour la mesure de SNR. Le signal d’intérêt est le SNR des circuits. L’objectif est de diminuer ce nombre de points, donc de diminuer le temps de test, sans détériorer les métriques de test.

Dans le cadre du test de circuit, l’analyse des métriques de test telle qu’elle est proposée permet de diminuer les contraintes sur la mesure de manière contrôlée, en anticipant l’impact de cette diminution sur ces métriques de test. Dans le cadre de la méthode mise en œuvre, l’objectif est de fournir un indice quant au paramètre de sensibilité  $\delta$  utilisé dans CDC-FOCUS.

### III.D Induction du modèle

Dans la section précédente, nous nous sommes intéressés à la notion de sélection de caractéristiques, et donc à la sélection des mesures les plus pertinentes pour le test, et les moins chères. À présent, afin d'évaluer la pertinence des caractéristiques sélectionnées, nous utilisons l'étape d'induction de modèle pour construire un modèle de classification. Cette induction n'utilise que les caractéristiques sélectionnées par l'algorithme CDC-FOCUS pour construire le modèle de classification. Les métriques du modèle sont évaluées comme spécifié en section II.E.1.

Dans nos travaux, aucun inducteur n'a été développé spécifiquement pour réaliser la classification des circuits à l'aide des caractéristiques extraites par CDC-FOCUS. Ainsi, des inducteurs issus de l'état de l'Art du test de circuits sont utilisés, en particulier ceux présentés en section II.D, à savoir :

- k-Plus Proches Voisins ;
- Arbre de Classification ;
- Machine à Vecteurs de Support ;
- *Multivariate Adaptive Regression Splines* (Régression Multivariée par Splines Adaptative) MARS.

Ces algorithmes sont comparés sur la base de la matrice de confusion présentée en section II.E.2. A cet effet, nous appliquons, dans le chapitre suivant, ces algorithmes de classification à des exemples concrets de circuits. La construction des matrices de confusions pour chaque algorithmes de classification permet alors de les comparer. Nous retenons comme critères de comparaison la précision de la classification et le taux de faux positifs. Dans le cadre de nos travaux, les métriques de test permettent d'évaluer la qualité des observations sélectionnées précédemment par CDC-FOCUS, et elles évaluent les erreurs de classification. *In fine*, la qualité du modèle est déterminée d'une part par les métriques de test, et d'autre part par le coût des observations sélectionnées par CDC-FOCUS.

### III.E Synthèse, avantages et limites de la méthode proposée

La méthode proposée est une méthode en filtre, simple d'implémentation et dont les étapes sont spécifiées indépendamment les unes des autres. Elle s'articule autour de l'algorithme CDC-FOCUS développé à partir de l'algorithme C-FOCUS afin de prendre explicitement en compte le coût de la mesure dans la recherche des caractéristiques pertinentes. Par ailleurs, cet algorithme ne fait pas appel à un classifieur pour sélectionner les caractéristiques pertinentes,

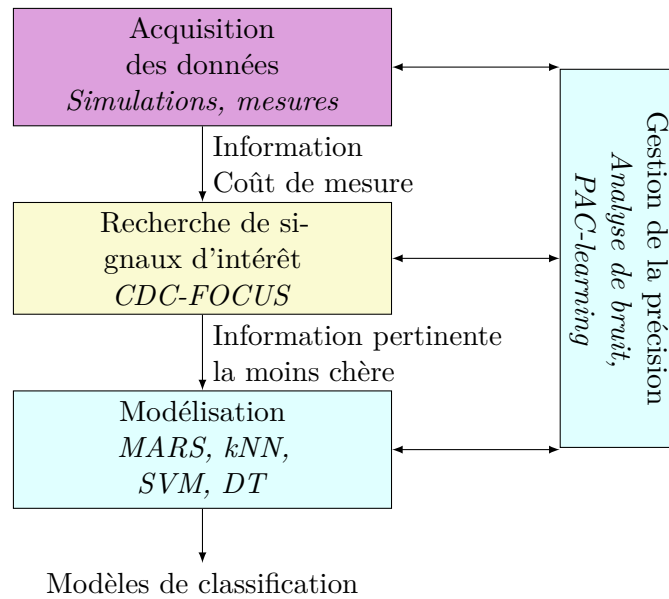


FIGURE III.12 – Les différents blocs de la Vision Unifiée (rappel)

comme le ferait une approche wrapper ou hybride.

CDC-FOCUS hérite de C-FOCUS la capacité à traiter des caractéristiques continues. Ceci implique l'introduction d'un seuil  $\delta$  utilisé pour discriminer des valeurs de caractéristiques "semblables" et des valeurs "différentes". L'évaluation de ce seuil est complexe et impacte les performances de l'algorithme, comme on le verra au chapitre IV. Dans nos travaux, l'analyse du bruit proposée permet de fixer une limite basse à  $\delta$ .

Les algorithmes de classification utilisés pour construire le modèle de classification sont issus de l'état de l'Art. Cependant, la théorie permettant d'exprimer leur efficacité est délicate à mettre en œuvre (elle demande une étude approfondie des algorithmes, notamment de l'espace des hypothèses, comme proposé dans section II.E.1). Néanmoins, ils permettent d'anticiper les erreurs de précision et de faux-positifs, et ainsi de connaître la qualité des modèles construits avant leur mise en œuvre.

Notre méthode de conception d'une stratégie de test à coût réduit est construite de façon à s'appliquer indépendamment de la technologie utilisée pour le circuit. De plus, la base de données peut être construite soit en simulation, soit à partir d'une campagne de mesures.

Cette double origine des données initiales utilisées permet de mettre en œuvre la méthode proposée dans différents cas typiques. Elle permet, en simu-

lation, d'anticiper les caractéristiques à mesurer de manière à assurer le test. La méthode assiste alors le concepteur pour la mise en place d'une stratégie de test du circuit. D'autre part, elle peut être utilisée au moment du test, sans connaissance de la conception, soit pour extraire les mesures les plus pertinentes pour le test si suffisamment d'échantillons sont disponibles, soit pour assurer que les mesures réalisées sont suffisantes pour expliquer la classification des circuits.

Dans le chapitre suivant, la méthode présentée ici va être mise en œuvre dans différents cas d'utilisation.



## Chapitre IV

---

# Validation de la méthode

---

Trois critères permettent de considérer une affirmation comme valide : la vérification par l'expérience directe, la déduction irréfutable, et le témoignage digne de confiance.

---

Jean-François REVEL in *Le Moine et le Philosophe*

LA VALIDATION de la démarche proposée au chapitre précédent est maintenant présentée au travers d'exemples d'utilisation. L'objectif est de montrer que la méthode proposée est utilisable dans différents cas d'utilisation.

Pour valider la méthode, les métriques utilisées sont les suivantes :

- le coût de test ;
- la précision des modèles de classification construits ;
- le taux de faux positifs constaté pour les modèles de classification construits.

La première partie de ce chapitre introduit des outils et méthodes pratiques pour l'acquisition de base de données utilisables par notre méthode. Ces outils permettent d'utiliser des bases de données incomplètes. La deuxième section présente les résultats obtenus par simulations. L'objectif est de remplacer des mesures de performance par des mesures alternatives. Enfin, la troisième section utilise des données issues de mesures pour valider la méthode présentée. Ce cas est particulier car nous ne disposons pas de connaissance sur les mesures (ce qu'elles observent, comment elles sont reliées entre elles) et la base de données contruite peut présenter des mesures manquantes ou

redondantes.

L'intérêt de proposer plusieurs cas d'utilisation est de montrer la généralité de la méthode permise par le flot de traitement de données dont les résultats peuvent être exploités pour des cas d'utilisation différents.

## IV.A Acquisition des bases de données

Les bases de données ne sont pas directement accessibles et leur obtention implique la mise en place de processus de traitement que nous allons détailler ici. En premier lieu, l'acquisition par simulations consiste à déterminer les caractéristiques à extraire, à les simuler, et à organiser la grande quantité de données ainsi générée. Le cas de l'acquisition en mesure, traité en second lieu, est différent. Nous ne reviendrons pas sur la technique de la mesure, mais nous analyserons le traitement de la base de données obtenue par ce moyen. Celle-ci est incomplète du fait de l'approche "*Stop on Fail*" qui est l'arrêt de la mesure lorsqu'un test échoue (ceci pour réduire le temps du test). Nous verrons comment valider la méthode proposée même dans ce cas particulier.

### IV.A.1 Acquisition de bases de données par simulation

Pour l'acquisition de données par simulation, nous supposons que nous avons à disposition un circuit défini par sa netlist et ses spécifications, une technologie et un simulateur électrique capable de réaliser des simulations de type Monte Carlo en respectant la distribution technologique. L'objectif de la simulation est de construire une base de données contenant des mesures alternatives des circuits et une classification de ceux-ci entre "fonctionnels" et "défectueux". Ainsi, nous devons générer des instances du circuit dont le simulateur extraiera les performances et des mesures alternatives. Compte tenu de la taille des bases de données, de nombreuses simulations sont requises.

La définition des mesures pour les spécifications doit être réalisée avec l'aide du concepteur du circuit, qui aura souvent établi des bancs de tests montrant que le circuit répond aux spécifications. La détermination des mesures alternatives peut être automatisée par la lecture de la netlist du circuit, et la mesure systématique des caractéristiques des composants électriques ou des nœuds définis dans cette netlist. Les caractéristiques électriques des composants peuvent être décrites dans la documentation du simulateur utilisé. Pour nos travaux, nous avons écrit un script en langage python pour lire la netlist et créer un fichier de commandes compréhensibles par le simulateur permettant d'extraire les mesures alternatives d'un circuit. Par la suite, le simulateur utilise la netlist, le fichier de commande et la définition technologique pour générer le fichier des résultats. Comme nos simulations sont de type Monte Carlo, nous indiquons de plus au simulateur l'"initialisation

aléatoire” à utiliser. La simulation d’un circuit est schématisée sur la figure IV.1.

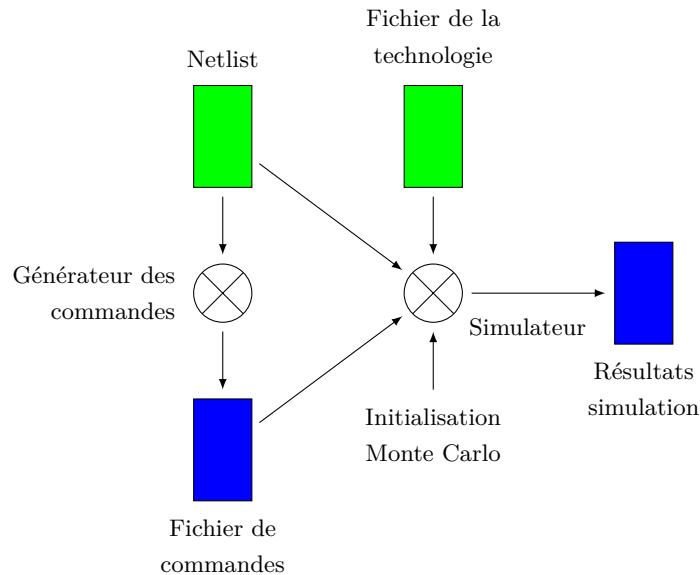


FIGURE IV.1 – Diagramme de génération des simulations. Les rectangles verts sont les fichiers d’origines, les bleus sont les fichiers générés par les outils.

L’inconvénient de cette technique est que le nombre de commandes d’extraction et d’enregistrement de grandeurs électriques est important. Malgré le soin apporté dans la réalisation des logiciels de simulation, il est possible que le fichier de résultats généré soit corrompu sans qu’aucun message ne le signale. Cette corruption apparaît lorsqu’un trop grand nombre de commandes d’écriture de fichier sont utilisées dans une simulation. Il est donc impératif, avant de démarrer une campagne de simulation, de s’assurer que le simulateur est bien capable d’écrire la grande quantité de résultats sans corrompre le fichier. Sinon, il faut diminuer le nombre de caractéristiques à extraire.

Pour établir la base de données, il est nécessaire d’instancier et de simuler une grande quantité de circuits. La simulation d’un circuit est trop longue pour pouvoir simuler les instances les unes après les autres. Nous avons à notre disposition un ensemble d’ordinateurs géré par un ordonnanceur de tâche. Nous avons exploité cet outil pour paralléliser les simulations électriques. Deux problèmes interviennent alors : la limite des licences logicielles disponibles et la limite des capacités de calcul. Concernant les simulations électriques, la contrainte est avant tout une contrainte sur les licences. Ainsi, afin de ne pas bloquer les activités du laboratoire, nous avons développé une sur-couche (bash) à l’ordonnanceur de tâches afin de limiter l’utilisation des licences pendant les heures de bureau et de les exploiter au maximum pendant les périodes



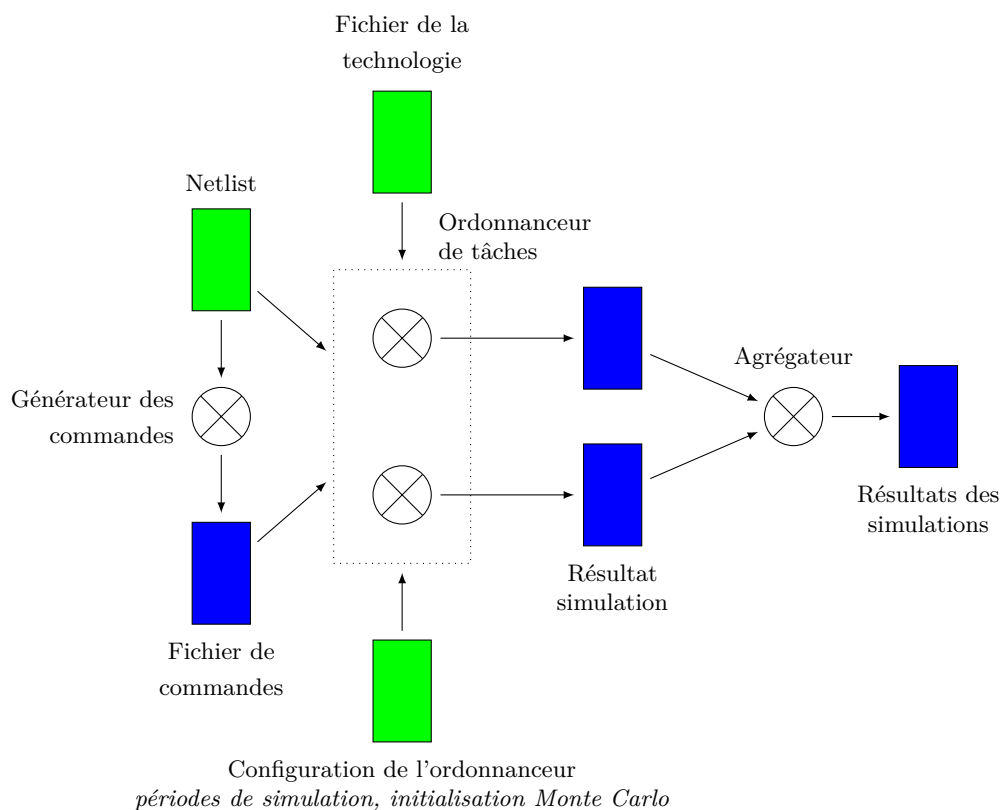


FIGURE IV.2 – Diagramme des simulations parallèles. Les rectangles verts sont les fichiers d'origines, les bleus sont les fichiers générés par les outils.

en-dehors des horaires de travail. Cette démarche exige tout d'abord la gestion des erreurs afin d'éviter que les simulations ne se bloquent et, de plus, puisque ces dernières ont lieu sur des machines différentes, une stratégie d'agrégation de l'état des simulations est nécessaire. La figure IV.2 récapitule la solution de simulation utilisée. Cet outil de simulation parallèle a aussi été utilisé pour faire fonctionner le flot de la méthode propagée dans nos travaux.

Cette stratégie permet de rapidement obtenir les bases de données de simulation avec plusieurs milliers d'instances du circuit simulé. Le problème suivant est le traitement des données ainsi obtenues. Nous avons essentiellement utilisé Matlab pour le traitement des données. Cet outil atteint ses limites avec des tableaux de données de plusieurs centaines de lignes et plusieurs milliers de colonnes (pour des nombres flottants). Cependant, la machine virtuelle Java qui héberge l'exécution de Matlab peut être paramétrée pour accepter des tableaux de taille plus grande. De plus, lors du développement des fonctions de traitement, il convient de soigner le passage des arguments, afin de ne pas réaliser de copies de tableau inutiles. Enfin Matlab propose des bibliothèques en langage C permettant d'intégrer du code C dans ses scripts après une étape de

compilation<sup>1</sup>. Cette technique permet d'accélérer l'exécution d'un script.

Le déploiement de ces stratégies de développement demande une maîtrise de l'environnement Matlab et des outils de développement. On retiendra en particulier la commande `profile` pour analyser l'exécution du code et détecter les fonctions les plus coûteuses en temps de calcul. Une fonctionnalité non documentée est `profile -memory on`<sup>2</sup>, qui permet d'analyser l'occupation mémoire des différentes fonctions d'un script. De plus, il conviendra de maîtriser le développement en langage C dans le *framework* Matlab, ainsi que son débogage<sup>3</sup>. Le gain en vitesse d'exécution des fonctionnalités implémentées en langage C est représenté en figure IV.3. Cette figure nous montre le temps d'exécution d'une fonction (ici le test de suffisance de CDC-FOCUS) en Matlab et en C wrappé<sup>4</sup> selon différents arguments d'appel. En moyenne, l'implémentation en C wrappée est 400 fois plus rapide que son équivalent Matlab.

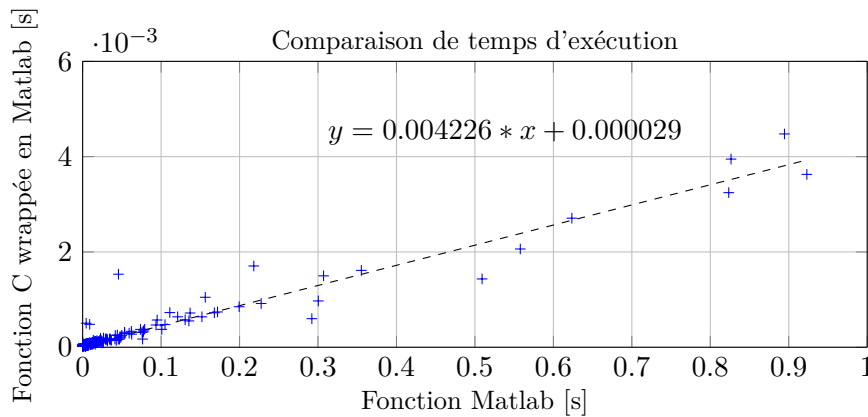


FIGURE IV.3 – Gain en vitesse d'exécution pour une fonctionnalité implémentée et wrappée en langage C plutôt qu'en Matlab

Les contraintes de l'acquisition de données en simulation sont donc celles de l'implémentation informatique des simulations et des traitements de données. Les solutions apportées ici sont les simulations parallèles automatisées et des implémentations exploitant le langage C et le langage Matlab. Dans la suite, nous allons discuter de l'acquisition de bases de données par mesures, que la méthode proposée adresse également.

---

1. Le code compilé est alors appelé {em Mexfile  
 2. <http://undocumentedmatlab.com/blog/undocumented-profiler-options>  
 3. [http://fr.mathworks.com/help/matlab/matlab\\_external/debugging-on-linux-platforms.html](http://fr.mathworks.com/help/matlab/matlab_external/debugging-on-linux-platforms.html)  
 4. C'est-à-dire codée en langage C dans le *framework* Matlab

## IV.A.2 Acquisition de bases de données par mesure

Pour utiliser correctement le flot de données présenté au chapitre III, certaines précautions doivent être observées. Les données obtenues par mesures peuvent en effet être incomplètes (mesures en *Stop on Fail*), et des mesures peuvent être fortement corrélées, voire identiques. Cette section fournit des recommandations sur la construction de la base de données dans le cas de l'acquisition d'informations par la mesure.

### IV.A.2.a Cas des mesures manquantes

Le cas de l'acquisition de données par la mesure est particulier, comme expliqué en section section II.B.2. L'inconvénient traité ici est le comportement *Stop On Fail*, qui permet de stopper le test d'un circuit dès qu'une spécification est violée, réduisant ainsi le temps de test. Cette technique ne permet cependant pas de construire une base de données complète, car des mesures sont alors manquantes pour les circuits dont une mesure (au moins) est en-dehors des spécifications.

La figure IV.4 est une représentation graphique issue d'une base de données contenant tous les tests pour tous les circuits. Les pixels bleus représentent des valeurs mesurées, et les pixels blancs des valeurs manquantes, aussi dites NaN (*Not a Number*). Les lignes blanches horizontales sont en fait la représentation graphique du comportement *Stop on Fail* qui arrête la séquence de test d'un circuit dès qu'une mesure est en-dehors des spécifications. Par ailleurs, la figure IV.4 comporte aussi des lignes verticales blanches. Celles-ci représentent des mesures qui n'ont jamais été effectuées ou des valeurs qui ne sont pas représentées sous forme numérique.

