Systèmes de numérisation hautes performances à base de modulateurs sigma delta passe-bande

Ali Beydoun

ÉCOLE
SUPÉRIEURE
D'ÉLECTRICITÉ

Thèse

présentée pour obtenir le grade de

DOCTEUR EN SCIENCES DE l'UNIVERSITÉ PARIS XI ORSAY

Spécialité : électronique

par

Ali BEYDOUN

SYSTÈME DE NUMÉRISATION HAUTES PERFORMANCES À BASE DE MODULATEURS SIGMA-DELTA PASSE-BANDE

soutenue le 27/05/08 devant le jury composé de :

MM.	Jean-Paul Gilles	Président
	Maher Kayal	Rapporteurs
	Dominique Dallet	
	Patrick Loumeau	Examinateurs
	Andreas Kaiser	
	Philippe Bénabès	Directeur de thèse

SUPÉLEC

Thèse préparée au Département Signaux et Systèmes Électroniques

À ma mère,

merci pour ton soutien,
merci pour ta confiance,
rien n'aurait été possible sans toi.

Remerciements

Ce travail de thèse a été réalisé au département Signaux et Systèmes Électroniques (SSE) de SUPÉLEC. Je remercie Monsieur Jacques Oksman, ancien chef du département, pour son accueil au sein de l'équipe. Je remercie bien sûr également Monsieur Gilles Fleury qui a plus tard pris le relais à ce poste.

Je remercie Monsieur Jean-Paul Gilles, professeur à l'université Paris XI, d'avoir présidé mon jury de thèse ainsi que Messieurs Maher Kayal, professeur à l'EPFL, et Dominique Dallet, professeur des universités à l'ENSEIRB, d'avoir accepté de rapporter sur mon travail de thèse. Je remercie également Messieurs Patrick Loumeau, professeur à l'institut TELECOM Paris-Tech et Andreas Kaiser directeur de recherche CNRS à l'IEMN d'avoir fait partie du jury en qualité d'examinateurs.

J'exprime toute ma gratitude à mon directeur de thèse Monsieur Philippe Bénabès, professeur au département SSE, qui a su m'orienter dans mon travail, se montrant disponible et me laissant une grande liberté dans mes recherches. J'ai particulièrement apprécié son talent et son honnêteté scientifiques.

En outre, je souhaite remercier toutes les personnes du département qui ont, de près ou de loin, contribué à ce travail de thèse. Je remercie Richard Kielbasa, SylvieGuessab, Jérôme Juillard, Caroline Lelandais-Perrault, Morgan Roger et Alain Bonnoît d'avoir apporté les nécessaires critiques constructives à mon manuscrit.

Un grand merci également à Messieurs Francis Trélin et Luc Batalie pour leur incontournable soutien technique. Je remercie Fabienne Suraud pour sa gentillesse et son efficacité et je remercie Karine el Rassi également pour sa gentillesse et son aide appréciable dans la préparation de la soutenance. Merci aussi à toutes les personnes du laboratoire et en particulier les thésards : Sorore, Tudor, Esmaiel, Yoan, Davud, Alexia, Mohammad, Rawad, et Zhiguo. Parmi ceux-là, je remercie plus particulièrement Emilie, Karine et Hassan pour la préparation du pot. Un remerciement spécial à Emilie pour son amitié et son soutien.

Je remercie encore une fois Monsieur Patrick Loumeau de m'avoir accordé sa confiance à la fin de mon travail de thèse. Je suis par ailleurs très heureux d'avoir intégré son équipe pour poursuivre mon parcours scientifique.

Je pense enfin à toute ma famille, et en particulier mes parents, ma sœur, mes frères et leurs épouses, qui m'ont donné l'amour, le courage et les moyens de faire cette thèse.

Résumé

Les systèmes de communications numériques mobiles tendent à intégrer de plus en plus d'applications (GSM, radio, TV, GPS, etc.) tout en fonctionnant sur plusieurs normes. Cette évolution implique une reconfigurabilité en ligne des récepteurs à l'aide d'une programmation logicielle justifiant le terme de radio-logicielle.

Par ailleurs, les normes de communication actuelles exigent des débits élevés. Les bandes de fréquence nécessaires sont donc étendues (jusqu'à plusieurs centaines de mégahertz). Ainsi, les systèmes de réception doivent être à très large bande.

La reconfigurabilité logicielle implique la numérisation des signaux au plus près de l'antenne, les hauts débits imposent une large bande passante. Une solution pour répondre à ces exigences est l'utilisation de convertisseurs analogique-numérique à base de modulateurs sigma-delta en parallèle. Parmi les architectures possibles, on compte : l'architecture à entrelacement temporel, l'architecture à base de modulation de *Hadamard* et l'architecture à décomposition fréquentielle. Ces architectures sont susceptibles de traiter toute la bande de fréquences possible. Cependant, pour une norme donnée, le bon fonctionnement du récepteur ne nécessite pas la conversion de la bande totale à la résolution maximale. La largeur de bande pourra être adaptée au signal à convertir.

Au cours de ce travail de thèse, nous avons proposé une nouvelle architecture de numérisation large bande constituée de modulateurs sigma-delta en parallèle, fonctionnant sur le principe de la décomposition fréquentielle FBD (*Frequency Band Decomposition*). Les modulateurs sont de type passe-bande à temps continu afin de permettre le fonctionnement à des fréquences élevées. Pour la partie numérique, nous avons développé un système de reconstruction du signal numérique adapté à la sortie des différents modulateurs.

L'incertitude due aux dispersions technologiques dans la réalisation de circuits analogiques est l'une des causes de la dégradation de la précision des modulateurs sigma-delta à temps continu. Afin d'adapter l'architecture aux imperfections de la partie analogique, nous proposons une modification de cette architecture en ajoutant deux modulateurs supplémentaires (EFBD *Extended Frequency Band Decomposition*). Grâce à cette architecture à bande étendue, combinée à des algorithmes de calibration, nous corrigeons les erreurs sur le module et la phase introduites par les dispersions analogiques.

Finalement, nous avons implanté le traitement numérique dans une technologie CMOS 0.12 μm afin d'évaluer la surface nécessaire pour le traitement numérique.

Cette étude théorique a permis de proposer des solutions nouvelles de conversion large bande et de les valider en vue d'une implémentation future sous forme intégrée.

Abstract

Mobile Communication systems tend to integrate more and more applications (GSM, radio, TV, GPS, etc.) and different standards (GSM, UMTS, WIMAX,..). This evolution requires a flexible receiver able, with a single channel, to deal with each different standard and application. The principle of such a receiver is based on the concept of the Software Radio.

The basic idea of the software radio is to integrate the analog-to-digital converter in the channel receiver directly after the antenna. This allows the receiver to adapt itself to different standards by reprogramming the functionality of all digital components in the channel receiver.

However, the current standard communications require high flow, so the useful signal frequency bands must be extended (up to several hundred megahertz). Therefor, the A/D bandwidth must be expanded.

One way to meet these requirements is the use of analog-to-digital converters based on parallel sigma-delta modulators. Three architectures were proposed on the state of the art based on this principle : Time Interleaved Sigma-Delta ($TI\Sigma\Delta$), Parallel Sigma-Delta ($\Pi\Sigma\Delta$) based on *Hadamard* modulation and Frequency Band Decomposition (FBD). These architectures convert the entire frequency band. However, for multistandard applications, a useful signal has a limited bandwidth and thus the conversion of the entire frequency band is not optimal.

This thesis proposes a new architecture for bandpass A/D converter using parallel band pass sigma-delta modulator based on the principle of the frequency band decomposition. We have used continuous time modulators to reach the high operating frequency. Moreover, a digital reconstruction system was proposed to reconstruct the digital input signal using all modulators output.

Technological dispersions on analog components decrease considerably the expected resolution of the converter. Actually, they shift resonator central frequencies of the modulator from their nominal value. This leads to mismatch the digital reconstruction system already calibrated to work with nominal values. In order to overcome this problem, the idea is to extend the usual FBD architecture by adding two additional modulators (EFBD Extended Frequency Band Decomposition). The EFBD architecture allows a 5% relative error on central frequencies without a large degradation of the resolution. Moreover, three calibration algorithms were developed to achieve the expected resolution and correct mistakes on the amplitude and the phase with the new configuration (EFBD)

Finally, the digital reconstruction system was implemented in 0.12 μm CMOS technology in order to evaluate their performances in term of area and maximum operating frequency.

Table des matières

Liste des abréviations

Pour des raisons de lisibilité, la signification d'une abréviation ou d'un acronyme n'est souvent rappelée qu'à sa première apparition dans le texte d'un chapitre. Par ailleurs, puisque nous utilisons toujours l'abréviation la plus usuelle, il est fréquent qu'il s'agisse du terme anglais.

ADC	*Analog-to-Digital Converter*
AM	*Amplitude Modulation*
AOP	Amplificateur OPérationnel
ARMAX	*Auto Regressive Moving Average with eXternal inputs*
ASIC	*Application-Specific Integrated Circuit*
BP	*Bandpass*
BW	*Bandwidth*
CAN	Convertisseur Analogique-Numérique
CDMA	*Code Division Multiple Access*
CMOS	*Complementary Metal Oxide Semiconductor*
CNA	Convertisseur Numérique-Analogique
CT	*Continuous Time*
CTΣΔ	*Continuous Time Sigma Delta*
DAC	*Digital-to-Analog Converter*
DECT	*Digital Enhanced Cordless Telephone*
DEM	*Dynamic Element Matching*
DR	*Dynamic Range*
DSP	*Digital Signal Processors*
DT	*Discret Time*
EDGE	*Enhanced Data Rates for* GSM *Evolution*
EFBD	*Extended Frequency Band Decomposition*
ENOB	*Effective Number Of Bits*
FBD	*Frequency Band Decomposition*
FIR	*Finite Impulse Response*
FM	*Frequency Modulation*
GFSK	*Gaussian Frequency Shift Keying*
GMSK	*Gaussian Minimum Shift Keying*
GPRS	*General Packet Radio Service*
GPS	*Global Positioning System*
GSM	*Global System for Mobile Communications*
HZ	*Half-delay Zero*

IF	*Intermediate Frequency*
LP	*Lowpass*
LNA	*Low Noise Amplifier*
LSB	*Less Significant Bit*
MASH	*Multi stAge noise SHaping*
MOSFET	*Metal Oxide Semiconductor Field-Effect Transistor (ou* MOS*)*
MSB	*Most Significant Bit*
MSCL	*Multi Stage Closed Loop modulators*
NRZ	*Non Return to Zero*
NTF	*Noise Transfert Function*
OFDM	*Orthogonal Frequency Division Multiplexing*
OSR	*Oversampling Ratio*
ΠΣΔ	*Parallel Sigma Delta*
PEM	*Prediction Error identification Method*
PIFA	*Planar Inverted Folded Antenna*
PSD	*Power Spectral Density*
QAM	*Quadrature Amplitude Modulator*
QPSK	*Quadrature Phase-Shift Keying*
RF	*Radio Frequency*
RPEM	*Recursive Prediction Error identification Method*
RZ	*Return to Zero*
SAH	*Sample-And-Hold*
SDR	*Software Defined Radio*
SFDR	*Spurious Free Dynamic Range*
SNDR	*Signal to Noise and Distortion Ratio*
SNR	*Signal to Noise Ratio*
SR	*Software Radio*
STF	*Signal Transfer Function*
THD	*Total Harmonic Distortion*
TIΣΔ	*Time-Interleaved Sigma Delta*
UMTS	*Universal Mobile Telecommunications System*
WIMAX	*Worldwide Interoperability for Microwave Access*
WI-FI	*WIreless FIdelity*

Introduction générale

Contexte

Le marché des télécommunications offre de nos jours de plus en plus de services et de fonctionnalités nécessitant le recours à des normes de réception variées. Outre la multiplication de ces normes, l'élargissement des bandes passantes de transmission conduit à la nécessité de concevoir des récepteurs qui soient multistandard mais aussi large bande.

Les architectures de réception actuelles manquent de flexibilité pour l'adaptation à de nouveaux services ou de nouvelles normes. En effet, chaque norme impose des caractéristiques techniques (largeur de bande des filtres analogiques, fréquence de démodulation,etc.) sur le récepteur. Un changement de la norme impose le remplacement des différents blocs dans l'architecture. Par conséquent, pour chaque norme de communication, nous avons besoin d'une architecture de réception indépendante, ce qui augmente la surface et le coût de fabrication d'un récepteur multistandard.

La solution à ce problème consiste à déplacer, dans une architecture de réception, la réalisation des blocs matériels analogiques dans le domaine numérique permettant ainsi une reconfiguration complète de l'architecture de réception. C'est le principe de la radio-logicielle (« *Software Radio* »). Ce principe a vu le jour à la fin des années 70 dans le cadre de la recherche militaire, a ensuite été présenté par J. Mitola en 1992 pour des applications grand public. Le schéma synoptique de cette approche idéale est présenté sur la figure suivante :

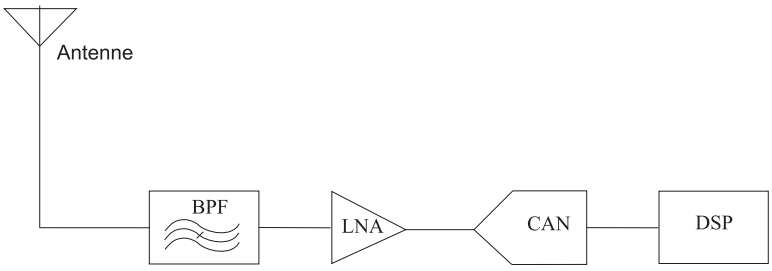

Fig. 1 – Chaîne théorique de réception « radio-logicielle ».

Cette approche se base sur la numérisation directe des signaux après l'antenne. Ensuite, les signaux numérisés sont traités par un ensemble de processeurs intégrant les fonctions appropriées à chaque norme. L'intérêt d'une telle architecture réside dans la flexibilité qu'elle offre sur la reconfigurabilité à de nouvelles normes via des mises à jour de logiciels gérant la partie numérique sans avoir recours à des modifications matérielles. Cette solution permet une adaptation facile aux évolutions technologiques et permet également une meilleure interopérabilité des systèmes de radiocommunication entre les différents standards de communication.

La numérisation des signaux dans le domaine RF (Radio Fréquence) impose au convertisseur analogique numérique d'avoir :

1. une fréquence de fonctionnement élevée,
2. un fonctionnement passe-bande et large bande,
3. une grande dynamique (écart entre la puissance minimale et la puissance maximale) du signal en entrée car le récepteur doit traiter plusieurs canaux (plusieurs signaux utiles). Par conséquent, le nombre de bits de quantification nécessaires pour coder l'échantillon doit être important.

Aujourd'hui, un convertisseur vérifiant les contraintes ci-dessus n'existe pas. Parmi les convertisseurs classiques, les convertisseurs à base de modulateurs sigma-delta ($\Sigma\Delta$) présentent un intérêt notable. En effet, ils assurent une grande précision à fréquence centrale très élevée avec peu de composants. Cependant, la largeur de bande de conversion est très petite au regard des bandes passantes des nouvelles normes. L'une des voies de recherche pour l'élargissement de la bande de fonctionnement et l'augmentation de la dynamique est la mise en parallèle de modulateurs $\Sigma\Delta$. Plusieurs architectures, basées sur ce principe, ont été proposées [1, 2, 3, 4] :

- l'architecture à entrelacement temporel TI$\Sigma\Delta$ (*Time-Interleaved Sigma Delta*),
- l'architecture à base de modulation de *Hadamard* $\Pi\Sigma\Delta$ (*Parallel Sigma Delta*),
- l'architecture à base de décomposition fréquentielle FBD (*Frequency Band Decomposition*).

Malgré leur simplicité d'implantation, les deux premières architectures sont très sensibles aux erreurs de gain et de décalage en tension dues aux imperfections des composants analogiques. Ces erreurs génèrent des raies dans le spectre de bruit de quantification et dégradent par conséquent la performance du convertisseur. Plusieurs méthodes de calibration ont été proposées dans [5, 6, 7] pour atténuer leurs effets. À la différence des deux premières architectures, l'architecture FBD est moins sensible aux erreurs de gain et de décalage en tension. En effet, elles n'introduisent pas de raies spectrales mais elles se manifestent par des erreurs de phase qui se répercutent sous forme d'ondulations sur le spectre du signal utile et ne dégradent pas la performance. Cependant, cette architecture est plus complexe à réaliser et exige des ressources matérielles importantes. Les trois architectures citées ci-dessus proposent la conversion de toute la plage fréquentielle. Cette solution n'est pas optimale car dans un récepteur multistandard, chaque norme se situe dans une bande limitée de plage fréquentielle.

Sujet de la thèse

C'est dans le cadre de cette nécessité actuelle de trouver des structures de réception multistandard large bande que s'inscrit ce travail de thèse. Nous nous sommes intéressés plus particulièrement à la recherche d'une nouvelle structure de convertisseurs analogique-numérique (CAN) permettant la conversion des signaux passe-bande à très hautes fréquences. Pour cela nous avons exploré la voie des structures à décomposition fréquentielle (FBD) en raison de leur robustesse aux imperfections analogiques. Nous avons proposé une nouvelle architecture de type FBD qui permet la conversion d'une bande limitée. Le principe de cette architecture consiste à placer N modulateurs $\Sigma\Delta$ à temps continu équi-répartis dans la bande du signal utile. Chaque modulateur traite une bande de $\frac{1}{N}$ de la bande utile. Nous avons validé le fonctionnement de cette architecture par simulation sur un exemple concret.

La principale faiblesse de cette architecture est le décalage des fréquences centrales des modulateurs en raison de dispersions technologiques dans la réalisation des filtres à temps continu. Ce décalage dégrade la performance (rapport signal sur bruit) de l'architecture car il désadapte

le traitement numérique en aval des modulateurs. Afin de résoudre ce problème, nous avons proposé d'étendre l'architecture FBD en ajoutant un modulateur à chaque extrémité de la bande du signal utile. L'architecture obtenue est l'architecture EFBD (*Extended Frequency Band Decomposition*). Nous avons proposé également des algorithmes de calibration permettant d'adapter le traitement numérique dédié à la reconstruction du signal utile pour compenser les erreurs analogiques. Ces algorithmes se caractérisent par leur simplicité d'implantation et leur fonctionnement en temps réel. Finalement, nous avons codé le traitement numérique au niveau portes logiques en langage VHDL et nous avons procédé à la synthèse avec une technologie CMOS 0.12 μm.

Plan du mémoire

Le mémoire se compose de cinq chapitres.

Le chapitre 1 est consacré à l'étude des architectures de réception de radiocommunication multistandard et aux problèmes technologiques à résoudre pour les mettre en œuvre. L'intérêt d'un convertisseur analogique-numérique large bande fonctionnant à très haute fréquence dans un tel récepteur est démontré. Nous nous sommes intéressés aux CAN à base de modulateurs $\Sigma\Delta$ en raison de leur précision élevée et leur faible surface de réalisation. Cependant, leur bande de fonctionnement étant restreinte, nous présentons à la fin de ce chapitre une voie de recherche basée sur le parallélisme, destinée à augmenter la bande de fonctionnement des modulateurs actuels.

Le chapitre 2 expose le principe de parallélisme pour l'augmentation des bandes de fonctionnement des CAN classiques. Nous évoquons plus particulièrement trois architectures à base de modulateurs $\Sigma\Delta$: l'architecture à entrelacement temporel TI$\Sigma\Delta$ (*Time Interleaved Sigma-Delta*), l'architecture à base de modulation de *Hadamard* $\Pi\Sigma\Delta$ (*Parallel Sigma-Delta*) et l'architecture à base de décomposition fréquentielle FBD (*Frequency Band Decomposition*). Ces trois architectures font l'objet d'une étude sur leur principe de fonctionnement et leur robustesse en termes de puissance de bruit en sortie vis-à-vis des imperfections des composants analogiques constituant le modulateur.

Le chapitre 3 propose une nouvelle architecture de CAN passe-bande à base de décomposition fréquentielle. Cette architecture est constituée de modulateurs $\Sigma\Delta$ à temps continu. Une étude théorique est développée dans ce chapitre permettant d'estimer la performance de cette architecture, puis son fonctionnement est validé par simulation sur un exemple concret. Bien que les erreurs dues aux imperfections analogiques n'introduisent pas de non-linéarité (raies spectrales) sur le spectre du bruit en sortie, elles introduisent un décalage des fréquences centrales des résonateurs constituant les modulateurs et peuvent par conséquent dégrader la performance attendue par une telle architecture.

Le chapitre 4 est consacré à la recherche d'une méthode de correction permettant de compenser les effets des erreurs analogiques en particulier le décalage des fréquences centrales des modulateurs. Nous proposons une solution qui consiste à étendre l'architecture actuelle en ajoutant un modulateur à chaque extrémité de la bande du signal utile. Ensuite, des algorithmes de calibration sont appliqués pour adapter le traitement numérique en aval des modulateurs et corriger les défauts engendrés par les imperfections analogiques. Ces algorithmes nécessitent une faible complexité matérielle pour leur réalisation et peuvent être appliqués en temps réel. Ils sont développés en détail au cours de ce chapitre.

Le chapitre 5 est consacré à l'optimisation de l'architecture matérielle de chacun des blocs du traitement numérique appliqué en aval des modulateurs. Cette optimisation tient compte tout d'abord de la fonction réalisée par chacun des blocs et de sa vitesse de fonctionnement afin de

profiter du partage des ressources. Ensuite, nous avons cherché à déterminer les longueurs des mots binaires minimaux permettant d'assurer la performance souhaitée par chacun des blocs. Finalement, l'architecture du traitement numérique global est développée en tenant compte des différents blocs optimisés séparément. Cette architecture a été testée et synthétisée au niveau portes logiques avec une technologie CMOS 0.12 μm.

Chapitre 1

Présentation de l'approche multistandard

Objectif

Ce chapitre introduit les enjeux de l'approche multistandard pour les récepteurs en télécommunication. Dans ce contexte, la radio-logicielle s'impose comme la solution permettant la réalisation de récepteur pouvant s'adapter à de multiples standards. Son principe repose sur une conversion des signaux de radiocommunication dans le domaine numérique directement après l'antenne et la reconfigurabilité des entités de traitement numérique. Nous envisageons dans ce chapitre plusieurs architectures de réception dont l'architecture *IF-sampling* semble la plus adaptée pour la solution radio-logicielle. Elle nécessite cependant un CAN rapide, large bande et à grande dynamique d'entrée. Pour ces exigences, les convertisseurs $\Sigma\Delta$ à temps continu s'avèrent une bonne solution, en raison de leur rapidité et leur facilité de réalisation. Ils font l'objet d'une brève présentation en fin de chapitre.

1.1 Introduction

L'évolution des systèmes de télécommunications exige des bandes de fréquences contenant les informations utiles de plus en plus larges. Au début des télécommunications mobiles la largeur des bandes utiles était réservée à la transmission de voies humaines [0, 20kHz]. Aujourd'hui les exigences de transmission ont changé avec :

- l'augmentation du nombre d'applications dans un système de radiocommunication mobile (Internet, GPS, TV, Bluetooth, etc.),
- l'augmentation de la densité d'information,
- la diversité des normes de transmission (plusieurs types de modulations, de codage, etc.).

Compte tenu de ces évolutions, les efforts se concentrent à l'heure actuelle sur la conception et l'intégration de structures de réception universelles. De telles structures devraient pouvoir s'adapter à de nombreuses normes (standard) pour des applications diverses dans un appareil de surface et de consommation raisonnables. Ce type de récepteur est appelé récepteur multistandard.

Pour mieux cerner le cadre de ce travail de thèse, portant sur le CAN large bande fonctionnant à haute fréquence, nous allons présenter en détail le concept de la radio-logicielle et l'intérêt des architectures de réception **large bande** et **multibande**.

1.2 Enjeux et implications technologiques de la réception multistandard

L'objectif est d'avoir un récepteur qui, par ses caractéristiques aussi bien physiques (technologies, architectures des circuits) que logicielles (programmation du fonctionnement sur différentes normes), permette de traiter une multitude de signaux de radiocommunication. Les paragraphes qui suivent constituent une introduction aux principaux enjeux de la réception multistandard ainsi que la solution de la radio-logicielle [8].

1.2.1 Définition et caractéristiques des normes de radiocommunication

Les **normes** sont des accords documentés contenant des spécifications techniques ou autres critères précis destinés à être utilisés systématiquement en tant que règles, lignes directrices ou définitions de caractéristiques pour assurer que des matériaux, produits, processus et services soient aptes à leur emploi.

Dans le cas des télécommunications mobiles, les spécifications relatives à une norme sont :

- le **type de modulation**. Le premier type de modulation apparu est la modulation analogique AM (Amplitude Modulation), puis FM (Frequency Modulation). Avec le développement des technologies intégrées sur silicium, la modulation analogique a rapidement été remplacée par des modulations de type numérique. Le traitement numérique permet une plus grande précision, le développement de codes correcteurs d'erreurs et une plus grande simplicité de réalisation. Au nombre des modulations numériques, on compte aujourd'hui le CDMA (Code Division Multiple Access), le QPSK (Quadrature phase-shift keying), etc.
- la **largeur de bande** : c'est la bande de fréquence qui contient les informations utiles. Elle est directement liée au débit [9]. La création de nouveaux services (multimédia, internet, etc.) exige l'utilisation de largeurs de bandes de plus en plus importantes.
- la **largeur du canal** : la bande est découpée en sous bandes, ces sous bandes constituent les canaux. Chaque canal est réservé à un seul utilisateur, sauf dans le cas récent de la modulation CDMA qui permet à plusieurs utilisateurs d'employer le même canal.
- les voies **montantes** et **descendantes** correspondent aux fréquences d'émission et de réception.

La figure 1.1 illustre l'occupation spectrale de quelques normes de radiocommunication. La première norme mise au point, pour la téléphonie mobile, fut le GSM. Sa largeur de bande se révèle insuffisante pour des applications plus sophistiquées (TV, radio, etc.). D'autres normes ont été crées pour répondre à ces besoins (GPRS pour la génération 2.5G, UMTS pour la 3G et WIMAX pour la 4G sans oublier les réseaux locaux sans fil tels que bluetooth et le WI-FI).

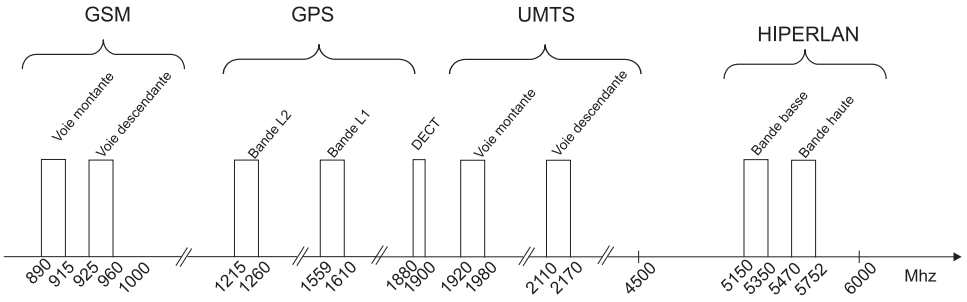

Fig. 1.1 – Occupation spectrale des normes de radiocommunication.

Le tableau 1.1 présente les caractéristiques de quatres normes de la téléphonie mobile. En se basant sur ces caractéristiques, le récepteur multistandard, capable de fonctionner avec toutes les normes, devrait pouvoir traiter des signaux dont la diversité est illustrée comme suit :

- Fréquence de radiocommunication variant de 890 MHz à 5825 MHz
- Bandes de radiocommunication variant de 35 MHz à 675 MHz
- Largeur de bande des canaux variant de 0.2 MHz à 20 MHz
- Modulation à bande étroite : GMSK, GFSK et large bande : QPSK/CDMA et QAM/OFDM.

Tab. 1.1 – Caractéristiques techniques de différentes normes de radiocommunication

	GSM	DECT	UMTS	IEEE802.11a
Voie montante (MHz)	890-915	1880-1900	1920-1980	5150-5825
Voie descendante (MHz)	825-960	-	2110-2170	-
Largeur de bande (MHz)	35	20	60	675
Largeur de canal (MHz)	0.2	1.4	3.84	20
Modulation	GMSK	GFSK	QPSK-CDMA	M-QAM-OFDM

La solution de la radio-logicielle pour la conception d'un tel récepteur est introduite au paragraphe suivant.

1.2.2 Concept de la radio-logicielle

Le concept de la radio-logicielle a été introduit par J.Mitola [10, 11]. Ce concept mettant en jeu des communications sans fil vise à rendre les réseaux et les terminaux intelligents et indépendants dans le but d'en simplifier l'utilisation. Ce concept repose sur la programmation logicielle des fonctionnalités de base de l'interface de radiocommunication réalisée actuellement par des circuits matériels dédiés (ASIC). Cette programmation doit définir : la fréquence porteuse, la largeur de bande du canal de radiocommunication, le type de modulation et de codage. Le schéma idéal d'un terminal de radiocommunication, selon le concept radio-logicielle, comporte (voir figure 1.2) :

- une antenne, un filtre passe-bande BPF et un amplificateur à faible bruit LNA, dont la réalisation se fait dans le domaine analogique
- un CAN et un processeur numérique DSP qui permet la reprogrammation pour l'adaptation à de multiples standards.

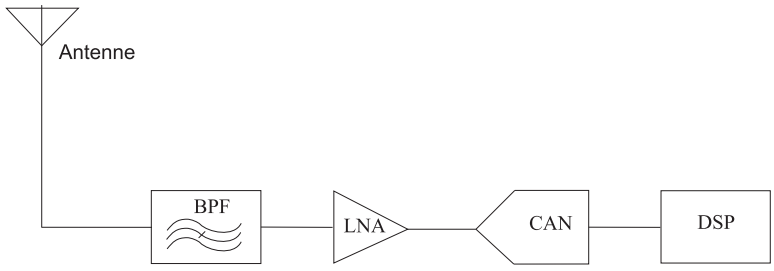

Fig. 1.2 – Schéma idéal d'un récepteur Software Radio.

La réalisation de ce type d'architecture est difficile. Les difficultés majeures résident dans :

- la mise en oeuvre d'un CAN large bande qui soit le plus en amont possible dans la chaîne de réception pour réduire la partie réalisée dans le domaine analogique du récepteur,

– la nécessité d'une dynamique de conversion plus grande, du fait de la présence de brouilleur potentiels devant les signaux utiles,
– le remplacement des circuits intégrés à application spécifique (ASICs) par des DSPs (implantation logicielle) dans le but de réaliser le plus possible de fonctionnalités de radiocommunication par logiciel [12, 13, 11].

Compte tenu de ces difficultés pour la réalisation du concept de radio-logicielle, une alternative nommée SDR (Software Defined Radio) a été présentée par les concepteurs en téléphonie mobile. Elle est constituée, comme le montre le schéma de la figure 1.3, d'un étage RF analogique, d'un étage à fréquence intermédiaire IF et d'un traitement en bande de base sur des DSP pour assurer la programmabilité du récepteur. Cependant, cette architecture ne garantit un fonctionnement multistandard que pour le traitement en bande de base puisqu'il est programmable grâce à son implantation logicielle. Ainsi, de nouvelles architectures font l'objet de recherches actives [14].

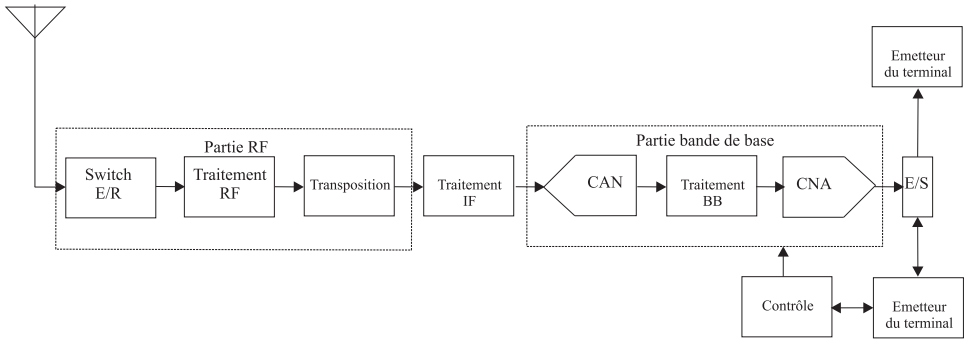

Fig. 1.3 – Schéma de l'architecture Software Defined Radio (SDR).

L'objectif de notre de travail concerne la première des difficultés évoquée à savoir la numérisation le plus tôt possible dans la chaîne de réception.

1.3 Architecture du récepteur multistandard

En raison de la multitude des normes, le récepteur multistandard exige un fonctionnement avec des signaux RF **multi-bande** et **large bande**. Nous allons détailler dans la suite plusieurs architectures de récepteurs afin de déterminer la plus adaptée à des applications multistandards.

1.3.1 État de l'art sur les architectures des récepteurs de radiocommunication

Le récepteur le plus utilisé dans les mobiles de deuxième génération est le récepteur **hétérodyne** dont le schéma de fonctionnement est donné par la figure 1.4.

Dans cette architecture, le signal reçu par l'antenne est filtré par un filtre RF pour isoler la bande de réception, puis amplifié par un amplificateur faible bruit LNA. Ensuite, le signal est filtré par un filtre de réjection d'image avant le mélangeur qui effectue une première démodulation. Puis, le signal est ramené en bande de base grâce au deuxième mélangeur. La sélection du canal se fait au niveau du deuxième oscillateur local f_{ol_2}, programmable en général, associé au mélangeur. Cette deuxième démodulation est complexe, ceci est dû aux techniques de modulation utilisées dans les standards actuels de radiocommunication (ex : QPSK). En effet, la partie positive du spectre du signal modulé et sa partie négative ne portent pas la même information. Cette démodulation complexe se fait sur deux voies I (In phase) et Q (Quadrature Phase) pour ne

pas perdre d'informations. Enfin, la sélection de canal est assurée sur les deux voies, de manière analogique avant la numérisation du signal (figure 1.4).

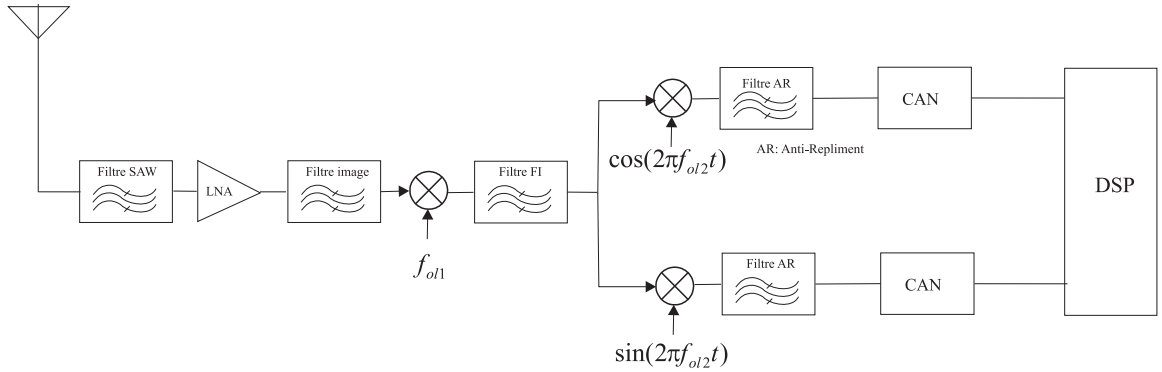

Fig. 1.4 – Architecture d'un récepteur hétérodyne.

Cette architecture exige des filtres très sélectifs pour supprimer les fréquences images et permettre une bonne réception du signal utile. Ceci constitue un obstacle majeur pour l'utilisation de telles architectures dans un récepteur multistandard. En effet, le récepteur multistandard doit fonctionner avec des signaux de fréquences très élevées. Ceci nécessite, dans le cadre d'une architecture hétérodyne, des filtres de réjection d'images de fréquence et de facteur de qualité très élevés afin de bien supprimer les images. Ces filtres sont difficilement réalisables en pratique.

Pour éviter le problème de la réjection de la fréquence image, l'architecture homodyne permet de ramener directement le signal utile en bande de base. Dans ce cas, la fréquence de l'oscillateur local f_{ol} est la même que celle de la porteuse radio fréquence du signal f_{RF} ainsi, la fréquence intermédiaire est zéro $f_{RF} - f_{ol} = 0$ (figure1.5).

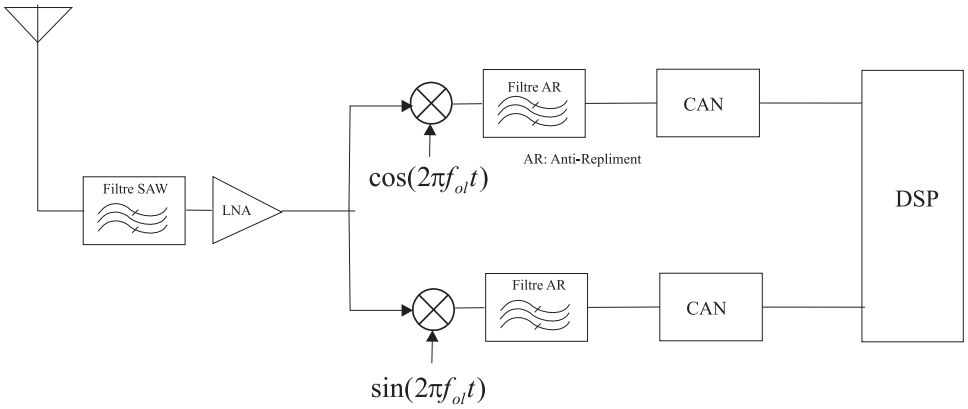

Fig. 1.5 – Architecture d'un récepteur homodyne.

Toutefois, le problème de la fréquence image n'est pas totalement résolu car la transposition se faisant par rapport à la fréquence centrale du canal, le spectre à gauche de la fréquence centrale se superpose à celui de droite, devenant indissociables. Nous pouvons nous affranchir de cet inconvénient en séparant les composantes en phase (I) et en quadrature (Q) du signal (figure 1.5).

En première approche, l'architecture homodyne semble intéressante mais elle souffre d'un certain nombre de problèmes à surmonter :

- Le **décalage en tension** : en effet, le mélangeur n'isole pas parfaitement l'oscillateur local du LNA et inversement. Ces fuites sont mélangées à elles-mêmes, produisant un offset relativement important qui peut saturer les étages suivants.
- Le **désappariement** entre les deux voies I et Q : cette source d'erreur est présente dans toutes les architectures à deux branches puisqu'elle provient d'un appariement imprécis entre les deux voies. En effet, les deux signaux de démodulation n'ont pas la même amplitude et le déphasage entre eux n'est pas exactement $\frac{\pi}{2}$. Cette erreur se traduit par une dégradation de la qualité du signal reçu.
- Le **bruit de scintillement** : il est aussi nommé « bruit en 1/f ». Il est toujours présent dans les composants actifs et dans certains composants passifs. Il peut être dû à des impuretés dans le matériau pour un transistor, par exemple, qui libèrent aléatoirement des porteurs de charge, ou bien à des recombinaisons électron-trou parasites. Ce bruit prédominant en basses fréquences est caractérisé par un spectre dont l'amplitude des raies évolue en 1/f.

Ces inconvénients font que l'architecture homodyne est peu adaptée à la réception de signaux de radiocommunication. Il est alors intéressant de chercher des structures de mélangeurs permettant la réjection d'image sans nécessiter l'utilisation de filtres externes. Deux méthodes ont été proposées par « Hartley » et « Weaver » basées sur la démodulation en quadrature et le changement de la polarité du spectre du signal image afin de le supprimer. La méthode de Weaver est la plus performante vis à vis des erreurs d'appariement entre les voies de démodulation et elle se base sur la double démodulation en quadrature pour supprimer le signal image [9]. En se basant sur la méthode de « Weaver », une topologie d'un récepteur nommé **Wide-Band-IF** (figure 1.6) a été introduite dans [15].

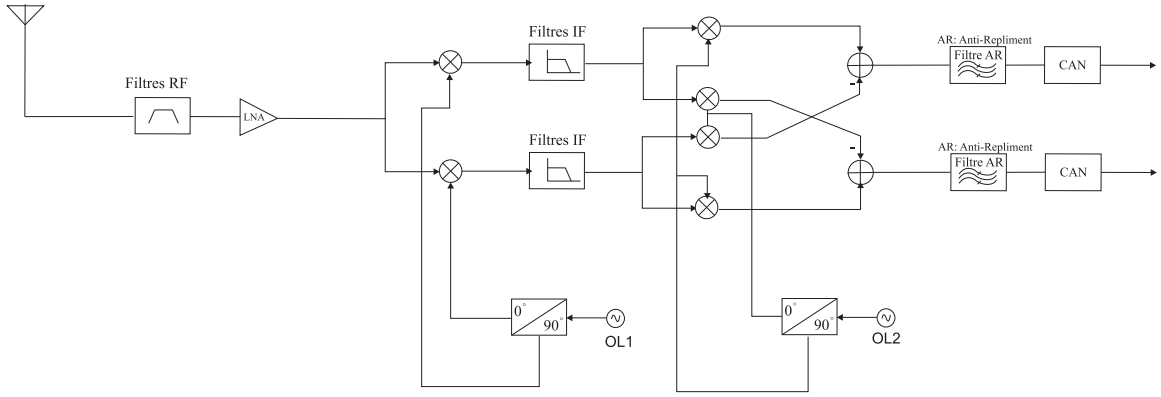

Fig. 1.6 – Architecture de réception Wide band IF.

Cette architecture permet une bonne réjection du signal image [16] et par conséquent de bonnes performances de réception. Par contre, elle exige l'utilisation de six mélangeurs analogiques ce qui pose un problème matériel au niveau de la surface d'intégration et du coût.

En se basant sur la facilité de réalisation d'une démodulation dans le domaine numérique, une alternative à cette architecture est l'architecture **IF-Sampling** (figure 1.7). Cette architecture effectue une première démodulation en quadrature dans le domaine analogique à la fréquence intermédiaire IF. Ensuite, le signal est numérisé avant d'être ramené en bande de base suite à une deuxième démodulation en quadrature effectuée dans le domaine numérique.

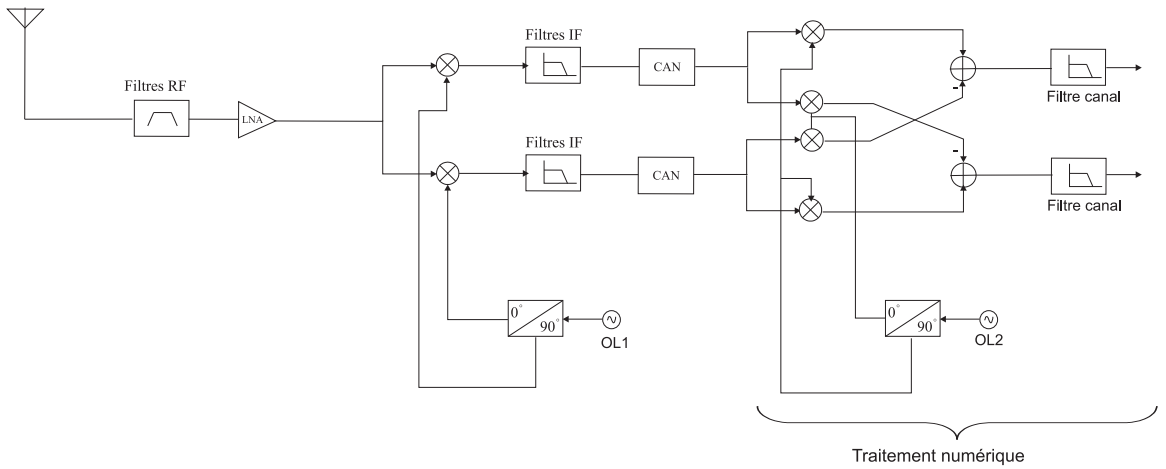

Fig. 1.7 – Architecture de réception IF Sampling.

L'architecture **IF-Sampling** présente :

- une bonne réjection d'image avec la méthode de Weaver par rapport au récepteur hétérodyne,
- une surface d'implantation raisonnable par rapport à l'architecture Wide band IF grâce à la réalisation du deuxième démodulation dans le domaine numérique.

Ces atouts rendent cette architecture envisageable pour des applications multi-standard. Par contre, l'enjeu essentiel d'une telle architecture est la performance du CAN. Celui-ci doit convertir un signal large bande à la fréquence intermédiaire IF, avoir une grande dynamique d'entrée et une faible puissance de consommation. Ces exigences rendent impossible l'utilisation de CAN classiques de type *Nyquist*. La recherche d'une nouvelle architecture de convertisseur compatible avec l'architecture IF-Sampling fait l'objet de ce travail de thèse.

1.3.2 Récepteur multibande

La réception multistandard exige en plus du caractère large bande, un fonctionnement multibande. Ainsi, nous nous proposons dans ce paragraphe de montrer l'adaptation de l'architecture large bande IF-Sampling (figure 1.7) au fonctionnement multibande. Parmi les éléments constitutifs de cette chaîne de réception, certains nécessitent des modifications :

- L'antenne : premier élément de la chaîne de réception, elle peut être réalisée grâce à la technique PIFA (Planar Inverted Folded Antenna) [17, 18, 19] pour un fonctionnement multibande.
- Le filtre RF peut être doté d'une fonctionnalité multibande si les bandes des normes visées sont assez proches. Dans le cas où les normes possèdent des bandes assez éloignées, la solution est d'utiliser un banc de filtres RF dont chacun est sélectionné pour une norme donnée. Le choix entre les différentes normes se fait par le choix du filtre RF correspondant grâce à des switches analogiques commandés par la partie numérique.
- Le LNA doit avoir une bande passante qui couvre toutes les largeurs de bande des normes utilisées. Certains fabricants proposent des LNA large bande tel que le ADL5523 d'Analog Devices qui a une bande de fonctionnement de 400 MHz à 4 GHz. Il couvre toute les bandes du GSM, GPS et UMTS.

– l'oscillateur local doit avoir une gamme de fréquence à synthétiser assez large pour ramener les bandes des différentes normes autour de la fréquence intermédiaire. Les synthétiseurs à base de modulateurs ΣΔ sont capables de synthétiser une bande suffisamment large [20, 21].

La figure 1.8 présente le schéma bloc de l'architecture d'un récepteur multistandard.

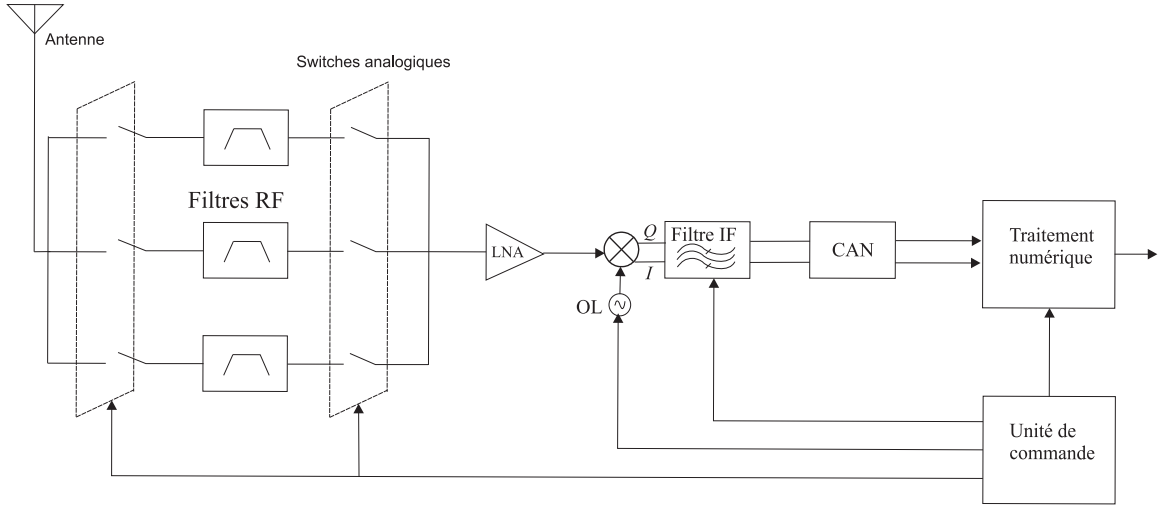

Fig. 1.8 – Architecture de réception multistandard.

Cette architecture multibande possède une unité de commande permettant d'agir simultanément sur le choix du filtre RF, sur la fréquence de l'oscillateur local et sur la bande du filtre IF afin d'assurer un fonctionnement multistandard. Une variante à cette architecture utilisant la technique de réception homodyne a été présentée dans [8]. Compte tenu de la diversité des normes traitées par l'architecture multibande, les signaux à l'entrée du CAN occupent une bande très importante et présentent des dynamiques en amplitude élevées. D'où la nécessité de disposer d'un CAN répondant aux trois exigences suivantes :

– des fréquences de fonctionnement élevées,
– une bande de fonctionnement large adaptée à toutes les normes,
– une résolution élevée pour avoir une bonne qualité du signal converti.

1.4 Le Convertisseur analogique-numérique

Les convertisseurs analogique-numérique de type Nyquist (Pipeline, à Approximations successives) s'avèrent très lents et ainsi ne sont pas adaptés aux récepteurs multistandards. Un autre type de convertisseur de Nyquist pouvant en terme de vitesse s'adapter aux applications de notre intérêt est le convertisseur Flash. Cependant, malgré sa vitesse, ce type de convertisseur présente deux inconvénients majeurs :

– Sa précision est limitée. Aujourd'hui les CAN du commerce de type Flash les plus rapides (750 MHz pour le MAX108 chez Maxim) ne dépassent pas une résolution de 7.5 bits.
– Sa consommation et la surface requise pour son implantation sont très élevées. En effet, ce CAN nécessite $2^{N_b} - 1$ comparateurs pour réaliser un CAN de N_b bits. Ceci est le principal inconvénient. Son utilisation est limitée aux applications qui n'exigent qu'une résolution modérée, n'excédant pas 8 bits.

En plus d'une vitesse de fonctionnement élevée, le nombre de bits nécessaires au CAN dédié aux applications multistandards, doit être élevé à cause de la grande dynamique des signaux à son entrée. Les CAN à base de modulateurs ΣΔ à temps continu présentent un certain nombre d'avantages : leur vitesse de conversion (fréquence centrale des signaux convertis jusqu'à 2 GHz [22]), leur précision avec un faible nombre de composants et leur dynamique d'entrée. La figure 1.9 présente le schéma bloc d'un modulateur ΣΔ à temps continu.

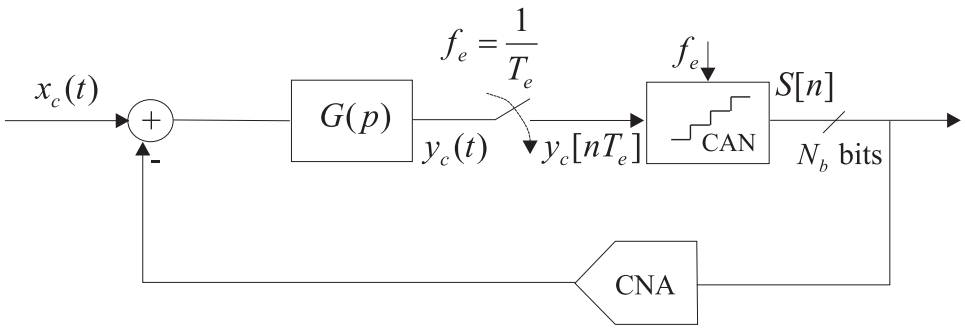

Fig. 1.9 – Modulateur ΣΔ à temps continu.

Son principe de fonctionnement repose sur :

- Le suréchantillonnage : l'échantillonnage à une fréquence égale à plusieurs fois la fréquence de Nyquist f_N ($f_N = 2f_B$, f_B est largeur de bande du signal) permet d'étaler la puissance de bruit de quantification introduit par le CAN dans la boucle sur une gamme de fréquences plus large en diminuant ainsi la puissance de bruit dans la bande du signal utile.
- La mise en forme du bruit de quantification du CAN par la boucle de rétroaction (figure 1.9). Elle consiste à atténuer la densité spectrale de bruit de quantification dans la bande du signal utile et par conséquent à améliorer la résolution.

Le modulateur à temps continu est caractérisé par :

- Le type du filtre de boucle G(p) : passe-bas ou passe-bande. Ce filtre est réalisé avec des éléments actifs ou passifs (Gm-C, Gm-LC, etc.) permettant de travailler à des fréquences élevées.
- l'ordre L du filtre G(p). Plus l'ordre du modulateur augmente, moins le bruit est important dans la bande, ce qui améliore la résolution du modulateur. Cependant, dès que le nombre de résonateurs est supérieur à 2, le modulateur peut être instable [23]. Des travaux de recherche menés au laboratoire [24, 25] ont montré la possibilité de réalisation de modulateurs d'ordre 6 comme un bon compromis entre la performance et la stabilité.
- Le nombre de bits N_b du CAN. Une augmentation du nombre de bits peut améliorer la résolution. Cependant, le nombre de bits du CAN doit rester raisonnable afin de conserver l'intérêt de la modulation qui est d'obtenir une résolution élevée au moyen de composants moins performants. Typiquement, un quantificateur interne n'excède pas 4 voire 3 bits de résolution.
- la fréquence d'échantillonnage (f_e), on définit alors le facteur de suréchantillonnage ($OSR = \frac{f_e}{2f_B}$),

Dans la suite de ce travail de thèse nous utiliserons des architectures composées de modulateurs à temps continu d'ordre 6 avec des CAN et CNA 3 bits.

Une présentation détaillée sur le principe de fonctionnement et la méthode de synthèse des modulateurs à temps continu se trouve en annexe A.

Bien que les modulateurs ΣΔ à temps continu présentent une vitesse de fonctionnement élevée et une bonne résolution, leur bande de fonctionnement restent faible par rapport à la largeur de

bande des signaux utiles dans un récepteur multistandard. La mise en parallèle de modulateurs à temps continu constitue une voie de recherche prometteuse pour l'élargissement de la bande du convertisseur. Les différentes architectures en parallèle ainsi qu'une comparaison entre elles seront détaillées au chapitre 2.

1.5 Conclusion

Dans ce chapitre nous avons présenté le concept de réception multistandard et la nécessité d'une architecture reconfigurable afin de répondre à la diversité des traitements à exécuter. L'un des verrous à lever dans une telle architecture est le CAN. Il remplit un rôle primordial pour atteindre la flexibilité du récepteur aux différentes normes de communication. Il doit être placé le plus proche possible de l'antenne pour assurer cette mission et fonctionner ainsi à très haute fréquence avec des signaux large bande.

Le modulateur $\Sigma\Delta$ passe bande à temps continu est le plus adapté à ce type d'application en raison de sa vitesse de fonctionnement et de sa résolution élevées. Cependant, sa bande de fonctionnement reste étroite par rapport aux exigences du récepteur multistandard. La mise en parallèle des modulateurs à temps continu permet d'augmenter la bande de fonctionnement. Cette solution sera détaillée au prochain chapitre où seront présentées les différentes approches parallèles.

Chapitre 2

Présentation des ADCs parallèles à base de modulateurs ΣΔ

Objectif

Ce chapitre présente le principe de fonctionnement des CAN à base de modulateurs ΣΔ en parallèle. Il existe plusieurs solutions basées sur ce principe à savoir l'architecture à entrelacement temporel TIΣΔ (Time Interleaved Sigma-Delta), l'architecture à base de modulation de *Hadamard* ΠΣΔ (Parallel Sigma-Delta) et l'architecture à base de décomposition fréquentielle FBD (Frequency Band Decomposition). L'imperfection des composants analogiques présente une source d'erreur inévitable qui peut dégrader la performance des architectures en parallèle. La robustesse de chacune de ces architectures face à ces erreurs sera développée et l'architecture FBD retenue comme la plus robuste.

2.1 Introduction

La modulation ΣΔ présente un certain intérêt pour la conversion A/N des signaux. Elle permet d'atteindre de hautes performances en terme d'ENOB (Effectif Number of bits). Cependant, la largeur de bande de fonctionnement de ce type de convertisseur est faible. Il ne peut pas être utilisé dans le cadre d'un récepteur multistandard où le critère large bande est un critère primordial. La mise en parallèle de modulateurs ΣΔ est un moyen prometteur pour l'augmentation de la bande de conversion. Plusieurs architectures basées sur ce principe ont déjà été proposées telle que l'architecture à entrelacement temporel TIΣΔ, l'architecture à base de modulation de *Hadamard* ΠΣΔ et l'architecture à base de décomposition fréquentielle FBD. Un convertisseur avec M modulateurs en parallèle assure en principe, quelque soit la technique utilisée, temporelle ou fréquentielle, une bande de fonctionnement M fois plus large que celle obtenue avec un seul modulateur. Dans la suite, nous allons étudier chacune de ces architectures et évaluer leurs performances vis-à-vis des imperfections des composants analogiques, et différentes méthodes de correction.

2.2 Architecture à entrelacement temporel

L'entrelacement temporel a été introduit par Black et al. [26] comme un moyen pour augmenter la vitesse de conversion du CAN en disposant en parallèle des CAN de type *Nyquist* fonctionnant à des fréquences plusieurs fois plus faibles que la fréquence de conversion du convertisseur global. La figure 2.1 présente la structure générale de l'architecture TIΣΔ.

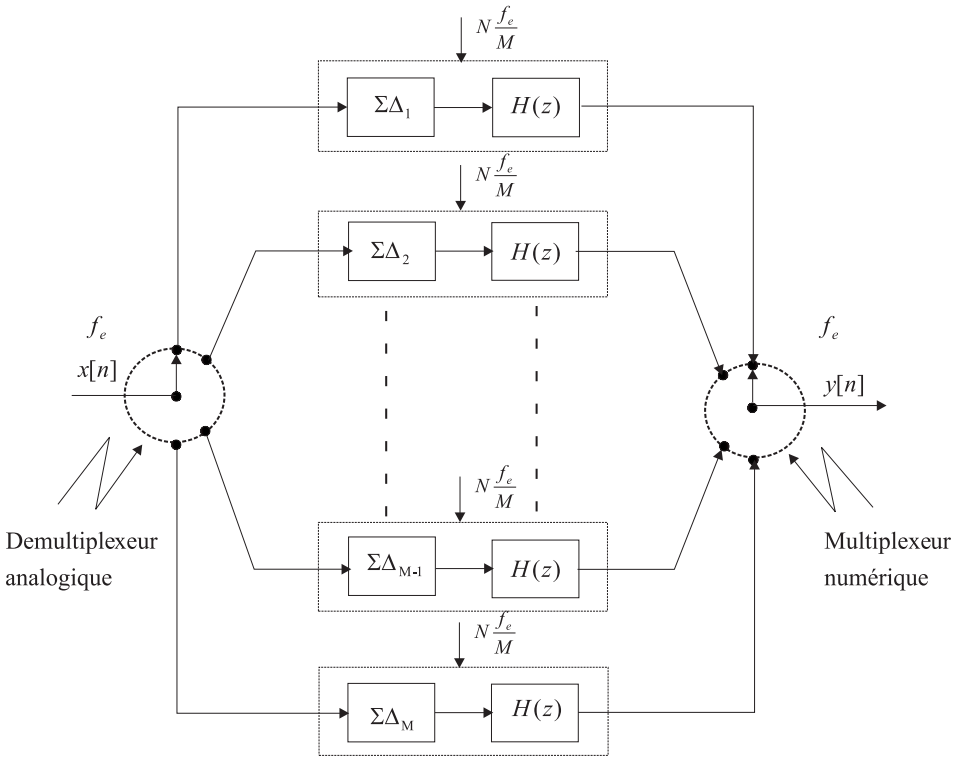

Fig. 2.1 – Architecture parallèle à entrelacement temporel TIΣΔ.

Son principe de fonctionnement est le suivant :

- un démultiplexeur analogique distribue le signal d'entrée entre les M modulateurs ΣΔ identiques en parallèle. Cela signifie que le signal à l'entrée du modulateur a une fréquence M fois plus faible que le signal d'entrée de l'architecture globale x [n]. On note que si le signal d'entrée x [n] n'est pas appliqué au modulateur, l'entrée du modulateur est connectée à la masse.
- un filtre passe-bas H(z) élimine le bruit de quantification hors bande. Le choix des coefficients de ce filtre dépend de la reconstruction complète du signal d'entrée et de l'atténuation du bruit de quantification introduit par le modulateur. Sa fonction de transfert est donnée dans le domaine en z par $H(z) = \sum_{i=0}^{P-1} h[i]z^{-i}$ où P est l'ordre du filtre.
- un multiplexeur numérique reconstruit le signal numérisé.

Le fonctionnement des modulateurs ΣΔ nécessite un échantillonnage du signal à leur entrée à une fréquence N fois plus grande que la fréquence de *Nyquist*. D'où une fréquence de fonctionnement égale à $N\frac{f_e}{M}$ où N est le rapport de suréchantillonnage. Pour N=M, le convertisseur TIΣΔ est un convertisseur de *Nyquist*.

2.2.1 Performances théoriques

Afin d'évaluer les performances d'un convertisseur TIΣΔ, l'architecture numérique équivalente dans le domaine discret de l'architecture TIΣΔ est présentée sur la figure 2.2.

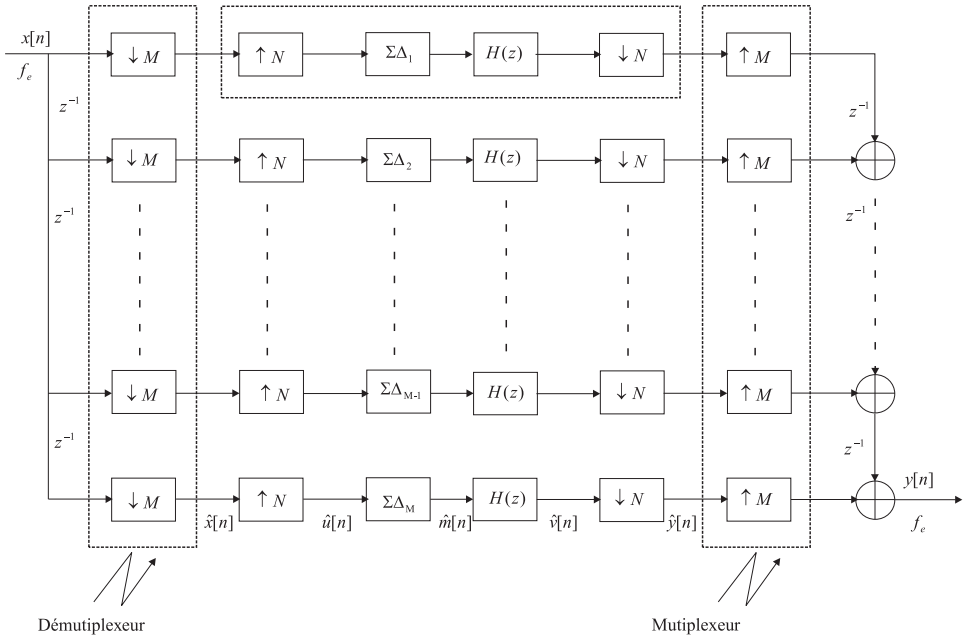

Fig. 2.2 – Architecture parallèle à entrelacement temporel TIΣΔ dans le domaine z.

Le démultiplexeur en entrée est remplacé par des décimateurs de facteur M à l'entrée de chaque voie avec un retard de z^{-1} entre deux voies adjacentes [27]. Le multiplexeur est représenté de la même façon mais avec des interpolateurs d'ordre M. Le décimateur et l'interpolateur d'ordre N modélise le suréchantillonnage du modulateur ΣΔ. On exprime le signal $\hat{M}(z)$ en sortie du modulateur conformément au modèle linéaire du quantificateur dans la boucle du modulateur [28], par :

$$\hat{M}(z) = STF(z)\hat{U}(z) + NTF(z)E(z) \tag{2.1}$$

E(z) : est la transformée en z du bruit de quantification du CAN dans la boucle du modulateur.

Dans le cas d'un modulateur passe-bas d'ordre L avec une faible largeur de bande, on peut approcher la fonction de transfert par rapport au signal STF(z) par z^{-L} et celle par rapport au bruit NTF(z) par $\left(1 - z^{-1}\right)^L$. En se basant sur l'hypothèse de linéarité du modulateur, nous pouvons utiliser le principe de superposition pour calculer la composante utile $y_x[n]$ du signal y[n] en sortie qui est due au signal d'entrée x[n] et la puissance de la composante $y_e[n]$ qui est due au bruit de quantification e[n].

Expression de $Y_x(z)$ et condition de reconstruction parfaite du signal d'entrée

Afin de simplifier l'établissement de l'expression de $Y_x(z)$, nous allons omettre le terme z^{-L} dû au passage du signal dans le modulateur. Ceci n'influe pas sur le résultat final car ce n'est qu'un retard identique sur les signaux des différentes voies. La figure 2.3 présente dans ce cas l'architecture simplifiée de l'architecture TIΣΔ du point de vue du signal utile.

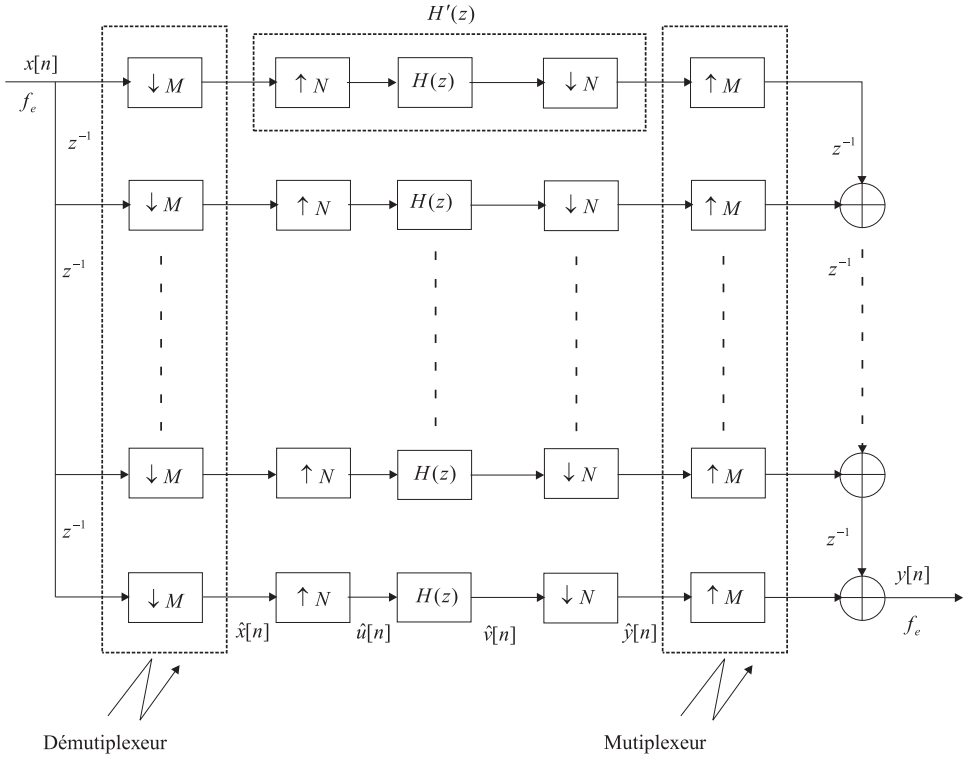

Fig. 2.3 – Modèle simplifié de l'architecture TIΣΔ du point de vue du signal utile.

L'interpolateur (↑ N), le filtre H(z) et le décimateur (↓ N) dans la figure 2.3 sont équivalents à la fonction H′(z). Pour déterminer H′(z), nous partons de l'expression du signal Û(z) en sortie du premier interpolateur donnée par :

$$\hat{U}(z) = \hat{X}(z^N)$$

Le signal V̂(z) en sortie du filtre passe-bas s'exprime par :

$$\hat{V}(z) = \hat{X}(z^N)H(z) \tag{2.2}$$

Ensuite, ce signal est décimé d'un facteur N. L'expression du signal décimé est donnée par :

$$\hat{Y}(z) = \frac{1}{N}\sum_{l=0}^{N-1} \hat{V}\left(z^{\frac{1}{N}}W_N^l\right) \quad avec \quad W_N = e^{-j\frac{2\pi}{N}} \tag{2.3}$$

En insérant l'équation (2.2) dans (2.3), nous obtenons l'équation suivante :

$$\begin{aligned} \hat{Y}(z) &= \frac{1}{N}\sum_{l=0}^{N-1} \hat{X}\left(z^{\frac{N}{N}}W_N^{lN}\right) H\left(z^{\frac{1}{N}}W_N^l\right) \\ &= \hat{X}(z)\frac{1}{N}\sum_{l=0}^{N-1} H\left(z^{\frac{1}{N}}W_N^l\right) \end{aligned} \tag{2.4}$$

La fonction de transfert H′(z) est donnée, en utilisant la définition de la transformée en z, par :

$$H'(z) = \frac{\hat{Y}(z)}{\hat{X}(z)} = \frac{1}{N}\sum_{l=0}^{N-1} H\left(z^{\frac{1}{N}}W_N^l\right) = \frac{1}{N}\sum_{l=0}^{N-1}\sum_{i=-\infty}^{\infty} h[i]\, z^{\frac{-i}{N}} W_N^{-li} \tag{2.5}$$

H′(z) peut encore être mise, en inversant les deux sommes, sous la forme suivante :

$$H'(z) = \sum_{i=-\infty}^{\infty} h\,[i]\, z^{\frac{-i}{N}} \left(\frac{1}{N} \sum_{l=0}^{N-1} W_N^{-li} \right) = \sum_{i=-\infty}^{\infty} h\,[i]\, z^{\frac{-i}{N}} C_N\,[i] \tag{2.6}$$

Le terme C_N [i] vérifie (Comb sequence) l'égalité suivante [27] :

$$C_N\,[i] = \frac{1}{N} \sum_{l=0}^{N-1} W_N^{-li} = \begin{cases} 1 & \text{si } i \text{ est multiple de } N \\ 0 & \text{sinon} \end{cases} \tag{2.7}$$

Nous pouvons constater d'après les équations 2.6 et 2.7 que seuls les coefficients h [i] dont l'indice i est multiple de N entrent en jeu pour la détermination de la fonction de transfert H′(z). Afin de ne pas modifier le spectre du signal d'entrée, le plus judicieux est de prendre H′(z) = 1. Ceci est simple à réaliser en imposant sur les coefficients h [i] la contrainte donnée par :

$$h\,[i] = \begin{cases} 1 & \text{si } i = 0 \\ 0 & \text{si } i \text{ est multiple de } N \end{cases} \tag{2.8}$$

Les autres coefficients du filtre H(z) sont choisis de façon à minimiser la puissance de bruit en sortie [1]. En se basant sur cette condition, l'architecture TIΣΔ peut se mettre sous la forme présentée à la figure 2.4.

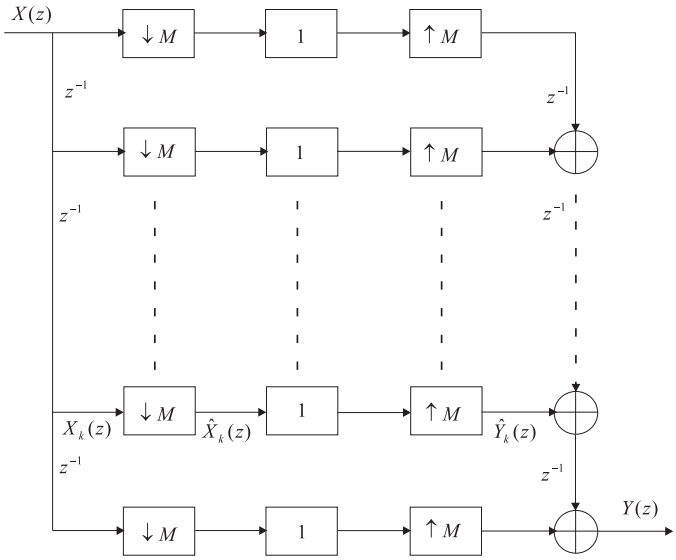

Fig. 2.4 – Architecture TIΣΔ avec $H'(z) = 1$.

Le signal à l'entrée de chaque voie k ($X_k(z)$), présente un retard z^{-k} par rapport au signal d'entrée ($X_k(z) = z^{-k}X(z)$). Ensuite, ce signal est décimé d'un facteur M. Le signal après la décimation $\hat{X}_k(z)$ s'exprime par :

$$\hat{X}_k(z) = \frac{1}{M} \sum_{l=0}^{M-1} z^{\frac{-k}{M}} W_M^{-lk} X\left(z^{\frac{1}{M}} W_M^{l} \right) \tag{2.9}$$

Ensuite le signal décimé $\hat{X}_k(z)$ est interpolé d'un facteur M pour obtenir le signal $\hat{Y}_k(z)$. Le signal $\hat{Y}_k(z)$ s'exprime par :

$$\hat{Y}_k(z) = \frac{1}{M} \sum_{l=0}^{M-1} z^{-k} W_M^{-lk} X\left(z^{1} W_M^{l} \right) \tag{2.10}$$

Le signal Y(z) en sortie est la somme des sorties de toutes les voies $\hat{Y}_k(z)$ en tenant compte du retard propre à chaque voie. Il est donné par :

$$\begin{aligned} Y(z) &= \sum_{k=0}^{M-1} z^{-(M-1-k)}\hat{Y}_k(z) = \sum_{k=0}^{M-1} z^{-(M-1-k)}\frac{1}{M}\sum_{l=0}^{M-1} z^{-k}W_M^{-lk}X\left(zW_M^l\right) \\ &= \frac{1}{M}z^{-(M-1)}\sum_{l=0}^{M-1} X\left(zW_M^l\right)\underbrace{\sum_{k=0}^{M-1} W_M^{-lk}}_{MC_M(l)} = z^{-(M-1)}\sum_{l=0}^{M-1} X\left(zW_M^l\right)C_M(l) \\ &= z^{-(M-1)}X\left(z\right) \end{aligned} \tag{2.11}$$

Donc le signal en sortie Y(z) est une version retardée de M − 1 cycles d'horloge du signal d'entrée X(z). Les termes $\mathrm{X}\left(\mathrm{zW_M^l}\right)\mathrm{C_M(l)} = \mathrm{X}\left(\mathrm{e^{j\left(w-\frac{2\pi l}{M}\right)}}\right)$ dans l'équation (2.11) pour l ≠ 0 représentent le repliement spectral dû à la décimation. Ces termes de repliement spectral sont annulés en raison de la contrainte 2.8 sur la spécification du filtre H′(z). On voit donc que l'architecture à entrelacement temporel se comporte comme un filtre passe-tout par rapport au signal d'entrée si on impose la contrainte 2.8 sur le filtre passe-bas H(z).