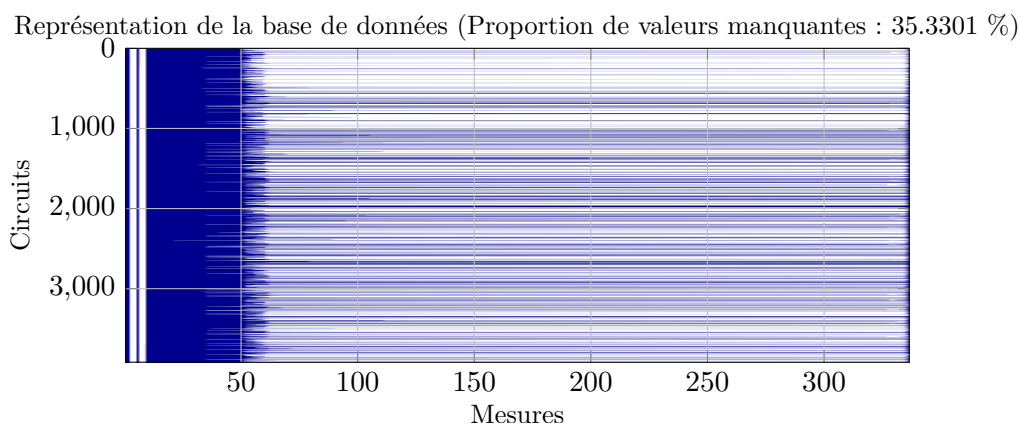


FIGURE IV.4 – Représentation graphique des mesures de test

Cette base de données est inexploitable en l'état par la méthode que nous

avons proposée car cette dernière ne peut pas traiter les données manquantes (NaN). La première façon de contourner ce problème est de réaliser une nouvelle campagne de mesures évitant le comportement *Stop on Fail*. Si ce n'est pas possible, on peut modifier la base de données pour obtenir une nouvelle base exploitable : l'idée est de ne conserver que les circuits dont "suffisamment" de mesures ont été réalisées et d'analyser la base de données seulement sur les mesures conservées. Ces modifications amènent à la figure IV.5 où 52 mesures sont conservées.

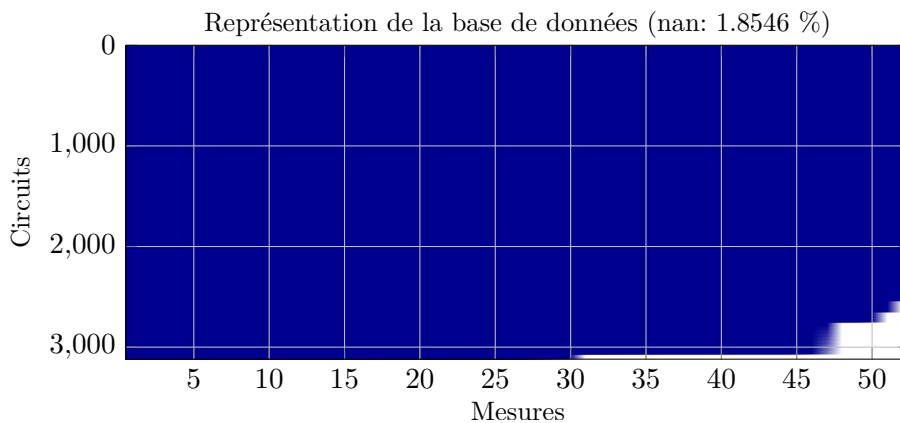


FIGURE IV.5 – Représentation graphique d'une base de données reconstruite à partir de mesures

La base de données contient alors nettement moins de séquences de test incomplètes, mais le nombre de dimensions et d'échantillons disponibles est également nettement réduit. Les conclusions issues de l'analyse de cette base de données reconstruite ne sont alors valables que pour cette dernière, et non pas pour la base de donnée d'origine. En d'autres termes, la séquence de test pertinente au coût le plus faible extrait dans la base de donnée reconstruite ne sera pas nécessairement la même que celle de la base de donnée d'origine.

On peut étudier le cas particulier des bases de données avec des valeurs manquantes sous l'angle de la reconstruction de données. Par exemple, [Parthasarathy and Aggarwal, 2003] propose de reconstruire une base de données avec des valeurs manquantes à partir d'une analyse en composantes principales (PCA). Dans le cas particulier des données issues de mesures *Stop on Fail*, les données manquantes sont systématiquement des données de circuits défaillants. Elles correspondent aux mesures réalisées en fin de séquence de test. Par ailleurs, la méthode proposée par [Parthasarathy and Aggarwal, 2003] utilisant PCA (voir section II.C.1.a), elle, exige un noyau déterminant les relations qu'il est possible de reconstruire. Nous n'utilisons pas de reconstruction de données ici.

#### IV.A.2.b Réduction *a priori* de la dimension des bases de données

Dans les bases de données de mesures, il est possible d'avoir des mesures dont les valeurs sont systématiquement identiques, ou fortement corrélées linéairement. Ces dimensions peuvent être détectées avant la mise en place de la méthode proposée. Par ailleurs, le temps d'exécution de l'algorithme CDC-FOCUS augmente avec le nombre de mesures dans la base de données. Il est donc utile de réduire la dimension de la base de données en utilisant des connaissances *a priori* lorsque cela est possible. L'insertion d'une étape d'analyse de corrélation avant la sélection de caractéristiques va permettre cela.

La figure IV.6 propose une stratégie en filtre pour réduire *a priori* le nombre de caractéristiques dans la base de données par analyse de corrélation comme analyse préliminaire à la sélection de caractéristiques. Des groupes de corrélation sont construits et la caractéristique la moins chère de chaque groupe est conservée.

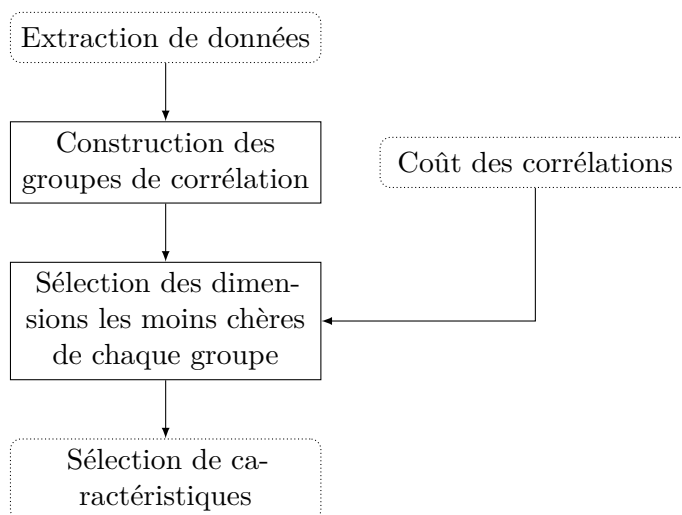


FIGURE IV.6 – Diminution de la dimension du problème par étude des corrélations

L'utilisation du critère de corrélation permet de diminuer sensiblement le nombre de caractéristiques dans la base de données. À titre d'exemple, la base de données réduite représentée en figure IV.5 contenant 52 caractéristiques peut encore être réduite à 41 caractéristiques par le critère de corrélation.

Les coûts des mesures sont, quant à eux, utilisés de manière relative les uns par rapport aux autres : plus une dimension aura un coût relatif élevé, plus il faudra de caractéristiques pour la remplacer. Dans le cas des bases de données issues de mesures, l'absence de connaissance sur le circuit rend

souvent impossible le paramétrage correct des coûts. En effet, les données issues des mesures peuvent avoir été anonymisées de sorte qu'il n'y ait pas de connaissances sur le circuit au travers des mesures. Pour contourner cette difficulté, nous avons fait le choix de définir les coûts selon la nature de la mesure réalisée. Par exemple, les mesures de fréquence ont un coût "élevé"(5), alors que les mesures de tension ont un coût "faible"(1).

## IV.B Validation avec des données de simulation

Cette section valide la méthode présentée au chapitre III sur des données obtenues en simulation pour deux circuits différents. Dans ces deux cas, il s'agit de détecter l'ensemble de caractéristiques pertinentes (au sens de la section III.B.1) extraites par CDC-FOCUS, puis de construire un modèle de classification pour les circuits. Il s'agit également d'éliminer les observations à coût élevé afin de les remplacer par des mesures alternatives dans le cas du LNA traité en section IV.B.1. On cherche également à supprimer une mesure de performance coûteuse au profit d'un modèle conforme au respect des spécifications du circuit. Ce dernier cas concerne un convertisseur analogique numérique en temps continu et il est plus complexe car il n'existe pas de modèle analytique permettant d'expliquer son comportement.

### IV.B.1 Utilisation de la méthode avec un circuit LNA

Dans cet exemple, un LNA est simulé pour générer une base de données, utilisée ensuite en entrée de la méthode proposée au chapitre précédent, afin d'obtenir *in fine* un modèle de classification. L'intérêt du LNA est qu'il ne s'agit que d'un bloc d'une chaîne de réception. Il représente donc un exemple suffisamment simple pour un premier exemple d'utilisation de notre méthode. Cependant, nous utilisons ici un LNA différentiel à deux niveaux de fonctionnement. Cela permet de complexifier ce bloc et d'ajouter des mesures supplémentaires pour valider son comportement.

La conception du LNA n'a pas été abordée dans nos travaux. Cette conception est telle que les performances du LNA soient respectées lors d'une fabrication idéale du circuit. Cependant, une déviation des paramètres technologiques peut entraîner la réalisation d'un circuit défaillant. Le LNA est donc simulé en utilisant une simulation de type Monte-Carlo, où les paramètres technologiques suivent une dispersion induite par les modèles statistiques de la technologie.

Par ailleurs, la simulation est réalisée au niveau de la netlist idéale. Elle n'inclut pas l'extraction de *layout*. Seules les erreurs de déviation technologique sont prises en compte.

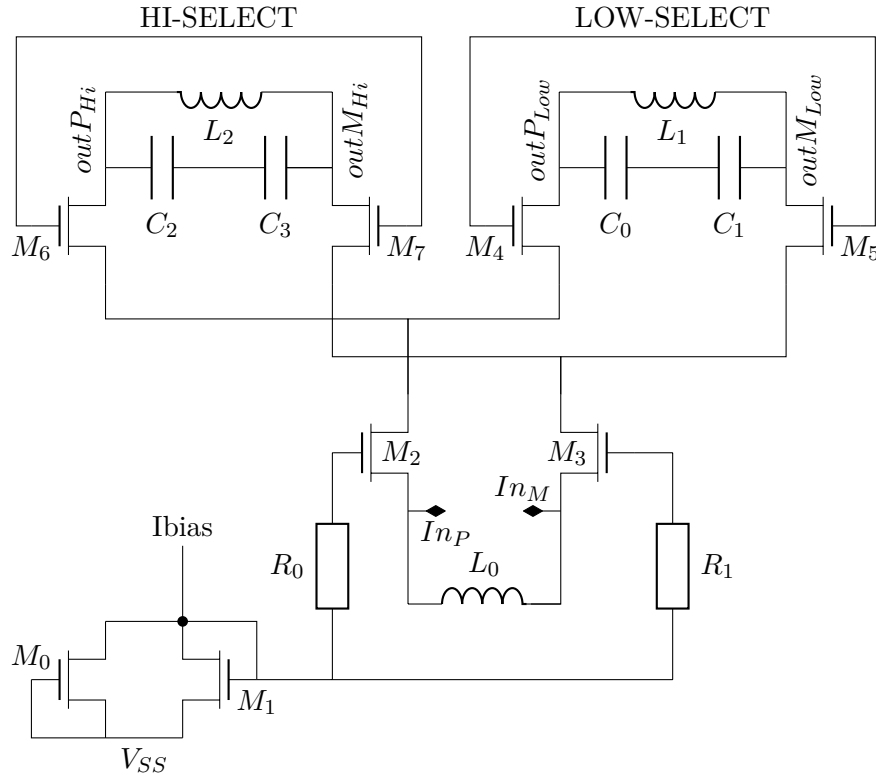


FIGURE IV.7 – Schéma du LNA

#### IV.B.1.a Présentation du LNA

Le LNA utilisé pour valider nos travaux est un LNA différentiel à deux niveaux de fonctionnement, *HI* et *LOW*, en technologie 130nm. Ces deux niveaux de fonctionnement permettent d'amplifier des signaux correspondant à deux fréquences différentes. Son schéma est présenté figure IV.7, et il est utilisé par [Ouvry et al., 2014].

Chacun des niveaux de fonctionnement peut recevoir des signaux sur trois bandes de fréquence. Les performances considérées sont les gains à fréquence de fonctionnement, les gains maximaux et leur fréquence associée, les bandes passantes, les coefficients de réflexion en entrée adaptée ( $S_{1,1}$ ), les figures de bruit et les points de compression.

Le test de ce circuit est long car il exige de valider les mêmes performances sur les trois bandes de fréquence et les deux niveaux de fonctionnement. La validation traditionnelle de ce circuit consiste donc à mettre le LNA dans l'état de fonctionnement *LOW*, à paramétrer l'outil de test pour transmettre et recevoir des signaux de l'état *LOW*, à envoyer les signaux de test et à

analyser les signaux de sortie, et ceci pour les trois bandes de fréquences. Puis, après validation de l'état *LOW*, il est nécessaire de mettre le LNA en état de fonctionnement *HI*, de paramétrer l'outil de test pour transmettre et recevoir des signaux de l'état *HI*, et enfin de transmettre et d'analyser les signaux reçus. La redondance des opérations de test qui peut alors apparaître doit être évitée car elle est coûteuse en temps. Ainsi, la méthode présentée est utilisée pour éliminer ces redondances.

#### IV.B.1.b Procédure d'application de la méthode proposée sur le LNA

La procédure utilisée consiste à simuler par Monte-Carlo des échantillons du LNA pour en extraire 58 observations (48 performances et 10 mesures alternatives). Dans nos travaux, ce sont 5000 simulations que nous avons utilisé. Les observations extraites des simulations sont des performances et des mesures alternatives. Les performances couvrent les gains, le bruit et les non-linéarités pour tous les modes de fonctionnement du circuit (chacune des 3 bandes de fréquence des 2 niveaux de fonctionnement). Les mesures alternatives, elles, contiennent des mesures de tension source-grille et d'inductance.

On associe ensuite un coût à chacune des mesures. Cela permet, comme précisé en section III.B.1, de calculer le coût d'un ensemble de mesures par la somme des coûts de chacune des mesures le constituant. Le coût associé aux mesures de performances est de 10, et celui associé aux mesures alternatives est de 1. Les spécifications sont définies sur un ensemble de 48 performances. De cette façon, la mesure de toutes les performances permettant de valider les spécifications a un coût de 480. Par ailleurs, les spécifications définies conduisent à un rendement de 96.16% sur la simulation Monte Carlo.

Ensuite, la base de données obtenue est utilisée en entrée de CDC-FOCUS. Dans ce cas d'utilisation, le seuil de CDC-FOCUS n'est pas fixé *a priori*, mais plusieurs valeurs du seuil sont utilisées (près de 3000 valeurs différentes seront utilisées pour cet exemple). Cette approche "brute force" permet d'apprécier l'impact du seuil sur l'approche proposée. Enfin, pour chaque valeur de seuil de CDC-FOCUS, quatre modèles de classification sont construits, avec les approches MARS, kNN, SVM et Arbre de Décision (voir section II.D) respectivement. Nous utilisons les caractéristiques extraites par CDC-FOCUS. Ainsi, ce sont  $2982 \times 4 = 11928$  modèles qui sont construits à partir de 75% des échantillons disponibles. Les matrices de confusion de chaque modèle sont établies en utilisant les 25% d'échantillons restants. Ces matrices permettent de tracer les résultats obtenus, à savoir les précisions et faux-positifs de classification. Enfin, la réduction du coût est calculée directement par la somme des



coûts des caractéristiques extraites par CDC-FOCUS. La démarche globale est représentée sur la figure IV.8.

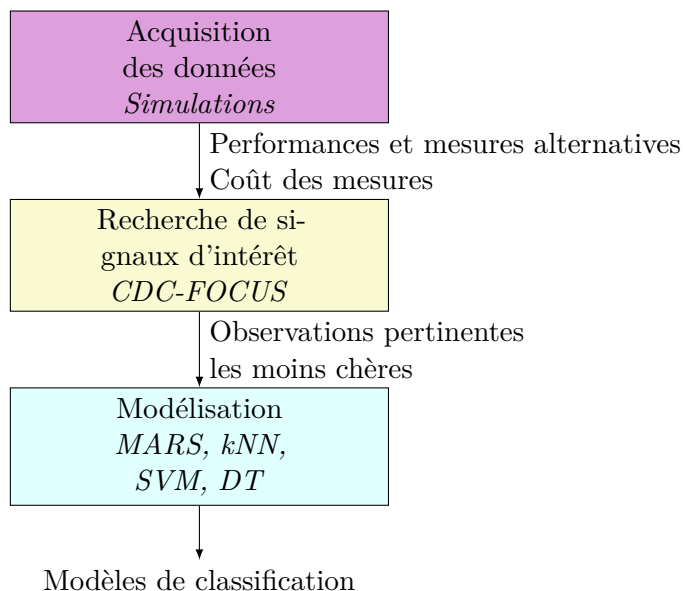


FIGURE IV.8 – Algorithmes utilisés pour traiter l'exemple du LNA

#### IV.B.1.c Résultats de l'application de la méthode sur les simulations du LNA avec le flot proposé

Les métriques d'intérêt sont représentées en figure IV.9 page suivante. Sur cette figure, par souci de lisibilité, seulement une partie des modèles construits sont représentés. Cette figure indique aussi le coût associé aux caractéristiques extraites par CDC-FOCUS, qui sont utilisées pour obtenir le modèle considéré.

Un premier constat est le positionnement des modèles selon les métriques de précision et de faux positifs. Pour ce problème de classification du LNA, les performances des classifieurs sont groupées par zone. Ainsi, SVM est le classifieur qui donne la meilleure précision, et MARS est celui qui donne le meilleur compromis entre la précision et les faux positifs. Dans nos résultats, SVM a été paramétré pour utiliser un noyau de type fonction à base radiale qui permet d'efficacement classer des échantillons d'une classe précise, ici les circuits sains, ce qui explique la précision des résultats obtenus. MARS propose un modèle de régression linéaire par morceaux qui permet de régresser les performances à partir des mesures alternatives sélectionnées par CDC-FOCUS. La classification est ensuite opérée sur les régressions. La performance de la régression MARS permet alors d'obtenir les résultats présentés.

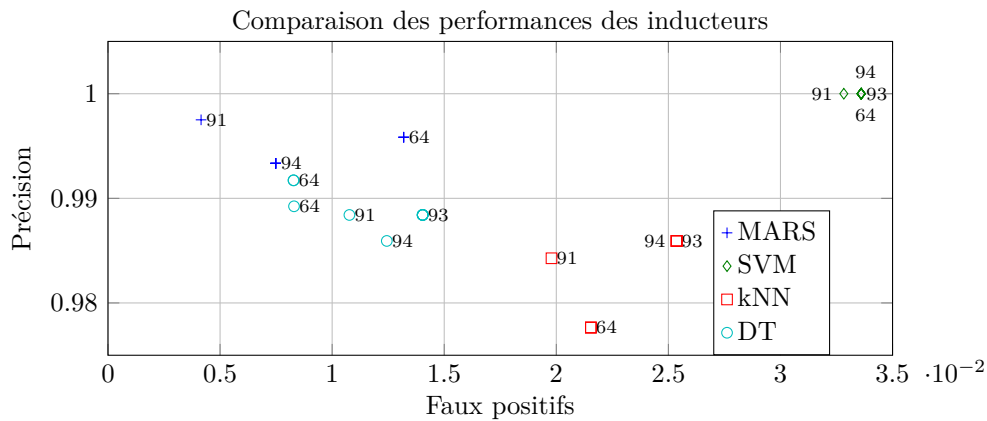


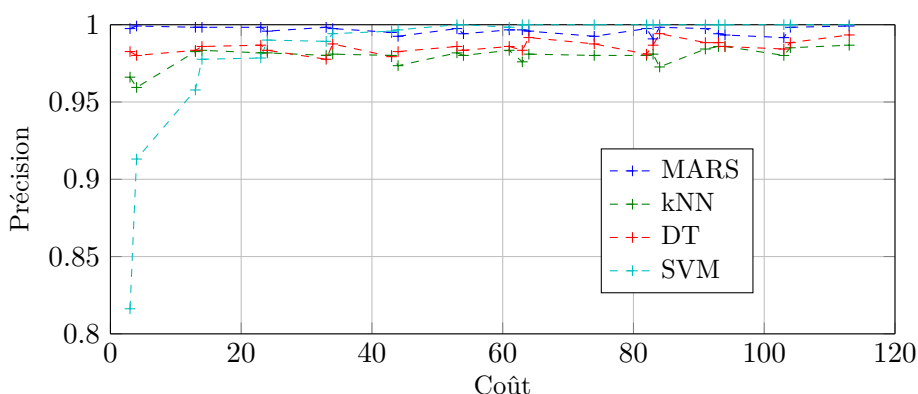
FIGURE IV.9 – Comparaison des inducteurs, selon la précision et les faux-positifs

Une bonne précision ne signifie pas que le nombre de faux positifs est faible. Ce raccourci rapide est faux en raison du fort déséquilibre des classes représentées. Les figures IV.10a et IV.10b synthétisent les résultats obtenus pour montrer l'évolution des métriques de test en fonction du coût total des mesures associées. Cette synthèse est rendue possible par la modification du seuil de CDC-FOCUS. Pour ce cas d'utilisation, les algorithmes atteignent une précision élevée très rapidement, pour un coût faible (20). Rappelons que le coût maximal est de 480.

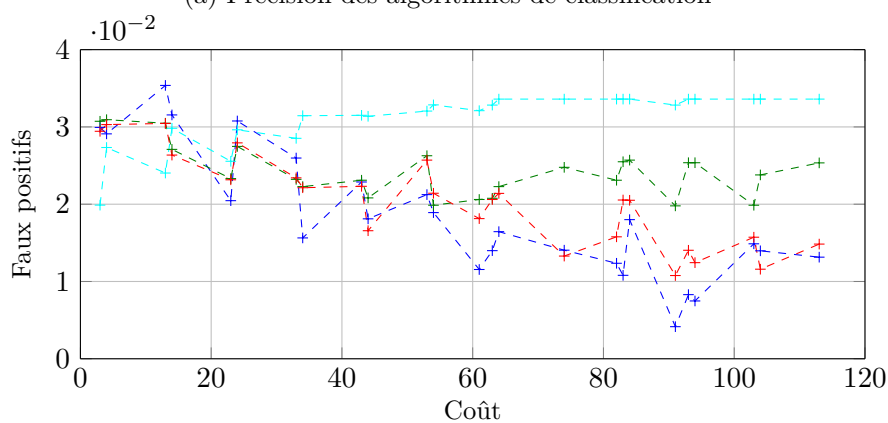
Cette réduction importante du coût est nettement moins marquée pour les faux positifs, et en particulier pour les algorithmes SVM et kNN, qui ne semblent pas voir leur taux de faux positifs diminuer avec l'augmentation du coût. Ces algorithmes sont en effet fortement influencés par la composition de l'ensemble d'entraînement et le déséquilibre des classes dégrade leurs métriques. *A contrario*, MARS conserve de "bonnes" métriques. Cet algorithme est en effet un algorithme de régression, à la différence des trois autres qui sont des algorithmes de classification. De ce fait, il n'est pas sensible au déséquilibre des classes.

Pour apprécier les métriques de test obtenues, le tableau IV.1 fournit les métriques de test des différents algorithmes de classification lorsque toutes les performances spécifiées sont utilisées. Dans ce cas, toute l'information pertinente est disponible et aucune information non pertinente n'est présente.

Ces résultats montrent que les classificateurs, bien qu'ayant été entraînés sur toutes les informations pertinentes, n'atteignent pas des métriques de test particulièrement "élevées" comparé à celles obtenues suite à l'utilisation de



(a) Précision des algorithmes de classification



(b) Faux positifs des algorithmes de classification

FIGURE IV.10 – Précision et faux positifs des algorithmes de classification suivant le coût total des observations retenues par CDC-FOCUS

CDC-FOCUS. Ainsi, fournir toutes les mesures pertinentes pour l'entraînement d'un algorithme de classification n'est pas suffisant, il est aussi nécessaire de les fournir avec peu (ou pas) de mesures non pertinentes. CDC-FOCUS, en plus de répondre au problème de réduction des coûts, traite aussi de la réduction de la dimension de l'espace d'entraînement des algorithmes, ce qui améliore leurs métriques.

Ce premier exemple du LNA montre aussi l'impact du choix du seuil de CDC-FOCUS. Ce seuil nous permet de faire varier la sélection des caractéristiques pour adresser des coûts différents et des métriques de test différentes, comme présenté sur la figure IV.10. Le choix d'une valeur de seuil particulière a été traité dans la section III.C, où nous avons choisi de relier le choix de ce seuil aux métriques de test par l'intermédiaire d'une étude de bruit. En effet, le seuil peut être interprété comme une limite d'incertitude sur les mesures. Nous

Classifieur	Précision	Taux de faux positifs
MARS	N'a pas convergé	
kNN	0.9640	0.6190
DT	0.9936	0.0714
SVM	0.9664	1

TABLE IV.1 – Métriques de test lorsque toutes les performances spécifiées sont utilisées

pouvons aussi interpréter différemment le seuil comme une incertitude, non pas sur les valeurs mesurées, mais sur la spécification. En effet, du point de vue de la base de données, la spécification d'une caractéristique est la limite entre les échantillons de la classe "défectueux" et de la classe "fonctionnels". Plus précisément, elle est située entre les deux échantillons les plus proches de la classe "défectueux" et de la classe "fonctionnelle". Cette considération, plutôt que de relier le choix du seuil aux métriques de test, le relie au contenu de la base de donnée. C'est-à-dire que nous pouvons fixer une limite au seuil au-dessus de laquelle nous ne pouvons pas obtenir plus d'informations. La figure IV.11 présente la détermination des seuils maximaux pour deux caractéristiques.

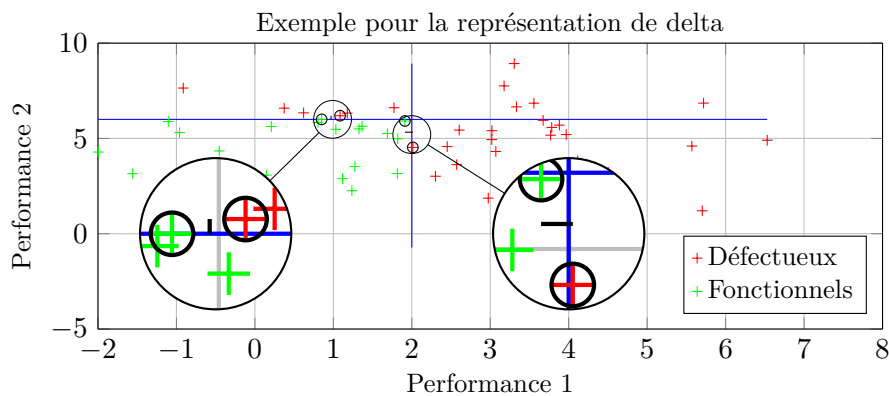


FIGURE IV.11 – Exemple de détermination du seuil maximal pour deux caractéristiques. Les croix vertes sont les échantillons "fonctionnels", les rouges sont ceux "défectueux". Les spécifications sont les lignes bleues. La plus petite distance entre deux échantillons de classes différentes autour des spécifications est matérialisée par un trait noir dans les loupes.

L'exemple nous montre qu'une limite de seuil peut être établie pour chaque caractéristique. Or, dans le cas d'utilisation de CDC-FOCUS, nous ne mentionnons qu'un même seuil s'appliquant identiquement sur toutes les caractéristiques. Or, dans notre exemple du LNA, chaque caractéristique exige une limite

de seuil différente. L'algorithme CDC-FOCUS devrait donc être paramétré autrement, en particulier en prenant en compte un seuil différent pour chaque caractéristique.

#### IV.B.1.d Conclusion quant à l'utilisation de la méthode sur le LNA

L'exécution de la méthode proposée sur les données issues de simulation du LNA a montré qu'il est possible de fortement réduire le coût de test par l'utilisation de l'algorithme CDC-FOCUS. Nous pouvons, à partir des résultats obtenus, proposer un modèle MARS ayant une précision de 99.75%, un taux de faux-positifs à 0.4% pour un coût égale à 18.96% du coût d'origine.

Le chapitre III a présenté différents classifieurs qui ont été mis en œuvre dans cet exemple. Il en ressort que MARS fournit de "bonnes" métriques de test, au contraire de SVM. L'avantage de MARS est de réaliser une régression des performances, qui sont ensuite comparées aux spécifications. MARS adresse nativement les problèmes de régression non linéaires et est adapté à la modélisation des performances de circuits analogiques, ce qui peut expliquer de meilleures performances par rapport aux autres classifieurs.

La méthode est utilisée ici de façon à montrer la relation qu'il est possible de déterminer entre le coût des observations réalisées et les métriques de test. Cette capacité est possible grâce à la méthode "brute force" consistant à parcourir plusieurs valeurs du seuil de CDC-FOCUS. Ici, augmenter ce seuil permet d'atteindre de meilleures métriques mais à des coûts plus élevés. Le choix *a priori* du seuil peut être dicté par l'analyse de la base de données. Nous avons proposé dans cet exemple d'établir comme limite maximale du seuil la distance minimale séparant deux échantillons de classes différentes pour chaque performance. Dans notre exemple, nous avons plusieurs performances donc plusieurs seuils maximaux. Or notre algorithme tel qu'il a été proposé au chapitre III ne propose qu'un unique seuil pour toutes les caractéristiques. Cette fonctionnalité peut cependant facilement être intégrée à l'algorithme, mais n'a pas été réalisée ici.

L'exemple suivant n'utilise qu'une performance, on montrera alors la pertinence du choix de la limite de seuil par l'analyse de la base de données.

#### IV.B.2 Convertisseur Analogique Numérique en Temps Continu

Cet exemple utilise un circuit de conversion analogique numérique en temps continu (ADC-CT). Il s'agit d'un circuit bien plus complexe que le LNA pré-

senté ci-avant, car il est fortement non-linéaire. Les applications visées pour ce type de circuit sont essentiellement des applications de communication à faible consommation orientées vers l'Internet des Objets.

L'objectif de cette section est d'assister la conception du test de l'ADC-CT en éliminant une mesure de performances chère pour la remplacer par d'autres mesures de performance.

#### IV.B.2.a Présentation du convertisseur en temps continu

Le circuit de conversion analogique numérique en temps continu (ADC-CT) est représenté sur la figure IV.12, et son fonctionnement est détaillé dans [Patil et al., 2015].

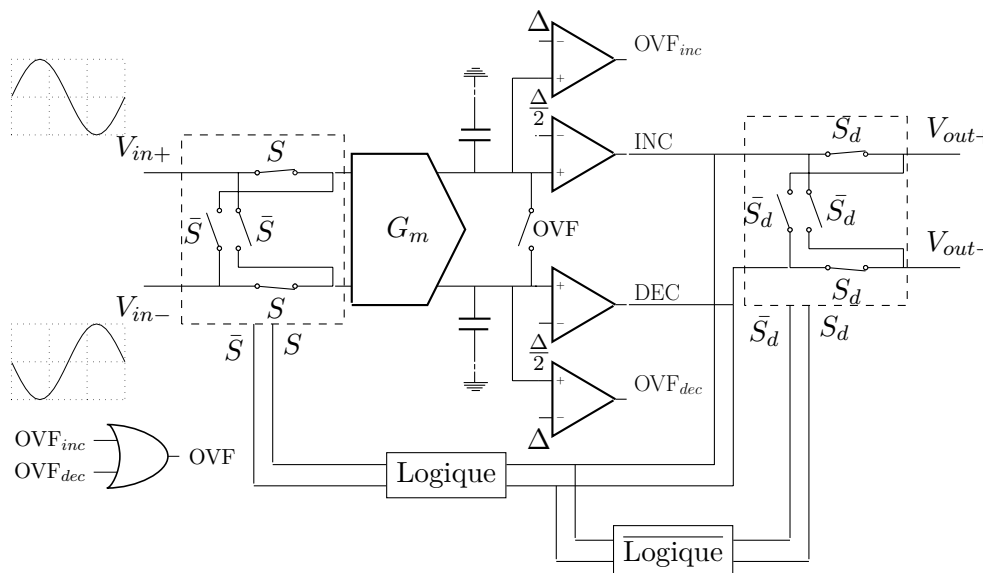


FIGURE IV.12 – Schéma du convertisseur analogique numérique en temps continu, reproduit depuis [Patil et al., 2015]

Il réalise une conversion analogique numérique rapide, à faible consommation, idéale pour les applications de type IoT (*Internet of Things*, Internet des Objets). Celles-ci exigent la capacité de convertir un signal analogique en un signal numérique avec une consommation électrique faible. Or, une source de consommation électrique importante provient de l'arbre d'horloge. Ainsi, une solution technique pour les circuits IoT consiste à construire des solutions complètes sans arbre d'horloge (i.e. en temps continu) dédiées, pour le cas particulier de l'ADC, à des applications de type *Wake up Radio* (Radio de réveil).

Pour le fonctionnement de ce circuit ADC-CT, les éléments d'intérêt sont le bruit, le gain et la consommation. Ils sont accessibles via les performances suivantes :

- Plage dynamique exempte de parasites, *Spurious Free Dynamic Range* (SFDR) ;
- Rapport signal à bruit, *Signal to Noise Ratio* (SNR) ;
- Gain ;
- Consommation.

De plus, les grandeurs internes d'intérêt pour ce circuit sont les commutations réalisées par le signal  $S$ , et les commutations suivantes :

- Nombre de commutations de l'interrupteur OVF ;
- Nombre de commutations des comparateurs DEC et INC ;
- Nombre de commutations des comparateurs  $OVF_{inc}$  et  $OVF_{dec}$ .

Parmi les performances à valider, le SFDR requiert une analyse du circuit dans le domaine de Fourier. La précision de cette analyse dépend en partie du nombre de points pour les signaux d'entrée et de sortie. L'objectif est d'identifier des mesures alternatives permettant d'éviter la mesure directe du SFDR, puis de construire un modèle de classification permettant de discriminer les circuits fonctionnels et non-fonctionnels.

Pour ce faire, la méthode proposée est mise en œuvre, comme expliqué dans la procédure qui suit.

#### IV.B.2.b Procédure d'application de la méthode proposée sur l'ADC-CT

Tout d'abord, une base de données est construite par simulation Monte Carlo du circuit. Elle contient 2780 échantillons de 25 observations chacun. Parmi ces observations, on retrouve celles des performances décrites ci-avant ainsi que des grandeurs internes (en particulier des commutations des comparateurs).

Cette base de données est ensuite utilisée en entrée de CDC-FOCUS pour extraire les observations pertinentes de coût le plus faible. Comme pour le cas du LNA, plusieurs valeurs de seuil sont utilisées, ce qui permet d'obtenir différentes sélections d'observations, à des coûts et avec des métriques différentes. En tout, ce sont 5000 valeurs de seuil différentes qui sont utilisées<sup>5</sup>. Un coût de 1 est associé à toutes les observations, sauf l'observation du SFDR qui a un coût de 30.

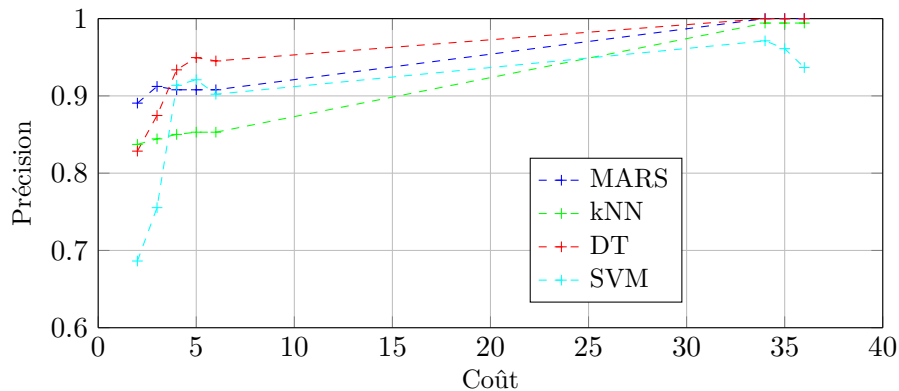
---

5. Ces 5000 valeurs de seuil sont issues de différentes campagnes d'essai "brute force" compilées. C'est la raison de leur grand nombre.

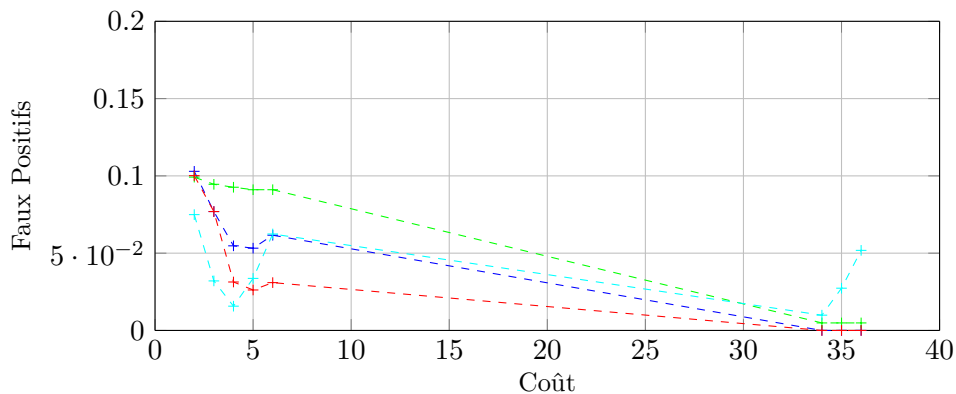
Enfin, les modèles de classification sont construits pour les 5000 sélections réalisées par CDC-FOCUS. 75% des échantillons sont utilisés pour l'entraînement, et 25% pour la validation et la construction des matrices de confusion. Comme pour le cas du LNA, les métriques de test sont obtenues par les matrices de confusion et les coûts des tests par la somme des coûts unitaires des observations extraites par CDC-FOCUS.

#### IV.B.2.c Résultats de l'application de la méthode sur l'ADC-CT

Les résultats obtenus à la suite de l'exécution de la méthode proposée sont présentés en figure IV.13.



(a) Précision



(b) Faux positifs

FIGURE IV.13 – Métriques des algorithmes de classification pour l'ADC-CT lorsque le SFDR est présent

Les courbes représentent les métriques de précision et de faux positifs en fonction des coûts associés aux observations réalisées. Un saut est visible entre le coût de 6 et le coût de 30. Ce saut indique que la méthode sélectionne



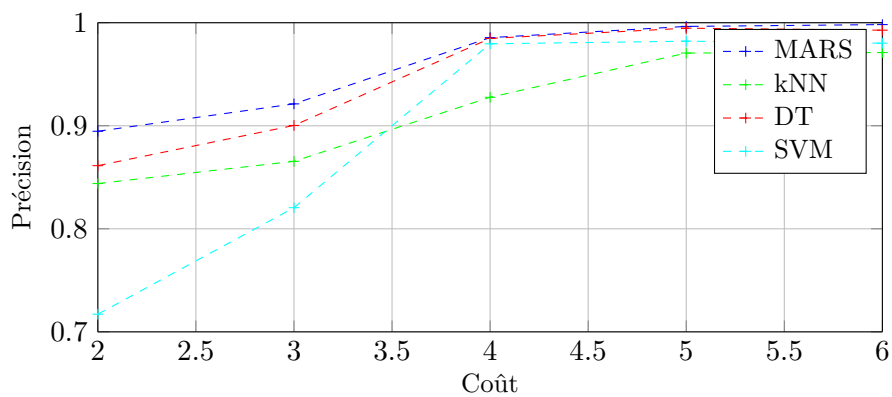
alors le SFDR et ne le remplace plus. En conséquence, les coûts supérieurs à 30 montrent une haute précision et un faible taux de faux positifs. Ces améliorations visibles sont le résultat de la définition simple des performances, basée seulement sur le SFDR. Ainsi, les modèles contruits incluant le SFDR ont naturellement une précision élevée et un taux de faux positifs faible, mais ils ont aussi un coût élevé.

Cependant, la méthode propose des modèles dont les coûts sont supérieurs à 30. Ce n'est pas attendu puisqu'un modèle contenant le SFDR contient déjà toute l'information pertinente. Ce comportement est le résultat d'une contrainte trop forte sur l'algorithme CDC-FOCUS (un seuil appliqué trop élevé). Dans ce cas, l'algorithme tente d'ajouter des observations qui sur-spécifient le SFDR. Ce comportement entraîne une dégradation de la précision et du taux de faux positifs des modèles qui y sont sensibles, ici l'algorithme SVM.

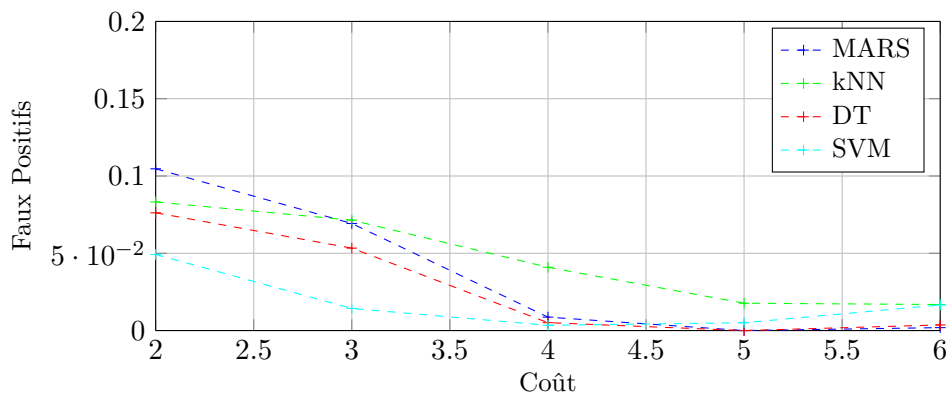
Dans ce cas d'utilisation, l'objectif est l'élimination d'une observation précise et connue, le SFDR. Il est donc possible d'utiliser la méthode proposée en éliminant le SFDR dès la phase d'acquisition de données. La procédure d'exécution est alors similaire à celle de la base de données avec SFDR, mais l'observation du SFDR est éliminée des mesures avant l'exécution de l'algorithme CDC-FOCUS. De même, il n'est pas utilisé pour la génération des modèles de classification. L'intérêt d'une telle analyse est de forcer l'algorithme à ne pas proposer la sélection du SFDR (puisque'il est absent), même lorsque le seuil est élevé (c'est-à-dire que la contrainte est forte). De cette façon, il est possible de savoir si CDC-FOCUS "ne voit pas" des ensembles de mesures dont le coût serait inférieur à 30. Les résultats obtenus dans ce cas sont présentés en figure IV.14.

Nous déduisons de la comparaison des figures IV.13 et IV.14 que le coût maximal avant d'inclure le SFDR est de 6 dans les deux cas. En effet, les deux exécutions conduisent à l'extraction des mêmes observations. Cependant, lorsque l'observation du SFDR est possible, l'algorithme propose de l'ajouter quand le seuil appliqué est trop élevé. Les métriques restent meilleures lorsque le SFDR est absent de la base de données. L'analyse des observations sélectionnées lorsque le SFDR est présent et lorsqu'il est absent montre que ce sont les mêmes observations qui sont sélectionnées pour des coûts identiques. Les écarts de métriques observés sont donc induits par des constructions de modèle différentes, notamment dans la composition des ensembles d'entraînement et de validation.