Étant donné que la fonction de transfert du signal du modulateur est un simple retard $\mathrm{STF(z)} = \mathrm{z^{-L}}$, son insertion dans l'architecture TIΣΔ n'influe pas sur le raisonnement développé ci-dessus. Comme ce retard est identique pour toutes les voies, il apparaîtra sur le signal en sortie conformément à l'équation suivante :

$$Y(z) = z^{-(M-1+L)}X\left(z\right) \tag{2.12}$$

Choix du filtre H(z)

Les coefficients du filtre passe-bas H(z) doivent vérifier la contrainte imposée par l'équation (2.8) afin d'avoir une reconstruction parfaite du signal d'entrée. Un filtre passe-bas idéal est un filtre qui permet de satisfaire cette contrainte. En effet, la fonction de transfert d'un filtre passe-bas idéal est donnée par :

$$H\left(e^{jw}\right) = \begin{cases} 1, & \text{pour} \quad |w| < w_c \\ 0 & \text{sinon} \end{cases} \Rightarrow h(i) = \begin{cases} \frac{w_c}{\pi}\left(\frac{\sin(w_c i)}{w_c i}\right) & \text{pour} \quad i \neq 0 \\ \frac{1}{N} & \text{pour} \quad i = 0 \end{cases} \tag{2.13}$$

Nous pouvons remarquer que le choix d'une fréquence de coupure de type $\mathrm{f_c} = \frac{1}{2\mathrm{N}}$ annule les coefficients du filtre dont l'indice est un multiple de N. Ce filtre joue également le rôle d'un filtre anti-repliement dans la bande $\left[-\frac{1}{2\mathrm{N}}, \frac{1}{2\mathrm{N}}\right]$ avant d'appliquer la décimation ↓ N. La figure 2.5 montre les valeurs des coefficients en fonction de l'indice i ainsi que les réponses fréquentielles du filtre H(z) pour différents rapports de suréchantillonnage N = 4, N = 8.

Le filtre idéal impose une réponse fréquentielle de gain 0 dB dans la bande passante $\left[-\frac{1}{2\mathrm{N}}, \frac{1}{2\mathrm{N}}\right]$. Or, dans une architecture TIΣΔ, la reconstruction parfaite du signal d'entrée nécessite que tous les $N^{ème}$ coefficients du filtre soient nuls à l'exception du coefficient du centre (équation (2.8)). Ceci permet de relâcher les contraintes sur la bande passante du filtre H(z) et de choisir les coefficients restants de façon à minimiser l'erreur de quantification totale introduite par le modulateur ΣΔ [1, 29].

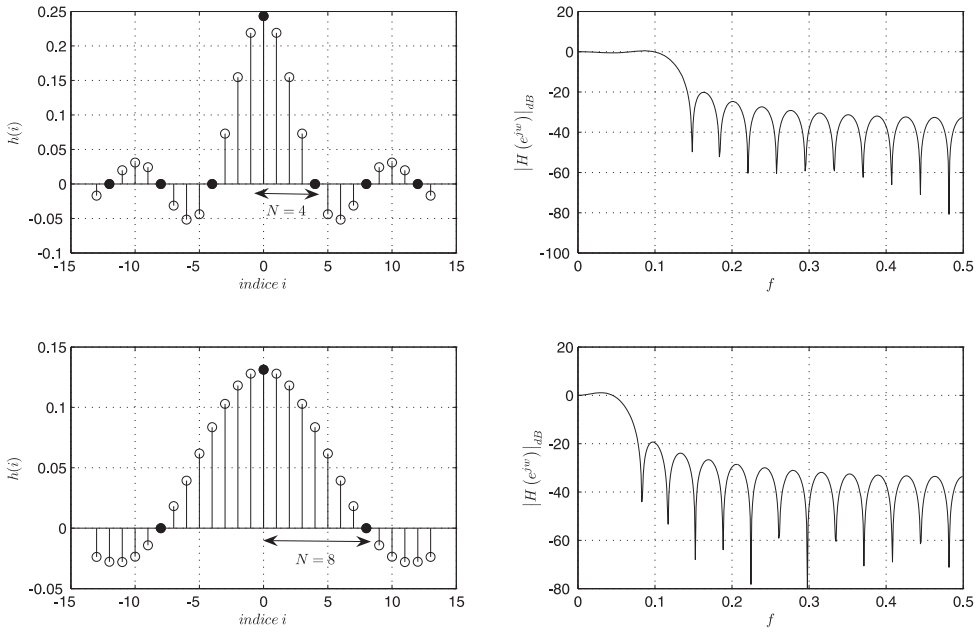

Fig. 2.5 – Coefficients h(i) et réponse fréquentielle du filtre H(z) pour N = 4 et N = 8.

Expression de la puissance de bruit $y_e[n]$ en sortie

Le bruit de quantification introduit par le CAN dans la boucle du modulateur ΣΔ est un bruit blanc si toutes les conditions nécessaires de *Bennet* sont respectées [28]. Ce bruit a une densité spectrale constante de $\frac{q^2}{12f_e}$ où q est le pas de quantification du CAN. En se basant sur le modèle linéaire du modulateur, la densité spectrale du bruit en sortie du modulateur est donnée par :

$$S_{ee}\left(e^{jw}\right) = \frac{q^2}{12f_e}\left|NTF\left(e^{jw}\right)\right|^2 \tag{2.14}$$

La figure 2.6 présente le chemin parcouru par chacun des signaux de bruit $e_k[n]$. Après le passage dans le filtre passe-bas, le bruit présente la densité spectrale donnée par l'équation suivante :

$$H_{ee}\left(e^{jw}\right) = \frac{q^2}{12f_e}\left|NTF\left(e^{jw}\right)H\left(e^{jw}\right)\right|^2 = \frac{q^2}{12f_e}\left|V\left(e^{jw}\right)\right|^2 \tag{2.15}$$

Après la décimation ↓ N, la densité spectrale de bruit s'exprime par :

$$H_{dd}\left(e^{jw}\right) = \frac{q^2}{12f_e}\frac{1}{N}\sum_{L=0}^{N-1}\left|V\left(e^{jw/N}W_N^L\right)\right|^2 \tag{2.16}$$

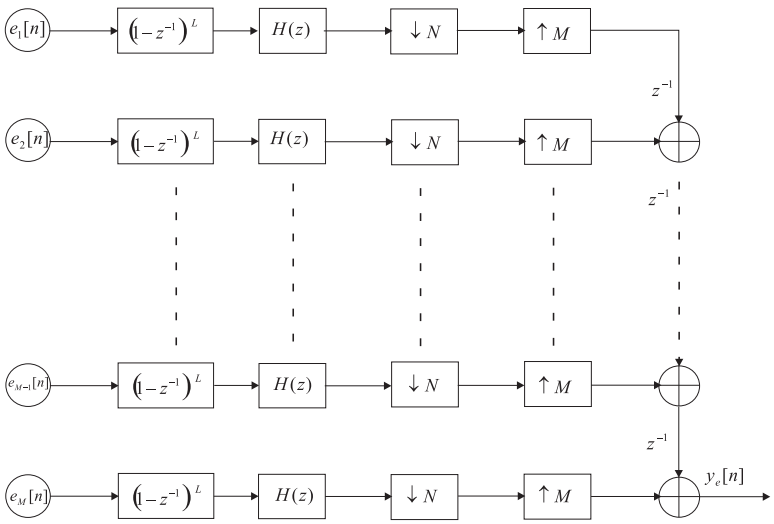

Fig. 2.6 – Parcours du bruit $e_k[n]$ dans l'architecture TIΣΔ.

Ensuite, l'interpolation ↑ M permet de compresser le spectre en créant M répliques du spectre initial entre 0 et 1. Ceci se traduit par une modification de la densité spectrale comme le montre l'équation suivante :

$$P_{ee}\left(e^{jw}\right) = \frac{q^2}{12 f_e} \frac{1}{M \times N} \sum_{L=0}^{N-1} \left| V\left(e^{jwM/N} W_N^L\right)\right|^2 \tag{2.17}$$

La densité spectrale a été divisée par M car l'interpolation consiste à insérer M − 1 zéros entre deux échantillons successifs du signal à interpoler. La puissance de bruit intervenant à la sortie de chaque voie est donnée par :

$$P_k = \frac{1}{2\pi} \frac{q^2}{12} \frac{1}{M \times N} \sum_{L=0}^{N-1} \int_{-\pi}^{\pi} \left| V\left(e^{jwM/N} W_N^L\right)\right|^2 dw \tag{2.18}$$

En supposant que les sources de bruit de quantification des différents modulateurs sont décorrélées entre elles, la puissance de bruit totale en sortie s'exprime par :

$$P = M \times P_k = \frac{1}{2\pi} \frac{q^2}{12} \frac{1}{N} \sum_{L=0}^{N-1} \int_{-\pi}^{\pi} \left| V\left(e^{jwM/N} W_N^L\right)\right|^2 dw \tag{2.19}$$

En supposant que le filtre passe-bas H(z) est un filtre idéal de fréquence de coupure $\frac{1}{2N}$ et en se basant sur la périodicité de la transformée de Fourier de V de $\frac{2\pi}{N}$, la puissance de bruit totale en sortie s'exprime par :

$$P = \frac{1}{2\pi} \frac{q^2}{12} \int_{-\pi/N}^{\pi/N} \left| NTF\left(e^{jw}\right)\right|^2 dw \tag{2.20}$$

Le calcul de cette puissance peut être approché à l'ordre 1 pour un taux d'interpolation N élevé par :

$$P \approx \frac{q^2}{12} \frac{1}{2L+1} \frac{1}{N} \left(\frac{\pi}{N}\right)^{2L} \tag{2.21}$$

On note qu'en doublant N, la puissance est atténuée de 6L + 3 dB, ce qui revient à améliorer approximativement la résolution en sortie de L bits, où L est l'ordre des modulateurs ΣΔ [3]. La

figure 2.7 montre, à titre d'exemple, le module de la NTF(z) = $(1 - z^{-1})^2$ après décimation et interpolation pour différentes combinaisons M et N.

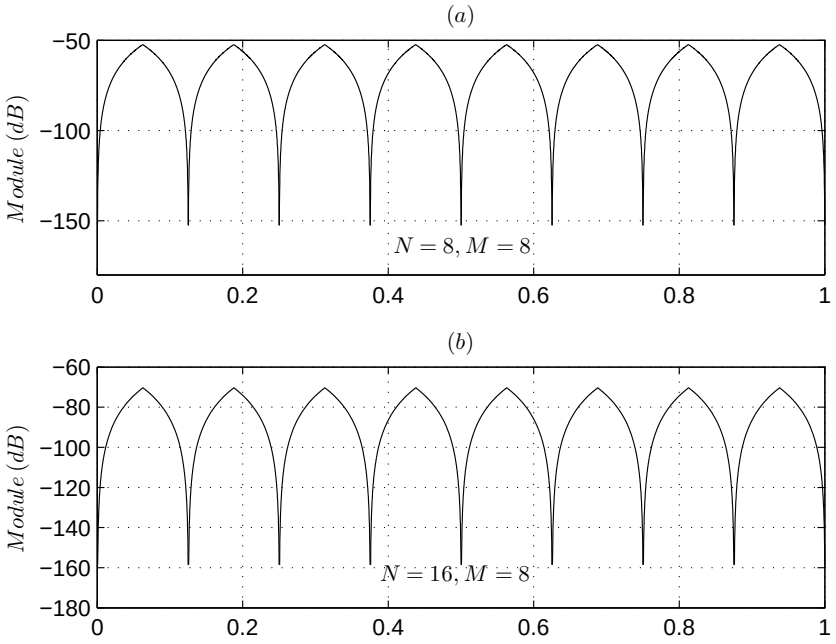

Fig. 2.7 – Module de la NTF(z) = $(1 - z^{-})^2$ après décimation et interpolation pour différentes valeurs de N et M.

On peut noter comme le montre l'équation (2.21) que la puissance totale de bruit dépend uniquement du rapport de suréchantillonnage N pour un ordre L donné. L'augmentation du nombre de voies M permet de compresser davantage le spectre en augmentant le nombre de répliques dans la bande [0, 1] et d'élargir par conséquent la bande de fonctionnement du convertisseur.

Ce paragraphe nous a permis de montrer que les principaux paramètres qui déterminent la performance de l'architecture TIΣΔ sont :

- l'ordre L du modulateur ΣΔ,
- le taux de suréchantillonnage N qui détermine la puissance de bruit en sortie,
- le nombre de voies M qui détermine la bande passante du convertisseur,
- l'ordre P du filtre passe-bas qui coupe le bruit hors bande.

2.2.2 Architecture à entrelacement temporel à la fréquence de Nyquist

Un convertisseur ΣΔ ne peut pas fonctionner à la fréquence de *Nyquist*. En effet, comme la bande du signal utile se trouve entre $\left[0, \frac{f_e}{2}\right]$, la puissance du bruit de quantification récupérée en sortie est plus grande que celle du CAN dans la boucle à cause de la mise en forme de bruit par le modulateur. Cependant, l'architecture à entrelacement temporel permet, par sa structure multi-cadence, le fonctionnement à la fréquence de *Nyquist*. Ceci correspond au cas où N = M dans l'architecture TIΣΔ (figure 2.2). Pour comprendre la possibilité de fonctionnement du modulateur à la fréquence de *Nyquist*, nous présentons le spectre du signal utile aux différents endroits de l'architecture TIΣΔ sur la figure 2.8.

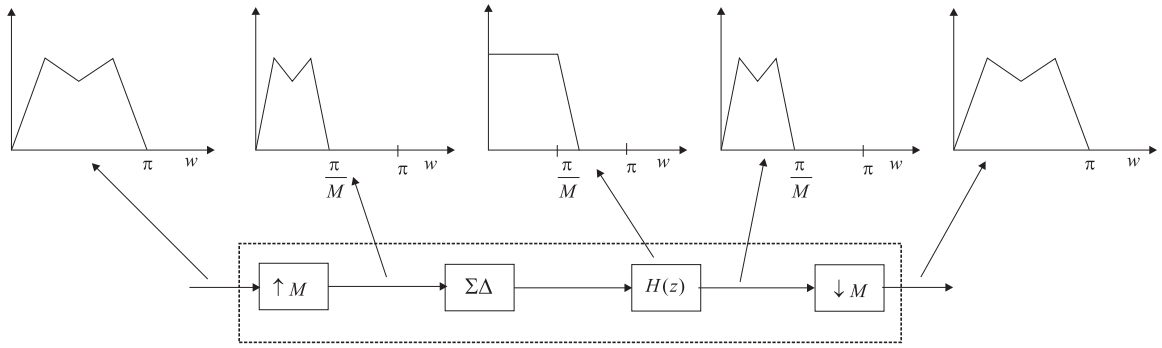

Fig. 2.8 – Spectre du signal utile dans une architecture TIΣΔ de *Nyquist*.

Le spectre du signal d'entrée est compressé d'un facteur M suite à l'interpolation d'un facteur M. Ensuite, après le passage dans le modulateur ΣΔ, un filtre passe-bas de bande plus large que $\frac{\pi}{\mathrm{M}}$ permet de laisser passer le signal utile. Finalement, la décimation d'un facteur M permet de récupérer le spectre initial du signal utile. Contrairement au signal utile, le bruit de quantification ne subit pas l'interpolation d'un facteur M. Il est filtré par le filtre passe-bas dans une bande de $\frac{\pi}{\mathrm{M}}$ et décimé par le facteur M permettant d'étaler la puissance de bruit sur une bande M fois plus large. Le point clef de ce fonctionnement est la compression du spectre du signal utile en entrée d'un facteur M. On note que le nombre de modulateurs M dans l'architecture TIΣΔ de *Nyquist* joue le rôle du suréchantillonnage dans le convertisseur ΣΔ conventionnel.

2.2.3 Sensibilité vis-à-vis des non idéalités du circuit

La reconstruction parfaite du signal d'entrée dans l'architecture TIΣΔ évoquée au dessus ne tient pas compte des imperfections des composants analogiques constituant le modulateur. Les sources d'erreur sont diverses. Elles comprennent le bruit en 1/f, l'erreur sur les valeurs nominales des composants (capacités, résistances) et le gain fini des amplificateurs. Ces sources d'erreurs se manifestent entre autre par un gain et un décalage en tension sur le signal de chaque voie. Comme ces sources d'erreurs ne sont pas identiques sur toutes les voies, la reconstruction du signal d'entrée ne peut pas être parfaite, c'est ce que l'on appelle la disparité entre les voies. En tenant compte de ces sources d'erreurs, le signal $y_k[n]$ en sortie du modulateur n'est pas une version retardée du signal d'entrée. Il s'exprime par [30, 31] :

$$y_k\,[n] = (1 + a_k)\,x_k\,[n] + b_k \tag{2.22}$$

Où $\mathrm{a_k}$ est l'erreur sur le gain et $\mathrm{b_k}$ l'erreur de décalage en tension. Dans la suite, en raison de la nature stochastique des erreurs analogiques, $\mathrm{a_k}$ et $\mathrm{b_k}$ sont supposées aléatoires décorrélées entre elles, de distribution gaussienne, de moyenne nulle et de variance respective σ^2_{a} et σ^2_{b}.
En considérant l'erreur de gain $(1 + a_k)\,x_k\,[n]$ dans l'équation (2.22), le signal en sortie de l'architecture TIΣΔ s'exprime par [31] :

$$\begin{aligned} &Y(z) = z^{-(M-1)}X(z) + z^{-(M-1)}A_0X(z) + z^{-(M-1)}L(z) \\ &\text{avec} \quad A_l = \tfrac{1}{M}\sum_{k=0}^{M-1} a_kW^{lk} \quad \text{et} \quad L(z) = \sum_{l=0}^{M-1} X\left(zW^{-l}\right)A_l \end{aligned} \tag{2.23}$$

Nous pouvons noter que :

– Le premier terme est une version retardée du signal d'entrée. Ce retard est dû au temps nécessaire pour reconstruire le signal d'entrée à partir des différentes voies.
– Le deuxième terme est le signal global en sortie multiplié par un gain A_0, A_0 est la moyenne des erreurs de gain a_k pour toutes les voies. Ce terme n'est pas gênant parce qu'il ne déforme pas le spectre du signal utile.
– Le troisième terme représente les recouvrements spectraux introduits par cette erreur de gain. Ce dernier terme déforme le spectre du signal utile et il est difficile à corriger.

En supposant que les différentes sources d'erreur a_k sont décorrélées entre elles, l'amplitude moyenne des répliques spectrales par rapport au fondamental est une variable aléatoire de distribution gaussienne [31, 7] de moyenne μ_g et d'écart-type σ_g données respectivement par :

$$\begin{aligned} \mu_g &= \tfrac{\sigma_a}{2}\sqrt{\tfrac{\pi}{M}} \\ \sigma_g &= \sigma_a\sqrt{\tfrac{\left(1-\frac{\pi}{4}\right)}{M}} \end{aligned} \tag{2.24}$$

On peut noter que μ_g et σ_g sont proportionnelles à σ_a et inversement proportionnelles à $\sqrt{M}$. L'augmentation du nombre de voies M augmente le nombre de composants du recouvrement spectral à cause des erreurs de gain mais elle diminue la moyenne et la variance de leur amplitude.

L'effet du décalage en tension a été calculé de la même façon dans [32]. L'expression du signal en sortie en tenant compte seulement du décalage en tension b_k est donnée par :

$$Y(z) = z^{-(M-1)}X(z) + \frac{1}{M}\sum_{k=0}^{M-1} b_k W_M^k \tag{2.25}$$

Cette source d'erreur crée des raies aux fréquences multiples de $\frac{f_e}{M}$ sur le spectre du signal en sortie, ce qui amène à une diminution notable de la résolution du convertisseur TIΣΔ. L'amplitude de ces raies est une variable aléatoire gaussienne de moyenne $\mu_{dc} = \sigma_b M\sqrt{\pi\sqrt{2}}$. On note que plus le nombre de voies augmente, plus l'erreur de décalage en tension augmente, ce qui rend indispensable la calibration lorsque le nombre de voies est élevé. La calibration consiste à supprimer l'effet de ces erreurs comme le montre la figure 2.9.

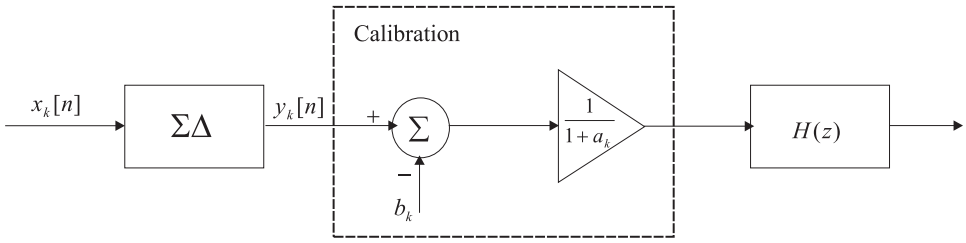

Fig. 2.9 – Modulateur ΣΔ suivi d'une calibration corrigeant les erreurs de gain et de décalage en tension.

2.2.4 Méthode de calibration avec un modulateur ΣΔ numérique

La correction des erreurs analogiques de gain et de décalage en tension dans un modulateur ΣΔ peut s'effectuer en ajustant la tension de référence du CNA dans la boucle de retour. Cette solution s'avère complexe à réaliser dans le domaine analogique en raison de la vitesse de fonctionnement très élevée (fréquence de suréchantillonnage). Cette correction peut se faire dans le domaine numérique avec un multiplieur derrière le modulateur (figure 2.9). Or, la réalisation de

la multiplication à une fréquence très élevée exige des ressources matérielles importantes et par conséquent augmente la surface et le coût. Une solution pour contourner ce problème de vitesse et de surface est d'utiliser un modulateur ΣΔ numérique [33] après le modulateur analogique. Ce modulateur est moins encombrant en surface que le multiplieur et fonctionne à très grande vitesse. Le principe de cette méthode est présenté sur la figure 2.10.

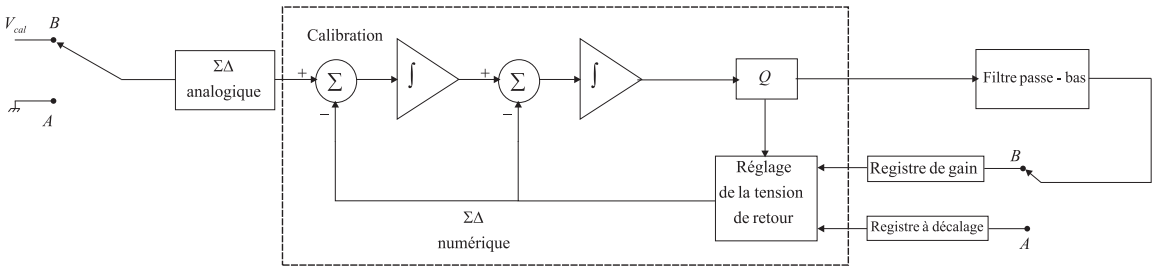

Fig. 2.10 – Méthode de calibration avec un modulateur ΣΔ numérique.

L'implantation du modulateur numérique avec un niveau de retour ajustable est simple à réaliser. Le fonctionnement de cette méthode de calibration consiste à :

1. connecter l'entrée du modulateur à la masse (point A) pour déterminer la composante continue b_k et la stocker dans le registre dédié au décalage.
2. connecter l'entrée du modulateur à la tension continue de calibration V_{cal} (point B) pour estimer le gain a_k du modulateur et le stocker dans le registre dédié au gain.

Le filtre passe-bas après le modulateur sert à estimer le gain et la tension de décalage. L'ordre du modulateur, de type numérique, doit être choisi de façon à ce que le niveau de bruit introduit par ce modulateur soit inférieur à celui du modulateur analogique. En pratique, pour une calibration en temps réel, l'architecture TIΣΔ doit avoir un étage en plus qui sert de remplaçant pour chacune des voies lorsqu'elles sont en phase de calibration. Cette solution donne de bons résultats [7] mais elle nécessite des ressources matérielles importantes.

2.2.5 Architecture à entrelacement temporel à multiplexage aléatoire

La disparité de gain entre les voies de l'architecture à entrelacement temporel introduit du recouvrement spectral à cause des répliques du spectre générées par cette erreur (équation (2.23)). Pour éliminer l'effet de ces répliques, l'idée est de multiplier le signal d'entrée par une séquence pseudo-aléatoire $\{+1, -1\}$ puis de multiplier le signal en sortie avec la même séquence d'entrée mais avec un certain retard. La multiplication par la séquence pseudo-aléatoire permet d'étaler le spectre du signal d'entrée sur toute la plage fréquentielle et ainsi de le noyer dans le spectre de bruit. Les répliques de ce spectre crées par la disparité du gain entre les voies sont aussi étalées sur toute la plage fréquentielle. Le spectre du signal utile est alors récupéré en sortie en multipliant par la même séquence aléatoire. C'est le processus du désétalement. Cette technique est efficace pour diminuer l'effet de recouvrement spectral mais elle n'est pas efficace pour la suppression des raies spectrales périodiques générées par les décalages en tension [32].

Une autre technique pour supprimer les effets indésirables de la disparité de gain et de décalage en tension entre les voies consiste à rendre aléatoire le multiplexage entre les différentes voies. Le bon fonctionnement de l'architecture à entrelacement temporel impose que deux échantillons successifs à l'entrée de chaque voie doivent être séparés au minimum de M cycles d'horloge. Cette condition est réalisée en ajoutant une voie supplémentaire et en effectuant le choix entre les voies

d'une façon aléatoire. La figure 2.11 présente le mécanisme de sélection aléatoire dans le cas de 4 voies.

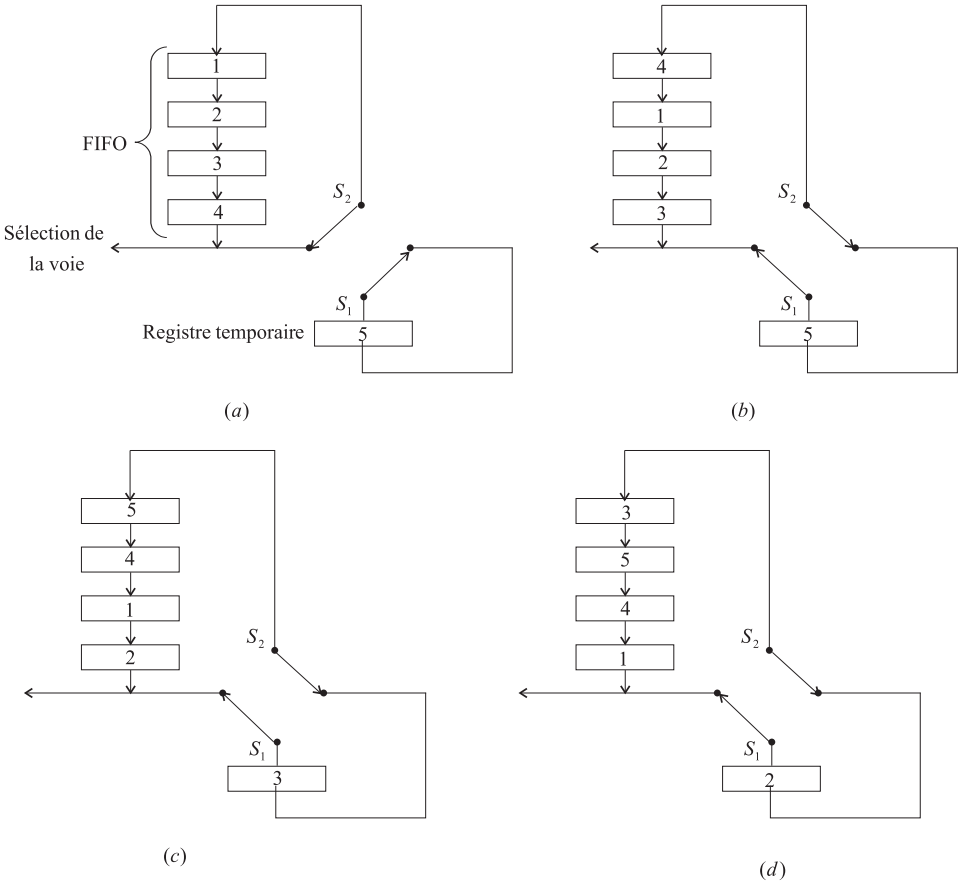

Fig. 2.11 – Mécanisme de sélection aléatoire des voies.

Ce mécanisme repose sur :

- un registre à décalage de M valeurs. La lecture de ce registre à décalage se fait suivant le principe « premier entré, premier sorti »(FIFO « First In, First Out »). Ce registre permet de garantir que deux utilisations successives de la même voie sont séparées au moins par M cycles d'horloge.
- deux interrupteurs S_1 et S_2. Ces deux interrupteurs sont commandés par un générateur aléatoire monobit. Ils permettent de mettre la sortie du registre soit dans la queue du registre soit dans le registre tampon et celle du registre tampon dans la queue du registre.

Avec cette technique, le terme A_l dans l'équation (2.23) apparaît comme un bruit blanc et par conséquent le bruit introduit par le recouvrement spectral est blanc et sera étalé dans toute la plage fréquentielle [34, 32]. Cette technique a permis aussi de casser la périodicité et de blanchir le spectre de l'erreur introduite par le décalage en tension qui se manifeste par des raies spectrales tous les $\frac{f_e}{M}$ [34, 5].

2.2.6 Architecture à entrelacement temporel à base de modulateurs ΣΔ passe-haut

La présence de bruit basses fréquences dans les modulateurs ΣΔ passe-bas est incontournable. L'élimination de ce bruit est une préoccupation majeure à cause de la dégradation de la résolution du convertisseur qu'il provoque. Une technique a été proposée dans [35] pour résoudre ce problème. Cette technique consiste à multiplier le signal d'entrée du modulateur par la séquence $\cos(\pi n)$. Ceci a pour effet de translater le signal d'entrée vers les hautes fréquences autour de la fréquence $\frac{f_e}{2}$. Ensuite le signal est traité tout en étant à l'écart du bruit basses fréquences. Puis, il est multiplié par la même séquence pour la ramener de nouveau en basses fréquences.

Dans le cadre de l'architecture TIΣΔ, une architecture à base de modulateurs ΣΔ passe-haut, inspirée de la technique décrite ci-dessus, a été proposée dans [6]. Cette architecture est présentée sur la figure 2.12.
Dans cette architecture, la multiplication par $\cos(\pi n)$ a été remplacée par une décimation et une interpolation d'un facteur M pour créer une réplique du signal d'entrée autour de $\frac{f_e}{2}$. Cette technique exige que le nombre de voies M soit pair pour avoir une réplique autour de $\frac{f_e}{2}$. Après avoir été translaté à la fréquence $\frac{f_e}{2}$, le signal est traité par un modulateur ΣΔ passe-haut et un filtre passe-haut $H_{ph}(z)$. Le bruit basses fréquences, le bruit de quantification introduit par le modulateur passe-haut et les répliques du signal d'entrée sont éliminés par le filtre passe-haut. La figure 2.13 présente les signaux dans le domaine fréquentiel après chaque bloc pour une seule voie avec des modulateurs passe-bas et passe-haut.
On peut noter que :

- le filtre numérique $H_{ph}(z)$ est lié au nombre de voies M. Plus le nombre de voies M est grand, plus les contraintes sur le filtre numérique sont sévères pour supprimer les répliques du signal d'entrée.
- le signal d'entrée se trouve à l'abri des erreurs de décalage en tension qui se trouve en basses fréquences. En revanche, les erreurs de gain persistent toujours.
- le convertisseur proposé est un convertisseur de type *Nyquist* (N = M).

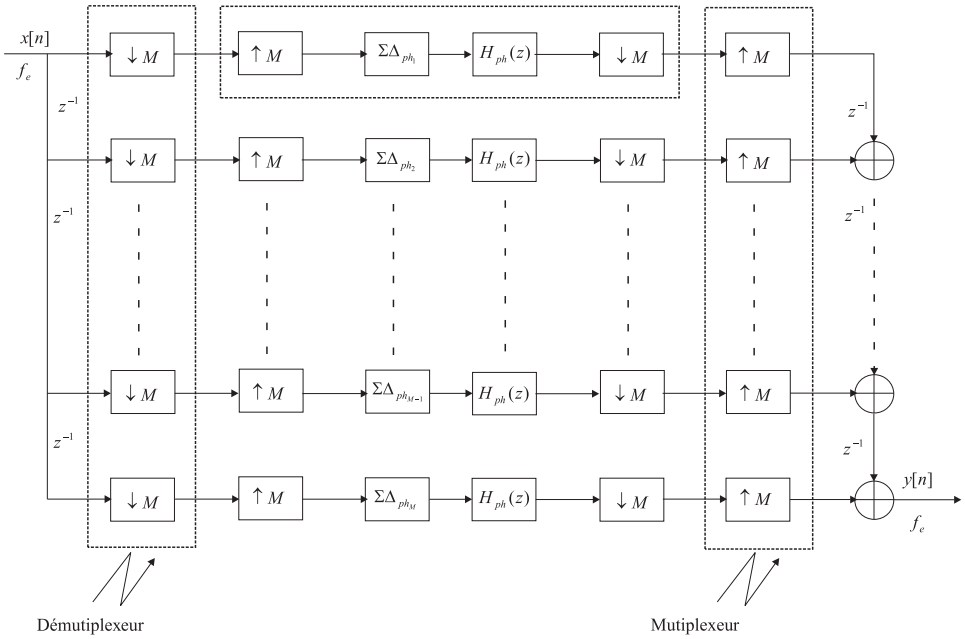

Fig. 2.12 – Architecture TIΣΔ de *Nyquist* à base de modulateurs ΣΔ passe-haut.

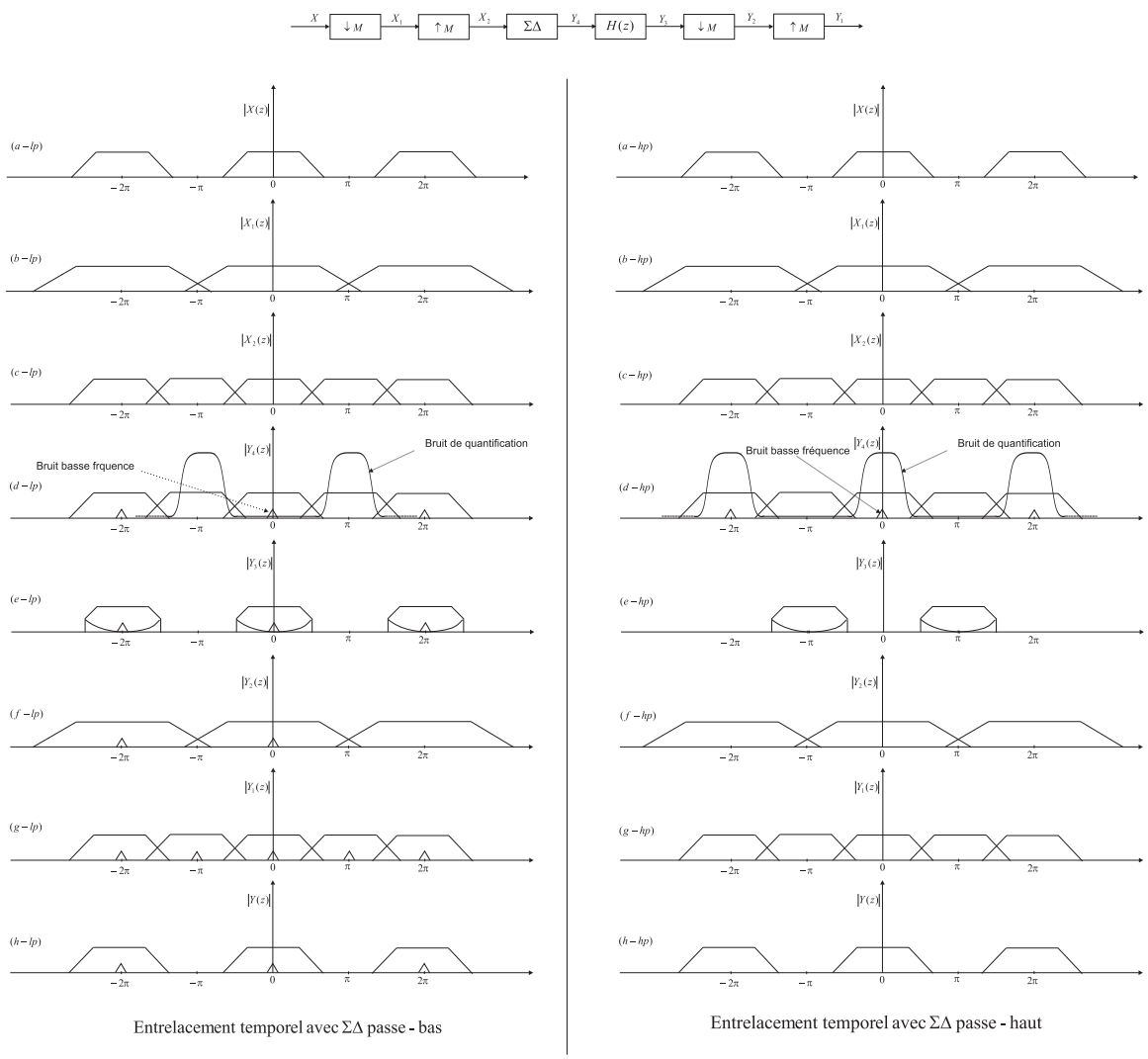

Fig. 2.13 – Représentation dans le domaine fréquentiel des signaux dans une voie avec TIΣΔ passe-bas et passe-haut.

Le problème lié à la disparité entre les voies due aux erreurs de gain n'a pas été résolu avec l'utilisation des modulateurs passe-haut. La correction de ce gain peut être effectuée en ajoutant un multiplieur après la sortie de chaque voie comme illustré par la figure 2.14.
Lorsque les coefficients multiplicatifs $W = [w_1, w_2 \dots w_M]$ sont choisis pour être égaux à l'inverse du gain de chaque voie, le gain de voie total sera unitaire et donc les effets de la disparité entre les gains seront éliminés. La détermination des valeurs du vecteur W est basée sur l'algorithme des moindres carrés. L'algorithme des moindres carrés avec signe du signal d'entrée $SD - LMS$ (Sign Data Least Mean Square) présente une performance meilleure que d'autres versions de l'algorithme telles que $SE - LMS$ et $SS - LMS$ (Sign-Error et Sign-Sign) [29, 6]. L' algorithme des moindres carrés présente une remarquable simplicité d'implantation. Avec cet algorithme, les poids w sont calculés d'une façon itérative par l'équation :

$$\hat{W}_{k+1} = \hat{W}_k + \mu \left(y_{ideal}\left[n\right] - y\left[n\right] \right) \hat{Y} \tag{2.26}$$

avec :
$\hat{Y}$: vecteur des sorties des différentes voies $\hat{Y} = [y_1, y_2, \dots y_M]$,

k : indice de temps,
y_{ideal} : sortie idéale qui est juste une version décalée du signal d'entrée,
μ : pas de l'algorithme. Il détermine la vitesse de convergence.

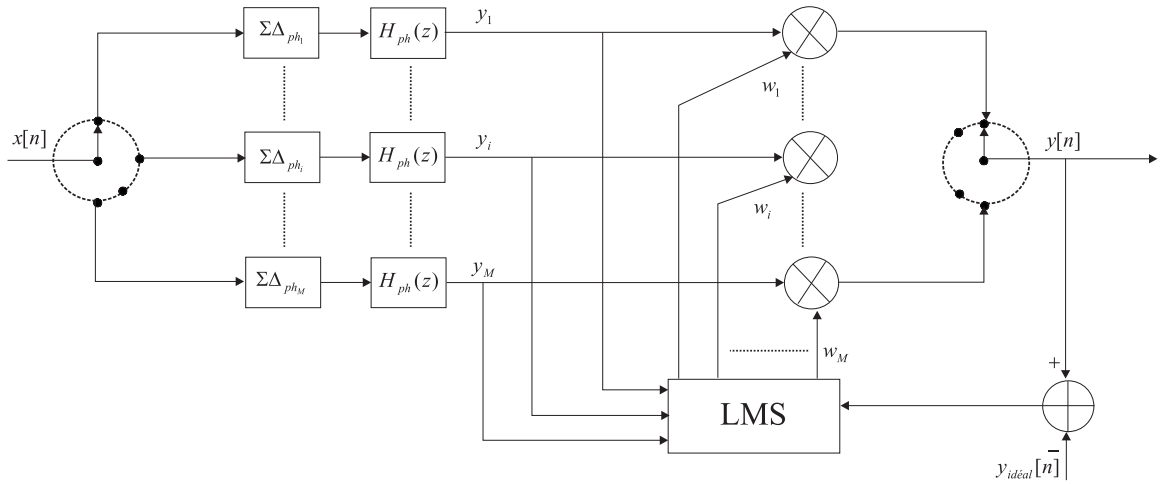

Fig. 2.14 – Correction de gains de l'architecture TIΣΔ passe-haut.

2.3 Architecture à base de modulation de *Hadamard*

L'architecture à base de modulation de *Hadamard* ΠΣΔ [1] est une autre voie pour la parallélisation des modulateurs ΣΔ. Cette architecture est présentée sur la figure 2.15.

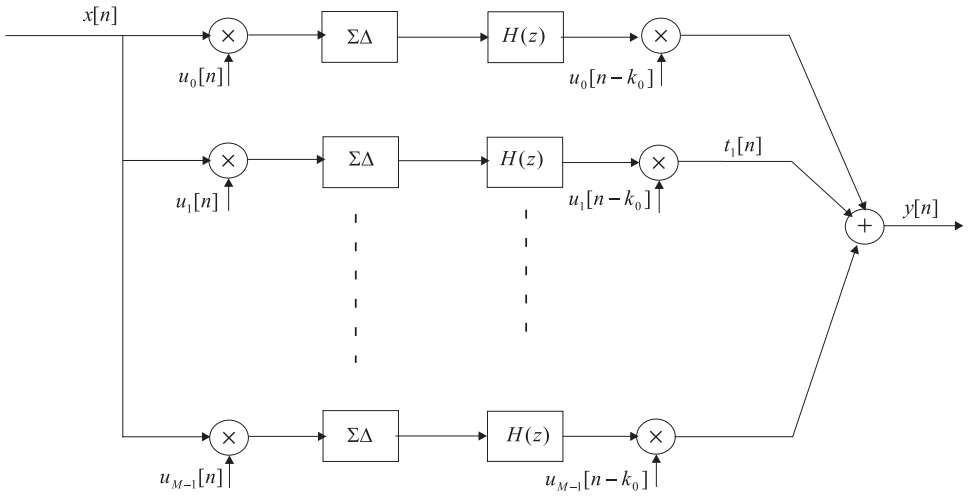

Fig. 2.15 – Architecture parallèle à base de la modulation de *Hadamard*.

Le signal d'entrée est appliqué à tous les modulateurs en même temps. Ensuite, le signal est multiplié par la séquence de *Hadamard* de valeurs ±1 à l'entrée de chaque voie. Cette multiplication est simple à réaliser. Elle nécessite juste un changement de signe quand la valeur d'un élément de la séquence est égale à −1. Après le passage dans le modulateur ΣΔ et le filtre passe-bas H(z), le signal est multiplié par la même séquence mais retardé avant d'être sommé aux sorties

des autres voies pour reconstruire le signal global en sortie. La séquence de *Hadamard* $u_r[n]$ $(0 \leq r \leq M-1)$ est déterminée à partir de la matrice carré de *Hadamard* [1]. La séquence $u_r[n]$ est la ligne d'indice r de la matrice H_d répétée d'une façon cyclique comme le montre l'équation suivante :

$$u_r[n] = m[r, n \bmod M] \tag{2.27}$$

Où $m[i,j]$ est l'élément de la ligne i et de la colonne j de la matrice H_d.
La matrice de *Hadamard* existe si et seulement si sa dimension M est une puissance de 2. Elle est construite récursivement de la façon suivante :

$$H_{d_i} = \begin{bmatrix} H_{d_{i-1}} & H_{d_{i-1}} \\ H_{d_{i-1}} & -H_{d_{i-1}} \end{bmatrix} \quad \text{avec} \quad H_{d_0} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \tag{2.28}$$

Cette condition implique que si pour un nombre de voies M, la résolution obtenue n'est pas satisfaisante, il faut au moins multiplier par 2 le nombre de voies afin de construire une autre matrice de *Hadamard*. Ceci multiplie par 2 les ressources matérielles et par conséquent la surface d'implantation.

Principe de fonctionnement et performances théoriques

Comme dans l'architecture TIΣΔ, l'explication du principe de fonctionnement se base sur le modèle linéaire du modulateur ΣΔ. Ce modèle linéaire permet d'exprimer le signal en sortie de l'architecture parallèle y[n] (figure 2.16) par $y[n] = y_x[n] + y_e[n]$, où $y_x[n]$ est la sortie qui correspond au signal d'entrée x[n] et $y_e[n]$ la sortie qui correspond au bruit de quantification $e_i[n]$ des différentes voies.

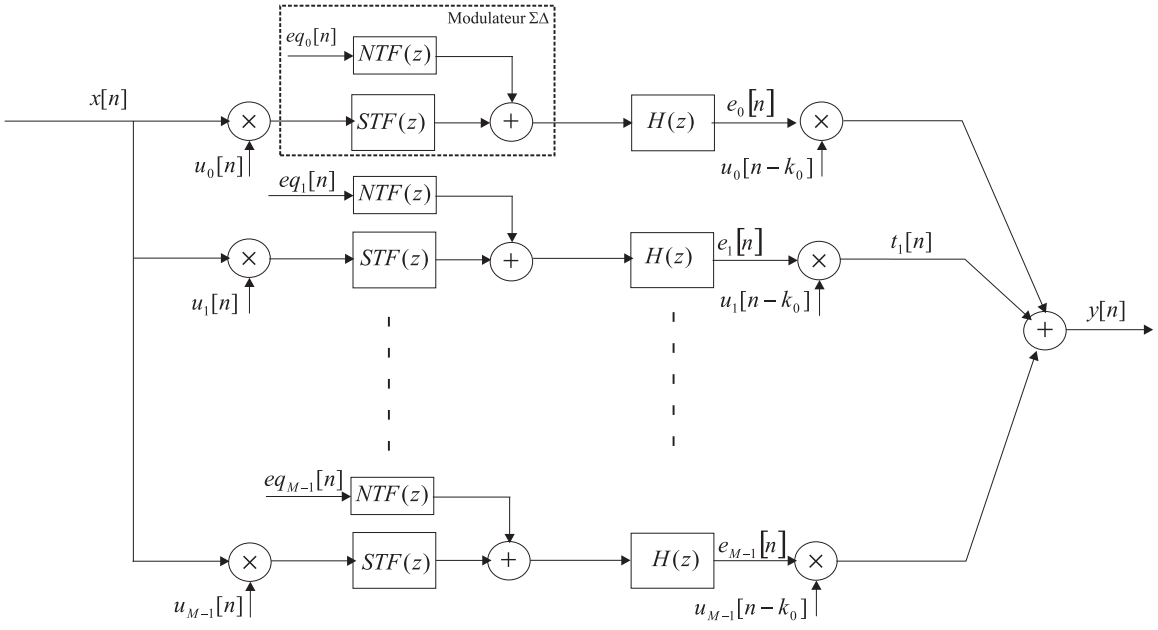

Fig. 2.16 – Architecture ΠΣΔ avec le modèle linéaire du modulateur ΣΔ.

[1] La matrice de *Hadamard* H_d est une matrice unitaire composée de valeurs +1 et −1 et telle que $H_d{}^T H_d = MI$ où M est la dimension de la matrice

Expression du signal utile $y_x[n]$

La fonction de transfert par rapport au signal STF(z) du modulateur ΣΔ peut être assimilée à un simple retard. En se basant sur cette hypothèse, nous présentons, à titre d'exemple, sur la figure 2.17 une architecture à deux voies où l'on a supposé que le retard introduit par le modulateur est d'une période d'échantillonnage [36]. Le filtre passe-bas H(z) est à 3 coefficients.

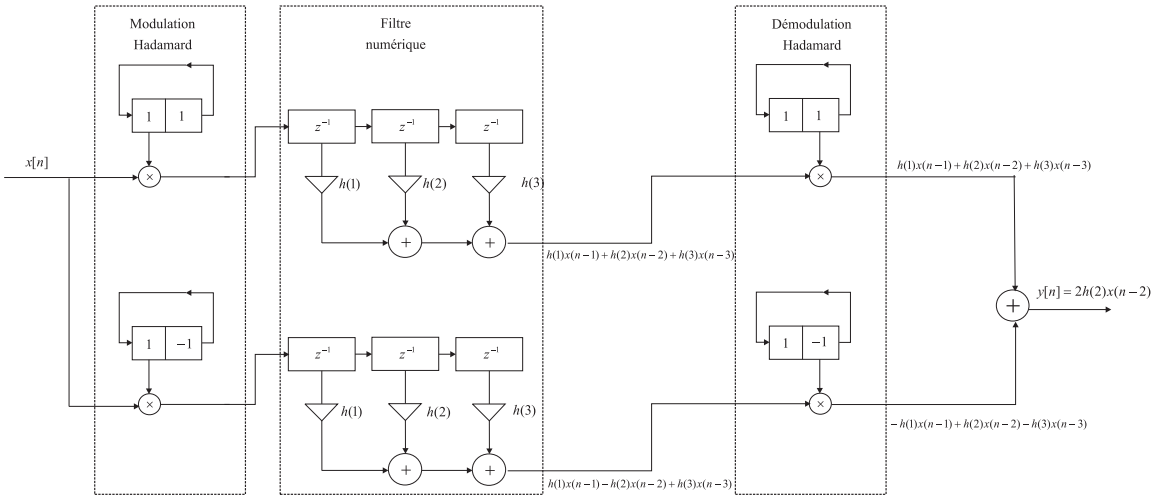

Fig. 2.17 – Exemple d'illustration pour deux voies avec $STF(z) = z^{-1}$.

Nous pouvons constater que grâce à la modulation et à la démodulation par la séquence de *Hadamard*, le signal d'entrée ne voit que le coefficient central dans ce cas particulier. Donc en choisissant $h(2) = \frac{1}{2}$, le signal d'entrée x[n] subit seulement un retard lors de son passage dans l'architecture parallèle ΠΣΔ. Afin de généraliser la condition sur les coefficients du filtre passe-bas H(z) pour assurer une reconstruction parfaite du signal en sortie, nous partons de l'expression du signal $t_r[n]$ à la sortie de la voie d'indice r (voir figure 2.16). Le signal $t_r[n]$ s'exprime par :

$$\begin{aligned} t_r[n] &= [x[n] \times u_r[n] * (s[n] * h[n])]\, u_r[n-k_0] \\ &= [x[n] \times u_r[n] * (q[n])]\, u_r[n-k_0] \\ &= \sum_{k=0}^{\infty} q[k]\, x[n-k]\, u_r[n-k]\, u_r[n-k_0] \end{aligned} \tag{2.29}$$

où s[n] est la réponse impulsionnelle de la fonction STF(z) et h[n], la réponse impulsionnelle de la fonction H(z). La composante du signal en sortie due au signal utile est la somme des sorties des différentes voies. Elle s'exprime par :

$$\begin{aligned} y_x[n] &= \sum_{r=0}^{M-1} t_r[n] \\ &= \sum_{k=0}^{\infty} q[k]\, x[n-k] \sum_{r=0}^{M-1} u_r[n-k]\, u_r[n-k_0] \end{aligned} \tag{2.30}$$

Or, d'après la définition de la matrice de *Hadamard*, la ligne $u_r[n]$ se répète d'une façon cyclique et vérifie la relation suivante :

$$\sum_{r=0}^{M-1} u_r[n-k]\, u_r[n-k_0] = MC_M[k-k_0] = \begin{cases} M & \text{si } k-k_0 \text{ est multiple de } M \\ 0 & \text{sinon} \end{cases} \tag{2.31}$$

Donc, le signal $y_x[n]$ s'exprime par :

$$y_x[n] = M \sum_{k=0}^{\infty} q[k]\, x[n-k]\, C_M[k-k_0] \tag{2.32}$$

On peut noter facilement d'après l'équation (2.32) que le signal d'entrée x[n] voit seulement, grâce à la modulation et à la démodulation de *Hadamard*, le coefficient central et les coefficients de la réponse impulsionnelle q[n] ($s[n] * h[n]$) dont l'indice est multiple de M. Par conséquent, en imposant le coefficient central de q[n] à $\frac{1}{M}$ et les coefficients d'indices multiples de M à zéro, le signal en sortie $y_x[n]$ est une version retardée du signal d'entrée. Comme la fonction de transfert STF(z) introduit un simple retard ($s[n] = \delta[n-L]$, L étant le retard), la condition sur q[n] est vérifiée en choisissant les coefficients du filtre passe-bas H(z) d'ordre P suivant la relation :

$$h[n] = \begin{cases} M & \text{si} \quad n = \frac{P-1}{2} \\ \\ 0 & \text{si} \quad n = \frac{P-1}{2} + mM,\ m = \pm 1, \ldots \pm \left\lfloor \frac{P-1}{2M} \right\rfloor \end{cases} \tag{2.33}$$

Le respect de la condition 2.33 avec l'effet de la modulation et de la démodulation de *Hadamard* permet de supprimer l'effet du filtrage et de récupérer en sortie une version retardée du signal d'entrée.

Expression du signal de bruit $y_e[n]$

Au contraire du signal utile, le bruit de quantification introduit par les modulateurs subit seulement la démodulation par la séquence de *Hadamard*. Chacune de ces sources de bruit traverse la fonction de mise en forme de bruit NTF(z) et le filtre passe-bas H(z) avant d'être multipliée par $u_r[n-k_0]$ et sommée à d'autres signaux de bruit (voir figure 2.16) pour former le signal global donné par :

$$y_e[n] = \sum_{r=0}^{M-1} u_r[n-k_0]\, e_r[n] \tag{2.34}$$

Afin de déterminer la densité spectrale du bruit en sortie, nous exprimons d'abord la fonction d'autocorrélation de ce bruit $R_{y_e y_e}[k]$ en supposant que les sources de bruit des différents modulateurs sont décorrélées entre elles. Elle est donnée par :

$$\begin{aligned} R_{y_e y_e}[k] &= \mathrm{E}[y_e[n]\, y_e[n+k]] \\ &= \sum_{r=0}^{M-1} \sum_{q=0}^{M-1} u_r[n-k_0]\, u_q[n-k_0+k]\, \mathrm{E}[e_r[n]\, e_q[n+k]] \\ &= M C_M[k]\, R_{e_r e_r}[k] \end{aligned} \tag{2.35}$$

La densité spectrale de puissance est obtenue à partir de la transformée de Fourier de la fonction d'autocorrélation $R_{y_e y_e}[k]$. Étant donné que la séquence $C_M[k]$ peut être écrite aussi sous la forme suivante (voir annexe B, § B.4) :

$$C_M[k] = \frac{1}{M} \sum_{l=0}^{M-1} W_M^{-lk}, \quad \text{avec} \quad W_M = e^{-j\frac{2\pi}{M}} \tag{2.36}$$

La densité spectrale du bruit en sortie s'exprime par :

$$S_{y_e y_e}\left(e^{jw}\right) = \sum_{k=0}^{M-1} S_{e_r e_r}\left(e^{j\left(w-2\pi\frac{k}{M}\right)}\right) \tag{2.37}$$

où $S_{e_r e_r}\left(e^{jw}\right)$ est la densité spectrale du bruit e_r [n]. Elle s'exprime, en utilisant le modèle linéaire du modulateur par :

$$S_{e_r e_r}\left(e^{jw}\right) = \frac{q^2}{12}\left|NTF\left(e^{jw}\right) H\left(e^{jw}\right)\right|^2 \tag{2.38}$$

À l'instar de l'architecture à entrelacement temporel, l'architecture à base de la modulation de *Hadamard* crée M répliques de la densité spectrale $S_{e_r e_r}\left(e^{jw}\right)$ dans toute la plage fréquentielle.

La performance de cette architecture en terme de puissance de bruit dépend du choix du filtre passe-bas H(z) dont le rôle est d'éliminer le bruit en dehors de la bande utile. Ce filtre est conçu de façon à respecter la contrainte 2.33 et les autres coefficients sont choisis de façon à minimiser le bruit [1]. L'ordre du filtre ne peut pas être augmenté indéfiniment. Il a été vérifié par simulation qu'un filtre d'ordre supérieur à 2LM − 1 n'apporte pas d'amélioration significative de la performance [36].

Cette architecture est aussi sensible aux imperfections des composants analogiques du modulateur. L'effet de ces erreurs est le même que pour l'architecture à entrelacement temporel. Nous pouvons utiliser les méthodes de correction utilisées dans les architectures à entrelacement temporel [7, 32].

Cette architecture parallèle peut être utilisée également avec des modulateurs ΣΔ avec un facteur de suréchantillonnage N. Le principe de fonctionnement reste le même. Cela permet d'ajouter un degré de liberté sur le contrôle du niveau de la puissance de bruit en sortie. Dans ce cas, chaque colonne de la matrice de *Hadamard* est répétée N fois.

2.4 Architecture à base de décomposition fréquentielle

L'architecture à base de décomposition fréquentielle FBD (Frequency Band Decomposition) a été inspirée de la structure de banc de filtres [27].

Cette architecture se compose (figure 2.18) d'un modulateur passe-bas dans la première voie, de modulateurs passe-bande et d'un modulateur passe-haut dans la dernière voie. Elle permet d'obtenir une densité spectrale de bruit en sortie identique à celle obtenue avec les deux architectures précédentes en découpant toute la plage fréquentielle en M sous bandes [2, 37, 38]. Après chaque modulateur, un filtre passe-bande centré au minimum de la fonction de transfert de bruit élimine le bruit de quantification hors de cette bande. Ainsi, chaque voie laisse passer une portion de la bande totale, et en additionnant les signaux sur tous les canaux, on retrouve le signal initial numérisé.

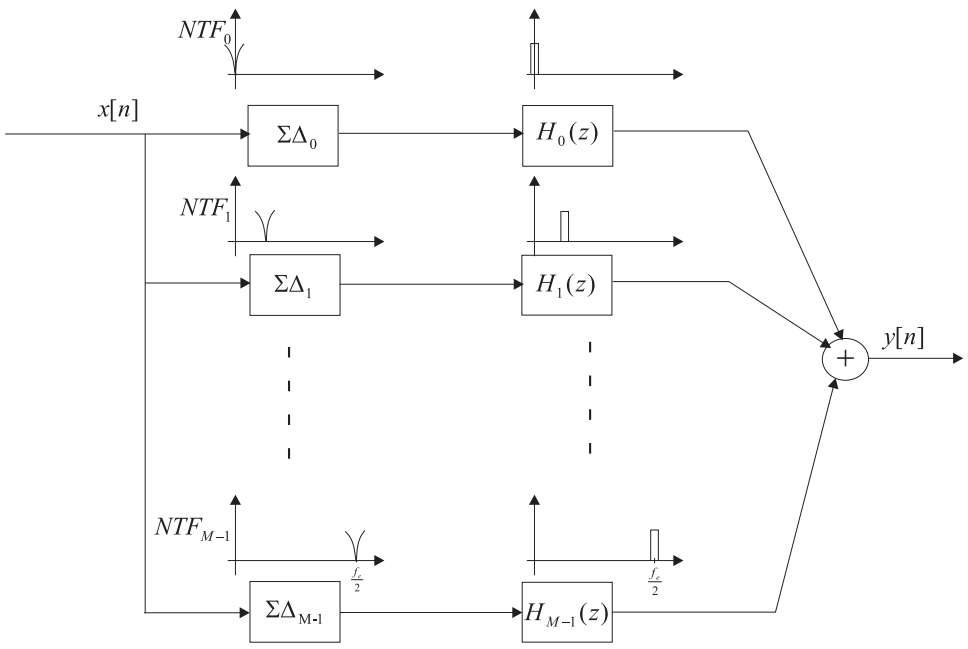

Fig. 2.18 – Architecture parallèle à décomposition fréquentielle FBD.

Cette architecture présente deux avantages majeurs :

- Elle est robuste aux erreurs de décalage en tension, en raison de l'utilisation de modulateurs passe-bande, à l'exception du premier modulateur.
- La disparité de gain entre les différentes voies n'a pas d'effet sur la linéarité du convertisseur. Elle n'introduit pas de raies spectrales qui peuvent dégrader les performances. Elle se manifeste sous la forme d'ondulations dans l'amplitude du spectre en sortie. L'amplitude de ces ondulations est directement liée à l'amplitude de l'erreur de gain de chaque voie.

En revanche, Cette architecture est plus complexe à réaliser. Elle nécessite l'implantation de M modulateurs différents. De plus, les filtres passe-bande numériques derrière les modulateurs doivent être d'ordre élevé pour éliminer le bruit de quantification en raison de leur faible zone de transition [2].

2.5 Comparaison des trois architectures parallèles

Les trois architectures parallèles développées ci-dessus peuvent êtres comparées suivant les critères annoncés ci-après :

- **Les performances théoriques :** les trois architectures présentent des niveaux de puissance de bruit en sortie plus ou moins identiques. En effet, les trois architectures peuvent être décrites avec le principe de modulation et de démodulation (voir paragraphe 2.3). La seule différence entre les différentes architectures est le choix de la matrice unitaire. Dans le cas de l'architecture à entrelacement temporel, la matrice de *Hadamard* est remplacée par la matrice identité. À titre d'exemple, la matrice identité d'ordre M = 4 est donnée par :

$$\mathrm{H}_{I_4} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{2.39}$$

On peut noter d'après cette matrice que chaque voie est utilisée une fois sur 4 (chaque ligne de la matrice représente la séquence de modulation). Dans l'architecture à décomposition fréquentielle, la matrice unitaire est la matrice de la transformée de Fourier discrète. Elle est donnée par :

$$\mathrm{H}_{FD} = \frac{1}{\sqrt{M}} \begin{bmatrix} 1 & 1 & \dots & 1 \\ 1 & W & \dots & W^{M-1} \\ \vdots & \vdots & & \vdots \\ 1 & W^{M-1} & \dots & W^{(M-1)^2} \end{bmatrix} \quad \text{avec} \quad W = e^{j\frac{2\pi}{M}} \tag{2.40}$$

La ligne d'indice r de la matrice $\mathrm{H_{FD}}$ est la séquence de modulation qui va translater le spectre du signal d'entrée autour de la fréquence normalisée $\frac{\mathrm{r}}{\mathrm{M}}$.

- **Sensibilité aux erreurs de gain :** La disparité entre les gains des différentes voies avec les architectures ΠΣΔ et TIΣΔ introduit des raies spectrales très gênantes. Ces architectures nécessitent donc une méthode de calibration qui peut être coûteuse. En revanche, avec l'architecture FBD, cette source d'erreur se manifeste sous la forme d'ondulations sur l'amplitude du spectre du signal de sortie qui ne dégradent pas le rapport signal sur bruit. Par conséquent, la résolution attendue par cette architecture est maintenue.
- **Sensibilité aux erreurs de décalage en tension :** À cause de la démodulation dans les architectures ΠΣΔ et TIΣΔ, l'erreur de décalage en tension localisée à la fréquence nulle, se trouve répétée tous les $\frac{\mathrm{f_e}}{\mathrm{M}}$. Ceci augmente la puissance de bruit et dégrade ainsi la performance de ces architectures. Cependant, l'architecture à décomposition fréquentielle ne présente pas ce problème. L'erreur de décalage en tension est introduite uniquement par le premier modulateur.
- **Dynamique d'entrée :** Les trois architectures parallèles présentent une dynamique élevée par rapport à celle obtenue avec un seul modulateur. Cependant, l'architecture TIΣΔ présente la plus grande dynamique car l'énergie du signal d'entrée est partagée sur les M voies. Ce qui n'est pas le cas dans les autres architectures.
- **Complexité de réalisation :** Les architectures ΠΣΔ et TIΣΔ présentent un notable avantage vis-à-vis de l'implémentation pour deux raisons : premièrement, les modulateurs ΣΔ passe-bas sont identiques pour toutes les voies, ce qui facilite l'implémentation. Deuxièmement, les filtres passe-bas H(z) sont identiques et n'exigent pas un gain de 0 dB dans la bande passante du signal utile de chaque voie. Il impose seulement que le coefficient du centre soit égal à $\frac{1}{\mathrm{M}}$ et que tous les coefficients d'indices multiples de M soient nuls.
  En revanche, l'architecture FBD nécessite des modulateurs ΣΔ passe-bande différents pour chaque voie et un filtre passe-bas ou passe-bande de gain 0 dB dans la bande passante du signal utile de chaque voie. Cette contrainte nécessite des filtres d'ordres élevés pour éliminer le bruit de quantification en dehors de la bande utile car la zone de transition entre les bandes adjacentes est très étroite.
- **Gigue d'horloge :** L'erreur sur les instants d'échantillonnage est très gênante avec les architectures ΠΣΔ et TIΣΔ. Elle se manifeste par des raies spectrales qui peuvent dégrader le rapport signal sur bruit. Avec l'architecture FBD, cette erreur se traduit par un bruit blanc qui augmente le niveau de la densité spectrale de bruit en sortie sans introduire de raies spectrales.
- **Bruit en $\frac{1}{\mathrm{f}}$:** L'architecture FBD est la plus robuste car seule la première voie est sensible à ce bruit qui est concentré vers les basses fréquences. Les deux autres architectures récupèrent le bruit en $\frac{1}{\mathrm{f}}$ sur toutes les voies car elles n'utilisent que des modulateurs passe-bas.

Le tableau 2.1 résume la sensibilité de chacune des architectures aux différentes erreurs. La dernière colonne récapitule le degré de complexité de l'implémentation [5, 4].

Tab. 2.1 – Comparaison de la sensibilité aux erreurs des trois architectures parallèles TIΣΔ, ΠΣΔ et FBD.

Architecture	TIΣΔ	ΠΣΔ	FBD
Sensibilité à l'erreur de gain	forte	forte	faible
Sensibilité à l'erreur de décalage en tension	forte	forte	faible
Sensibilité à la gigue d'horloge	forte (raies spectrales)	moyenne (raies spectrales)	faible (bruit blanc)
Sensibilité au bruit en $^1/_f$	forte	forte	faible
Complexité d'implémentation	faible	faible	forte

2.6 Conclusion

Nous avons présenté dans ce chapitre le concept du parallélisme pour l'élargissement de la bande de fonctionnement des CAN à base de modulateur ΣΔ. Nous avons comparé plus précisément l'architecture à entrelacement temporel TIΣΔ, l'architecture à base de modulation de *Hadamard* ΠΣΔ et l'architecture à décomposition fréquentielle FBD. Les deux premières architectures (TIΣΔ et ΠΔΣ) présentent une faible complexité de réalisation. Cependant, elles nécessitent des méthodes de calibration pour améliorer les performances dégradées par la présence de raies spectrales introduites par les imperfections des composants analogiques du modulateur. La troisième architecture (FBD) est plus robuste aux imperfections analogiques en raison de l'utilisation de modulateurs ΣΔ passe-bande.

Les trois architectures présentées ci-dessus traitent toute la bande fréquentielle entre 0 et $\frac{f_e}{2}$. Étant donné que, dans une application multistandard, le signal utile a une bande de fréquence limitée, il est inutile d'utiliser des architectures qui traitent toute la plage fréquentielle. Nous proposons dans le prochain chapitre, une architecture à base de décomposition fréquentielle permettant de traiter une bande plus étroite autour d'une certaine fréquence centrale.

Chapitre 3

Architecture FBD passe-bande et traitement numérique associé

Objectif

Ce chapitre présente le principe de fonctionnement d'un CAN passe-bande à base de modulateurs $\Sigma\Delta$ en parallèle utilisant la méthode de décomposition fréquentielle (FBD). La performance d'une telle architecture est évaluée par le calcul et par des simulations en fonction du nombre de modulateurs utilisés et du facteur de qualité des résonateurs. Ensuite, nous exposons deux méthodes de reconstruction numérique du signal : la méthode de reconstruction directe et la méthode de reconstruction avec démodulation. La méthode de reconstruction avec démodulation sera retenue en raison de sa moindre complexité matérielle par rapport à la méthode de reconstruction directe.

3.1 Introduction

La mise en parallèle de CAN à base de modulateurs $\Sigma\Delta$ présente un moyen prometteur pour l'augmentation de la bande de conversion. Nous avons vu au chapitre 2 que l'architecture FBD [37, 2] est plus robuste aux imperfections des composants analogiques. En effet, elle est moins sensible à la tension de décalage statique et au bruit en $\frac{1}{f}$ qui s'avèrent gênants avec les architectures à base de modulateurs passe-bas. Par contre, elle traite toute la bande fréquentielle et est plus complexe à réaliser que les architectures $T\Sigma\Delta$ et $\Pi\Sigma\Delta$ du fait que les modulateurs utilisés ne sont pas identiques.

Dans un récepteur radiocommunication, le signal utile possède une bande de fréquence limitée autour de la fréquence intermédiaire IF. Dans ce cas, l'utilisation de N modulateurs, avec les architectures citées ci-dessus, pour convertir toute la bande fréquentielle n'est pas optimale. L'architecture FBD passe-bande permet d'optimiser l'utilisation de N modulateurs en les plaçant dans la bande du signal utile à convertir. Dans la suite, nous présentons le principe de fonctionnement de cette architecture et évaluons ses performances.

3.2 Architecture FBD passe-bande

3.2.1 Principe de fonctionnement et paramètres de calcul

L'architecture FBD passe-bande consiste à optimiser le fonctionnement pour une bande limitée en plaçant les N modulateurs ΣΔ dans la bande du signal utile comme le montre la figure 3.1.

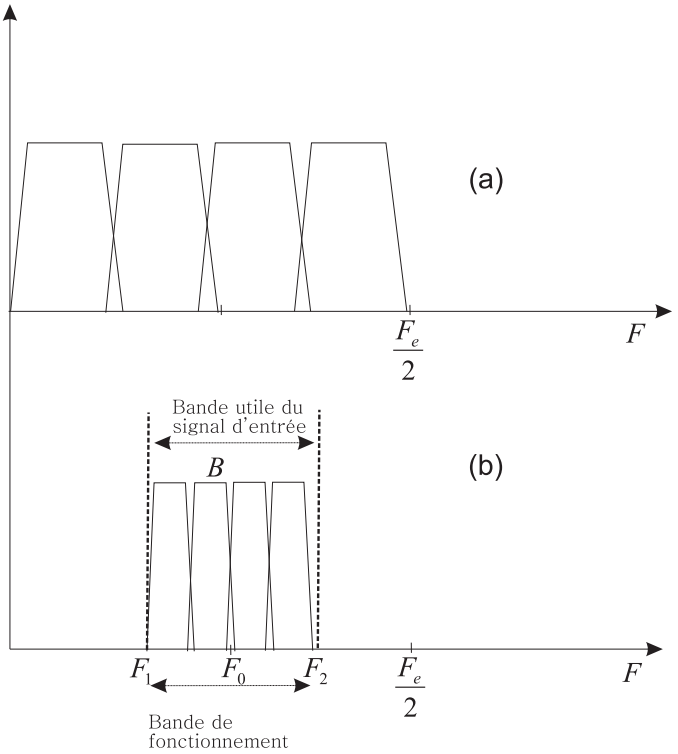

Fig. 3.1 – Répartition des modulateurs (N = 4) avec : (a) l'architecture FBD, (b) l'architecture FBD passe-bande.

Ceci permet d'éviter la conversion de toute la plage fréquentielle et d'augmenter la résolution en diminuant la largeur de bande du signal utile autour de la fréquence centrale de chaque modulateur. Cette architecture diffère de l'architecture *FBD* présentée dans [37] par le fait que :

- les N modulateurs ΣΔ sont équi-répartis dans la bande du signal utile au lieu de partager toute la bande fréquentielle (figure 3.1),
- les modulateurs ΣΔ sont à temps continu pour répondre à l'exigence de fréquences de fonctionnement élevées pour un récepteur multistandard (figure 3.2),
- elle ne contient que des modulateurs passe-bande d'architecture identiques dont les fréquences centrales diffèrent en fonction de leur emplacement dans l'architecture FBD,
- la reconstruction numérique du signal n'est pas identique. Une description complète de la reconstruction numérique du signal sera détaillée au paragraphe 3.3.

Le rapport de suréchantillonnage OSR (*OverSampling Ratio*) de chaque modulateur k est égal à N fois celui du système définit par $\mathrm{OSR_{sys}} = \frac{\mathrm{F_e}}{2\mathrm{B}}$; B étant la largeur de bande du signal d'entrée.

La figure 3.3 présente la répartition des bandes des différents modulateurs dans le cas d'une architecture FBD passe-bande à 4 modulateurs.

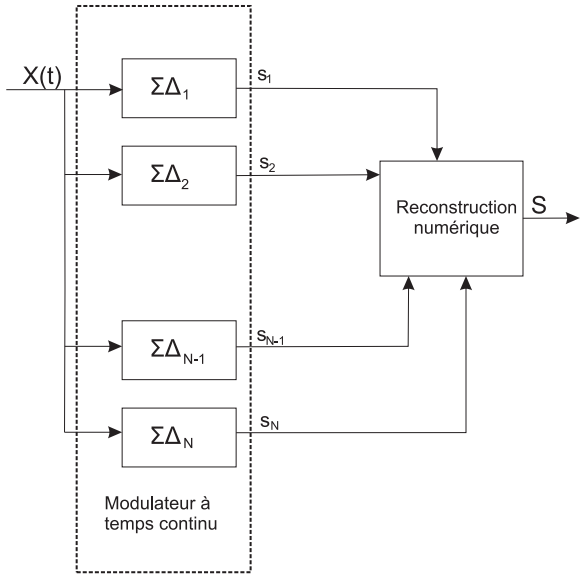

Fig. 3.2 – Architecture FBD passe-bande.

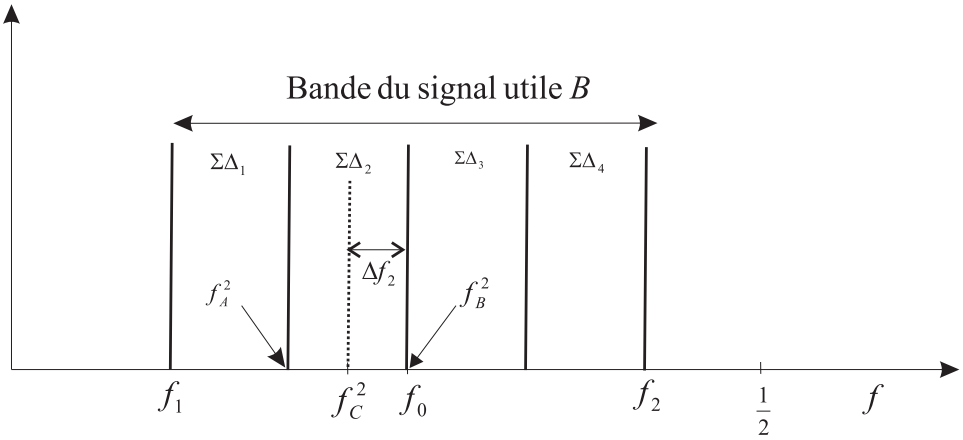

Fig. 3.3 – Répartition et paramètres des modulateurs ΣΔ.

Le signal utile en entrée a une bande limitée B dont les bornes sont les fréquences F_1 et F_2. La fréquence centrale de la bande utile est égale à $F_0 = \frac{F_1+F_2}{2}$. Dans le souci de généraliser nos résultats, nous préférons travailler avec des fréquences réduites définies par $f_0 = \frac{F_0}{F_e}$, $f_1 = \frac{F_1}{F_e}$, $f_2 = \frac{F_2}{F_e}$. Tous les modulateurs doivent traiter une largeur de bande du signal utile identique $\frac{B}{N}$. La bande de fonctionnement de chaque modulateur k est limitée par la plage fréquentielle $[f_A^k, f_B^k]$ avec :

$$\begin{cases} f_A^k = f_1 + (k-1)\frac{f_2-f_1}{N} \\ f_B^k = f_1 + k\frac{f_2-f_1}{N} \\ f_C^k = \frac{f_A^k+f_B^k}{2} \\ \Delta f_k = \frac{f_B^k-f_A^k}{2} \end{cases} \tag{3.1}$$

- f_C^k : est la fréquence centrale de la bande utile traitée par chaque modulateur,
- Δf_k : représente la moitié de la bande utile traitée par le modulateur.

3.2.2 Expression de la NTF des modulateurs à temps continu

La synthèse des modulateurs à temps continu se fait à partir des modulateurs à temps discret par la méthode de l'invariance impulsionnelle (voir annexe A, § A.4) de façon à avoir la même fonction de transfert par rapport au bruit NTF [39, 40, 41]. Nous avons choisi comme structure de départ en temps discret, la structure MSCL (Multi Stage Closed Loop) [42] (voir annexe B, § B.1) en raison de la simplicité d'expression de sa NTF. Celle-ci s'exprime par l'équation suivante :

$$NTF(z) = \prod_{j=1}^{n} NTF_j(z) \text{avec} \begin{cases} NTF_j = \frac{1}{1+c_j G_j(z)} \\ G_j(z) = \frac{\frac{p_j}{2} z^{-1} - z^{-2}}{1 - p_j z^{-1} + z^{-2}}, \quad p_j = 2\cos(2\pi \frac{F_{cr_j}}{F_e}) \end{cases} \tag{3.2}$$

F_{cr_j} étant la fréquence centrale du résonateur j.
L'ordre du modulateur doit être choisi de façon à rejeter le maximum de bruit de quantification tout en garantissant la stabilité du modulateur. L'utilisation d'un CAN multibit dans la boucle du modulateur permet d'améliorer la résolution tout en maintenant la stabilité. Cependant, pour être profitable, la solution multibit doit s'associer à un traitement numérique adapté pour corriger les erreurs de non linéarité dans le CNA. C'est pourquoi un nombre de bits très élevé ne peut être envisagé sous peine de complexifier fortement l'architecture. Dans le cadre de cette thèse, nous avons choisi de travailler avec des modulateurs à temps continu d'ordre 6 (3 résonateurs passe-bande) et un CAN de 3 bits. Ce choix est justifié par la possibilité de réalisation d'un modulateur d'ordre 6 présentant des critères de stabilité corrects [25, 24] et la faisabilité d'un algorithme de brassage de source pour compenser les non-idéalités du CNA dans la boucle de rétroaction [43].

Calcul de la NTF

Le calcul de la NTF est indispensable pour évaluer la puissance résiduelle de bruit de quantification en sortie de chaque modulateur et évaluer par la suite la performance globale attendue par l'architecture FBD passe-bande. Nous rappelons que la synthèse des modulateurs à temps continu se fait à partir de leurs équivalents à temps discret de façon à avoir la même fonction de transfert par rapport au bruit NTF. Dans la suite, nous présentons le calcul de la NTF en fonction du type de résonateur utilisé : idéal (facteur de qualité $Q = \infty$) ou réel (facteur de qualité fini).

NTF avec Q infini

L'expression de la NTF pour un modulateur d'ordre 6 se calcule à partir de l'expression suivante :

$$NTF(z) = \prod_{j=1}^{3} NTF_j(z) = \prod_{j=1}^{3} \frac{1}{1 + c_j G_j(z)} \quad avec \quad G_j(z) = \frac{\frac{p_j}{2} z^{-1} - z^{-2}}{1 - p_j z^{-1} + z^{-2}} \tag{3.3}$$

Étant donné que la bande du signal utile pour chaque modulateur est étroite autour de la fréquence centrale, le développement au premier ordre des différentes $NTF_j(z)$ permet d'écrire le module au carré de la NTF(z) par la formule suivante (voir annexe B, § B.1.1) :

$$\left|NTF^k(z)\right|^2 = \left|\prod_{j=1}^{3} NTF_j^k(z)\right|^2 = (4\pi)^6 \prod_{j=1}^{3} \left(\frac{f - f_{cr_j}}{c_j}\right)^2 \tag{3.4}$$

NTF **avec Q fini**

La fonction de transfert du résonateur à temps discret $G_j(z) = \frac{\frac{p_j}{2}z^{-1}-z^{-2}}{1-p_j z^{-1}+z^{-2}}$ a pour équivalent en temps continu la fonction $G_j(p) = \frac{\alpha p+a}{p^2+w_j^2}$ qui est la fonction de transfert d'un résonateur à temps continu avec un facteur de qualité infini. Ce résonateur n'est pas réalisable en pratique. Les résonateurs réalisables ont un facteur de qualité Q fini et une fonction de transfert du type $G_j(p) = \frac{p+a}{p^2+\frac{w_j}{Q}p+w_j^2}$. Pour tenir compte du facteur de qualité fini des résonateurs, nos calculs ont montré (voir annexe B, § B.1.2) que la fonction de transfert $G_j(z)$ doit être écrite sous la forme suivante :

$$G_j(z) = \frac{\frac{p_j}{2}z^{-1} - z^{-2}}{1 - p_j\left(1 - \frac{\pi f_{cr_j}}{Q_j}\right)z^{-1} + \left(1 - \frac{2\pi f_{cr_j}}{Q_j}\right)z^{-2}} \tag{3.5}$$

Le facteur de qualité Q_j introduit dans cette formule est le même que celui du résonateur à temps continu. On peut remarquer que quand Q_j devient très grand (voire infini), on retrouve la fonction $G_j(z)$ de départ dans le cas d'un Q infini.

En tenant compte de la nouvelle expression de $G_j(z)$, le module de la NTF est calculé de la même façon par l'équation (3.3) en appliquant une approximation au premier ordre autour des fréquences centrales des résonateurs (voir annexe B, § B.1.3). Le module au carré de la NTF avec des résonateurs non idéaux (Q fini) est donné par l'équation suivante :

$$\left|NTF^k(z)\right|^2 = \left|\prod_{j=1}^{3} NTF_j^k(z)\right|^2 = (4\pi)^6 \prod_{j=1}^{3}\left(\frac{(f - f_{cr_j})^2 + \left(\frac{f_{cr_j}}{2Q_j}\right)^2}{(c_j)^2}\right) \tag{3.6}$$

3.2.3 Optimisation des fréquences centrales

Le module de la NTF dépend des fréquences centrales des résonateurs f_{cr_j}, des coefficients c_j et des facteurs de qualité Q_j. Il faut déterminer par le calcul ces différents paramètres de façon à récupérer une puissance de bruit résiduelle minimale dans la bande du signal utile. Les coefficients c_j conditionnent la stabilité du modulateur et par conséquent leur choix est restreint. Les facteurs de qualité Q_j sont à leur tour imposés par la technologie de réalisation. Le seul degré de liberté qui reste est le choix des fréquences centrales f_{cr_j}. La puissance de bruit dans la bande du signal utile $[f_A^k, f_B^k]$ du modulateur k est donnée par l'équation :

$$P_{NTF^k} = \int_{f_A^k}^{f_B^k} \left|NTF^k(z)\right|^2 \Gamma_k(f)df = \int_{f_c^k-\Delta f_k}^{f_c^k+\Delta f_k} \left|\prod_{j=1}^{3} NTF_j^k(z)\right|^2 \Gamma_k(f)df \tag{3.7}$$

$\Gamma_k(f)$ est la densité spectrale du bruit de quantification du CAN dans la boucle du modulateur. Elle est donnée, conformément aux conditions de *Bennett*, par l'équation suivante :

$$\Gamma(f) = \frac{1}{3 \times 4^{N_b} f_e} \tag{3.8}$$

N_b étant le nombre de bits du CAN dans la boucle.
Le choix optimal des fréquences centrales des résonateurs permettant de minimiser la puissance

de bruit exprimée par l'équation (3.7) est obtenu par la résolution du système d'équations suivant :

$$\left\{ \frac{\partial P_{NTF^k}}{\partial f^k_{cr_1}} = 0, \frac{\partial P_{NTF^k}}{\partial f^k_{cr_2}} = 0, \frac{\partial P_{NTF^k}}{\partial f^k_{cr_3}} = 0 \right\} \tag{3.9}$$

Dans le cas d'un modulateur d'ordre 1, la fréquence centrale du résonateur f_{cr_1} doit être égale à la fréquence centrale de la bande utile à traiter [44]. Pour un modulateur passe-bande d'ordre 6, on suppose que la fréquence de l'un des résonateurs est égale à la fréquence centrale de la bande utile f_c^k et que les deux autres fréquences sont symétriques par rapport à f_c^k (équation (3.10)).

$$\left\{ \begin{array}{l} f^k_{cr_1} = f^k_c - \lambda \Delta f_k \\ f^k_{cr_2} = f^k_c \\ f^k_{cr_3} = f^k_c + \lambda \Delta f_k \end{array} \right. \tag{3.10}$$

λ étant un coefficient variant entre 0 et 1.
La résolution du système d'équations 3.9 conduit à distinguer deux cas :

1. **Résonateurs idéaux (Q infini)** : en remplaçant, dans l'équation (3.7), l'expression de la NTF pour $Q = \infty$ avec un coefficient c identique pour tous les résonateurs, la puissance de bruit dans la bande utile s'exprime par :

$$P_{NTF^k} = \left(\frac{4\pi}{c}\right)^6 \Gamma(f) \int_{f_c^k - \Delta f_k}^{f_c^k + \Delta f_k} \left(f - f^k_{cr_1}\right)^2 \left(f - f^k_{cr_2}\right)^2 \left(f - f^k_{cr_3}\right)^2 df$$

   L'intégration de l'expression précédente[1] permet d'exprimer la puissance de bruit par l'équation :

$$P_{NTF^k} = \left(\frac{4\pi}{c}\right)^6 \left(\frac{2(\Delta f_k)^7}{105}\right) \left(35\lambda^4 - 42\lambda^2 + 15\right) \Gamma_k(f) \tag{3.11}$$

   L'obtention de la valeur de λ minimisant la puissance de bruit se fait par l'annulation de la dérivée partielle $\frac{\partial P_{NTF^k}}{\partial \lambda}$. Ceci donne deux solutions pour λ :

   (a) $\lambda = 0$: Cette solution correspond au cas où les trois résonateurs ont des fréquences centrales égales à f_c^k. La puissance de bruit dans ce cas est égale à :

$$P_{NTF^k} = \left(\frac{4\pi}{c}\right)^6 \left[\frac{2(\Delta f_k)^7}{7}\right] \Gamma(f) \tag{3.12}$$

   (b) $\lambda = \sqrt{\frac{3}{5}}$: Dans ce cas, deux des fréquences sont symétriques par rapport à la fréquence centrale f_c^k et la puissance de bruit est donnée par l'équation suivante :

$$P_{NTF^k} = \left(\frac{4\pi}{c}\right)^6 \left[\frac{8}{175} (\Delta f_k)^7\right] \Gamma(f) \tag{3.13}$$

   Cette solution pour λ assure une puissance de bruit plus faible. Cette valeur de $\lambda = \sqrt{\frac{3}{5}}$ correspond à la valeur qui a été trouvée par *Schreier* dans le cas d'un modulateur d'ordre 6 [23].

[1] Le calcul est réalisé avec le logiciel de calcul symbolique MAPLE.

2. **Résonateurs non idéaux (Q fini)** : l'expression de la puissance de bruit avec des résonateurs réels est donnée par l'équation suivante :

$$P_{NTF^k} = \Gamma(f) \int_{f_c^k - \Delta f_k}^{f_c^k + \Delta f_k} \prod_{j=1}^{3} \left(\frac{4\pi}{c_j}\right)^2 \left((f - f_{cr_j})^2 + \left(\frac{f_{cr_j}}{2Q_j}\right)^2\right) df \tag{3.14}$$

Pour obtenir les fréquences optimales, nous appliquons la même démarche de calcul qu'avec des résonateurs idéaux : les trois fréquences sont symétriques par rapport à $\mathrm{f_c^k}$ (équation (3.10)) et la solution s'obtient à partir de $\frac{\partial \mathrm{P_{NTF^k}}}{\partial \lambda} = 0$. L'expression de $\frac{\partial \mathrm{P_{NTF^k}}}{\partial \lambda}$ obtenue avec MAPLE se révèle très compliquée. Afin d'alléger le calcul, nous supposons que :

– les résonateurs ont le même facteur de qualité Q et le même coefficient c. Cette condition est souvent respectée en raison de l'implémentation des résonateurs sur la même puce avec la même technologie.
– les deuxièmes termes $\frac{\mathrm{f_{cr_j}}}{\mathrm{2Q_j}}$ des différentes $\mathrm{NTF_j}$ peuvent être approchés par $\frac{\mathrm{f_c^k}}{\mathrm{2Q}}$. Cette approximation est valide tant que la différence entre les fréquences centrales $\mathrm{f_{cr_j}}$ ne dépasse pas $\mathrm{\Delta f_k}$ et que $\mathrm{f_{cr_j}}$ est très petit devant 2Q.

En partant de ces deux conditions, la résolution de l'équation $\frac{\partial \mathrm{P_{NTF^k}}}{\partial \lambda} = 0$ donne deux solutions :

(a) $\lambda = 0$: C'est le cas où les trois fréquences centrales sont identiques ($\mathrm{f_{cr_j}} = \mathrm{f_c^k}$). Dans ce cas, la puissance de bruit dans la bande du signal utile de largeur $\frac{\mathrm{B}}{\mathrm{N}}$ est donnée par l'équation suivante :

$$P_{NTF^k} = 2\left(\frac{4\pi}{c}\right)^6 \Delta f_k \left[\alpha_k^6 + \Delta f_k^2 \alpha_k^4 + \frac{3}{5}\Delta f_k^4 \alpha_k^2 + \frac{1}{7}\Delta f_k^6\right] \Gamma_k(f) \tag{3.15}$$

(b) $\lambda = \sqrt{\frac{3}{5}}\sqrt{\frac{1-\frac{5\alpha^4}{\mathrm{\Delta f_k^4}}}{1+\frac{3\alpha^2}{\mathrm{\Delta f_k^2}}}}$ avec $\alpha = \frac{\mathrm{f_c}}{\mathrm{2Q}}$: Cette expression de λ, en tenant compte du facteur de qualité, correspond plus à la réalité que celle calculée par *Schreier* [23, 45]. Lorsque le facteur de qualité tend vers l'infini, cette expression de λ tend vers la valeur classique de $\sqrt{\frac{3}{5}}$.

La puissance de bruit calculée avec une répartition des fréquences centrales en tenant compte du facteur de qualité des résonateurs est donnée par l'équation (3.16). Elle se révèle bien plus faible que celle obtenue avec la première solution ($\lambda = 0$).

$$P_{NTF^k} = \left(\frac{4\pi}{c}\right)^6 \left[\frac{8\Delta f_k(\alpha_k^6 + \Delta f_k^2 \alpha_k^4 + \frac{9}{35}\Delta f_k^4 \alpha_k^2 + \frac{1}{175}\Delta f_k^6)}{1 + \frac{3\alpha_k^2}{\Delta f_k^2}}\right] \Gamma_k(f) \tag{3.16}$$

Ainsi, c'est cette expression de λ qui sera préférentiellement retenue pour le choix des fréquences centrales, à condition cependant de satisfaire un critère sur l'ordre de grandeur du facteur de qualité. Ce critère peut s'expliciter en considérant l'expression de λ dans le cas 2b. En effet, la deuxième solution n'est valide que si le terme sous la racine est positif :

$$\left(1 - \frac{5\alpha^4}{\Delta f_k^4}\right) > 0$$

Si cette condition n'est pas vérifiée, il n'y a pas de solution réelle, la solution se trouve dans le domaine complexe. Dans ce cas, la seule solution est celle qui correspond à $\lambda = 0$. La résolution de $\left(1 - \frac{5\alpha^4}{\Delta f_k^4}\right) = 0$ permet de déterminer le facteur de qualité limite Q_{lim} entre les deux solutions. Il est donné par l'équation (3.17).

$$Q_{lim} = \frac{\sqrt[4]{5}}{2} \frac{f_c^k}{\Delta f_k} \tag{3.17}$$

Avec des résonateurs de facteur de qualité $Q < Q_{lim}$ et une bande de fonctionnement étroite (Δf_k petit), le facteur dominant dans l'expression de la puissance (équation (3.15)) est $\left[2\left(\frac{4\pi}{c}\right)^6 \alpha_k^6 \Delta f_k\right]$. Dans ce cas, la puissance de bruit est approchée par $P_{NTF^k} \approx \Gamma_k(f) \times \Delta f_k \times$ cte. Ce qui signifierait que le bruit du CAN de boucle n'a pas été mis en forme par le modulateur $\Sigma\Delta$ (i.e $\left|NTF^k(z)\right|^2 \approx$ cte). Avec cette configuration, il n'y a pas d'intérêt à utiliser un modulateur $\Sigma\Delta$. Avec un facteur de qualité infini ($\alpha = 0$), on retrouve l'expression de la puissance donnée par l'équation (3.13).

Après avoir déterminé l'expression de la puissance de bruit dans la bande de fonctionnement du modulateur à temps continu en tenant compte du facteur de qualité des résonateurs, nous allons estimer, dans le prochain paragraphe, la performance de l'architecture FBD passe-bande en terme de puissance de bruit résiduelle en sortie.

3.2.4 Performances de l'architecture FBD passe-bande

La performance du modulateur $\Sigma\Delta$ se mesure par la puissance de bruit résiduelle dans la bande du signal utile à convertir autour de sa fréquence centrale. Un moyen d'exprimer cette performance est la résolution, appelée aussi nombre de bit effectif ENOB (Effective Number Of Bits). Cette grandeur est obtenue en calculant le nombre de bits équivalents à partir de la puissance du bruit de quantification obtenue avec un CAN classique. Elle a pour expression (3.18) :

$$P_{bruit} = \frac{1}{3 \times 4^{N_b}} \tag{3.18}$$

À partir de cette équation, on peut écrire la P_{NTF^k} sous la forme suivante :

$$P_{NTF^k} = \frac{1}{3 \times 4^{ENOB}}$$

Ceci permet d'exprimer le nombre de bits effectif :

$$ENOB = -\frac{\ln\left(3P_{NTF^k}\right)}{\ln\left(4\right)} \tag{3.19}$$

L'ENOB peut aussi être exprimé en fonction du nombre de bits N_b du CAN dans la boucle du modulateur. En effet, à partir de l'expression de P_{NTF^k}, on peut faire l'équivalence suivante :

$$P_{NTF^k} = \int_{f_A^k}^{f_B^k} \left|NTF^k(z)\right|^2 \Gamma_k(f) df = \frac{1}{3 \times 4^{N_b}} \int_{f_A^k}^{f_B^k} \left|NTF^k(z)\right|^2 df \equiv \frac{1}{3 \times 4^{ENOB}}$$

En se basant sur cette expression, l'ENOB s'exprime également par :

$$ENOB = N_b - \frac{\ln\left(\frac{P_{NTF^k}}{\Gamma_k(f)}\right)}{\ln(4)} \tag{3.20}$$

La puissance de bruit résiduelle en sortie du modulateur dépend du facteur de qualité Q des résonateurs. L'évolution de cette puissance P_{NTF^k} en fonction du facteur de qualité Q, pour une fréquence centrale $f_c^k = 1/4$, des coefficients $c_j = \frac{1}{2}$ (ce choix pour c_j sera détaillé au paragraphe 3.2.5) et une demie-bande de fonctionnement $\Delta f_k = 1/160$ est représentée sur la figure 3.4.

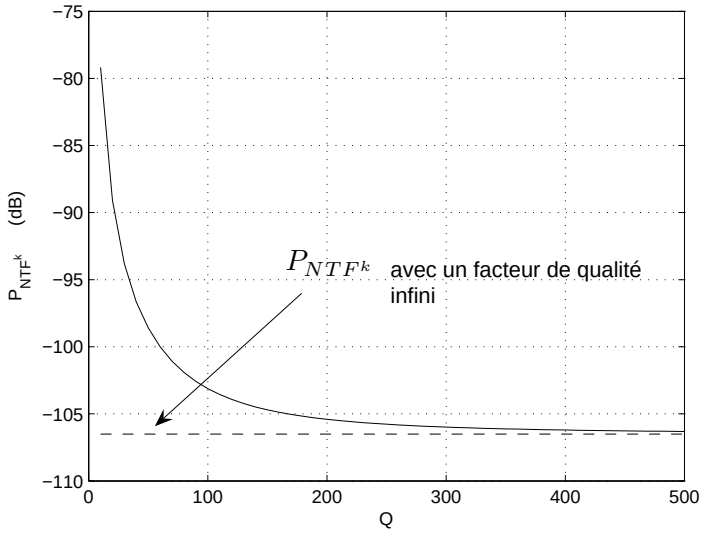

Fig. 3.4 – Puissance de bruit en fonction de Q avec $\Delta f_k = 1/160$, $f_c^k = 1/4$ et $c_j = \frac{1}{2}$.

Celle-ci montre que plus le facteur de qualité est grand plus le P_{NTF^k} est faible et meilleur est l'ENOB. Un facteur de qualité de 50 entraîne approximativement une chute de l'ENOB de 1 bit (8 dB) par rapport à la valeur théorique (obtenue avec $Q = \infty$). On peut noter également qu'à partir du facteur de qualité $Q = 100$, l'augmentation de ce facteur n'apporte pas un gain significatif sur l'ENOB.

Un autre élément qui entre en jeu dans le calcul de la puissance de bruit est le terme Δf_k. La figure 3.5 présente la puissance en fonction de Δf_k pour différents facteurs de qualité.

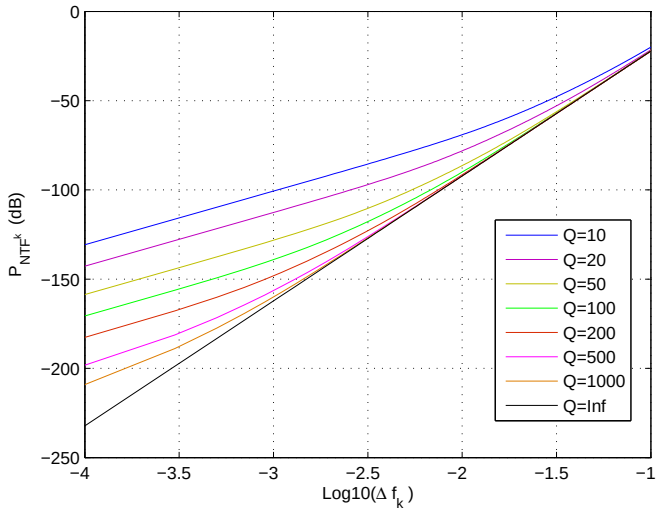

Fig. 3.5 – Puissance de bruit en fonction de Δf_k pour différents Q avec $c_j = \frac{1}{2}$.

On peut noter que :
- plus faible est la largeur de demie-bande de fonctionnement Δf_k, meilleure est la résolution. Ce résultat s'explique par le fait que l'augmentation de la bande de fonctionnement entraîne une récupération de bruit plus grande,
- pour des bandes de fonctionnement Δf_k larges, l'influence du facteur de qualité devient minime (à partir de $\Delta f_k = 10^{-2}$). En effet, aux extrémités de cette bande le bruit de quantification est amplifié par la NTF. Dans ce cas une augmentation du facteur de qualité ne peut pas entraîner une diminution significative de la puissance de bruit.

Une fois estimée la puissance de bruit en sortie de chaque modulateur dans sa bande de fonctionnement, il reste à déterminer la puissance de bruit totale délivrée par l'architecture FBD passe-bande dans la bande du signal utile B à convertir. En partant de l'hypothèse que les sources de bruit de quantification des différents modulateurs sont blanches ($\Gamma_k(f)$ est constante) et décorrélées entre elles, la puissance de bruit dans la bande du signal utile B est la somme des puissances délivrées par les N modulateurs ayant chacun une bande utile de $\frac{B}{N}$:

$$P_{bruit\_totale} = \sum_{k=1}^{N} P_{NTF^k} \tag{3.21}$$

Puisque tous les modulateurs ont la même largeur de bande de fonctionnement, ils délivrent la même puissance de bruit. De ce fait, l'ENOB de l'architecture FBD passe-bande est exprimé directement à partir de l'ENOB obtenu avec un seul modulateur comme le montre l'expression suivante :

$$ENOB_{FBD} = \frac{-\ln\left(3 \times P_{bruit\_totale}\right)}{\ln(4)} = \frac{-\ln\left(3 \times N \times P_{NTF^k}\right)}{\ln(4)} = ENOB - \frac{\ln(N)}{\ln(4)} \tag{3.22}$$

Donc, le passage d'un seul modulateur à N modulateurs adjacents de même largeur de bande de fonctionnement entraîne une perte en résolution de $\frac{\ln(N)}{\ln(4)}$. L'ENOB de l'architecture FBD passe-bande en fonction du nombre de modulateurs pour différents facteurs de qualité est représenté sur la figure 3.6.

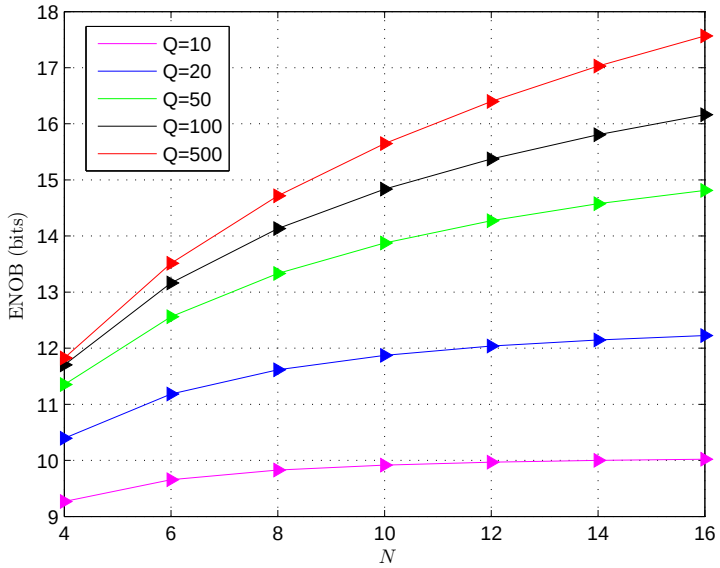

Fig. 3.6 – ENOB en fonction du nombre de modulateurs N pour différents Q avec $c_j = \frac{1}{2}$.

On note que :
- Avec une architecture FBD à huit modulateurs, la résolution théorique est de 13.3 bits pour un facteur de qualité de 50 et de 14.7 pour Q = 500,
- Pour un facteur de qualité très grand ($Q = 500$), un dédoublement du nombre de modulateurs entraîne l'amélioration de la résolution d'un facteur m où m est le nombre de résonateurs dans chaque modulateurs (voir annexe B.2),
- pour un facteur de qualité faible (Q < 20), il est inutile d'augmenter le nombre de modulateurs pour améliorer la performance.

3.2.5 Influence du coefficient c sur la performance de l'architecture

La valeur du coefficient c détermine à la fois la résolution (ENOB) en sortie du modulateur $\Sigma\Delta$ et la dynamique maximale du signal en son entrée. Le calcul de la résolution à partir de l'expression théorique du bruit (équation (3.16)) en supposant que tous les modulateurs ont le même coefficient c montre que cette résolution croît avec c (figure 3.7). À première vue, plus c est grand, meilleure est la résolution, mais il ne faut pas oublier qu'une valeur de c grande diminue la dynamique du signal en entrée et peut rendre le modulateur instable. Sa valeur est également liée au nombre de bits du CAN dans la boucle de modulateur.

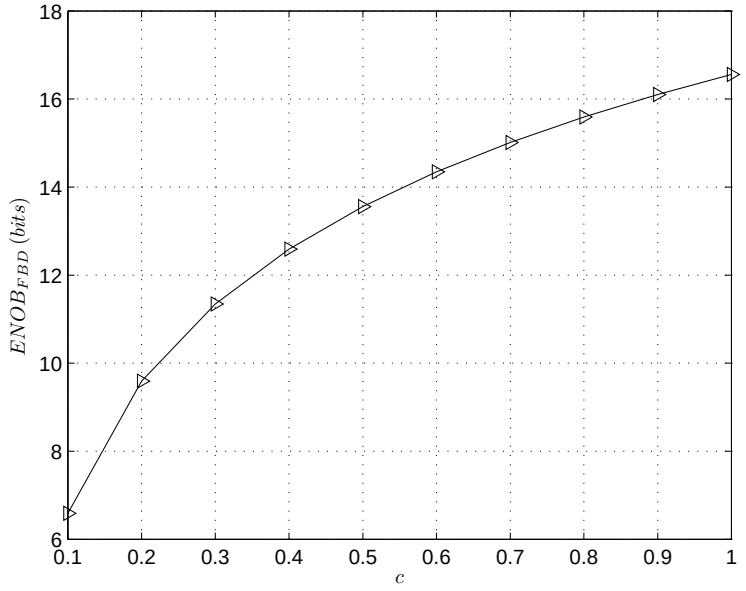

Fig. 3.7 – Résolution en fonction du coefficient c.

Dans le cas d'un modulateur monobit, un compromis entre la résolution et la stabilité, trouvé d'une façon heuristique dans [42], montre que le choix optimal du coefficient c est donné par :

$$c = \frac{2}{L} = \frac{1}{m} \tag{3.23}$$

L : est l'ordre du modulateur.
m : est le nombre de résonateurs.
Pour mettre en évidence l'influence du coefficient c dans l'architecture FBD avec des modulateurs 3 bits, nous avons calculé le SNR en sortie de l'architecture FBD en utilisant la méthode

de reconstruction numérique développée au paragraphe 3.3.2. Dans cette simulation, le signal en entrée est un signal sinusoidal dont l'amplitude varie dans l'intervalle $\left[10^{-6}, 1\right]$. La figure 3.8 montre l'évolution du SNR en fonction de la puissance du signal d'entrée P_e et ceci pour différentes valeurs de c.

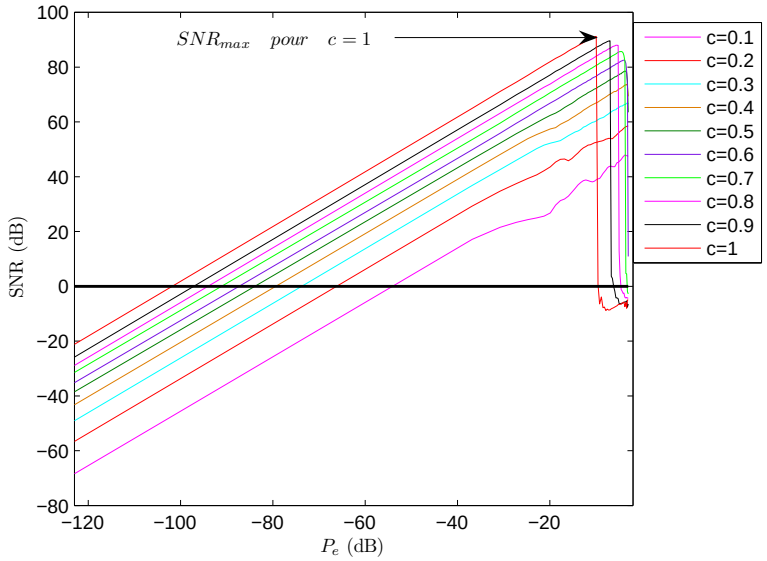

Fig. 3.8 – Dynamique du CAN en fonction de la puissance d'entrée pour différents c.

La chute du SNR, après avoir atteint sa valeur maximale $\mathrm{SNR_{max}}$, signifie que le modulateur devient instable et saturé à partir de cette valeur de la puissance d'entrée. Nous pouvons constater qu'avec des modulateurs 3 bits, la condition sur c est moins sévère et qu'une augmentation de c peut assurer une amélioration du SNR sans diminuer énormément la dynamique en entrée. Un meilleur compromis entre le SNR et la dynamique du signal d'entrée est obtenu pour $c = 0.5$. Dans ce cas, nous avons une dynamique de 78 dB avec un $\mathrm{SNR_{max}}$ de 80 dB (figure 3.9).

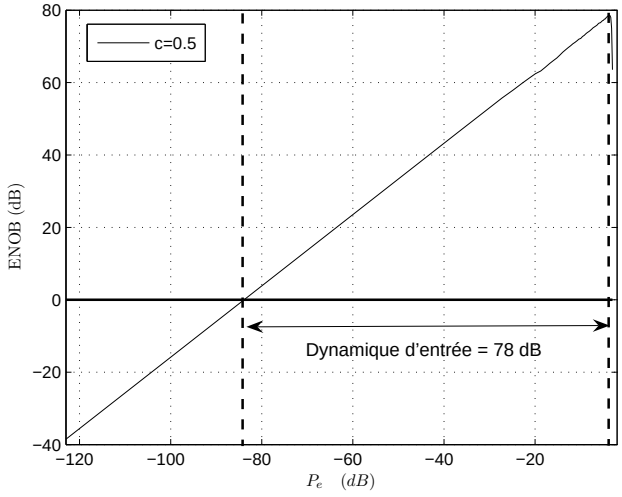

Fig. 3.9 – Dynamique du CAN $c = 0.5$.

Comme la stabilité du modulateur ΣΔ est directement liée au nombre de bits N_b du CAN dans la boucle du modulateur, le choix du coefficient c l'est aussi. Pour montrer cette dépendance, nous nous sommes intéressés à l'évolution du SNR_{max} en fonction de c pour des modulateurs à 1, 2 et 3 bits. Les résultats de simulations sont représentés par la figure 3.10.

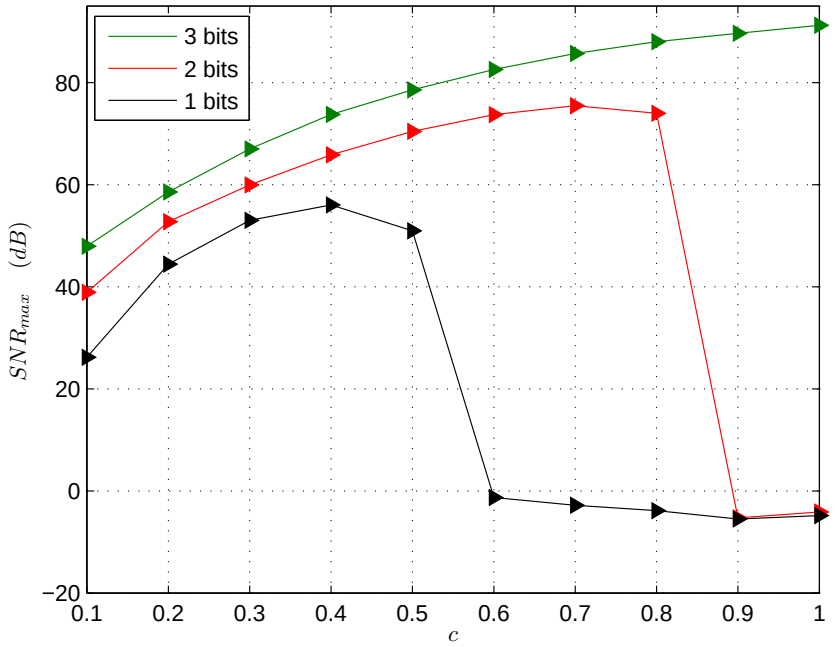

Fig. 3.10 – SNR_{max} en fonction de c.

On peut noter que pour un CAN monobit, un modulateur d'ordre 6 devient instable à partir de c = 0.4 quelque soit la puissance du signal d'entrée. En effet, on obtient un spectre du signal en sortie très riche en harmonique dégradant le SNR. On peut vérifier également que la valeur heuristique c = $1/3$ assure un bon fonctionnement dans le cas monobit. En augmentant le nombre de bits N_b du CAN, la stabilité du modulateur est assurée pour des coefficients c plus grands. Dans notre cas, où l'on utilise des modulateurs avec un CAN 3 bits, la stabilité est assurée même pour c = 1. Nous avons trouvé, par simulation, que le choix de c avec un modulateur 3 bits réalisant un compromis entre le SNR_{max} (la stabilité) et la dynamique en entrée est donné par la condition heuristique suivante :

$$\frac{1}{m} < c \leq \frac{1}{m-1} \tag{3.24}$$

m : est le nombre de résonateur.
Cette étude a justifié le choix de l'utilisation, dans l'architecture FBD passe-bande, de modulateurs ΣΔ d'ordre 6 avec un CAN 3 bits et un coefficient c = 0.5.

3.3 Reconstruction numérique du signal

Le système de reconstruction numérique (figure 3.2) est l'élément clef de l'architecture FBD passe-bande. En effet, chaque modulateur ΣΔ contribue à la mise en forme du bruit de quan-

tification du CAN autour de sa fréquence centrale. La figure 3.11 montre un exemple avec 4 modulateurs.

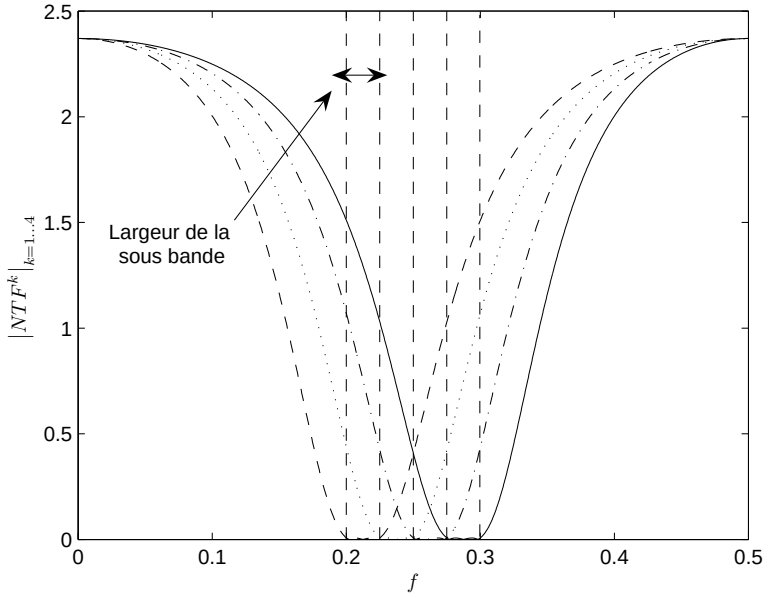

Fig. 3.11 – Les NTF avec une architecture FBD à 4 modulateurs $\Sigma\Delta$ passe-bande.

On cherche ensuite à effectuer à la sortie de chaque modulateur un filtrage passe-bande de façon à supprimer le bruit de quantification en dehors de la bande utile. Le spectre obtenu dans ce cas est idéal et constitue notre spectre de référence (figure 3.12).

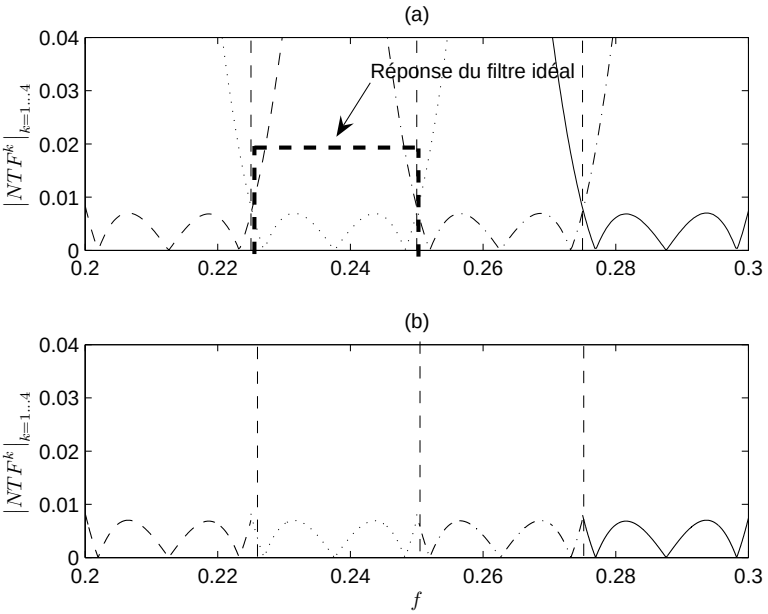

Fig. 3.12 – (a) Agrandissement de la figure 3.11 , (b) le spectre de référence en sortie.

En réalité, il n'est pas possible de réaliser de filtrage passe-bande idéal. En effet, la pente du gain d'un filtre réel n'est pas infinie aux bornes de la bande passante. Ceci conduit à une atténuation moindre du bruit de quantification en dehors de la bande passante et ainsi à une dégradation des performances globales de l'architecture FBD passe-bande. Cette atténuation dépend de la largeur de la bande passante et du nombre de coefficients du filtre. Le filtre idéal possède un nombre de coefficients infini et par conséquent il est irréalisable. En pratique, les filtres numériques ont un nombre de coefficients fini afin d'être implantables sur des calculateurs numériques.

Pour évaluer la performance du filtrage numérique, nous définissons le **facteur de dégradation** FD comme étant le rapport entre la puissance du bruit de quantification récupérée en sortie d'un traitement numérique réel et celle en sortie d'un traitement numérique idéal.

$$FD = \frac{P_{bruit\_mesuré}}{P_{bruit\_référence}} \tag{3.25}$$

Ce facteur de dégradation peut aussi s'exprimer à travers une perte en nombre de bits effectifs (ENOB). D'autres critères peuvent aussi être pris en compte pour évaluer la performance d'un traitement numérique. Dans la suite du travail, nous considérons les trois critères suivants :

- le facteur de dégradation FD,
- la minimisation des ondulations dans le spectre du signal utile,
- la complexité de réalisation et la fréquence de fonctionnement.

Deux approches sont possibles pour reconstruire le signal : la **reconstruction directe** et la **reconstruction avec démodulation**. Dans la suite, nous présentons le principe de chacune de ces méthodes avec leurs avantages et leurs inconvénients.

3.3.1 Reconstruction directe

Cette méthode est intuitive. Elle consiste à placer un filtre passe-bande derrière chaque modulateur pour couper le bruit de quantification en dehors de sa bande de fonctionnement. Ensuite, les signaux en sortie des filtres numériques sont sommés pour construire le signal de sortie $S[n]$ (figure 3.13).

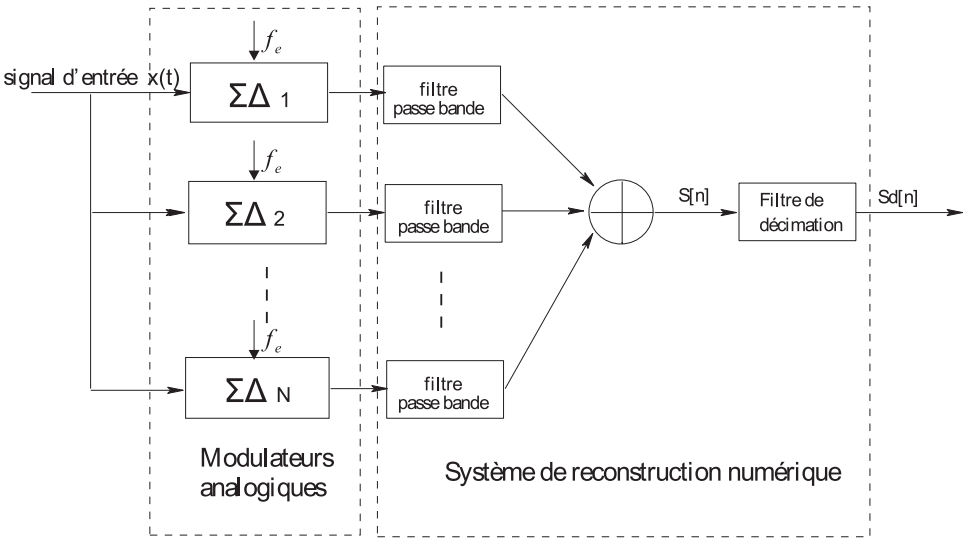

Fig. 3.13 – Reconstruction directe avec des filtres passe-bande.

Le signal en sortie est ensuite décimé pour diminuer son débit qui correspond à la fréquence de suréchantillonnage des modulateurs. Les filtres passe-bande utilisés sont des filtres à Réponse Impulsionnelle Finie FIR (Finite Impulse Response). Ce choix est conditionné par leur simplicité d'implantation, leur stabilité et leur faible sensibilité vis-à-vis de la quantification des coefficients.

Dans le cadre de cette thèse, nous avons considéré un exemple concret sur la base duquel nous effectuons un certain nombre de calculs qui nous permettent d'évaluer la performance du système de reconstruction numérique. Les résultats obtenus seront généralisables à toutes les architectures du même type. En effet, tous les paramètres de l'exemple sont normalisés par rapport à la fréquence d'échantillonnage. Une mise à jour de ces paramètres ne changerait pas le principe de fonctionnement. Pour l'évaluation, nous avons choisi une fréquence d'échantillonnage de 800 MHz, valeur acceptable au regard des capacités des technologies actuelles. Les spécifications de cet exemple sont données dans le tableau 3.1.

Tab. 3.1 – Paramètres de l'exemple considéré.

Bande du signal utile	$B = [0.2,\ 0.3] \times F_e = [160 \dots 240]\,\text{MHz}$
Fréquence d'échantillonnage	$F_e = 800\text{MHz}$
Nombre de modulateurs	$N = 8$
Rapport de sur-échantillonnage	$OSR = \frac{N \times F_e}{2B} = 40$
Facteur de décimation	$R_d = \lfloor OSR_{sys} \rfloor = 5$

Avec ce système de reconstruction, l'ENOB a été calculé en fonction du nombre de coefficients des filtres et ceci avec différents types de fenêtres de pondération. Les résultats de ce calcul sont représentés sur la figure 3.14.

Nous constatons que :

- nous avons besoin de filtres FIR possédant au moins 500 coefficients pour pouvoir atténuer suffisamment le bruit de quantification et atteindre la résolution théorique définie à partir du spectre de bruit de référence. Cette architecture de reconstruction exige une puissance de calcul énorme. Elle nécessite l'implémentation des 8 filtres passe-bande à 500 coefficients fonctionnant à la fréquence d'échantillonnage. Ce qui rend cette architecture irréalisable.
- La résolution calculée avec des filtres passe-bande ayant une fenêtre de Hamming présente des ondulations de période 160 coefficients. Pour comprendre ce phénomène, nous avons tracé la réponse impulsionnelle idéale du filtre passe-bande et les fenêtres de pondérations de différents ordres et ceci pour les fenêtres de type *Hamming* (figure 3.15) et *Hanning* (figure 3.16).

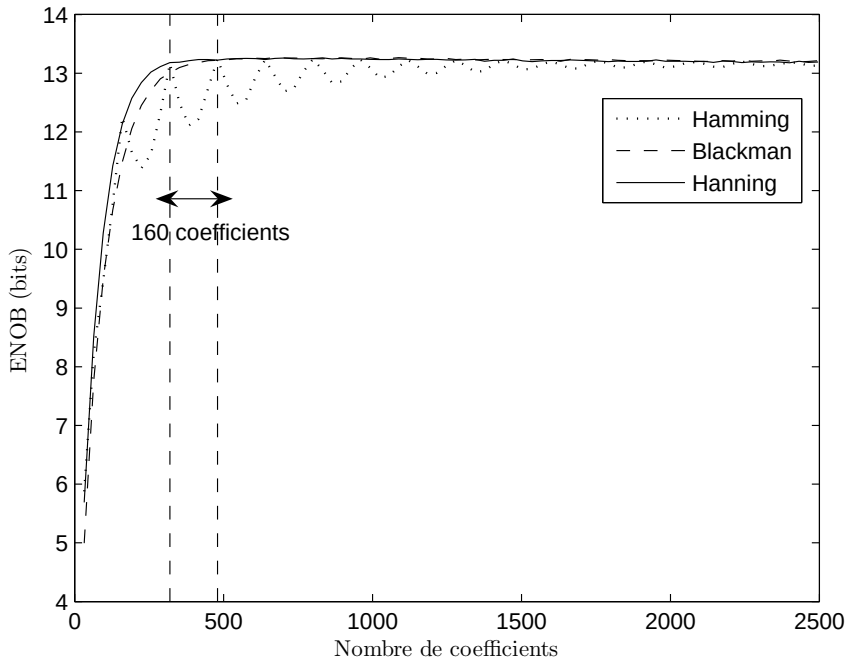

Fig. 3.14 – Résolution en fonction du nombre de coefficients des filtres passe-bande.

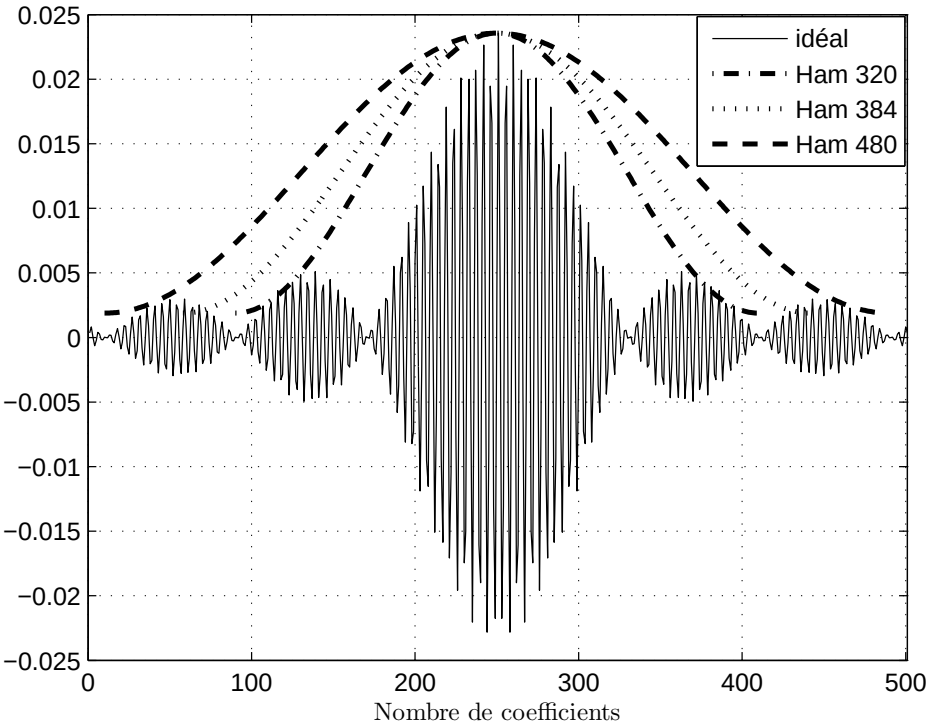

Fig. 3.15 – Réponse impulsionnelle idéale et fenêtre de *Hamming* de différents ordres.

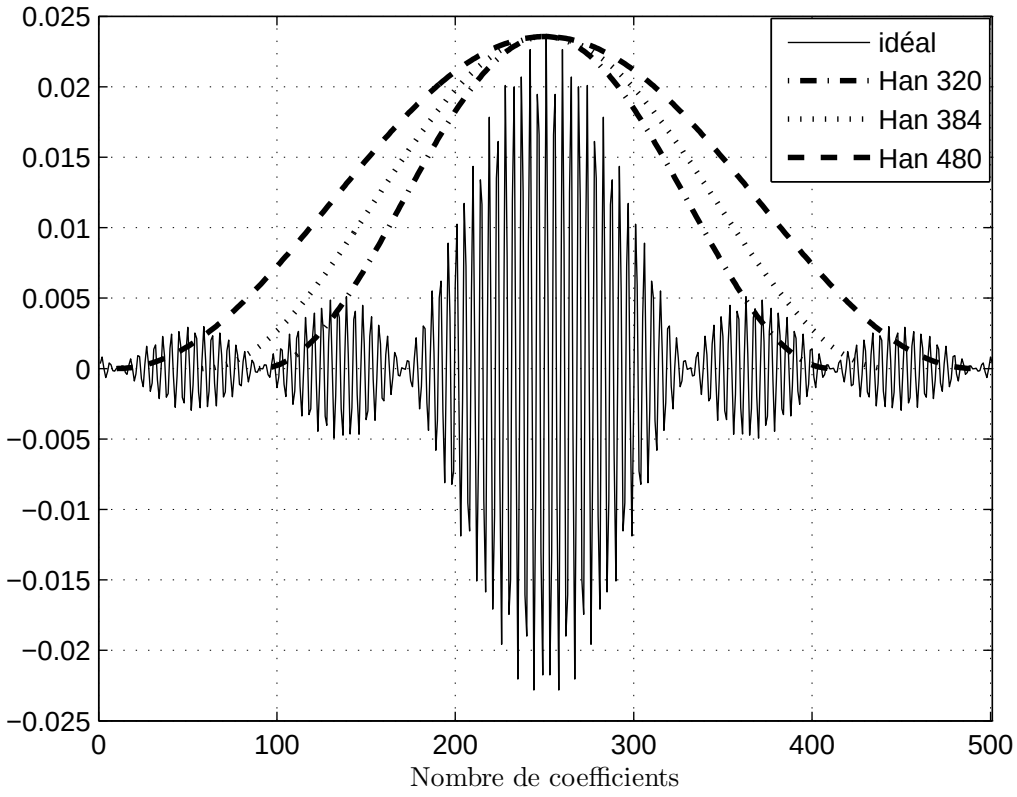

Fig. 3.16 – Réponse impulsionnelle idéale et fenêtre de *Hanning* de différents ordres.

La réponse impulsionnelle idéale est une fonction du type sinus cardinal dont la largeur du lobe principal et des lobes secondaires est proportionnelle à la bande du filtre. La largeur du lobe principal est $\frac{2}{L_B}$ et celle des lobes secondaires est $\frac{1}{L_B} = \frac{N}{B}$, L_B est la largeur de la bande passante. Soit respectivement 160 et 80 coefficients pour notre exemple. Nous constatons, pour la fenêtre de Hamming, que lorsque la réponse impulsionnelle présente des coefficients non nuls à ses bords, le filtre présente des lobes secondaires élevés dans le domaine fréquentiel. Ce qui explique la perte en résolution lorsque l'ordre du filtre augmente de 320 à 384.

La figure 3.17 présente la réponse fréquentielle définie par $\mathrm{Hf}(e^{j2\pi f}) = \sum_{i=0}^{ordre+1} a_i \left(e^{j2\pi f}\right)^{-i}$ du filtre passe-bande avec la fenêtre de *Hamming* pour les deux ordres 320 et 384. Le filtre à 320 coefficients présente des lobes secondaires plus faibles que celui avec 384 coefficients. La période de 160 coefficients est due au fait qu'il faut dépasser le lobe secondaire de chaque côté du coefficient central pour retrouver des coefficients nuls aux bords.

En conclusion, une augmentation de la longueur du filtre, avec la fenêtre de *Hamming*, n'implique pas forcément une meilleure atténuation du bruit en dehors de la bande de fonctionnement. En revanche, la fenêtre de *Hanning* atténue les lobes secondaires de la réponse impulsionnelle quel que soit l'ordre. On note que ce phénomène n'apparaît pas quand la bande du filtre est faible ou le filtre est du type passe-bas car dans ce cas les lobes secondaires sont très faibles.

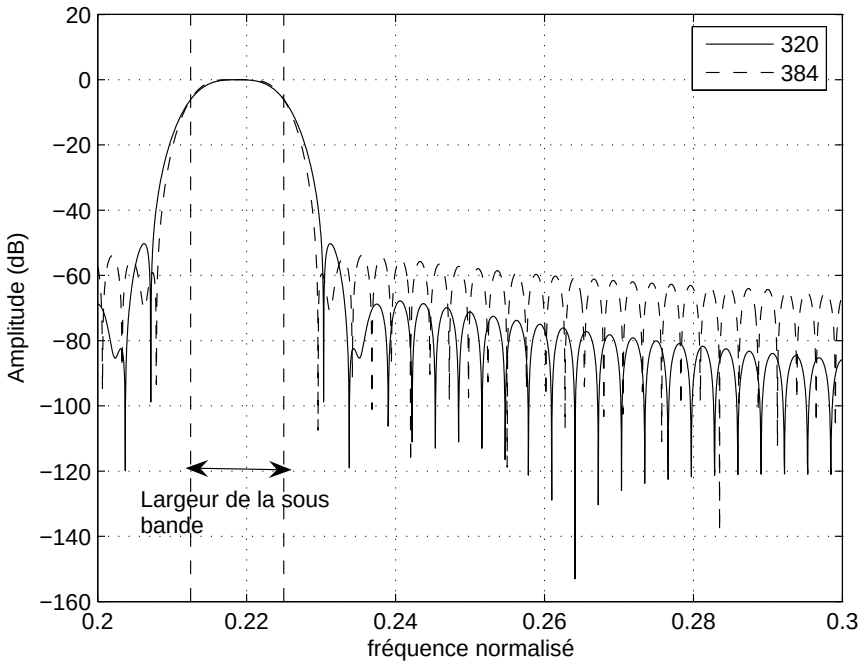

Fig. 3.17 – Réponse fréquentielle des deux filtres FIR passe-bande.

3.3.2 Reconstruction avec démodulation

L'idée de cette méthode de reconstruction est de ramener le signal de sortie de chaque modulateur en bande de base, de faire le traitement nécessaire pour éliminer le bruit de quantification et de le ramener de nouveau à sa bande initiale. Cette méthode de reconstruction avec démodulation est illustrée par la figure 3.18.

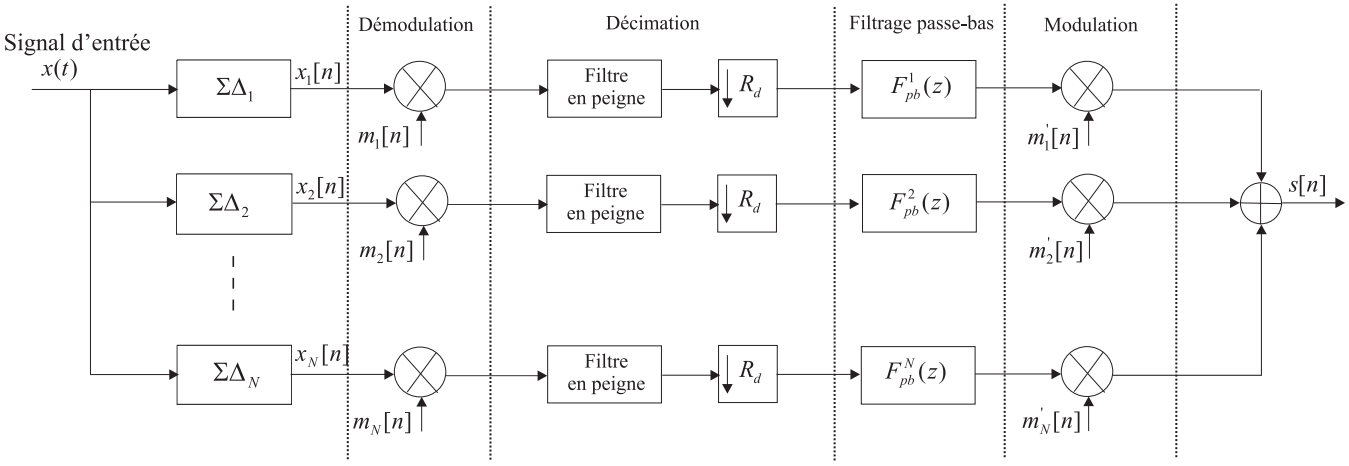

Fig. 3.18 – Reconstruction en bande de base.

Elle comporte les étapes suivantes :

1. **Démodulation** : elle consiste à ramener les signaux en sortie des différents modulateurs $x_k[n]$ en bande de base grâce à une multiplication par la séquence complexe $m_k[n] = e^{-j2\pi f_c^k n}$. La démodulation complexe permet de séparer, en bande de base, la partie positive et la partie négative du spectre du signal utile parce qu'elles ne portent pas la même information selon les types de modulation numérique utilisés dans les récepteurs de radiocommunication. Le signal résultant est l'enveloppe complexe du signal utile situé dans la bande de fonctionnement de chaque modulateur (voir annexe B, § B.3). On note que si la fréquence centrale f_c^k est un nombre rationnel $f_c^k = \frac{p}{q}$ où p et q sont des entiers, la séquence m_k est une séquence périodique de période q. Cette propriété est très importante pour l'implémentation numérique de cette architecture. La séquence de démodulation est de q valeurs et peut être calculée à l'avance et stockée dans une mémoire ROM.
2. **Décimation** : le signal démodulé, ayant un débit correspondant à la fréquence de suréchantillonnage du modulateur, est décimé d'un facteur R_d après avoir été suréchantillonné par le modulateur. Il est important d'effectuer un choix judicieux de ce facteur R_d et de déterminer sa valeur maximale. Pour cela, nous considérons l'allure du spectre du signal décimé. Ce spectre s'obtient en répétant le motif X(f) tous les $\frac{1}{R_d}$ (dans l'intervalle fondamental $[0, 1]$), en additionnant les recouvrements de spectre éventuels, en multipliant la représentation par $\frac{1}{R_d}$ puis en normalisant l'axe fréquentiel de sorte que $f = \frac{1}{R_d}$ corresponde à la fréquence $f' = 1$. Ceci est illustré par la figure 3.19.

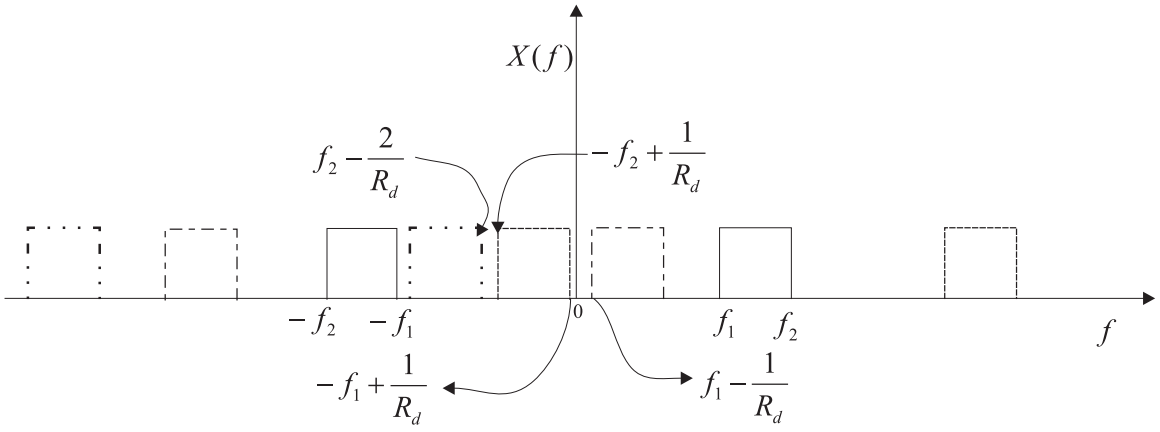

Fig. 3.19 – Spectre du signal décimé.

Un bon facteur de décimation est celui qui permet de diminuer le débit sans créer de recouvrement de spectre permettant ainsi de récupérer le signal utile. En se basant sur la figure 3.19, le rapport de décimation permettant d'éviter le recouvrement spectral doit vérifier la relation 3.26.

$$\begin{cases} f_1 - \frac{1}{R_d} \geq -f_1 + \frac{1}{R_d} \\ -f_2 + \frac{1}{R_d} \geq f_2 - \frac{2}{R_d} \end{cases} \Rightarrow 2(f_2 - f_1) \leq \frac{1}{R_d} \Rightarrow R_d \leq \frac{f_e}{2(f_2 - f_1)} \tag{3.26}$$

À partir de l'équation (3.26) et du fait que R_d est un entier, on peut en déduire que le rapport de décimation maximal est égal à la partie entière du rapport de suréchantillonnage (équation (3.27)).

$$R_{d_{\max}} = \left\lfloor \frac{f_e}{2(f_2 - f_1)} \right\rfloor = \left\lfloor \frac{F_e}{2B} \right\rfloor = \lfloor OSR_{sys} \rfloor \tag{3.27}$$

3. **Filtrage passe-bas** : après la décimation, le signal est filtré par un filtre FIR passe-bas $F_{pb}^{k}(z)$ pour éliminer le bruit de quantification en dehors de la bande de fonctionnement.
4. **Modulation** : une fois filtré, le signal est modulé en le multipliant par la séquence $m_{k}^{'}[n] = e^{j2\pi f_{c}^{k} R_{d} n}$ et est ajouté aux autres sorties pour reconstruire le signal passe-bande numérisé.

La décimation avant le filtrage passe-bas est le point clé de cette architecture. Il permet de diminuer les contraintes sur les filtres passe-bas en augmentant la zone de transition entre la bande passante et la bande de réjection de ces filtres d'un facteur R_d. Ce qui permet de réaliser des filtres FIR avec :

- un nombre de coefficients beaucoup plus faible que celui demandé par la première solution pour atteindre la résolution théorique,
- une fréquence de fonctionnement R_d fois plus faible.

Cette technique (décimation avant filtrage) est inapplicable avec la première solution car la décimation passe-bande exige des filtres passe-bande anti-repliement. Cependant cette architecture exige quelques étapes de corrections numériques qui seront détaillées dans la suite.

3.4 Corrections appliquées à la reconstruction numérique avec démodulation

La méthode de reconstruction avec démodulation présente une bonne performance en terme de nombre de bits effectifs par rapport à la puissance de calcul demandée pour atteindre cette résolution. Cependant, cette architecture dans son état actuel n'assure pas la minimisation des ondulations dans le spectre du signal utile. Ces ondulations sont dûes au fait que :

- le gain de certains composants de cette architecture (modulateur $\Sigma\Delta$, filtre passe-bas, filtre en peigne) n'est pas constant dans la bande de fonctionnement. Ceci a pour effet de favoriser l'amplification de certains signaux et l'atténuation d'autres signaux.
- les phases entre les bandes de fonctionnement adjacentes ne sont pas accordées. En effet, si un signal se trouve à la limite entre la bande de fonctionnement du modulateur k et celle du modulateur $k+1$, ce signal va avoir deux phases différentes introduites par chacune des bandes de fonctionnement. Ceci se répercute sur le spectre du signal utile en sortie sous la forme d'ondulations.

La solution à ce problème consiste à corriger le module et la phase séparément. Dans la suite, nous allons détailler chacune des sources d'imperfections ainsi que la méthode de correction correspondante.

3.4.1 Correction du module de la STF du modulateur

La conception d'un modulateur $\Sigma\Delta$ à temps continu à partir de son équivalent à temps discret par la méthode de l'invariance impulsionnelle (voir annexe A, § A.4) conserve la fonction de transfert par rapport au bruit (NTF). La fonction de transfert par rapport au signal STF n'est plus un simple retard comme dans le cas discret. Elle s'exprime dans le domaine fréquentiel par l'équation (A.7). Le module de cette fonction de transfert (figure 3.20) montre qu'elle n'est pas plate dans la bande de fonctionnement du modulateur. Elle présente plutôt une forme parabolique qui n'est pas centrée obligatoirement au milieu de la bande.

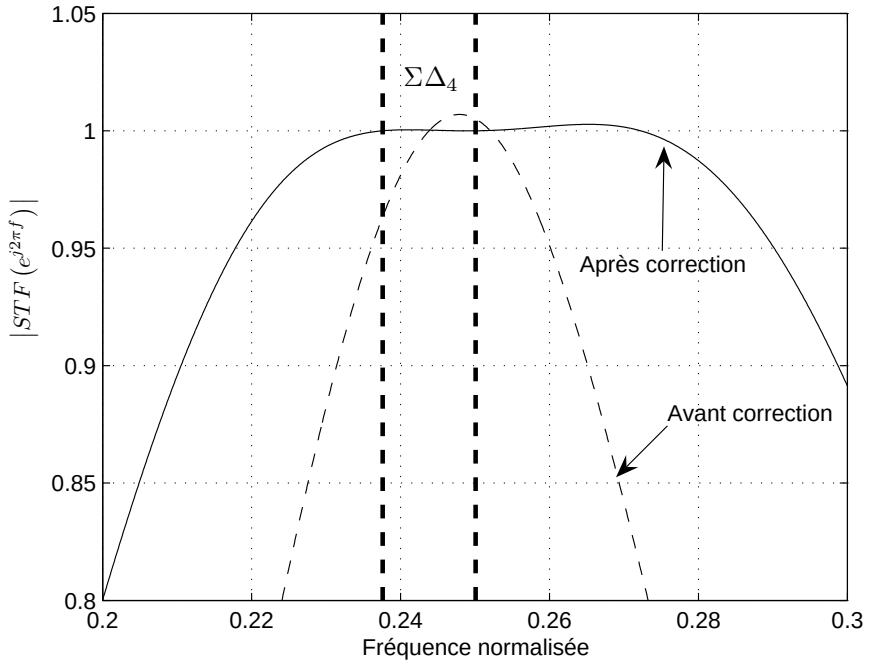

Fig. 3.20 – Fonction de transfert du modulateur à temps continu avant et après correction.

De ce fait, le signal utile dans la bande de fonctionnement du modulateur n'est pas uniformément amplifié. De plus, le gain de la STF à la frontière entre les bandes de fonctionnement de deux modulateurs adjacents n'est pas uniforme. Ces phénomènes introduisent des distortions qui se manifestent par des ondulations sur le spectre en sortie.

Afin de corriger ces imperfections, nous avons observé que le module de la STF présente une forme parabolique de concavité négative (figure 3.20), et nous avons proposé, pour corriger cet effet, d'utiliser un filtre numérique de réponse fréquentielle symétrique (i.e de réponse fréquentielle parabolique de concavité positive) permettant de compenser le module de la STF et d'avoir un gain proche de 0 dB dans la bande de fonctionnement. Ce type de réponse fréquentielle peut être réalisé, par exemple, avec un filtre numérique à 3 coefficients. L'expression de ce filtre, notée $C_1^k(z)$, est donnée par :

$$C_1^k(z) = g\left(-\varepsilon e^{-j2\pi\nu} + (1+2\varepsilon)\, z^{-1} - \varepsilon e^{j2\pi\nu} z^{-2}\right) \tag{3.28}$$

avec :

- ε : est la concavité du module du filtre,
- ν : est la différence entre la fréquence du milieu de la bande de fonctionnement f_c^k et la fréquence pour laquelle le module de la STF atteint son maximum,
- g : est l'inverse du module maximum de la STF permettant d'assurer un gain unitaire.

Le filtre $C_1^k(z)$ est appliqué aux signaux en bande de base, mais nous avons présenté le module de la STF avant et après correction dans la bande de fonctionnement du modulateur (figure 3.20) afin de bien visualiser l'effet de la correction. Les paramètres du filtre $C_1^k(z)$ (ε, φ, g) sont déterminés à partir d'un algorithme de calcul en temps réel dont les détails sont présentés au chapitre 4.

3.4.2 Correction du module du filtre de décimation

La décimation d'un signal par un facteur R_d engendre des repliements créés par les répétitions spectrales tous les $\frac{k}{R_d}$ $(0 \leq k \leq R_d - 1)$. Il est donc nécessaire de filtrer avant de décimer afin de limiter la bande des signaux à $\pm\frac{1}{2R_d}$ autour de la fréquence centrale [46, 47] comme le montre la figure 3.21.

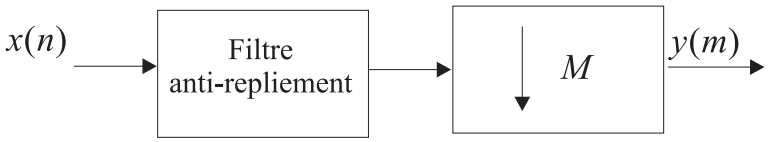

Fig. 3.21 – Schéma bloc du filtre décimateur.