Dans cet exemple, c'est encore une méthode *brute force* de parcours des différents seuils qui permet d'obtenir les métriques en fonction du coût. Comme nous l'avons signalé dans l'exemple précédent, il est possible de donner une limite maximale de seuil au-dessus de laquelle il n'y a pas assez de points



(a) Précision

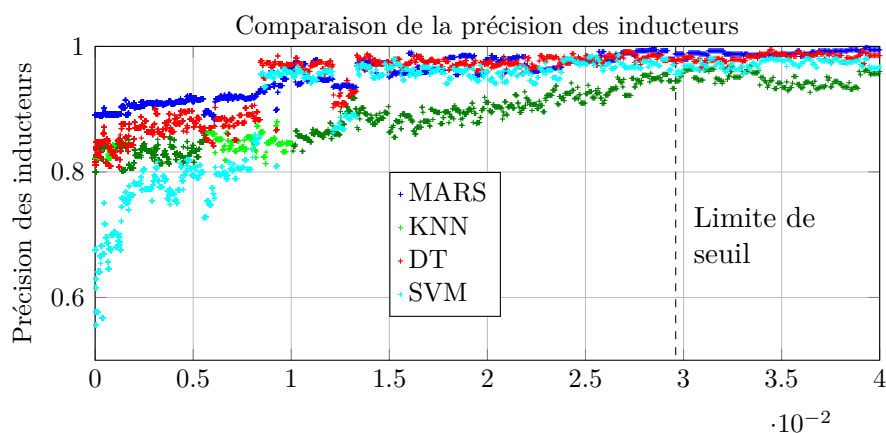


(b) Faux positifs

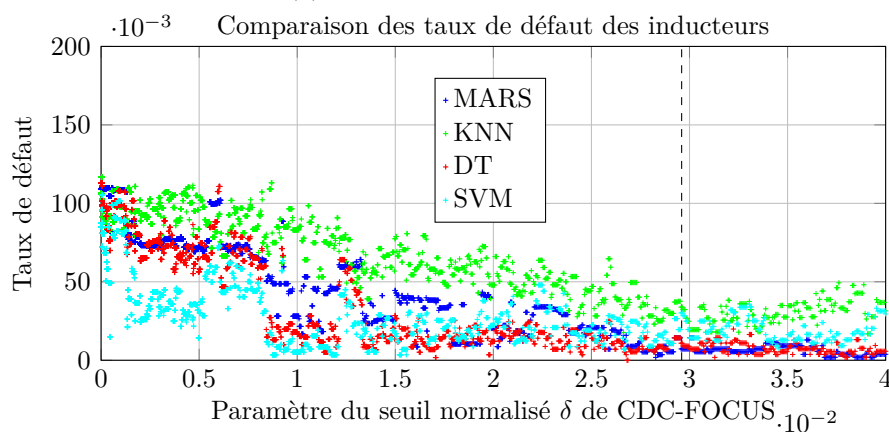
FIGURE IV.14 – Métriques des algorithmes de classification pour l'ADC-CT lorsque le SFDR est absent de la base de données

dans la base de données pour améliorer la classification. Il est simple d'utiliser l'exemple de l'ADC-CT pour appliquer cette approche car d'une part il n'y a qu'une performance et donc un seul seuil maximal à calculer, et d'autre part l'algorithme CDC-FOCUS tel qu'utilisé ne considère qu'un seul seuil de paramétrage. Ainsi, la figure IV.15 page suivante donne l'évolution des métriques de test en fonction du seuil choisi. La limite de seuil qui y apparaît est calculée à partir de la plus petite distance entre deux points de classes différentes, normalisée sur les valeurs prises par le SFDR.

Les courbes ainsi obtenues montrent que au-delà de la limite de seuil, les modèles de classification ne deviennent pas meilleurs et peuvent même se détériorer. Ce phénomène est en particulier visible sur les métriques du kNN. Il est donc possible d'évaluer une limite maximale du seuil de paramétrage de CDC-FOCUS à partir de laquelle il n'est pas possible d'obtenir une meilleure précision en raison de la dispersion des circuits.



(a) Précision des classifieurs



(b) Taux de défaut des classifieurs

FIGURE IV.15 – Évolution des métriques de test en fonction du seuil de CDC-FOCUS. La limite de seuil est matérialisée par le trait discontinu noir.

#### IV.B.2.d Synthèse sur l'application de la méthode à l'ADC-CT

La méthode appliquée sur l'ADC-CT a défini des ensembles de mesures permettant de ne pas mesurer directement le SFDR, ce qui entraîne une forte réduction du coût associé au test. Une mesure de coût 30 est remplacée par une série de mesures de coût total égal à 6.

La dégradation de la précision et du taux de faux positifs liée à la substitution de la mesure de SFDR par d'autres mesures, est contrôlée par la contrainte (c'est-à-dire le seuil) appliquée à CDC-FOCUS. L'expérience consistant à faire disparaître le SFDR de la base de données montre que CDC-FOCUS ne propose pas d'ensemble de mesures dont le coût est supérieur à 6 pour éliminer la mesure du SFDR. Ce résultat permet de prouver la cohérence de la sélection des

caractéristiques pertinentes par rapport à la composition de la base de données.

Cet exemple ne comportant qu'une seule performance pour la classification, nous l'avons utilisé pour évaluer la méthode d'établissement de la limite maximale du seuil à partir de la base de données. Nous avons donc établi la limite maximale comme étant la plus petite distance normalisée entre deux échantillons de classes différentes. Notre exemple a montré qu'au delà de cette limite, les métriques de test ne s'améliorent pas, et peuvent même se détériorer. Nous formulons alors l'hypothèse que les algorithmes de classification subissent alors un sur-apprentissage.

### IV.B.3 Conclusion sur l'utilisation de la méthode proposée avec des simulations

Les bases de données issues de simulations ont l'avantage de pouvoir contenir un grand nombre d'échantillons et un grand nombre de signaux extraits. Les limites à la taille des bases de données traitées sont les capacités de calcul, la mémoire à disposition et le nombre de licences disponibles par programme. Par ailleurs, les choix d'implémentation informatique des programmes de simulation et d'analyse des données sont déterminant dans la vitesse d'exécution de ceux-ci.

Les exemples du LNA et de l'ADC-CT utilisés ici ont permis de valider la méthode de réduction du coût de test et la conservation de métriques de test "acceptables". Le paramétrage de l'algorithme CDC-FOCUS permet d'obtenir un équilibre entre métriques de test et coût de test mais son choix est empirique et obtenu par un balayage exhaustif d'un intervalle donné. Rappelons que la nécessité du seuil est héritée de C-FOCUS et que ses auteurs en proposent une évaluation empirique également.

Nous avons, au travers des exemples, proposé une évaluation de la limite du seuil fondée sur l'analyse de la base de données. Plus particulièrement, nous proposons comme seuil la distance la plus petite entre deux échantillons de classes différentes. L'exemple de l'ADC-CT valide cette approche dans le cas d'une seule performance.

Le choix d'un algorithme de classification avait été laissé ouvert. Les exemples du LNA et de l'ADC-CT montrent que les meilleures métriques sont obtenues par MARS et par les Arbres de Décision et les Machines à Vecteurs de Support (SVM). L'algorithme kNN montre des métriques inférieures aux métriques des autres algorithmes ce qui provient probablement de son approche basée sur la distance euclidienne. Une approche à noyaux (à condition d'utiliser un noyau adéquat) pourrait conduire à de meilleures métriques.

## IV.C Validation de la méthode sur des mesures

Cette section propose d'utiliser notre méthode dans le cas particulier des données obtenues par mesure de circuits. La base de données, présentée en section IV.C.1, est fournie par Presto Engineering.

Les données issues de mesures sont particulières car elles sont généralement obtenues par le principe du *Stop on Fail*. Comme expliqué dans la section IV.A, ce *Stop on Fail* réduit le temps de test d'un circuit en arrêtant l'exécution des tests dès qu'une mesure montre que le circuit est défaillant. Néanmoins, ce principe ne permet pas d'avoir une base de données complète.

Par ailleurs, compte tenu du caractère confidentiel des circuits mesurés, ni les schématiques ni les définitions des tests ne sont disponibles. Il n'est donc pas possible de connaître le fonctionnement du circuit. Dans ces conditions, l'objectif de la méthode est d'extraire des mesures qui semblent pertinentes et qui pourraient être suffisantes pour évaluer la classification des circuits, tout en évitant la réalisation de toutes les mesures.

### IV.C.1 Présentation de la base de données et procédure d'application de la méthode

La base de données est issue d'une campagne de 336 mesures réalisées sur 3920 circuits. Comme nous l'avons expliqué ci-avant, elle n'est pas directement utilisable par notre méthode en raison des données manquantes. En effet, CDC-FOCUS ne peut fonctionner que si les échantillons contiennent toutes les mesures. Ainsi, contrairement aux cas étudiés précédemment en simulation, une première transformation des données est opérée pour obtenir une base de données complète comme expliqué en section IV.A.2.a.

Le nombre d'échantillons dans la base de données complète obtenue est donc de 2218 circuits et 283 mesures. Par la suite, notre méthode est appliquée de la même manière que pour les cas de simulation. L'algorithme CDC-FOCUS analyse la base de données selon plusieurs valeurs de seuil, et pour chaque résultat fourni par CDC-FOCUS, quatre modèles de classification sont construits. Il est à noter que les modèles basés sur les Arbres de Décision (DT) ont été identifiés en section II.D.1 comme inadaptés aux données issues de mesures en raison de leur dépendance au bruit. Ils seront utilisés dans cette section pour évaluer la dégradation des métriques, dans le cadre d'une utilisation sur des données issues de mesures.

Pour fixer les coûts unitaires des mesures, nous avons choisi de relier l'unité de la mesure au coût associé car nous n'avons pas accès à la schématique du circuit. Ainsi, une mesure de temps ou de fréquence a un coût élevé, alors

qu'une mesure de tension ou de courant a un coût faible. Ces coûts sont reportés dans le tableau IV.2.

Unité	Coût associé
mV	2
V	2
$\mu\text{A}$	2
LSB	1
nA	1
mA	1
KHz	5
$\mu\text{s}$	6
%	3
<i>Sans unité</i>	2

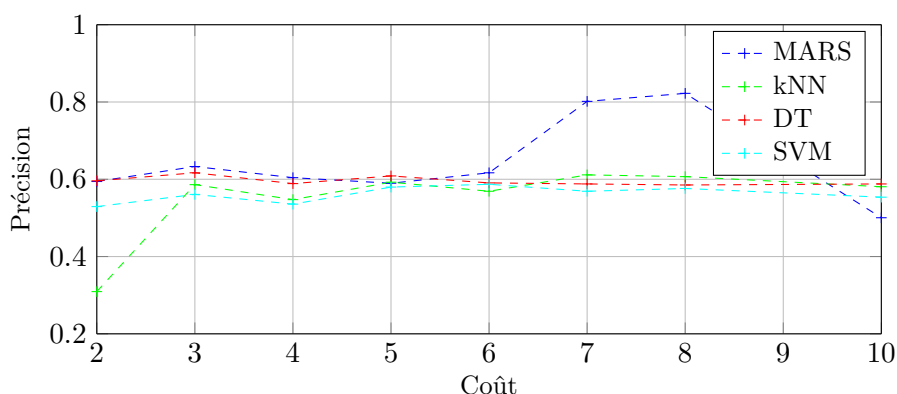
TABLE IV.2 – Coûts des mesures selon leur unité

#### IV.C.2 Résultats de l'application de la méthode sur la base de données issue de mesures

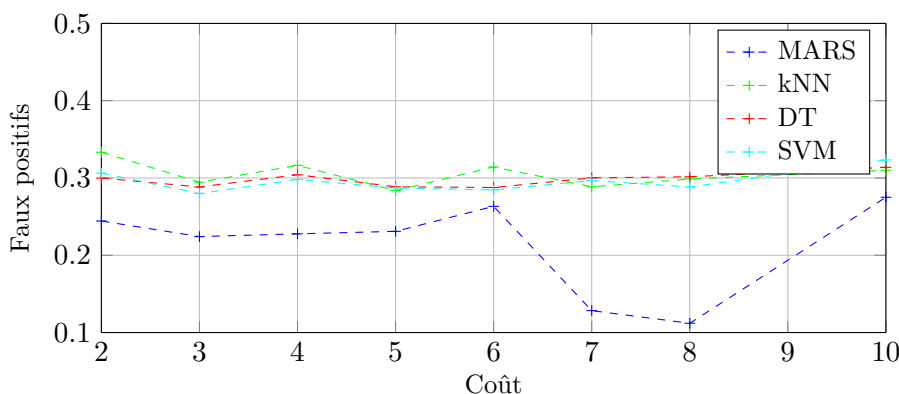
Comme pour les expériences en simulation de la section IV.B, nous nous intéressons aux métriques de précision, de faux positifs et de coût. Elles sont représentées en figure IV.16.

Contrairement aux cas en simulation, les résultats ne présentent pas une amélioration des métriques de test avec l'augmentation du coût des mesures. En effet, dans ce cas, seules deux valeurs de coût fournissent des métriques de test d'intérêt. Par ailleurs, pour la plupart des coûts (exceptés les coûts 7 et 8), la précision et les faux positifs sont stables à des valeurs correspondant au rendement des circuits. Ainsi, pour ces valeurs de coût, les modèles ne sont pas plus précis qu'un classifieur qui considère tous les circuits comme fonctionnels.

La méthode de parcours exhaustif d'un intervalle donné de valeurs de seuil de CDC-FOCUS permet cependant de montrer que les coût 7 et 8 proposent une amélioration des métriques de test et seulement pour l'algorithme de classification MARS. Dans la présentation des résultats issus de simulations, nous avons proposé de sélectionner le paramètre de configuration de CDC-FOCUS *a priori* à l'aide d'une analyse de la base de données. L'application de cette analyse ici est représentée sur la figure IV.17. Elle montre l'évolution de la précision des modèles de classification selon le paramètre CDC-FOCUS utilisé, ainsi que les seuils limites obtenus suite à l'analyse de la base de données.



(a) Précision des algorithmes de classification



(b) Faux positifs des algorithmes de classification

FIGURE IV.16 – Précision et faux positifs des algorithmes de classification suivant le coût des observations retenues par CDC-FOCUS

Nous obtenons plusieurs seuils car plusieurs performances sont utilisées dans cet exemple. Comme expliqué en section précédente, l'utilisation de cette méthode de sélection de seuil limite s'applique directement lorsque plusieurs seuils peuvent être utilisés par CDC-FOCUS, ce qui n'est pas le cas dans l'implémentation que nous avons choisi ici. Cependant, l'un de ces seuils permet d'obtenir le maximum de précision.

### IV.C.3 Conclusion de la validation de la méthode proposée sur la base de données de mesures

Le cas de validation sur les mesures est rendu complexe en raison tout d'abord du pré-traitement de la base de données. Ce pré-traitement est nécessaire car les bases de données issues de mesures sont incomplètes. Il élimine alors de la base de données les circuits qui ne parcourent pas tous les tests. Ainsi, ce sont les circuits défectueux qui sont éliminés. Cependant, la base de

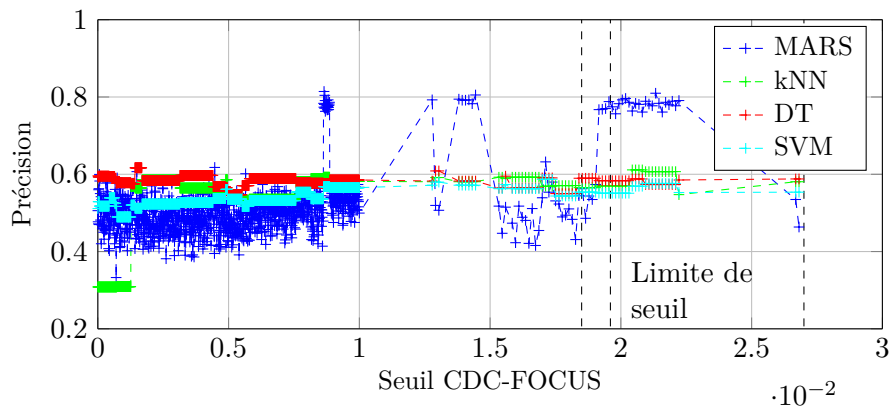


FIGURE IV.17 – Évolution de la précision en fonction du seuil de CDC-FOCUS

données obtenue est utilisée dans le but de remplacer les mesures les plus chères par d'autres mesures moins chères. La notion de mesures chères ou peu chères est liée ici à la nature des signaux mesurés. En particulier, nous cherchons à éliminer les mesures de fréquence et à conserver les mesures de tension ou courant.

Les résultats des modèles de classification obtenus avec la base de données issue de mesures ne suivent pas le même profil que ceux obtenus avec des données issues de simulation. En effet, les performances ne sont améliorées que pour les coûts 7 et 8 et uniquement lors de la modélisation par MARS.

La figure IV.18 présente l'évolution du coût des caractéristiques sélectionnées en fonction du seuil CDC-FOCUS utilisé. On y montre que plus la contrainte sur l'algorithme CDC-FOCUS est élevée, plus le coût des caractéristiques sélectionnées est élevé, ce qui reste cohérent avec les résultats obtenus précédemment.

Enfin, nous avons appliqué notre analyse de base de données pour déterminer le paramètre "idéal" de CDC-FOCUS, c'est-à-dire le paramètre qui permet d'obtenir les métriques de test les plus élevées. Nous avons obtenus plusieurs seuils dont l'un est dans la zone d'amélioration de la précision observée. Nous avons donc montré que notre approche d'analyse de la base de données est cohérente, aussi bien en simulation qu'en mesures.

## IV.D Conclusion sur la validation de la méthode

Ce chapitre a mis en place différents cas d'utilisation mettant en œuvre notre méthode. Le point commun entre tous ces cas d'utilisation est de réduire le coût associé à l'observation d'un ensemble de performances, tout en conservant des métriques de test satisfaisantes. Les métriques de test utilisées ici sont celles définies au chapitre I. Il s'agit de la précision et du taux de faux positifs,



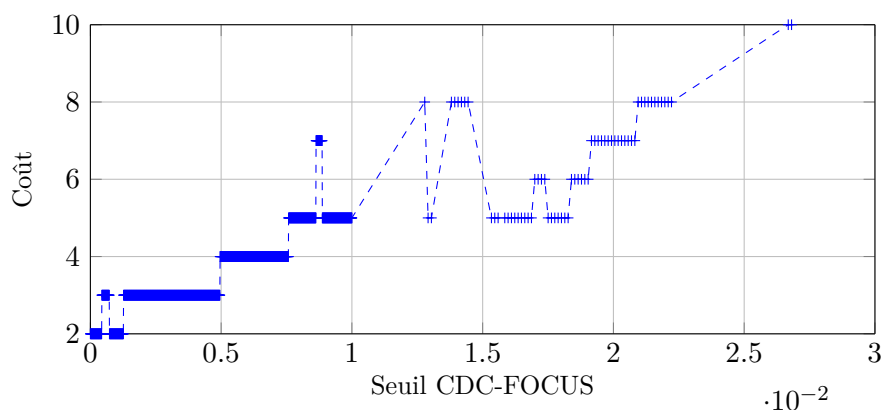


FIGURE IV.18 – Évolution du coût des caractéristiques sélectionnées en fonction du paramètre de CDC-FOCUS

deux notions reliées aux problématiques de performance des tests.

En simulation, l'utilisation de notre méthode permet de remplacer une mesure que l'on sait chère par un ensemble de mesures de coût plus faible, ou de sélectionner un sous-ensemble de mesures de performances permettant de réduire le coût du test sans détériorer les métriques de test.

La méthode est appliquée sur des données issues de mesures pour désigner des mesures à faible coût permettant de remplacer des mesures à coût élevé. Ces dernières sont pour la plupart des mesures en fréquence ou des mesures en temps. En effet, celles-ci monopolisent le circuit testé et le testeur pendant une longue durée, et demandent des capacités de calcul particulières (entre autres, conversion et calcul dans le domaine de Fourier) qui sont elles aussi onéreuses. Appliquer notre méthode permet ainsi de définir des mesures de substitution permettant de réaliser les tests à un coût plus faible, et ce sans avoir de connaissances *a priori* sur la conception du circuit.

Les résultats présentés dans ce chapitre montrent que le choix du paramètre CDC-FOCUS est critique car il impacte fortement la qualité des résultats obtenus. Rappelons que ce seuil a été introduit dans CDC-FOCUS par C-FOCUS afin d'utiliser des données continues. Les auteurs de C-FOCUS ont alors gardé ouvert la question du choix de ce seuil. Dans nos travaux, nous montrons qu'une analyse de la base de données permet d'obtenir *a priori* une valeur de paramètre fournissant à la fois de "bonnes" métriques de test et un coût de test réduit. Cependant, l'implémentation que nous avons choisi pour CDC-FOCUS ne permet pas d'utiliser des seuils différents par caractéristiques, ce qui serait requis pour appliquer correctement les résultats de l'analyse de la base de données. L'algorithme CDC-FOCUS doit donc être légèrement modifié

pour complètement exploiter cette capacité.