Comme la fréquence du signal en sortie du modulateur ΣΔ est très élevée, il est indispensable d'avoir un filtre à faible complexité. Le filtre en peigne (*Comb-Filter*), en raison de sa simplicité de réalisation, est le meilleur choix pour ce type d'applications [45]. Ce filtre est un filtre moyenneur réalisant la moyenne simple sur R_d échantillons. Sa réalisation ne nécessite aucune opération de multiplication.

Le filtre en peigne introduit des repliements autour de la fréquence $\frac{1}{R_d}$ après la périodisation du spectre. Ces repliements se produisent loin de la bande du signal utile (autour de $f = 0$) et n'interfèrent pas avec le spectre du signal utile qui se trouve en basse fréquence. Ces repliements seront éliminés par la suite par le filtre FIR passe-bas qui suit la décimation. La fonction de transfert en z de ce filtre est donnée par l'équation (3.29), où K est le nombre de moyenneurs mis en cascade.

$$C(z) = \left(\frac{1}{R_d}\frac{1 - z^{-R_d}}{1 - z^{-1}}\right)^K \tag{3.29}$$

Pour un modulateur ΣΔ passe-bande d'ordre L, un filtre en peigne d'ordre $K \geq \frac{L}{2} + 1$ est suffisant pour atténuer le bruit de quantification qui se replira sur la bande désirée [45, 48]. Le module de la fonction de transfert de C(z) est donné par l'équation suivante :

$$\left|C\left(e^{j2\pi f}\right)\right| = \left|\frac{\sin(\pi f R_d T_e)}{R_d \sin(\pi f T_e)}\right|^K \tag{3.30}$$

Le tracé du module de C(z) est présenté sur la figure 3.22. On constate en observant la réponse fréquentielle du filtre en peigne que certains signaux seront atténués. Cette atténuation est maximale pour les signaux qui se trouvent à la limite supérieure de la bande de fonctionnement soit $\frac{f_2 - f_1}{2N} \times R_d$, ce qui se répercute sous forme d'ondulations sur le spectre du signal en sortie.

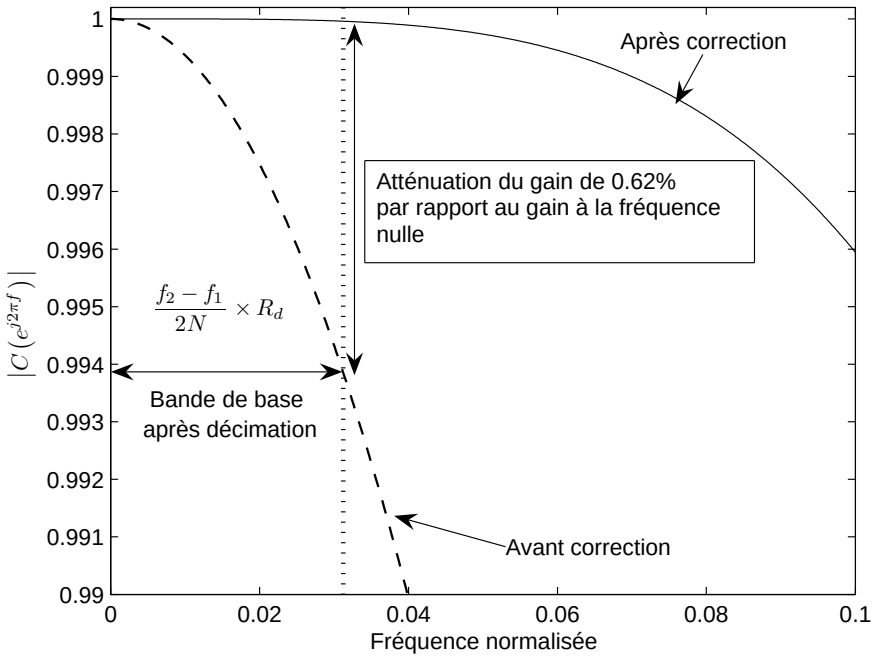

Fig. 3.22 – Module du filtre en peigne avant et après correction.

Pour corriger ce problème, l'idée est d'utiliser un filtre $C_2(z)$ dont le module est égal à $\left|\frac{1}{C(z)}\right|$ pour compenser l'atténuation. Une contrainte est que ce filtre doit corriger le gain dans la bande du signal sans avoir un gain élevé en dehors de cette bande car il produirait une amplification du bruit de quantification en dehors de la bande et dégraderait par la suite la performance du système. Il est donc préférable d'introduire le filtre correcteur $C_2(z)$ après la décimation pour bénéficier de l'élargissement de la bande utile d'un facteur R_d. Le gain du filtre en dehors de la bande utile est alors plus faible. Par ailleurs, cette solution facilite l'implémentation en travaillant à fréquence plus faible. Un filtre correcteur $C_2(z)$ à 3 coefficients est capable de réaliser la correction du module du filtre $C(z)$ dans la bande de fonctionnement. La fonction de transfert de $C_2(z)$ est donnée par :

$$C_2(z) = -\varepsilon + (1 + 2\varepsilon)z^{-1} - \varepsilon z^{-2} \tag{3.31}$$

Pour déterminer la concavité ε de ce filtre, nous exprimons tout d'abord le développement limité à l'ordre 1 du module de $C_2(z)$ et de celui de $\frac{1}{C(z)}$ en tenant compte du facteur de décimation (équation (3.32)).

$$\begin{aligned} \left|C_2(e^{j2\pi f})\right| &= 1 + 4\varepsilon \sin^2(\pi f) \approx 1 + 4\pi^2 f^2 \varepsilon + o(f^2) \\ \left|\frac{1}{C(z)}\right| &= \left|\frac{R_d \sin(\pi \frac{f}{R_d} T_e)}{\sin(\pi f T_e)}\right|^K \approx 1 + \frac{K\pi^2(R_d^2-1)}{6R_d^2} f^2 + o(f^2) \end{aligned} \tag{3.32}$$

Ensuite, en identifiant $|C_2(z)|$ à $\left|\frac{1}{C(z)}\right|$ nous obtenons l'expression de ε donnée par l'équation suivante :

$$\varepsilon = \frac{K\left(1 - \frac{1}{R_d^2}\right)}{24} \tag{3.33}$$

La largeur de la bande de base du signal après démodulation et décimation est égale à $\frac{f_2-f_1}{2N} \times R_d$ soit 0.0313 en valeur normalisée. Le filtre en peigne introduit une atténuation du gain du signal utile en bande de base allant jusqu'à 0.62% à la limite de la zone de transition entre les modulateurs adjacents (voir figure 3.22). Avec la correction, cette atténuation a diminué jusqu'à 0.01% dans la zone de transition. En pratique, ce filtre de correction sera inclus dans le filtre passe-bas $F_{pb}^k(z)$.

3.4.3 Raccordement des phases des modulateurs ΣΔ

La fonction de transfert du signal (STF) du modulateur ΣΔ présente une phase presque linéaire autour de la fréquence centrale f_C^k de la bande de fonctionnement $\left[f_A^k, f_B^k\right]$ (figure 3.23).

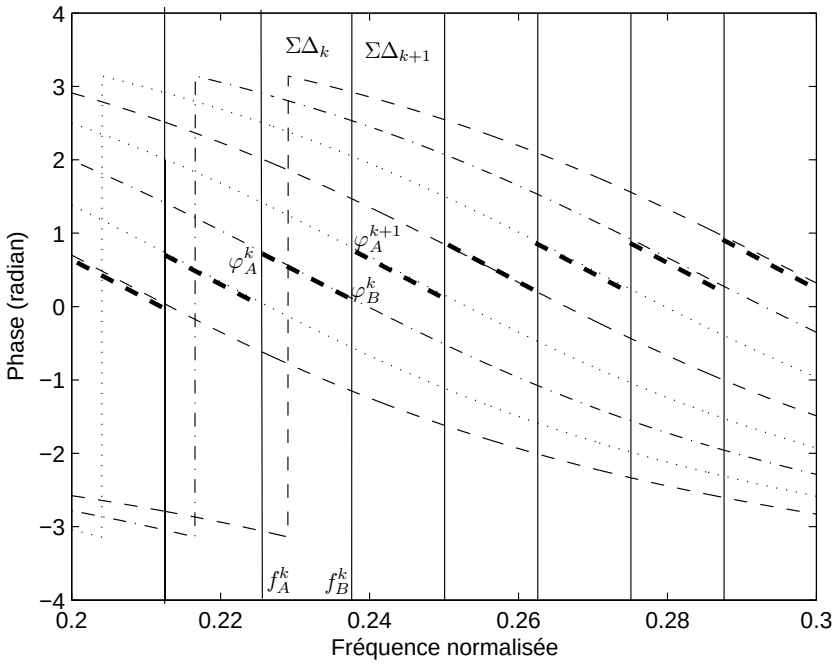

Fig. 3.23 – Réponse de Phase des modulateurs ΣΔ avant correction.

L'égale répartition des N modulateurs ΣΔ dans la bande du signal utile ($[0.2F_e, 0.3F_e]$) montre que la phase de la STF de chaque modulateur est linéaire dans sa bande de fonctionnement. Cependant, les STF des modulateurs adjacents ne possèdent pas la même valeur de phase à la fréquence limite f_B^k entre les deux bandes de fonctionnement. Soit φ_A^k et φ_B^k les phases correspondantes aux fréquences limites f_A^k et f_B^k du $k^{ième}$ modulateur. Un signal situé à la fréquence f_B^k aura deux phases différentes selon son traitement par le modulateur k ou k + 1. S'il est traité par le modulateur k, sa phase est φ_B^k, sinon c'est φ_A^{k+1} ($f_A^{k+1} = f_B^k$). Comme le traitement numérique se fait dans le domaine complexe (sur deux voies I et Q), le raccordement de phase dans la zone de transition entre les deux modulateurs adjacents k et k + 1 se fait en multipliant la sortie du modulateur k + 1 par $e^{-j\Delta\varphi}$ avec $\Delta\varphi = \varphi_A^{k+1} - \varphi_B^k$. La sortie du modulateur k est à son tour multipliée par $e^{-j\left(\varphi_A^k - \varphi_B^{k-1}\right)}$ pour se raccorder au modulateur d'ordre k − 1. En se basant sur ce

principe, le terme de correction est exprimé d'une façon générale par l'équation suivante :

$$C_3^k = \begin{cases} 1 & k = 1 \\ e^{-j \sum_{m=2}^{k} \left(\varphi_A^m - \varphi_B^{m-1}\right)} & k \geq 2 \end{cases} \tag{3.34}$$

La figure 3.24 montre les phases des différentes STF après la correction de phase par la multiplication par le terme C_3^k.

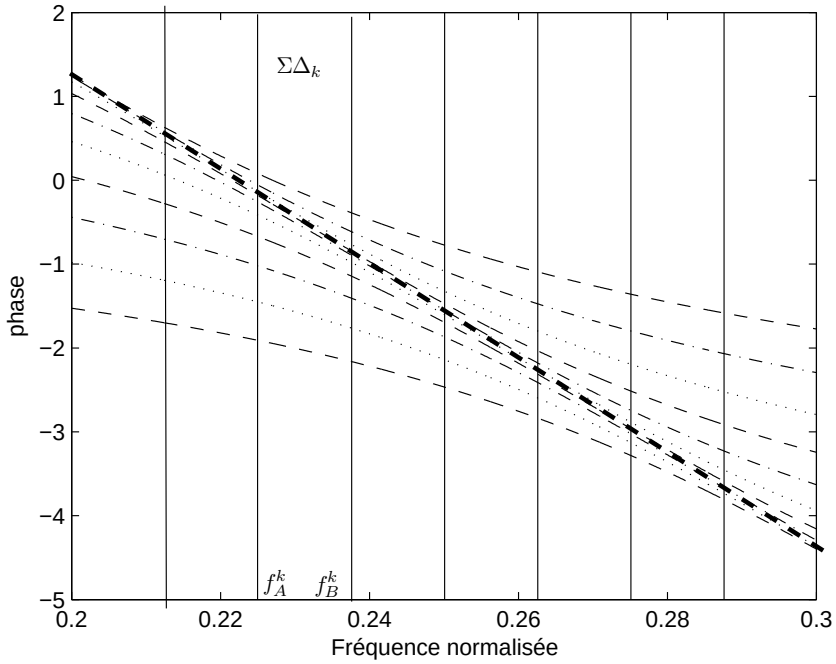

Fig. 3.24 – Réponse de Phase des modulateurs $\Sigma\Delta$ après correction.

Nous pouvons constater que les phases sont bien raccordées à la frontière des bandes de fonctionnement adjacentes et que le signal utile présente une phase presque linéaire dans toute sa bande [$0.2F_e$, $0.3F_e$].

3.4.4 Raccordement des phases des filtres passe-bas

Un autre problème de raccordement des phases est dû à la démodulation et à la modulation du signal de chaque modulateur (cf figure 3.18). En effet, le signal de sortie de chaque modulateur est ramené en bande de base suite à une démodulation complexe. Ensuite, il est filtré par le filtre en peigne avant d'être décimé d'un facteur R_d. Après la décimation, le signal est traité par les trois filtres passe-bas $C_1(z)$, $C_2(z)$ et $F_{pb}^k(z)$. Ensuite, il est modulé pour le ramener à sa bande initiale pour reconstruire à nouveau le signal utile passe-bande.

Le fait de ramener le signal en bande de base de chaque étage de traitement à une fréquence différente pour reconstruire le signal utile, crée un déphasage, introduit par les filtres passe-bas, à la limite des bandes de fonctionnement. Ce phénomène est illustré par la figure 3.25.

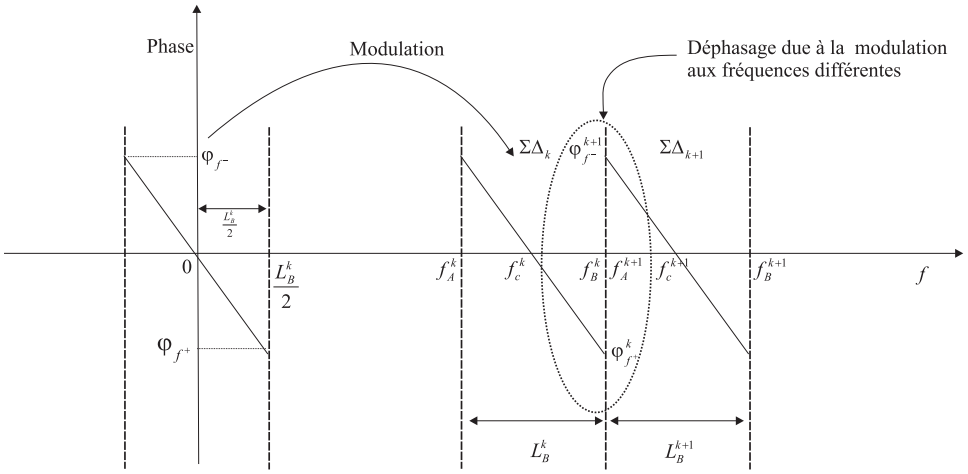

Fig. 3.25 – Déphasage introduit par les filtres passe-bas suite à la modulation.

Pour exprimer ce déphasage, nous partons de l'expression de la fonction de transfert des filtres passe-bas utilisés. Ces filtres sont de type I dont l'expression de la fonction de transfert pour un ordre P est donnée par :

$$H\left(e^{j2\pi f}\right) = \sum_{i=0}^{P} a_i \left(e^{j2\pi f}\right)^{-i} = e^{-j2\pi fM} \sum_{n=0}^{M} a_n \cos(2\pi fn) \quad \text{avec} \quad M = \frac{P-1}{2} \tag{3.35}$$

Ce type de filtre introduit un déphasage linéaire $\varphi = -2\pi \mathrm{Mf}$ au signal. La largeur de bande de base du signal traité par la voie k est égale à $\frac{\mathrm{L_B^k}}{2}$, où $\mathrm{L_B^k}$ est la largeur de la bande de fonctionnement $\mathrm{L_B^k} = \mathrm{f_B^k} - \mathrm{f_A^k}$.

La modulation consiste à replacer le spectre du signal utile de chaque voie k autour de la fréquence centrale $\mathrm{f_C^k}$ après avoir été traité en basse fréquence. Un signal utile situé à la fréquence $\mathrm{f_B^k}$ à la limite entre les bandes de fonctionnement des modulateurs k et k + 1 va avoir une phase $\varphi_{\mathrm{f+}}^{\mathrm{k}} = 2\pi\left(\mathrm{f_B^k} - \mathrm{f_C^k}\right)\mathrm{M}$, engendrée par la voie k, et une phase $\varphi_{\mathrm{f-}}^{\mathrm{k+1}} = 2\pi\left(\mathrm{f_A^{k+1}} - \mathrm{f_C^{k+1}}\right)\mathrm{M} = 2\pi\left(\mathrm{f_B^k} - \mathrm{f_C^{k+1}}\right)\mathrm{M}$ engendrée par la voie k + 1 (voir figure 3.25). Le raccordement de phase à la fréquence limite $\mathrm{f_B^k}$ consiste à décaler la phase de la voie k + 1 vers le bas de $\varphi_{\mathrm{f-}}^{\mathrm{k+1}} - \varphi_{\mathrm{f+}}^{\mathrm{k}}$. En tenant compte du fait que la voie k doit à son tour se raccorder à la voie k − 1, le facteur de correction de phase avant la décimation (c'est le cas du filtre en peigne), permettant d'avoir une phase linéaire dans toute la bande de fonctionnement, est donné par :

$$C_4^k = e^{-j2\pi\left(f_C^k - f_C^1\right)M} \tag{3.36}$$

Si le filtrage passe-bas est appliqué après la décimation, il suffit de multiplier la fréquence par $\mathrm{R_d}$ et le terme de correction devient :

$$C_4^k = e^{-j2\pi\left(f_C^k - f_C^1\right)MR_d} \tag{3.37}$$

En pratique, le filtre de correction $\mathrm{C_2(z)}$ est indépendant du modulateur ΣΔ. Il sera intégré dans le filtre passe-bas $\mathrm{F_{pb}^k(z)}$. Le nouveau filtre passe-bas aura pour fonction de transfert $\mathrm{G_{pb}^k(z)}$ donnée par :

$$G_{pb}^k(z) = F_{pb}^k(z) \times C_2(z) \tag{3.38}$$

La correction de la phase est réalisée en multipliant le signal en sortie de chaque voie par un nombre complexe regroupant les différents termes $\mathrm{C_3^k}$ et $\mathrm{C_4^k}$ de chacun des filtres passe-bas $\mathrm{C_1^k(z)}$,

$C_2(z)$ et $F^k_{pb}(z)$. En tenant compte des étapes de correction, le synoptique de l'architecture FBD avec le traitement numérique équivalent est représenté sur la figure 3.26.

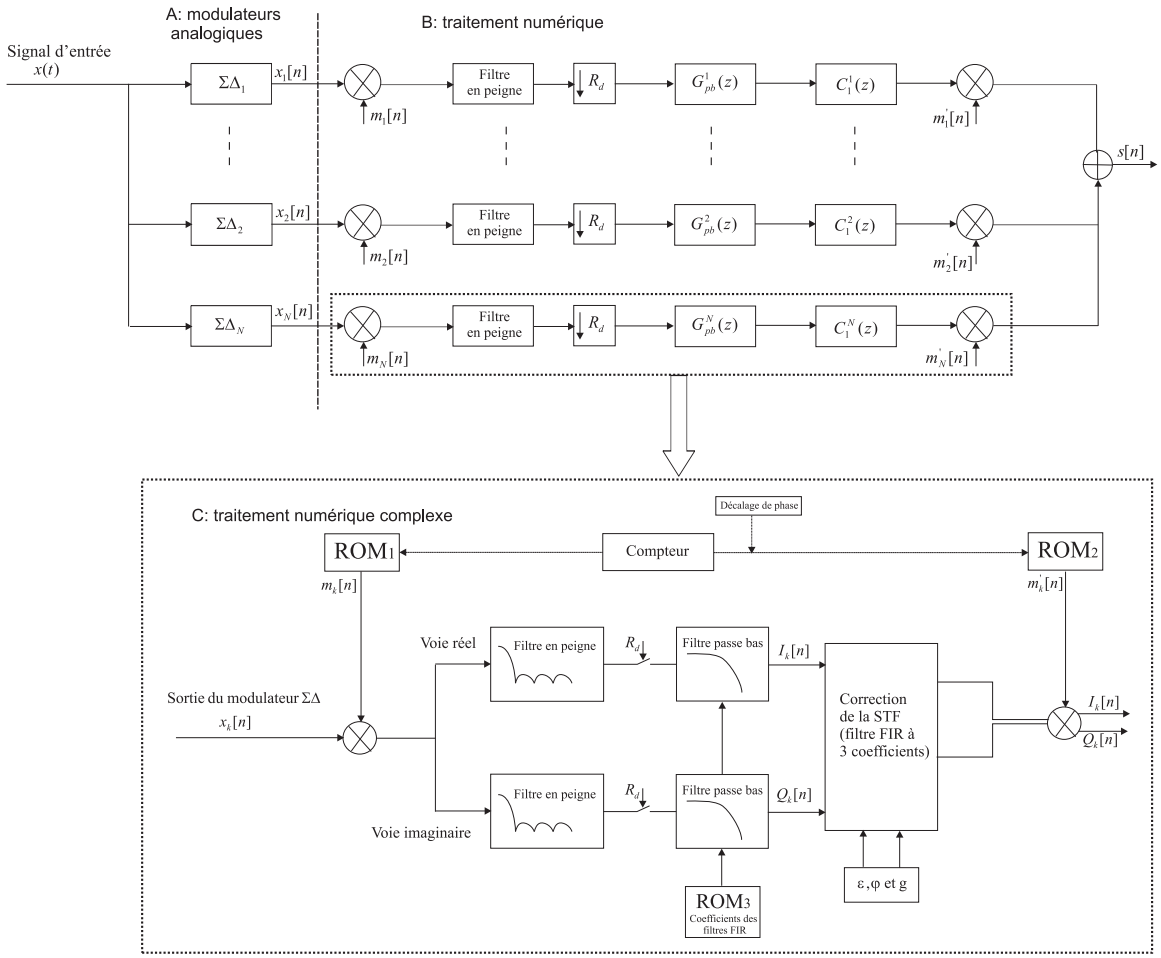

Fig. 3.26 – Architecture FBD avec des filtres de corrections.

3.5 Résultats de simulation avec la méthode de reconstruction avec démodulation

Nous avons considéré, pour évaluer cette méthode de reconstruction, l'exemple dont les spécifications ont été données dans le tableau 3.1. Les filtres passe-bas $F^k_{pb}(z)$ utilisés sont des filtres à réponse impulsionnelle finie FIR, choisis pour leur stabilité. La figure 3.27 représente la réponse fréquentielle des différents filtres $F^k_{pb}(z)$ à 64 coefficients après modulation (a), pour comprendre leur effet sur le spectre du signal utile en sortie, ainsi que leur somme (b). Le spectre du signal en sortie est compris dans l'intervalle $\left[0,\ \frac{F'_e}{2}\right]$, soit $\left[0,\ \frac{1}{2}\right]$, suite à la décimation du signal d'entrée ($[0.2F_e,\ 0.3F_e]$) d'un facteur de $R_d = 5$ ($F'_e = 5F_e$). Nous constatons que le module de la réponse fréquentielle de la somme des différents filtres $\left|\sum_{k=1}^{N} F^k_{pb}\left(e^{j2\pi f}\right)\right|$ présente :

- une atténuation vers les fréquences de bord (0 et 0.5) allant jusqu'à 50%,
- des ondulations dans la bande du signal utile de l'ordre de 10^{-3}.

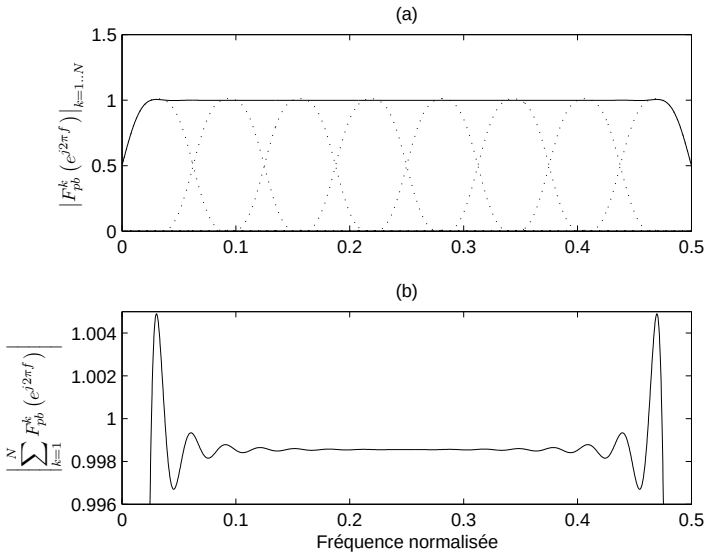

Fig. 3.27 – (a) Module des filtres passe-bas d'ordre 64, (b) Module de la somme des filtres passe-bas.

Nous avons utilisé un signal en entrée de type chirp d'amplitude normalisée 0.9 balayant toute la plage fréquentielle [$0.2F_e$, $0.3F_e$]. En appliquant la méthode de reconstruction avec démodulation, le spectre du signal en sortie du convertisseur FBD passe-bande est représenté par la figure 3.28. Le spectre du signal en sortie présente des atténuations sur les fréquences de bord ainsi que des ondulations qui sont dues aux filtres passe-bas $F_{pb}^{k}(z)$. Ces ondulations sont très faibles de sorte qu'elles n'influent pas sur les performances attendues du système de reconstruction. Comme dans le cas de la reconstruction directe, la performance, mesurée par l'ENOB, dépend essentiellement du filtre passe-bas $F_{pb}^{k}(z)$ dont le rôle est la suppression du bruit de quantification en dehors de la bande utile.

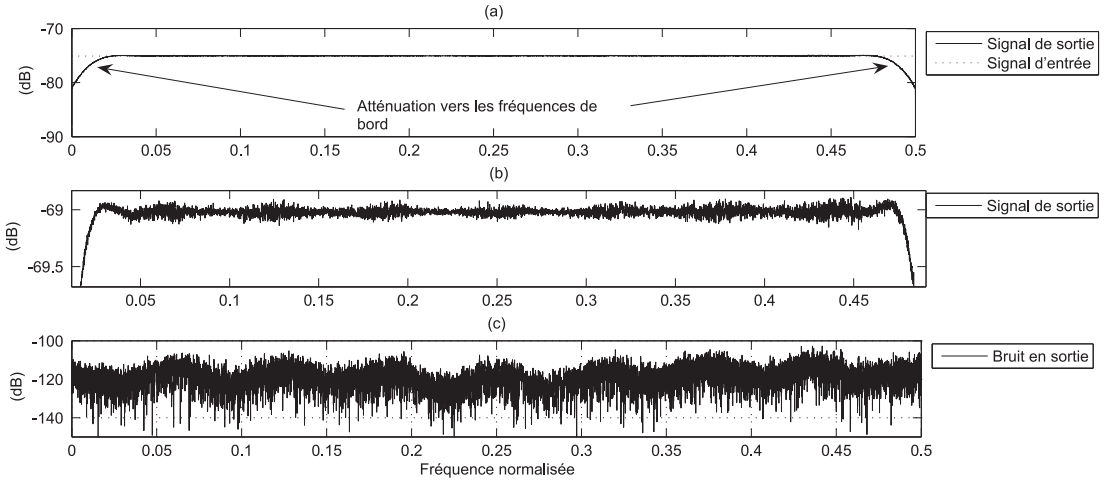

Fig. 3.28 – (a,b) Densité spectrale du signal d'entrée et du signal en sortie. (c) Densité spectrale du bruit de quantification résiduel en sortie.

Pour mettre en évidence le rôle de ce filtre passe-bas, nous avons calculé la perte en nombre de bits entre l'ENOB mesuré en sortie et l'ENOB de référence (13.3 bits) calculé à partir du spectre de bruit de référence. Cette perte en ENOB a été calculée pour différents ordres avec trois types de fenêtre de pondération (*Hamming*, *Hanning*, *Blackman*). Les valeurs obtenues sont récapitulées dans le tableau 3.2.

Tab. 3.2 – Perte en ENOB pour différents type de fenêtres en fonction du nombre de coefficients.

Nombre de coefficients	48	56	64	96	128	256
Perte en résolution (bits), *Hamming*	-1.17	-0.64	-0.04	0.05	0.06	0.05
Perte en résolution (bits),*Blackman*	-0.46	-0.28	-0.15	0.03	0.08	0.07
Perte en résolution (bits), *Hanning*	-0.22	-0.12	-0.04	0.06	0.08	0.05

Nous constatons que :

- l'augmentation du nombre de coefficients au delà de 64 coefficients n'améliore pas la résolution de plus de 0.5 bits,
- pour des ordres faibles (≤ 64), la fenêtre de *Hanning* présente une meilleure résolution par rapport aux autres fenêtres parce qu'elle possède des lobes secondaires plus faibles.

Cette méthode de reconstruction présente un grand avantage par rapport à la méthode de reconstruction directe : elle nécessite 8 filtres passe-bas de 64 coefficients fonctionnant à une fréquence 5 fois plus faible que la fréquence d'échantillonnage pour atteindre la résolution de 13.3 bits, alors que pour une résolution similaire, la méthode de reconstruction directe nécessite 8 filtres passe-bande de 512 coefficients fonctionnant à la fréquence d'échantillonnage.

3.6 Conclusion

Nous avons présenté dans ce chapitre le concept du CAN passe-bande basé sur l'architecture FBD passe-bande. Cette architecture se compose de N modulateurs $\Sigma\Delta$ passe-bande distribués de façon équi-réparti dans la bande du signal utile à convertir. La performance théorique de cette architecture a été estimée en nombre de bits effectifs (ENOB) calculée à partir de la puissance de bruit sur la base d'un spectre de référence.

Deux méthodes de reconstruction numérique du signal ont été proposées :

- la première consiste à placer un filtre passe-bande derrière chaque modulateur pour reconstruire le signal utile.
- la deuxième consiste à ramener le signal en bande de base, à filter le bruit de quantification hors bande utile et à moduler le signal dans sa bande de départ. Dans ce cas, des processus de raccordement de phases et de correction de modules ont été nécessaires pour améliorer la performance du signal numérique en sortie.

La deuxième solution se révèle très avantageuse par rapport à la première, car pour obtenir un même nombre de bits effectifs, elle nécessite beaucoup moins de puissance de calcul.

Dans ce chapitre, les modulateurs $\Sigma\Delta$ ont été supposés idéaux. Les erreurs des imperfections des composants analogiques sur la performance du système ont été négligées. L'impact de ces sources d'erreurs sur l'architecture FBD passe-bande fait l'objet d'une étude détaillée dans le prochain chapitre. Des modifications à apporter à l'architecture FBD pour compenser l'effet de ces erreurs y sont également présentées.

Chapitre 4

Architecture EFBD : une architecture robuste aux imperfections de l'analogique

Objectif

Ce chapitre présente l'étude de l'influence du décalage des fréquences centrales des modulateurs dû aux dispersions technologiques sur la performance d'un CAN avec l'architecture FBD passe-bande présentée au chapitre 3. Ensuite, nous proposons une version étendue de l'architecture FBD passe-bande. Cette nouvelle version présente une plus grande robustesse aux imperfections de fabrication. La robustesse est caractérisée au travers de plusieurs simulations avec différents types d'erreurs analogiques (dispersions locales et globales). Une méthode de calibration basée sur la minimisation de la puissance de bruit de quantification est développée pour adapter le traitement numérique aux imperfections des modulateurs analogiques. D'autres méthodes de calibration du module et de la phase du signal numérisé sont également développées.

4.1 Introduction

Le principe et les performances dans le cas idéal d'un CAN large bande à base de décomposition fréquentielle (FBD) passe-bande ont été étudiés au chapitre 3. Les fréquences centrales des modulateurs ont été choisies de façon à minimiser la puissance de bruit. Le traitement numérique a été défini (largeur de bande des filtres FIR, fréquence de démodulation et coefficients de correction de phase, etc.), en se basant sur les caractéristiques nominales des modulateurs, de façon à avoir une puissance de bruit minimale et à reconstruire le signal avec le minimum de distorsion possible.

En pratique, les fréquences centrales des modulateurs ne correspondent pas aux valeurs nominales. Le processus de fabrication des composants analogiques introduit des erreurs sur la valeur attendue de ces composants. Ces erreurs se répercutent sur la valeur de la fréquence centrale de chacun des résonateurs constituant le modulateur $\Sigma\Delta$. La figure 4.1 représente le module de la NTF dans le cas idéal (a) et dans le où les fréquences centrales des résonateurs ont subi un décalage d'une demie sous-bande (b). La puissance de bruit récupérée par chacune des voies est beaucoup plus grande que celle obtenue dans le cas idéal.

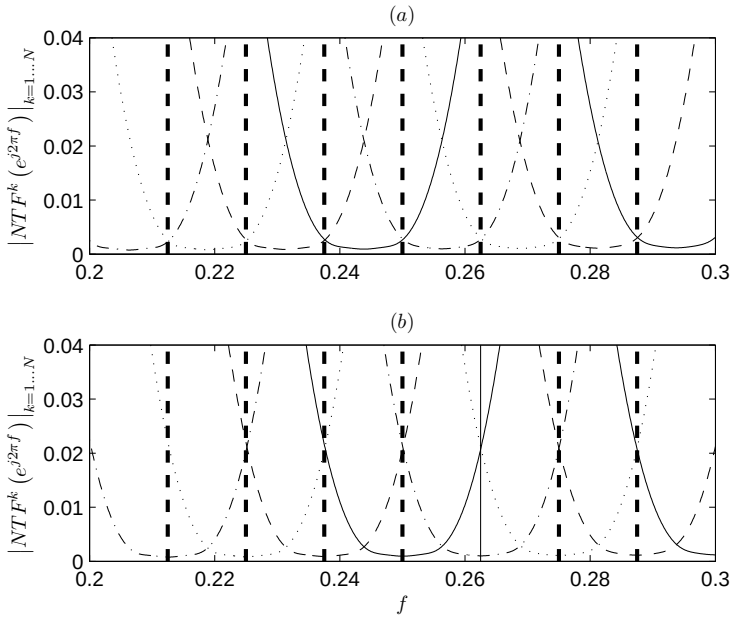

Fig. 4.1 – Module des NTF, (a) cas idéal, (b) décalage identique de toutes les fréquences centrales d'une demie sous-bande.

Pour estimer la dégradation des performances de l'architecture, introduite par l'imperfection des composants analogiques, nous avons calculé l'ENOB en faisant varier toutes les fréquences centrales des résonateurs de façon identique. Le traitement numérique est toujours adapté au cas idéal (largeur de bande et fréquence de démodulation). La figure 4.2 représente l'ENOB calculé et ceci, pour deux facteurs de qualité.

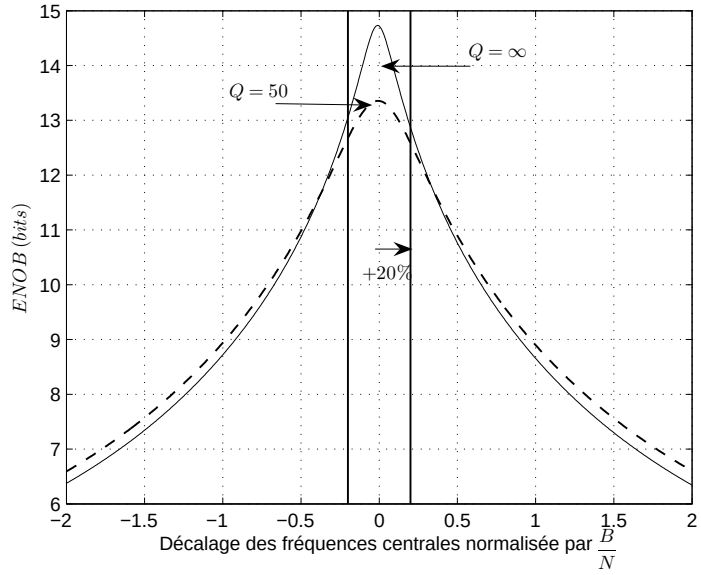

Fig. 4.2 – L'ENOB en fonction d'un décalage identique sur les fréquences centrales des résonateurs.

Nous constatons qu'avec un facteur de qualité infini, la résolution est meilleure mais le système est très sensible. Les résonateurs à facteur de qualité infini, en plus des difficultés rencontrées pour les réaliser, ne présentent donc pas d'intérêt. En effet, si l'architecture comporte des résonateurs ayant un facteur de qualité de 50, un décalage de $20\% \times \left(\frac{B}{N}\right)$ des fréquences centrales entraîne une chute de l'ENOB de 1 bit tandis que cette perte est de 3 bits avec $Q = \infty$.

4.2 Architecture EFBD

La performance attendue de l'architecture FBD est conditionnée par la précision de réalisation des composants analogiques, qui dépend du processus de fabrication. Comme il est impossible d'obtenir avec ces processus les valeurs nominales des composants, l'idée est d'étendre l'architecture FBD passe-bande pour la rendre plus robuste aux imperfections analogiques. Le principe de cette extension d'architecture est d'ajouter deux modulateurs $\Sigma\Delta$ à chaque extrémité de la bande du signal utile (voir figure 4.3).

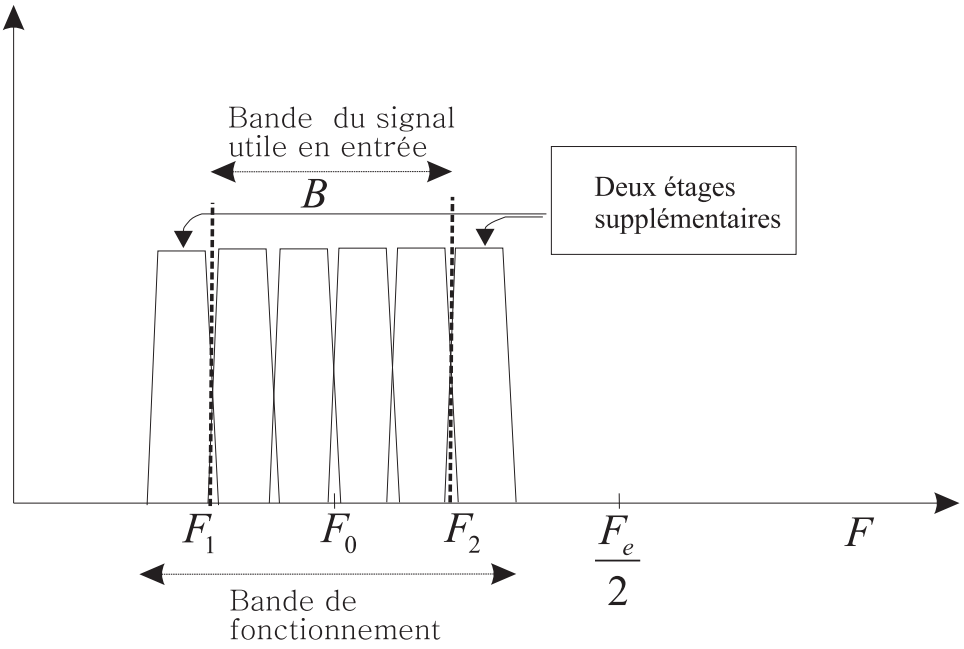

Fig. 4.3 – Bande de fonctionnement de l'architecture EFBD.

Les deux nouveaux modulateurs seront numérotés « 0 » et « N + 1 » (voir figure 4.4). Nous appellerons cette nouvelle architecture l'architecture EFBD (Extended Frequency Band Decomposition).

Avec l'architecture EFBD (voir § 4.2.2), les performances restent proches du cas idéal même pour un décalage des fréquences centrales allant jusqu'à $\frac{B}{N}$ vers la droite ou vers la gauche. Avec l'architecture FBD, nous avions pour un décalage similaire une perte en résolution d'environ 5 bits. Dans notre exemple de simulation (un banc de 8 modulateurs avec une bande du signal utile entre $F_1 = 0.2F_e$ et $F_2 = 0.3F_e$), le nombre total de modulateurs $\Sigma\Delta$ sera de 10 $(N + 2)$. Pour cet exemple, l'erreur relative maximale admissible pour ne pas avoir de chute de résolution, varie entre -4% $(-\frac{B/N}{f_2+B/N})$ et $+6.67\%$ $(+\frac{B/N}{f_1-B/N})$.

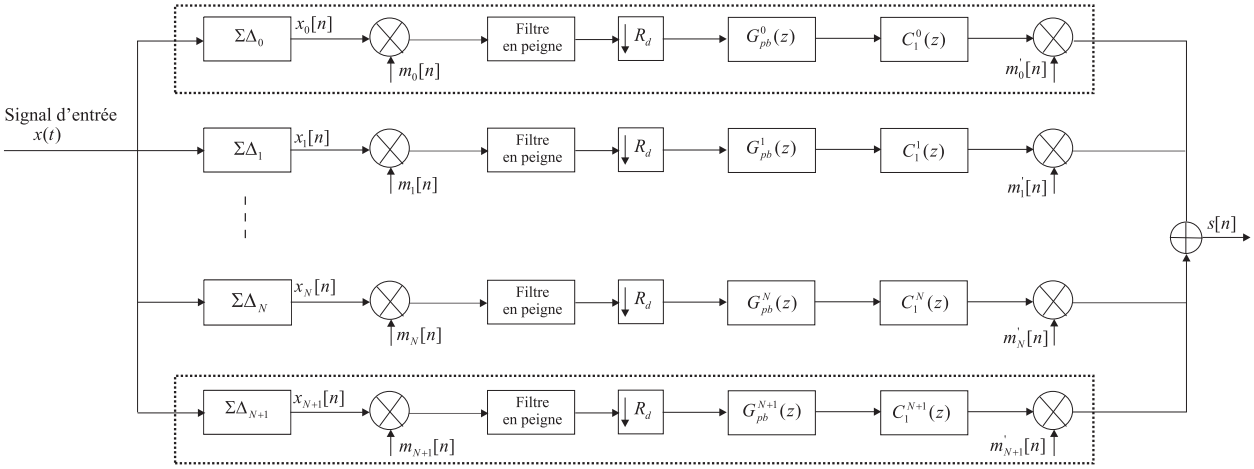

FIG. 4.4 – Banc de modulateurs de l'architecture EFBD.

4.2.1 Adaptation du traitement numérique

Nous avons vu que l'erreur absolue maximale admissible sur les fréquences centrales pour l'architecture EFBD est $\left|\frac{B}{N}\right|$. Il reste à adapter le traitement numérique et à déterminer les bandes des filtres FIR de façon à reconstruire le signal avec la puissance minimale de bruit. La règle est d'utiliser pour chaque bande de fréquences le modulateur qui présente le niveau de bruit de quantification le plus faible. En se basant sur cette règle, le choix logique est de définir la bande de chaque étage k comme étant la plage fréquentielle qui se trouve entre les deux points d'intersection de $\mathrm{NTF^k}$ avec les NTF des étages adjacents soit $\mathrm{NTF^{k-1}}$ et $\mathrm{NTF^{k+1}}$ (voir figure 4.5).

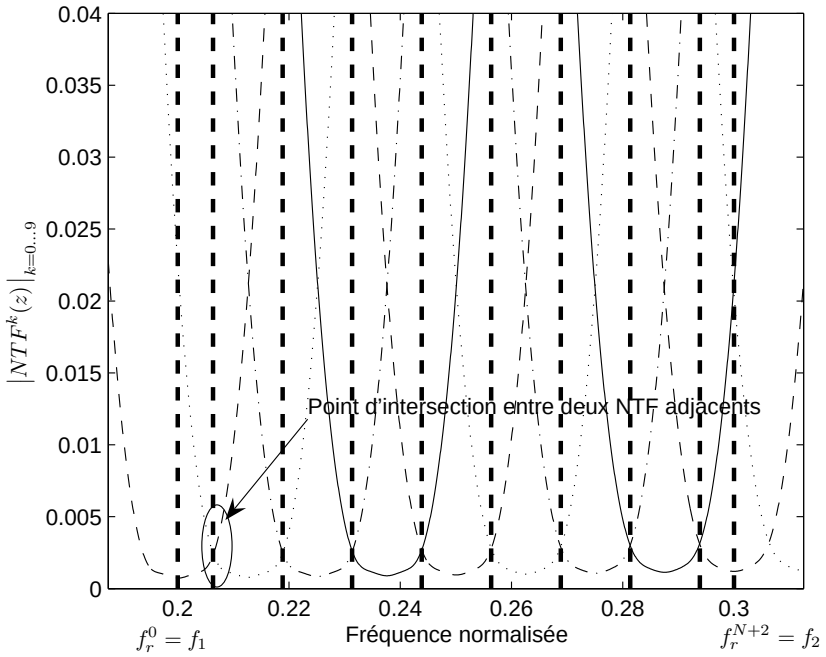

FIG. 4.5 – Modules des $\mathrm{NTF^k}$ pour un décalage de $\frac{B}{2N}$ pour toutes les fréquences centrales.

La limite inférieure de la bande du premier étage utilisé (f_r^0) est toujours égale à f_1, la limite supérieure de la bande du dernier étage utilisé (f_r^{N+2}) est toujours égale à f_2 et la relation $f_r^k \leq f_r^{k+1}$ est toujours vérifiée. Quand un modulateur n'est pas utilisé, les deux fréquences limites de sa bande de fonctionnement sont égales. Dans le cas idéal, les limites des différentes bandes de fonctionnement sont données par :

$$f_r^k = f_1 + (k-1)\frac{f_2 - f_1}{N}, k = 1 \ldots N+1 \tag{4.1}$$

La figure 4.5 montre le module de la NTF des 10 modulateurs dans le cas où toutes les fréquences centrales ont été augmentées de $\frac{B}{2N}$ de leurs valeurs initiales (le facteur de qualité Q a été considéré égal à 50). Dans ce cas particulier, les fréquences limites entre les différentes bandes sont données par :

$$\begin{cases} f_r^0 = f_1 \qquad (= 0.2) \\ f_r^k = f_1 + (k-0.5)\frac{f_2 - f_1}{N}, \qquad k = 1..N \\ f_r^{N+1} = f_r^{N+2} = f_2 \qquad (= 0.3) \end{cases} \tag{4.2}$$

La largeur de bande normalisée des premier et neuvième modulateurs est égale à (0.00625), c'est-à-dire la moitié de la bande des autres étages ($0.0125 = \frac{B}{N}$). Dans ce cas, le dernier modulateur n'est pas utilisé car le minimum de sa NTF se trouve en dehors de la bande utile du signal.

La figure 4.5 permet d'observer que la largeur de bande optimale pour chaque modulateur (assurant une puissance de bruit minimale) est celle qui a pour bornes les fréquences d'intersection de la NTF du modulateur en question avec la NTF des modulateurs adjacents. Ces fréquences d'intersection devront donc être déterminées. Pour cela, l'équation suivante doit être résolue :

$$\left|NTF^k(e^{j2\pi f_r^{k+1}})\right|^2 = \left|NTF^{k+1}(e^{j2\pi f_r^{k+1}})\right|^2 \tag{4.3}$$

où f_r^{k+1} est la fréquence limite entre les bandes de fonctionnement de l'étage k et $k+1$.

Dans notre architecture, la fonction de transfert par rapport au bruit du modulateur à temps continu d'ordre 6 est approchée, dans le domaine z, par (voir démonstration en annexe B.1.3) :

$$\left|NTF^k(e^{j2\pi f})\right|^2 \approx \prod_{j=1}^{3}\left[\frac{4\pi^2}{(c_j)^2}\left(4(f-f_j)^2 + \left(\frac{f_j}{Q_j}\right)^2\right)\right] \tag{4.4}$$

La solution de l'équation 4.3 nécessite la connaissance exacte de la NTF^k de chaque modulateur. D'autre part, la fonction NTF^k évolue quand les fréquences centrales des résonateurs changent de valeur suite aux erreurs de l'analogique. D'où la nécessité d'utiliser des méthodes numériques permettant de déterminer les limites entre les bandes des différents étages. Le choix de la méthode numérique dépend de :

- la puissance de calcul nécessaire et de la surface d'implantation,
- la précision souhaitée sur les fréquences centrales.

Parmi les méthodes numériques envisageables pour la détermination des fréquences d'intersection, deux approches seront présentées :

- La première consiste à identifier le modulateur $\Sigma\Delta$ à travers un modèle mathématique pour déterminer les fréquences centrales des résonateurs puis les fréquences d'intersection entre les NTF adjacentes. Cette partie d'identification sera développée en détail au paragraphe 4.4.
- La deuxième se base sur la minimisation de la puissance de bruit en sortie en faisant varier chaque fréquence f_r^k dans un intervalle donné autour de sa valeur théorique. Cette voie de recherche fait l'objet du paragraphe 4.5.

L'une de ces méthodes numériques permettra de déterminer les limites entre les bandes adjacentes des différents étages. Par conséquent, nous connaissons aussi les valeurs des fréquences centrales des bandes de fonctionnement :

$$f_c^k = \left.\frac{f_r^{k+1} + f_r^k}{2}\right|_{k=0...N+1} \tag{4.5}$$

Une fois connues les fréquences centrales, les corrections sur le traitement numérique pourront être effectuées. Ces corrections sont au nombre de 4 et concernent :

- le démodulateur avec la séquence de démodulation $\mathrm{m_k}\,[\mathrm{n}]$,
- le modulateur avec la séquence de modulation $\mathrm{m_k^{'}}\,[\mathrm{n}]$,
- le filtre FIR passe-bas avec sa largeur de bande qui doit maintenant être donnée par $\mathrm{R_d} \times \Delta \mathrm{f_k}$, où $\Delta \mathrm{f_k}$ est la largeur de demie-bande donnée par :

$$\Delta f_k = \frac{f_r^{k+1} - f_r^k}{2} \tag{4.6}$$

- les coefficients de raccordement de phase C_4^k qui sont calculés par la même formule (3.37) mais avec les nouvelles valeurs des fréquences centrales f_c^k.

4.2.2 Robustesse de l'architecture EFBD

L'objet de ce paragraphe est d'évaluer l'erreur maximale admissible sur les fréquences centrales des résonateurs par rapport aux corrections pouvant être apportées par l'adaptation du traitement numérique. Cette erreur maximale sera évaluée en plusieurs étapes. En premier lieu, elle est déterminée théoriquement à l'aide d'une simulation pour laquelle toutes les fréquences centrales des résonateurs sont décalées identiquement (figure 4.6).

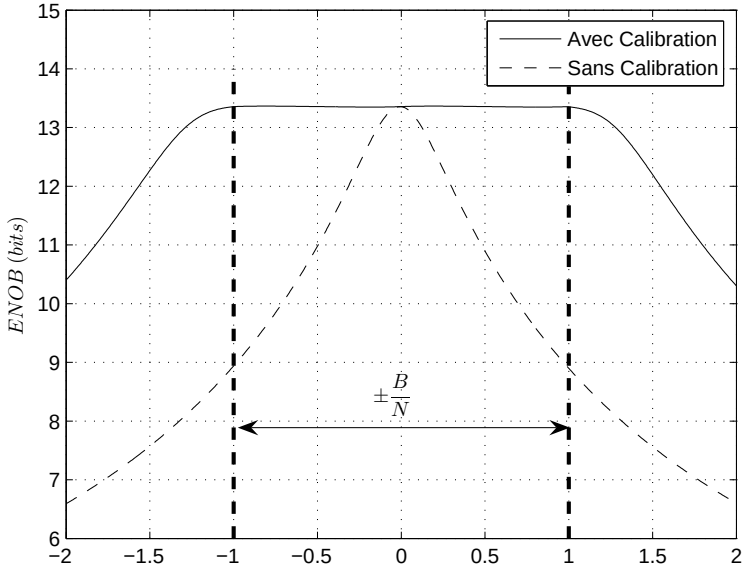

Fig. 4.6 – La résolution de l'architecture EFBD en fonction du décalage.

D'après ces résultats de simulation, l'erreur maximale admissible est un décalage de $\pm\frac{B}{N}$ (largeur de sous-bande). Par contre, avec le traitement sans calibration, un décalage de $\frac{B}{N}$ entraîne une perte supérieure à 4 bits.

En second lieu, l'erreur maximale est évaluée à partir de considérations sur les filières de fabrication et sur les différents types de résonateurs implantables.

En effet, l'erreur sur les fréquences centrales des résonateurs ne se limite pas à un simple décalage. Les dispersions sur les composants analogiques à l'origine de la variation des fréquences centrales peuvent être décomposées en une somme de deux termes indépendants [49] :

- Les **dispersions globales**, jusqu'à 20%, représentent les dispersions entre tranches du wafer et les dispersions entre chaque circuit sur la même tranche. Tous les composants identiques (capacités, inductances) sur la même tranche sont décalés dans le même sens. Les dispersions globales peuvent être modélisées par un facteur constant multipliant toutes les fréquences centrales des résonateurs.
- Les **dispersions locales**, de l'ordre de 1%, représentent les dispersions entre les composants d'un même circuit. Elles peuvent être considérées, avec des technologies classiques (CMOS 0.35 μm), comme négligeables par rapport aux dispersions globales. Cependant, l'augmentation des finesses de gravure les rend de plus en plus importantes. Les dispersions locales sont dues à des phénomènes physiques et au processus de lithographie : non uniformité des températures de recuit, non uniformité des concentrations de dopage, erreurs sur les largeurs de transistor, etc. Ces erreurs interviennent au niveau des transistors et des différents éléments passifs constituant le modulateur $\Sigma\Delta$. Les dispersions locales sont caractérisées par leur répartition aléatoire.

Avec l'évolution des technologies, il est aujourd'hui possible d'intégrer des résonateurs, de type Gm-C [50, 51] ou Gm-LC [52, 53], pour la réalisation de modulateurs passe-bande à temps continu. Des circuits de ce type, pour le travail à hautes fréquences, ont vu le jour dans des technologies diverses telles que le BiCMOS sur SiGe [53] et les technologies III-V(InP,GaAs). L'erreur introduite sur les fréquences centrales dépend du type de résonateur et de la technologie utilisée.

Les résonateurs Gm-LC comportent le plus souvent une résistance négative pour augmenter leur facteur de qualité. Dans ce cas, la fréquence centrale dépend des valeurs de L, C mais aussi de la résistance négative.

Les résonateurs Gm-C sont le plus souvent constitués de deux amplificateurs à transconductance et de capacités intégrées. Au premier ordre, la fréquence centrale dépend des valeurs de transconductances et capacités, mais d'autres paramètres interviennent comme les résistances de sortie des amplificateurs à transconductance.

Dans le cadre de ce travail de thèse, l'architecture des modulateurs $\Sigma\Delta$ et la technologie de réalisation n'ont pas encore été définis. Afin d'estimer la robustesse de l'architecture EFBD face aux dispersions analogiques, nous nous sommes basés sur une erreur relative de 20% sur les fréquences centrales des résonateurs. Cette erreur résulte des dispersions globales sur les composants passifs (L et C) dans une technologie classique (CMOS 0.35 μm) et se répercute sur les fréquences centrales comme une erreur du même ordre de grandeur.

D'une façon générale, les erreurs sur la fréquence centrale proviennent en grande partie des composants passifs. C'est pourquoi des méthodes de rattrapage grossier sont utilisées comme l'ajout de composants supplémentaires pouvant être sélectionnés par des interrupteurs, ou bien des corrections au Laser. Avec ces corrections, nous ferons l'hypothèse qu'il est possible de diminuer l'erreur sur les fréquences centrales jusqu'à 5%.

L'étude de la robustesse de l'architecture s'est faite en deux temps. Dans un premier temps nous avons étudié l'effet des dispersions globales sur les performances de l'architecture EFBD en modélisant cette dispersion par un facteur multiplicatif appliqué aux fréquences centrales des résonateurs. Dans un second temps, nous avons étudié l'impact de la dispersion globale plus la dispersion locale sur la performance de l'architecture EFBD. Les dispersions locales sont modélisées par un bruit blanc gaussien $\mathrm{N}\left(0, \sigma^2\right)$ affectant chaque fréquence centrale de chaque résonateur.

La figure 4.7 montre l'impact des dispersions globales sur l'ENOB. Nous pouvons noter que les dispersions globales ont une faible influence sur la performance attendue. Avec un coefficient multiplicatif légèrement inférieur à 1, la résolution est légèrement améliorée car les fréquences centrales sont alors plus serrées, ce qui permet d'avoir une puissance de bruit légèrement plus faible. En limitant cette erreur à 5%, la perte en résolution peut atteindre au maximum 0.1 bit avec la calibration du traitement. Par contre, la perte en résolution est de 4.5 bits sans la calibration.

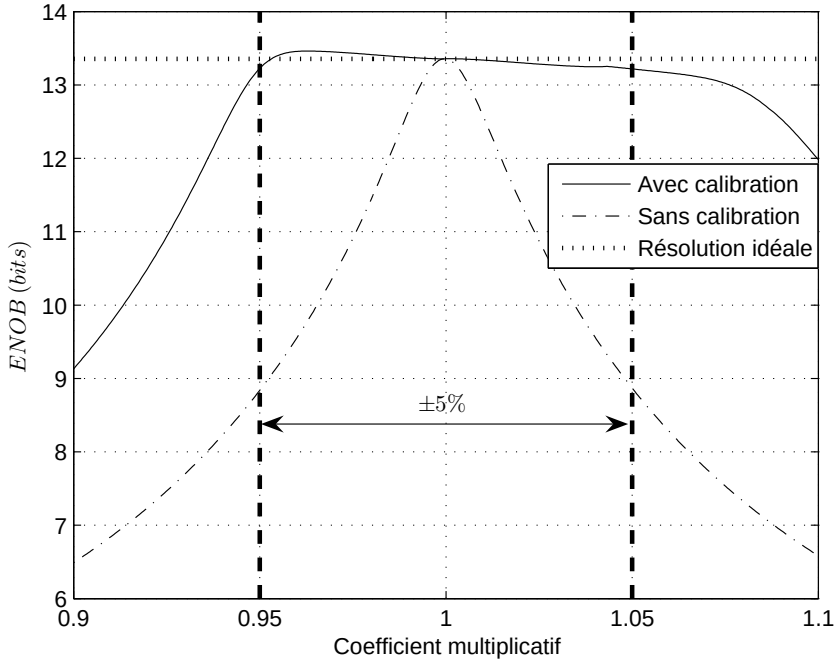

Fig. 4.7 – Effet du coefficient multiplicatif sur l'ENOB.

Dans un deuxième temps, nous avons considéré à la fois les dispersions globales et locales. La figure 4.8 représente la densité de probabilité estimée de l'ENOB obtenue en considérant les deux types d'erreur. Ces résultats proviennent de trois simulations de Monte-Carlo à 300 tirages. Nous avons considéré pour ces simulations une dispersion globale de 1% (toutes les fréquences centrales sont multipliées par 1.01) et une dispersion locale aléatoire modélisée par des sources d'erreur d'écart-type $\sigma = (5\%, 10\%, 20\%)$ de $\frac{\mathrm{B}}{\mathrm{N}}$, soit 0.25%, 0.5% et 1% d'écart type relatif par rapport à la fréquence centrale de la bande utile $\mathrm{f} = \frac{1}{4}$. Avec un écart-type de 10%, la perte en résolution est au maximum de 0.2 bit ce qui reste acceptable. Avec un écart-type de 20%, la perte en résolution peut atteindre 1 bit et représente un pire-cas.

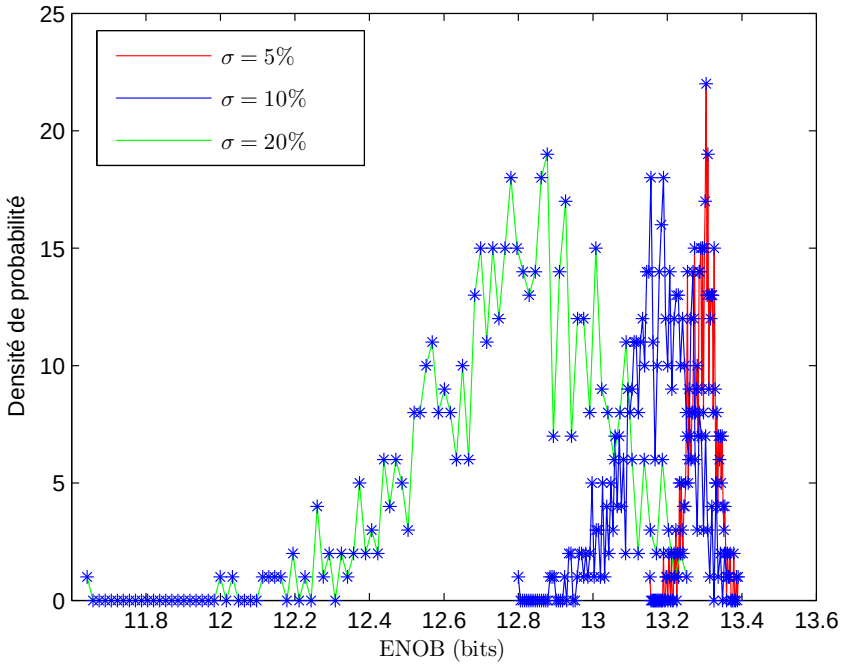

Fig. 4.8 – Densité de probabilité de l'ENOB estimée avec $\sigma = 5\%$, 10% et 20%.

4.3 Détermination des bandes de fonctionnement

L'architecture EFBD permet, avec la calibration du traitement numérique, de compenser les erreurs de fabrication des composants analogiques et de maintenir la résolution prévue par l'architecture FBD. La calibration du traitement numérique nécessite la connaissance, a priori, des fréquences limites entre les différentes bandes du traitement. La détermination de ces fréquences limites nécessite l'application de méthodes numériques car la solution analytique de l'équation (4.3) exige la connaissance de la forme exacte de la NTF de chaque modulateur.

Avant de présenter nos deux axes de recherche pour la détermination de ces fréquences limites f_r^k, la précision avec laquelle celles-ci doivent être déterminées est évaluée. Pour cela, nous avons considéré le cas où tous les modulateurs sont idéaux (les fréquences centrales sont égales à leurs valeurs théoriques). Ensuite, nous avons calculé la résolution en introduisant une erreur identique sur les valeurs théoriques des fréquences limites des bandes de fonctionnement. La figure 4.9 représente la résolution obtenue en fonction de l'erreur introduite normalisée à $\frac{B}{N}$. Nous pouvons constater qu'une erreur de $\pm 4\%$ de la largeur de sous-bande $\frac{B}{N}$ ($\pm 0.05\%$ de la fréquence d'échantillonnage) entraîne une chute de la résolution inférieure à 0.1 bit. Par conséquent, une erreur de 0.1% ($2 \times 0.05\%$) sur les fréquences limites entraîne une erreur moins de 0.1 bit. De ce fait, les fréquences limites f_r^k peuvent être quantifiées avec un pas $q_f = 10^{-3} F_e$. Ce résultat est très important pour l'implantation du traitement numérique et des algorithmes de calibration qui seront évoqués plus tard.

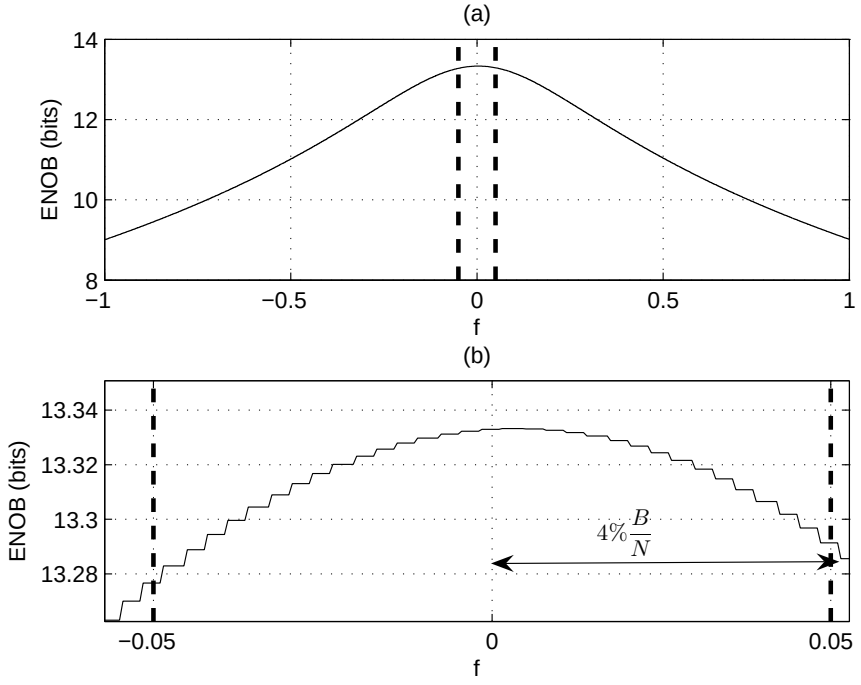

Fig. 4.9 – (a) Effet de l'erreur sur les fréquences limites sur la performance globale, (b) zoom autour d'une erreur de $\pm 4\%\frac{\mathrm{B}}{\mathrm{N}}$.

4.4 Identification de la NTF

La première méthode proposée pour définir les bandes de fonctionnement de chaque modulateur consiste à trouver, en première étape, la forme exacte de chaque $\mathrm{NTF^k(z)}$, et ensuite de résoudre l'équation (4.3). La $\mathrm{NTF^k(z)}$ peut être reconstruite par une interpolation polynomiale à partir de la connaissance de ses trois minimums, *i.e.*, à partir des trois zéros de la $\mathrm{NTF^k(z)}$.

L'identification du modulateur $\Sigma\Delta$ consiste à chercher un modèle mathématique permettant d'avoir le même comportement s'il est excité par le même signal d'entrée. La recherche de ce modèle nécessite le choix d'une structure mathématique qui pourrait décrire le système et la détermination des paramètres de la structure choisie. Le choix de la structure dépend de l'architecture physique du modulateur en question. En se basant sur le modèle linéaire du modulateur, le modèle mathématique qui correspond le mieux à la structure du modulateur est le modèle $ARMAX$(Auto Regressive Moving Average with eXternal inputs).

Notre but est de déterminer les trois zéros de la $\mathrm{NTF^k(z)}$. De ce fait, il est plus judicieux de mettre à zéro l'entrée du modulateur. Dans ce cas, la sortie du modulateur n'est que le bruit de quantification filtré par NTF(z). D'une part, cette hypothèse permet de réduire le modèle mathématique du modulateur à un modèle $ARMA$ (sans entrée exogène) et par conséquent de diminuer la quantité de calculs nécessaire pour l'identification. D'autre part, elle facilite l'implantation en reliant l'entrée du modulateur à la masse au lieu de générer un signal spécial pour l'identification.

Après le choix du modèle mathématique, l'algorithme d'identification qui permet de déterminer ses paramètres doit être développé. Nous avons étudié les différents algorithmes de calcul

adaptés au modèle ARMAX. Une présentation détaillée des différents algorithmes ainsi que des résultats de simulation concernant l'identification de la NTF d'un modulateur d'ordre 6 se trouve en annexe C.

D'après les résultats de simulation sur l'identification de la NTF d'un modulateur d'ordre 6, nous pouvons conclure que :

- Les algorithmes de calcul en temps différé (*Off Line*) nécessitent le stockage de 15000 points de mesure pour l'identification de la NTF de chaque modulateur sans compter les ressources matérielles nécessaires pour leur implantation (inversion et multiplication matricielle),
- Les algorithmes en temps réel (*On Line*) permettent de réduire la complexité matérielle en calculant les paramètres du modèle d'une façon itérative à la récolte de chaque point de mesure. Cependant, la convergence de ces algorithmes devient difficile voire impossible lorsque le nombre de paramètres devient grand et que les paramètres recherchés ont des valeurs très proches. Ce qui est le cas du modulateur $\Sigma\Delta$ d'ordre 6 dont les fréquences centrales des résonateurs sont très proches (les zéros de la NTF).

4.5 Détermination des bandes de fonctionnement à partir de la puissance de bruit

La détermination de la largeur de bande de fonctionnement de chaque modulateur de l'architecture EFBD en passant par l'identification de la $NTF^k(z)$ est difficile. En effet, les méthodes d'identification proposées en annexe C permettent de déterminer les zéros de la $NTF^k(z)$ avec des algorithmes en temps différé « *Off Line* » à partir de 15 000 points de mesure. La difficulté majeure de cette méthode réside dans la complexité de l'implantation numérique (multiplication et inversion matricielle). Par ailleurs, elle exige des ressources mémoire importantes. Les méthodes « *On Line* » présentées au § C.3.2 posent des problèmes de convergence en raison du nombre de paramètres à estimer. Dans notre cas, avec des modulateurs $\Sigma\Delta$ passe-bande d'ordre 6, la convergence vers les vraies valeurs des paramètres (12 paramètres) n'est pas assurée. On note aussi que, même si la convergence est assurée, cette méthode exige l'implémentation d'un algorithme numérique pour déterminer les fréquences d'intersection entre les NTF adjacentes (équation (4.3)) et déterminer ensuite les limites entre les bandes de fonctionnement f_r^k.

Compte tenu des difficultés de mise en œuvre des méthodes d'identification, nous avons reconsidéré l'objectif de base qui est de déterminer les bandes de fonctionnement qui minimisent la puissance de bruit résiduelle en sortie. Dans le but de définir la méthode que nous avons proposée pour la détermination des bandes de fonctionnement des modulateurs, nous considérons le cas d'une architecture *EFBD* avec 10 modulateurs. Dans cette architecture, les fréquences centrales des résonateurs sont décalées de $+20\% \times \frac{B}{N}$ (dispersion globale). De plus, nous prenons en compte une erreur aléatoire gaussienne d'écart type $\sigma = 10\% \times \frac{B}{N}$ (dispersion locale) sur les fréquences centrales des résonateurs (figure 4.10).

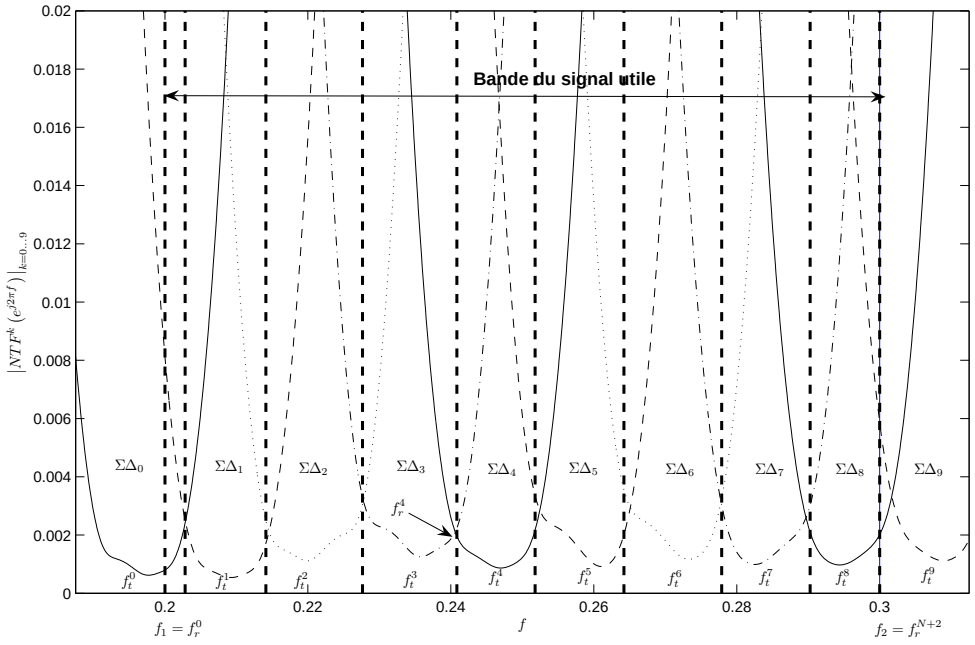

Fig. 4.10 – Le module de la NTF des 10 modulateurs.

La bande de fonctionnement de chaque modulateur $\left[\mathrm{f_r^k},\ \mathrm{f_r^{k+1}}\right]$ (voir § 4.2.1) est la zone limitée par les deux lignes verticales en pointillés représentant ses bornes $\mathrm{f_r^k}$ et $\mathrm{f_r^{k+1}}$ (figure 4.10). Dans ce cas particulier, le point d'intersection entre les NTF des modulateurs 8 et 9 se trouve en dehors de la bande du signal utile. Par conséquent, le dernier modulateur $\Sigma\Delta_9$ ne sera pas utilisé car le minimum de sa NTF se trouve en dehors de la bande du signal utile définie dans l'intervalle $[\mathrm{f_1},\ \mathrm{f_2}]$.

4.5.1 Calcul de la puissance de bruit

La puissance de bruit totale en sortie de l'architecture EFBD passe-bande est donnée par :

$$P_{bruit\_total} = \sum_{k=0}^{N+1} P_{NTF^k}\ , \tag{4.7}$$

où $\mathrm{P_{NTF^k}}$ est la puissance de bruit de chaque modulateur. Elle est calculée, conformément au traitement présenté sur la figure 3.26, par la formule suivante :

$$\begin{aligned} P_{NTF^k} &= \int_{-\frac{1}{2}}^{+\frac{1}{2}} \left|NTF^k\left(e^{j2\pi R_d f}\right)\right|^2 \left|G_{pb}^k\left(e^{j2\pi R_d\left(f-f_c^k\right)}\right)\right|^2 \left|C_1^k\left(e^{j2\pi R_d\left(f-f_c^k\right)}\right)\right|^2 \Gamma(f)df \qquad (4.8)\\ &\approx \int_{-\frac{1}{2}}^{+\frac{1}{2}} \left|NTF^k\left(e^{j2\pi R_d f}\right)\right|^2 \left|G_{pb}^k\left(e^{j2\pi R_d\left(f-f_c^k\right)}\right)\right|^2 \Gamma(f)df \end{aligned}$$

On note que :

- $C_1^k(z)$ est un filtre de correction à 3 coefficients du module de la STF dont l'influence sur la puissance de bruit en sortie est très faible. Il a donc été omis dans l'expression de P_{NTF^k} (formule 4.8),
- $\Gamma(f)$ est la densité spectrale du bruit de quantification et elle est constante si les conditions de *Bennett* sont respectées [28], c'est-à-dire :

$$\Gamma(f) = \Gamma = \frac{1}{3 \times 4^{N_b}} \tag{4.9}$$

Afin de simplifier la compréhension du principe de l'algorithme que nous allons proposer, nous présentons le produit $\left|NTF^k\left(z^{R_d}\right)\right|^2 \left|G_{pb}^k\left(z^{R_d}\right)\right|^2$ pour $k = 4$ sur la figure 4.11. Nous constatons que la majorité de la puissance de bruit se trouve dans la bande de fonctionnement $\left[f_r^k,\, f_r^{k+1}\right] \times R_d$. En effet, le filtre passe-bas $G_{pb}^k(z)$ a un gain presque constant dans la bande de fonctionnement et atténue suffisamment le bruit de quantification modulé par NTF^k en dehors de cette bande. De ce fait, on peut omettre également le terme $\left|G_{pb}^k\left(z^{R_d}\right)\right|^2$ dans l'expression (4.8) de P_{NTF^k}. Dans ce cas, la puissance de bruit totale en sortie s'exprime par :

$$P_{bruit\_totale} = \sum_{k=0}^{N+1} P_{NTF^k} = \Gamma \sum_{k=0}^{N+1} \int_{f_r^k}^{f_r^{k+1}} \left|NTF^k\left(e^{j2\pi f}\right)\right|^2 df \tag{4.10}$$

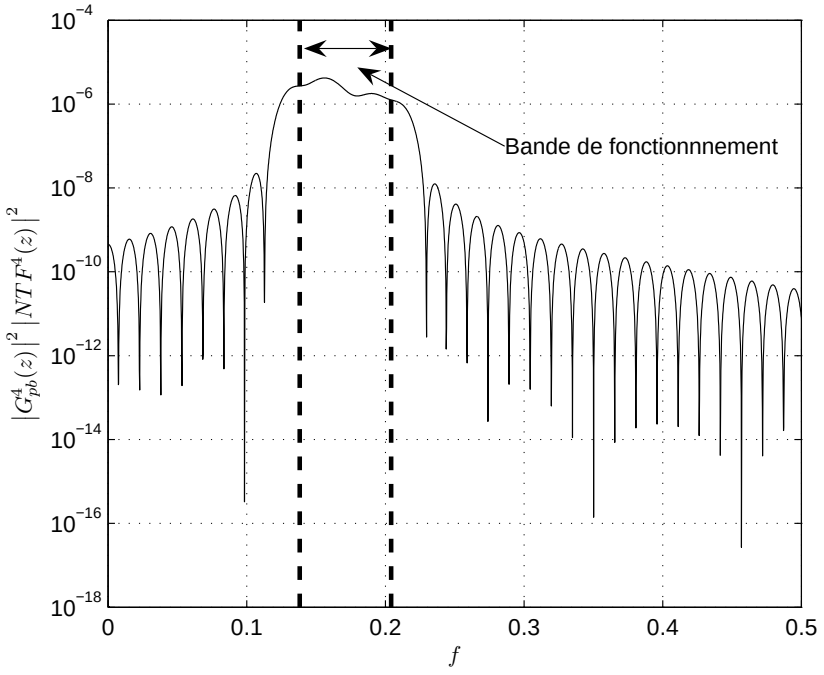

Fig. 4.11 – Produit du module au carré de la réponse fréquentielle du filtre passe-bas avec celui de la réponse fréquentielle de la NTF.

Soient $\left\{f_t^k\right\}_{k=0\ldots N+1}$ des fréquences constantes choisies de façon à ce que, pour chaque modulateur k, f_t^k soit égale à la fréquence de résonance théorique f_{cr_2} du deuxième résonateur, centré au milieu

de la bande. En introduisant la fréquence constante f_t^k dans l'expression (4.10) de la puissance totale de bruit, l'intégrale

$$\int_{f_r^k}^{f_r^{k+1}} \left|NTF^k\left(e^{j2\pi f}\right)\right|^2 df$$

peut se décomposer en une somme de deux intégrales

$$\int_{f_r^k}^{f_t^k} \left|NTF^k\left(e^{j2\pi f}\right)\right|^2 df + \int_{f_t^k}^{f_r^{k+1}} \left|NTF^k\left(e^{j2\pi f}\right)\right|^2 df$$

dont chacune a une borne constante f_t^k. Ensuite, en faisant la somme de toutes les intégrales, la puissance totale de bruit peut être ré-écrite sous la forme donnée par :

$$\begin{aligned}
P_{bruit_totale} &= \sum_{k=0}^{N+1} P_{NTF^k} = \Gamma \sum_{k=0}^{N+1} \int_{f_r^k}^{f_r^{k+1}} \left|NTF^k\left(e^{j2\pi f}\right)\right|^2 df \\
&= \Gamma \int_{f_r^0=f_1}^{f_t^0} \left|NTF^0\left(e^{j2\pi f}\right)\right|^2 df \\
&+ \sum_{k=1}^{N+1} \Gamma \underbrace{\left(\int_{f_t^{k-1}}^{f_r^k} \left|NTF^{k-1}\left(e^{j2\pi f}\right)\right|^2 df + \int_{f_r^k}^{f_t^k} \left|NTF^k\left(e^{j2\pi f}\right)\right|^2 df \right)}_{T^k\left(f_r^k\right)} \\
&+ \int_{f_t^{N+1}}^{f_r^{N+2}=f_2} \Gamma \left|NTF^{N+1}\left(e^{j2\pi f}\right)\right|^2 df
\end{aligned} \tag{4.11}$$

Nous constatons dans la formule (4.11) que :

- le premier et le dernier terme sont constants,
- le deuxième terme est la somme de $N+1$ termes dont chaque terme $T^k\left(f_r^k\right)$ dépend de la fréquence limite f_r^k entre les bandes de fonctionnement des modulateurs $k-1$ et k.

Pour montrer le choix de f_r^k dans la minimisation de la puissance de bruit totale, nous avons représenté le terme $T^4\left(f_r^4\right)$ en faisant varier f_r^4 entre f_t^3 et f_t^4 (figure 4.12).

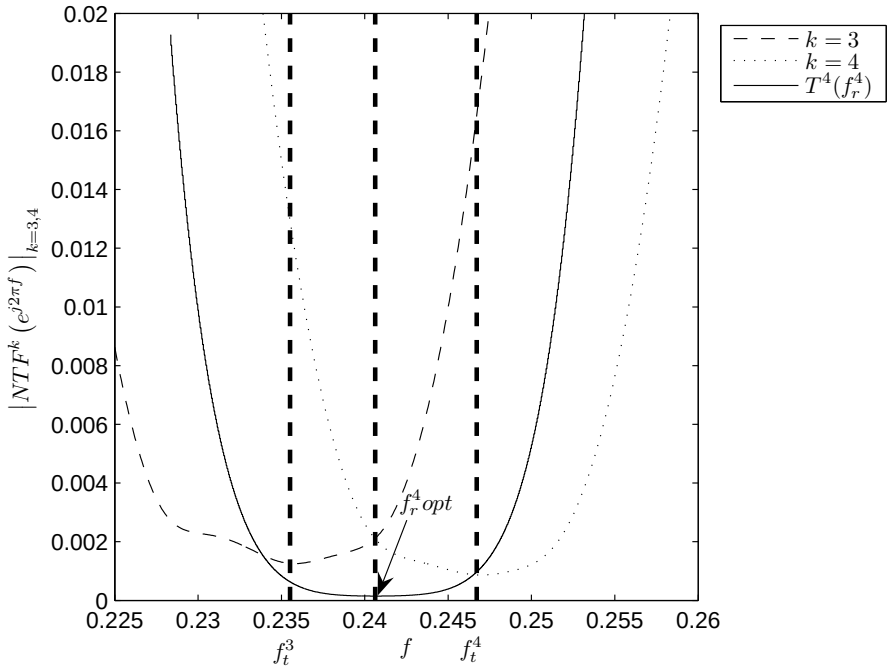

Fig. 4.12 – Critère d'optimisation $\mathrm{T}^4\left(\mathrm{f}_\mathrm{r}^4\right)$.

L'allure de $\mathrm{T}^4\left(\mathrm{f}_\mathrm{r}^4\right)$ est convexe. Son minimum est atteint à la fréquence d'intersection entre NTF^3 et NTF^4. Cette fréquence d'intersection est la valeur optimale de f_r^4, elle est notée dans l'algorithme d'adaptation par $\mathrm{f}_\mathrm{r}^4\mathrm{opt}$. Nous avons vérifié également que l'allure de $\mathrm{T}^4\left(\mathrm{f}_\mathrm{r}^4\right)$ ne change pas si on tient compte de la réponse fréquentielle du filtre passe-bas $\mathrm{G}_{\mathrm{pb}}^{\mathrm{k}}(\mathrm{z})$.

NB : Le module de la fonction $\mathrm{T}^4\left(\mathrm{f}_\mathrm{r}^4\right)$ *sur la figure 4.12 a été multiplié par* 10^3 *pour la mettre à l'échelle par rapport au tracé des modules des NTF et montrer que son minimum est atteint au point d'intersection entre les deux* NTF.

La puissance de bruit totale exprimée par la formule (4.11) est composée de deux termes constant et d'un terme dépendant des fréquences limites $\mathrm{f}_\mathrm{r}^\mathrm{k}$. La recherche des valeurs optimales de ces fréquences minimisant la puissance de bruit totale peut se faire pour chaque fréquence $\mathrm{f}_\mathrm{r}^\mathrm{k}$ séparément. En effet, pour une fréquence $\mathrm{f}_\mathrm{r}^\mathrm{k}$ donnée, la variation de la puissance de bruit en fonction de $\mathrm{f}_\mathrm{r}^\mathrm{k}$ présente la même allure que $\mathrm{T}^\mathrm{k}\left(\mathrm{f}_\mathrm{r}^\mathrm{k}\right)$ sachant que les autres fréquences $\left\{\mathrm{f}_\mathrm{r}^\mathrm{k}\right\}$ sont maintenues fixes (équation (4.11)). Cette constatation constitue l'idée sur laquelle se base l'algorithme d'adaptation du traitement numérique aux imperfections des composants analogiques dont la description complète sera détaillée au paragraphe suivant.

4.5.2 Algorithme d'adaptation du traitement numérique

Le but de l'algorithme d'adaptation est de déterminer les bandes de fonctionnement $\mathrm{L}_\mathrm{B}^\mathrm{k} = 2 \times \Delta \mathrm{f_k}|_{\mathrm{k=0...N+1}}$ de façon à minimiser la puissance de bruit de quantification résiduelle en sortie de l'architecture EFBD. Nous avons vu ci-dessus que la puissance de bruit dépend des fréquences limites $\mathrm{f}_\mathrm{r}^\mathrm{k}$ entre les bandes de fonctionnement et que cette puissance est minimale quand les fréquences $\mathrm{f}_\mathrm{r}^\mathrm{k}$ sont égales à la fréquence d'intersection entre les fonctions de mise en forme de bruit (NTF) des modulateurs adjacents.

Dans la procédure d'exploitation de l'algorithme, nous avons tenu compte du fait que :

- L'algorithme se base sur le calcul de la puissance de bruit totale P_{bruit_totale} en sortie. De ce fait, l'entrée de l'architecture EFBD est reliée à la masse pour n'avoir que ce bruit en sortie. Un générateur de bruit monobit (dither) est toujours présent à l'entrée du CAN dans la boucle du modulateur afin de régler le problème des cycles limites (voir annexe C, § C.5.1).
- L'erreur maximale sur les fréquences centrales des résonateurs ne doit pas excéder $\left|\frac{B}{N}\right|$ afin de garantir une bonne performance de l'architecture EFBD.
- Les fréquences limites f_r^k peuvent être quantifiées avec un pas $q_f = 10^{-3} f_e$ (voir § 4.3) sans dégradation notable de la performance. Avec cette quantification, la bande du signal utile de l'exemple utilisé au cours de cette thèse (voir tableau 3.1) $B = [0.2,\ 0.3] \times F_e$ représente 102 pas de quantification ($\lceil 0.1/q_f \rceil = 102$) et la largeur de chaque sous-bande, dans le cas idéal, vaut 12 ou 13 pas de quantification ($\left\lceil \frac{0.1}{N \times q_f} \right\rceil$). Étant donné que l'erreur maximale sur les fréquences centrales ne doit pas excéder $\left|\frac{B}{N}\right|$, la largeur de bande de fonctionnement définie par $L_B^k = 2 \times \Delta f_k$ varie entre 0 et $25 q_f$. Ceci permet d'avoir des filtres FIR précalculés stockés dans une mémoire ROM (figure 4.13, ROM_3).

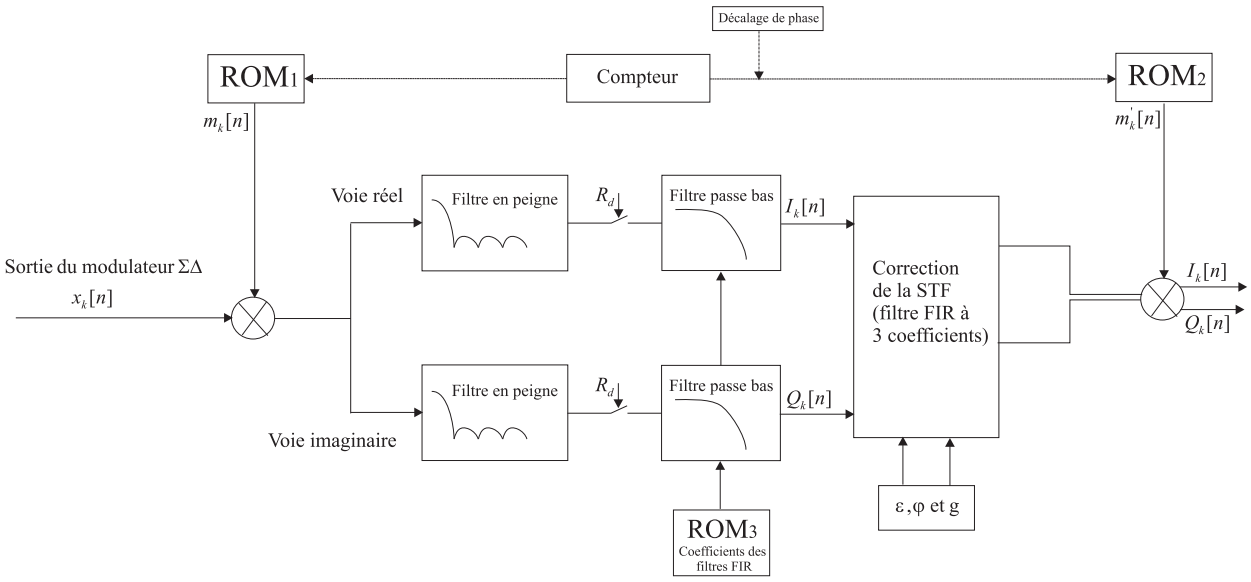

Fig. 4.13 – Traitement complexe derrière chaque modulateur.

L'algorithme d'adaptation consiste à :

1. mettre l'entrée de l'architecture $EFBD$ à la masse (figure 4.14),
2. initialiser les valeurs de $\left\{f_c^k, \Delta f_k\right\}_{k=0\ldots N+1}$ à leurs valeurs théoriques,
3. faire varier chaque fréquence $f_r^k\big|_{k=1\ldots N+1}$ entre f_r^{k-1} et f_r^{k+1} en posant :

$$\begin{cases} f_r^0 = f_1 \\ f_r^{N+2} = f_2 \\ f_r^k \leq f_r^{k+1} \end{cases} \tag{4.12}$$

4. adapter le traitement numérique en calculant :

– les séquences de modulation et de démodulation à partir de la fréquence f_c^k exprimée par :

$$f_c^k = \left.\frac{f_r^{k-1} + f_r^k}{2}\right|_{k=1...N+1} \tag{4.13}$$

– les largeurs de bande des filtres passe-bas définies par :

$$\Delta f_k = \left.\frac{f_r^k - f_r^{k-1}}{2}\right|_{k=1...N+1} \tag{4.14}$$

5. estimer la puissance de bruit en sortie à partir de :

$$\hat{P}_{bruit}(i) = \frac{1}{N_s}\sum_{n=0}^{N_s}\left[\sum_{k=0}^{N+1}\left(I_k\,[n]^2 + Q_k\,[n]^2\right)\right] \tag{4.15}$$

où N_s est le nombre d'échantillons permettant d'estimer la puissance. I_k et Q_k sont les sorties en phase et en quadrature de phase à la sortie de chaque k-ième étage de traitement (figure 4.13). Le calcul de $\hat{P}_{bruit}$ se fait d'une façon itérative à la récolte de chaque échantillon et n'exige pas d'importantes resources matérielles.

6. après plusieurs itérations, choisir la fréquence f_r^k pour laquelle la puissance $\hat{P}_{bruit}$ est minimale.

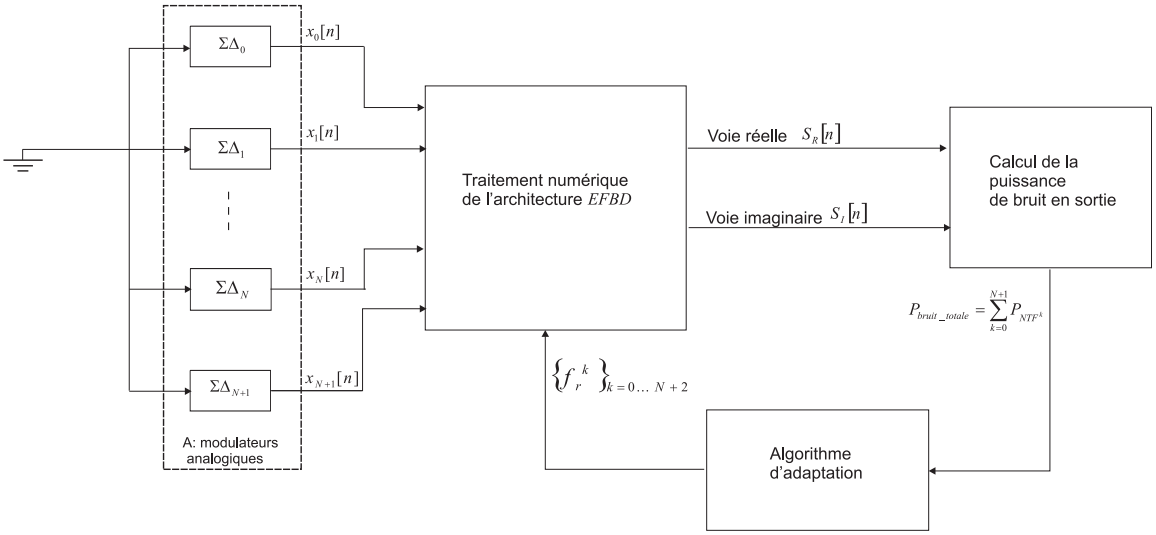

Fig. 4.14 – Schéma fonctionnel de l'algorithme d'adaptation numérique.

L'organigramme de cet algorithme est représenté par la figure 4.15.

On appellera séquence l'ensemble des puissances de bruit mesurées après la calibration de chacune des fréquences $f_r^k|_{k=1...N+1}$. Elle est de longueur N + 1 (9 dans notre exemple). L'exécution de l'algorithme (figure 4.15) permet de générer une séquence.

La puissance de bruit en sortie est estimée à partir d'un nombre fini d'échantillon N_s. Ceci introduit une erreur dans l'estimation de la vraie valeur de la puissance de bruit et influence la vitesse de convergence de l'algorithme vers les valeurs optimales f_r^kopt.

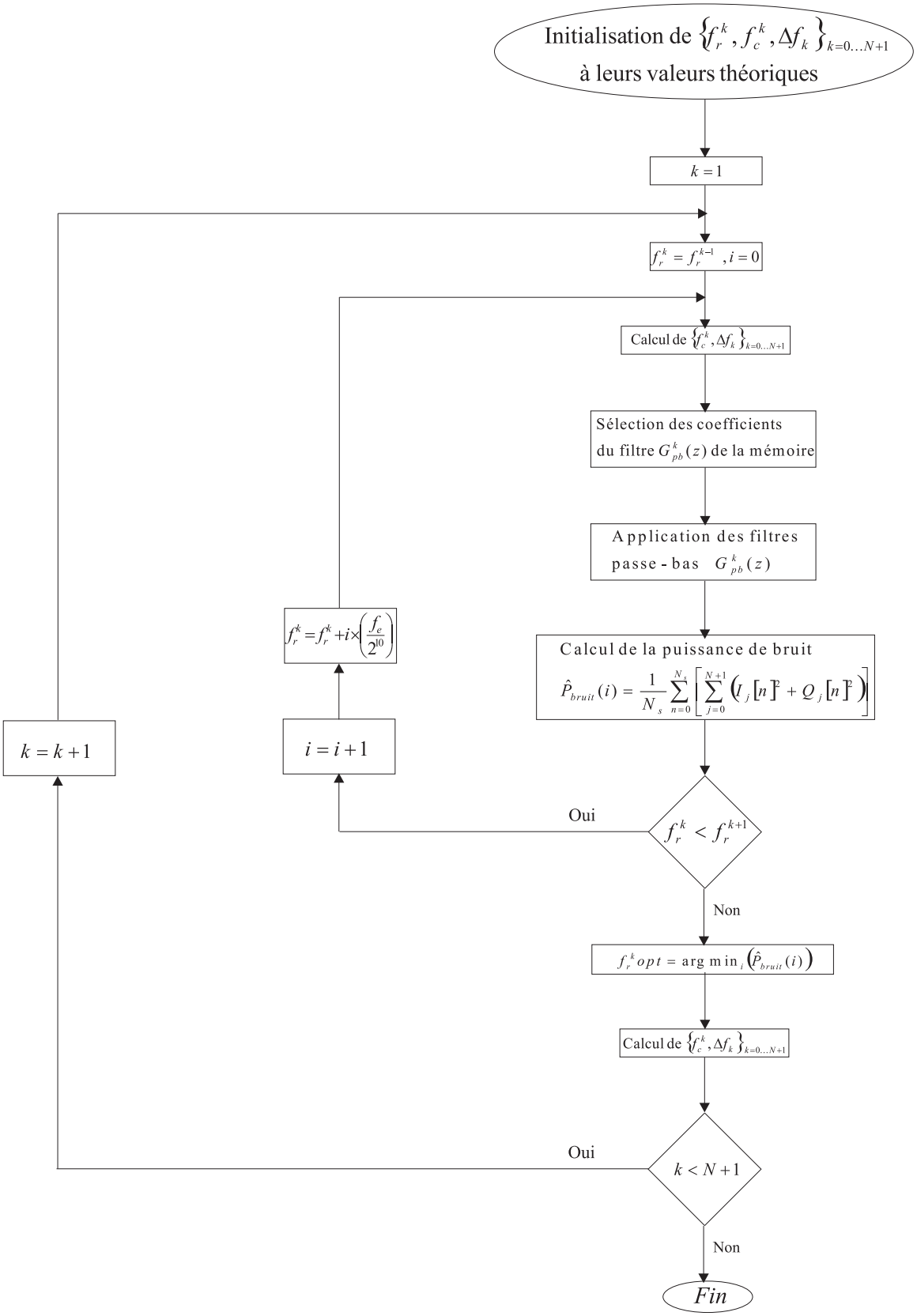

Fig. 4.15 – Organigramme de l'algorithme d'adaptation numérique basé sur la minimisation de la puissance de bruit.

Nous avons appliqué l'algorithme d'adaptation à l'architecture EFBD (figure 4.3) composé de 10 modulateurs sur lesquels nous avons imposé un décalage des fréquences centrales des résonateurs de $\left(+\frac{B}{2N}\right)$. Comme pour les autres méthodes, l'algorithme peut s'appliquer « *Off Line* » ou « *On Line* » :

Calcul en temps différé « *Off Line* »

les signaux en sortie $x_k[n]$ des différents modulateurs (figure 4.14) de N_s échantillons sont stockés en mémoire. Ensuite, ces signaux en mémoire seront utilisés à chaque itération de l'algorithme d'adaptation. La figure 4.16 présente trois séquences de l'ENOB (calculées à partir de la puissance estimée $\hat{P}_{bruit}$) de l'algorithme d'adaptation (l'algorithme a été exécuté 3 fois successivement) avec un nombre d'échantillons de 2^{12}.

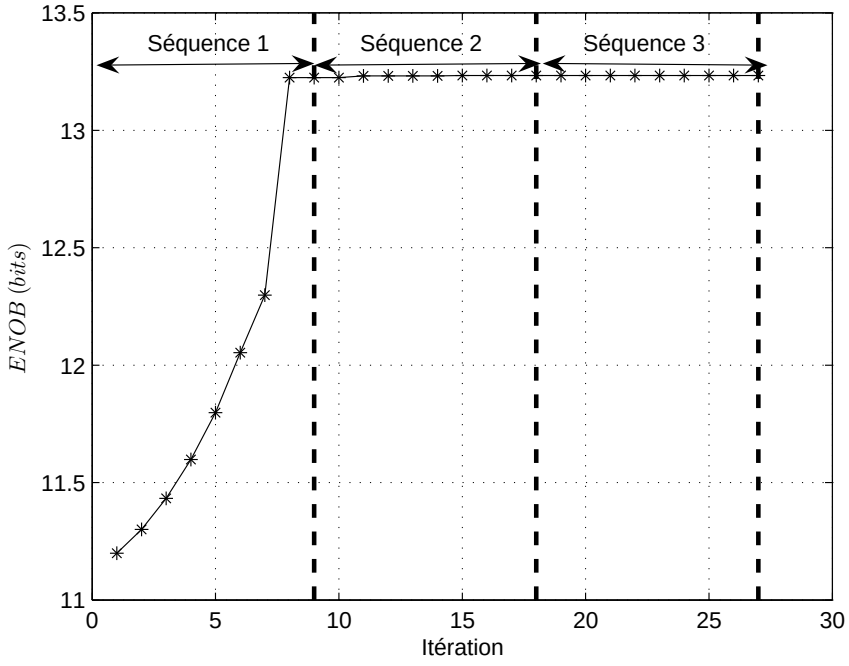

Fig. 4.16 – Évolution de l'ENOB pour l'algorithme d'adaptation avec $N_s = 2^{12}$ en temps différé.

Nous remarquons que la convergence est atteinte à la fin de la première séquence. La convergence vers les vraies valeurs f_r^k est déduite de la convergence de la résolution ENOB vers sa valeur théorique qui est égale dans notre exemple à 13.3 bits.

Calcul en temps réel « *On Line* »

Afin d'éviter l'utilisation d'une mémoire pour le stockage de l'ensemble des valeurs $x_k[n]$, pour chaque itération de l'algorithme, l'estimation de la puissance de bruit se fait de façon itérative par la formule (4.15) à la récolte de chaque échantillon $x_k[n]$ sur un nombre d'échantillons N_s. Les figures 4.17 et 4.18 présentent l'évolution de la résolution (ENOB) en fonction du nombre d'itérations de l'algorithme et ceci pour une estimation de la puissance $\hat{P}_{bruit}$ respectivement sur 2^{12} et 2^{14} échantillons.

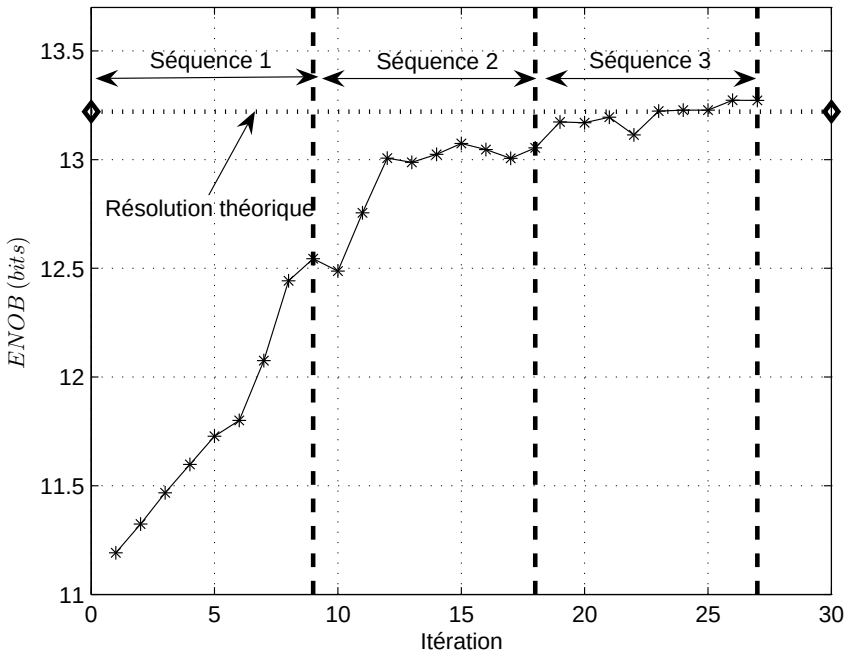

Fig. 4.17 – Évolution de l'ENOB pour l'algorithme d'adaptation avec $N_s = 2^{12}$ en temps réel.

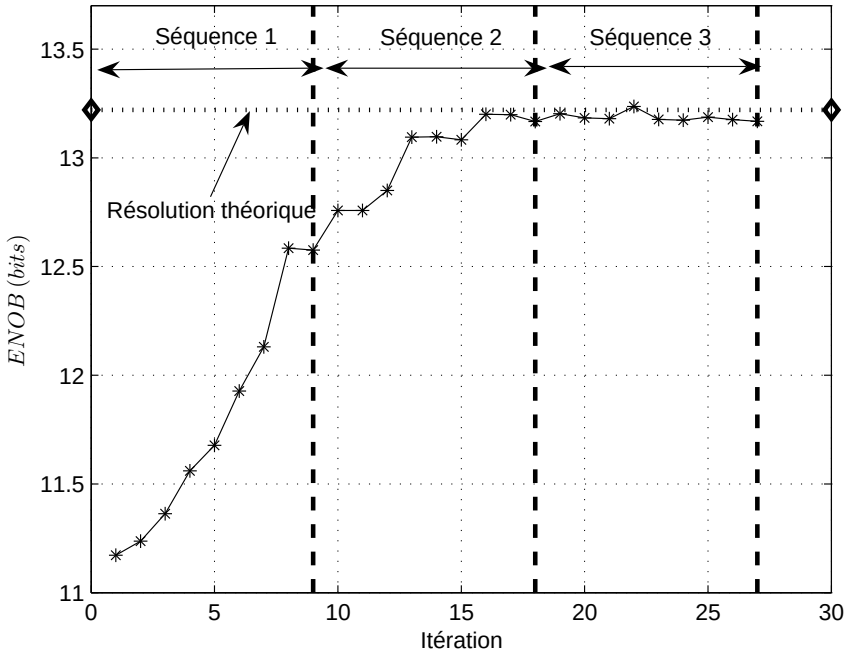

Fig. 4.18 – Évolution de l'$ENOB$ pour l'algorithme d'adaptation avec $N_s = 2^{14}$ en temps réel.

Nous remarquons que la convergence est assurée à la fin de la troisième séquence même avec l'introduction d'une erreur aléatoire sur les fréquences centrales comme celle présentée sur la

figure 4.10. La fréquence d'échantillonnage dans notre exemple est égale à 800 MHz et le rapport de décimation R_d est égal à 5. Avec $N_s = 2^{12}$, le temps de calcul nécessaire pour estimer la puissance de bruit pour chaque valeur de f_r^k est égal à N_s/F_e soit 5 μs. Le pire-cas par rapport au temps de calcul est celui où les fréquences centrales des résonateurs ont été déplacées d'une largeur $\left(\frac{B}{N}\right)$ de leur valeur théorique. Dans ce cas, la fréquence f_r^k doit balayer 25 valeurs entre f_r^{k-1} et f_r^{k+1}. Le temps nécessaire pour optimiser la fréquence f_r^k dans ce pire-cas est égal à $5 \times 25\ \mu s = 125\ \mu s$. Avec $N + 1$ fréquences limites, le temps nécessaire pour exécuter l'algorithme et réaliser ainsi une séquence est égal à $(N + 1) \times 125\ \mu s = 1.12$ ms. Le temps nécessaire pour adapter le traitement avec cet algorithme au bout de trois séquences est estimé alors à 3.3 ms.

4.6 Calibration de la STF des modulateurs ΣΔ

Nous avons vu au § 3.4 que la fonction de transfert par rapport au signal $STF^k(f)$ d'un modulateur ΣΔ passe-bande à temps continu n'est pas un simple retard comme dans le cas discret. Cette fonction de transfert réalise l'interface entre deux domaines : un signal en sortie dans le domaine discret et un signal en entrée dans le domaine continu. Elle s'exprime dans le domaine fréquentiel [40, 41] (voir annexe A § A.4) par :

$$STF(2\pi f) = \frac{S(e^{j2\pi f})}{X(j2\pi f)} = \frac{G(j2\pi f)}{1 + F(e^{j2\pi f})} \tag{4.16}$$

Cette fonction n'est pas constante dans la bande de fonctionnement du modulateur. Elle a localement une forme parabolique. Nous avons vu au chapitre 3 que la correction de ce défaut peut se faire par un filtre de correction $C_1^k(z)$ à trois coefficients. L'expression dans le domaine z de ce filtre est donnée par l'équation (3.28).

$$C_1^k(z) = g\left(-\varepsilon e^{-j2\pi\nu} + (1 + 2\varepsilon)\, z^{-1} - \varepsilon e^{j2\pi\nu} z^{-2}\right)$$

avec :

- ε : la concavité du filtre. Elle est de signe opposé à la concavité de $STF^k(f)$,
- ν : la différence entre la fréquence du milieu de la bande de fonctionnement f_c^k et la fréquence pour laquelle le module de la $STF^k(f)$ atteint son maximum,
- g : le gain du filtre. Il est égal à l'inverse du maximum du module de $STF^k(f)$ afin d'obtenir un gain égal à 1 dans la bande de fonctionnement.

Les largeurs de bande de fonctionnement des différents modulateurs varient en fonction des imperfections analogiques. Ces largeurs de bande sont déterminées par l'algorithme présenté au paragraphe 4.5 de façon à minimiser la puissance de bruit sur la bande utile pour conserver les performances initiales de l'architecture EFBD. Or les coefficients $\{\varepsilon, \nu, g\}$ du filtre dépendent de la largeur de bande de fonctionnement de chaque modulateur. Il est donc indispensable d'avoir un algorithme qui permette de calculer les coefficients $\{\varepsilon, \nu, g\}$ du filtre $C_1^k(z)$ après la détermination de la bande de fonctionnement de chaque modulateur.

4.6.1 Simplification du filtre de correction $C_1^k(z)$

Le filtre de correction $C_1^k(z)$ est complexe à implanter à cause de ses coefficients en $\cos(2\pi\nu)$ et $\sin(2\pi\nu)$ sur les deux voies I_k et Q_k du traitement complexe (figure 4.13). Afin de simplifier ce filtre, on a effectué une approximation au premier ordre de l'exponentielle dans l'expression de

$C_1^k(z)$ ($e^x \approx x$ pour x petit). Cette approximation est valable pour une différence en fréquence ν petite. Dans ce cas, la fonction de transfert du filtre $C_1^k(z)$ peut s'approcher à l'ordre 1 par :

$$C_1^k(z) \approx g\left(-\varepsilon\left(1 - j2\pi\nu\right) + \left(1 + 2\varepsilon\right) z^{-1} - \varepsilon\left(1 + j2\pi\nu\right) z^{-2}\right) \tag{4.17}$$

La condition sur ν (valeur petite) est respectée en supposant que le maximum de la $STF^k(f)$ se trouve dans la bande de fonctionnement du modulateur.

Après avoir simplifié la réalisation de $C_1^k(z)$, il reste à trouver une méthode numérique qui permette de déterminer les valeurs de l'ensemble des paramètres $\{\varepsilon, \nu, g\}$ après avoir déterminé les bandes de fonctionnement par l'algorithme décrit au paragraphe 4.5.2.

4.6.2 Détermination des paramètres du filtre de correction $C_1^k(z)$

Les paramètres ε et ν du filtre de correction $C_1^k(z)$ doivent permettre de corriger la $STF^k(f)$ afin que celle-ci présente une ondulation minimale dans la bande de fonctionnement.

Dans ce but, les paramètres $\{\varepsilon, \nu\}$ sont initialisés aux valeurs théoriques calculées à partir du modèle mathématique (4.16) de la STF pour un modulateur idéal (les fréquences centrales des résonateurs sont égales à leurs valeurs théoriques). Le gain g est fixé à 1. Les fonctions G(jw) et $F(e^{jw})$ dans (4.16) dépendent de l'architecture de réalisation du modulateur. Dans le cas présent, nous considérons, à titre d'exemple, une architecture avec des résonateurs en série dans laquelle le signal d'entrée est injecté seulement à l'entrée du premier résonateur.

Pour déterminer les valeurs de $\{\varepsilon, \nu\}$ pour lesquelles l'ondulation est minimale, nous définissons les intervalles $\Delta\varepsilon$ et $\Delta\nu$, découpés en N_v sous-intervalles, et qui constituent le pas de calcul de notre algorithme de calibration. La largeur des intervalles $\Delta\varepsilon$ et $\Delta\nu$ dépend du choix de l'architecture de réalisation des modulateurs. Nous avons choisi, dans le cadre de nos simulations, pour ν l'intervalle $\Delta\nu = \left[-\frac{4}{1000},\ \frac{4}{1000}\right]$ et pour $\Delta\varepsilon = [-0.1,\ 0.1]$. Ces valeurs ont été définies pour une architecture temps continu avec des résonateurs en série. On note que :

- pour de faibles bandes de fonctionnement, l'approximation (4.17) de $C_1^k(z)$ reste valable mais il faut augmenter l'intervalle de recherche de la fréquence ν afin d'avoir une correction de la $STF^k(f)$ dans cette bande.
- le nombre de valeurs N_v influe sur la finesse du critère de recherche des valeurs ν et ε et ainsi sur leurs pas de quantification.

En faisant varier ν et ε indépendamment dans les intervalles $\Delta\nu$ et $\Delta\varepsilon$ autour de leurs valeurs théoriques avec $N_v = 64$, nous avons calculé l'ondulation (différence entre la valeur minimale et la valeur maximale de l'amplitude du spectre) qu'il peut y avoir dans la bande de fonctionnement de chaque modulateur entre f_r^k et f_r^{k+1} à partir de la réponse fréquentielle de la $STF^k(f)$ et du filtre $C_1^k(z)$. Le résultat obtenu est représenté en trois dimensions par la figure 4.19.

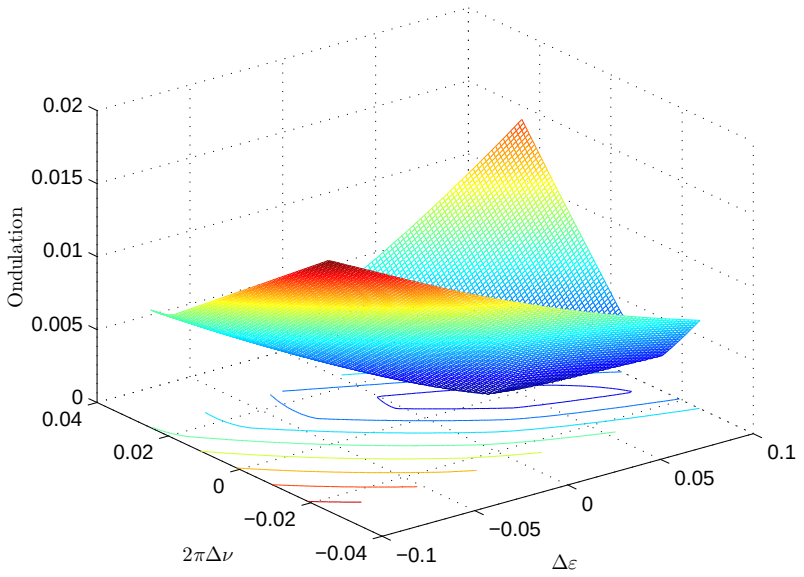

Fig. 4.19 – Ondulation dans la bande en fonction de $\Delta\nu$ et $\Delta\varepsilon$.

Nous remarquons que l'ondulation mesurée présente un seul minimum et par conséquent le critère de la recherche de ν et ε minimisant l'ondulation dans la bande de fonctionnement est un critère convexe. La figure 4.20 présente les lignes de niveau de ce critère.

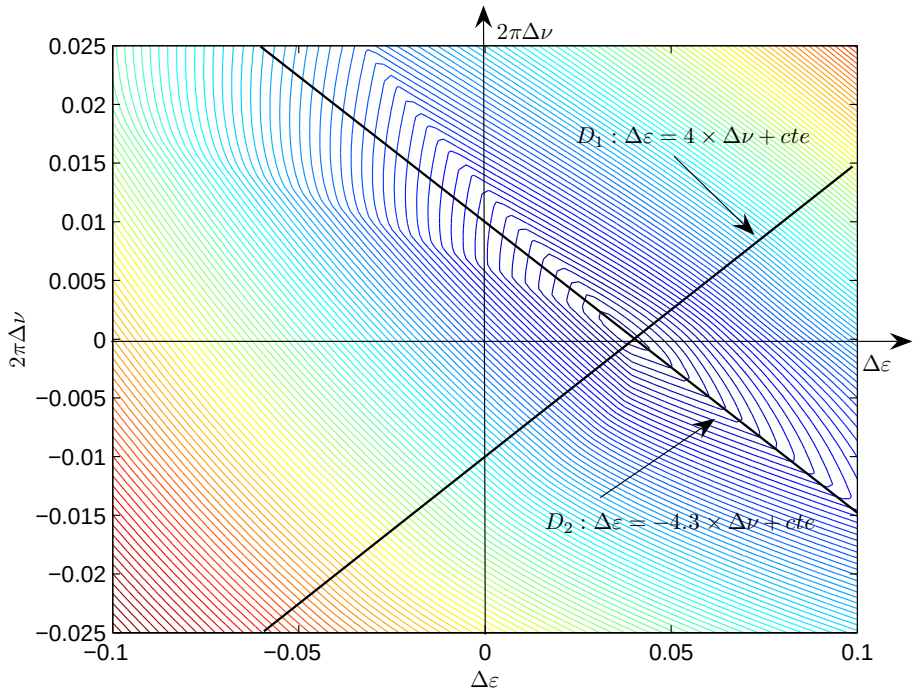

Fig. 4.20 – Lignes de niveau du critère de minimisation de l'ondulation en fonction de $(\Delta\varepsilon, \Delta\nu)$).

La convergence vers le minimum peut-être accélérée en définissant deux directions de recherche

au lieu de faire varier ν et ε indépendamment chacune sur N_v valeurs. Ces deux directions sont définies par les pentes λ_1 et λ_2 des droites $D_1 : \Delta\varepsilon = \lambda_1 \times (2\pi\Delta\nu)$ et $D_2 : \Delta\varepsilon = \lambda_2 \times (2\pi\Delta\nu)$ (voir figure 4.20). Quelles que soient les valeurs de λ_1 et de λ_2, l'algorithme va converger en raison de la nature convexe du critère à minimiser (ondulation $= f(\varepsilon, \nu)$ figure 4.19). Il existe cependant des valeurs optimales de λ_1 et de λ_2 pour lesquelles l'algorithme va converger le plus rapidement. Elles dépendent de la largeur de bande de fonctionnement. En pratique, nous n'avons pas une connaissance exacte des largeurs de bande. Pour donner un encadrement pour les valeurs de λ_1 et λ_2, nous avons mené une simulation paramétrique où nous avons fait varier la largeur de bande de fonctionnement pour un modulateur. Les modules de λ_1 et λ_2 se trouvent dans l'intervalle $[3 \dots 5]$. Nous avons choisi $\lambda_1 = 4$ et $\lambda_2 = -4$ pour tous les modulateurs. Avec ces valeurs, nous avons vérifié que la convergence vers le minimum, pour tous les modulateurs, est assurée au bout de deux ou trois itérations. Chaque itération comporte la recherche du minimum suivant la direction λ_1 puis suivant la direction λ_2.

La figure 4.21 montre les lignes de niveaux obtenues en divisant chacun des intervalles $\Delta\varepsilon$ et $\Delta\nu$ sur $N_v = 8, 16, 32, 64$ valeurs. Nous remarquons que $N_v = 32$ est une valeur acceptable afin d'obtenir un critère fin et convexe en fonction de $\Delta\varepsilon$ et $\Delta\nu$. Nous verrons, dans la suite, l'influence du nombre de valeurs N_v sur l'erreur de quantification de ε et ν.

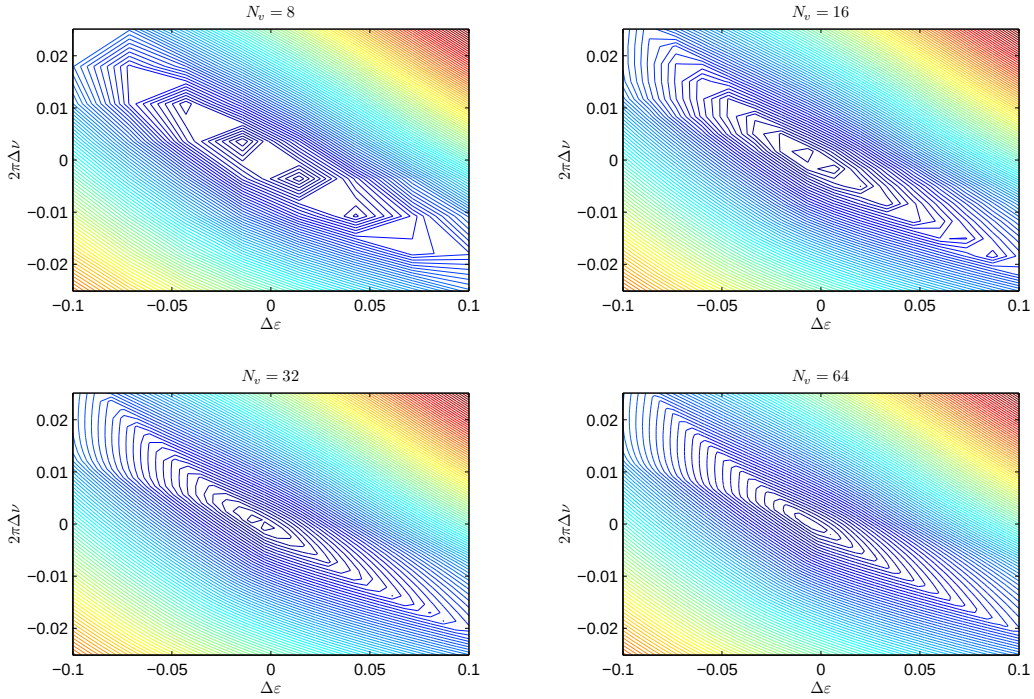

Fig. 4.21 – Lignes de niveaux du critère de minimisation de l'ondulation pour $N_v = 8, 16, 32, 64$.

Pour mettre en œuvre la méthode de recherche des valeurs optimales ε_{opt} et ν_{opt}, nous avons besoin de calculer l'ondulation dans la bande de fonctionnement introduite par la $SFT^k(f)$ de chaque modulateur. De ce fait, nous avons besoin d'un signal de référence qui a une densité spectrale constante dans la bande de fonctionnement $\left[f_r^k, f_r^{k+1}\right]$. Ce signal peut être un signal en sinus cardinal ou un signal de type chirp. Nous avons choisi de travailler avec un signal de type chirp (signal sinusoïdal dont la fréquence varie linéairement en fonction du temps) en raison

de la simplicité qu'il présente pour l'extraction de l'information dans le domaine fréquentiel à partir de sa représentation temporelle si sa fréquence varie lentement en fonction du temps. Son expression est donnée par :

$$x(t) = A \times \cos\left(2\pi\left(f_i + \beta t\right) t\right) \quad avec \quad \beta = \frac{f_f - f_i}{t_f} \tag{4.18}$$

f_i : fréquence initiale à $t = 0$,
β : vitesse de variation de la fréquence,
f_f : fréquence à l'instant $t = t_f$.
Comme le traitement numérique de l'architecture EFBD se fait dans le domaine complexe, il est judicieux d'exprimer le signal analytique (voir annexe B.3.2) correspondant à ce signal chirp afin de pouvoir interpréter le signal en sortie. Ce signal analytique s'exprime par :

$$x_a(t) = Ae^{j(2\pi(f_i+\beta t)t)} = \underbrace{Ae^{j\left(2\pi\beta t^2\right)}}_{x_b(t)} e^{j(2\pi f_i t)} \tag{4.19}$$

où $x_b(t)$ est l'enveloppe complexe du signal chirp.

Le module de l'enveloppe complexe du signal chirp est constant. Il est égal à A. L'amplitude du spectre d'un signal chirp linéaire est approchée, si la fréquence varie lentement en fonction du temps ($(f_f - f_i) \times t_f \gg 1$) [54] par :

$$PSD(f) = \begin{cases} A\sqrt{\frac{t_f}{4\left(f_f - f_i\right)}} = \text{cte} & \text{pour} \quad |f| \leq f_f \\ 0 \quad \text{ailleurs} \end{cases} \tag{4.20}$$

La figure 4.22 présente, à titre d'exemple, un signal chirp généré dans la bande [0.2, 0.3] sur 500 points. Elle présente également le module du signal analytique correspondant (a) et le module normalisé de sa transformée de Fourier (b).

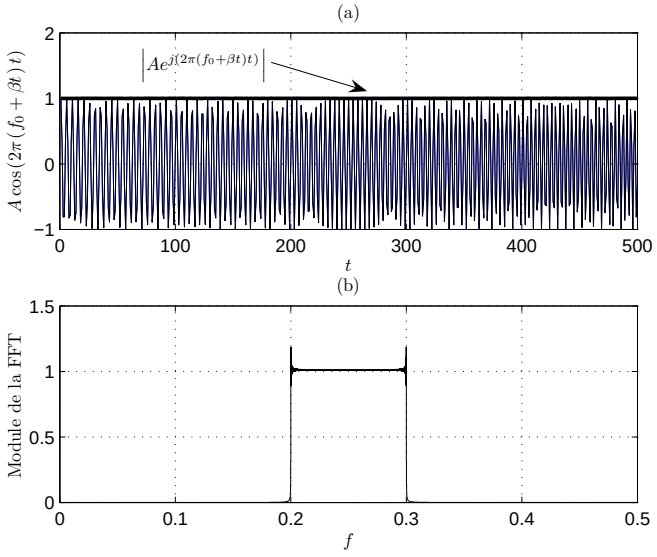

Fig. 4.22 – (a) Signal chirp de bande [0.2, 0.3] avec le module de son enveloppe complexe, (b) module normalisé de la transformée de Fourier du signal chirp.

L'avantage du signal chirp est qu'il permet, si sa fréquence varie lentement en fonction du temps, de mesurer le module de son spectre à la fréquence f' à partir de la mesure du module de son enveloppe complexe à l'instant t' défini par $t' = \frac{f'-f_i}{\beta}$.

Ce résultat va nous permettre de mesurer l'ondulation introduite par la $STF^k(f)$ sur le signal d'entrée en mesurant l'amplitude de la réponse fréquentielle aux fréquences f_r^k et f_r^{k+1} à partir du module de l'enveloppe complexe aux instants correspondants. Le schéma fonctionnel de cette méthode de calibration est donné sur la figure 4.23.

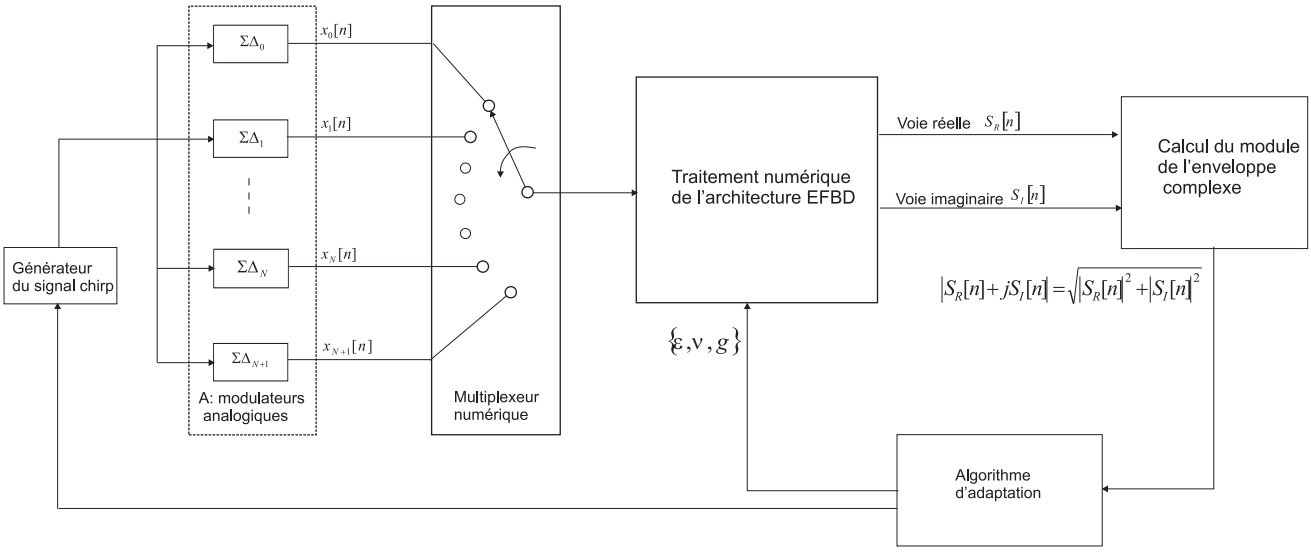

Fig. 4.23 – Schéma fonctionnel de la méthode de calibration de la STF(f).

Cette méthode nécessite :

- Un générateur de signal chirp qui couvre toute la bande du signal utile $[f_1,\ f_2]$. Ce générateur est très simple à réaliser en numérique [55, 56, 57]. Un CNA monobit à base de modulateur $\Sigma\Delta$ [45] permet la mise en forme à l'entrée des des modulateurs à temps continu sans exiger de grandes resources matérielles.
- Un calculateur pour évaluer le module de l'enveloppe complexe afin de pouvoir mesurer l'effet de la fonction de transfert STF(f).
- Un algorithme d'adaptation qui permet, à partir du module de l'enveloppe complexe, de mesurer l'effet de la STF(f) et ensuite de mettre à jour les paramètres $\{\varepsilon, \nu, g\}$ du filtre $C_1^k(z)$ pour mieux corriger le module de la STF(f). Cet algorithme d'adaptation sera détaillé dans la suite.

4.6.3 Algorithme d'adaptation pour la calibration de la STF

Les conditions préalables à l'application de cet algorithme sont :

- La connaissance de la bande de fonctionnement $\left[f_r^k,\ f_r^{k+1}\right]$ de chaque modulateur. Pour cela, la calibration de la $STF^k(f)$ doit être appliquée après la détermination des bandes par l'algorithme de détermination des bandes développé au § 4.5.2.
- La connaissance des directions de recherche λ_1 et λ_2 optimales pouvant être déterminées par simulations paramétriques (voir § 4.6.2).

Les étapes de calcul de cet algorithme, illustrées à la figure 4.24, sont les suivantes :

1. Initialiser les paramètres du filtre correcteur $C_1^k(z)$ $\{\varepsilon, \nu, g\}_{k=0...N+1}$ à leurs valeurs théoriques calculées avec des modulateurs idéaux pour une bande de fonctionnement de largeur $\frac{B}{N}$.

2. Faire varier, pour chaque direction de recherche λ_1 et λ_2, la valeur de $\nu(k)$ autour de sa valeur théorique dans l'intervalle $\Delta\nu$ sur N_v valeurs.

3. Générer un signal chirp en entrée de l'architecture EFBD d'amplitude A et de bande $\left[f_r^k,\ f_r^{k+1}\right]$.

4. Calculer le signal en sortie en utilisant un filtre passe-bas de largeur de bande 4 fois plus grande que la valeur nominale $\frac{B}{2N}$ afin d'assurer un gain unitaire dans la bande $\left[f_r^k,\ f_r^{k+1}\right]$. Ce dernier filtre est à 256 coefficients pour assurer un gain constant dans la bande souhaitée. Il peut être réalisé en utilisant une décomposition polyphase en quatres branches, chacune de 64 coefficients. Les coefficients des 4 branches seront stockés dans la mémoire ROM_3 du traitement complexe (figure 4.13).

5. Calculer l'ondulation définie comme la différence entre la valeur minimale et la valeur maximale M du module de l'enveloppe complexe dans la bande de fonctionnement. L'utilisation d'un filtre passe-bas de largeur de bande 4 fois plus grande entraîne une plus grande récupération du bruit hors la bande de fonctionnement et par conséquent le module de l'enveloppe complexe sera très bruité. Ceci introduit de l'erreur sur la vraie valeur de l'ondulation et influe sur la convergence vers les vraies valeurs des paramètres. La solution à ce problème est d'utiliser un filtre passe-bas de 32 ou 64 coefficients pour lisser le module de l'enveloppe complexe. Au total, cet algorithme d'adaptation exige l'ajout de 5 jeux de 64 coefficients dans la mémoire ROM_3 du traitement complexe (figure 4.13).

6. Déterminer la valeur optimale $\nu_{opt}(k)$ à partir de la valeur de l'intervalle $\Delta\nu$ qui assure l'ondulation minimale. La valeur de $\varepsilon_{opt}(k)$ se déduit à partir de $\nu_{opt}(k)$ par la formule $\varepsilon(k) \rightarrow \varepsilon(k) + 2\pi \times \nu_{opt}(k) \times \lambda(l)$.

7. Calculer la valeur du gain g(k) à la fin de la recherche de $\nu_{opt}(k)$ dans les deux directions de recherche. Cette valeur du gain doit être égale à $\frac{1}{M}$ où M est la valeur maximale du module de l'enveloppe complexe obtenue avec les valeurs optimales $\nu_{opt}(k)$ et $\varepsilon_{opt}(k)$. Ceci revient à remplacer la valeur théorique de départ g(k) par $g(k) - \left(\frac{M}{A} - 1\right)$ en utilisant un développement limité à l'ordre 1 du gain.

Même si cet algorithme est assez riche en termes de nombre d'opérations à effectuer, son implantation ne présente pas de réelles complexités matérielles. Il nécessite des compteurs, des registres, des comparateurs et le calculateur d'amplitude. Nous avons estimé, à titre d'exemple, le temps nécessaire pour réaliser la calibration de toutes les $STF^k(f)$ en considérant un signal de 40 000 échantillons en sortie du modulateur k à calibrer. Le temps nécessaire pour générer ce type de signal est de 50 μs (40000/800MHz). L'algorithme n'exige pas la mémorisation des 40 000 échantillons car le calcul de l'ondulation se fait de façon récursive. La calibration de chaque $STF^k(f)$ nécessite de parcourir N_v valeurs de ν et ceci dans deux directions de recherche. Avec $N_v = 32$, le temps nécessaire pour calibrer un seul modulateur est 3.2 ms (50 μs $\times$ 32 $\times$ 2). Par conséquent, le temps nécessaire pour calibrer la STF(f) des 10 modulateurs est de 32 ms.

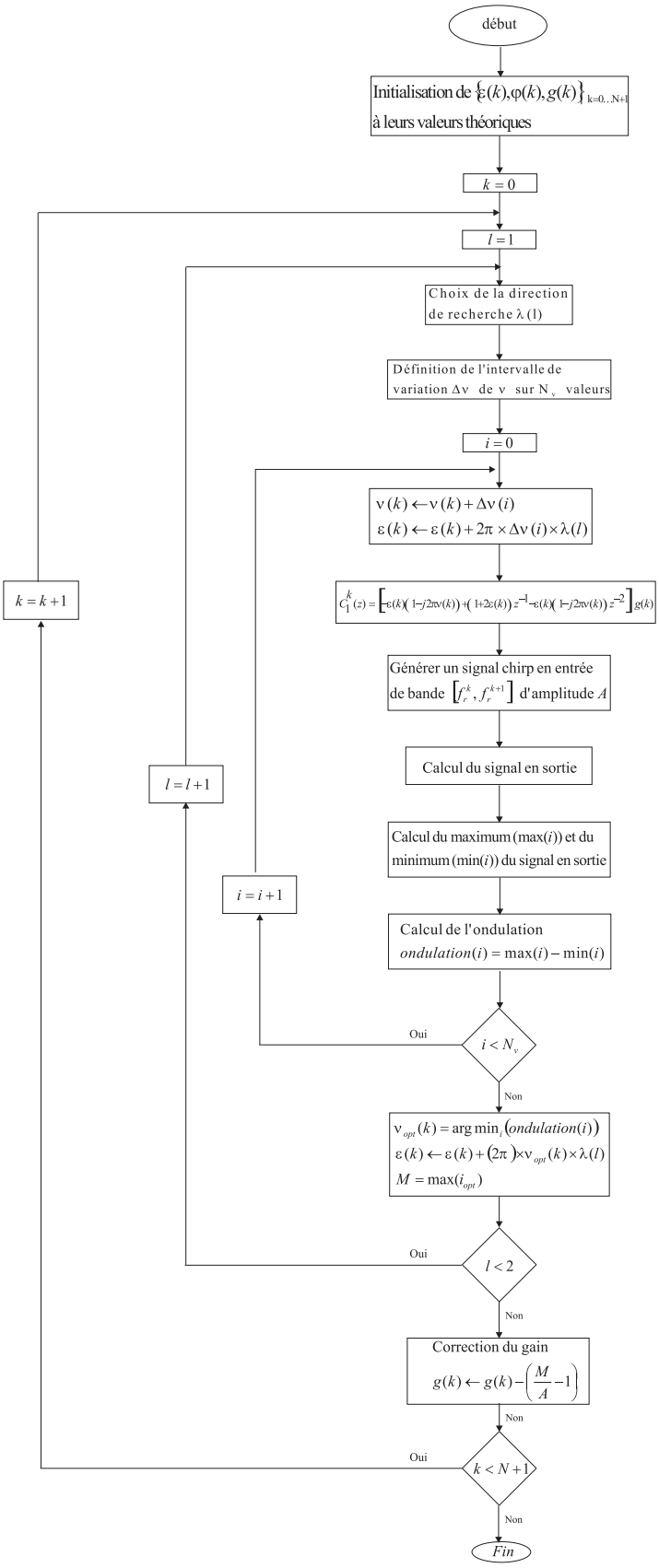

Fig. 4.24 – Organigramme de l'algorithme d'adaptation pour la correction de la STF(f).

4.6.4 Exemple : calibration de la STF du quatrième modulateur

Nous avons procédé à la calibration de la $STF^4(f)$ en sélectionnant la sortie du modulateur $\Sigma\Delta_4$ à l'aide du multiplexeur numérique (figure 4.23) puis en appliquant l'algorithme d'adaptation (§ 4.6.3). Le nombre de sous-intervalles N_v était fixé à 64. Nous avons considéré un signal de type chirp en entrée d'amplitude 0.5 et de bande $\left[f_r^3,\ f_r^6\right]$. La figure 4.25 montre le module de l'enveloppe complexe du signal en sortie calculé avec le filtre passe-bas, utilisé pendant la période d'exécution de l'algorithme de calibration, de largeur de bande 4 fois plus grande que la valeur théorique $\frac{B}{2N}$.

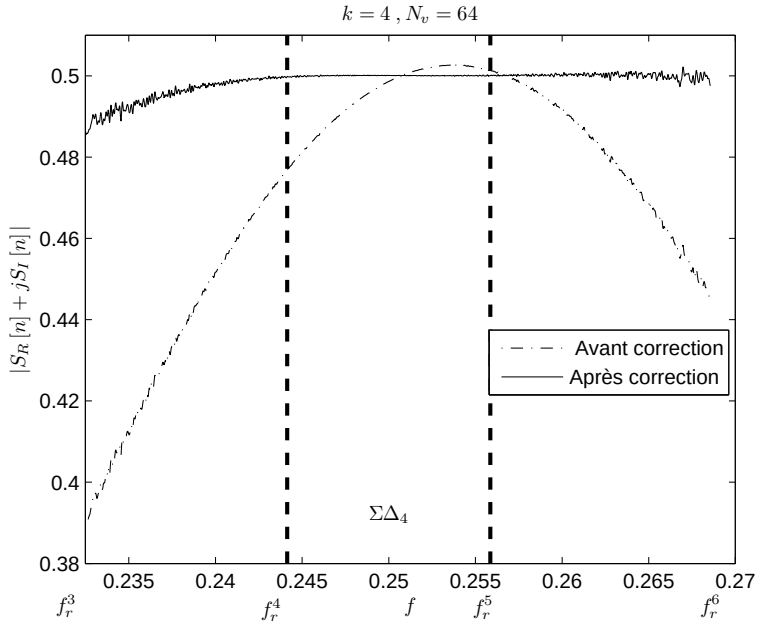

Fig. 4.25 – Module de l'enveloppe complexe du signal en sortie avec le filtre passe-bas de calibration.

Bien que le module de l'enveloppe complexe soit calculé dans le domaine temporel, nous avons choisi de représenter en abscisse le vecteur de fréquence généré à chaque instant par la relation $f = f_i + \beta t$ pour le signal de type chirp. Ceci dans le but de visualiser le module en fonction de la fréquence. Nous remarquons qu'après la correction, le module de l'enveloppe complexe est presque constant et égal à 0.5 dans la bande de fonctionnement, compensant ainsi l'atténuation provoquée par la $STF^4(f)$.

La figure 4.26 montre le module de l'enveloppe complexe du signal en sortie du modulateur $\Sigma\Delta_4$ calculé avec le filtre passe-bas correspondant à la largeur de sa bande de fonctionnement $\left[f_r^4,\ f_r^5\right]$. Nous pouvons remarquer l'apport du filtre de correction $C_1^4(z)$ sur la compensation de l'atténuation produite par la $STF^4(f)$. Les atténuations du signal après correction vers les fréquences de bords f_r^4 et f_r^5 sont dues à la réponse fréquentielle du filtre passe-bas $G_{pb}^4(z)$. Elles seront compensées par le gain apporté par les filtres passe-bas des bandes adjacentes $G_{pb}^3(z)$ et $G_{pb}^5(z)$.

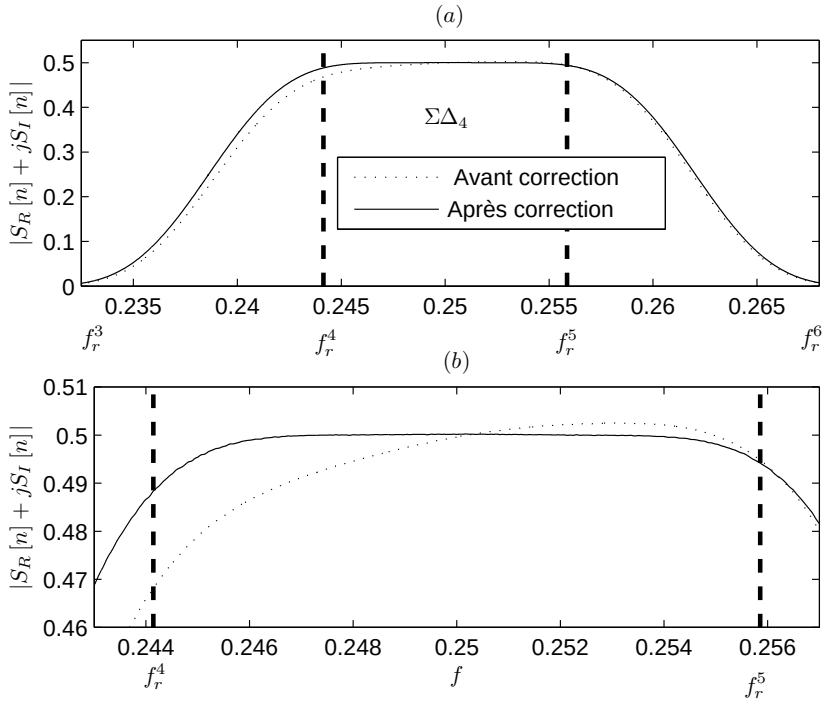

FIG. 4.26 – (a) Module de l'enveloppe complexe du signal en sortie avant et après correction, (b) zoom sur le module autour de la valeur théorique 0.5.

L'algorithme d'adaptation pour la calibration de la $\mathrm{STF^k(f)}$ permet de déterminer les valeurs de $\{\varepsilon, \nu, \mathrm{g}\}$ du filtre de correction $\mathrm{C_1^k(z)}$. Il s'agit cependant de savoir avec quelle précision ces valeurs doivent être calculées. Nous avons noté au § 4.6.2 que le nombre de sous-intervalles $\mathrm{N_v}$ influe à la fois sur la finesse du critère à minimiser et sur la précision de calcul de $\{\varepsilon, \nu, \mathrm{g}\}$. Nous avons également remarqué d'après la figure 4.21 que le nombre d'intervalles $\mathrm{N_v}$ minimum permettant d'obtenir un critère fin et convexe est égal à 32. Pour illustrer l'influence du nombre d'intervalles $\mathrm{N_v}$ sur la précision, nous avons calculé le module de l'enveloppe complexe pour l'exemple précédent (modulateur $\Sigma\Delta_4$ et un signal de type chirp en entrée d'amplitude 0.5 et de bande $\left[\mathrm{f_r^3},\ \mathrm{f_r^6}\right]$) en utilisant :

1. le filtre $\mathrm{C_1^4(z)}$ pour lequel les paramètres $\{\varepsilon, \nu, \mathrm{g}\}$ sont fixés à leurs valeurs idéales,
2. le filtre $\mathrm{C_1^4(z)}$ pour lequel les paramètres $\{\varepsilon, \nu, \mathrm{g}\}$ sont décalés de leurs valeurs idéales par une erreur maximale d'un demi pas de quantification :

$$\begin{cases} \varepsilon(k) \leftarrow \varepsilon(k) + \frac{q_\varepsilon}{2} \quad \text{avec} \quad q_\varepsilon = \frac{\max(\Delta\varepsilon) - \min(\Delta\varepsilon)}{N_v} \\ \nu(k) \leftarrow \nu(k) + \frac{q_\nu}{2} \quad \text{avec} \quad q_\nu = \frac{\max(\Delta\nu) - \min(\Delta\nu)}{N_v} \end{cases} \tag{4.21}$$

q_ε : est le pas de quantification sur la valeur ε.
q_ν : est le pas de quantification sur la valeur ν.

La figure 4.27 présente le module de l'enveloppe complexe calculé avec les valeurs idéales de $\{\varepsilon, \nu, \mathrm{g}\}$ et celui calculé avec les valeurs $\left\{\varepsilon(\mathrm{k}) + \frac{\mathrm{q}_\varepsilon}{2}, \nu(\mathrm{k}) + \frac{\mathrm{q}_\nu}{2}\right\}$ et ceci pour un nombre de sous-intervalles $\mathrm{N_v} = 8,\ 16,\ 32,\ 64$.

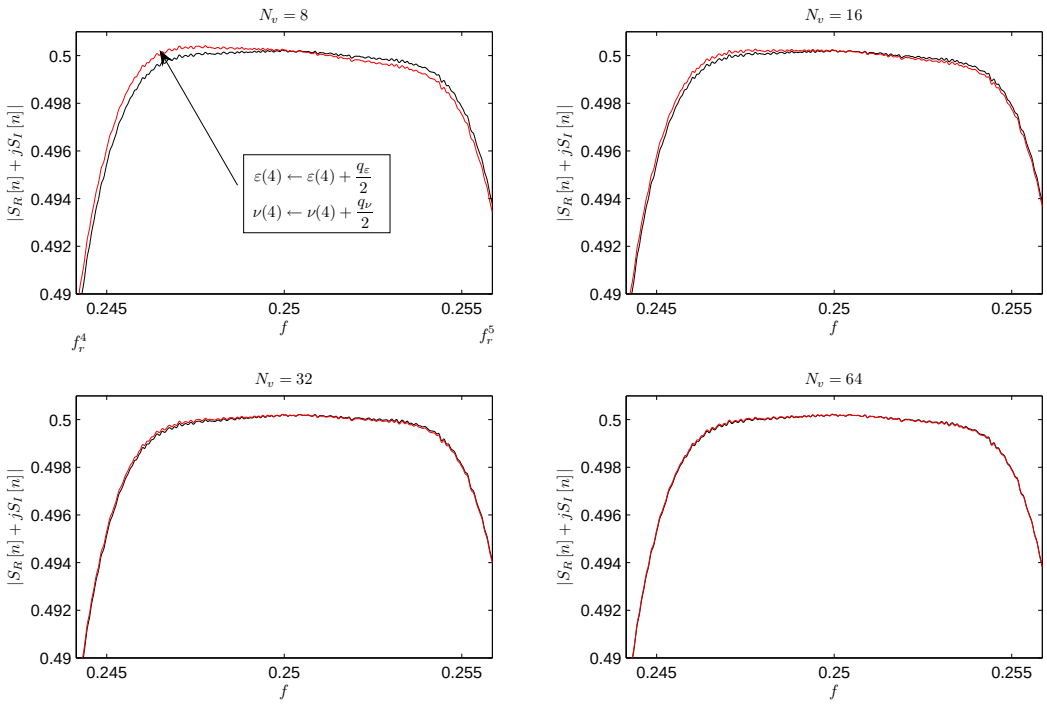

Fig. 4.27 – Module de l'enveloppe complexe avec $N_v = 8, 16, 32, 64$.

Nous avons choisi d'ajouter un demi-pas de quantification $\frac{q}{2}$ aux valeurs idéales car l'erreur de quantification est bornée entre $\pm\frac{q}{2}$. Nous remarquons que pour $N_v = 8$, la performance du filtre correcteur est détériorée lorsqu'on ajoute un demi-pas de quantification. Par contre, à partir $N_v = 32$ la différence entre les deux courbes de module est très minime. En conclusion, le nombre de sous-intervalles optimal permettant d'obtenir un critère convexe et une bonne précision sur les valeurs de $\{\varepsilon, \nu, g\}$ est au moins égal à 32.