---

## Conclusion et Perspectives

---

Avec l'évolution des technologies, la caractérisation des performances des circuits et leur mesure demandent des outils plus rapides et plus précis. Parallèlement, le temps de test de chaque circuit augmente en raison de performances plus strictes et d'un nombre de performances à tester plus élevé. Le coût total du test est donc affecté et il devient nécessaire d'un point de vue économique de chercher à le réduire.

Nos travaux ont proposé une méthodologie de réduction du coût de test des circuits analogiques par sélection des caractéristiques pertinentes, le coût du test étant pris en compte au moment de cette sélection. Nous avons établi une démarche construite en trois étapes fondamentales, représentée sur la figure V.1. Cette démarche a été établie à l'issue du chapitre I où une rapide revue de l'état de l'Art a été menée. L'approche proposée est suffisamment générique pour répondre à différentes situations.

### V.A Résumé de la Méthodologie complète

Cette démarche en trois étapes permet de rendre notre méthodologie applicable en simulation pendant la conception d'un circuit, mais aussi au moment du test en utilisant des mesures du circuit. Dans le premier cas, le but recherché est l'établissement d'une stratégie de test qui minimise le coût du test. Dans le second, l'objectif est la diminution du coût d'une stratégie de test déjà établie.

Les trois étapes fondamentales de la méthode proposée sont l'acquisition d'informations sur le circuit, la recherche de signaux d'intérêts, et la construction de modèles de classification. L'acquisition d'informations est réalisée soit par simulations pendant le développement du circuit, soit par mesures. La recherche de signaux d'intérêts permet d'extraire les mesures d'intérêt pour valider le bon fonctionnement du circuit. Enfin, la dernière étape construit

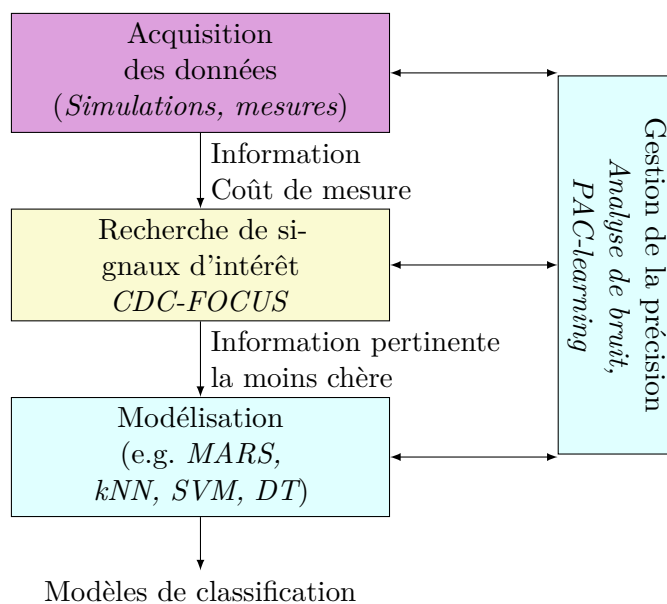


FIGURE V.1 – Rappel de la méthodologie proposée

un modèle permettant de relier les mesures sélectionnées à l'état du circuit, à savoir fonctionnel ou défectueux.

Nous avons analysé et comparé des algorithmes existant dans l'état de l'Art et déjà utilisé dans le cadre du test des circuits pour répondre à notre problématique. Les métriques que nous avons considéré sont la précision des modèles et le taux de faux positifs. Ces grandeurs sont intéressantes car elles sont fortement liées à l'aspect économique du test, à savoir vendre un produit fonctionnel ou vendre un produit défectueux. La vente de produits défectueux entraîne généralement un coût élevé induit par le "retour client" qu'il faut traiter. D'autres métriques sont aussi proposées par l'analyse des matrices de confusion, en particulier le taux de faux négatifs. Le coût associé est alors la perte sèche des produits fonctionnels invendus.

Les algorithmes de modélisation que nous avons utilisés sont les "k-plus proches voisins", les "Machine à Vecteurs de Supports" (*Support Vector Machine*), les "Arbres de Décision" et le "MARS" (*Multivariate Adaptive Regression Splines*). Le chapitre IV a montré que les résultats les plus satisfaisants ont été obtenus par MARS. Dans nos travaux nous utilisons MARS pour établir le modèle de classification mais MARS est avant tout un modèle de régression. Ainsi, nous régressons les performances des circuits à partir des mesures d'intérêt et comparons ces performances régressées aux spécifications.

Cette méthodologie a été présentée dans une publication sous le titre de

“Weight Based Feature Selection Algorithm for Alternative Measurements” (M. Verdy *et al.*) (STEM workshop, 2014).

## V.B Sélection de caractéristique guidée par le coût

Nous avons proposé un algorithme permettant de sélectionner les mesures dont l’objectif est de tester un circuit analogique, tout en diminuant le coût du test. Cet algorithme, CDC-FOCUS (Cost Driven Continuous FOCUS) est hérité de l’algorithme C-FOCUS auquel nous proposons d’intégrer la prise en compte du coût dans la sélection de caractéristiques. Il a fait l’objet du chapitre III. Son fonctionnement repose sur l’analyse de tableaux de données et la recherche de conflits, correspondant à la situation de deux circuits “proches” qui sont pourtant de classes différentes.

Une limitation à cet algorithme est de ne pas pouvoir traiter de tableaux de données avec des informations manquantes. Dans le domaine du test, ce cas intervient lorsqu’on utilise une base de données incomplète, acquise par exemple à l’aide de la technique *Stop on Fail* qui permet d’arrêter le test d’un circuit dès qu’une mesure indique qu’il est défectueux. Par ailleurs, la sélection de caractéristiques par CDC-FOCUS ne prend pas en compte explicitement la précision de mesure associée aux caractéristiques. La prise en compte de ce paramètre pourrait permettre de sélectionner, parmi les mesures de même coûts, les mesures les plus précises.

Cet algorithme CDC-FOCUS a fait l’objet d’une publication sous le titre “Cost Driven Statistical Analysis for Selection of Alternative Measurements of Analog Circuits” (M. Verdy *et al.*) (21<sup>st</sup> International Conference on Electronics, Circuits, and Systems, 2014).

## V.C Validation de la méthode proposée sur des circuits

Dans le cadre de nos travaux, nous avons utilisé des circuits développés par le laboratoire LAIR et des mesures fournies par Presto-Engineering pour valider notre méthodologie. Les bases de données construites à partir des simulations ou des mesures contiennent une grande quantité de données, dont l’acquisition et le traitement soulèvent des problèmes auxquels nous avons apportés des solutions. En particulier, l’automatisation de simulations et l’agrégation de résultats pour les mesures en section IV.A.1, et l’étude des bases de données incomplètes en section IV.A.2.

En effet, l’acquisition de données par simulation de circuits complets nécessite beaucoup de temps et de puissance de calcul, ainsi que de la mémoire. La

solution adaptée consiste à distribuer les simulations sur plusieurs machines à n'aggréger que les données utilisées par la suite.

La construction de la base de données peut être également soumise à l'anonymisation des données. Dans ce cas, il n'est pas possible d'accéder au schéma fonctionnel du circuit mesuré. De plus, la méthode *Stop on Fail* est utilisée pour réduire le temps de mesure des circuits. Cette méthode a l'inconvénient majeur de conduire à des mesures manquantes. Nous proposons de contourner cette difficulté par une réduction de la base de données. Une autre solution envisageable est la reconstruction de données. Nous n'avons pas exploité ici cette solution car les techniques de reconstruction traditionnelles demandent une connaissance *a priori* des relations entre les données présentes et les données manquantes.

## V.D Évaluation de l'erreur dans la méthode

Pour nos travaux, l'évaluation et la prédiction de l'erreur commise est un indicateur de la qualité des résultats. De plus, notre méthode propose la possibilité de contrôler le compromis entre la précision des modèles de classification et le coût associé au test. Ce compromis est utile lorsque le dispositif testé doit l'être à faible coût avec une relaxation possible des contraintes de test, ou au contraire, si le test doit être très précis quitte à avoir un coût élevé. Pour juger de la qualité de nos modèles de classification, nous avons choisi les matrices de confusion car elles représentent les métriques économiques du test (pertes associées aux faux-positifs en particulier). Nous avons aussi proposé une étude de l'influence du bruit de mesure sur les métriques de test. Cette étude est limitée par l'hypothèse de bruit additif. Elle a été publiée sous le titre de "Balancing test cost reduction vs. measurements accuracy at test time" (M. Verdy *et al.*) (NEWCAS, 2015).

## V.E Synthèse des perspectives

Nous proposons maintenant les pistes de poursuite de nos travaux.

Tout d'abord, la prédiction de l'impact du bruit de mesure sur les métriques de test ne peut être conduite qu'avec l'hypothèse de bruit additif. L'intégration de ce résultat dans la méthode proposée pourrait être améliorée en relâchant cette hypothèse. L'hypothèse de bruit additif doit être faite en raison de l'utilisation d'une simple "déconvolution". L'utilisation de technique de régularisation de ce problème de déconvolution est une piste à étudier pour élargir cette hypothèse.

Cette analyse peut être utilisée *in fine* dans l'algorithme de sélection de caractéristiques pour sélectionner les mesures les moins chères, qui seront le moins sensible au bruit de mesure. Cette donnée peut être prise en compte dans l'état actuel de CDC-FOCUS par l'attribution d'un coût plus élevé des mesures qui nécessitent une précision élevée de mesure.

Concernant la construction des modèles de classification, nous avons fait le choix de ne pas explorer les modèles de réseaux de neurones. Ce choix est fondé sur la nature "boite noire" des réseaux de neurones et notre volonté de conserver la connaissance "physique" des circuits. Nous pouvons exploiter les notions de distance et de conflits utilisés par CDC-FOCUS pour construire un modèle de classification. La théorie des graphes permet de manipuler les bases de données que nous utilisons et de définir des relations particulières entre les échantillons pour les lier entre eux. De telles relations pourraient être l'appartenance à la même classe. Il serait alors possible de représenter la frontière entre les circuits défectueux et les circuits fonctionnels à l'aide de plusieurs graphes construits à partir de la base de données d'échantillons et des caractéristiques conservées par CDC-FOCUS.





# Annexes



## Annexe A

---

# Nombre d'échantillons en simulation Monte-Carlo

---

Tout d'abord,  $N$  échantillons sont générés. Par la suite,  $G_i, i \in \{1, \dots, N\}$  est définie de sorte à ce que  $G_i = 1$  si les performances du  $i^{\text{ème}}$  circuit correspondent aux spécifications, et  $G_i = 0$  dans le cas contraire. Notant  $Y$  le rendement, il vient :

$$n_g = \sum_{i=1}^N G_i \quad (\text{A.1})$$

$n_g$  suit une distribution binomiale de paramètres  $N$  et  $Y$ . Dans ce cas, l'espérance et la variance de l'observateur de Monte-Carlo  $\hat{Y}$  sont exprimées par :

$$E[\hat{Y}] = \frac{NY}{N} = Y \quad (\text{A.2})$$

$$\text{Var}[\hat{Y}] = \frac{N(1-Y)Y}{N^2} = \frac{(1-Y)Y}{N} \quad (\text{A.3})$$

Ensuite, par le théorème de de Moivre-Laplace, la variable aléatoire  $\hat{Y}$  converge vers la loi normale centrée réduite :

$$\begin{aligned} Z &= \frac{n_g - NY}{\sqrt{NY(1-Y)}} \\ &= \frac{\hat{Y} - Y}{\sqrt{\frac{Y(1-Y)}{N}}} \end{aligned} \quad (\text{A.4})$$

Il est admis de considérer heuristiquement que  $Z$  suit la loi normale centrée réduite lorsque la condition  $NY(1-Y) \gg 1$  est vérifiée.

Ensuite, en notant  $\epsilon$  l'erreur admissible, la probabilité que la différence entre le rendement observé et le rendement réel soit supérieure à  $\epsilon$  est exprimée par :

$$\begin{aligned}\rho &= P \left[ |\hat{Y} - Y| \leq \epsilon \right] = P \left[ \frac{\hat{Y} - Y}{\sqrt{\frac{(1-Y)Y}{N}}} \leq \frac{\epsilon}{\sqrt{\frac{(1-Y)Y}{N}}} \right] \\ &= P \left[ Z \leq \frac{\epsilon}{\sqrt{\frac{(1-Y)Y}{N}}} \right]\end{aligned}\tag{A.5}$$

Par développement des intervalles de confiance de  $Z$  qui suit la loi normale centrée, il vient :

$$\rho = 2\phi \left( \frac{\epsilon\sqrt{N}}{\sqrt{(1-Y)Y}} \right) - 1\tag{A.6}$$

Il s'en suit [Brockman and Director, 1989] :

$$N = \frac{(1-Y)Y}{\epsilon^2} \left[ \phi^{-1} \left( \frac{1+\rho}{2} \right) \right]^2\tag{A.7}$$

avec :

- $N$  le nombre d'échantillons minimal nécessaire pour assurer la précision requise sur le calcul du rendement
- $Y$  la vraie valeur du rendement ou, à défaut, une évaluation de la valeur minimale du vrai rendement
- $\epsilon$  l'erreur requise sur le rendement
- $\rho$  la confiance dans l'erreur sur le rendement
- $\phi^{-1}$  la fonction inverse de la fonction de distribution normale standard

---

# Analyse en Composante Principale

---

Ce chapitre propose la démonstration du résultat utilisé dans la méthode d'Analyse en Composante Principale. La matrice de projection  $\mathbf{Q}$  est construite à partir de la décomposition spectrale de la matrice de covariance des données d'origine.

Les notations utilisées sont celles de la section II.C.1.a : avec :

- $\mathbf{C}$  la matrice de dimension  $[P \times N]$  contenant les observations des circuits
- $\mathcal{C}$  l'espace d'observation dans lequel est décrit  $\mathbf{C}$
- $\mathcal{L}$  l'espace de projection, de dimension inférieure à celle de  $\mathcal{C}$
- $\mathbf{L}$  la matrice projetée de  $\mathbf{C}$  dans  $\mathcal{L}$
- $N$  le nombre de circuits
- $P$  la dimension des observations

L'objectif est alors d'explicitier la matrice de projection  $\mathbf{Q}$  telle que :

$$\mathbf{L} = \mathbf{Q}^T \mathbf{C} \quad (\text{B.1})$$

Pour ce faire, l'expression de la variance du nuage centré sur une direction  $\mathbf{u}$  est :

$$\sigma_{\mathbf{u}} = \frac{1}{N} \sum_{i=1}^N \left( \mathbf{u}^T (\mathbf{c}_i - \bar{\mathbf{c}}) \right)^2 = \mathbf{u}^T \mathbf{C}_C \mathbf{u} \quad (\text{B.2})$$
$$\mathbf{C}_C = \frac{1}{N} \sum_{i=1}^N (\mathbf{c}_i - \bar{\mathbf{c}}) (\mathbf{c}_i - \bar{\mathbf{c}})^T = \frac{1}{N} \mathbf{C} \mathbf{C}^T - \bar{\mathbf{c}} \bar{\mathbf{c}}^T$$

avec :

- $\mathbf{u}$  le vecteur unitaire définissant la direction sur laquelle est projetée le nuage
- $\sigma_{\mathbf{u}}$  la variance du nuage projeté sur la direction  $\mathbf{u}$
- $\bar{\mathbf{c}}$  le vecteur de taille  $P$  contenant les valeurs moyennes des  $P$  observations sur l'ensemble des  $N$  circuits
- $\mathbf{C}_{\mathbf{C}}$  la matrice de covariance de  $\mathbf{C}$

Dès lors, la variance maximale s'obtient à partir de l'équation (B.2) par un multiplicateur de Lagrange sous la contrainte de  $\mathbf{u}$  unitaire :

$$\mathbf{u}^* = \arg \max_{\mathbf{u}} \left( \mathbf{u}^T \mathbf{C}_{\mathbf{C}} \mathbf{u} + \lambda (1 - \mathbf{u}^T \mathbf{u}) \right) \quad (\text{B.3})$$

$$\Leftrightarrow \mathbf{C}_{\mathbf{C}} \mathbf{u}^* = \lambda \mathbf{u}^* \quad (\text{B.4})$$

équation (B.4) introduit les vecteurs propres et valeurs propres de  $\mathbf{C}_{\mathbf{C}}$ . En utilisant l'équation (B.4) dans (B.2), il vient :

$$\sigma_{\mathbf{u}} = \lambda \quad (\text{B.5})$$

Les couples  $(\lambda, \mathbf{u}^*)$  solutions de l'équation (B.4) sont donc les valeurs propres et vecteurs propres de  $\mathbf{C}_{\mathbf{C}}$ . De plus, (B.5) montre que la variance maximale est obtenue pour la valeur propre la plus grande, et la direction est donnée par le vecteur propre associé. Ainsi, rechercher la direction de variance la plus grande revient à réaliser la décomposition spectrale de la matrice de covariance de  $\mathbf{C}$ . De la même façon, la  $k^{ieme}$  direction de variance maximale est le  $k^{ieme}$  vecteur propre correspondant à la  $k^{ieme}$  plus grande valeur propre. Ainsi, il est possible d'observer le nuage de points dans un espace défini par les  $k$  premiers vecteurs propres correspondant aux  $k$  valeurs propres les plus grandes. Il est alors possible d'explicitier la projection (II.8) page 44 :

$$\mathbf{Q}^T = [\mathbf{u}_1 \cdots \mathbf{u}_k] \quad (\text{B.6})$$

Le nombre  $k$  de vecteurs à conserver est alors à définir. Des critères sont proposées en section II.C.1.a.

## Annexe C

---

# Régression Inverse par Tranche

---

Dans cette annexe, un rappel sur la Régression Inverse par Tranche est proposé. L'objectif est de justifier l'expression de la base de l'espace EDR  $[\beta_1, \dots, \beta_K]$  dans l'expression :

$$y = f(\beta_1 \mathbf{c}_i, \beta_2 \mathbf{c}_i, \dots, \beta_K \mathbf{c}_i, \epsilon) \quad (\text{C.1})$$

La suite du développement consiste à montrer que les  $K$  plus grandes valeurs propres non nulles de la matrice de covariance (II.15) page 47 sont associées aux  $K$  vecteurs propres qui sont les vecteurs de l'espace EDR  $\beta_n, n \in \{1, \dots, K\}$  [Li, 2012]. Pour ce faire, il est nécessaire d'introduire l'hypothèse suivante.

**Hypothèse 2.** Soit  $\mathbf{b} \in \mathbb{R}^P$ , alors l'espérance  $E[\mathbf{bc}_i | \beta_1 \mathbf{c}_i, \dots, \beta_K \mathbf{c}_i]$  est linéaire en  $\beta_1 \mathbf{c}_i, \dots, \beta_K \mathbf{c}_i : \exists \alpha_k \in \mathbb{R}, k \in \{1, \dots, K\}$  tels que

$$E[\mathbf{bc}_i | \beta_1 \mathbf{c}_i, \dots, \beta_K \mathbf{c}_i] = \sum_{k=1}^K \alpha_k \beta_k \mathbf{c}_i \quad (\text{C.2})$$

L'hypothèse 2 est une contrainte sur la distribution des  $\mathbf{c}_i$ . Elle est complexe à vérifier et donc souvent remplacée par la contrainte plus forte de distribution elliptique symétrique des  $\mathbf{c}_i$  (comme par exemple la loi normale). En pratique, il est souvent simplement requis que la distribution des  $\mathbf{c}_i$  ne soit pas ouvertement non elliptique symétrique.

Lorsque l'hypothèse 2 est vérifiée, il vient :  $\forall \mathbf{b}$  orthogonal au sous-espace formé par  $\beta_k \Sigma_{\mathbf{c}_i}, k \in \{1, \dots, K\}$ ,

$$\begin{aligned} \mathbf{b} E[\mathbf{c}_i | y] &= E[E[\mathbf{bc}_i | \beta_1 \mathbf{c}_i, \dots, \beta_K \mathbf{c}_i, y] | y] \\ &= E[E[\mathbf{bc}_i | \beta_1 \mathbf{c}_i, \dots, \mathbf{bc}_i] | y] \end{aligned} \quad (\text{C.3})$$



Dès lors,

$$\begin{aligned} \mathbf{b}E[\mathbf{c}_i|y] = 0 &\Leftrightarrow E[\mathbf{b}\mathbf{c}_i|\beta_1\mathbf{c}_i, \dots, \beta_K\mathbf{c}_i] = 0 \\ &\Leftrightarrow E\left[(E[\mathbf{b}\mathbf{c}_i|\beta_1\mathbf{c}_i, \dots, \beta_K\mathbf{c}_i])^2\right] = 0 \end{aligned} \quad (\text{C.4})$$

$$\Leftrightarrow E[E[\mathbf{b}\mathbf{c}_i|\beta_1\mathbf{c}_i, \dots, \beta_K\mathbf{c}_i]\mathbf{c}_i'\mathbf{b}'] = 0 \quad (\text{C.5})$$

$$\Leftrightarrow E\left[\left(\sum_{k=1}^K \alpha_k \beta_k \Sigma_{\mathbf{c}_i\mathbf{c}_i}\right) \mathbf{c}_i'\mathbf{b}'\right] = 0 \quad (\text{C.6})$$

$$\Leftrightarrow \sum_{k=1}^K \alpha_k \beta_k \Sigma_{\mathbf{c}_i\mathbf{c}_i} \mathbf{b}' = 0 \quad (\text{C.7})$$

Or, (C.7) est vrai car  $\mathbf{b}$  est orthogonal au sous-espace formé par les  $\beta_k \Sigma_{\mathbf{c}_i\mathbf{c}_i} \forall k \in \{1, \dots, K\}$  donc  $\forall k \in \{1, \dots, K\}, \beta_k \Sigma_{\mathbf{c}_i\mathbf{c}_i} \mathbf{b}' = 0$ . Par ailleurs (C.6) est valable par l'hypothèse 2.