4.7 Raccordement de phase entre bandes de fonctionnement adjacentes

La méthode de reconstruction numérique avec démodulation nécessite, comme nous l'avons vu aux § 3.4.3 et 3.4.4, une correction de la phase de chaque modulateur. Cette correction consiste à raccorder les phases dans la zone de transition à la limite entre les bandes de fonctionnement des modulateurs adjacents. Ce raccordement assure une phase presque linéaire dans la bande du signal utile et par conséquent diminue les ondulations dans le spectre du signal en sortie qui sont dues à la différence de phase dans la zone de transition. Nous distinguons deux types de raccordement :

– Le raccordement entre les phases des fonctions de transfert $STF^k(f)$ des modulateurs dans la zone de transition (voir § 3.4.3). Ce raccordement se fait en multipliant le signal en sortie de chaque modulateur par le terme complexe C_3^k de module égal à 1. L'expression de ce terme est donnée par l'équation (3.34). La multiplication par C_3^k a pour effet d'introduire un déphasage sur la séquence de démodulation complexe $m_k[n] = e^{j2\pi f_c^k n}$.

– Le raccordement de phase dans les zones de transition pour corriger le déphasage introduit par les filtres passe-bas suite au processus de démodulation et de modulation (voir § 3.4.4). De la même façon, le raccordement se fait en multipliant par le terme complexe C_4^k. Ce terme est exprimé par l'une des équations (3.36) ou (3.37) suivant l'emplacement du filtre dans la chaîne du traitement avant ou après la décimation :

$$\begin{cases} C_4^k = e^{-j2\pi\left(f_C^k - f_C^1\right)M} & \text{correction avant décimation} \\ C_4^k = e^{-j2\pi\left(f_C^k - f_C^1\right)MR_d} & \text{correction après décimation} \end{cases}$$

Ce terme complexe introduit un déphasage sur la séquence de démodulation $m_k[n] = e^{j2\pi f_c^k n}$ ou de modulation $m_k^{'}[n] = e^{j2\pi f_c^k R_d n}$.

Le calcul des termes de correction de phase C_4^k exige la connaissance de la fréquence centrale f_c^k de chacune des bandes de fonctionnement. Ces fréquences centrales ne peuvent pas être calculées à l'avance à cause des imperfections des composants analogiques qui font changer la largeur de bande optimale pour le traitement. Elles sont calculées à partir de la largeur de bande de fonctionnement à la fin de l'exécution de l'algorithme de détermination des bandes de fonctionnement développé au § 4.5.2.

Nous avons montré au § 4.5 que les bandes de fonctionnement peuvent être quantifiées avec un pas $q_f = \frac{f_e}{2^{10}}$ sans avoir une perte considérable en ENOB. Comme les fréquences centrales f_c^k se trouvent au milieu de la bande de fonctionnement, la valeur de f_c^k peut se trouver entre deux valeurs successives séparées par $q_f = \frac{f_e}{2^{10}}$. D'où la nécessité de quantifier les fréquences centrales sur 11 bits avec un pas $q_{f_c} = \frac{f_e}{2^{11}} = \frac{f_e}{2048}$.

La quantification des fréquences centrales f_c^k (avec un pas $q_{f_c} = \frac{f_e}{2^{11}}$) apporte un avantage remarquable à l'implémentation du traitement numérique. En effet, cette quantification permet de calculer à l'avance les valeurs possibles de l'exponentielle $e^{j2\pi f_c^k}$ dans l'expression de $m_k[n]$ en quantifiant la plage fréquentielle entre 0 et 1 sur 2048 valeurs. Ces valeurs seront stockées dans un tableau en mémoire ROM. La séquence $m_k[n]$ est obtenue en lisant les valeurs stockées dans la ROM de façon cyclique. Par exemple, à l'instant $n = 0$, $m_k[0]$ est égal au premier élément du tableau et pour $n = 2048$ on revient sur le premier élément du tableau. En fait, cela revient à dire qu'un tour complet du cercle unité avec un pas de $\frac{2\pi}{2048}$ a été effectué. Un déphasage positif ou négatif sur la séquence de démodulation se traduit par un décalage en avant ou en arrière de la valeur de $m_k[n]$ dans le tableau. La séquence de modulation $m_k^{'}[n] = e^{j2\pi f_c^k R_d n}$ utilise les mêmes valeurs de $e^{j2\pi f_c^k}$ stockées dans le tableau mais avec un pas de $R_d \times f_c^k$ entre deux valeurs successives du tableau.

4.7.1 Détermination des coefficients C_3^k et C_4^k

Après avoir déterminé la largeur des bandes de fonctionnement des différents modulateurs suite à l'exécution de l'algorithme développé au § 4.5.2, la correction de phase par le terme C_4^k se fait facilement par le calcul du nouveau décalage à appliquer sur la séquence de démodulation ou de modulation. En ce qui concerne le terme de correction C_3^k, le calcul du décalage à effectuer sur la séquence de modulation à partir de son expression mathématique (équation (3.34)) n'est pas trivial. En effet, l'accès aux valeurs de la phase de la STF de chaque modulateur n'est pas possible. De ce fait, nous proposons une méthode numérique qui permet de trouver le bon décalage pour raccorder les phases des STF à la fréquence limite entre deux bandes de fonctionnement adjacentes. L'idée de cette méthode repose sur les deux postulats suivants :

- Le raccordement de phase des STF à la fréquence limite entre les bandes adjacentes permet de diminuer l'ondulation autour de cette fréquence. Par conséquent, la bonne valeur de C_3^k est celle qui minimise l'ondulation autour de la fréquence limite.
- La correction par le terme C_3^k consiste à appliquer un décalage (dec) positif ou négatif à la séquence $m_k[n]$.

Comme la correction de phase n'est qu'un simple décalage de la séquence $m_k[n]$, la méthode numérique consiste à fixer le décalage à sa valeur théorique calculée lors de la conception pour des modulateurs idéaux avec des bandes de fonctionnement égales à $\frac{B}{N}$, puis à faire varier ce décalage autour de sa valeur théorique pour garder la valeur qui assure une ondulation minimale autour de la fréquence limite. Le calcul de l'ondulation est réalisé, comme dans le cas de la calibration de la STF, avec la mesure du module de l'enveloppe complexe en injectant un signal chirp en entrée.

4.7.2 Algorithme de calcul pour le raccordement de phase

Le schéma fonctionnel de cette méthode numérique est le même que celui utilisé pour la calibration de la STF (figure 4.23) dans laquelle tous les modulateurs sont connectés au traitement numérique. Le multiplexeur ne sera pas utilisé durant le raccordement de phase. Les étapes de calcul de l'algorithme, illustrées à la figure 4.28, sont les suivantes :

1. Initialiser le décalage à appliquer pour chaque modulateur à sa valeur théorique $dec_{th}(k)$. Cette valeur est estimée lors de la conception en considérant que les modulateurs sont idéaux et que la bande de fonctionnement est égale à $\frac{B}{N}$. Le décalage appliqué au premier modulateur est toujours considéré égal à zéro ($dec_{th}(0) = 0$). La phase du premier modulateur est considéré comme une référence de phase.
2. Tester si le modulateur est utilisé. Ce test concerne le premier modulateur en testant si $f_r^k = f_r^{k-1}$ et le dernier modulateur en testant si $f_r^k = f_r^{k+1}$.
3. Définir l'intervalle Δ_{dec} autour duquel le décalage va varier autour de sa valeur théorique. Cet intervalle est défini sur N_{dec} valeurs. Il est généralement symétrique de type $[-n_1,\ n_1]$ avec $N_{dec} = 2n_1 + 1$.
4. Générer un signal de type chirp de bande $\left[f_r^{k-1},\ f_r^{k+1}\right]$ et d'amplitude A pour chaque valeur du décalage défini par $dec(k) = dec_{th}(k) + \Delta_{dec}(i)$. La bande du signal a été choisie de façon à assurer, dans le cas où le raccordement est effectué, un module constant de l'enveloppe complexe en sortie autour de la fréquence f_r^k pour laquelle on cherche le bon décalage de raccordement.
5. Calculer l'ondulation de l'enveloppe complexe autour de la fréquence f_r^k dans une zone définie par $[f_{inf},\ f_{sup}]$ avec $f_{inf} = \frac{f_r^k + f_r^{k-1}}{2}$ et $f_{sup} = \frac{f_r^k + f_r^{k+1}}{2}$.
6. Choisir la valeur optimale de l'intervalle Δ_{dec} pour laquelle l'ondulation est minimale.
7. Ajouter la valeur optimale de $\Delta_{dec_{opt}}$ au décalage des modulateurs d'indice supérieur à k.

Nous avons testé cet algorithme en considérant une architecture EFBD avec 10 modulateurs. Les fréquences centrales des résonateurs sont décalées de $\frac{B}{2N}$ (dispersions globales) de leurs valeurs théoriques. Nous avons choisi $N_{dec} = 63$ avec une plage de variation du décalage autour de sa valeur théorique définie par $\Delta_{dec} = [-31,\ 31]$.

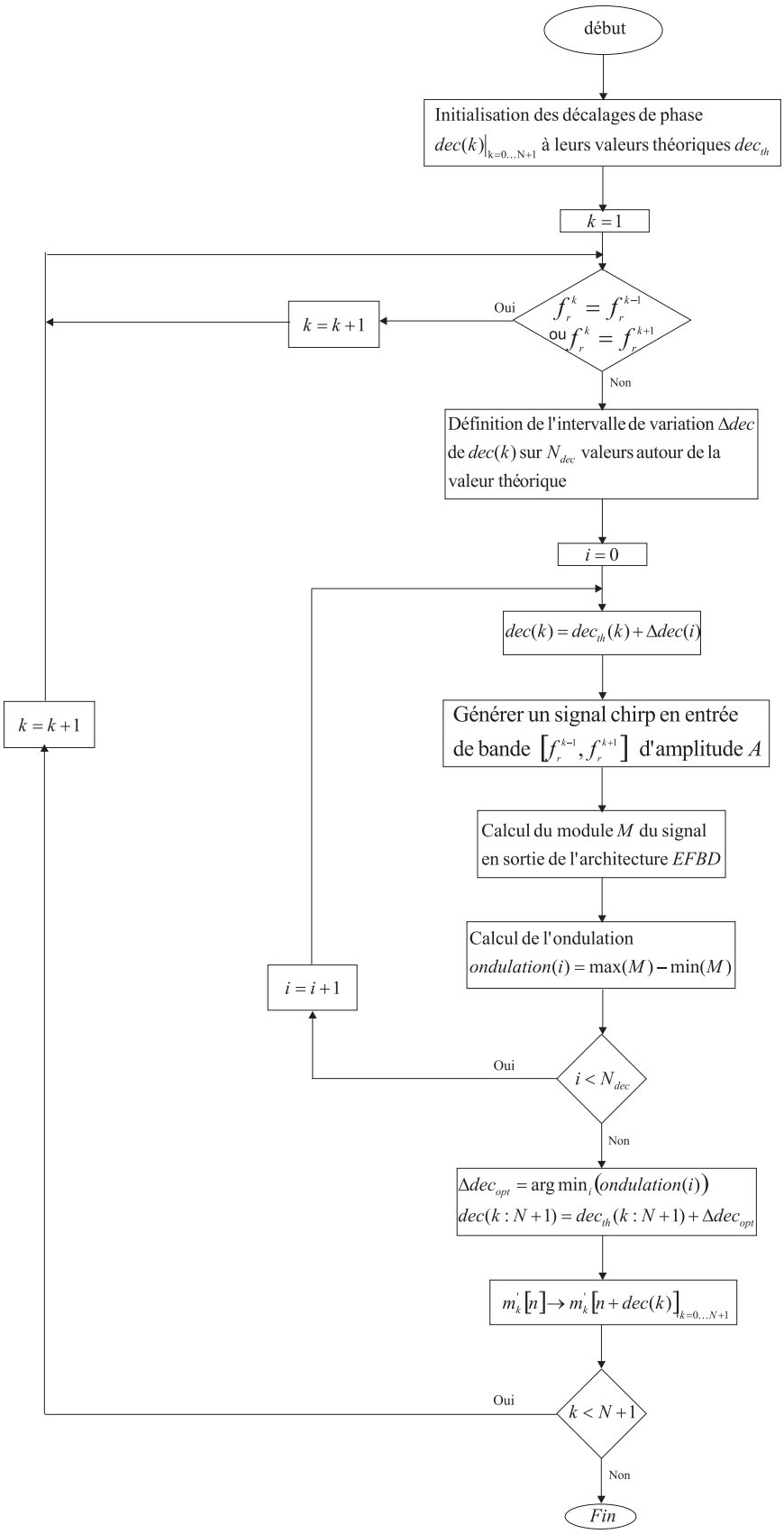

Fig. 4.28 – Organigramme de l'algorithme de raccordement de phase de la STF.

La figure 4.29 (a) montre le module de l'enveloppe complexe avant le raccordement de phase des modulateurs ΣΔ et celui calculé après le raccordement de phase en utilisant les valeurs de décalage délivrées par l'algorithme de calcul (§ 4.7.2).

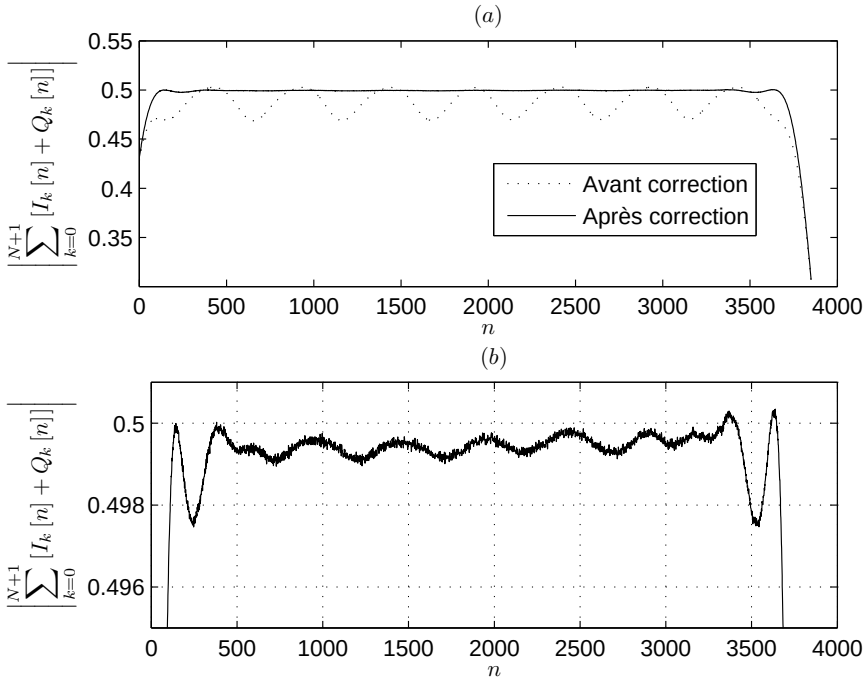

Fig. 4.29 – (a) Module de l'enveloppe complexe en sortie avant et après le raccordement de phase des modulateurs. (b) Agrandissement du tracé du module de l'enveloppe complexe après raccordement.

Nous remarquons qu'avant le raccordement, le module de l'enveloppe complexe présente une ondulation maximale de 6.4% en dessous de la valeur théorique du module (0.5). Par contre, avec le raccordement de phase, le module de l'enveloppe complexe est presque constant et l'ondulation maximale est 0.4% en dessous de la valeur théorique du module. Le raccordement de phase apporte une correction du gain remarquable vers les fréquences de bords (autour des instants t = 0 et t = 3500) (figure 4.29 (a)). Cependant, nous remarquons, en agrandissant cette partie, que le module de l'enveloppe complexe présente un gain plus élevé vers les fréquences de bords. Ceci est dû à la réponse fréquentielle des filtres FIR utilisés dans ces bandes avec 64 coefficients (figure 4.30).

Dans ce cas particulier, nous avons remarqué que le nombre d'échantillons minimum en entrée pour pouvoir mesurer l'ondulation autour de la fréquence limite est de l'ordre de 20 000 échantillons. Par conséquent, le temps nécessaire pour générer un signal chirp dans la bande $\left[\mathrm{f_r^{k-1}},\ \mathrm{f_r^{k+1}}\right]$ est égal à 25 μs (20000/800 MHz). En faisant varier chaque valeur du décalage sur $\mathrm{N_{dec}}$ valeurs autour de la valeur théorique, le temps nécessaire pour raccorder la phase de chaque modulateur est 25 $\mu\mathrm{s} \times \mathrm{N_{dec}}$. Si tous les modulateurs sont utilisés, le temps nécessaire pour raccorder la phase de tous les modulateurs est de 25 $\mu\mathrm{s} \times \mathrm{N_{dec}} \times \mathrm{N}$. Dans notre cas, avec $\mathrm{N_{dec}} = 64$, le temps nécessaire est de 16 ms.

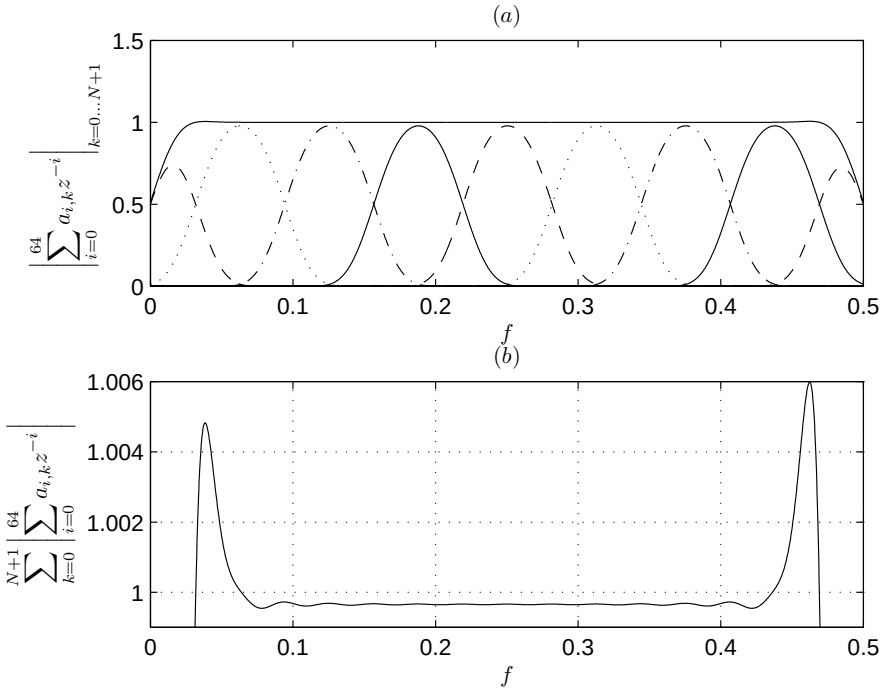

Fig. 4.30 – (a) Module de la réponse fréquentielle des filtres $F_{pb}^{k}(z)$, (b) somme de la réponse fréquentielle de tous les filtres passe-bas.

4.8 Conclusion

Nous avons étudié dans ce chapitre le principe de l'architecture EFBD ainsi que sa robustesse face aux imperfections des composants analogiques. Nous avons montré qu'un décalage d'une largeur de sous-bande $\frac{B}{N}$ sur les fréquences centrales des résonateurs (dispersion globale) est acceptable sans avoir une chute remarquable de l'ENOB. L'adaptation du traitement numérique aux imperfections des composants analogiques nécessite la détermination de la bande de fonctionnement des modulateurs. Nous avons suivi deux démarches pour déterminer ces bandes. La première est basée sur l'identification de la NTF de chaque modulateur. Elle est limitée, en appliquant des algorithmes de calcul « *On Line* », par le nombre de paramètres à estimer (8 au maximum). Cette démarche n'est pas applicable dans notre cas où le nombre de paramètres à estimer est égal à 12. La deuxième démarche, que nous avons proposée, consiste à déterminer les bandes de fonctionnement à partir de la puissance de bruit en sortie. Elle présente une précision de $10^{-3}f_e$ et est facile à implanter.

Nous avons proposé également deux algorithmes pour la calibration de la STF et le raccordement de phases des modulateurs autour de la fréquence limite entre deux bandes de fonctionnement adjacentes. Ces deux algorithmes présentent une facilité d'implantation, n'exigeant que des fonctions de base de l'électronique numérique.

Dans la suite, nous allons nous intéresser à l'implantation du traitement numérique associé à l'architecture EFBD.

Chapitre 5

Implémentation du traitement numérique

Objectif

Ce chapitre présente l'implémentation du traitement numérique présent à la sortie de chaque modulateur pour reconstruire le signal en sortie. Chaque bloc constitutif de ce traitement a été optimisé sur des critères de vitesse et de surface. Une méthode de détermination du nombre de bits optimal des entrées/sorties de chaque bloc est exposée. Après l'optimisation de chaque bloc élémentaire, les performances de l'architecture globale ont pu être évaluées, dont sa fréquence maximale de fonctionnement. Cette dernière a pu être augmentée en utilisant le principe de pipelinage pour découper le chemin critique de l'architecture.

5.1 Introduction

Nous avons adopté au chapitre 3 (§ 3.3.2) une méthode de reconstruction du signal numérisé par le banc de modulateurs $\Sigma\Delta$ basée sur la démodulation et la modulation du signal. Le signal de chaque modulateur est ramené en bande de base, puis traité pour éliminer le bruit de quantification et ramené à nouveau à sa bande initiale. La figure 5.1 (partie C) présente le traitement numérique à la sortie d'un seul modulateur.

Dans la suite, nous allons présenter une architecture pour chaque bloc élémentaire (démodulateur complexe, décimateur, filtre passe-bas $G_{pb}^{k}(z)$, filtre de calibration $C_{1}^{k}(z)$ et modulateur) qui permet d'optimiser la surface d'implantation et de garantir la vitesse de fonctionnement (800 MHz dans l'exemple considéré au cours de cette thèse (voir tableau 3.1)). Ensuite, nous allons optimiser l'architecture globale du traitement d'un seul étage en assemblant les différents composants constitutifs.

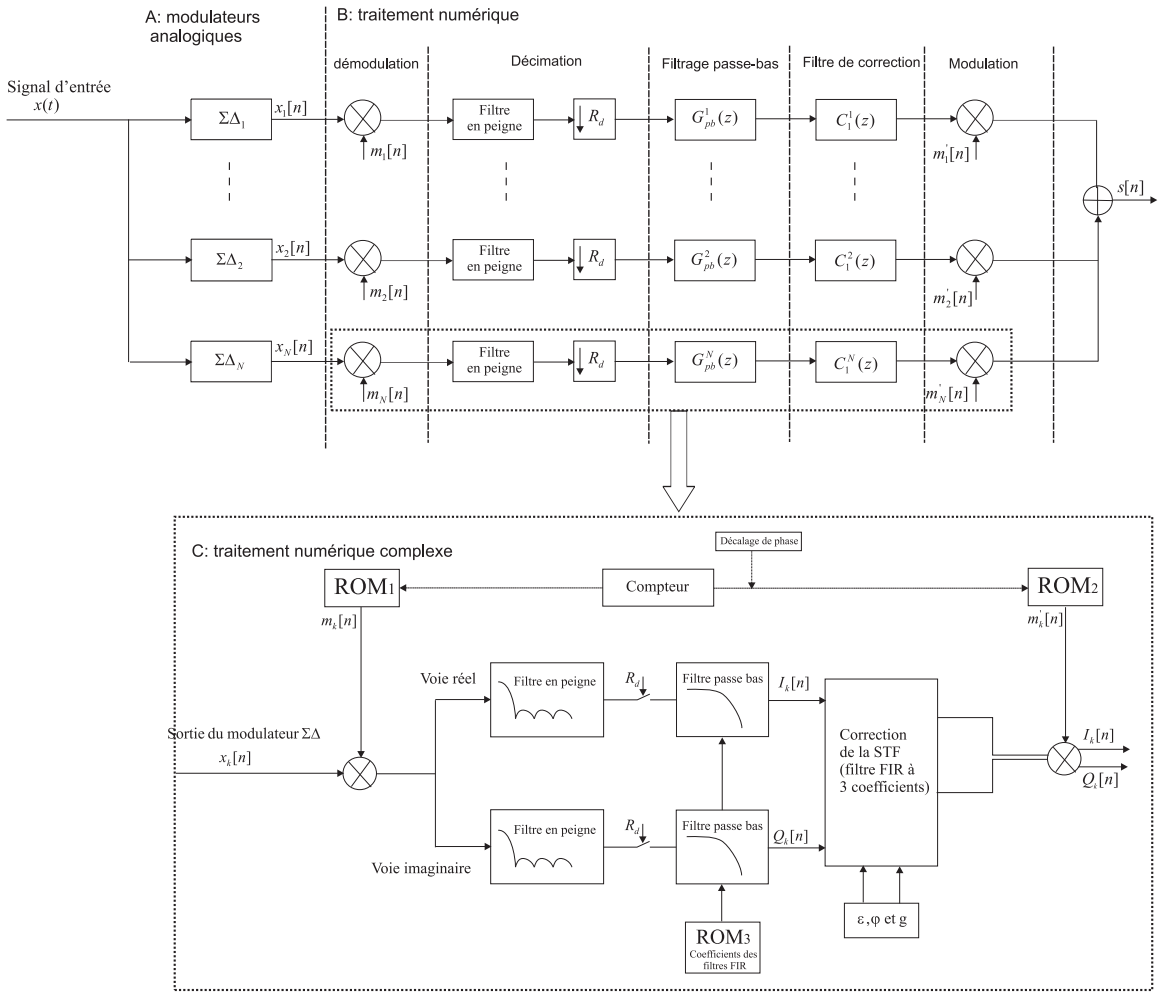

Fig. 5.1 – Architecture EFBD avec filtres de corrections et traitement numérique de sortie.

5.2 Démodulateur

La démodulation complexe consiste à ramener en bande de base les signaux $x_k[n]$ en sortie des différents modulateurs grâce à une multiplication par la séquence complexe $m_k[n] = e^{-j2\pi f_c^k n}$. Nous avons vu au chapitre 4 (§ 4.5.2) que la largeur de bande de fonctionnement de chaque modulateur peut être quantifiée avec un pas $q_f = \frac{f_e}{2^{10}}$ sans avoir une chute notable en résolution. De ce fait, la fréquence de démodulation f_c^k qui se trouve au milieu de la bande de fonctionnement doit être quantifiée avec un pas deux fois plus faible soit $q_{f_c} = \frac{f_e}{2^{11}} = \frac{f_e}{2048}$. Par conséquent, la valeur de la fréquence centrale f_c^k est un nombre rationnel de la forme $f_c^k = \frac{i \times f_e}{2048}$, $i \in [0 \dots 2047]$. Ceci implique que la séquence de démodulation $m_k[n]$ est périodique de période 2048. Cette périodicité a un effet considérable sur l'implantation du bloc démodulateur. En effet, la séquence de démodulation $m_k[n]$ peut être obtenue à partir d'un tableau où sont stockées les valeurs de l'exponentielle à la fréquence $\frac{f_e}{2048}$ $\left(e^{j2\pi \frac{f_e}{2048} n} \Big|_{n=0 \cdots 2047} \right)$. On note que :

- le premier élément $m_k[0]$ de la séquence de démodulation est le premier élément du tableau,
- pour une fréquence $f_c^k = \frac{i \times f_e}{2048}$, les valeurs de la séquence $m_k[n]$ sont obtenues suite à une lecture cyclique du tableau avec un pas de i. La figure 5.2 présente, à titre d'exemple, le principe de fonctionnement avec un pas de quantification $\frac{f_e}{64}$ pour faciliter la présentation.

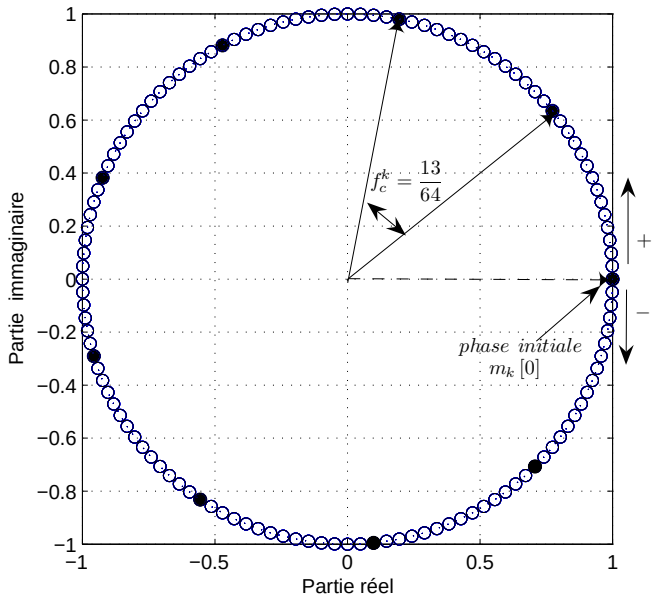

Fig. 5.2 – Valeurs de l'exponentielle $e^{j2\pi\frac{l}{64}}\Big|_{l=0...64}$.

La multiplication par la séquence complexe $m_k[n] = e^{-j2\pi f_c^k n} = \cos(2\pi f_c^k n) - j\sin(2\pi f_c^k n)$ engendre deux signaux : un signal en phase résultant de la multiplication par $\cos(2\pi f_c^k n)$ et un autre en quadrature de phase résultat de la multiplication par $-j\sin(2\pi f_c^k n)$. La figure 5.3 présente la partie réelle (a) et la partie imaginaire (b) de l'exponentielle $e^{j2\pi\frac{f_e}{2048}n}\Big|_{n=0\cdots2047}$.

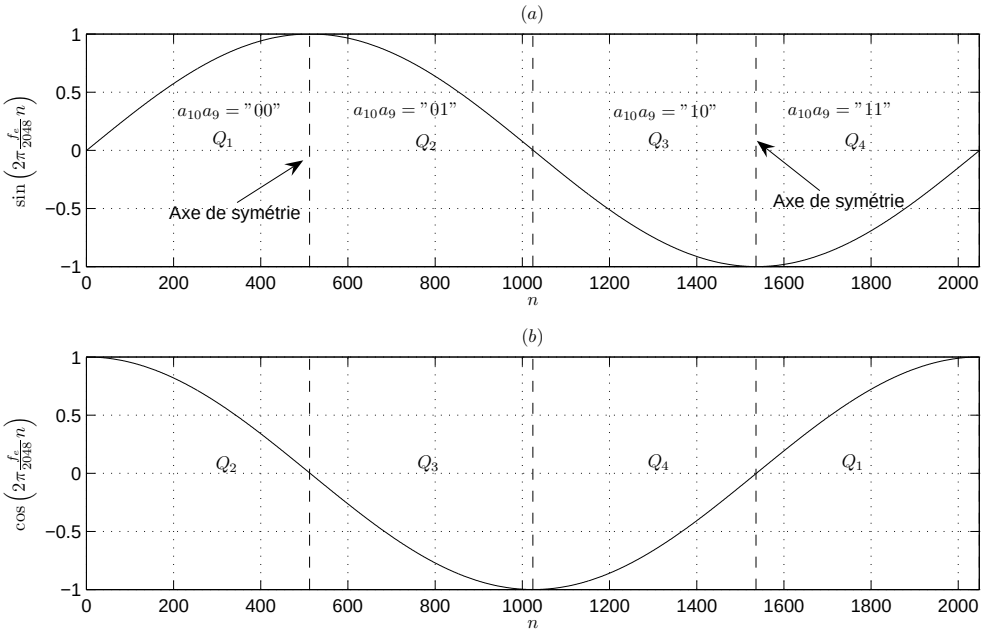

Fig. 5.3 – (a) $\sin\left(2\pi\frac{n}{2048}\right)$,(b) et $\cos\left(2\pi\frac{n}{2048}\right)$ en fonction du temps.

On peut noter que :

- Le calcul des valeurs de l'exponentielle n'est pas nécessaire pour toutes les phases. En effet, il suffit de calculer ces valeurs dans un quadrant, par exemple dans le premier quadrant Q_1. Les valeurs des autres quadrants se déduisent facilement de la première par symétrie. Le deuxième quadrant Q_2 est symétrique par rapport au premier. Les valeurs du troisième quadrant Q_3 sont multipliées par -1 par rapport à celles du quadrant Q_1. Les valeurs du quatrième quadrant sont symétriques et négatives par rapport à celles du quadrant Q_1.
- En pratique, les valeurs du premier quadrant sont calculées à partir de l'expression $\sin\left(2\pi\frac{i+0.5}{2048}\right)$. Le fait d'introduire, dans la fonction sinusoïdale, un décalage fréquentiel de $\frac{0.5}{2048}$ permet d'assurer la symétrie entre les quadrants Q_1 et Q_2 , Q_3 et Q_4. Ce décalage fréquentiel d'un demi-pas de quantification de la fréquence centrale($\frac{0.5}{2048}$) n'a pas d'influence sur la résolution attendue par l'architecture EFBD.

En se basant sur cette idée, nous n'avons besoin de stocker en mémoire que les valeurs du premier quadrant. Pour déterminer le nombre de bits sur lesquels ces valeurs sont quantifiées, nous avons calculé l'ENOB de l'architecture EFBD passe-bande en fonction du nombre de bits de quantification de ces valeurs en tenant compte du fait que la fréquence centrale f_c^k est codée sur 11 bits (figure 5.4). Nous pouvons noter qu'une quantification des coefficients de démodulation sur 12 bits est suffisante pour maintenir la résolution attendue par l'architecture EFBD passe-bande. La figure 5.5 montre le premier quadrant Q_1 dont les valeurs sont codées sur 12 bits.

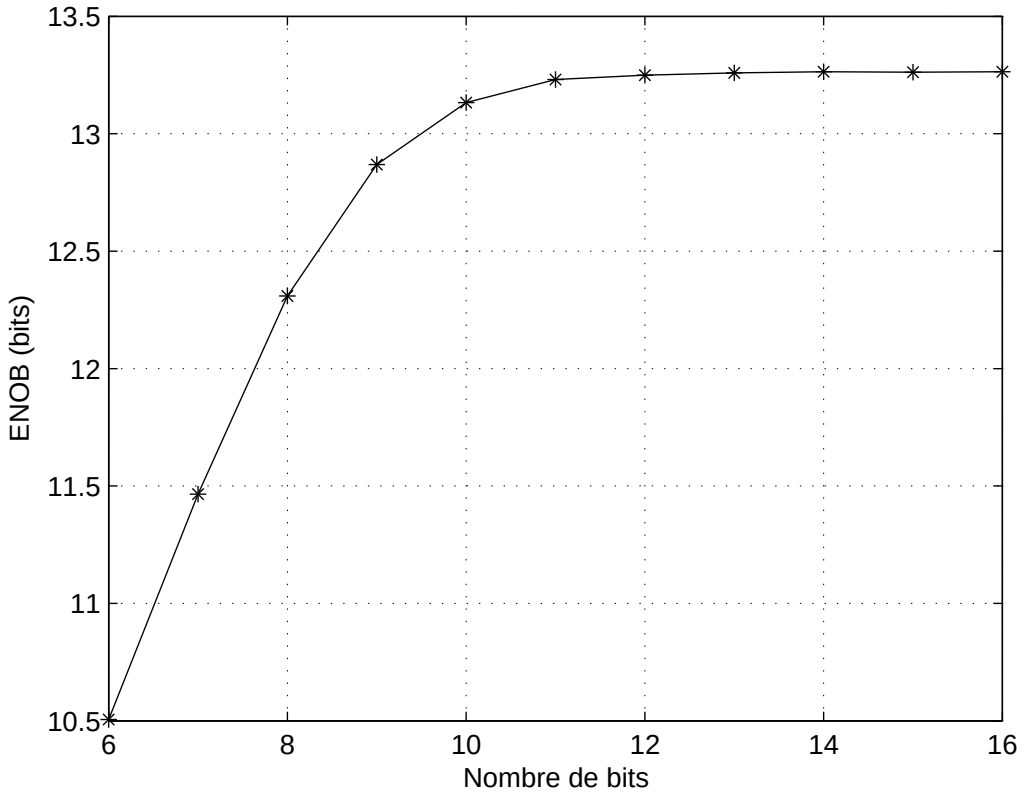

Fig. 5.4 – ENOB en fonction du nombre de bits des coefficients de démodulation.

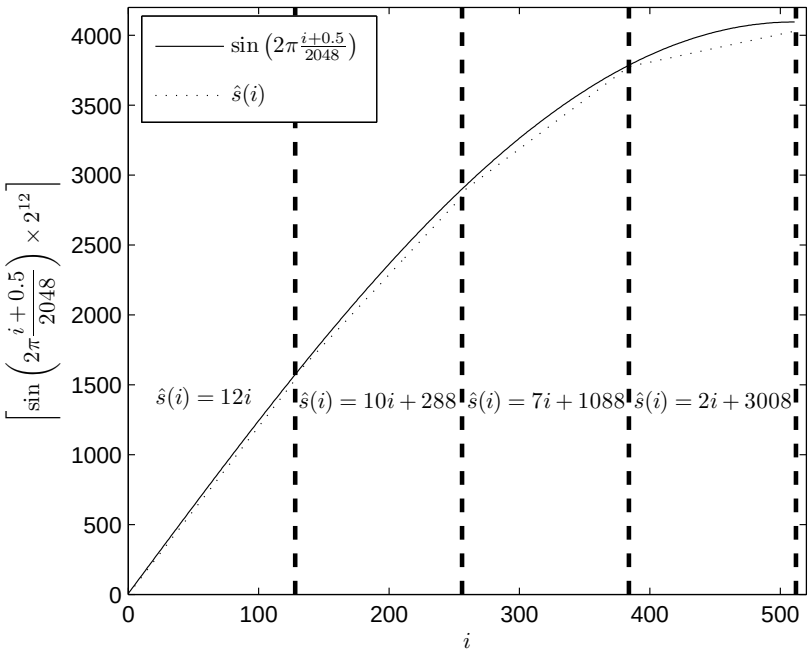

Fig. 5.5 – Valeurs du premier quadrant codées sur 12 bits ainsi que les droites d'approximation ŝ(i) dans chaque partie.

Afin d'optimiser l'espace mémoire nécessaire pour le stockage des valeurs du premier quadrant, nous avons divisé le premier quadrant en quatre parties égales. Ce choix est un compromis entre l'erreur obtenue et l'économie des ressources matérielles. Ensuite, nous avons approché, dans chacune des parties, le premier quadrant de la fonction sinusoïdale par une droite ŝ(i) (figure 5.5). L'erreur e_r(i) entre les vraies valeurs du premier quadrant et les différentes droites ŝ(i) est représenté par la figure 5.6.

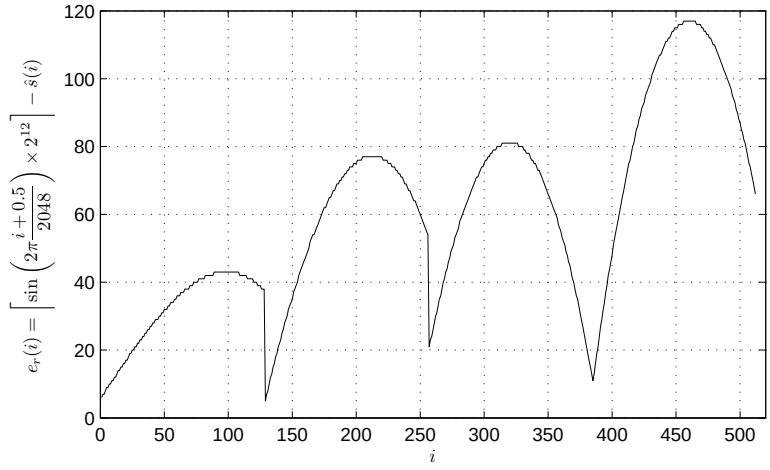

Fig. 5.6 – Différence entre $\sin\left(2\pi\frac{\mathrm{i+0.5}}{2048}\right)$ et ŝ(i) dans le premier quadrant.

Cette erreur est relativement faible par rapport aux vraies valeurs (3%). Le maximum de l'erreur est de 120, 7 bits sont donc nécessaires pour la coder ($2^7 = 128$). D'où l'idée de stocker l'erreur $e_r(i)$ dans la mémoire au lieu de la vraie valeur permettant ainsi de gagner 5 bits par valeur. La vraie valeur se déduit facilement à partir de $\hat{s}(i)+e_r(i)$. Le calcul de $\hat{s}(i)$ dans les différents intervalles ne nécessite aucune multiplication. En effet, chacune des multiplications par l'indice i peut se mettre sous la forme suivante : $12i = 8i + 4i$, $10i = 8i + 2i$, $7i = 8i - i$. Étant donné que la multiplication par un nombre en puissance de 2 n'est qu'un simple décalage, chacune des multiplications se résume à une addition.

La fréquence de démodulation f_c^k est codée sur 11 bits ($q_{f_c} = \frac{f_e}{2^{11}} = \frac{f_e}{2048}$). La détermination des coefficients de démodulation de la voie imaginaire (en quadrature de phase) à partir des valeurs du tableau du premier quadrant se fait à partir des 2 bits de poids fort. Soit $a_{10}a_9a_8 \ldots a_1a_0$ le mot binaire représentant la fréquence f_c^k. Le bit a_{10} indique le signe de la fonction sinusoïdale, si on est dans la partie positive ($a_{10} = '0'$, c'est-à-dire la valeur de f_c^k est inférieure à 1024) ou bien dans la partie négative ($a_{10} = '1'$, c'est-à-dire la valeur de f_c^k est supérieur à 1024) (voir figure 5.3). Le bit a_9 détermine si on est dans le premier quadrant Q_1 ($a_9 = '0'$) ou bien dans le quadrant Q_2 symétrique par rapport à Q_1 ($a_9 = '1'$). En ce qui concerne la voie réelle (en phase), les valeurs du cosinus sont symétriques par rapport à celles du sinus dans tous les quadrants à un signe négatif près (figure 5.3). Le signe des valeurs de démodulation de la voie réelle n'est plus déterminé par le bit a_{10} mais par $a_{10} \oplus a_9$.

La figure 5.7 et le tableau 5.1 présentent les paramètres et les ports de l'entité démodulateur.

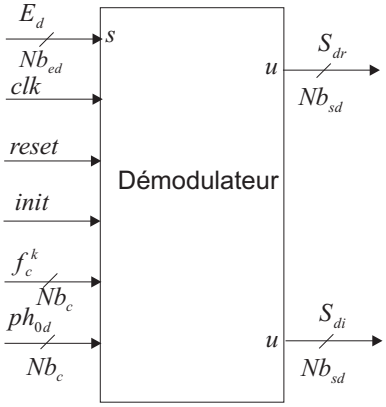

Fig. 5.7 – Schéma bloc de l'entité démodulateur.

Avec cette technique de calcul des coefficients de démodulation, nous avons réussi à diminuer les contraintes sur la capacité de la mémoire en passant de 2048 valeurs de 12 bits à 512 valeurs de 7 bits.

La multiplication par la séquence m_k est simple à réaliser car le signal en sortie du modulateur est un signal sur $Nb_{ed} = 4$ bits (1 bit de signe). En tenant compte du fait que les valeurs de la séquence de démodulation sont codées sur 12 bits, le signal en sortie est un signal signé (*signed*) de 16 bits. Comme la représentation en nombre non signé (*unsigned*) exige moins de ressources matérielles, nous avons adopté une représentation non signée en ajoutant 2^{15} ($2^{Nb_{sd}-1}$) au signal de sortie afin de faciliter la réalisation des blocs qui suivent (décimateur, filtre passe-bas). Au niveau matériel, cet ajout n'est qu'une inversion du bit de poids fort du signal signé. Cet ajout sera retranché avant le filtre de calibration dans l'architecture globale pour enlever la composante continue. Cette suppression tiendra compte des gains des blocs qu'elle traverse.

Tab. 5.1 – Paramètres et ports d'entrées/sorties du démodulateur.

Démodulateur	
Paramètres de configuration	
Nb_{ed}	Nombre de bits du signal d'entrée.
Nb_{sd}	Nombre de bits du signal de sortie.
Nb_c	Nombre de bits de la fréquence centrale et de la phase d'initialisation.
Ports d'entrées/sorties	
clk	Horloge de commande de tous les registres.
reset	Entrée de remise à zéro de tous les registres.
init	Entrée de validation de lecture de la phase initiale ph_{0d}.
E_d	Entrée signée du démodulateur.
f_c^k	Fréquence de démodulation.
ph_{0d}	Phase initiale de démodulation.
S_{dr}, S_{di}	Sorties réelle et imaginaire non signée.

5.3 Filtre en peigne

La décimation d'un facteur R_d consiste à diminuer la cadence d'un signal en ne conservant qu'une valeur sur R_d valeurs à l'entrée du décimateur. Nous avons vu au § 3.3.2 que la décimation se manifeste dans le domaine fréquentiel par une répétition du spectre du signal à décimer tous les $\frac{k}{R_d}$ $(0 \leq k \leq R_d - 1)$. Il est donc indispensable de filtrer avant de décimer afin de limiter le spectre du signal d'entrée à $\left[-\frac{1}{2R_d} \dots \frac{1}{2R_d}\right]$ et d'éviter par conséquent le repliement spectral autour des fréquences $\frac{k}{R_d}$. Un filtre FIR peut être envisagé pour réaliser ce filtrage. Étant donné que les signaux à traiter par ce filtre ont une fréquence élevée (fréquence de suréchantillonnage 800 MHz), les structures les plus simples ne nous permettraient pas d'atteindre la fréquence de travail souhaitée. Il faut donc choisir des structures plus complexes (parallélisation, pipelinage), ce qui conduit à des circuits dont la surface est accrue et donc plus coûteux.

Le facteur de suréchantillonnage étant élevé (OSR = 40), le signal utile se trouve dans une bande de largeur Δf_k très faible par rapport à $\frac{1}{2R_d}$. De ce fait, le filtre passe-bas le plus adapté pour cette application est un filtre moyenneur ou un filtre en peigne. La fonction de transfert du filtre en peigne d'ordre K, où K est le nombre de moyenneurs en cascade, est donnée par :

$$C(z) = \left(\frac{1}{R_d}\sum_{i=0}^{R_d-1} z^{-i}\right)^K = \left(\frac{1}{R_d}\frac{1-z^{-R_d}}{1-z^{-1}}\right)^K \tag{5.1}$$

Le filtre moyenneur possède les deux propriétés suivantes qui font de lui un bon candidat pour ce type d'application :

- L'implantation de ce filtre est très simple. Il ne nécessite aucune opération de multiplication. Sa réalisation exige un registre à décalage et des additionneurs qui peuvent fonctionner à haute fréquence.
- Le module de la réponse fréquentielle $\left|C\left(e^{j2\pi f}\right)\right| = \left|\frac{\sin(\pi f R_d T_e)}{R_d \sin(\pi f T_e)}\right|^K$ possède des zéros aux mêmes endroits que les répliques du spectre du signal utile décimé (figure 5.8). Ce qui permet d'atténuer les repliements dans le spectre du signal utile à ces endroits.

La figure 5.8 montre le module de la réponse fréquentielle pour les deux ordres K = 2 et K = 4. Nous remarquons que plus l'ordre est grand plus l'atténuation des repliements spectraux est forte. Par contre, l'augmentation de l'ordre atténue également davantage le spectre du signal

utile se situant vers les basses fréquences. Cette atténuation est corrigée, comme nous l'avons indiqué au § 3.4.2, par un filtre de correction passe-bas $C_2(z)$ à trois coefficients.

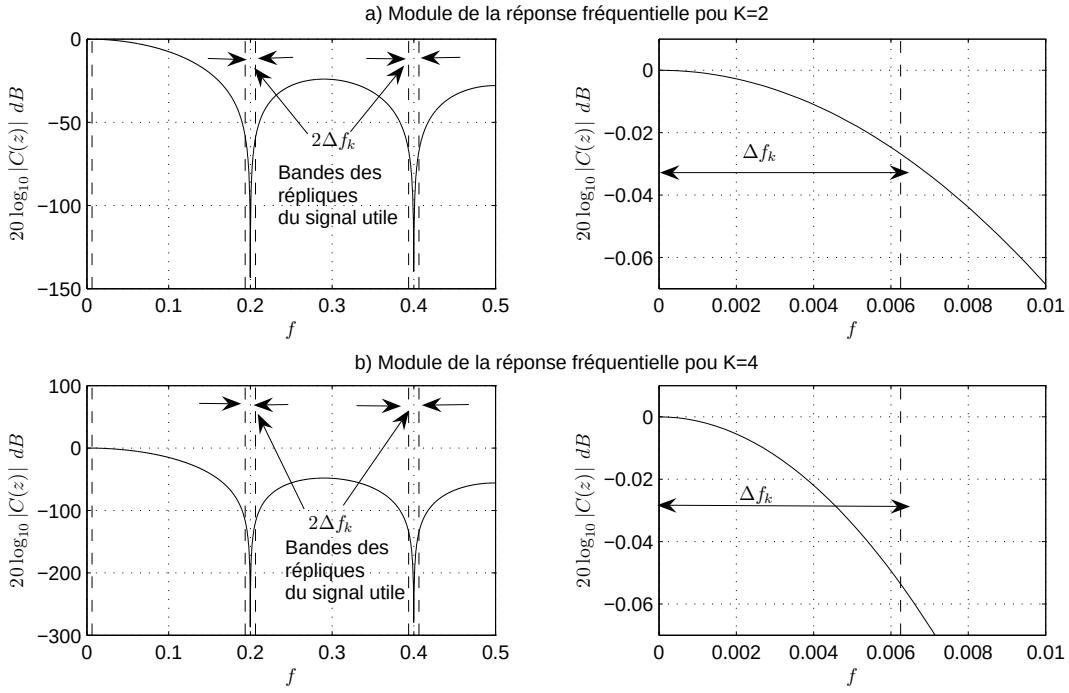

Fig. 5.8 – Module de la réponse fréquentielle des filtres C(z) pour les ordres a) $K = 2$ et b) $K = 4$.

À la sortie du modulateur ΣΔ, le spectre du signal utile est limité par $\pm\Delta f_k$. Ce spectre contient le signal utile plus le bruit de quantification modulé par la NTF de ce modulateur. Pour que le bruit introduit par les repliements spectraux, lors du processus de décimation, soit inférieur au bruit de quantification déjà présent, il suffit que le filtre en peigne possède des zéros d'ordre $K \geq \frac{L}{2} + 1$ aux fréquences $\frac{k}{R_d}$ $(0 \leq k \leq R_d - 1)$ [58], où L est l'ordre du modulateur ΣΔ passe-bande. Dans notre cas, on prendra $K = 4$ pour un modulateur passe-bas d'ordre 6.

Architecture de réalisation

La réalisation du filtre moyenneur C(z) peut se faire de façon directe comme le montre la figure 5.9.

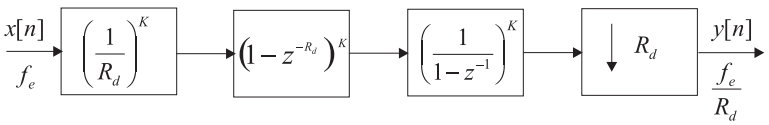

Fig. 5.9 – Réalisation récursive directe du filtre moyenneur C(z).

Cette architecture nécessite la réalisation du filtre dérivateur $\left(1 - z^{-R_d}\right)^K$ qui nécessite $R_d \times K$ registres. Cette architecture peut être simplifiée en changeant l'emplacement de réalisation du

filtre dérivateur, ce qui ne change pas la fonction de transfert globale [27]. En effet, une diminution de la cadence du signal d'un facteur R_d se traduit dans le domaine fréquentiel par la substitution $z^{R_d} \rightarrow z$. De ce fait, la réalisation du numérateur $\left(1 - z^{-R_d}\right)^K$ avant la décimation est équivalente à la réalisation du terme $\left(1 - z^{-1}\right)^K$ après la décimation. En se basant sur cette propriété, le filtre en peigne peut être réalisé de manière efficace en séparant l'équation (5.1) en un numérateur et un dénominateur :

$$C(z) = \left(\frac{1}{1 - z^{-1}}\right)^K \left(1 - z^{-R_d}\right)^K \left(\frac{1}{R_d}\right)^K \tag{5.2}$$

et en déplaçant la partie numérateur après la décimation [59](voir figure 5.10).

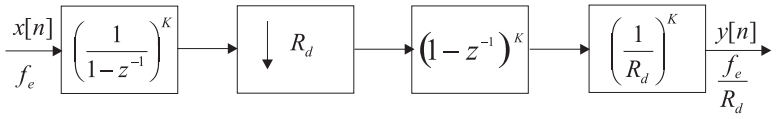

Fig. 5.10 – Réalisation récursive optimale du filtre moyenneur C(z).

L'architecture du filtre tel qu'il est représenté sur la figure 5.10 réduit la complexité par rapport à la réalisation illustrée par la figure 5.9. En effet, le terme dérivateur exige $(R_d - 1) \times K$ registres en moins et il fonctionne à une cadence inférieure à celle qui se trouve à la sortie du modulateur. La figure 5.11 et le tableau 5.2 présentent les paramètres et les ports de l'entité filtre décimateur.

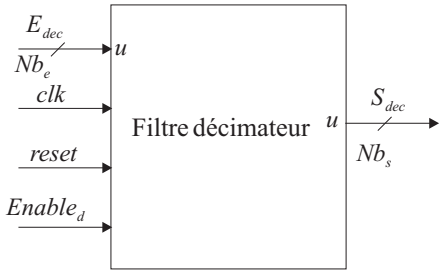

Fig. 5.11 – Schéma-bloc de l'entité filtre décimateur.

Tab. 5.2 – Paramètres et ports d'entrées/sorties du filtre décimateur.

Filtre décimateur	
Paramètres de configuration	
Nb_e	Nombre de bits du signal d'entrée.
Nb_s	Nombre de bits du signal de sortie.
Nb_i	Nombre de bits des registres internes.
Ports d'entrées/sorties	
clk	Horloge de commande de tous les registres du filtre.
reset	Entrée de remise à zéro des différents registres du filtre décimateur.
$Enable_d$	Entrée de verrouillage des registres du dérivateur.
E_{dec}	Entrée non signée du filtre décimateur.
S_{dec}	Sortie non signée du filtre.

La figure 5.12 montre l'architecture de réalisation du filtre décimateur.

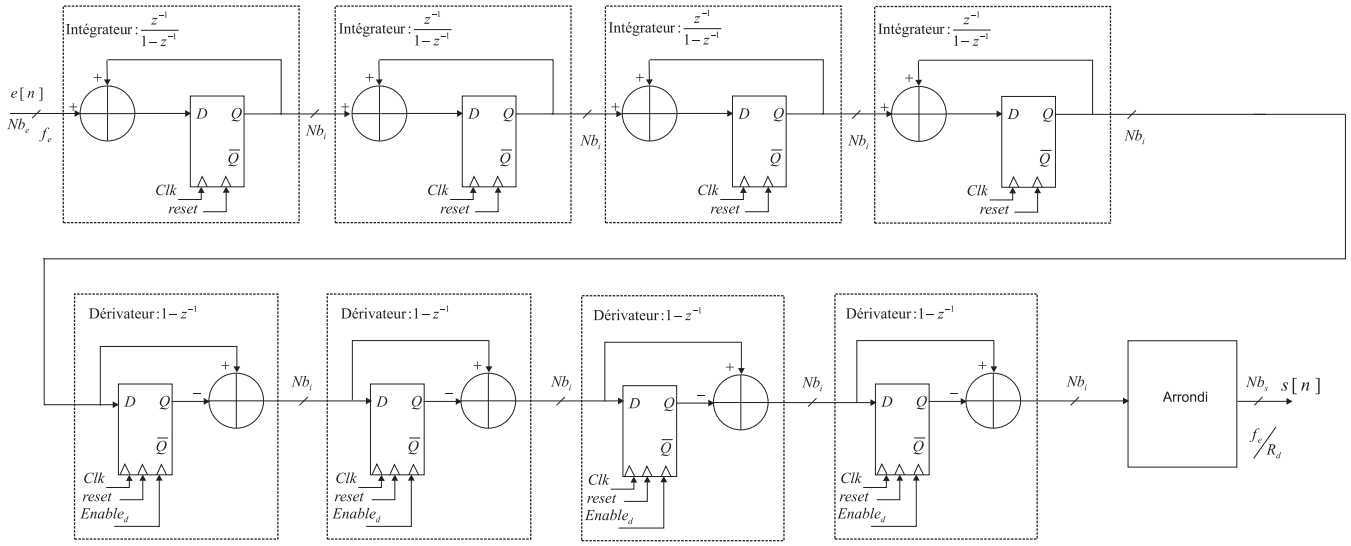

Fig. 5.12 – Réalisation récursive optimale.

Il comprend :

- 4 intégrateurs successifs de type $\frac{z^{-1}}{1-z^{-1}}$ avant la décimation (K = 4). Plutôt que de réaliser des intégrateurs de fonction de transfert $\frac{1}{1-z^{-1}}$, nous avons choisi d'ajouter un retard et de réaliser la fonction de transfert $\frac{z^{-1}}{1-z^{-1}}$ pour simplifier la réalisation. La dynamique du signal augmente au fur et à mesure qu'il traverse les intégrateurs. Pour cela, il faut déterminer la valeur optimale de la taille des registres internes pour ne pas avoir de dépassement de la capacité de calcul. En partant de l'expression de la fonction de transfert du filtre réalisée sur la figure 5.12, nous pouvons écrire :

$$(R_d)^K C(z) = \left[\sum_{i=0}^{R_d-1} z^{-i}\right]^K \leq \left|\sum_{i=0}^{R_d-1} z^{-i}\right|^K \leq \left[\sum_{i=0}^{R_d-1} \left|z^{-i}\right|\right]^K \leq [R_d]^K \tag{5.3}$$

  L'expression ci-dessus montre que le gain maximum apporté par le filtre en peigne est égal à $[R_d]^K$. En tenant compte de ce gain, si le signal d'entrée est codé sur un nombre de bits Nb_e, la taille des registres Nb_i doit vérifier la relation :

$$Nb_i = Nb_e + \lceil K \log_2 (R_d) \rceil \tag{5.4}$$

- 4 dérivateurs successifs fonctionnant à une vitesse R_d fois plus faible que les intégrateurs grâce au signal de verrouillage des registres $Enable_d$,
- 1 bloc permettant de diminuer la taille des données en sortie. À l'inverse des intégrateurs, les dérivateurs font diminuer l'amplitude du signal les traversant. De ce fait, il est inutile de garder le signal en sortie sur Nb_i bits. Il faut arrondir le signal de sortie pour avoir le même nombre de bits que le signal d'entrée. Cette opération d'arrondi consiste à garder les Nb_s bits de poids forts puis à ajouter le bit qui suit la troncature :

$$\overbrace{a_{Nb_i-1}a_{Nb_i-2}\ldots a_{Nb_i-Nb_s}}^{Nb_s} a_{Nb_i-Nb_s-1}\ldots a_1a_0 + \underbrace{00\ldots00}_{Nb_s-1} a_{Nb_i-Nb_s-1} = r_{Nb_s-1}r_{Nb_s-2}\ldots r_1r_0 \tag{5.5}$$

Dans cette architecture, le coefficient multiplicatif $\left(\frac{1}{\mathrm{R_d}}\right)^{\mathrm{K}}$ dans C(z) (équation 5.2) n'est pas pris en compte. Ce coefficient sera intégré dans le calcul des coefficients du filtre FIR $\mathrm{G_{pb}^{k}(z)}$ (voir § 5.4).

5.4 Filtre numérique $G_{pb}^{k}(z)$

Le filtre FIR passe-bas $\mathrm{G_{pb}^{k}(z)}$ est composé de deux filtres : le filtre $\mathrm{F_{pb}^{k}(z)}$ pour couper le bruit de quantification en dehors de la bande utile et le filtre $\mathrm{C_2(z)}$ pour corriger l'atténuation introduite par le filtre en peigne vers les basses fréquences. Nous avons vu au § 3.5 (tableau 3.2) que l'ordre du filtre $\mathrm{F_{pb}^{k}(z)}$ nécessaire pour avoir l'ENOB en sortie est de 64. Comme le filtre $\mathrm{C_2(z)}$ est à 3 coefficients, le filtre $\mathrm{G_{pb}^{k}(z)}$ possède 67 coefficients.

Les filtres $\mathrm{G_{pb}^{k}(z)}$ n'ont pas tous la même largeur de bande. Ils ne possèdent pas à la fréquence de coupure un gain normalisé $\mathrm{g_{f_c}}$ de 0.5. De ce fait, la somme des réponses fréquentielles des différents filtres $\mathrm{G_{pb}^{k}(z)}$ pour tous les étages ne présente pas un gain de 0 dB dans toute la bande de fonctionnement comme le montre la figure 5.13 (a). L'idée la plus simple pour corriger cet effet et avoir un gain de 0.5 à la fréquence de coupure est de diviser les coefficients du filtre par $2 \times \mathrm{g_{f_c}}$. La figure 5.13 (b) montre l'effet de la division par $2 \times \mathrm{g_{f_c}}$ sur la suppression des bosses dues aux filtres de faibles largeurs de bande. Le gain $\mathrm{g_{f_c}}$ dépend de la largeur de bande du filtre $\mathrm{G_{pb}^{k}(z)}$.

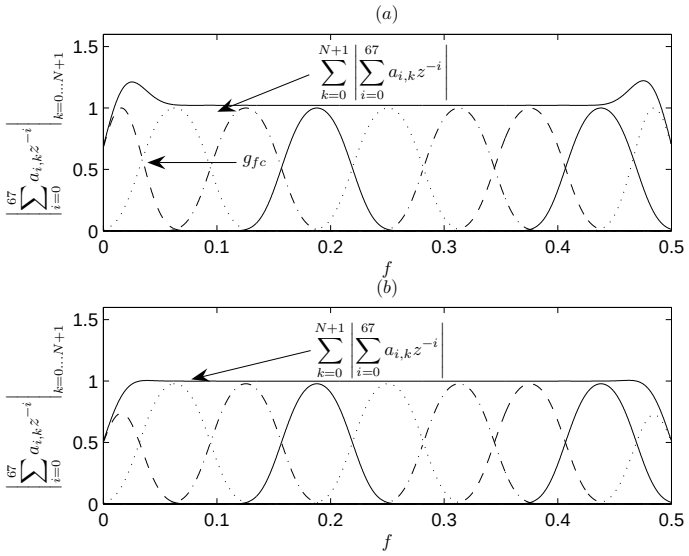

Fig. 5.13 – (a) Module de la réponse fréquentielle des filtres $\mathrm{G_{pb}^{k}(z)}$, (b) somme de la réponse fréquentielle de tous les filtres passe-bas.

Une démarche importante consiste à évaluer sur combien de bits peuvent être représentés les coefficients de ce filtre sans avoir une chute notable de l'ENOB. La figure 5.14 présente l'ENOB calculé par simulation en faisant varier la largeur des mots binaires des coefficients $\mathrm{Nb_{coef}}$. Nous constatons qu'à partir de $\mathrm{Nb_{cf}} = 11$ bits, l'ENOB est proche de sa valeur théorique de 13.3 bits. Donc nous allons choisir de représenter les coefficients du filtre $\mathrm{G_{pb}^{k}(z)}$ sur 11 bits.

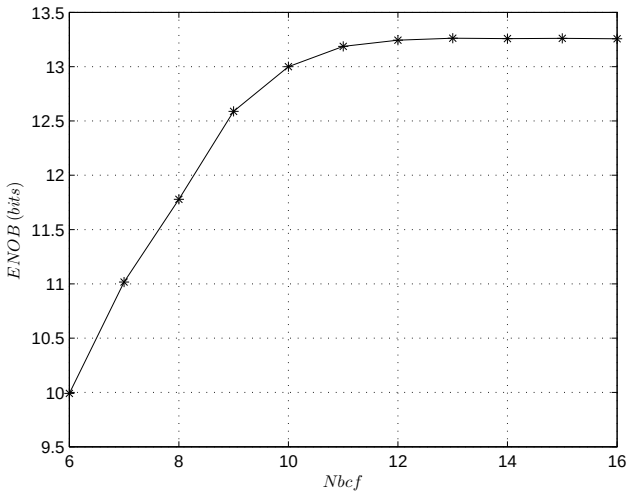

Fig. 5.14 – ENOB en fonction du nombre de bits des coefficients du filtre $G^{k}_{pb}(z)$.

Dans la suite, nous définissons l'architecture de réalisation optimale du filtre $G^{k}_{pb}(z)$. Cette optimisation doit tenir compte de deux critères :

- la vitesse de fonctionnement : elle est égale dans l'exemple considéré tout au long de cette thèse à 800 MHz (voir tableau 3.1).
- la surface d'implantation.

5.4.1 Architecture de réalisation

Le filtre passe-bas $G^{k}_{pb}(z)$ est un filtre FIR de type I. Ce filtre est caractérisé par :

- une réponse impulsionnelle présentant un axe de symétrie coïncidant avec le coefficient central comme le montre la figure 5.15,
- un nombre de coefficients impairs.

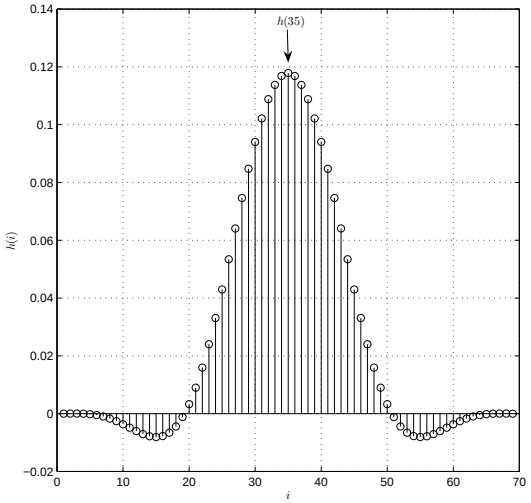

Fig. 5.15 – Réponse impulsionnelle du filtre $G^{k}_{pb}(z)$.

La fonction de transfert du filtre $G_{pb}^{k}(z)$ est donnée par :

$$G_{pb}^{k}(z) = \sum_{i=0}^{P} h(i) z^{-i} \quad \text{avec} \quad \text{P} \quad \text{impair} \tag{5.6}$$

Pour un signal E_f en entrée de ce filtre, la sortie S_f à l'instant n s'exprime par :

$$S_f[\text{n}] = \sum_{\text{i=0}}^{\text{P}} h(i) E_f[n-i] \tag{5.7}$$

Compte tenu de la symétrie des coefficients h du filtre $G_{pb}^{k}(z)$, la sortie s(n) peut être ré-écrite :

$$S_f[\text{n}] = \sum_{\text{i=0}}^{\frac{\text{P}-3}{2}} h(i)\left[E_f(n-i) + E_f(n-(P-1-i))\right] + h\left(\frac{\text{N}-1}{2}\right) E_f\left(n - \frac{\text{N}-1}{2}\right) \tag{5.8}$$

En se basant sur l'expression ci-dessus, il est possible d'optimiser l'architecture de réalisation du filtre FIR en partageant les ressources des multiplieurs jusqu'à diminuer leur nombre pratiquement de moitié comme le montre la figure 5.16.

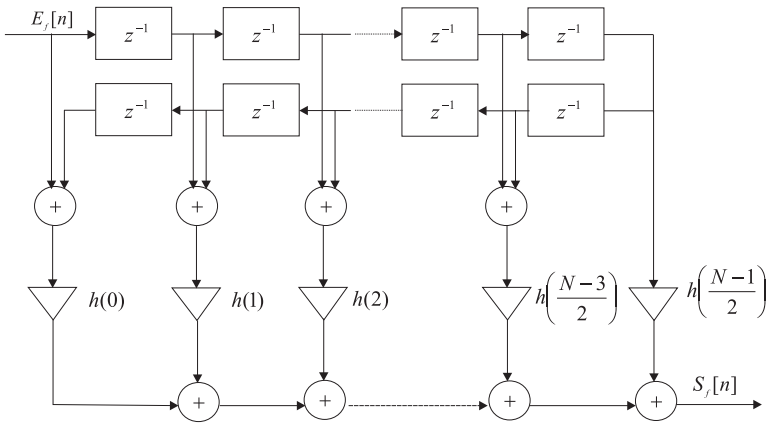

FIG. 5.16 – Architecture de réalisation du filtre FIR de type I.

Dans notre cas, le filtre $G_{pb}^{k}(z)$ a 67 coefficients. En principe, 33 multiplieurs seraient nécessaires pour implanter l'architecture parallèle (figure 5.16) au lieu de 67 avec une réalisation directe. Cependant, l'implantation de 33 multiplieurs avec une vitesse de fonctionnement de 160 MHz exige beaucoup de ressources matérielles et rend notre circuit encombrant. Pour résoudre ce problème, nous prenons en considération le fait que le filtre $G_{pb}^{k}(z)$ se situe après le filtre décimateur qui diminue la cadence du signal d'un facteur $R_d = 5$. Ainsi, l'opération de filtrage (équation 5.8) peut s'effectuer entre deux valeurs successives avec une vitesse 5 fois plus grande. Ceci permet de répartir le calcul de 33 multiplications en 5 coups d'horloge en effectuant 7 multiplications à chaque coup d'horloge. La nouvelle architecture que nous proposons avec la diminution de cadence du signal par le décimateur est présentée sur la figure 5.17.

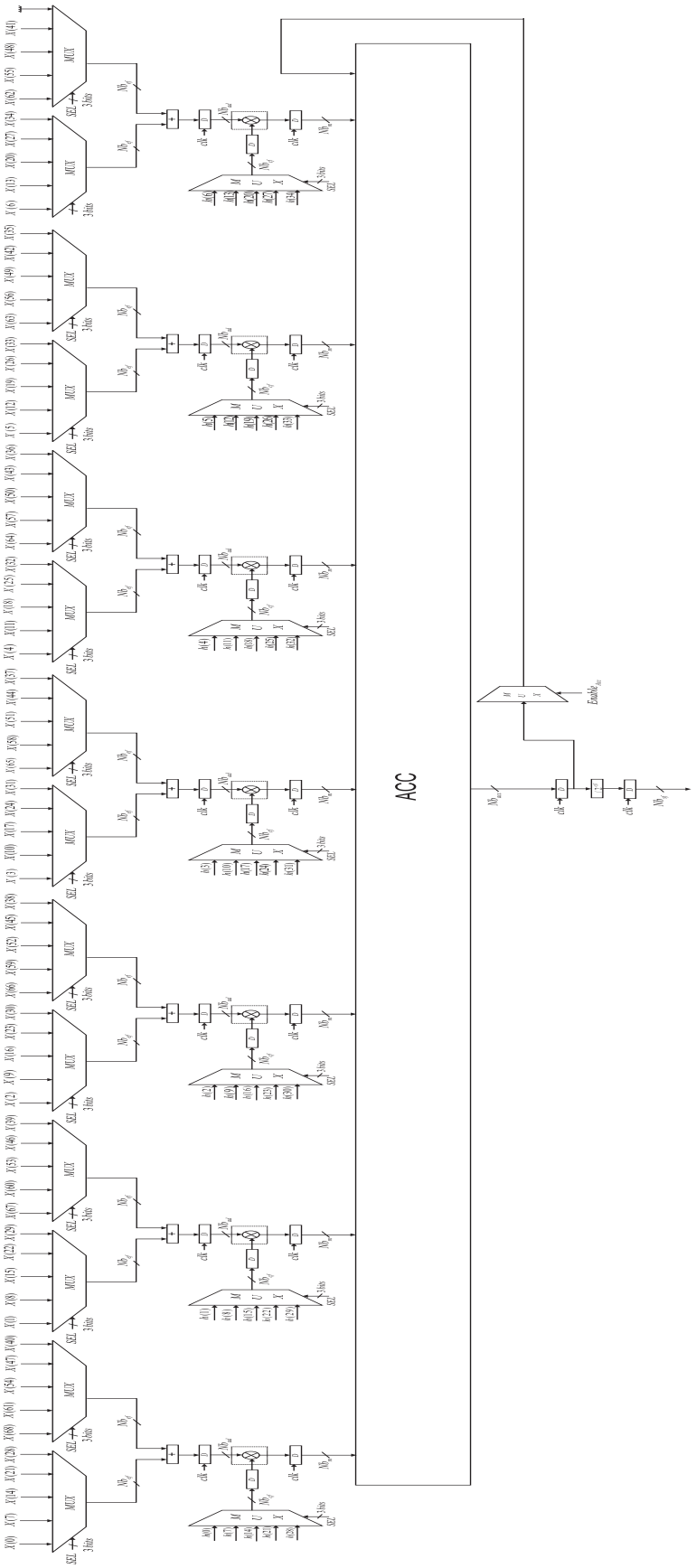

Fig. 5.17 – Architecture optimale de réalisation du filtre $G^{k}_{pb}(z)$ avec 7 multiplieurs.

Cette architecture optimise la surface de silicium nécessaire pour l'implantation en diminuant de 33 à 7 le nombre de multiplieurs. Le nombre maximum de multiplications que l'on peut effectuer avec cette architecture est de 35. Afin d'exploiter au maximum ces ressources, nous avons augmenté l'ordre du filtre $F^k_{pb}(z)$ à 66 au lieu de 64 pour obtenir un filtre $G^k_{pb}(z)$ à 69 coefficients. Cette architecture comporte :

1. Un registre à décalage permettant de sauvegarder les P dernières valeurs du signal d'entrée à une cadence 5 fois plus faible que la vitesse de traitement du filtre $G^k_{pb}(z)$(figure 5.18).

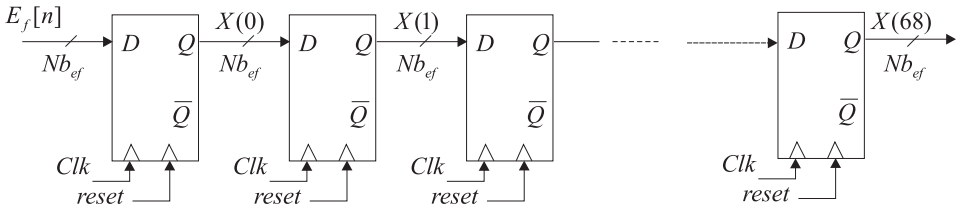

FIG. 5.18 – Registre à décalage du signal d'entrée.

2. Des multiplexeurs permettant de sélectionner les entrées X(i) et les coefficients h(i) convenables pour effectuer le calcul à travers un signal de commande (SEL) sur 3 bits. Ce signal sera délivré par une unité de commande réalisée avec une machine d'état.
3. Une mémoire ROM où sont stockés les coefficients des 26 filtres $G^k_{pb}(z)$ calculés pour les largeurs de bandes $L_B = \frac{i}{1024}\big|_{i=0...25}$ (voir § 4.5.2).
4. Des additionneurs pour effectuer la somme $X(n-i) + X(n-(P-1-i))$.
5. Des multiplieurs pour réaliser le produit $h(i)\,[X(n-i) + X(n-(P-1-i))]$. Au cours de la synthèse de ce filtre, nous avons constaté qu'il ne peut pas fonctionner à la fréquence de 800 MHz car le multiplieur possède un chemin critique (Le chemin critique est le chemin de propagation qui limite la fréquence maximale de fonctionnement ; il dépend du retard introduit par chaque composant) plus grand que la période de calcul. La vitesse de fonctionnement de 800 MHz peut être atteinte en utilisant le pipelinage. Le pipelinage consiste à couper le chemin critique du multiplieur en parallélisant le calcul. La multiplication de deux nombres peut être décomposée en plusieurs multiplications partielles possédant des chemins critiques plus petits que la période de fonctionnement. En effet soit A et B les deux opérandes de multiplication représentés sur des mots binaires de longueurs respectives N et N′ comme le montre l'expression 5.9.

$$A = \overbrace{a_{N-1}\dots a_{N_1}}^{A_H}\underbrace{a_{N_1-1}\dots a_0}_{A_L} \quad , \quad B = \overbrace{b_{N'-1}\dots b_{N'_1}}^{B_H}\underbrace{b_{N'_1-1}\dots b_0}_{B_L} \tag{5.9}$$

La multiplication A × B peut être écrite de la façon suivante :

$$\begin{aligned} A \times B &= \left(A_L + A_H \times 2^{N_1}\right) \times \left(B_L + B_H \times 2^{N'_1}\right) \\ &= A_L B_L + A_L B_H \times 2^{N'_1} + A_H B_L \times 2^{N_1} + A_H B_H \times 2^{N_1+N'_1} \end{aligned}$$

Chacune des multiplications partielles est ainsi réalisée avec des opérandes de plus petites tailles et possède par conséquent un chemin critique inférieur à la période de fonctionnement. Le pipelinage nous a permis de faire fonctionner le multiplieur à 800 MHz, même

s'il a ajouté un retard d'un coup d'horloge qui n'influe pas sur le fonctionnement global du filtre car le calcul est synchrone.

6. Un accumulateur pour faire la somme des résultats de 7 multiplications obtenues au bout de 5 coups d'horloge.
7. Un multiplexeur commandé par le signal $\text{Enable}_{\text{acc}}$ qui sert à mémoriser l'état de l'accumulateur au cours du calcul de la sortie du filtre pendant les 5 coups d'horloge et à remettre à zéro cette mémoire au début du calcul de la nouvelle sortie.
8. Une division par le coefficient 2^{cfi}. En effet, les coefficients du filtre FIR sont inférieurs à 1, leur somme est égale à l'unité $\sum_{\text{i=0}}^{\text{P}} \text{h(i)} = 1$ pour avoir un gain de 0 dB à la fréquence nulle. Un nombre réel A inférieur à 1 est représenté en notation en virgule fixe sur N bits par :

$$A|_b = a_{-1}2^{-1} + a_{-2}2^{-2} + \ldots + a_{-N}2^{-N}$$

Afin de déterminer la taille adaptée de ces coefficients et d'éviter au maximum des zéros au début des mots binaires représentant les coefficients, nous multiplions les coefficients, avant de les stocker en mémoire, par 2^{cfi} où cfi est le nombre de zéros précédant le premier bit non nul. il est égal à $\text{cfi} = \left\lfloor -\log_2 \left(\max \left(\text{h(i)}|_{\text{i=0...P}-1} \right) \right) \right\rfloor$. À titre d'exemple, le nombre 0.0605 est représenté sur 8 bits par $0.\overbrace{\underbrace{0000}_{cfi=4}1111}^{\text{N=8}} \equiv 0.0586$. Par contre si on multiplie par $2^{\text{cfi}} = 16$, le nombre 0.9680 (0.0605×16) est représenté sur 8 bits par $0.\underbrace{11111000}_{\text{N=8}} \equiv 0.9688$. On voit que la multiplication par 16, équivalente à un décalage de 4 bits à gauche du mot binaire représentant 0.0605, nous a permis de gagner 1 bit de précision sur la représentation du nombre en obtenant 5 bits de $'1'$ au lieu de 4.

Dans notre architecture, la division par 2^{cfi} n'exige aucune resource supplémentaire car ce n'est qu'un décalage à droite de cfi bits.

La figure 5.19 et le tableau 5.3 présentent les paramètres et les ports de l'entité du filtre $\text{G}^{\text{k}}_{\text{pb}}(\text{z})$.

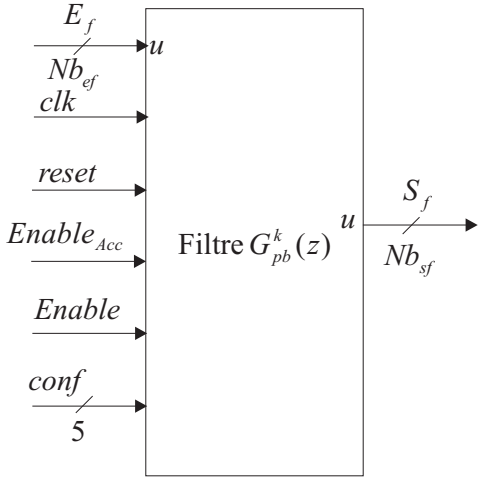

Fig. 5.19 – Schéma-bloc de l'entité filtre $\text{G}^{\text{k}}_{\text{pb}}(\text{z})$.

Tab. 5.3 – Paramètres et ports d'entrées/sorties du filtre $G_{pb}^{k}(z)$.

Filtre $G_{pb}^{k}(z)$	
Paramètres de configuration	
Nb_{ef}	Nombre de bits du signal d'entrée.
Nb_{ad}	Nombre de bits du signal à la sortie de l'additionneur.
Nb_{cf}	Nombre de bits des coefficients du filtre.
Nb_{m}	Nombre de bits de poids fort restant à la sortie de multiplieur.
Nb_{acc}	Nombre de bits de l'accumulateur pour faire l'addition des différents résultats sans être saturé.
Nb_{sf}	Nombre de bits du signal de sortie.
Ports d'entrées/sorties	
clk	Horloge de commande de tous les registres du filtre.
reset	Entrée de remise à zéro des différents registres.
Enable	Entrée de verrouillage des registres à décalage contrôlant le débit du signal d'entrée.
$Enable_{acc}$	Entrée de remise à zéro de la mémoire de l'accumulateur à la fin du calcul.
conf	Entrée de sélection des coefficients du filtre souhaité à partir de la ROM .
E_f	Entrée non signée du filtre.
S_f	Sortie non signée du filtre.

5.4.2 Optimisation

L'implantation du filtre numérique est sujette à trois sources d'erreur qui peuvent limiter sa performance. Ces sources d'erreur sont :

– **L'erreur de quantification du signal d'entrée** : La quantification du signal d'entrée $e_f[n]$ introduit une erreur uniformément distribuée $\varepsilon_{ef}[n]$. Elle est considérée comme un bruit blanc uniformément distribué dans $\left[-\frac{q_{ef}}{2} \dots \frac{q_{ef}}{2}\right]$ de moyenne nulle et de variance $\frac{q_{ef}^2}{12}$ (q_{ef} est le pas quantification) qui se superpose au signal d'entrée [60, 61, 62]. Le signal en sortie est alors la somme de deux signaux :

$$\begin{aligned} s_f[n] &= \sum_{i=0}^{P} h(i)\left[e_f(n-i) + \varepsilon_{ef}(n-i)\right] \\ &= \sum_{i=0}^{P} h(i) e_f(n-i) + \sum_{i=0}^{P} h(i)\varepsilon_{ef}(n-i) \\ &= y[n] + y_{\varepsilon_{ef}}[n] \end{aligned} \tag{5.10}$$

Le signal $y[n]$ correspond au signal d'entrée non quantifié $e_f[n]$ et $y_{\varepsilon_{ef}}[n]$ correspond au bruit de quantification $\varepsilon_{ef}[n]$. Le terme $y_{\varepsilon_{ef}}[n]$ est aléatoire car c'est la somme de $P+1$ termes aléatoires ε_{ef}. En se basant sur le théorème de la limite centrale, $y_{\varepsilon_{ef}}[n]$ a une loi gaussienne, donc il suffit de déterminer sa moyenne et sa variance pour définir cette loi.

Moyenne :

$$E\left[y_{\varepsilon_{ef}}[n]\right] = E\left[\sum_{i=0}^{P} h(i)\varepsilon_{ef}(n-i)\right] = \sum_{i=0}^{P} h(i) E\left[\varepsilon_{ef}(n-i)\right] = 0$$

Variance :

$$E\left[y_{\varepsilon_{ef}}^2[n]\right] = E\left[\sum_{i=0}^{P} h(i)\varepsilon_{ef}(n-i)\sum_{t=0}^{P} h(t)\varepsilon_{ef}(n-t)\right] = \frac{q_{ef}^2}{12}\sum_{i=0}^{P} h^2(i) \tag{5.11}$$

Comme le gain du filtre est normalisé à 0 dB ($\sum_{\mathrm{i=0}}^{\mathrm{P}} \mathrm{h(i)} = 1$), la variance $\mathrm{E}\left[\mathrm{y}^2_{\varepsilon_{\mathrm{ef}}}[\mathrm{n}]\right]$ est majorée :

$$E\left[y^2_{\varepsilon_{ef}}[n]\right] = \frac{q^2_{ef}}{12}\sum_{i=0}^{P} h^2(i) \leq \frac{q^2_{ef}}{12} \tag{5.12}$$

La puissance de bruit maximale introduite par la quantification du signal d'entrée est égale à $\frac{\mathrm{q}^2_{\mathrm{ef}}}{12}$. Cette source d'erreur est faible pour un nombre de bits assez élevé.

– **L'erreur de quantification des coefficients du filtre** : la quantification des coefficients du filtre introduit une erreur sur le module de la réponse fréquentielle du filtre et par conséquent influe sur l'atténuation du bruit en dehors de la bande utile. La réponse fréquentielle du filtre $\mathrm{G}^{\mathrm{k}}_{\mathrm{pb}}(\mathrm{z})$ est donnée par :

$$\begin{aligned} G^k_{pb}\left(e^{jw}\right) &= \sum_{n=0}^{\frac{P-3}{2}} h(n)\left[e^{-jwn} + e^{-jw(P-1-n)}\right] + h\left(\frac{P-1}{2}\right)e^{-jw\left(\frac{P-1}{2}\right)} \\ &= \left[\sum_{n=0}^{\frac{P-3}{2}} 2h(n)\cos\left(\frac{P-1}{2}-n\right)w + h\left(\frac{P-1}{2}\right)\right]e^{-jw\left(\frac{P-1}{2}\right)} \end{aligned} \tag{5.13}$$

On peut noter que, dans un filtre à coefficients symétriques, les coefficients h(n) n'influent que sur le module de la réponse fréquentielle et non pas sur sa phase. La quantification des coefficients h(n) introduit une erreur de quantification $\varepsilon_{\mathrm{c}}[\mathrm{n}]$ blanche uniformément distribuée dans l'intervalle $\left[-\frac{\mathrm{q_c}}{2}, \frac{\mathrm{q_c}}{2}\right]$ de moyenne nulle et de variance $\frac{\mathrm{q}^2_{\mathrm{c}}}{12}$, $\mathrm{q_c}$ est le pas de quantification. Les coefficients quantifiés $\mathrm{h}^*[\mathrm{n}]$ sont exprimés par $\mathrm{h}^*[\mathrm{n}] = \mathrm{h}[\mathrm{n}] + \varepsilon_{\mathrm{c}}[\mathrm{n}]$ où $\mathrm{h}[\mathrm{n}]$ sont les coefficients idéaux et $\varepsilon_{\mathrm{c}}[\mathrm{n}]$ l'erreur de quantification. En remplaçant les coefficients du filtre h(n) par leurs valeurs quantifiées $\mathrm{h}^*[\mathrm{n}]$, le module de la réponse fréquentielle est la somme de deux termes : le module de la réponse idéale $\mathrm{G}^{\mathrm{k}}_{\mathrm{pb}}\left(\mathrm{e}^{\mathrm{jw}}\right)$ du filtre et le module de la réponse fréquentielle $\mathrm{E_c}\left(\mathrm{e}^{\mathrm{jw}}\right)$ dû aux erreurs de quantification :

$$\begin{aligned} \left|G^k_{pb}\left(e^{jw}\right)\right|^* &= \left[\sum_{n=0}^{\frac{P-3}{2}} 2h^*(n)\cos\left(\frac{P-1}{2}-n\right)w + h^*\left(\frac{P-1}{2}\right)\right] \\ &= \left[\sum_{n=0}^{\frac{P-3}{2}} 2h(n)\cos\left(\frac{P-1}{2}-n\right)w + h\left(\frac{P-1}{2}\right)\right] \\ &+ \left[\sum_{n=0}^{\frac{P-3}{2}} 2\varepsilon_c(n)\cos\left(\frac{P-1}{2}-n\right)w + \varepsilon_c\left(\frac{P-1}{2}\right)\right] \\ &= \left|G^k_{pb}\left(e^{jw}\right)\right| + \left|E_c\left(e^{jw}\right)\right| \end{aligned} \tag{5.14}$$

Comme l'erreur de quantification est majorée par $\frac{q_c}{2}$, le module de $\mathrm{E_c}\left(\mathrm{e^{jw}}\right)$ est majoré :

$$\begin{aligned}\left|E_c\left(e^{jw}\right)\right| &\leq \sum_{n=0}^{\frac{P-3}{2}} 2\left|\varepsilon_c(n)\right|\left|\cos\left(\frac{P-1}{2}-n\right)w\right| + \left|\varepsilon_c\left(\frac{P-1}{2}\right)\right| \\ &\leq \frac{q_c}{2}\left[1+2\sum_{n=0}^{\frac{P-1}{2}}\left|\cos nw\right|\right] \leq P\frac{q_c}{2}\end{aligned} \tag{5.15}$$

Cette borne supérieure est indépendante de la fréquence w. En pratique nous pouvons obtenir la performance en terme de puissance de bruit avec un nombre de bits de quantification des coefficients $\mathrm{Nb_{cf}}$ inférieur à celui imposé par la relation 5.15.
Afin d'alléger cette condition pour la détermination du nombre de bits des coefficients $\mathrm{Nb_{cf}}$, il a été démontré dans [60] que $\left|\mathrm{E_c}\left(\mathrm{e^{jw}}\right)\right|$, pour w fixé, est une variable aléatoire gaussienne (théorème de la limite centrale) de moyenne et d'écart type donnés par :

$$\begin{aligned}&E\left[\left|E_c\left(e^{jw}\right)\right|\right] = 0 \\ &E\left[\left|E_c\left(e^{jw}\right)\right|^2\right]^{\frac{1}{2}} = \frac{q_c}{2}\sqrt{\frac{2P-1}{3}}\sqrt{\frac{1}{2}+\frac{1}{2P-1}\left(-\frac{1}{2}+\frac{\sin(Pw)}{\sin(w)}\right)} \leq \frac{q_c}{2}\sqrt{\frac{2P-1}{3}}\end{aligned} \tag{5.16}$$

Cette nouvelle borne est moins exigeante que la première. Elle permet de déterminer le nombre de bits nécessaire des coefficients pour un écart type donné sur $\left|\mathrm{E_c}\left(\mathrm{e^{jw}}\right)\right|$.
Dans le cadre de notre travail, nous avons effectué des simulations paramétriques pour déterminer le nombre de bits $\mathrm{Nb_{cf}}$ minimal pour représenter les coefficients du filtre. En effet, le critère le plus important à prendre en compte en fonction du nombre de bits $\mathrm{Nb_{cf}}$ des coefficients est la valeur de l'ENOB en sortie. La figure 5.14 montre l'évolution de l'ENOB en fonction de $\mathrm{Nb_{cf}}$. Nous pouvons observer qu'à partir de $\mathrm{Nb_{cf}} = 11$, l'ENOB est proche de la valeur théorique soit 13.3 bits. Cette valeur sera adoptée dans le reste de notre travail.
La figure 5.20 montre la réponse fréquentielle du filtre $\mathrm{G_{pb}^{k}(z)}$ avant et après quantification des coefficients. On note que les deux réponses fréquentielles sont presque identiques pour les trois premiers lobes secondaires, ce qui assure avec $\mathrm{Nb_{cf}} = 11$ une performance proche de celle prévue.

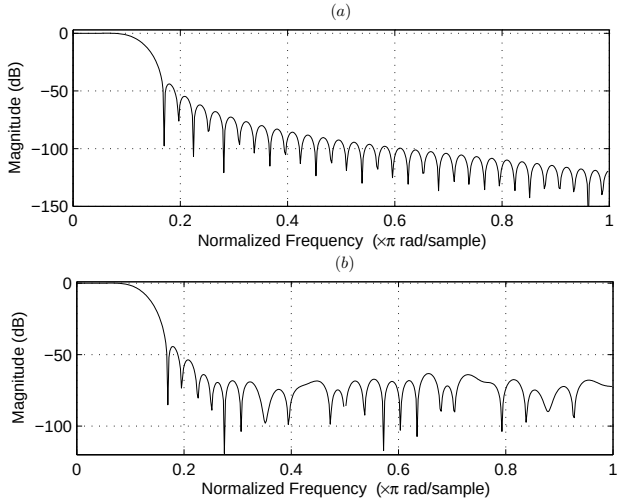

FIG. 5.20 – (a) Réponse fréquentielle idéale, (b) réponse fréquentielle avec $\mathrm{Nb_{cf}} = 11$.

– **L'erreur d'arrondi des résultats intermédiaires.** L'arrondi consiste à garder, à la fin d'une opération arithmétique (addition, multiplication), un certain nombre de bits de poids fort afin d'éviter la saturation des registres qui suivent dans le traitement. Le fait de garder un certain nombre de bits les plus significatifs revient à quantifier le résultat obtenu sur un nombre de bits plus faible. Cette opération d'arrondi introduit une erreur de quantification $\varepsilon_{\mathrm{a}}[\mathrm{n}]$ sur le résultat obtenu. Cette source d'erreur est considérée blanche, uniformément distribuée dans l'intervalle $\left[-\frac{\mathrm{q_a}}{2}, \frac{\mathrm{q_a}}{2}\right]$, de moyenne nulle et de variance $\frac{\mathrm{q_a^2}}{12}$, avec $\mathrm{q_a}$ le nouveau pas de quantification.
L'impact de cette source d'erreur sur le signal en sortie dépend de l'architecture de réalisation du filtre et de l'endroit où est effectué l'arrondi. Nous allons exprimer, à titre d'exemple, l'impact de l'arrondi du résultat de multiplication de l'architecture présentée sur la figure 5.16 sur le signal en sortie. L'arrondi ajoute une erreur $\varepsilon_{\mathrm{a_i}}[\mathrm{n}]$ sur le résultat de chaque multiplication. En supposant que ces erreurs sont décorrélées entre elles et avec le signal d'entrée, la contribution de ces erreurs sur le signal en sortie est exprimée par $\mathrm{y_a}[\mathrm{n}] = \sum_{\mathrm{i}=0}^{\frac{\mathrm{P}-1}{2}} \varepsilon_{\mathrm{a_i}}[\mathrm{n}]$. Pour un ordre P assez grand, $\mathrm{y_a}[\mathrm{n}]$ est approximativement une variable aléatoire gaussienne (théorème de la limite centrale). Elle est définie par sa moyenne et sa variance.

Moyenne :

$$E\left[y_a[n]\right] = \sum_{i=0}^{\frac{P-1}{2}} E\left[\varepsilon_{a_i}[n]\right] = 0 \tag{5.17}$$

Variance

$$E\left[y_a^2[n]\right] = \sum_{i=0}^{\frac{P-1}{2}} E\left[\varepsilon_{a_i}^2[n]\right] = \left(\frac{P+1}{2}\right)\frac{q_a^2}{12} \tag{5.18}$$

Nous constatons d'après ces trois sources d'erreur que la contribution de l'erreur de quantification du signal d'entrée est la plus faible. L'erreur de quantification des coefficients du filtre a été fixée en choisissant $\mathrm{Nb_{cf}} = 11$ afin d'avoir un ENOB proche de la valeur théorique 13.3 bits. En ce qui concerne l'erreur d'arrondi, elle dépend de l'architecture et de l'endroit où elle est effectuée dans cette architecture. Dans l'architecture du filtre présentée sur la figure 5.17, l'arrondi est indispensable à la fin de chaque opération arithmétique afin d'optimiser le traitement. Nous avons effectué des simulations paramétriques où nous avons fait varier la longueur des mots à différents endroits entre une valeur minimale et la valeur maximale théorique comme indiqué dans le tableau 5.4.

Tab. 5.4 – Intervalle de variation des longueurs des mots binaires.

	bits
$\mathrm{Nb_{ef}}$	[11, 18]
$\mathrm{Nb_{ad}}$	$\mathrm{Nb_{ef}} + 1$
$\mathrm{Nb_{cf}}$	11
$\mathrm{Nb_m}$	[18, 22]
$\mathrm{Nb_{acc}}$	[16, 22]

Nous avons calculé pour chaque combinaison $\{\mathrm{Nb_{ef}}, \mathrm{Nb_m}, \mathrm{Nb_{acc}}\}$ la puissance de bruit en sortie en injectant en entrée un signal sinusoïdal d'amplitude 1 dans la bande passante du filtre. La figure 5.21 présente la puissance de bruit en sortie en fonction de la somme des différentes combinaisons.

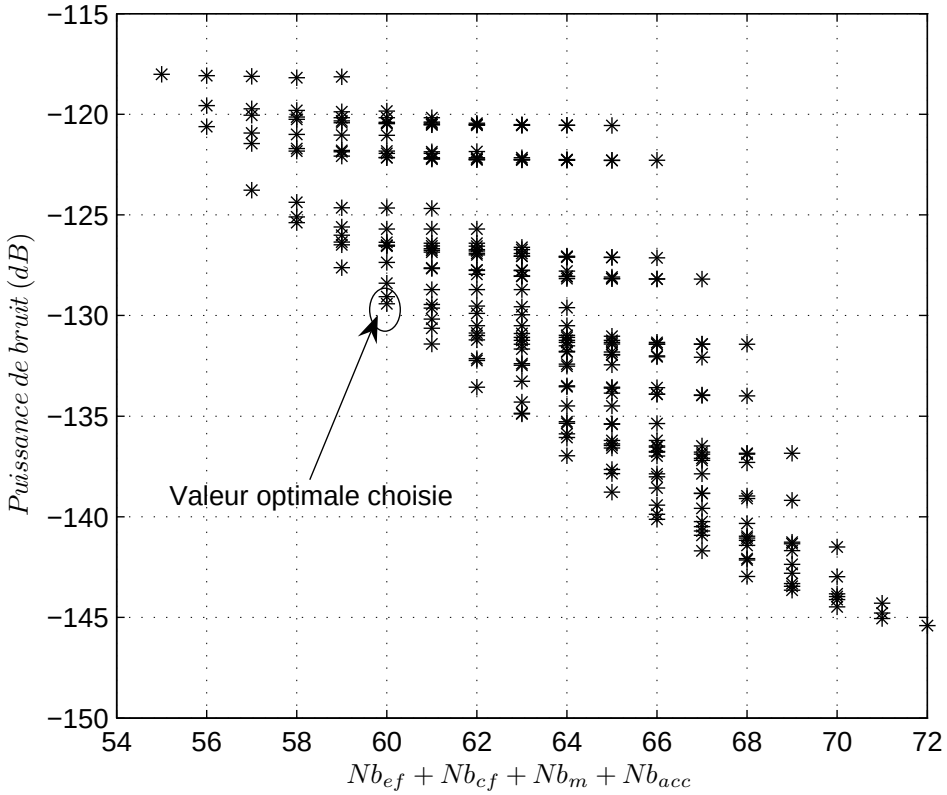

Fig. 5.21 – Puissance de bruit en fonction de la somme des largeurs des registres du traitement.

Comme le niveau souhaitable de la densité spectrale de bruit de quantification à la sortie de chaque étage de traitement est de -120 dB (voir chapitre 3, figure 3.28) afin d'atteindre un ENOB de 13.3 bits en sortie, nous avons choisi la combinaison
$\{\mathrm{Nb_{ef}} = 13, \mathrm{Nb_{cf}} = 11, \mathrm{Nb_m} = 18, \mathrm{Nb_{acc}} = 18\}$ permettant d'avoir un niveau de PSD autour de -129 dB. D'autres choix sont envisageables avec une puissance de bruit plus grande et une combinaison de somme plus faible. Ces choix doivent être également validés dans l'architecture globale du traitement.

La figure 5.22 présente la PSD du signal idéal (a) et celle du signal quantifié sur $\mathrm{Nb_{ef}} = 13$ (b). Le niveau de bruit de quantification a pour valeur approximative -120 dB. La figure 5.23 présente la PSD du signal en sortie du filtre idéal (a) avec un signal quantifié en entrée, et celle obtenue avec l'architecture parallèle avec le choix des longueurs de mots intermédiaires $\{\mathrm{Nb_{ef}} = 13, \mathrm{Nb_{cf}} = 11, \mathrm{Nb_m} = 18, \mathrm{Nb_{acc}} = 18\}$. Le niveau de la densité spectrale de bruit en sortie est aux alentours de -130 dB, ce qui est conforme au niveau souhaité de -120 dB afin d'avoir un ENOB de 13.3 bits.

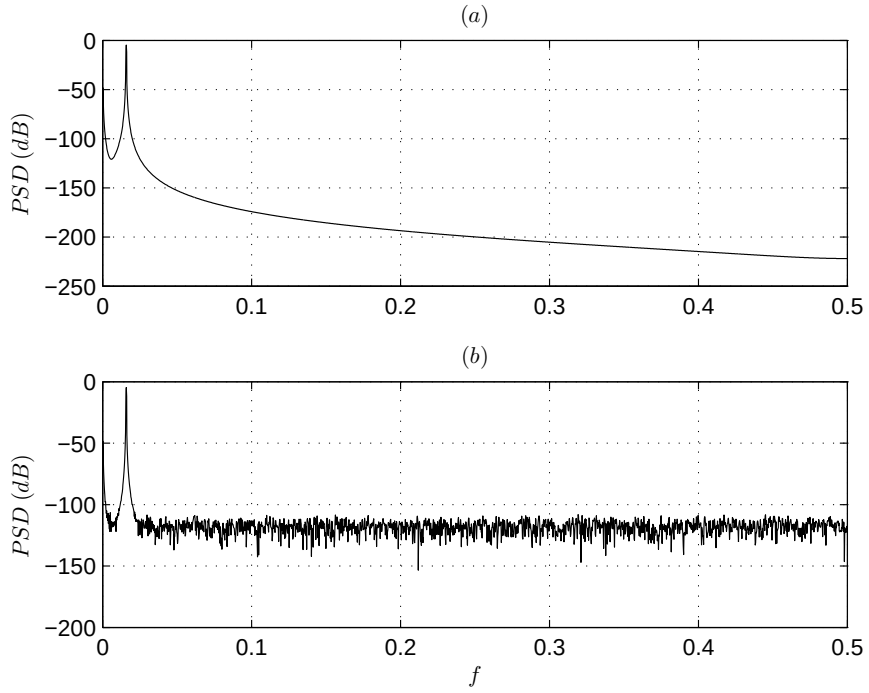

Fig. 5.22 – (a) PSD du signal d'entrée : sans quantification, (b) avec quantification sur $\mathrm{Nb}_{\mathrm{ef}} = 13$.

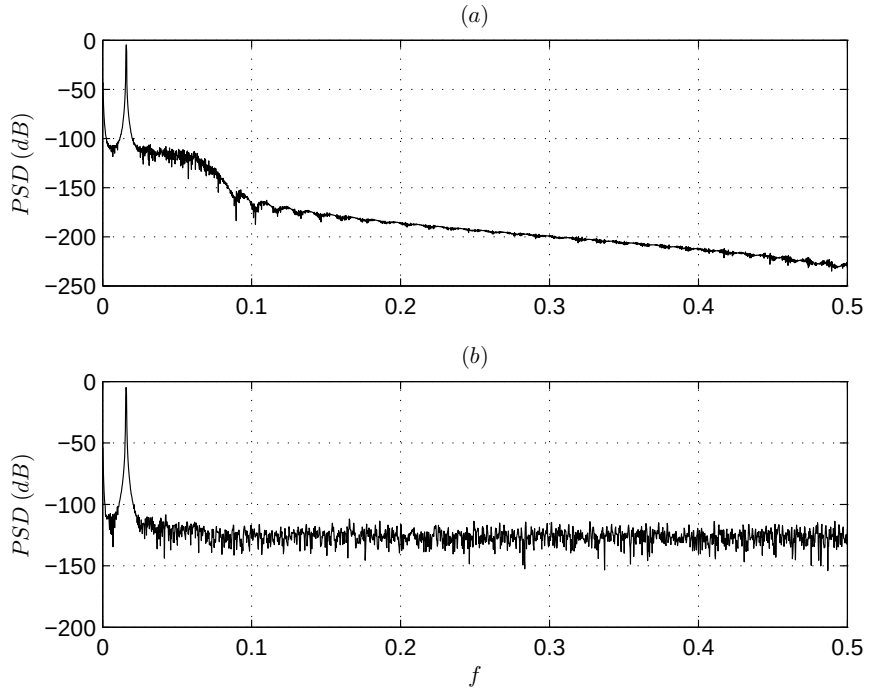

Fig. 5.23 – (a) Sortie du filtre idéal avec une entrée quantifiée, (b) sortie du filtre quantifié avec une entrée quantifiée.

5.5 Soustracteur

Le passage de la représentation signée à la représentation non signée de la sortie du démodulateur a permis de simplifier la réalisation du filtre décimateur et du filtre $G^k_{pb}(z)$ fonctionnant à la vitesse de 800 MHz (§ 5.2). Cependant, ce passage a ajouté aux signaux de sortie du démodulateur une composante continue de valeur $2^{Nb_{sd}-1}$. Cette composante continue doit être retranchée avant la modulation du signal afin de ne pas récupérer des raies sur le spectre du signal en sortie. Nous avons choisi d'effectuer cette soustraction directement après le filtre $G^k_{pb}(z)$, puisque c'est le premier élément fonctionnant à une vitesse R_d fois plus faible. La valeur de la composante continue a été modifiée lors de son passage par le filtre décimateur et le filtre $G^k_{pb}(z)$ qui n'ont pas un gain unité. Ce gain dépend :

- des différents arrondis effectués à l'intérieur de chaque entité.
- de la division des coefficients du filtre $G^k_{pb}(z)$ par $2 \times g_{f_c}$ lors de la conception (voir § 5.4). Le gain g_{f_c} dépend de la largeur de bande du filtre $G^k_{pb}(z)$. Comme 25 largeurs de bande sont possibles pour le filtre $G^k_{pb}(z)$, les valeurs à retrancher peuvent être calculées auparavant et stockées dans une mémoire ROM. Le calcul de ces valeurs impose leur représentation dans la ROM sur 16 bits dans notre exemple.

La figure 5.24 et le tableau 5.5 présentent les paramètres et les ports de l'entité soustracteur.

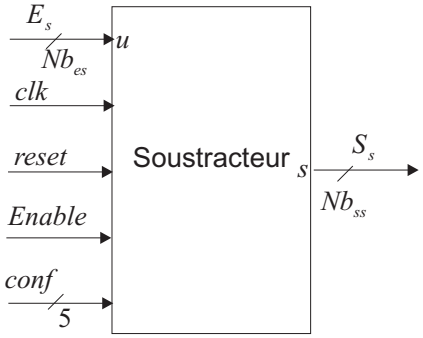

Fig. 5.24 – Schéma bloc de l'entité soustracteur.

Tab. 5.5 – Paramètres et ports d'entrées/sorties du soustracteur

Soustracteur	
Paramètres de configuration	
Nb_{es}	Nombre de bits du signal non signé en entrée.
Nb_{ss}	Nombre de bits du signal signé de sortie.
Ports d'entrées/sorties	
clk	Horloge de commande de tous les registres du filtre.
reset	Entrée de remise à zéro des différents registres.
Enable	Entrée de verrouillage des registres à décalage.
E_s	Entrée non signée du soustracteur.
S_f	Sortie signée du soustracteur.

5.6 Filtre de correction $C_1^k(z)$

La fonction de transfert du filtre de correction du module de la STF du modulateur ΣΔ $C_1^k(z)$ a été exprimée d'une façon approchée au chapitre 4 par l'équation (4.17). Cette fonction $C_1^k(z)$ qui a pour forme :

$$C_1^k(z) \approx g\left(-\varepsilon\left(1-j2\pi\nu\right)+\left(1+2\varepsilon\right)z^{-1}-\varepsilon\left(1+j2\pi\nu\right)z^{-2}\right)$$

ne peut pas être intégrée dans le filtre passe-bas $G_{pb}^k(z)$. En effet, l'ensemble des coefficients $\{\varepsilon, \nu, g\}$ varie en fonction des imperfections des composants analogiques du modulateur. Elles seront mises à jour par l'algorithme de calibration développé au § 4.6.3.

Architecture de réalisation

Le filtre $C_1^k(z)$ est un filtre à coefficients complexes. Son entrée E_c est un signal complexe de type $E_c = E_{cr} + jE_{ci}$. Le signal en sortie de ce filtre s'exprime dans le domaine z par :

$$\begin{aligned} S_c(z) &= \left(E_{cr}(z)+jE_{ci}(z)\right)C_1^k(z) \\ &= \left(E_{cr}(z)+jE_{ci}(z)\right)\left(g\left(-\varepsilon\left(1-j2\pi\nu\right)+\left(1+2\varepsilon\right)z^{-1}-\varepsilon\left(1+j2\pi\nu\right)z^{-2}\right)\right) \end{aligned} \tag{5.19}$$

Le développement de cette expression permet d'exprimer la partie réelle et la partie imaginaire du signal en sortie :

$$\begin{aligned} &S_c(z) = S_{cr}(z) + jS_{ci}(z) \\ &\text{avec} \\ &S_{cr}(z) = \left[E_{cr}(z)z^{-1} + \varepsilon\left[E_{cr}(z)\left(-z^{-2}+2z^{-1}-1\right)+E_{ci}(z)2\pi\nu\left(z^{-2}-1\right)\right]\right]g \\ &S_{ci}(z) = \left[E_{ci}(z)z^{-1} + \varepsilon\left[E_{ci}(z)\left(-z^{-2}+2z^{-1}-1\right)+E_{cr}(z)2\pi\nu\left(-z^{-2}+1\right)\right]\right]g \end{aligned} \tag{5.20}$$

D'après ces expressions, la partie réelle $S_{cr}(z)$ et la partie imaginaire $S_{ci}(z)$ présentent pratiquement la même architecture de réalisation. La figure 5.25 présente l'architecture de réalisation de $S_{cr}(z)$. Cette architecture comprend :

- deux registres à décalages X_r et X_i permettant de sauvegarder les trois dernières valeurs des deux signaux d'entrée E_{cr} sur la voie réelle et E_{ci} sur la voie imaginaire.
- trois multiplieurs. Afin d'assurer une fréquence de fonctionnement à 160 MHz, nous avons découpé chaque multiplication en 4 multiplications partielles afin d'avoir un chemin critique inférieur à la période de fonctionnement. Les résultats des différentes multiplications sont arrondis afin d'éviter la saturation des registres qui suivent dans le traitement. Nb_{ar1}, Nb_{ar2} et Nb_{sc} sont les longueurs des mots après l'arrondi des différents résultats de multiplication indiquées sur la figure 5.25.

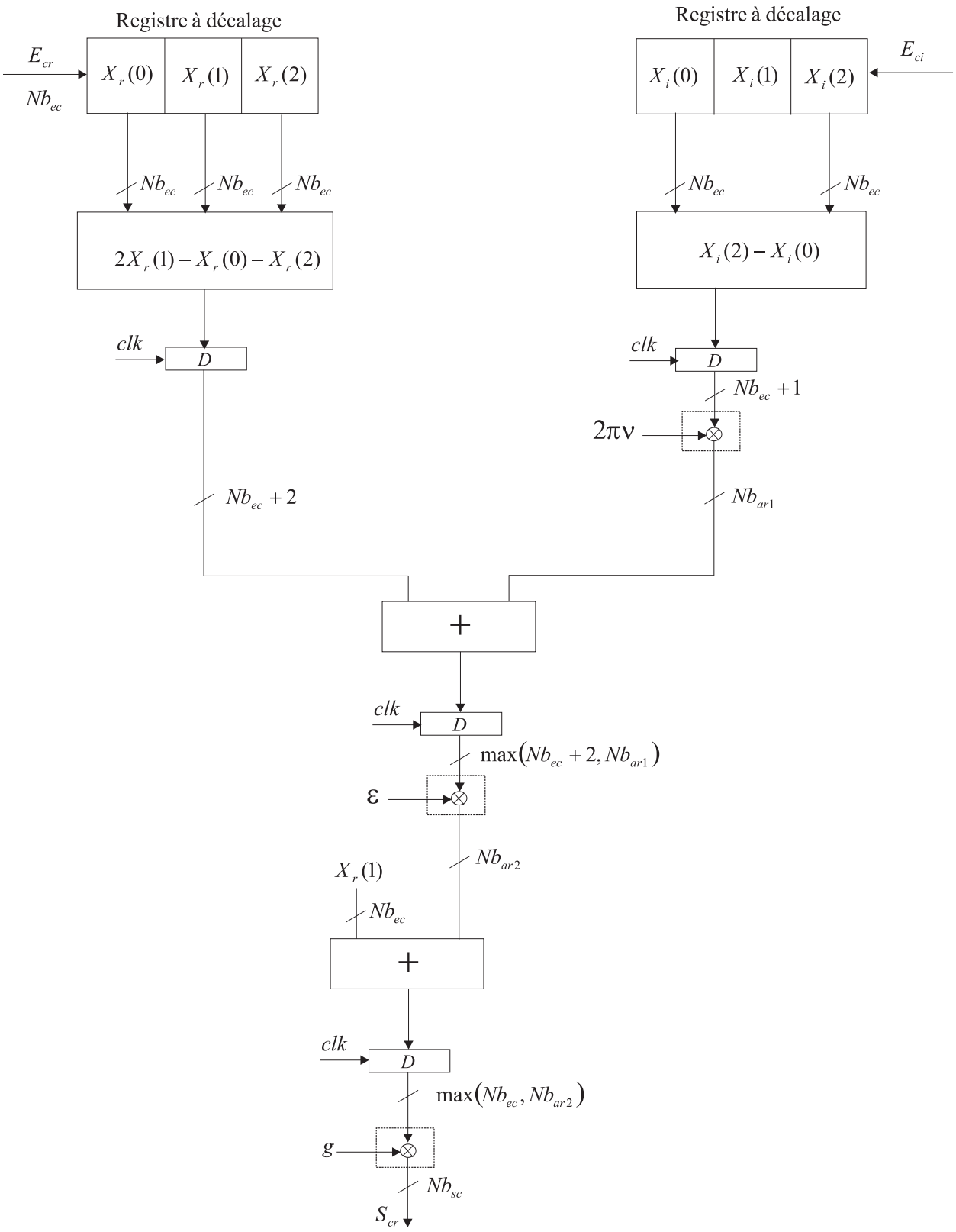

FIG. 5.25 – Architecture de réalisation de la partie réelle $S_{cr}(z)$ du filtre $C_1^k(z)$.

La figure 5.26 et le tableau 5.6 présentent les paramètres et les ports de l'entité filtre $C_1^k(z)$.

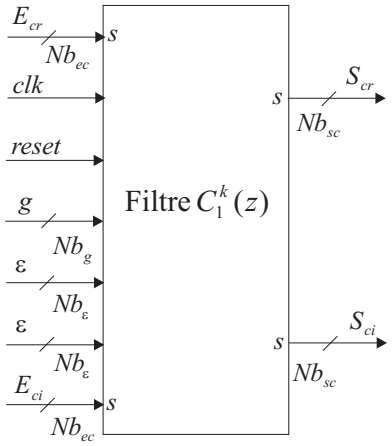

Fig. 5.26 – Schéma-bloc de l'entité filtre $C_1^k(z)$.

Tab. 5.6 – Paramètres et ports d'entrées/sorties du filtre $C_1^k(z)$.

Filtre $C_1^k(z)$	
Paramètres de configuration	
Nb_{ec}	Nombre de bits du signal signé en entrée.
Nb_g	Nombre de bits du gain du filtre g.
Nb_ε	Nombre de bits de la concavité du filtre ε.
Nb_ν	Nombre de bits du déphasage du filtre ν.
Nb_{ar1}	Nombre de bits de l'arrondi après multiplication par $2\pi\nu$.
Nb_{ar2}	Nombre de bits de l'arrondi après multiplication par ε.
Nb_{sc}	Nombre de bits du signal de sortie.
Ports d'entrées/sorties	
clk	Horloge de commande de tous les registres du filtre.
reset	Entrée de remise à zéro des différents registres.
E_{cr}, E_{ci}	Entrées réelle et imaginaire signées du filtre.
g	Gain du filtre.
ε	Concavité du filtre.
ν	Déphasage du filtre.
S_{cr}, S_{ci}	Sorties réelle et imaginaire signées du filtre.

Afin de déterminer la longueur optimale des mots dans l'architecture présentée sur la figure 5.25, des simulations paramétriques où nous avons fait varier indépendamment la longueur de chaque mot entre une valeur minimale et la valeur maximale théorique comme indiqué dans le tableau 5.7 ont été effectuées. Pour chaque combinaison $\{Nb_{ec}, Nb_\varepsilon, Nb_g, Nb_\nu, Nb_{ar1}, Nb_{ar2}, Nb_{sc}\}$ la puissance de bruit en sortie est calculée en injectant en entrée un signal sinusoïdal d'amplitude 1 dans la bande passante du filtre. La combinaison choisie est $\{Nb_{ec} = 15, Nb_\varepsilon = 5, Nb_g = 8, Nb_\nu = 8, Nb_{ar1} = 17, Nb_{ar2} = 17, Nb_{sc} = 15\}$ permettant d'avoir un niveau de la densité spectrale de bruit de -120 dB.

TAB. 5.7 – Intervalles de variation des longueurs des mots binaires du filtre $C_1^k(z)$.

	bits
Nb_{ec}	[10, 18]
Nb_{ε}	[5, 8]
Nb_g	[8, 11]
Nb_{ν}	[7, 10]
Nb_{ar1}	[13, 20]
Nb_{ar2}	[13, 20]
Nb_{sc}	[13, 20]

5.7 Modulateur

Le modulateur permet de ramener le signal de la bande de base à sa bande initiale en tenant compte du facteur de décimation R_d. Cette modulation est réalisée en multipliant le signal en sortie du filtre $C_1^k(z)$ par la séquence complexe $m'_k[n] = e^{j2\pi f_c^k R_d n}$. En suivant le même raisonnement que celui du § 5.2, les valeurs de la séquence m'_k sont calculées à partir de l'erreur d'approximation sur la fonction sinus dans le premier quadrant. Cette erreur est stockée en mémoire ROM. L'architecture du modulateur diffère de celle du démodulateur par les deux facteurs suivants :

1. La différence entre les deux séquences m_k et m'_k est le facteur R_d. Ce facteur est lié directement au pas de lecture de la mémoire ROM. Dans le cas de la séquence m_k, la lecture est réalisée d'une façon cyclique avec un pas de f_c^k entre deux valeurs successives. Par contre, avec la séquence m'_k la lecture est cyclique également mais avec un pas égal au facteur $R_d \times f_c^k$.
2. L'entrée du modulateur est un signal complexe de type $E_m = E_{mr} + jE_{mi}$. Donc, à la différence du démodulateur, la multiplication à réaliser est une multiplication entre deux nombres complexes. Le signal en sortie du modulateur s'exprime de la façon suivante :

$$\begin{aligned} S_m[n] &= E_m[n]\, e^{j2\pi f_c^k R_d n} \\ &= (E_{mr} + jE_{mi})\left(\cos(2\pi f_c^k R_d n) + j\sin(2\pi f_c^k R_d n)\right) \\ &= \left(E_{mr}\cos(2\pi f_c^k R_d n) - E_{mi}\sin(2\pi f_c^k R_d n)\right) \\ &+ j\left(E_{mi}\cos(2\pi f_c^k R_d n) + E_{mr}\sin(2\pi f_c^k R_d n)\right) \\ &= S_{mr}[n] + jS_{mi}[n] \end{aligned} \tag{5.21}$$

Même si le modulateur fonctionne à une cadence R_d fois plus faible que le démodulateur, le chemin critique de chacune des multiplications dans l'expression (5.21) est long car le modulateur se trouve en fin de chaîne de traitement. Nous sommes amenés au cours de la synthèse à diviser chacune des multiplications du signal en sortie S_m en 4 multiplications partielles afin de diminuer la longueur du chemin critique pour assurer le fonctionnement à la vitesse souhaitée. La figure 5.27 et le tableau 5.8 présentent les paramètres et les ports de l'entité modulateur.

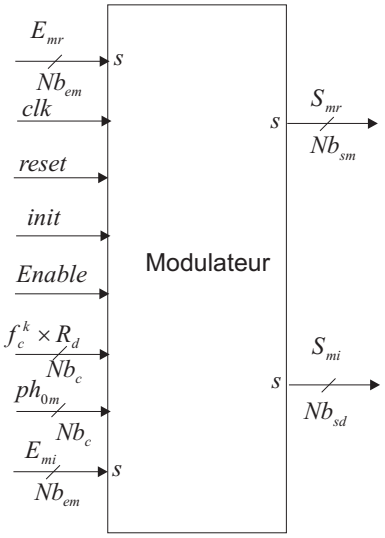

Fig. 5.27 – Schéma-bloc de l'entité du modulateur.

Tab. 5.8 – Paramètres et ports d'entrées/sorties du modulateur.

Bloc modulateur	
Paramètres de configuration	
Nb_{em}	Nombre de bits du signal d'entrée du modulateur.
Nb_{ss}	Nombre de bits du signal de sortie du modulateur.
Nb_c	Nombre de bits de la fréquence centrale et de la phase initiale ph_{0m}.
Ports d'entrées/sorties	
clk	Horloge de commande de tous les registres.
reset	Entrée de remise à zéro de tous les registres.
init	Entrée de validation de lecture de la phase ph_{0m}.
E_{mr}, E_{mi}	Entrées réelle et imaginaire signées du modulateur.
$f_c^k \times R_d$	Fréquence de modulation.
ph_{0m}	Phase initiale de modulation.
S_{mr}, S_{mi}	Sortie réelle et imaginaire signées du modulateur.

5.8 Architecture complète du traitement numérique

5.8.1 Résultats au niveau portes logiques

Le traitement numérique appliqué à la sortie de chaque modulateur ΣΔ est composé des différentes entités décrites dans les paragraphes précédents. L'architecture complète du traitement numérique est présenté sur la figure 5.28.

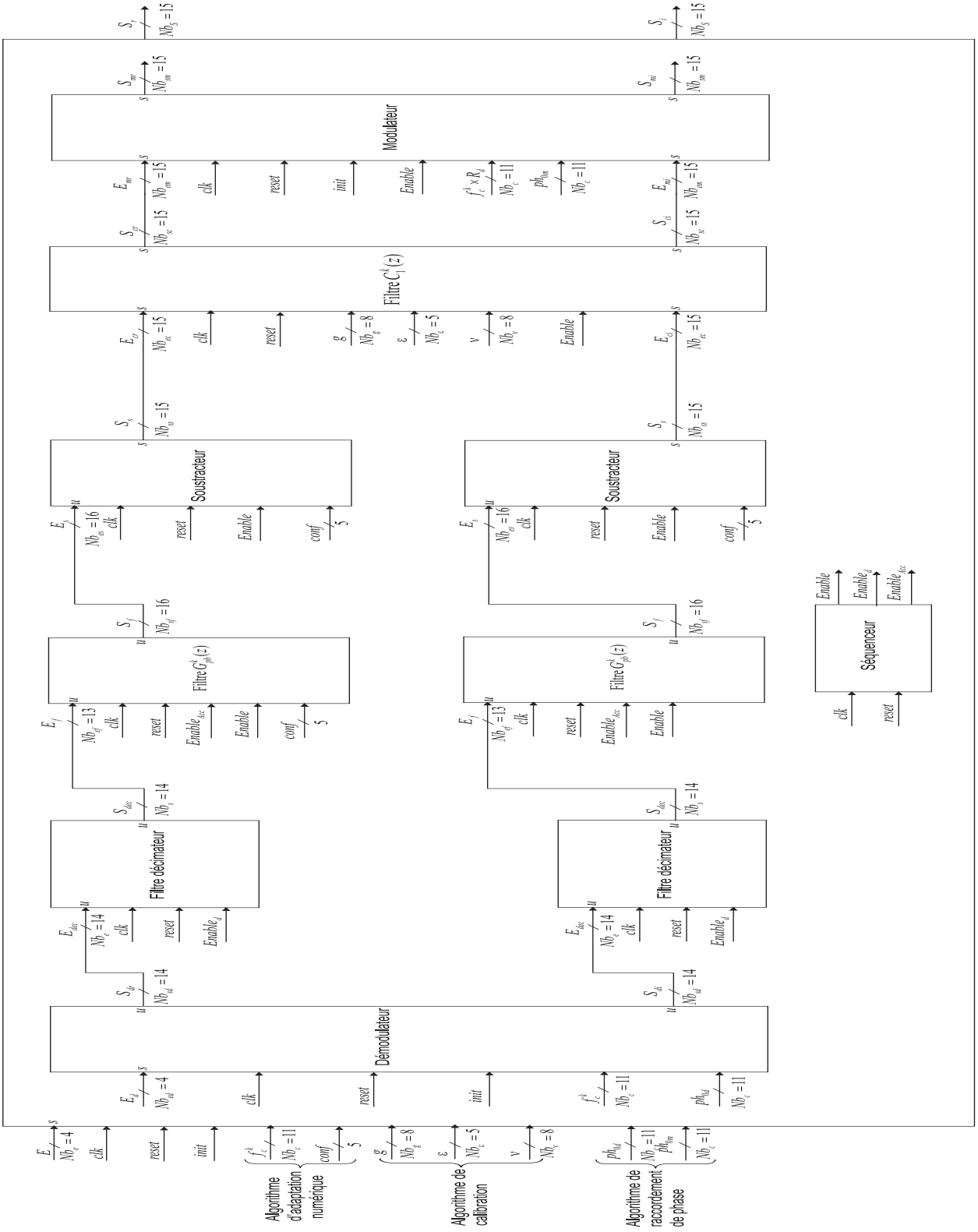

Fig. 5.28 – Architecture du traitement numérique.

L'assemblage des différents entités est effectué en gardant les longueurs des ports d'en-

trées/sorties optimales trouvées pour chaque entité. La figure 5.29 présente le signal en sortie du traitement numérique appliqué au deuxième modulateur conformément à l'exemple du tableau 3.1. Ce signal est obtenu à partir d'une simulation au niveau portes logiques. Nous avons utilisé un signal en entrée de type chirp d'amplitude normalisée 0.9 balayant toute la plage fréquentielle $[0.2F_e, 0.3F_e]$.

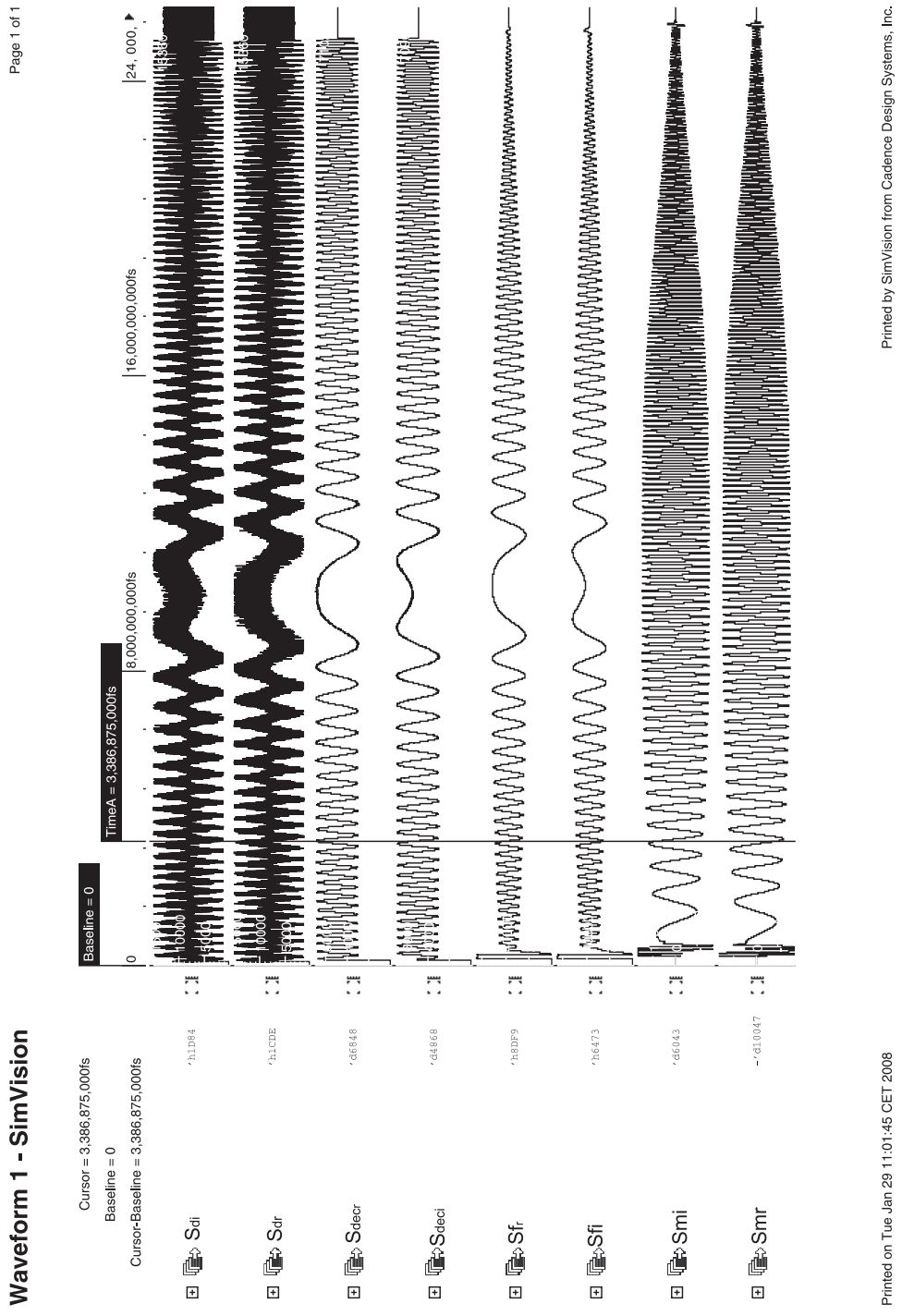

Fig. 5.29 – Résultat du traitement numérique au niveau portes logiques pour le deuxième modulateur.

La densité spectrale du signal complexe $S_r + jS_i$ à la sortie du deuxième modulateur (figure 5.30) montre que le niveau de bruit est légèrement plus élevé que celui attendu. Ceci est dû au fait qu'en plus du bruit de quantification du modulateur, un bruit est introduit par les différents arrondis appliqués lors du traitement numérique.

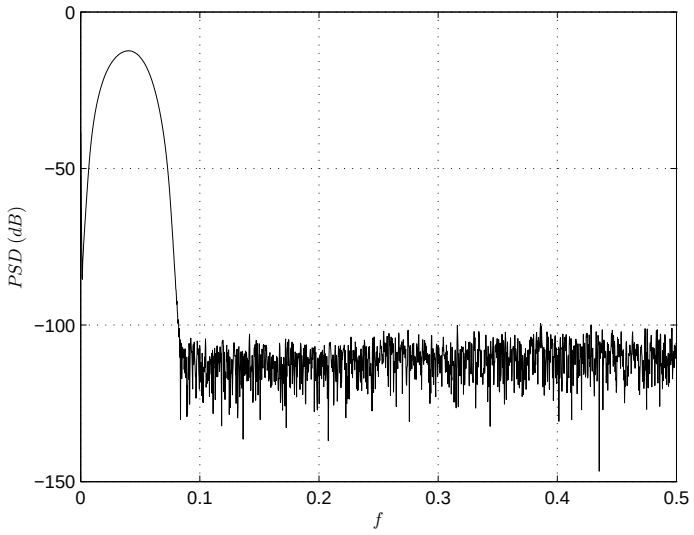

Fig. 5.30 – PSD du signal en sortie du deuxième démodulateur.

5.8.2 Synthèse de l'architecture

Nous avons procédé à la synthèse de cette architecture à une fréquence de fonctionnement de 800 MHz avec la technologie CMOS 0.12 μm en utilisant le design Kit de *STMicroelectronics*. La figure 5.31 présente le schéma des différentes entités après la synthèse. La vitesse maximale de fonctionnement du circuit dépend du chemin critique. La figure 5.32 présente le chemin critique de cette architecture. Il se trouve au niveau de l'accumulateur du filtre passe-bas $G^k_{pb}(z)$ qui somme les différents résultats de multiplication. Une éventuelle augmentation de la vitesse de fonctionnement nécessite le découpage de ce chemin suivant le principe de pipelinage.

Le tableau 5.9 présente les ressources matérielles nécessaires pour l'implantation de chaque entité et de l'entité de l'architecture finale du traitement numérique. Il présente également la vitesse de fonctionnement ($F_e = 800$ MHz ou $F_{ed} = 160$ MHz) et la surface de silicium nécessaire en mm^2. Nous pouvons noter que :

- le démodulateur, le filtre $C^k_1(z)$, le soustracteur et le modulateur représente et 28% de la surface totale,
- le filtre passe-bas $G^k_{pb}(z)$ occupe la plus grande partie du circuit : 64%,
- le filtre en peigne $C(z)$ occupe 8%.

Par conséquent, la surface totale du traitement numérique pour l'architecture EFBD avec 10 modulateurs est de 6.2 mm^2.

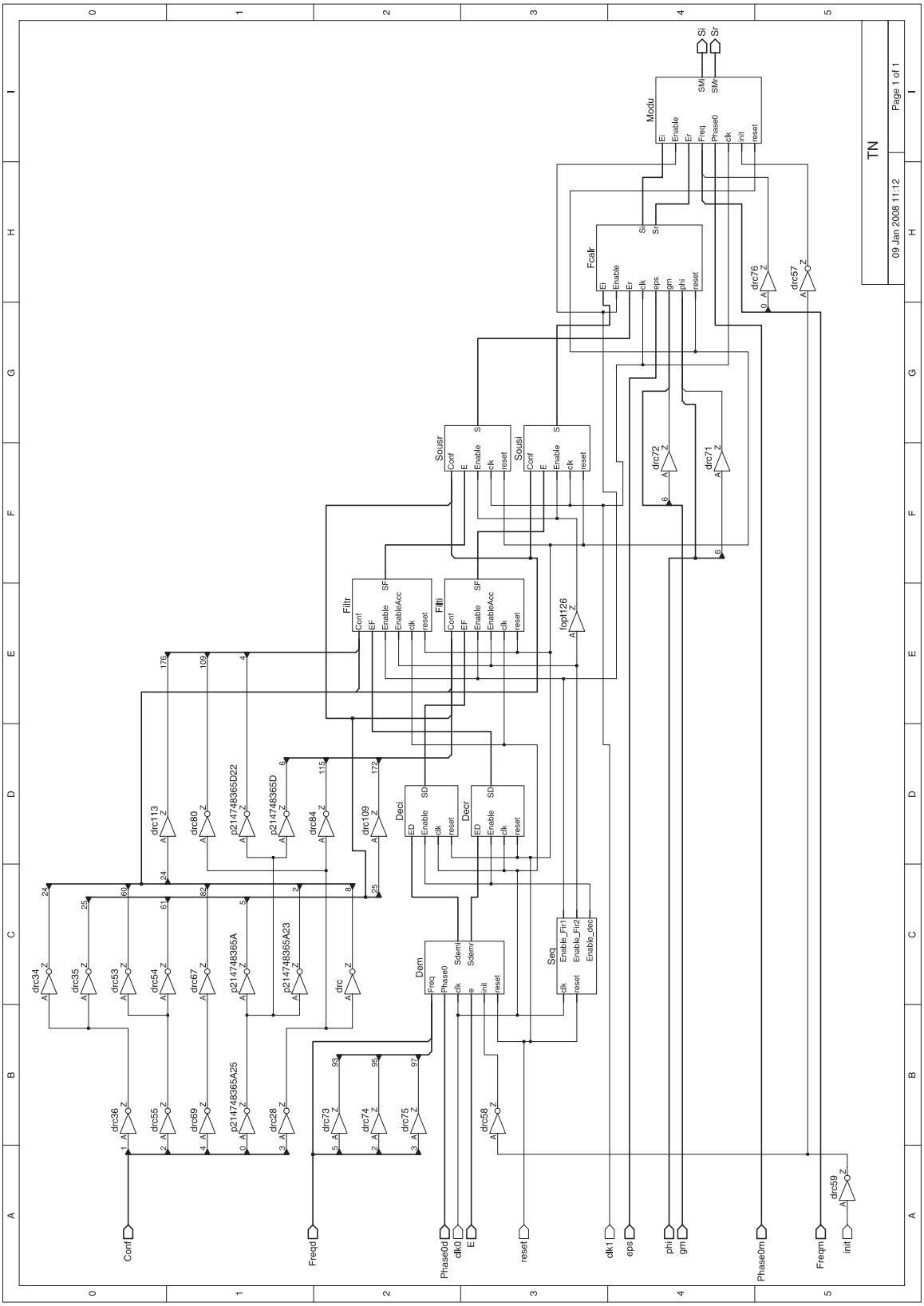

Fig. 5.31 – Architecture du traitement après synthèse.

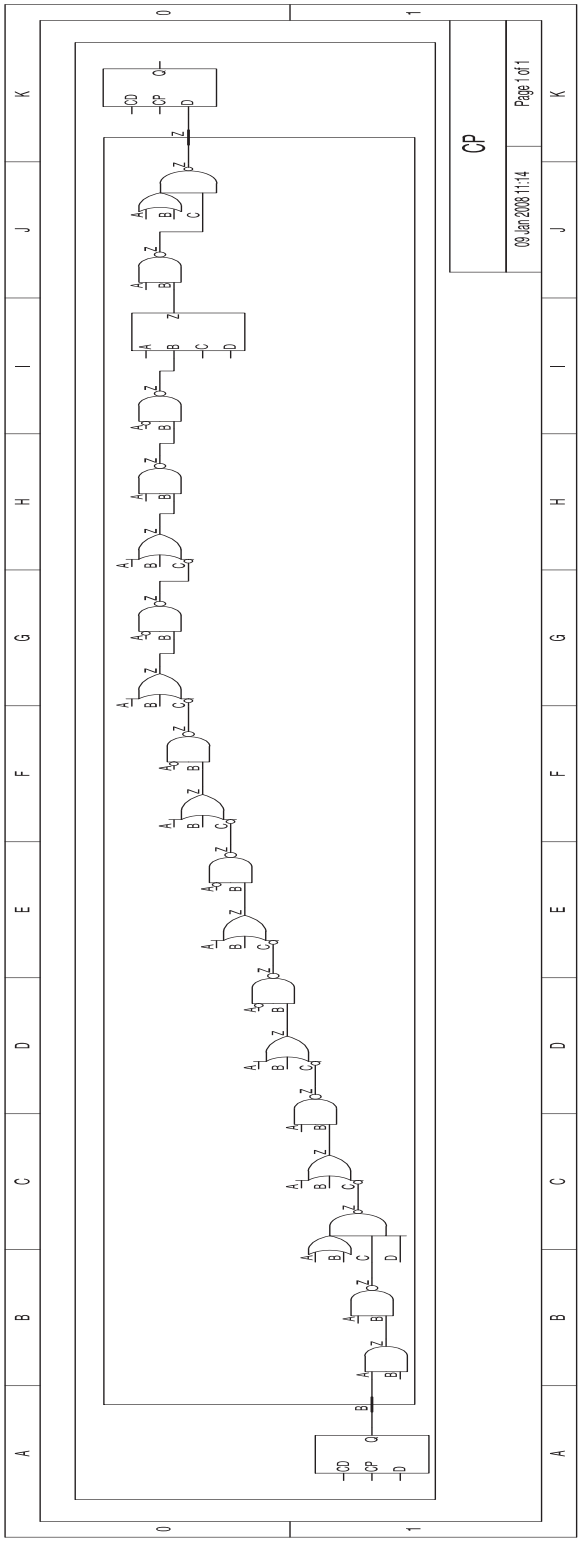

Fig. 5.32 – Chemin critique de l'architecture du traitement.

Tab. 5.9 – Estimation de la puissance de calcul nécessaire pour un seul modulateur.

Entité	Multiplieurs	Additionneurs	Fréquence	Bascules D	Nombre de portes	Surface (mm^2)
Démodulateur	0	2	F_e	167	2509	0.027
Filtre-décimateur	0 0	4×2 4×2	F_e F_{ed}	293×2	2193×2	0.026×2
Filtre $G^k_{pb}(z)$	7×2	7×2	F_e	1797×2	15024×2	0.2×2
Soustracteur	0	1×2	F_{ed}	47×2	195×2	0.003×2
$C^k_1(z)$	6	10	F_{ed}	914	3401	0.07
Modulateur	4	2	F_{ed}	555	4894	0.07
TN	-	-	F_e, F_{ed}	5887	43853	0.62

5.9 Conclusion

Ce chapitre a présenté le développement des architectures numériques retenues pour l'architecture EFBD passe-bande jusqu'au niveau portes logiques. Chaque entité du traitement a été optimisée en termes de vitesse et de surface d'implantation. Pour cela, une méthode de calcul des coefficients de démodulation et de modulation basée sur le stockage de l'erreur d'approximation au lieu des vraies valeurs permettant ainsi de diminuer la surface de l'architecture globale. Une nouvelle architecture pour le filtre FIR permettant de diminuer la surface en divisant le nombre de multiplieurs pratiquement par 5 a également été proposée. Nous avons réalisé la synthèse de l'architecture globale, au niveau portes logiques, à la fréquence de fonctionnement de 800 MHz avec la technologie CMOS 0.12 μm. Le nombre de portes nécessaires pour ce traitement est 43853, ce qui occuperait une surface de silicium de 6.2 mm^2 avec la technologie considérée.

Conclusion générale et perspectives

Apport de la thèse

Le thème de cette thèse concerne le développement d'architectures innovantes pour la réalisation de récepteurs multistandards, c'est-à-dire pouvant s'adapter à plusieurs normes de communication. De tels récepteurs se basent sur le concept de la radio-logicielle, c'est-à-dire la conception d'entités numériques reprogrammables par logiciels, plutôt que l'implantation de circuits complexes adaptés à chacune des normes. Pour permettre d'appliquer un tel concept, il est nécessaire de pouvoir numériser des signaux large bande le plus près possible de l'antenne. Toutes les fonctions exigées dans un récepteur classique peuvent ensuite être réalisées dans le domaine numérique.

Afin de répondre à ces exigences nous avons choisi la voie du parallélisme pour l'élargissement de la bande de fonctionnement des convertisseurs. Nous avons choisi de travailler avec des architectures à décomposition fréquentielle en raison de leur robustesse aux erreurs de gain et de décalage en tension par rapport aux architectures à entrelacement temporel et à base de modulation de *Hadamard*.

Les architectures proposées et le post-traitement numérique

Dans le cadre du travail de thèse, nous avons proposé une nouvelle architecture à base de décomposition fréquentielle pour la conversion de signaux de type passe-bande. Cette architecture est appelée architecture FBD passe-bande. Elle présente comme avantages par rapport aux autres architectures de l'état de l'art :

- de convertir une bande adaptée au signal utile plutôt que toute la bande,
- une plus grande rapidité grâce à des modulateurs à temps continu.

Avec l'architecture FBD passe-bande, les erreurs analogiques ne présentent pas d'effets non-linéaires (raies spectrales) sur le spectre de bruit en sortie. Cependant, elles peuvent déplacer les fréquences centrales des résonateurs, ce qui conduit à désadapter le traitement numérique qui était calibré en fonction de la position idéale de ces fréquences centrales, et par conséquent à dégrader la résolution attendue par cette architecture.

Architecture $EFBD$ passe-bande et algorithmes de calibration

Afin de résoudre le problème du déplacement des fréquences centrales dû aux dispersions technologiques, nous avons proposé :

- d'étendre l'architecture FBD en ajoutant deux modulateurs de chaque côté de la bande du signal utile. Ceci nous a permis de garantir que le signal utile se trouve toujours dans la bande de fonctionnement des modulateurs. Nous avons nommé cette architecture EFBD passe-bande (*Extended Frequency Band Decomposition*).

– d'utiliser des algorithmes de calibration pour adapter à nouveau le traitement et assurer la performance de l'architecture EFBD. Nous avons proposé trois algorithmes de calibration permettant d'adapter le traitement numérique, de corriger les problèmes liés aux module et phase du signal en sortie. Ces algorithmes se caractérisent par leur facilité de réalisation (avec peu de composants) et leur fonctionnement en temps réel. Ils ont été testées et validées par simulation sur un exemple réaliste (architecture EFBD à 10 modulateurs).

Optimisation et implantation du traitement numérique

Nous avons proposé, dans le cadre de l'architecture EFBD passe-bande, un traitement numérique simple et performant pour reconstruire le signal numérisé. Ce dernier se base sur la démodulation et la modulation, et permet d'atteindre la performance souhaitée en termes de résolution avec des ressources matérielles raisonnables. En comparaison, l'architecture FBD présentée dans [2, 37] propose l'utilisation des filtres passe-bande après les modulateurs pour reconstruire le signal et éliminer le bruit de quantification hors bande. Cette solution s'avère très compliquée en raison de l'ordre élevé de ces filtres pour atteindre la performance théorique de cette architecture.

Nous avons optimisé le traitement numérique en termes de surface et de vitesse de fonctionnement. Nous avons mené également des simulations pour obtenir la longueur optimale des flux de données aux différents endroits du traitement. L'architecture optimisée du traitement a fait l'objet d'une synthèse au niveau portes logiques avec la technologie CMOS 0.12 μm. La surface totale du traitement numérique est de 6.2 mm^2.

Perspectives

L'architecture EFBD passe-bande est plus adéquate que les architectures proposées dans l'état de l'art au concept de la radio-logicielle pour sa bande de fonctionnement réduite à la bande utile du signal et sa fréquence de fonctionnement plus élevée.

Nous avons pu montrer au cours de cette thèse la faisabilité du traitement numérique d'une architecture EFBD passe-bande en développant une méthode de reconstruction simplifiée et en optimisant l'architecture de chaque bloc du traitement numérique, au niveau surface et au niveau vitesse de fonctionnement. Dans les perspectives à court terme de ce travail de thèse, il reste à évaluer au niveau portes logiques les ressources matérielles pour l'implémentation des algorithmes de calibration, afin d'estimer la surface totale du traitement numérique.

La méthode de reconstruction et les algorithmes de calibration sont génériques. Ces algorithmes peuvent, par une transposition aux technologies les plus avancées, ouvrir la voie à des applications très hautes fréquences.

Les normes de radiocommunication présentent des bandes de fonctionnement plus ou moins larges et la précision de conversion exigée dépend de ces largeurs. Une bande large nécessite une précision plus faible qu'une bande étroite. Dans ce contexte, il est indispensable d'agir sur la largeur de bande du convertisseur analogique/numérique afin d'optimiser son fonctionnement.

Dans les perspectives à long terme de ce travail, nous proposons d'étudier des modulateurs à fréquences centrales reconfigurables. De nouvelles méthodes de calibration nous permettraient d'interagir sur les fréquences centrales pour compenser les erreurs analogiques et améliorer par conséquent la résolution.

Annexe A

Modulateur sigma-delta

A.1 Le convertisseur analogique numérique Sigma-Delta (ΣΔ)

Le concept du modulateur ΣΔ est d'employer une boucle de rétroaction pour améliorer la résolution d'un quantificateur grossier. Il fut breveté par Cutler en 1954. Nombreux furent ceux qui apportèrent des modifications et des améliorations dans les années qui suivirent, mais le changement le plus substantiel fut proposé par Ritchie en 1977. Il proposa d'inclure plusieurs intégrateurs en cascade pour augmenter l'ordre du modulateur et par conséquent améliorer la résolution. Suivirent plusieurs travaux sur la stabilité des modulateurs d'ordre élevé qui favorisèrent le développement de cette technique auprès des fabricants de circuits intégrés.

Hayashi proposa en 1986 la topologie MASH (multi-stages noise shaping) pour s'affranchir des problèmes de stabilité pour les modulateurs d'ordre élevé [63]. Bénabès a proposé en 1996 la structure MSCL [42]. Elle consiste à mettre plusieurs modulateurs simples en cascade avec une rétroaction supplémentaire de sorte que le signal de sortie global soit directement la somme des signaux de sorties des divers étages. Ces structures n'ont pas besoin de prétraitement numérique, cependant une calibration est souvent nécessaire pour compenser certains défauts du bouclage dans le modulateur.

L'architecture générale d'un CAN de type sigma-delta est présentée sur la figure A.1. Elle est constituée d'une boucle appelée modulateur, suivi d'un filtre numérique. Le modulateur contient un filtre analogique ou numérique (dit "filtre de boucle") suivi d'un CAN dans le chemin direct, ainsi qu'un CNA dans le chemin de retour. L'architecture est dite monobit ou multibit en fonction de la résolution du CAN de boucle (ainsi, si le CAN est un comparateur, l'architecture est monobit).