Ainsi, par l'hypothèse 2, la courbe de régression inverse  $E[\mathbf{c}_i|y]$  est supportée par le sous-espace généré par les  $\beta_k \Sigma_{\mathbf{c}_i\mathbf{c}_i}$  [Li, 2012].

Cette dernière assertion permet d'obtenir les vecteurs EDR, et reste valable pour toute transformation monotone  $T$  telle que II.12 puisse s'exprimer par l'équation (C.8).

$$T(y) = T \circ f(\beta_1\mathbf{x}, \beta_2\mathbf{x}, \dots, \beta_K\mathbf{x}, \epsilon) \quad (\text{C.8})$$

La courbe de régression inverse est alors exprimée par :

$$T(y) = E[\mathbf{x}|T(y)] \quad (\text{C.9})$$

L'introduction de la transformation monotone  $T$  autorise la simplification du calcul de la matrice de covariance II.15 page 47. En effet, la transformation à l'origine de la méthode SIR est la transformation par tranches où  $T$  est une discrétisation des valeurs de  $y$ . Lorsque  $T$  est une discrétisation de  $y$  en  $H$  tranches  $s_1, \dots, s_H$ , alors la matrice de covariance  $\Sigma_{\mathbf{c}_i\mathbf{c}_i}$  est exprimée par :

$$\Sigma_{\mathbf{c}_i\mathbf{c}_i} = \sum_{h=1}^H p_h m_h m_h' \quad (\text{C.10})$$

avec :

- $\Sigma_{\mathbf{c}_i\mathbf{c}_i}$  matrice de covariance de  $E[y|\mathbf{x}]$
- $H$  nombre de tranches définies par la transformation  $T$
- $p_h$  proportion des valeurs de  $y$  appartenant à la tranche  $s_h$
- $m_h$  moyenne des valeurs de  $\mathbf{x}_i$  telle que  $p_i \in s_h$

Les vecteurs de l'espace EDR sont alors les  $K$  vecteurs propres associés aux  $K$  plus grandes valeurs propres de  $\Sigma_{\mathbf{c}_i \mathbf{c}_i}$ .



## Annexe D

---

# Application du “Kernel Trick” à l’Analyse en Composante Principale

---

Dans la section II.C.1.a, il a été montré que la réduction possible de la dimension d’un nuage de points selon sa dispersion correspond à la décomposition en valeur propre de la matrice de covariance  $\mathbf{C}_{\mathbf{C}}$  de ce nuage (B.4). Or, l’expression de  $\mathbf{C}_{\mathbf{C}}$  nécessite le calcul des produits vectoriels de chaque élément du nuage de points  $\mathbf{C}$ .

Lors de la projection du nuage dans l’espace de redescription en utilisant  $\phi$ ,  $\mathbf{C}_{\mathbf{C}}$  devient :

$$\mathbf{C}_{\mathbf{C}} = \frac{1}{N} \sum_{i=1}^N \left( \phi(\mathbf{c}_i) - \overline{\phi(\mathbf{c})} \right) \left( \phi(\mathbf{c}_i) - \overline{\phi(\mathbf{c})} \right)^T \quad (\text{D.1})$$

La décomposition en valeurs propres dans l’espace de redescription est équivalente à résoudre [Scholkopf et al., 1996] :

$$\lambda \sum_{i=1}^N a_i (\phi(\mathbf{c}_k) \cdot \phi(\mathbf{c}_i)) \quad (\text{D.2})$$

$$= \frac{1}{N} \sum_{i=1}^N a_i \left( \phi(\mathbf{x}_k) \cdot \sum_{j=1}^N \phi(\mathbf{c}_j) \right) (\phi(\mathbf{c}_j) \cdot \phi(\mathbf{c}_i)) \quad (\text{D.3})$$

$$\Leftrightarrow N\lambda \mathbf{a} = \mathbf{K} \mathbf{a} \quad (\text{D.4})$$

avec :

- $N$  nombre de circuits
- $\lambda$  valeurs propres de  $\mathbf{C}_C$
- $K$  matrice du kernel telle que  $\mathbf{K}_{i,j} = (\phi(\mathbf{c}_i) \phi(\mathbf{c}_j)), \forall (i, j) \in \{1, \dots, N\}^2$
- $\mathbf{a}$  matrice de coefficients telle que  $\mathbf{u}_k = \sum_{i=1}^N a_i \phi(\mathbf{c}_i), \forall i \in \{1, \dots, N\}, \forall k \in \{1, \dots, K\}$

Le calcul de  $\mathbf{a}$  est réalisé par une condition de normalisation  $\lambda_k (\mathbf{a}_{\cdot,k} \cdot \mathbf{a}_{\cdot,k}) = 1$ .

De cette façon, le nuage de points d’origine  $\mathbf{C}$  est projeté dans un espace de redescription de dimension plus élevée, et la PCA est utilisée dans cet espace par la diagonalisation de  $\mathbf{K}$  (et la résolution de la condition ci-dessus). L’utilisation d’un noyau permet de ne plus calculer les projections pour chaque point du nuage de points  $\mathbf{C}$ , seul le noyau est utilisé. L’exemple classique suivant permet de mieux saisir le fonctionnement.

Soit la fonction de projection dans l’espace de redescription  $\phi$  telle que :

$$\begin{aligned} \phi : \quad \mathbb{R}^2 &\longrightarrow \mathbb{R}^3 \\ (x_1, x_2) &\longmapsto (x_1^2, x_2^2, \sqrt{2}x_1x_2) \end{aligned} \quad (\text{D.5})$$

En définissant le noyau par :

$$k(\mathbf{x}, \mathbf{y}) = (\mathbf{xy}^T)^2 \quad (\text{D.6})$$

$$\begin{aligned} &= (x_1y_1)^2 + (x_2y_2)^2 + 2x_1y_1x_2y_2 \\ &= \phi(\mathbf{x}) \phi(\mathbf{y})^T \end{aligned} \quad (\text{D.7})$$

Le noyau défini par l’équation (D.6) permet de ne pas calculer la projection (D.5) mais donne directement le produit scalaire des projections (D.7).

## Exemple d'utilisation d'algorithmes de sélection de caractéristiques

---

### E.A Exemple d'utilisation de RELIEF

Dans cet exemple, un nuage de points est construit de manière aléatoire (distribution normale) contenant 1000 échantillons de cinq caractéristiques. La classification binaire est construite en utilisant seulement les caractéristiques 1 et 3 par la fonction coût  $10 * c_{1,i} + 10 * c_{3,i} < 15$ . L'objectif de cet exemple est de détecter les 1<sup>ère</sup> et 3<sup>ème</sup> caractéristiques comme étant les plus pertinentes par application de RELIEF. En effet, d'après les définitions de la section II.C.2.a, les caractéristiques 1 et 3 sont des caractéristiques à pertinence forte, et toutes les autres caractéristiques ne sont pas pertinentes.

L'algorithme est exécuté avec  $M = 300$  et les résultats sont présentés sur la figure E.1. La limite  $\tau$  est calculée en utilisant

$$\tau = \min(W) + (\max(W) - \min(W)) * \text{tune} \quad (\text{E.1})$$

avec :

- $W$  vecteur des poids calculé par RELIEF
- $\text{tune}$  paramètre permettant de calculer  $\tau$  comme une proportion entre les valeurs extrêmes des poids  $W$ . Ici  $\text{tune} = 0.4$
- $\tau$  limite de pertinence des caractéristiques. Ici,  $\tau = 26.87$

La figure E.2 montre la convergence des poids au fur et à mesure de l'exécution de l'algorithme et des évaluations de  $W$ . Dans ce cas simple d'utilisation, l'algorithme identifie efficacement les contributions des caractéristiques 1 et 3 et les sépare clairement des autres caractéristiques. La figure E.2 montre que les poids sont clairement séparables à partir d'environ 30 itérations.

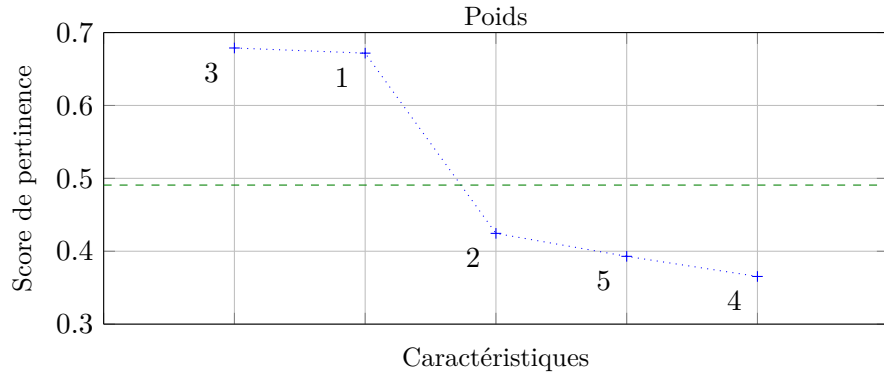


FIGURE E.1 – Résultats de RELIEF sur l'exemple d'application

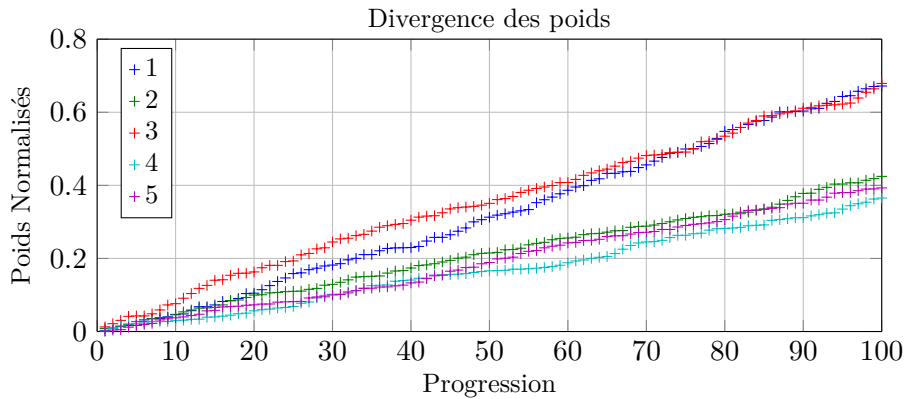


FIGURE E.2 – Divergence des poids de chaque caractéristique par RELIEF

## E.B Apports de FOCUS-2

L'exemple proposé dans cette section est extrait de [Almuallim and Dietterich, 1994]. Il permet de mettre en évidence l'apport de la queue d'exécution apportée par l'algorithme FOCUS-2.

Soit  $\mathbf{C} \in \mathbb{E}^N$  avec  $\mathbb{E} = \{0,1\}^P$  et  $P = 5$ .  $P$  représente le nombre de caractéristiques par mesure, et  $N$  est le nombre d'échantillons dans l'ensemble d'entraînement :  $c_i \in \mathbf{C}, i \in \{1, \dots, N\}$ .

D'après la recherche complète que réalise FOCUS (présenté en section II.C.2.d), toutes les combinaisons possibles de caractéristiques seront testées, par ordre de cardinal croissant, jusqu'à trouver une combinaison de caractéristiques qui ne présente pas de contradiction de classification dans l'ensemble d'entraînement. Dans cet exemple, le nombre maximal de test sera

$$T_{max} = 2^5 = 32 \quad (\text{E.2})$$

Dans la progression de l'algorithme, aucune information n'est extraite des échantillons de l'ensemble d'entraînement. En effet, considérons que le premier conflit soit  $G_1 = \{0, 0, 1, 1, 0\}$ . Ce conflit indique que l'ensemble pertinent de caractéristiques doit au moins contenir  $c_2$  et  $c_3$  pour vérifier au moins ce conflit  $G_1$ . Ainsi, il est inutile de tester les parties  $\{c_1\}$ ,  $\{c_4\}$ ,  $\{c_5\}$ ,  $\{c_1, c_4\}$ ,  $\{c_1, c_5\}$ ,  $\{c_4, c_5\}$  et  $\{c_1, c_4, c_5\}$ .

Il est ainsi possible, au fil de l'exécution de l'algorithme, d'enregistrer les parties qu'il n'est pas nécessaire de tester par cet argument. La progression de l'algorithme est sauvegardée dans la file  $Q$ .





---

# Machines à Vecteurs de Support

---

Cette annexe décrit la classification par Machine à Vecteurs de Support. Dans la suite, la séquence  $\{(\mathbf{c}_i, y_i)\}_{i=1}^N$  représente les  $N$  échantillons d'apprentissage, séparés en deux classes.  $y_i \in \{-1, +1\}$  représente la classe du circuit  $c_i$ .

## F.A Frontière et marge

Dans un premier temps, l'espace considéré est l'espace d'observation  $\mathbb{E} \subseteq \mathbb{R}^P$  de dimension  $P$ . Les  $N$  échantillons sont notés  $\mathbf{c}_i \in \mathbb{E}, i \in [1, N]$ , et ils sont classés entre échantillons de classe positive et échantillons de classe négative.  $\mathbf{w} \in \mathbb{E}$  est appelé *vecteur poids*. Il est alors possible de définir un hyperplan de  $\mathbb{E}$  par :

$$H_b = \left\{ \mathbf{x} \mid D(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + b = 0 \right\}, \forall \mathbf{x} \in \mathbb{R}^P \quad (\text{F.1})$$

où  $\mathbf{w}$  est orthogonal à l'hyperplan  $H_b$ . Cet hyperplan peut donc être considéré comme une frontière entre deux ensembles. L'objectif est de calculer  $(\mathbf{w}, b)$  pour définir la frontière  $H_b$  qui sépare l'espace entre les échantillons de classe positive et les échantillons de classe négative. En effet, l'appartenance d'un échantillon  $\mathbf{c}_i, i \in \{1, \dots, N\}$  à une classe est déduit du signe de  $D(\mathbf{c}_i)$ .

En définissant une marge  $M$  comme la plus petite distance entre la frontière et les échantillons, il vient :

$$\frac{y_i D(\mathbf{c}_i)}{\|\mathbf{w}\|} \geq M, \forall i \in \{1, \dots, N\} \quad (\text{F.2})$$

**Définition 12** (Vecteurs de support). *En utilisant les notations précédentes, les vecteurs de support sont les  $\mathbf{c}_k \in \mathbb{E}$  pour lesquels l'équation (F.2) est une égalité.*

L'objectif est alors de calculer  $\|\mathbf{w}\|$  qui maximise la marge  $M$ . Le vecteur  $\mathbf{w}$  peut par ailleurs être contraint en norme. En effet, il s'agit d'un vecteur directeur dont la norme ne change ni le problème ni la solution. Il est donc possible de le contraindre sous la forme  $M\|\mathbf{w}\| = 1$  de manière à transformer le problème de la recherche de frontière à marge maximale en un problème *minmax* :

$$\mathbf{w}^* = \max_{\mathbf{w}} \min_k (y_k D(\mathbf{c}_k) \|\mathbf{w}\|) \quad (\text{F.3})$$

Dès lors, maximiser la marge revient à minimiser  $\|\mathbf{w}\|$  sous la condition  $|D(c_k)| \geq 1$  :

$$\begin{aligned} \mathbf{w}^* &= \min_{\mathbf{w}} \|\mathbf{w}\|^2 & (\text{F.4}) \\ \text{t.q. } y_i D(\mathbf{c}_i) &\geq 1, \forall i \in \{1, \dots, N\} \end{aligned}$$

La marge maximale est alors  $M^* = \frac{1}{\|\mathbf{w}^*\|}$ .

La résolution de ce problème peut être réalisée par des méthodes numériques, mais n'est pas réalisable pour un espace  $\mathbb{E}$  de dimension élevée, ou lorsqu'il n'existe pas d'hyperplan permettant de séparer les données. Il faut dans ce cas introduire une résolution par une méthode à noyaux.

## F.B Utilisation des noyaux

L'utilisation des noyaux dans la Machine à Vecteurs de Support permet de calculer la frontière dans les cas où elle ne peut pas être représentée par un hyperplan, mais par une variété. La méthode la plus directe pour rechercher une frontière de la forme d'une variété est d'introduire une fonction  $\phi : \mathbb{E} \mapsto \mathcal{L}$  qui exprime les vecteurs de l'espace d'observation  $\mathbb{E}$  dans l'espace de redescription  $\mathcal{L}$ , de dimension  $L$ . Dès lors, l'équation (F.1) devient :

$$D(\mathbf{x}) = \mathbf{w}^T \phi(\mathbf{x}) + b \quad (\text{F.5})$$

La projection dans un espace de redescription, qui est de dimension plus élevée que l'espace d'observation, rend difficile, voire impossible, la résolution de la recherche de la frontière. Cependant, l'utilisation du “*kernel trick*” permet de simplifier la résolution à condition d'exprimer le noyau sous la forme (II.19) page 49, ce qui conduit à :

$$D(\mathbf{x}) = \sum_{i=1}^L \alpha_i k(\mathbf{x}, \mathbf{c}_i) + b \quad (\text{F.6})$$

où  $\alpha = (\alpha_1, \dots, \alpha_L)$  sont des paramètres à ajuster, tels que :

$$\mathbf{w} = \alpha^T \phi(\mathbf{x}) \quad (\text{F.7})$$

**Définition 13** (Paramètre direct, Paramètre dual). *En utilisant les notations de (F.7),  $\mathbf{w} = (w_1, \dots, w_N)$  est appelé paramètre direct.  $\alpha = (\alpha_1, \dots, \alpha_L)$  est appelé paramètre dual.*

Pour calculer les paramètres duaux, et donc les paramètres directs, la résolution par la méthode de Lagrange permet d'exprimer les  $\alpha$  comme multiplicateurs de Lagrange.

## F.C Résolution de la frontière à marge maximale

Le problème de minimisation sous contraintes de l'équation (F.4) est résolu par l'introduction des multiplicateurs de Lagrange (et plus précisément des conditions de Kuhn et Tucker) :

$$L(\mathbf{w}, \alpha) = \|\mathbf{w}\|^2 - \sum_{k=1}^L \alpha_k (y_k D(\mathbf{c}_k) - 1) \quad (\text{F.8})$$

$$\frac{\partial L}{\partial \mathbf{w}}(\mathbf{w}^*, \alpha) = 0 \quad (\text{F.9})$$

$$\frac{\partial L}{\partial \alpha}(\mathbf{w}, \alpha^*) = 0 \quad (\text{F.10})$$

Tout d'abord, (F.10) implique  $\|D(\mathbf{c}_k)\| - 1 = 0$  donc seuls les supports interviennent dans la solution (si elle existe). Pour les  $\mathbf{c}_k$  qui ne sont pas supports,  $\alpha_k^* = 0$ .

De plus, l'équation (F.9) implique

$$\begin{aligned} 2\|\mathbf{w}^*\| - \sum_{k=1}^L \alpha_k y_k \phi(\mathbf{c}_k) &= 0 \\ \Leftrightarrow \mathbf{w}^* &= \frac{1}{2} \sum_{k=1}^n \alpha_k^* y_k \phi(c_k) \end{aligned} \quad (\text{F.11})$$

L'équation (F.6) est alors réduite uniquement à la somme sur les supports  $\mathbf{c}_i$  :

$$D(\mathbf{x}) = \sum_i \alpha_i^* k(\mathbf{x}, \mathbf{c}_i) + b \quad (\text{F.12})$$

Par la suite, l'obtention des  $\alpha_i^*$  est réalisée par la minimisation de la quantité :

$$\begin{aligned} J(\alpha, b) &= \sum_{k=1}^n \alpha_k (1 - by_k) - \frac{1}{2} \alpha \mathbf{H} \alpha \quad (\text{F.13}) \\ H_{kl} &= y_k y_l k(\mathbf{c}_k, \mathbf{c}_l) \end{aligned}$$

Puis,  $b$  est calculé en sélectionnant deux vecteurs supports de classes différentes,  $\mathbf{c}_A$  et  $\mathbf{c}_B$ , et en définissant la frontière comme étant à la moitié de la distance

les séparant, ce qui donne :

$$b^* = -\frac{1}{2} \sum_{k=1}^n y_k a_k^* (k(\mathbf{c}_A, \mathbf{c}_k) + k(\mathbf{c}_B, \mathbf{c}_k)) \quad (\text{F.14})$$

D'autres méthodes sont proposées par [Hofmann et al., 2008] pour le calcul de  $b^*$ , comme l'optimisation de (F.13) en  $\alpha$  et en  $b$ .

De plus, [Ben-Hur and Weston, 2010] présente une méthode permettant d'augmenter la marge de classification en autorisant des erreurs de classification. Les auteurs introduisent des variables de détentes  $\xi$  dans la contrainte de l'équation (F.4) via  $y_i D(\mathbf{c}_i) \geq 1 - \xi_i, \forall i \in \{1, \dots, N\}$ .

---

# Classification par Arbre de Décision

---

La classification par arbre de décision est introduite dans la section II.D.1. L'algorithme d'induction d'un arbre de décision est rappelé ci-après :

---

**Algorithme 3** : Pseudo algorithme d'induction d'arbre de décision

---

```
input  :  $\mathbf{C}, \mathbf{y}$ 
output :  $T$  : Le model d'arbre de classification

1 Function split( $\mathbf{C}$  : Set of samples,  $\mathbf{y}$  : classes of every  $\mathbf{c}$ ) is
   /* Extract the best splitting feature  $F$  with limit  $f$  */
2    $(F, f) \leftarrow \text{bestSplitCriterion}(\mathbf{c}, \mathbf{y})$ ;
3    $T.\text{addNode}(Y)$ ;
   /* Split  $\mathbf{C}$  into two groups */
4    $(\mathbf{C}^+, \mathbf{y}^+) \leftarrow \{\mathbf{C} | \mathbf{C}(F) \geq f\}$ ;
5    $(\mathbf{C}^-, \mathbf{y}^-) \leftarrow \{\mathbf{C} | \mathbf{C}(F) < f\}$ ;
6   if  $\text{unique}(\mathbf{y}^+) > 1$  then
7      $\text{split}(\mathbf{C}^+, \mathbf{y}^+)$ 
8   else
9      $T.\text{addLeaf}(\text{unique}(\mathbf{y}^+))$ 
10  if  $\text{unique}(\mathbf{y}^-) > 1$  then
11     $\text{split}(\mathbf{C}^-, \mathbf{y}^-)$ 
12  else
13     $T.\text{addLeaf}(\text{unique}(\mathbf{y}^-))$ 
14  return  $T$ ;
15  $T = \text{split}(\mathbf{C}, \mathbf{y})$ ;
```

---

Cette annexe détaille les deux méthodes les plus répandues pour sélectionner les caractéristiques de séparation : la mesure d'entropie et l'évaluation de la fonction de Gini [Donald, 1994].

## G.A Séparation par mesure d'entropie

La sélection de caractéristiques de séparation par mesure d'entropie conduit à l'algorithme *ID3* et à ses descendants [Quinlan, 1986]. La séparation par entropie est réalisée pour conserver la proportion de classification qui apparaît dans les échantillons d'entraînement. En notant les deux classes possibles  $Y^+$  et  $Y^-$ , et  $y^+$  (resp.  $y^-$ ) le nombre d'échantillons d'entraînement classés  $Y^+$  (resp. classés  $Y^-$ ), alors un échantillon quelconque  $c$  a une probabilité d'appartenance à  $Y^+$  (resp.  $Y^-$ ) égale à  $\frac{y^+}{y^++y^-}$  (resp.  $\frac{y^-}{y^++y^-}$ ).

[Quinlan, 1986] introduit une mesure d'information par

$$I(y^+, y^-) = -\frac{y^+}{y^++y^-} \log\left(\frac{y^+}{y^++y^-}\right) - \frac{y^-}{y^++y^-} \log\left(\frac{y^-}{y^++y^-}\right) \quad (\text{G.1})$$

Par ailleurs, en notant une caractéristique  $F$ , dont les  $v$  valeurs possibles sont discrètes ( $F_1, \dots, F_v$ ), il est possible de décomposer les échantillons d'entraînement  $\mathbf{C}$  selon  $F$  par  $(C_1, \dots, C_v)$  où  $\forall i \in [1, v], C_i(F) = F_i$ . De plus, en notant  $y_i^+$  (resp.  $y_i^-$ ) le nombre d'échantillons de  $C_i$  appartenant à la classe  $Y^+$  (resp. appartenant à la classe  $Y^-$ ), alors l'espérance et le gain de  $F$  sont définis par [Quinlan, 1986] :

$$E(F) = \sum_{i=1}^v \frac{y_i^+ + y_i^-}{y^+ + y^-} I(y_i^+, y_i^-) \quad (\text{G.2})$$

$$G(F) = I(y^+, y^-) - E(F) \quad (\text{G.3})$$

La caractéristique utilisée pour séparer les échantillons peut naturellement être choisie comme étant celle qui maximise le gain  $G$  (ou qui minimise l'espérance d'information  $E$  puisque  $I(y^+, y^-)$  est le même pour toutes les caractéristiques).

L'algorithme propose une solution pour des caractéristiques discrètes (les branches découlant d'un nœud représentent alors chacune des valeurs que peut prendre  $F$ ). Une solution est aussi proposée pour les cas continus [Quinlan, 2006]. Un seuil est alors calculé pour séparer les caractéristiques de manière à augmenter le *rapport de gain* défini par :

$$\text{Info}(D) = -\sum_{j=1}^C p(D, j) \log(p(D, j)) \quad (\text{G.4})$$

avec :

- $C$  le nombre de classes dans les échantillons d'entraînement
- $D$  la partition des valeurs de la caractéristique  $F$
- $p(D, j)$  la proportion d'échantillons de  $D$  appartenant à la classe  $j$

## G.B Séparation par mesure de Gini

La mesure d'impureté de Gini représente la probabilité qu'un élément, choisi au hasard, soit mal classé lorsque la classification est aléatoirement choisie parmi un sous-ensemble de classes. Elle est définie sur un ensemble d'échantillons d'apprentissage  $\mathbf{T}$  par [Rokach and Maimon, 2010] :

$$I_G(\mathbf{T}) = 1 - \sum_i^C (p_i)^2 \quad (\text{G.5})$$

avec :

- $\mathbf{T}$  le sous-ensemble d'échantillons d'entraînement  $[T \times P]$
- $p_i$  la fréquence relative de la classe  $i$  dans  $\mathbf{T}$

La mesure d'impureté de Gini est utilisée dans le calcul d'un critère pour toutes les caractéristiques :

$$G_G(p_i, \mathbf{T}) = I_G(\mathbf{T}) - \sum_{\mathbf{T}_{i,j} \in \text{dom}(p_i)} \frac{|\mathbf{T}_{i,j}|}{|\mathbf{T}|} I_G(\mathbf{T}_{i,j}) \quad (\text{G.6})$$

avec :

- $G_G(p_i, \mathbf{T})$  la mesure de gain de Gini de la  $i^{\text{ème}}$  performance
- $p_i$   $i^{\text{ème}}$  caractéristique
- $\mathbf{T}_{i,j}$   $j^{\text{ème}}$  partition des valeurs de la caractéristique  $p_i$

Dans le cadre de l'application dans l'ALGORITHME 3, la meilleure caractéristique de séparation est celle dont le gain de Gini (G.6) est le plus faible. Puis l'ensemble des échantillons d'entraînement est subdivisé selon la caractéristique choisie, et l'algorithme itère jusqu'à ce que les échantillons d'entraînement soient correctement classés.





---

# Algorithmes dérivés de FOCUS

---

## H.A FOCUS-2

---

**Algorithme 4** : L'algorithme FOCUS-2 [Almuallim and Dietterich, 1994]

---

```
input  : C, y
output : A : Le plus petit ensemble pertinent

1 if Unique(y) then
2   return A = ∅;
   /* Extrait tous les conflits et initialise la queue de
      calcul */
3 G = GetAllConflicts(y);
4 Q = M∅,∅;
5 while True do
6   MA,B = Pop(Q);
7   Out = B;
8   i = 0;
   /* Cherche le conflit qui n'est couvert par aucune
      caractéristique de A et dont |Za - B| est minimal, Za
      étant le nombre de caractéristiques atteintes par le
      conflit a */
9   a = ExtractConflict(A, B, G);
10  for i < Size(a) do
11    x = (Za - B) {i};
12    if Test(A ∪ {x}) then
13      return A ∪ {x};
14    Push(Q, MA∪{x},Out);
15    Out = Out ∪ {x};
```

---

## H.B CDC-FOCUS

### H.B.1 Cœur de CDC-FOCUS

---

**Algorithme 5** : Algorithme CDC-FOCUS [Verdy et al., 2014]

---

```

input  : C : Matrice des échantillons
          y : Vecteur des classifications pour tous les échantillons
           $f_c(c_i) \forall i \in [1, O]$  : Fonction coût des caractéristiques
           $\Delta$  : Seuil introduit par C-Focus

output : Result : L'ensemble pertinent au coût le plus faible

1 Normalisation (C);
2 y  $\leftarrow$  SortedByCost (y,  $f_c$ );
   /* Trie des mesures par coût croissant */
3 H  $\leftarrow$  Horizon (C,  $f_c$ );           // Construit l'horizon
4 G  $\leftarrow$  ConflictMatrix (y);         // Construit les conflits
5 Q  $\leftarrow$   $\{\emptyset, \emptyset\}$ ;
6 horizonMax  $\leftarrow$  0;
7 Cost  $\leftarrow$  0;
8 Result  $\leftarrow$   $\emptyset$ ;
9 OUT  $\leftarrow$   $\emptyset$ ;
10 while True do
11   [A, B]  $\leftarrow$  Q.pop();
12   OUT  $\leftarrow$  B;
13    $c_t \leftarrow$  SearchConflict (A, B, G); // Recherche les conflits
   pouvant résoudre
   /* L'horizon maximal est-il atteint? */
14   if horizonMax  $\neq$  0 & card (A) + 1 > horizonMax then
15     return ;
   /* Teste les combinaisons de A et  $c_t$  */
16   for  $c_x \in c_t$  do
17     if  $C_S([A, c_x]) \leq$  Cost | Cost = 0 then
18       if isSufficient ( $[A, c_x]$ , C, y,  $\Delta$ ) then
19         Result  $\leftarrow$   $[A, c_x]$ ;
20         Cost  $\leftarrow$   $C_S$  (Result);
21         horizonMax  $\leftarrow$  updateHorizon (H, Cost);
22   Q.add( $[A, c_x]$ , OUT); // Alimente la FIFO
23   OUT  $\leftarrow$   $[OUT, c_x]$ ;

```

---

### H.B.2 Exemple de test de suffisance

---

**Algorithme 6** : Algorithme de test de suffisance
 

---

```

input   :  $A$  : un ensemble de caractéristiques dont il faut analyser la
           suffisance
            $C$  : le nuage de points
            $y$  : les classes des échantillons
            $\Delta$  : le seuil introduit par C-Focus

output : suffisant : Résultat du test de suffisance
1 def suffisant = isSufficuent ( $A, C, y$ ):
2   if Size( $A$ ) = 0 then // Si il n'y a aucune caractéristiques
   dans  $A$ , retourne faux
3   |   suffisant  $\leftarrow$  False;
4   |   return ;
5   if Size(Unique( $y(A)$ )) = 1 then // Si tous les échantillons
   de  $A$  sont dans la même classe, retourne vrai
6   |   suffisant  $\leftarrow$  True;
7   |   return ;
   /* Initialise la variable de retour à vrai          */
8   suffisant  $\leftarrow$  True;
9   for  $s \in C$  do // Pour tous les échantillons
10  |    $S+$  = SearchSamples( $s, A, C, \Delta$ )/* Extrait les
   échantillons qui sont proches de  $s$  d'après les
   caractéristiques  $A$                                 */
11  |    $y+$  =  $y(S+)$ /* Extrait les classes des échantillons
   extrait précédemment                                */
12  |   if Size(Unique( $y+$ ))  $\neq$  1 then // Si les échantillons ne
   sont pas tous dans la même classe,  $A$  n'est pas
   pertinent
13  |   |   suffisant  $\leftarrow$  False;
  
```

---



---

# Bibliographie

---

- [Abdallah et al., 2013] Abdallah, L., Stratigopoulos, H.-G., and Mir, S. (2013). True non-intrusive sensors for RF built-in test. *2013 IEEE International Test Conference (ITC)*, pages 1–10.
- [Akkouche, 2011] Akkouche, N. (2011). *Optimisation du test de production de circuits analogiques et RF par des techniques de modélisation statistique*. PhD thesis, Université de Grenoble, Grenoble.
- [Almuallim and Dietterich, 1991] Almuallim, H. and Dietterich, T. (1991). Learning with Many Irrelevant Features. *Proceedings of the 9th National conference on Artificial intelligence*, 2 :547–552.
- [Almuallim and Dietterich, 1994] Almuallim, H. and Dietterich, T. (1994). Learning boolean concepts in the presence of many irrelevant features. *Artificial Intelligence*, 69(1-2) :279–305.
- [Arauzo Azofra et al., 2003] Arauzo Azofra, A., Benitez Sanchez, J. M., and Castro Pena, J. L. (2003). C-focus : A continuous extension of focus. *Advances in Soft Computing*, pages 1–8.
- [Arinton and Korbicz, 2004] Arinton, E. and Korbicz, J. (2004). Dynamic High Order Neural Networks : Application for Fault Diagnosis. In Rutkowski, L., Siekmann, J. H., Tadeusiewicz, R., and Zadeh, L. A., editors, *Artificial Intelligence and Soft Computing - ICAISC 2004*, number 3070 in Lecture Notes in Computer Science, pages 145–150. Springer Berlin Heidelberg.
- [Ben-Hur and Weston, 2010] Ben-Hur, A. and Weston, J. (2010). A user’s guide to support vector machines. *Methods in molecular biology (Clifton, N.J.)*, 609 :223–39.
- [Bhattacharya and Chatterjee, 2004] Bhattacharya, S. and Chatterjee, A. (2004). A Built-In Loopback Test Methodology for RF Transceiver Circuits Using Embedded Sensor Circuits. In *13th Asian Test Symposium*, pages 68–73. IEEE.

- [Blumer et al., 1989] Blumer, A., Ehrenfeucht, A., Haussler, D., and Warmuth, M. K. (1989). Learnability and the Vapnik-Chervonenkis Dimension. *J. ACM*, 36(4) :929–965.
- [Boser et al., 1992] Boser, B. E., Guyon, I. M., and Vapnik, V. N. (1992). A Training Algorithm for Optimal Margin Classifiers. In *Proceedings of the 5th Annual ACM Workshop on Computational Learning Theory*, pages 144–152. ACM Press.
- [Bottoms, 2007] Bottoms, B. (2007). The International Roadmap for Semiconductors. IEEE.
- [Bousquet, 2001] Bousquet, O. (2001). Introduction aux "Support Vector Machines" (SVM). Available from : <http://www.math.u-psud.fr/~blanchard/gtsvm/intro.pdf>.
- [Breiman et al., 1984] Breiman, L., Friedman, J., Stone, C. J., and Olshen, R. A. (1984). *Classification and regression trees*. CRC press.
- [Brockman and Director, 1989] Brockman, J. and Director, S. (1989). Predictive subset testing : optimizing IC parametric performance testing for quality, cost, and yield. *IEEE Transactions on Semiconductor Manufacturing*, 2(3) :104–113.
- [Cho et al., 2008] Cho, C., Kim, D. D., Kim, J., Plouchart, J.-O., Lim, D., Cho, S., and Trzcinski, R. (2008). Decomposition and Analysis of Process Variability Using Constrained Principal Component Analysis. *IEEE Transactions on Semiconductor Manufacturing*, 21(1) :55–62.
- [Cybenko, 1989] Cybenko, G. (1989). Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals, and Systems*, 2(4) :303–314.
- [Daubert et al., 1988] Daubert, S., Vallancourt, D., and Tsividis, Y. (1988). Current copier cells. *Electronics Letters*, 24(25) :1560–1562.
- [Dhagat and Hellerstein, 1994] Dhagat, A. and Hellerstein, L. (1994). PAC Learning with Irrelevant Attributes. pages 64–74.
- [Donald, 1994] Donald, J. (1994). Rule induction-machine learning techniques. *Computing & Control Engineering Journal*, 6(October).
- [Dubois, 2011] Dubois, M. (2011). *Méthodologie d'estimation des métriques de test appliquée à une nouvelle technique de BIST de convertisseur SIGMA/DELTA*. PhD thesis, Université de Grenoble, Grenoble.
- [Friedman, 1991] Friedman, J. H. (1991). Multivariate Adaptive Regression Splines. *The Annals of Statistics*, 19(1) :1–67.
- [Friis, 1944] Friis, H. (1944). Noise Figures of Radio Receivers. *Proceedings of the IRE*, 32(7) :419–422.

- [Guestrin, 2007] Guestrin, C. (2007). SVMs, Duality and the Kernel Trick. Available from : <http://www.cs.cmu.edu/~guestrin/Class/15781/slides/kernels-learningtheory-annotated.pdf> [cited 2015-01-30].
- [Guyon, 1991] Guyon, I. (1991). Neural networks and applications tutorial. *Physics Reports*, 207(3–5) :215–259.
- [Haider et al., 2003] Haider, A., Bhattacharya, S., and Chatterjee, A. (2003). Automatic multitone alternate test-generator for rf circuits using behavioral models. In *International Test Conference, 2003. Proceedings. ITC 2003.*, volume 1, pages 665–673.
- [Halder et al., 2005] Halder, A., Bhattacharya, S., Srinivasan, G., and Chatterjee, A. (2005). A system-level alternate test approach for specification test of RF transceivers in loopback mode. In *18th International Conference on VLSI Design, 2005*, pages 289–294.
- [Hall et al., 2008] Hall, P., Park, B. U., and Samworth, R. J. (2008). Choice of neighbor order in nearest-neighbor classification. *The Annals of Statistics*, 36(5) :2135–2152.
- [Hastie et al., 2001] Hastie, T., Friedman, J., and Tibshirani, R. (2001). *The Elements of Statistical Learning*. Springer Series in Statistics. Springer, New York, NY.
- [Hoefflinger, 2012] Hoefflinger, B., editor (2012). *Chips 2020 : a guide to the future of nanoelectronics*. The frontiers collection. Springer, Heidelberg ; New York.
- [Hofmann et al., 2008] Hofmann, T., Schölkopf, B., and Smola, A. J. (2008). Kernel methods in machine learning. *The Annals of Statistics*, 36(3) :1171–1220.
- [Hopfield, 1982] Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences*, 79(8) :2554–2558.
- [Huss and Gyurcsik, 1991] Huss, S. D. and Gyurcsik, R. S. (1991). Optimal ordering of analog integrated circuit tests to minimize test time. In *Proceedings of the 28th conference on ACM/IEEE Design Automation Conference - DAC '91*, pages 494–499. ACM Press.
- [IEEE, 2005] IEEE (2005). IEEE Standard for Extensions to Standard Test Interface Language (STIL) (IEEE Std 1450-1999) for Semiconductor Design Environments. *IEEE Std 1450.1-2005*, pages 1–123.
- [Jacquelin et al., 2003] Jacquelin, E., Bennani, A., and Hamelin, P. (2003). Force reconstruction : analysis and regularization of a deconvolution problem. *Journal of Sound and Vibration*, 265(1) :81–107.



- [Jain et al., 1999] Jain, A. K., Murty, M. N., and Flynn, P. J. (1999). Data clustering : a review. *ACM Computing Surveys*, 31(3) :264–323.
- [Janecek et al., 2008] Janecek, A. G. K., Gansterer, W. N., Demel, M. A., and Ecker, G. F. (2008). On the Relationship Between Feature Selection and Classification Accuracy. *Journal of Machine Learning Research*, 4 :90–105.
- [John et al., 1994] John, G. H., Kohavi, R., and Pfleger, K. (1994). Irrelevant Features and the Subset Selection Problem. In Cohen, W. W. and Hirsh, H., editors, *International Conference on Machine Learning*, pages 121–129. Morgan Kaufmann Publishers.
- [Kira and Rendell, 1992] Kira, K. and Rendell, L. (1992). The feature selection problem : Traditional methods and a new algorithm. *Association for the Advancement of Artificial Intelligence (AAAI)*.
- [Kohavi and John, 1997] Kohavi, R. and John, G. H. (1997). Wrappers for feature subset selection. *Artificial intelligence*, 97(97) :273–324.
- [Kononenko, 1994] Kononenko, I. (1994). Estimating attributes : analysis and extensions of RELIEF. *Machine Learning : ECML-94*, 784.
- [Korbicz and Janczak, 1996] Korbicz, J. and Janczak, A. (1996). A neural network approach to identification of structural systems. In *Proceedings of the IEEE International Symposium on Industrial Electronics, 1996. ISIE '96*, volume 1, pages 98–103.
- [Kuentz and Saracco, 2010] Kuentz, V. and Saracco, J. (2010). Cluster-based Sliced Inverse Regression. *Journal of the Korean Statistical Society*, 39(2) :251–267.
- [Kupp et al., 2011a] Kupp, N., Stratigopoulos, H.-G., Drineas, P., and Makris, Y. (2011a). On proving the efficiency of alternative RF tests. In *2011 IEEE/ACM International Conference on Computer-Aided Design (ICCAD)*, pages 762–767.
- [Kupp et al., 2011b] Kupp, N., Stratigopoulos, H.-G., Drineas, P., and Makris, Y. (2011b). PPM-accuracy Error Estimates for Low-Cost Analog Test : A Case Study. In *2011 IEEE 17th International Mixed-Signals, Sensors and Systems Test Workshop*, pages 43–47.
- [Lettvin et al., 1959] Lettvin, J., Maturana, H., McCulloch, W., and Pitts, W. (1959). What the Frog’s Eye Tells the Frog’s Brain. *Proceedings of the IRE*, 47(11) :1940–1951.
- [Li, 2012] Li, K.-C. (2012). Sliced Inverse Regression for Dimension Reduction. *Journal of the American Statistical Association*, 86(414) :316–327.
- [Liu and Yu, 2005] Liu, H. and Yu, L. (2005). Toward integrating feature selection algorithms for classification and clustering. *IEEE Transactions on Knowledge and Data Engineering*, 17(4) :491–502.

- [Lupea et al., 2003] Lupea, D., Pursche, U., and Jentschel, H.-J. (2003). RF-BIST : loopback spectral signature analysis. In *Design, Automation and Test in Europe Conference and Exhibition (DATE2003)*, pages 478–483.
- [Milor and Sangiovanni-Vincentelli, 1990a] Milor, L. and Sangiovanni-Vincentelli, A. (1990a). Computing parametric yield accurately and efficiently. In *IEEE International Conference on Computer-Aided Design, 1990. ICCAD-90. Digest of Technical Papers*, pages 116–119.
- [Milor, 1998] Milor, L. S. (1998). A tutorial introduction to research on analog and mixed-signal circuit testing. *IEEE Transactions on Circuits and Systems II : Analog and Digital Signal Processing*, 45(10) :1389–1407.
- [Milor and Sangiovanni-Vincentelli, 1990b] Milor, L. S. and Sangiovanni-Vincentelli, A. (1990b). Optimal test set design for analog circuits. In *1990 IEEE International Conference on Computer-Aided Design. Digest of Technical Papers*, pages 294–297.
- [Nassery et al., 2012] Nassery, A., Byregowda, S., Ozev, S., Verhelst, M., and Slamani, M. (2012). Built-in-Self Test of transmitter I/Q mismatch using self-mixing envelope detector. In *2012 IEEE 30th VLSI Test Symposium (VTS)*, pages 56–61.
- [Neelamani et al., 2004] Neelamani, R., Choi, H., and Baraniuk, R. (2004). ForWaRD : Fourier-wavelet regularized deconvolution for ill-conditioned systems. *IEEE Transactions on Signal Processing*, 52(2) :418–433.
- [Negreiros et al., 2002] Negreiros, M., Carro, L., and Susin, A. (2002). A statistical sampler for a new on-line analog test method. In *On-Line Testing Workshop, 2002. Proceedings of the Eighth IEEE International*, pages 79–83.
- [Negreiros et al., 2004] Negreiros, M., Carro, L., and Susin, A. (2004). Low cost analogue testing of RF signal paths. In *Design, Automation and Test in Europe Conference and Exhibition, 2004. Proceedings*, volume 1, pages 292–297.
- [Niederreiter, 1988] Niederreiter, H. (1988). Low-discrepancy and low-dispersion sequences. *Journal of Number Theory*, 30(1) :51–70.
- [Ouvry et al., 2014] Ouvry, L., Masson, G., Pezzin, M., Piaget, B., Caillat, B., Bourdel, S., Dehaese, N., Fourquin, O., Gaubert, J., Meillere, S., and Vauche, R. (2014). A 4ghz CMOS 130 nm IR-UWB dual front-end transceiver for IEEE802.15 standards. In *2014 21st IEEE International Conference on Electronics, Circuits and Systems (ICECS)*, pages 798–801.
- [Parthasarathy and Aggarwal, 2003] Parthasarathy, S. and Aggarwal, C. (2003). On the use of conceptual reconstruction for mining massively incomplete data sets. *IEEE Transactions on Knowledge and Data Engineering*, 15(6) :1512–1521.

- [Patil et al., 2015] Patil, S., Ratiu, A., Morche, D., and Tsividis, Y. (2015). A 3–10fj/conv-step 0.0032mm<sup>2</sup> error-shaping alias-free asynchronous ADC. In *2015 Symposium on VLSI Circuits (VLSI Circuits)*, pages C160–C161.
- [Peres-Neto et al., 2005] Peres-Neto, P. R., Jackson, D. A., and Somers, K. M. (2005). How many principal components? Stopping rules for determining the number of non-trivial axes revisited. *Computational Statistics & Data Analysis*, 49(4) :974–997.
- [Peterson, 2009] Peterson, L. (2009). K-nearest neighbor. Available from : [http://www.scholarpedia.org/article/K-nearest\\_neighbor](http://www.scholarpedia.org/article/K-nearest_neighbor) [cited 2015-03-18].
- [Pichuka et al., 2007] Pichuka, C., Bapi, R. S., Bhagvati, C., Pujari, A. K., and Deekshatulu, B. L. (2007). A Tighter Error Bound for Decision Tree Learning Using PAC Learnability. *Proceedings of the 20th international joint conference on Artificial intelligence*, pages 1011–1016.
- [Quinlan, 1986] Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning*, 1(1) :81–106.
- [Quinlan, 2006] Quinlan, J. R. (2006). Improved Use of Continuous Attributes in C4.5. *Journal of Artificial Intelligence*, pages 77–90.
- [Rokach and Maimon, 2010] Rokach, L. and Maimon, O. (2010). Decision Tree. In *Data Mining and Knowledge Discovery Handbook*, pages 165–192. Springer US, Boston, MA.
- [Rémond and Nercessian, 2011] Rémond, E. and Nercessian, E. (2011). Mathematical approach based on a “Design of Experiment” to simulate process variations. *Design, Automation and Test in Europe (DATE2011)*.
- [Saerens et al., 2004] Saerens, M., Fouss, F., Yen, L., and Dupont, P. (2004). The principal components analysis of a graph, and its relationships to spectral clustering. *Machine Learning : ECML 2004*, (3201) :1–12.
- [Saracco et al., 1999] Saracco, J., Larramendy, I., and Aragon, Y. (1999). La regression inverse par tranches ou méthode SIR : presentation générale. *La revue de Modulad*.
- [Scholkopf et al., 1996] Scholkopf, B., Smola, A. J., and Muller, K.-R. (1996). Non-Linear Component Analysis as a Kernel Eigenvalue Problem. Technical 44, Max-Planck-Institut, Spemannstraße 38, 72076 Tübingen, Germany.
- [Singhee and Rutenbar, 2009] Singhee, a. and Rutenbar, R. (2009). Statistical Blockade : Very Fast Statistical Simulation and Modeling of Rare Circuit Events and Its Application to Memory Design. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 28(8) :1176–1189.

- [Singhee et al., 2008] Singhee, A., Singhal, S., and Rutenbar, R. a. (2008). Practical, fast Monte Carlo statistical static timing analysis : Why and how. In *2008 IEEE/ACM International Conference on Computer-Aided Design*, pages 190–195.
- [Spyronasios et al., 2011] Spyronasios, A., Abdallah, L., Stratigopoulos, H.-G., and Mir, S. (2011). On Replacing an RF Test with an Alternative Measurement : Theory and a Case Study. In *2011 Asian Test Symposium*, pages 365–370.
- [Srihari, 2004] Srihari, S. N. (2004). Computational Learning Theory (VC Dimension). Available from : <http://www.cedar.buffalo.edu/%7Esrihari/CSE574/Chap17/Chap17.Part2.pdf>.
- [Srinivasan et al., 2006] Srinivasan, G., Chatterjee, A., and Taenzler, F. (2006). Alternate loop-back diagnostic tests for wafer-level diagnosis of modern wireless transceivers using spectral signatures. In *24th IEEE VLSI Test Symposium, 2006. Proceedings*, pages 6 pp.–227.
- [Starzyk and El-Gamal, 1990] Starzyk, J. and El-Gamal, M. (1990). Artificial neural network for testing analog circuits. In *IEEE International Symposium on Circuits and Systems, 1990*, volume 3, pages 1851–1854.
- [Stehman, 1997] Stehman, S. V. (1997). Selecting and interpreting measures of thematic classification accuracy. *Remote Sensing of Environment*, 62(1) :77–89.
- [Stratigopoulos and Makris, 2008] Stratigopoulos, H.-G. and Makris, Y. (2008). Error Moderation in Low-Cost Machine-Learning-Based Analog/RF Testing. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 27(2) :339–351.
- [Stratigopoulos and Mir, 2010] Stratigopoulos, H.-G. and Mir, S. (2010). Fault diagnosis of analog circuits based on machine learning. In *2010 Design, Automation & Test in Europe Conference & Exhibition (DATE 2010)*, pages 1761–1766.
- [Stratigopoulos et al., 2009a] Stratigopoulos, H.-G., Mir, S., Acar, E., and Ozev, S. (2009a). Defect Filter for Alternate RF Test. *2009 14th IEEE European Test Symposium*, pages 101–106.
- [Stratigopoulos et al., 2009b] Stratigopoulos, H.-G., Mir, S., and Bounceur, A. (2009b). Evaluation of Analog/RF Test Measurements at the Design Stage. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 28(4) :582–590.
- [Stratigopoulos and Sunter, 2014] Stratigopoulos, H.-G. and Sunter, S. (2014). Fast Monte Carlo-Based Estimation of Analog Parametric Test Metrics. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 33(12) :1977–1990.

- [Sun et al., 2014] Sun, Y., Ma, L., Qin, N., Zhang, M., and Lv, Q. (2014). Analog filter circuits feature selection using MRMR and SVM. In *2014 14th International Conference on Control, Automation and Systems (ICCAS)*, pages 1543–1547.
- [Thiémard, 2001] Thiémard, E. (2001). An Algorithm to Compute Bounds for the Star Discrepancy. *Journal of Complexity*, 17(4) :850–880.
- [Valiant, 1984] Valiant, L. G. (1984). A theory of the learnable. *Communications of the ACM*, 27(11) :1134–1142.
- [Vapnik, 1995a] Vapnik, V. N. (1995a). Bounds on the Rate of Convergence of Learning Processes. In *The Nature of Statistical Learning Theory*, pages 65–87. Springer New York.
- [Vapnik, 1995b] Vapnik, V. N. (1995b). Introduction : Four Periods in the Research of the Learning Problem. In *The Nature of Statistical Learning Theory*, pages 1–14. Springer New York.
- [Vapnik, 1995c] Vapnik, V. N. (1995c). *The Nature of Statistical Learning Theory*. Springer, New York, NY.
- [Velasco-Medina et al., 1999a] Velasco-Medina, J., Rayane, I., and Nicolaidis, M. (1999a). AC/DC BIST for testing analog circuits. In *ASIC/SOC Conference, 1999. Proceedings. Twelfth Annual IEEE International*, pages 223–227.
- [Velasco-Medina et al., 1999b] Velasco-Medina, J., Rayane, I., and Nicolaidis, M. (1999b). On-line BIST for testing analog circuits. In *(ICCD '99) International Conference on Computer Design, 1999*, pages 330–332.
- [Verdy et al., 2015] Verdy, M., Morche, D., De Foucauld, E., Lesecq, S., Mallet, J.-P., and Mayor, C. (2015). Balancing test cost reduction vs. measurements accuracy at test time. In *New Circuits and Systems Conference (NEWCAS), 2015 IEEE 13th International*, pages 1–4.
- [Verdy et al., 2014] Verdy, M., Ratiu, A., Morche, D., De Foucauld, E., Lesecq, S., Mallet, J.-P., and Mayor, C. (2014). Cost-driven statistical analysis for selection of alternative measurements of analog circuits. In *2014 21st IEEE International Conference on Electronics, Circuits and Systems (ICECS)*, pages 104–107.
- [Weinzierl, 2000] Weinzierl, S. (2000). Introduction to Monte Carlo methods. *arXiv :hep-ph/0006269*. arXiv : hep-ph/0006269.
- [WU, 2007] WU, Q. (2007). Regularized sliced inverse regression for kernel models. pages 1–24.
- [Zhang and Styblinski, 1995] Zhang, J. C. and Styblinski, M. A. (1995). *Yield and Variability Optimization of Integrated Circuits*. Springer US, Boston, MA.

- [Zhou et al., 2015] Zhou, Z., Wen, C., and Yang, C. (2015). Fault Detection Using Random Projections and k-Nearest Neighbor Rule for Semiconductor Manufacturing Processes. *IEEE Transactions on Semiconductor Manufacturing*, 28(1) :70–79.

---

## Table des figures

---

I.1	Schéma de représentation du cycle de développement d'un circuit (reproduit depuis [Dubois, 2011]) . . . . .	3
I.2	Schéma de découpage et implémentation d'un circuit . . . . .	4
I.3	Évolution de la longueur du canal des transistors MOS de 1970 à 2020 [Hoefflinger, 2012] . . . . .	9
I.4	Évolution de la bande passante et de la vitesse de commutation de transistors NMOS de 1970 à 2020 [Hoefflinger, 2012] . . . . .	9
I.5	Illustration qualitative du compromis entre le coût du test et le coût des retours clients (d'après [Bottoms, 2007]) . . . . .	10
I.6	Schéma simplifié du test direct . . . . .	11
I.7	Schéma simplifié du test indirect . . . . .	12
I.8	Schéma simplifié du test indirect par BIST . . . . .	13
I.9	Schéma simplifié d'un Loopback . . . . .	14
I.10	Schéma simplifié de l'utilisation de la méthode BIMBO . . . . .	15
I.11	Schéma simplifié de l'ordonnancement des tests (cas des tests directs) . . . . .	15
I.12	Représentation des zones d'acceptabilité dans les domaines des paramètres technologiques, mesures et performances . . . . .	17
I.13	Élimination de mesure corrélées . . . . .	22
I.14	Conservation des mesures corrélées aux performances . . . . .	23
I.15	Sélection des points de mesure d'intérêt et construction d'un modèle de classification [Stratigopoulos and Makris, 2008] . . . . .	23
I.16	Vision Unifiée de la Problématique de construction de la procédure de test . . . . .	24
II.1	Vision Unifiée de la Problématique de construction de la procédure de test . . . . .	28
II.2	Acquisition d'information dans la Vision Unifiée . . . . .	31
II.3	Deux générations différentes d'échantillons . . . . .	35
II.4	Exemple de répartition de performances . . . . .	36

II.5 Organigramme du <i>Statistical Blockade</i> , d'après [Stratigopoulos and Sunter, 2014] . . . . .	37
II.6 Représentation de simulations utilisant le <i>Statistical Blockade</i> . . .	38
II.7 Histogramme d'une performance obtenue par <i>Statistical Blockade</i>	38
II.8 Illustration du comportement <i>Stop on Fail</i> en mesure . . . . .	40
II.9 Recherche de signaux d'intérêt dans la Vision Unifiée . . . . .	42
II.10 La <i>réduction de dimension</i> est une transformation . . . . .	43
II.11 La <i>sélection de caractéristiques</i> est une projection . . . . .	43
II.12 Exemple de représentation d'un nuage de points projetable sur une variété . . . . .	48
II.13 Algorithme <i>filtre</i> pour la sélection de caractéristiques (reproduit d'après [Liu and Yu, 2005]). $\mathbf{A}_{\text{best}}$ est le sous ensemble de caractéristiques le plus pertinent selon la mesure de pertinence $M$ . . . .	53
II.14 Algorithme <i>wrapper</i> de sélection de caractéristiques (reproduit d'après [Liu and Yu, 2005]). Dans ce cas, l'évaluation des sous ensembles est réalisée par un algorithme de classification externe. .	54
II.15 La modélisation dans la Vision Unifiée . . . . .	60
II.16 Exemple de classification par SVM . . . . .	64
II.17 Exemple illustrant l'impact du choix de $k$ . L'échantillon de test est le cercle vert. . . . .	65
II.18 Représentation d'un neurone [Guyon, 1991, Vapnik, 1995b] . . . .	69
II.19 Exemple de RNA . . . . .	70
II.20 Comparaison des erreurs de classifieurs . . . . .	71
II.21 Gestion de la précision dans la Vision Unifiée . . . . .	72
II.22 Construction et validation du modèle . . . . .	73
III.1 Les différents blocs de la Vision Unifiée . . . . .	84
III.2 La densité de probabilité d'un signal affecté par un bruit additif non corrélé est la convolution de leurs densités de probabilité respectives	95
III.3 Méthode d'évaluation de l'impact des incertitudes de mesure sur les métriques de test. <i>IFFT</i> signifie <i>Inverse Fast Fourier Transform</i> (inverse de la transformée de Fourier rapide) et <i>PDF</i> signifie <i>Probability Density Function</i> (Densité de probabilité). . . . .	96
III.4 Distribution théorique du SNR . . . . .	100
III.5 Schéma de l'évaluation du SNR. Les variations technologiques influent sur la puissance du bruit en sortie du dispositif. Le nombre de points de mesure utilisés pour le calcul de la FFT est source d'incertitude dans l'évaluation du SNR. . . . .	101
III.6 Dispersion des mesures sur un unique circuit, lorsque différentes quantités d'échantillons sont utilisées pour le calcul de la Transformée de Fourier . . . . .	101
III.7 Transformée de Fourier de la distribution du signal reconstruit par déconvolution, avec sa représentation réelle . . . . .	102
III.8 Comparaison des signaux reconstruit, d'origine et bruité . . . . .	102



III.9	Écart de rendement en fonction du nombre de points conservés . .	103
III.10	Évolution du nombre de faux négatifs suivant le modèle d'incertitude (III.14) . . . . .	103
III.11	Évolution du nombre de faux positifs suivant le modèle d'incertitude (III.12) . . . . .	104
III.12	Les différents blocs de la Vision Unifiée (rappel) . . . . .	106
IV.1	Diagramme de génération des simulations. Les rectangles verts sont les fichiers d'origines, les bleus sont les fichiers générés par les outils.	111
IV.2	Diagramme des simulations parallèles. Les rectangles verts sont les fichiers d'origines, les bleus sont les fichiers générés par les outils. .	112
IV.3	Gain en vitesse d'exécution pour une fonctionnalité implémentée et wrappée en langage C plutôt qu'en Matlab . . . . .	113
IV.4	Représentation graphique des mesures de test . . . . .	114
IV.5	Représentation graphique d'une base de données reconstruite à partir de mesures . . . . .	115
IV.6	Diminution de la dimension du problème par étude des corrélations	116
IV.7	Schéma du LNA . . . . .	118
IV.8	Algorithmes utilisés pour traiter l'exemple du LNA . . . . .	120
IV.9	Comparaison des inducteurs, selon la précision et les faux-positifs .	121
IV.10	Précision et faux positifs des algorithmes de classification suivant le coût total des observations retenues par CDC-FOCUS . . . . .	122
IV.11	Exemple de détermination du seuil maximal pour deux caracté- ristiques. Les croix vertes sont les échantillons "fonctionnels", les rouges sont ceux "défectueux". Les spécifications sont les lignes bleues. La plus petite distance entre deux échantillons de classes différentes autour des spécifications est matérialisée par un trait noir dans les loupes. . . . .	123
IV.12	Schéma du convertisseur analogique numérique en temps continu, reproduit depuis [Patil et al., 2015] . . . . .	125
IV.13	Métriques des algorithmes de classification pour l'ADC-CT lorsque le SFDR est présent . . . . .	127
IV.14	Métriques des algorithmes de classification pour l'ADC-CT lorsque le SFDR est absent de la base de données . . . . .	129
IV.15	Évolution des métriques de test en fonction du seuil de CDC-FOCUS. La limite de seuil est matérialisée par le trait discontinu noir. . . .	130
IV.16	Précision et faux positifs des algorithmes de classification suivant le coût des observations retenues par CDC-FOCUS . . . . .	134
IV.17	Évolution de la précision en fonction du seuil de CDC-FOCUS . .	135
IV.18	Évolution du coût des caractéristiques sélectionnées en fonction du paramètre de CDC-FOCUS . . . . .	136
V.1	Rappel de la méthodologie proposée . . . . .	140

<i>Table des figures</i>	185
E.1 Résultats de RELIEF sur l'exemple d'application . . . . .	158
E.2 Divergence des poids de chaque caractéristique par RELIEF . . . . .	158

## Résumé

Les circuits intégrés sont de plus en plus performants et les tâches qu'ils accomplissent toujours plus complexes. L'évolution des technologies permet une intégration poussée, et une diminution de leur consommation. Cependant, leur complexité rend difficile le contrôle des rendements de production. Le test complet des circuits est alors nécessaire pour assurer la vente de circuits fonctionnels, et rejeter les circuits qui ne respectent pas les spécifications. Le test complet est cependant cher en raison des fréquences de fonctionnement et du nombre important de signaux à analyser. Notre problématique se place dans le cadre du test des circuits analogiques et l'étude des méthodes permettant d'assurer les métriques de test des circuits à coût minimal.

Dans nos travaux, nous mettons en avant des scénarios qui répondent à des problèmes différents mais qui pourtant trouvent des éléments de solution communs. Nous proposons une architecture de solution commune que nous appelons la "vision unifiée". Une étude des outils de traitement de données nous permet d'identifier les briques utilisables dans notre vision unifiée. Nous nous intéressons à un algorithme de sélection de caractéristiques, à savoir C-FOCUS, pour la sélection des signaux d'intérêts dans le cadre du test. Afin d'assurer le coût minimal du test, nous proposons une version modifiée de cet algorithme, CDC-FOCUS, où nous incluons une fonction coût. Suite à la sélection des signaux d'intérêts, nous utilisons des algorithmes de l'état de l'Art pour construire des modèles de classification. Nous nous intéressons à la précision de notre méthode et nous étudions les matrices de confusion des différents algorithmes de classification que nous utilisons. De plus, nous réalisons une étude permettant de relier l'impact des incertitudes de mesures sur les métriques de test. Enfin, nous valorisons nos résultats sur des cas concrets de circuits, et ce dans différents cas : en simulations mais aussi par l'utilisation de bases de données issues de mesures de circuits.

**Mot clés** Test de circuit, sélection de caractéristiques, réduction de coût de test, classification

## Summary

Integrated circuits are more and more powerful, and the tasks they address are more and more complex. Technology evolution allows the increase of their integration while reducing their consumption. However, their functioning is getting complex and leads to a complicated control of production yield. It is then mandatory to ensure selling functional circuits, and rejecting out-of-specification circuits. Nevertheless, the complete test is expensive due to the operating frequency and the number of signals to be measured. Our study work is in the context of analog circuits test and the study of methods ensuring test metrics at minimal cost.

Here, we draw scenarios which different issues but expressing a common solution. We suggest a common solution architecture called "United Vision". A study of data mining tools lets us identify bricks usable into our method. We are investigating a feature selection algorithm C-FOCUS to select most interesting signals for test. In order to ensure minimal test cost, we propose a new version of this algorithm, CDC-FOCUS, into which we implement a cost function. After feature selection, we use state of the Art algorithms to build classification models. As we are interested into the accuracy of our method, we study the confusion matrices of the classification algorithms we use, and we conduct a study to rely measurement uncertainty to test metrics.

Last but not least, we exemplify the powerfulness of our proposal methodology using several circuits in different situations : using simulation and real measurement data base.

**Keywords** Circuit testing, feature selection, test cost reduction, classification