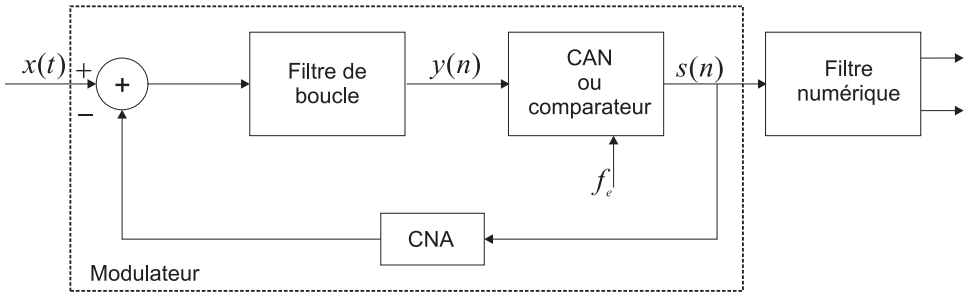

Fig. A.1 – Schéma général d'un convertisseur ΣΔ.

Cette architecture présente deux spécificités qui permettent d'atteindre des résolutions de conversion élevées :

- La fréquence d'échantillonnage f_e du CAN du modulateur est très supérieure à la fréquence de *Nyquist* (on parle de suréchantillonnage) afin de réduire le bruit de quantification sur la bande utile du signal à convertir. En effet, la densité spectrale du bruit est uniforme sur la bande et inversement proportionnelle à f_e (voir § A.2.1). Un ralentissement de cadence (décimation) est ensuite opéré par le filtre numérique.
- Le modulateur effectue une mise en forme du bruit de quantification en repoussant celui-ci en dehors de la bande utile (voir § A.2.3). Dans un second temps, le filtre numérique préserve la bande utile, mais réduit le bruit hors bande. On obtient alors une résolution plus élevée que celle du CAN de boucle.

A.2 Concepts élémentaires de la conversion ΣΔ

A.2.1 Quantificateur

Le quantificateur (CAN) dans la boucle du modulateur ΣΔ effectue la transition entre le monde analogique et le monde numérique en quantifiant le signal en entrée sur un nombre de bits fini N_b. Le quantificateur est un élément non linéaire : deux signaux à l'entrée du quantificateur dont la différence en valeur absolue est inférieure à $\frac{q}{2}$ sont représentés par la même valeur en sortie, q est le pas de quantification. Par conséquent, la fonction de quantification est une fonction irréversible. En supposant le bruit de quantification uniformément distribué dans ces segments, sa puissance moyenne est égale à sa variance [23, 45, 64]. Le bruit de quantification est assimilé à un bruit blanc si les conditions de Bennett [28] sont respectées :

1. le quantificateur ne sature pas,
2. les niveaux de quantification sont suffisamment grands et équi-répartis dans l'intervalle de fonctionnement,
3. l'amplitude du signal d'entrée est uniformément répartie sur toute la plage d'utilisation.

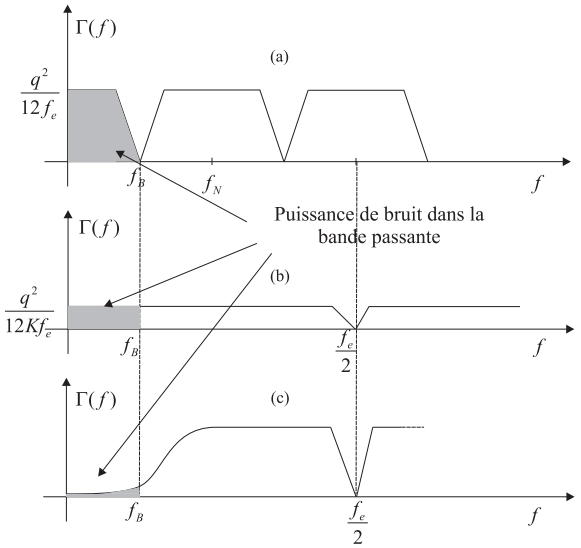

Fig. A.2 – Bruit de quantification dans le cas d'une conversion, (a) à la fréquence de *Nyquist*, (b) avec suréchantillonnage, (c) sigma-delta.

Dans ce cas, la densité spectrale de puissance Γ(f) est constante et uniformément répartie entre $-\frac{f_e}{2}$ et $+\frac{f_e}{2}$ (figure A.2 a), (f_e étant la fréquence d'échantillonnage). Elle s'exprime par :

$$\int_{\frac{-f_e}{2}}^{\frac{f_e}{2}} \Gamma(f)df = \sigma^2_{\varepsilon(n)} \Rightarrow \Gamma(f) = \frac{\sigma^2_{\varepsilon(n)}}{f_e} = \frac{q^2}{12f_e}$$

Pour un signal sinusoïdale en entrée occupant la plage d'amplitude $[-y_{max} \dots +y_{max}]$, la puissance de bruit s'exprime, en normalisant le signal d'entrée par rapport à y_{max}, par :

$$P_{bruit} = \sigma^2_{\varepsilon(n)} = \frac{q^2}{12} = \frac{1}{12}\left(\frac{2}{2^{N_b}-1}\right)^2$$

Pour un nombre de bits élevé ($N_b \geq 3$), la puissance de bruit est approchée par l'equation :

$$P_{bruit} = \frac{1}{12}\left(\frac{2}{2^{N_b}-1}\right)^2 \approx \frac{1}{12}\left(\frac{2}{2^{N_b}}\right)^2 = \frac{1}{3\times 4^{N_b}}$$

Dans ce cas, la densité spectrale de bruit s'exprime par :

$$\Gamma(f) = \frac{P_{bruit}}{f_e} = \frac{1}{3\times 4^{N_b} f_e} \tag{A.1}$$

Le rapport signal sur bruit maximal SNR_{max} (Signal to Noise Ratio) est donné par :

$$\begin{aligned} P_{signal} &= \frac{y^2_{max}}{2} \\ P_{bruit} &= \frac{q^2}{12} \approx \frac{1}{12}\left(\frac{2y_{max}}{2^{N_b}}\right)^2 \\ SNR_{max}(dB) &= 10log\left(\frac{P_{signal}}{P_{bruit}}\right) = 1.76 + 6.02N_b \end{aligned} \tag{A.2}$$

Selon l'équation A.2, un bit supplémentaire du CAN augmente le SNR de 6 dB.

A.2.2 Suréchantillonnage

Le suréchantillonnage consiste à échantillonner le signal d'entrée du convertisseur de boucle à une fréquence très grande par rapport à la fréquence de Nyquist f_N ($f_N = 2f_B$, f_B est largeur de bande du signal). Cette opération permet :

- d'étaler la puissance de bruit de quantification sur une gamme de fréquence plus large (figure A.2 (b)) en diminuant ainsi la puissance de bruit dans la bande du signal utile et par conséquent améliorer la résolution,
- de diminuer les contraintes sur le filtre anti-repliement à l'entrée du convertisseur en augmentant la zone de transition.

Si on échantillonne le signal d'entrée à une fréquence K fois supérieure à la fréquence de *Nyquist*, on va alors diviser la densité spectrale de puissance par K qui va cette fois s'étaler entre $\pm\frac{K f_N}{2}$ (figure A.2 (b)). Le SNR dans ce cas s'exprime par :

$$\begin{aligned} P_{bruit} &= \frac{q^2}{12K} \\ SNR_{max}(dB) &= 10Log\left(\frac{P_{signal}}{P_{bruit}}\right) = 1.76 + 6.02N_b + 10Log\,(K) \end{aligned}$$

On appelle K le rapport de suréchantillonnage qui est souvent noté OSR (*Over Sampling Ratio*) et défini par :

$$OSR = \frac{f_e}{f_N} = \frac{f_e}{2f_B} \tag{A.3}$$

où f_B est la bande utile du système.

Par exemple, si on utilise un OSR égal à 4, cela revient à diminuer le bruit de quantification dans la bande de 6 dB d'où un gain de 1 bit. Une augmentation de résolution de 10 bits revient à échantillonner le signal à une fréquence égale à $10^6 f_N$. Le suréchantillonnage permet d'augmenter la résolution effective d'un quantificateur. Toutefois, pour des raisons technologiques et de consommation, la fréquence d'échantillonnage ne peut être augmentée indéfiniment. Ainsi à fréquence de suréchantillonnage maximale, c'est l'ordre du filtre de boucle du modulateur qui devra être augmenter pour accroître les performances. C'est ce que nous abordons plus en détails au paragraphe suivant.

A.2.3 Principe de fonctionnement

Comme nous l'avons évoqué en introduction de cette annexe, le modulateur ΣΔ (figure A.1) est composé d'un filtre de boucle qui réalise la mise en forme du bruit de quantification, d'un quantificateur et d'une boucle de rétroaction avec un CNA. Ainsi, la sortie quantifiée est soustraite du signal d'entrée. Le filtre de boucle sert à minimiser l'écart moyen entre le signal d'entrée et sa valeur quantifiée ; de cette manière, le signal de sortie va tendre à suivre l'évolution du signal d'entrée. Ce filtre joue un double rôle : comme dans toute rétroaction, il assure le gain dans la boucle et détermine en plus la bande passante du bruit rejeté. Ce filtre peut être un filtre passe-bas (intégrateur) ou passe-bande (résonateur). Généralement, l'ordre du modulateur (L) est défini par celui du filtre de boucle. Celui-ci est directement lié au nombre d'intégrateurs ou de résonateurs (m) qu'il contient :

$$L = \begin{cases} m & \text{filtre passe-bas} \\ 2m & \text{filtre passe-bande} \end{cases}$$

Plus l'ordre du modulateur augmente, moins le bruit est important dans la bande ce qui améliore la résolution du modulateur. Cependant, dès que le nombre d'intégrateurs ou de résonateurs (m) est supérieur à 2, on risque d'avoir des problèmes de stabilité du modulateur [23]. Plus la résolution du CAN est grande, meilleure est la précision du modulateur ΣΔ. Cependant, le nombre de niveaux du CNA dans le chemin de retour augmente. Ceci a pour effet d'induire des erreurs de non-linéarité non négligeables qui vont dégrader les performances globales du modulateur. Les premiers modulateurs réalisés étaient des modulateurs monobit du fait de leur simplicité et de leur linéarité. Aujourd'hui, la réalisation des modulateurs multibits devient une priorité pour obtenir des performances qui ne peuvent jamais être atteintes au moyen de structures monobit. En résumé un modulateur ΣΔ est caractérisé par les paramètres suivants :

- le type de réalisation : discret ou continu,
- le type du filtre de boucle : passe-bas ou passe-bande,
- la résolution du CAN (N_b),
- la fréquence de suréchantillonnage (f_e),
- le facteur de suréchantillonnage ($OSR = \frac{f_e}{2f_B}$),
- l'ordre (L).

Le CAN dans la boucle du modulateur est un élément non linéaire. Pour exprimer la fonction de transfert de bruit et faciliter l'étude de la stabilité, le CAN peut être substitué par une source de bruit blanc additive et uniforme q(n) (figure A.3).

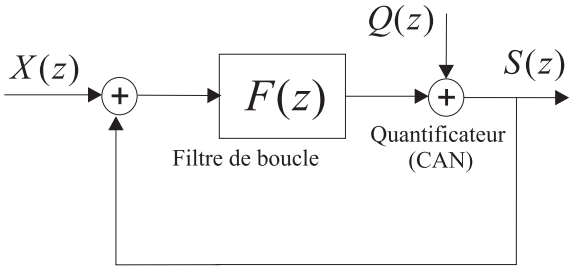

Fig. A.3 – Modèle linéaire du modulateur ΣΔ.

Dans le cas d'un modulateur ΣΔ, ces hypothèses ne sont pas totalement vérifiées, mais dans beaucoup de cas, les résultats obtenus forment un bon modèle de la réalité surtout pour des modulateurs d'ordre 2 et plus, ainsi que pour des modulateurs multibits. Dans la figure A.3 nous avons supposé que le CNA du chemin de retour est idéal sans défauts d'appariement. Avec ce modèle linéaire, la sortie du modulateur $S(z)$ peut s'exprimer à partir de la fonction de transfert du signal $STF(z)$ (Signal Transfer Function) ainsi que celle du bruit $NTF(z)$ (Noise Transfer Function) par :

$$S(z) = STF(z)X(z) + NTF(z)Q(z) \qquad avec \qquad \begin{cases} STF(z) = \frac{S(z)}{X(z)} = \frac{F(z)}{1+F(z)} \\ NTF(z) = \frac{S(z)}{Q(z)} = \frac{1}{1+F(z)} \end{cases} \quad (A.4)$$

Selon le choix du filtre de boucle F(z) (passe-bas ou passe-bande), La STF laisse passer le signal dans une certaine bande de fréquence, alors que la NTF atténue le bruit de quantification dans cette même bande. Généralement dans un modulateur d'ordre L où $L < 3$, la STF est une fonction retard de la forme z^{-L} et la NTF est une fonction passe-haut de la forme $(1 - z^{-1})^L$ pour le cas passe-bas. Pour des modulateurs d'ordre plus élevé, la formule A.4 reste toujours valable, mais l'expression des fonctions NTF(z) et STF(z) change. Le signal s(n) en sortie du modulateur ΣΔ contient le signal d'origine x(n) plus le bruit de quantification mis en forme. Un filtre numérique placé à la sortie du modulateur est donc nécessaire pour supprimer le bruit de quantification en dehors de la bande et faire passer uniquement le signal dans la bande utile . Un filtre décimateur ramène le débit de sortie à la fréquence de Nyquist.

A.3 Le modulateur ΣΔ passe-bande

La conception du modulateur passe-bande est basée sur la théorie développée précédemment avec le passe-bas. La différence entre un modulateur ΣΔ passe-bas et un modulateur passe-bande réside essentiellement dans la fonction de mise en forme de bruit NTF(z). Dans un modulateur passe-bande, le bruit de quantification doit être minimal non plus à la fréquence nulle, mais à une fréquence f_o correspondant le plus souvent à la fréquence centrale du signal d'entrée. Le passage du passe-bas au passe-bande s'obtient en remplaçant le filtre passe-bas de la boucle F(z) du modulateur ΣΔ par un filtre passe-bande (figure A.4).

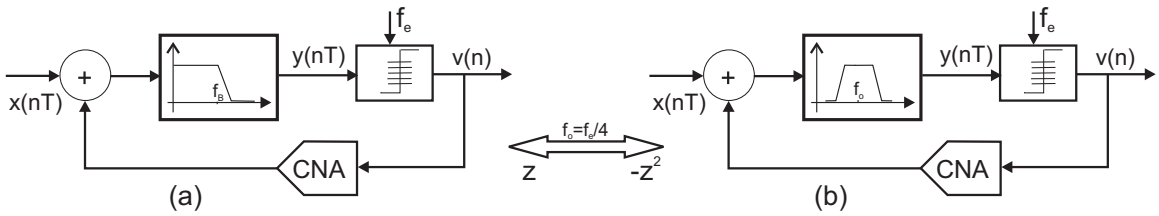

Fig. A.4 – Modulateurs ΣΔ : (a)passe-bas et (b) passe-bande.

Les modulateurs passe-bas ont une mise en forme du bruit de quantification de la forme passe-haut et sont réalisés avec un filtre composé d'intégrateurs. Si nous construisons le filtre de boucle, F(z), à partir de résonateurs, le bruit sera rejeté en dehors de la bande utile centrée sur la fréquence centrale du filtre f_c. Dans ce cas, la mise en forme du bruit de quantification est de type coupe-bande [23].

Un convertisseur ΣΔ passe-bande peut être obtenu à partir d'un convertisseur passe-bas en appliquant la transformation suivante :

$$z \rightarrow -z\frac{z+a}{az+1} \qquad avec \qquad a = -\cos\left(\frac{2\pi f_0}{f_e}\right) \qquad -1 < a < 1$$

On distingue trois cas pour a :

- a = 0, la transformation $z \rightarrow -z^2$ réalise un modulateur passe-bande de fréquence centrale de $\frac{f_e}{4}$,
- a < 0, on obtient des modulateurs passe-bande dont la fréquence centrale est proche des basses fréquences,
- a > 0, on obtient des modulateurs passe-bande dont la fréquence centrale est proche de $\frac{f_e}{2}$.

Pour a = 0, il en résulte un modulateur pour lequel le bruit de quantification est rejeté en dehors de la bande utile centrée à $\frac{f_e}{4}$ [65, 66]. Une application typique d'un tel convertisseur est la conversion analogique numérique d'un signal RF ou IF d'un récepteur de radiocommunication, dont le traitement est représenté à la figure.A.5.

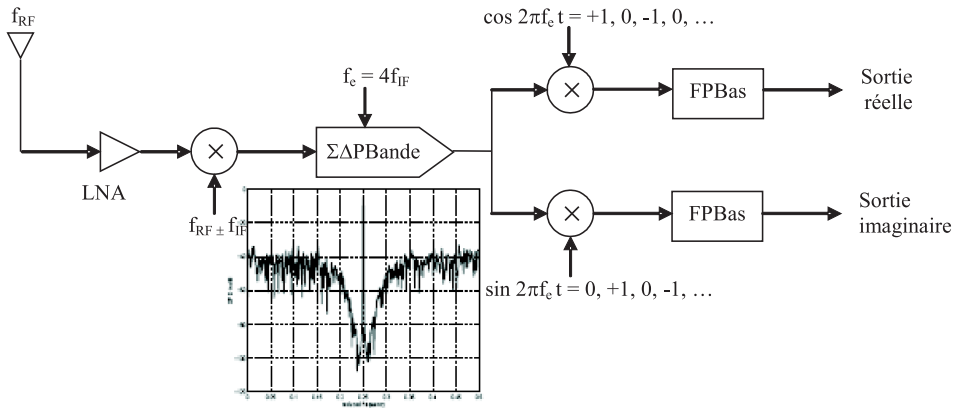

Fig. A.5 – Chaîne de réception radio utilisant un modulateur ΣΔ passe-bande.

En effet, les multiplications par les fonctions sinus et cosinus se résument alors à une multiplication par {0, 1 ou -1}. Ainsi l'étage complexe de démodulation analogique à réaliser est remplacé par quelques portes logiques.

Le spectre à la sortie d'un convertisseur ΣΔ passe-bande est représenté sur la figure.A.5. On peut remarquer que le bruit de quantification est minimal aux alentours de la fréquence centrale égale à $\frac{f_e}{4}$. En général, la possibilité pour un modulateur ΣΔ de convertir un signal en bande étroite à une fréquence non nulle, le rend particulièrement attractif dans les applications radio-fréquence.
La stabilité des modulateurs ΣΔ passe-bande [44] est déduite directement de la stabilité du modulateur correspondant passe-bas. Les conditions de stabilité du modulateur passe-bande seront identiques à celles du passe-bas. Cette constatation peut s'expliquer par le fait que le passage du passe-bas au passe-bande correspond uniquement à une variation de phase des pôles et des zéros dans le plan complexe. Or, les conditions de stabilité sont liées au module des pôles.

A.4 Modulateur ΣΔ à temps continu : passage temps discret-temps continu

Une illustration du modulateur ΣΔ à temps discret et à temps continu est représentée à travers la figure A.6.

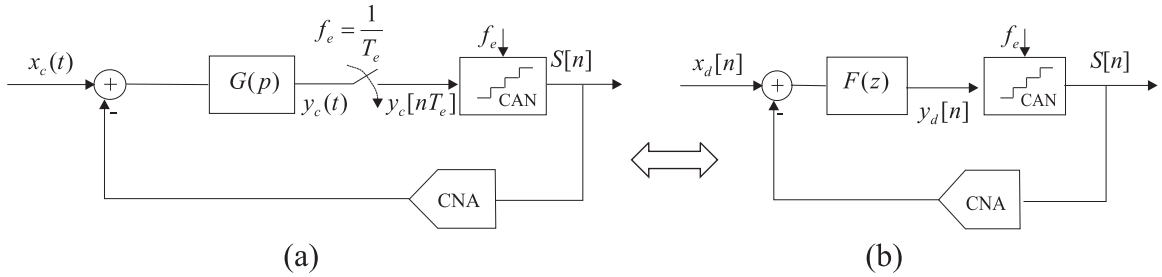

Fig. A.6 – Modulateur ΣΔ à temps continu (a) et son équivalent à temps discret(b).

La différence entre les deux modulateurs se trouve d'une part à l'entrée du modulateur. Celle-ci, dans le modulateur à temps continu, est un signal qui varie dans le temps $x_c(t)$, alors que dans le modulateur à temps discret, le signal d'entrée est échantillonné avant d'entrer dans la boucle : $x_d[n] = x_c(nT)$. D'autre part dans le modulateur à temps continu, l'échantillonnage s'effectue juste avant le quantificateur comme cela est indiqué sur la figure A.6.a.

Les modulateurs ΣΔ à temps continu sont des systèmes mixtes : l'entrée et le filtre de boucle sont à temps continu et leur sortie est à temps discret. Afin de surmonter les problèmes liés à la synthèse des circuits mixtes en raison des difficultés de simulation, un modulateur ΣΔ à temps continu peut être entièrement étudié dans le domaine discret puis être réexaminé en temps continu.

Afin d'obtenir l'équivalent à temps continu d'un modulateur à temps discret, une transformation $p \Rightarrow z$ est nécessaire. Les transformations classiques (transformation bilinéaire, transformation de la dérivation) ne réalisent qu'une correspondance approchée. La transformée de l'invariance impulsionnelle [67] réalise une correspondance exacte aux instants d'échantillonnage. Les deux modulateurs seront équivalents, s'ils produisent la même séquence de sortie, lorsqu'on leur applique à un instant donné le même signal d'entrée. Leurs signaux de sortie seront les mêmes, si on s'assure que les signaux d'entrée de leurs quantificateurs sont les mêmes à chaque instant d'échantillonnage, entre les deux modulateurs. Pour les modulateurs à temps continu et discret, cela signifie :

$$y_d[n] = y_c[nT_e]$$

Cette condition sera satisfaite si les valeurs des réponses impulsionnelles des boucles ouvertes des deux modulateurs sont égales aux instants d'échantillonnage (figure A.7).

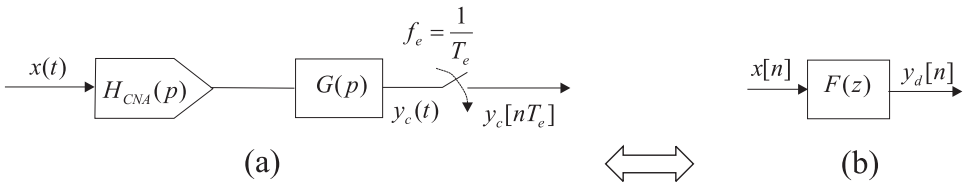

Fig. A.7 – (a) :Boucle ouverte du modulateur à temps continu, (b) : à temps discret.

Ceci sera vrai si la condition suivante est vérifiée [67, 40, 41] :

$$F(z) = Z_T\{L^{-1}[H_{CNA}(p) \cdot G(p)]\}, \tag{A.5}$$

avec

$H_{CNA}(p)$: transformée de Laplace de la réponse à un échantillon du CNA,

Z_T : la transformée en Z, après échantillonnage à la période T

L^{-1} : la transformée de Laplace inverse.

Le CNA du chemin de retour peut être du type NRZ (Non-Return to Zero), RZ (Return to Zero) ou HZ (Half-delay return to Zero). Selon le type de CNA, il y aura différentes fonctions F(z) équivalentes pour un même filtre de boucle G(p).

Dans le cas d'un modulateur à filtres continus, le calcul des fonctions de transfert est plus complexe puisque le signal évolue dans 2 mondes : le monde continu et le monde échantillonné. Pour faire ces calculs, il faut supposer que le signal d'entrée est à spectre borné, majoré par la moitié de la fréquence d'échantillonnage du système. L'expression du signal en sortie s'exprime comme la somme de 2 termes : la contribution par rapport au bruit qui a la même forme que précédemment, et la contribution du signal d'entrée.

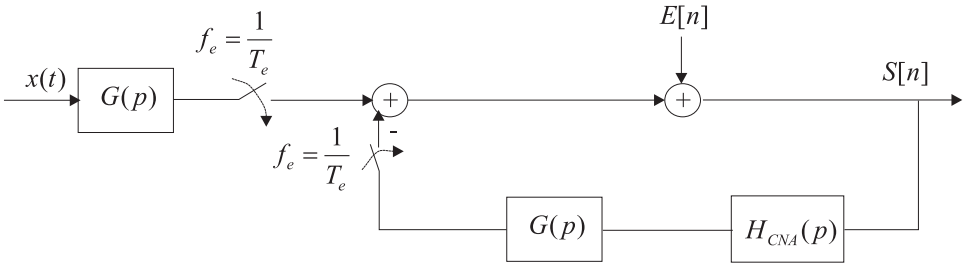

Fig. A.8 – Schéma équivalent du modulateur à temps continu.

En se basant sur le schéma équivalent du modulateur à temps continu représenté par la figure A.8, le signal en sortie s'exprime par :

$$S(z) = \frac{Z_T\left\{L^{-1}\left[G(p)X(p)\right]\right\}}{1 + Z_T\left\{L^{-1}\left[G(p)H_{CNA}(p)\right]\right\}} + \frac{E(z)}{1 + Z_T\left\{L^{-1}\left[G(p)H_{CNA}(p)\right]\right\}} \tag{A.6}$$

Or d'après la méthode de l'invariance impulsionnelle on a

$$Z_T\left\{L^{-1}\left[G(p)H_{CNA}(p)\right]\right\} = F(z)$$

On peut noter que le passage Temps Continu (TC)-Temps Discret (TD) conserve la même fonction de mise en forme du bruit NTF(z) = $\frac{1}{1+F(z)}$. Par contre, la fonction de transfert par rapport au signal ne peut pas s'exprimer dans le domaine z vu que la sortie Y(z) est discrète et l'entrée est continue x(t). La STF sera représentée dans le domaine fréquentielle par l'équation A.7 [40, 41].

$$STF(w) = \frac{S(e^{jw})}{X(jw)} = \frac{G(jw)}{1 + F(e^{jw})} \tag{A.7}$$

A.5 Critères de performances du modulateur ΣΔ

Plusieurs critères ont été définis pour évaluer la performance du modulateur et par conséquent celle du CAN. Les critères les plus importants sont :

1. **Le rapport signal sur bruit maximal** $\mathrm{SNR_{max}}$ (Signal to Noise Ratio) est le critère le plus utilisé pour caractériser la performance d'un modulateur. Ils est exprimé par le rapport entre la puissance du signal utile en entrée et celle du bruit de quantification dans la bande du signal (équation A.8).

$$SNR_{\max}|_{dB} = 10\log_{10}\left[\frac{P_{signal}}{P_{bruit}}\right] \tag{A.8}$$

2. **La Distorsion harmonique totale** THD (Total Harmonic Distortion) : un opérateur linéaire idéal ne modifie pas la composition spectrale du signal de sortie tel qu'il était à l'entrée. Toutes les non linéarités introduites par l'opérateur vont se manifester par une distorsion de la réponse dans le domaine temporel et par l'apparition de nouvelles fréquences dans le domaine fréquentiel. Lorsqu'un CAN convertit une sinusoïde pure, la représentation spectrale du signal de sortie comporte une raie à la fréquence du signal d'entrée mais aussi aux multiples du fondamental qui composent la distorsion harmonique. La distorsion harmonique totale (THD) est définie par :

$$THD_{dB} = 10\log\left[\sum_{i=2}^{\infty}(HD_i)^2\right] \tag{A.9}$$

où la distorsion harmonique $\mathrm{HD_i}$, est l'amplitude de la $i^{ème}$ harmonique ramenée sur l'amplitude de la fondamentale. Le SNDR (Signal to Noise and Distortion Ratio) est le rapport de la puissance du signal à la sortie du modulateur sur la somme de la puissance du bruit et des harmoniques :

$$SNDR_{dB} = 10\log_{10}\left[\frac{P_{signal}}{P_{bruit+harmoniques}}\right] \tag{A.10}$$

Il faut noter que la valeur maximale du SNDR dépend de la fréquence et peut être utilisée pour mesurer la dégradation des performances du modulateur quand la fréquence du signal d'entrée augmente. Le SNR et le SNDR sont représentés sur la figure A.9.

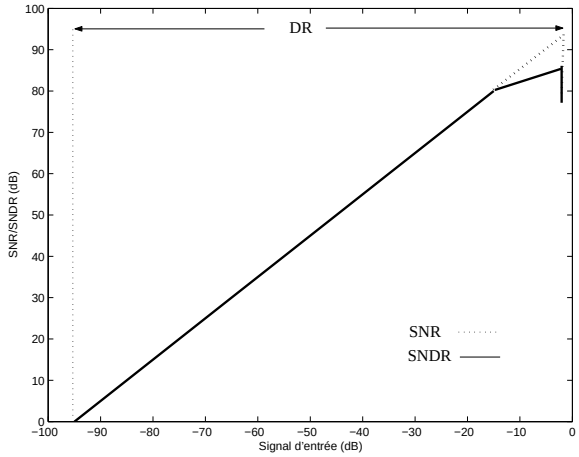

Fig. A.9 – SNR /SNDR et DR.

On peut constater, pour de faibles niveaux du signal d'entrée, que la distorsion harmonique n'est pas importante ce qui implique un SNR et un SNDR approximativement égaux. Plus le niveau du signal d'entrée augmente, plus la distorsion harmonique augmente et le SNDR sera donc plus faible que le SNR.

3. **La dynamique du signal d'entrée** DR (Dynamic Range) est le rapport de puissance entre le niveau maximal du signal d'entrée que le modulateur peut convertir et le niveau minimal du signal d'entrée détectable [48] (figure A.9). Le niveau minimal détectable est celui pour lequel le SNR vaut 0 dB.

Annexe B

Éléments de calcul pour l'architecture EFBD

B.1 Modulateur MSCL

Les structures MSCL (Multi Stage Closed Loop modulators) ont été imaginées par P.Bénabès [44] au cours de son travail de thèse. Ce sont des structures de modulateurs de degré élevé utilisant plusieurs modulateurs simples mis en cascade et rebouclés de sorte que le signal de sortie global soit directement la somme des signaux de sorties des divers étages. Ces structures n'ont pas besoin de prétraitement numérique, cependant une calibration est souvent nécessaire pour compenser certains défauts du bouclage dans le modulateur.

L'architecture de type MSCL d'ordre n (figure B.1) est construite de façon récursive. Elle est composée d'un modulateur d'ordre 1 suivi d'un modulateur d'ordre n-1. Le premier modulateur convertit le signal d'entrée, en donnant une première approximation s_1. Le deuxième modulateur, qui est un MSCL d'ordre $n-1$, échantillonne l'opposé du bruit de quantification du premier étage : $y_1 - s_1$, pour fournir $s_2 + .. + s_n$ et compenser partiellement le bruit apporté par le premier modulateur. La sortie S de l'ensemble est alors la somme $s_1 + .. + s_n$.

Cette structure peut se ramener à une architecture comportant un filtre de transmittance

$$G(z) = \prod_{j=1}^{n} (1 + c_j G_j(z)) - 1,$$

un CAN multibit et un CNA lui aussi multibit (figure B.2) où le nombre de niveaux correspond au nombre d'étages.

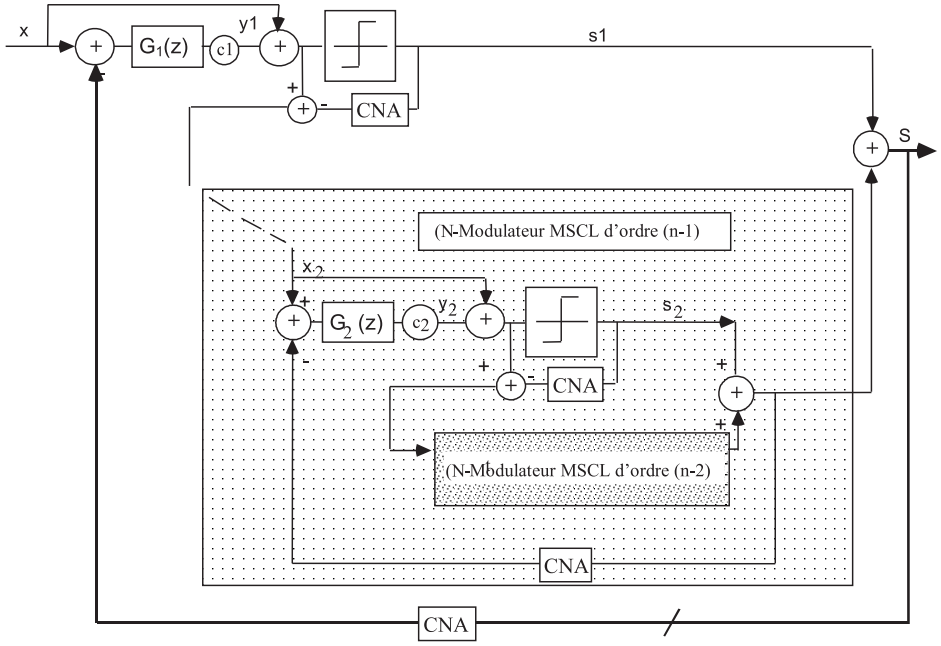

Fig. B.1 – Structure MSCL générale d'ordre n quelconque.

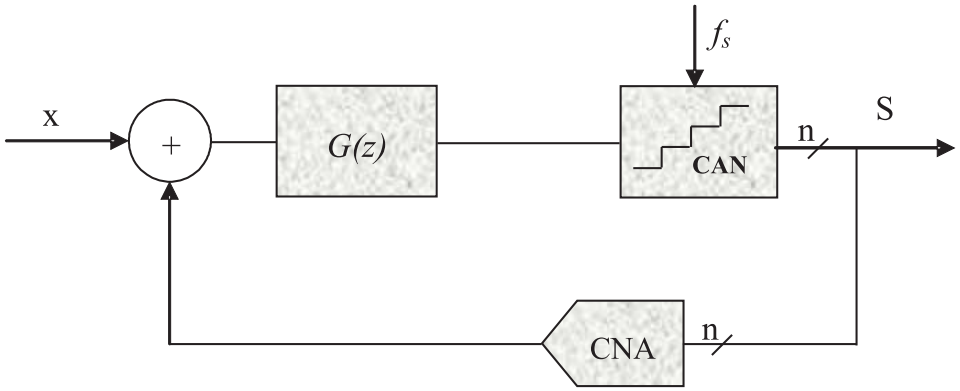

Fig. B.2 – Structure équivalente d'un modulateur MSCL.

Chaque filtre $G_j(z)$ passe-bande a pour transmittance :

$$G_j(z) = \frac{\alpha p_j z^{-1} - z^{-2}}{1 - p_j z^{-1} + z^{-2}} \quad avec \quad p_j = 2\cos\left(2\pi \frac{F_{cr_j}}{F_e}\right) \tag{B.1}$$

F_{cr_j} est la fréquence centrale de chaque résonateur (fréquences correspondant aux zéros de bruit du modulateur) et F_e la fréquence d'échantillonnage. La valeur optimale de α a été déterminée par des études paramétriques [44]. Il a été démontré que $\alpha = \frac{1}{2}$ convient pour les modulateurs dont les fréquences centrales sont proches de $\frac{f_e}{4}$. Pour les modulateurs à fréquences centrales proches de 0 ou de $\frac{f_e}{2}$, la valeur optimale de α doit être ré-évaluée. Les coefficients c_j sont des paramètres très importants dans la synthèse du modulateur. Ils déterminent la précision de celui-ci et sa marge de stabilité [42]. A partir du modèle simplifié du modulateur MSCL, obtenu par le

remplacement de chaque comparateur par son modèle linéaire [28], on obtient le modèle suivant :

$$S(z) = X(z) + B_n(z) \prod_{j=1}^{n} NTF_j(z) \quad avec \quad NTF_j(z) = \frac{1}{1 + c_j G_j(z)} \tag{B.2}$$

La fonction de transfert par rapport au signal (STF) de ce modulateur est égale à l'unité. On remarque que le bruit de quantification des n-1 premiers étages b_1 à b_{n-1} a été complètement éliminé. Seul subsiste le bruit du dernier étage b_n remis en forme par la transmittance NTF(z) :

$$NTF(z) = \prod_{j=1}^{n} NTF_j(z) \tag{B.3}$$

B.1.1 Calcul du module de la NTF pour une structure d'ordre 6 avec des résonateur idéaux (Q infini)

La fonction de transfert par rapport au bruit NTF d'un modulateur ΣΔ passe-bande avec une architecture *MSCL* est donnée par :

$$NTF(z) = \prod_{j=1}^{3} NTF_j(z) = \prod_{j=1}^{3} \frac{1}{1 + c_j G_j(z)} \quad avec \quad \begin{cases} G_j(z) = \frac{\frac{p_j}{2} z^{-1} - z^{-2}}{1 - p_j z^{-1} + z^{-2}} = \frac{A_j(z)}{B_j(z)} \\ NTF_j(z) = \frac{B_j(z)}{B_j(z) + c_j A_j(z)} \end{cases} \tag{B.4}$$

Le dénominateur $B_j(z)$ est égal à :

$$\begin{aligned} B_j(z) &= 1 - p_j z^{-1} + z^{-2} \\ &= z^{-1}(z - p_j + z^{-1}) \\ &= z^{-1}(2\cos(2\pi f) - 2\cos(2\pi f_{cr_j})) \\ &= 4z^{-1}\sin(\pi(f + f_{cr_j}))\sin(\pi(f - f_{cr_j})) \end{aligned}$$

Une approximation de $B_j(z)$ à l'ordre 1 autour de la fréquence centrale f_{cr_j} donne :

$$B_j(z) \approx 4\pi z^{-1} \sin(2\pi f_{cr_j})(f - f_{cr_j})$$

En utilisant l'hypothèse de la bande étroite, le terme $B_j(z)$ tend vers zéro autour de la fréquence centrale f_{cr_j}. Une approximation au premier ordre de la NTF donne :

$$NTF_j(z) \approx \frac{B_j(z)}{c_j A_j(z)} \Rightarrow |NTF_j(z)|^2 = \frac{|B_j(z)|^2}{|c_j A_j(z)|^2}$$

avec $|B_j(z)|^2 = (4\pi)^2 \sin^2(2\pi f_{cr_j})(f - f_{cr_j})^2$ et $|c_j A_j(z)|^2 = c^2 \sin^2(2\pi f_{cr_j})$. Le module au carré de la NTF d'un modulateur passe-bande avec un facteur de qualité infini est donné par l'expression suivante :

$$|NTF_j(z)|^2 = \left(\frac{4\pi}{c_j}\right)^2 (f - f_{cr_j})^2 \tag{B.5}$$

B.1.2 Fonction de transfert du filtre de boucle avec Q fini

La conception des modulateurs à temps continu a été faite par la méthode de l'invariance impulsionnelle dont la forme générale est donnée par l'équation (B.6) pour un CNA de type NRZ (Non Return to Zero).

$$G(z) = (1 - z^{-1}) Z_T \left\{ L^{-1} \left[\frac{G(p)}{p} \right] \right\} \tag{B.6}$$

Ce principe impose qu'à chaque instant d'échantillonnage $t = nT_e$, la réponse de G(z) soit strictement identique à celle de G(p), sans aucune approximation. L'équation B.6 ne permet pas un passage inverse de filtre en z à des filtres en p (G(z) → G(p)). Car avec l'hypothèse posée, il existe une infinité de filtre G(p) qui ont la même réponse aux instants d'échantillonnage que celle délivrée par le filtre G(z). Pour passer des filtres (G(z) → G(p)), des contraintes supplémentaires ont été imposées. Pour un filtre G(z) connu :

- on suppose que l'ordre de G(p) est le même que celui de G(z), les pôles de G(p) sont déduits des pôles de G(z) par la formule $p_k = \frac{1}{T_e} \ln(z_k)$,
- On fait alors le passage de G(p) → (G(z)) via la formule B.6,
- On identifie le résultat obtenu à G(z) pour obtenir le numérateur.

En se basant sur ce principe de synthèse, nous allons calculer le filtre à temps continu qui correspond au filtre de boucle d'une structure MSCL dont la fonction de transfert est donnée par $G_j(z) = \frac{\frac{p_j}{2} z^{-1} - z^{-2}}{1 - p_j z^{-1} + z^{-2}} = \frac{A_j(z)}{B_j(z)}$ avec $p_j = 2\cos(2\pi f_{cr_j})$.
Les pôles de $G_j(z)$ sont $z_{1,2} = \left\{ e^{j2\pi f_{cr_j}}, e^{-j2\pi f_{cr_j}} \right\}$.
La fonction de transfert G(p) correspondante est donnée par l'équation suivante :

$$G(p) = \frac{A(p)}{B(p)} = \frac{\alpha p + a}{(p - p_1)(p - p_2)} = \frac{\alpha p + a}{p^2 + (2\pi f_{cr_j})^2} = \frac{\alpha p + a}{p^2 + w_{cr_j}^2} \tag{B.7}$$

L'équation B.7 est la fonction de transfert d'un résonateur avec un facteur de qualité infini avec un terme passe-bas en a. Cette fonction est irréalisable en pratique. Les résonateurs réalisables possèdent un facteur de qualité Q fini et ont une fonction de transfert G(p) de la forme $G(p) = \frac{\alpha p + a}{p^2 + \frac{w_{cr_j}}{Q} p + w_{cr_j}^2} = \frac{A(p)}{B(p)}$. En tenant compte de cette hypothèse et en se basant sur le principe de la conservation des pôles, la fonction de transfert $G_j(z)$ possédera en dénominateur le polynôme $B_j(z)$ défini par :

$$\begin{aligned} B_j(z) &= (z - e^{p_1 T_e})(z - e^{p_2 T_e}) \\ &= (z - e^{\frac{j w_{cr_j} \sqrt{4Q^2 - 1} - w_{cr_j}}{2Q}})(z - e^{\frac{-j w_{cr_j} \sqrt{4Q^2 - 1} - w_{cr_j}}{2Q}}) \\ &= z^2 - e^{\frac{-w_{cr_j}}{2Q}} \left(2\cos\left(\frac{w_{cr_j} \sqrt{4Q^2 - 1}}{2Q} \right) \right) z + e^{\frac{-w_{cr_j}}{Q}} \end{aligned}$$

Pour un facteur de qualité assez grand, le dénominateur $B_j(z)$ est approché par :

$$B_j(z) \approx z^2 - p_j \left(1 - \frac{w_{cr_j}}{2Q} \right) z + \left(1 - \frac{w_{cr_j}}{Q} \right)$$

La fonction de transfert dans le domaine discret en tenant compte du facteur de qualité est donnée par l'équation suivante :

$$G_j(z) = \frac{\frac{p_j}{2} z^{-1} - z^{-2}}{1 - p_j \left(1 - \frac{\pi f_{cr_j}}{Q_j} \right) z^{-1} + \left(1 - \frac{2\pi f_{cr_j}}{Q_j} \right) z^{-2}} \tag{B.8}$$

B.1.3 Calcul du module de la NTF pour une structure d'ordre 6 avec des résonateurs réels (Q fini)

L'expression de la NTF d'un modulateur ΣΔ d'ordre 6 avec une architecture MSCL en tenant compte du facteur de qualité Q est donnée par l'équation suivante :

$$NTF(z) = \prod_{j=1}^{3} NTF_j(z) = \prod_{j=1}^{3} \frac{1}{1 + c_j G_j(z)} \quad avec \quad \begin{cases} G_j(z) = \frac{\frac{p_j}{2} z^{-1} - z^{-2}}{1 - p_j \left(1 - \frac{\pi f_{cr_j}}{Q_j}\right) z^{-1} + \left(1 - \frac{2\pi f_{cr_j}}{Q_j}\right) z^{-2}} = \frac{A_j(z)}{B_j(z)} \\ NTF_j(z) = \frac{B_j(z)}{B_j(z) + c_j A_j(z)} \end{cases} \tag{B.9}$$

Calcul du numérateur de la $NTF_j(z)$:

$$\begin{aligned} B_j(z) &= 1 - p_j \left(1 - \frac{\pi f_{cr_j}}{Q_j}\right) z^{-1} + \left(1 - \frac{2\pi f_{cr_j}}{Q_j}\right) z^{-2} \\ &= z^{-1} \left(z - 2\cos(2\pi f_{cr_j}) \left(1 - \frac{\pi f_{cr_j}}{Q_j}\right) + \left(1 - \frac{2\pi f_{cr_j}}{Q_j}\right) z^{-1} \right) \\ &= z^{-1} \left((z + z^{-1}) - \frac{2\pi f_{cr_j}}{Q_j} z^{-1} - 2\cos(2\pi f_{cr_j}) \left(1 - \frac{\pi f_{cr_j}}{Q_j}\right) \right) \\ &= z^{-1} \left(2\cos(2\pi f) - \frac{2\pi f_{cr_j}}{Q_j} e^{-j2\pi f} - 2\cos(2\pi f_{cr_j}) \left(1 - \frac{\pi f_{cr_j}}{Q_j}\right) \right) \\ &= z^{-1} \left(2(\cos(2\pi f) - \cos(2\pi f_{cr_j})) + (\cos(2\pi f_{cr_j}) - \cos(2\pi f)) \frac{2\pi f_{cr_j}}{Q_j} + j \frac{2\pi f_{cr_j}}{Q_j} \sin(2\pi f) \right) \\ &= z^{-1} \left(2(\cos(2\pi f) - \cos(2\pi f_{cr_j}))(1 - \frac{\pi f_{cr_j}}{Q_j}) + j \frac{2\pi f_{cr_j}}{Q_j} \sin(2\pi f) \right) \\ &= z^{-1} \left(4\sin(\pi(f - f_{cr_j}))\sin(\pi(f + f_{cr_j}))(1 - \frac{\pi f_{cr_j}}{Q_j}) + j \frac{2\pi f_{cr_j}}{Q_j} \sin(2\pi f) \right) \end{aligned}$$

Le calcul de la performance nécessite une bonne connaissance de la NTF de chaque modulateur autour de ses fréquences centrales. En utilisant une approximation à l'ordre 1 autour de la fréquence centrale de chaque résonateur f_{cr_j} on peut écrire $\sin(\pi(f - f_{cr_j})) \approx \pi(f - f_{cr_j})$ et alors le numérateur $B_j(z)$ s'exprime par :

$$B_j(z) \approx z^{-1} \left(4\pi(f - f_{cr_j}) \sin(2\pi f_{cr_j})(1 - \frac{\pi f_{cr_j}}{Q_j}) + j \frac{2\pi f_{cr_j}}{Q_j} \sin(2\pi f_{cr_j}) \right)$$

Pour un Q assez élevé, Le terme $B_j(z)$ est négligeable devant $A_j(z)$ autour de la fréquence centrale. La NTF est approchée par $NTF_{cr_j}(z) \approx \frac{B_j(z)}{c_j A_j(z)}$. Le module au carré de $B_j(z)$ est donné par l'équation suivante :

$$|B_j(z)|^2 \approx 4\pi^2 \sin^2(2\pi f_{cr_j}) \left(4(f - f_{cr_j})^2 + \left(\frac{f_{cr_j}}{Q_j}\right)^2 \right) \tag{B.10}$$

Calcul du dénominateur de la NTF_j :

$$\begin{aligned} c_j A_j(z) &= c_j \left(\frac{p_j}{2} z^{-1} - z^{-2}\right) = c_j z^{-1} \left(\frac{p_j}{2} - z^{-1}\right) \\ &= c_j e^{-j2\pi f} \left(\cos(2\pi f_{cr_j}) - \cos(2\pi f) + j \sin(2\pi f)\right) \end{aligned}$$

Le module au carré du dénominateur s'exprime par :

$$|c_j A_j(z)|^2 = c_j^2 \left(\left| \cos(2\pi f_{cr_j}) - \cos(2\pi f) \right|^2 + |\sin(2\pi f)|^2 \right)$$

Cette expression est approchée autour de la fréquence centrale par $|\mathrm{c_j A_j(z)}|^2 \approx \mathrm{c_j^2} \left|\sin(2\pi \mathrm{f_{cr_j}})\right|^2$. Le module de la $\mathrm{NTF_{cr_j}}$ au carré est donné par :

$$\left|NTF_{cr_j}\right|^2 \approx \frac{4\pi^2}{(c_j)^2} \left(4(f - f_{cr_j})^2 + \left(\frac{f_{cr_j}}{Q_j} \right)^2 \right) \tag{B.11}$$

Finalement, le module au carré de la NTF s'exprime par :

$$|NTF(z)|^2 = \prod_{j=1}^{3} \left|NTF_{cr_j}(z)\right|^2 \tag{B.12}$$

B.2 Amélioration de l'ENOB avec un dédoublement du nombre de modulateurs pour une architecture FBD passe-bande

La puissance de bruit d'un modulateur idéal (réalisé avec des résonateurs idéaux) a été établie au chapitre 3. Elle s'exprime par :

$$P_{NTF^k} = \left(\frac{4\pi}{c} \right)^{2m} \left[\frac{8}{175} \left(\Delta f_k \right)^{2m+1} \right]$$

m : est le nombre de résonateurs dans le modulateur.
La puissance de bruit totale en sortie avec une architecture à N modulateurs est égale à $\mathrm{P_{bruit} = N \times P_{NTF^k}}$. Lorsque l'on double le nombre de modulateurs, la largeur de bande de fonctionnement de chaque étage $\mathrm{\Delta f_k}$ est divisée par 2. Pour évaluer la performance lorsque le nombre de modulateurs est multiplié par deux, on calcule la différence entre les deux ENOB. Soit $\mathrm{ENOB_{2N}}$ et $\mathrm{ENOB_N}$ les deux résolutions respectivement aux architectures avec N et 2N modulateurs. La différence s'exprime par :

$$ENOB_{2N} - ENOB_N = -\frac{\ln(3 \times P_{NTF_{2N}^k} \times 2N)}{\ln(4)} + \frac{\ln(3 \times P_{NTF_N^k} \times N)}{\ln(4)} = \frac{\ln(\frac{P_{NTF_N^k}}{2 \times P_{NTF_{2N}^k}})}{\ln(4)}$$

or

$$\frac{P_{NTF_N^k}}{2 \times P_{NTF_{2N}^k}} = \frac{\left(\frac{4\pi}{c}\right)^{2m}}{2\left(\frac{4\pi}{c}\right)^{2m}} \frac{\left[\frac{8}{175}\left(\Delta f_k\right)^{2m+1}\right]}{\frac{8}{175}\left(\frac{\Delta f_k}{2}\right)^{2m+1}} = 2^{2m}$$

Par conséquent, le dédoublement du nombre de modulateur augmente la résolution d'une valeur égale au nombre de résonateur m.

$$ENOB_{2N} - ENOB_N = m \tag{B.13}$$

On retrouve ainsi, avec l'architecture FBD passe-bande avec des modulateurs ΣΔ à temps continu, le même résultat que celui obtenu avec les architectures parallèles passe-bas classiques ΤΙΣΔ et ΠΣΔ [4].

B.3 Signaux à bande étroite

B.3.1 Définition

Soit x(t) un signal réel. x(t) est un signal à bande étroite si son spectre est borné et qu'il est nul autour de f = 0. Un exemple de representation spectrale associée à un signal à bande étroite est donné sur la Figure B.3. Ainsi, le spectre du signal est nul en dehors des deux plages de largeur Δf centrées autour de $\pm f_0$.

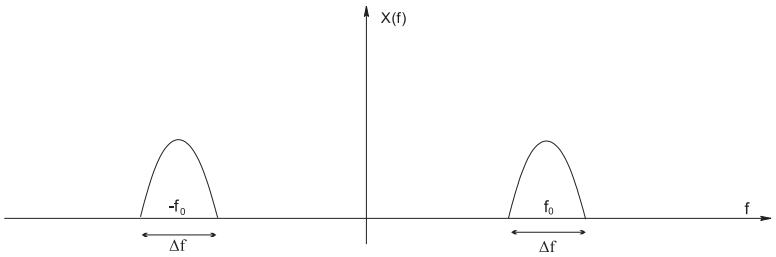

Fig. B.3 – Spectre d'un signal à bande étroite.

B.3.2 Notion de signal analytique

Physiquement, seul le spectre obtenu pour les fréquences positives a un sens. On peut néanmoins associer au signal réel x(t) un signal complexe noté $x_a(t)$ tel que :

$$X_a(f) = \begin{cases} 2X(f) & sif \geq 0 \\ 0 & sif < 0 \end{cases}$$

soit $X_a(f) = X(f) + \text{signe(f)}X(f)$ où la fonction signe(f) vaut - 1 pour f strictement négative et vaut +1 pour f positive. Le spectre de $X_a(f)$ est représenté sur la figure B.4.

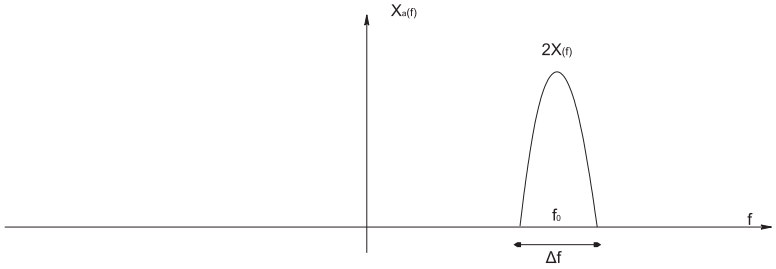

Fig. B.4 – Spectre du signal analytique.

En prenant la transformée de Fourier inverse de $X_a(f)$, on obtient :

$$x_a(t) = x(t) + jTF^{-1}\left[-jsigne(f)X(f)\right] = x(t) + j\frac{1}{\pi t} \otimes x(t)$$

Soit en introduisant TH la transformée de Hilbert de x(t) : $\text{TH}\left[x(t)\right] = \frac{1}{\pi t} \otimes x(t)$, on obtient :

$$x_a(t) = x(t) + jTH\left[x(t)\right]$$

Le signal $x_a(t)$ est, par definition, le signal analytique associé au signal réel x(t) qui représente la partie réelle de $x_a(t)$. Par exemple, le signal analytique associé à $x(t) = A\cos(2\pi f_0 t + \varphi)$ est $x_a(t) = Ae^{j(2\pi f_0 t + \varphi)}$. Soit le signal $x_b(t)$ dont le spectre est représente sur la figure B.5.

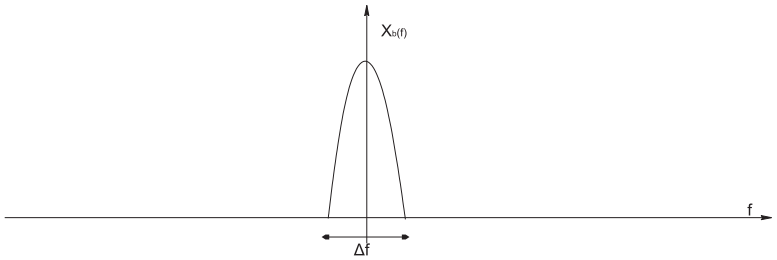

Fig. B.5 – Spectre du signal $x_b(t)$.

$X_a(f)$ apparaît alors comme le translaté de $X_b(f)$ autour de f_0 soit : $X_a(f) = X_b(f - f_0)$ ce qui donne dans le domaine temporel $x_a(t) = x_b(t)e^{j2\pi f_0 t}$ et $x(t) = \Re\left(x_a(t)\right) = \Re\left(x_b(t)e^{j2\pi f_0 t}\right)$. On constate alors que x(t) est un signal haute fréquence (HF) alors que $x_b(t)$ est un signal basse fréquence. $x_b(t)$ est appelé enveloppe complexe de x(t).

B.4 Principe de décimation d'un signal numérisé

Soit une suite d'échantillons notée x(n) qui subit une décimation pour obtenir un signal y(m) tel que y(m) = x(nM) (voir figure B.6), nous nous proposons de trouver la relation entre le spectre de x noté X(f) et le spectre de y noté Y(f).

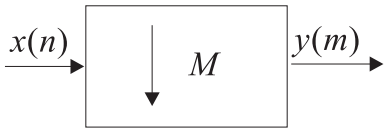

Fig. B.6 – Schéma bloc d'un décimateur.

Nous pouvons toujours construire un nouveau signal noté $x'(n)$ tel que $x'(n) = x(n)$ aux instants d'échantillonnage de y(m) et zéros ailleurs tel que décrit par l'équation suivante :

$$x'(n) = \begin{cases} x(n) & n = pM, \quad p \in Z \\ 0 & \text{sinon} \end{cases} \tag{B.14}$$

Nous pouvons écrire $x'(n)$ d'une manière plus commode grâce à l'équation suivante :

$$x'(n) = x(n)C_M(n), \quad -\infty < n < \infty \tag{B.15}$$

avec

$$C_M(n) = \frac{1}{M}\sum_{k=0}^{M-1} W_M^{-kn} \qquad ,W_M = e^{\frac{-j2\pi}{M}}$$

$C_M(n)$: représente un peigne de dirac de période M.
Nous avons donc l'égalité :

$$y(m) = x'(nM) = x(nM) \tag{B.16}$$

Nous pouvons maintenant écrire la transformée en z de y(m) telle qu'elle est décrite par l'équation suivante :

$$\begin{aligned} Y(z) &= \sum_{m=-\infty}^{+\infty} y(m)z^{-m} \\ &= \sum_{n=-\infty}^{+\infty} x^{'}(nM)z^{-n} \\ &= \sum_{k=-\infty}^{+\infty} x^{'}(k)z^{-\frac{k}{M}} \end{aligned} \tag{B.17}$$

On peut conclure que :

$$Y(z) = X^{'}\left(z^{1/M}\right) \tag{B.18}$$

La transformée en z de $x^{'}(n)$ est donnée par l'équation B.19.

$$\begin{aligned} X^{'}(z) &= \sum_{n=-\infty}^{+\infty} x^{'}(n)z^{-n} = \sum_{n=-\infty}^{+\infty} x(n)C_M(n)z^{-n} \\ &= \frac{1}{M}\sum_{n=-\infty}^{+\infty} x(n)\left(\sum_{k=0}^{M-1} W_M^{-kn}\right)z^{-n} = \frac{1}{M}\sum_{k=0}^{M-1}\sum_{n=-\infty}^{+\infty} x(n)(zW^k)^{-n} \\ &= \frac{1}{M}\sum_{k=0}^{M-1} X(zW^k) \end{aligned} \tag{B.19}$$

La transformée de Fourier du signal décimé y(m) s'exprime alors par :

$$Y(f) = \frac{1}{M}\sum_{k=0}^{M-1} X(z^{1/M}W^k) = \frac{1}{M}\sum_{k=0}^{M-1} X(\frac{f-k}{M}) \tag{B.20}$$

$$Y(f') = \frac{1}{M}\sum_{k=0}^{M-1} X(f' - \frac{k}{M}) \quad \text{on pose } f' = \frac{f}{M} \tag{B.21}$$

L'équation B.21 se traduit dans le domaine fréquentiel par une périodisation du spectre initial aux multiples de la nouvelle fréquence d'échantillonnage soit $\frac{1}{\mathrm{M}}$ et une division de la densité spectrale de puissance par M (voir figure B.7). La puissance totale du signal est conservée pendant cette opération. En fait l'amplitude du signal décimé est la même de celle du signal d'entrée. La division de la densité spectrale par M est compensée par l'élargissement du spectre d'un facteur M de façon à garder une puissance constante.

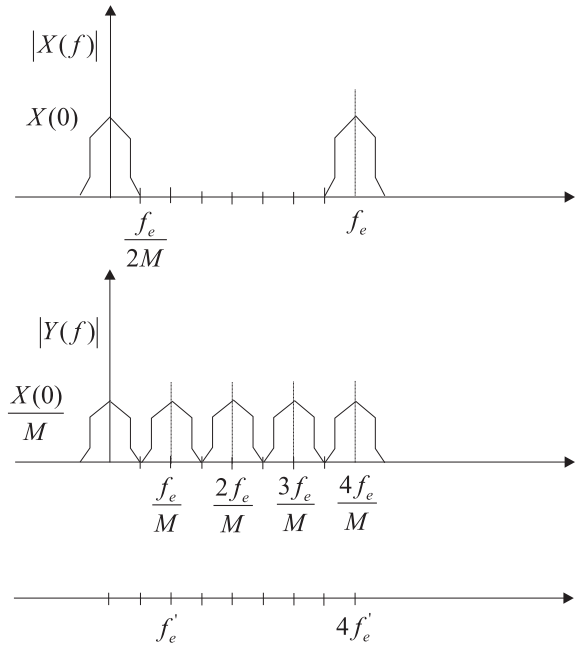

FIG. B.7 – Densité spectrale de puissance du signal avant et après décimation.

Annexe C

Méthodes d'identification paramétriques

C.1 Introduction

L'identification, ou la recherche de modèles à partir de données expérimentales, est une préoccupation majeure dans la plupart des disciplines scientifiques. Elle désigne à la fois une démarche scientifique et un ensemble de techniques visant à déterminer des modèles mathématiques capables de reproduire aussi fidèlement que possible le comportement d'un système physique, chimique, biologique, économique...

Identifier un processus (système), c'est chercher un modèle (dynamique) mathématique, appartenant à une classe de modèles connue, et qui, soumis à des signaux tests (en entrée), donne une réponse (dynamique et statique en sortie), la plus proche possible du système réel (figure C.1).

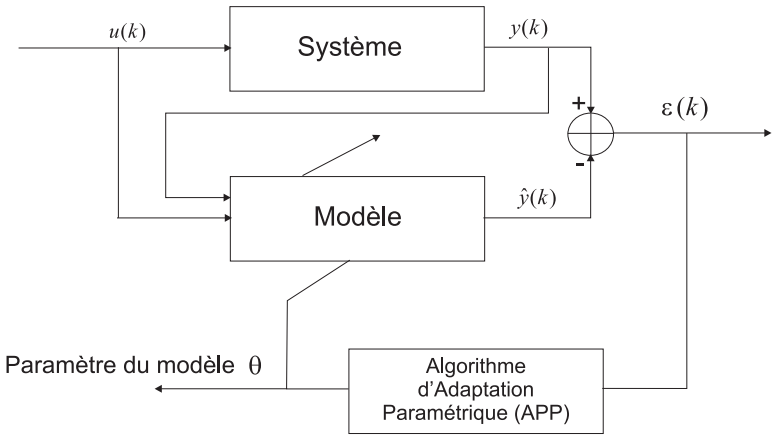

FIG. C.1 – Algorithme d'identification.

La notion de modèle mathématique d'un système, d'un processus ou d'un phénomène, est un concept fondamental. Il existe une multitude de modèles, chacun étant destiné à une application particulière. L'identification est une approche expérimentale pour la détermination du modèle dynamique d'un système. Cette approche peut être décomposée en quatre étapes (voir figure C.2).

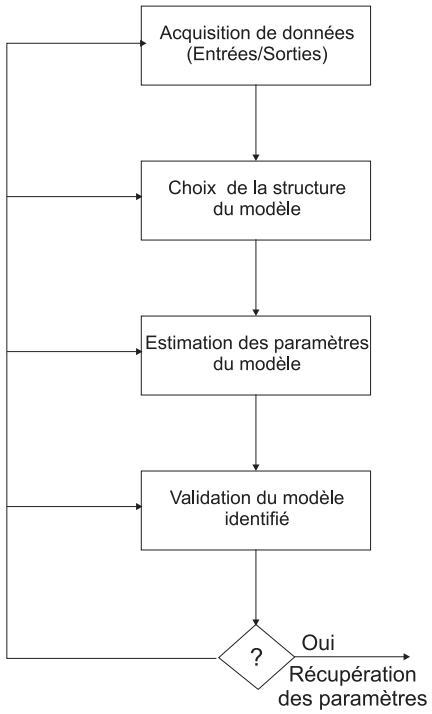

Fig. C.2 – Procédure d'identification d'un modèle de système.

C.2 Structure de modèles

Le modèle d'un système est une description de ses propriétés physiques. On assimile le système à un certain modèle pour qu'on puisse l'identifier. On présente dans ce paragraphe 2 structures permettant de représenter des systèmes physiques ayant une entrée (déterministe) u(k), une entrée stochastique v(k) et une sortie y(k) (figure C.3). Ces structures ont pour caractéristique remarquable de modéliser, avec une dynamique appropriée, l'influence du bruit agissant sur le système. L'ensemble des effets des bruits et perturbations sont représentées par le signal stochastique v(k), lui-même étant généré avec une dynamique H(z) par le signal également stochastique e(k), de type bruit blanc, de distribution normale $N(0, \sigma^2)$.

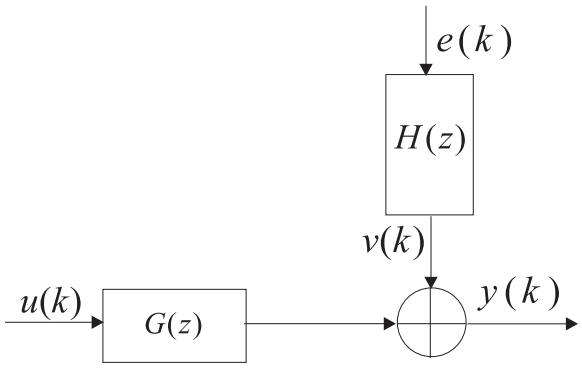

Fig. C.3 – Modèle de structure général.

Avec cette structure générale présentée sur la figure C.3, le signal en sortie s'exprime par :

$$Y(z) = G(z)U(z) + H(z)E(z) \tag{C.1}$$

On se limite ici à la présentation de 2 structures particulières, ARX et ARMAX. On se référera à [68] pour d'autres types de modèle.

C.2.1 Modèle de structure ARX

Dans le cas de la structure ARX « Auto Regressive model with eXternal inputs », le bruit e(k) perturbe la sortie brute de la fonction de transfert G(z) du système via la dynamique

$$H(z) = \frac{1}{A(z)} \tag{C.2}$$

alors que le système lui-même est représenté par :

$$G(z) = \frac{B(z)}{A(z)} = \frac{b_1 z^{-1} + b_2 z^{-2} + ...b_{n_b-1} z^{-n_b+1} + b_{n_b} z^{-n_b}}{1 + a_1 z^{-1} + ...a_{n_a-1} z^{-n_a+1} + a_{n_a} z^{-n_a}} \tag{C.3}$$

on a donc :

$$Y(z) = \frac{B(z)}{A(z)}U(z) + \frac{1}{A(z)}E(z) \tag{C.4}$$

et l'équation aux différences associée à cette structure est donc :

$$\begin{aligned} &\underbrace{y(k) + a_1 y(k-1) + ... + a_{n_a-1} y(k-n_a+1) + a_{n_a} y(k-n_a)}_{Partie \quad auto-régressive} = \\ &\underbrace{b_1 u(k-1) + ... + b_{n_b-1} u(k-n_b+1) + b_{n_b} u(k-n_b)}_{Partie \quad exogène} + \underbrace{e(k)}_{bruit \quad blanc} \end{aligned} \tag{C.5}$$

C.2.2 Modèle de structure ARMAX

Avec la structure ARMAX « Auto Regressive Moving Average with eXternal inputs », on offre comparativement à la structure ARX un degré de liberté supplémentaire pour modéliser la dynamique des perturbations e(k) en les faisant intervenir sur le système avec la fonction de transfert

$$H(z) = \frac{V(z)}{E(z)} = \frac{1 + c_1 z^{-1} + c_2 z^{-2} + ...c_{n_c-1} z^{-n_c+1} + c_{n_c} z^{-n_c}}{1 + a_1 z^{-1} + ...a_{n_a-1} z^{-n_a+1} + a_{n_a} z^{-n_a}} = \frac{C(z)}{A(z)} \tag{C.6}$$

Grâce a C(z), on peut avoir des dynamiques très différentes entre u(k) et y(k) et entre e(k) (bruit blanc gaussien $N(0, \sigma^2)$) et y(k), ce qui compense en partie les lacunes de la structure ARX. On a

$$Y(z) = \frac{B(z)}{A(z)}U(z) + \frac{C(z)}{A(z)}E(z) \tag{C.7}$$

L'équation aux différences correspondante est :

$$\begin{aligned} &\underbrace{y(k) + a_1 y(k-1) + ... + a_{n_a-1} y(k-n_a+1) + a_{n_a} y(k-n_a)}_{Partie \quad auto-régressive} = \\ &\underbrace{b_1 u(k-1) + ... + b_{n_b-1} u(k-n_b+1) + b_{n_b} u(k-n_b)}_{Partie \quad exogène} \\ &+ \underbrace{e(k) + c_1 e(k-1) + ... + c_{n_c-1} e(k-n_c+1) + c_{n_c} e(k-n_c)}_{Moyenne \quad Ajustée} \end{aligned} \tag{C.8}$$

C.3 Estimation des paramètres (PEM)

Après la sélection d'un modèle (ARX, ARMAX) potentiellement capable de représenter le système dynamique linéaire que l'on souhaite identifier ainsi que la nature des perturbations v(k) l'affectant, il reste à déterminer les valeurs numériques de ses paramètres, i.e. à effectuer une identification paramétrique. On présente ici la méthode PEM (« *Prediction-Error identification Method* »), une technique permettant d'obtenir les valeurs numériques des paramètres des fonctions de transfert G(z) et H(z) d'un modèle de structure générale (figure C.3).

La méthode PEM se base sur la comparaison du signal de sortie y(k) du vrai système et de celui du prédicteur $\hat{y}$(k) (sortie du modèle) (figure C.1). Comme son nom le sous-entend, ledit prédicteur $\hat{y}$(k) est conçu de façon à ce qu'il soit en mesure de prédire au mieux le signal de sortie y(k) à l'instant présent en ne se basant que sur les informations disponibles jusqu'à l'instant précédent, *i.e.* à l'instant k − 1. Le prédicteur $\hat{y}$(k) est donné dans sa forme générale ([68] §3.3), dans le domaine z, par :

$$\hat{Y}(z) = \frac{G(z)}{H(z)} U(z) + \left[1 - \frac{1}{H(z)}\right] Y(z) \tag{C.9}$$

La méthode PEM a pour objectif de trouver les paramètres des fonctions de transfert G(z) et H(z) de telle façon que l'erreur de prédiction

$$\varepsilon(k) = y(k) - \hat{y}(k)$$

soit minimisée. Partant d'un ensemble de N mesures y_N(k) correspondant aux entrées u_N(k), on réunit les paramètres de G(z) et H(z) à identifier dans le vecteur-colonne θ , lequel prend dans le cas de la structure ARX la forme :

$$\theta = [a_1, a_2, ..., a_{n_a}, b_1, b_2, ..., b_{n_b}]^\top$$

et on utilise la méthode PEM pour fournir une estimation $\hat{\theta}_N$ de θ minimisant la fonction

$$V_N\left(\theta, y_N(k), u_N(k)\right) = \frac{1}{N} \sum_{k=1}^{N} \ell\left(\varepsilon(k)\right) \tag{C.10}$$

où ℓ est la norme de ε(k) ce qui revient à exprimer le vecteur des paramètres estimés par :

$$\hat{\theta}_N = \arg\min \left\{V_N\left(\theta, y_N(k), u_N(k)\right)\right\} \tag{C.11}$$

L'estimateur $\hat{\theta}_N$ recherché doit donc minimiser la fonction $V_N\left(\theta, y_N(k), u_N(k)\right)$ à partir des signaux d'entrée u_N(k) et de sortie y_N(k), où N correspond au nombre de points de mesure. L'équation (C.11) représente la méthode d'estimation globale donnée par le concept PEM. Plusieurs algorithmes découlent de la méthode PEM suivant le choix de la norme ℓ et du modèle du prédicteur. Un cas particulier très important est celui où la fonction $\ell\left(\varepsilon(k)\right)$ est quadratique :

$$V_N\left(\theta, y_N(k), u_N(k)\right) = \frac{1}{N} \sum_{k=1}^{N} \frac{1}{2} \left(\varepsilon(k)\right)^2 \tag{C.12}$$

Dans la suite, nous allons présenter les algorithmes de calcul des paramètres dans le cas des deux modèles ARX et ARMAX en évoquant le calcul en temps différé (« Off Line ») (à partir d'un ensemble de points de mesure recueillis par le système) et le calcul en temps réel (« On Line ») (à la récolte de chaque point de mesure). Nous avons utilisé, pour les algorithmes de calcul, la boîte à outils *Matlab* « *System Identification Toolbox* » développé par Ljung [69] pour réaliser les simulations. Nous nous intéressons surtout aux deux fonctions *armax* et *rarmax* qui offrent la possibilité d'utiliser tous les algorithmes de l'état de l'art pour l'identification de modèle $ARMAX$ et en particulier du modèle $ARMA$.

C.3.1 Modèle de structure ARX : méthode des moindres carrés

Méthode « Off Line »

La sortie du modèle ARX est représentée par l'équation (C.4). L'estimation $\hat{y}$(k) « naturelle » (qui correspond à l'expression générale C.9) de la sortie du système considéré est fournie par :

$$\begin{array}{l}\hat{y}(k) = -a_1 y(k-1) - ... - a_{n_a-1} y(k-n_a+1) - a_{n_a} y(k-n_a) \\ +b_1 u(k-1) + ... + b_{n_b-1} u(k-n_b+1) + b_{n_b} u(k-n_b)\end{array}$$

La sortie du modèle $\hat{y}$(k) peut s'exprimer sous forme condensée par la formule suivante :

$$\hat{y}(k) = \varphi^\top(k)\theta \tag{C.13}$$

où :
$\varphi^\top$(k) : [−y(k − 1)... − y(k − n_a)u(k − 1)...u(k − n_b)] le vecteur régresseur et
θ : [$a_1 a_2 ... a_{n_a} b_1 b_2 ... b_{n_b}$]$^\top$ le vecteur de paramètre de dimension $n_a + n_b$.
Le prédicteur $\hat{y}$(k) est un régresseur linéaire en paramètre θ. Dans ce cas, où la fonction coût est quadratique , la recherche du vecteur $\hat{\theta}_N$ minimisant V_N est un problème standard en statistique ; il s'agit de la méthode des moindres carrés.

$$\frac{\partial V_N(\theta, y_N(k), u_N(k))}{\partial \theta} = \frac{2}{N}\sum_{k=1}^{N} \varphi(k)\left(y(k) - \varphi^\top(k)\theta\right) = 0$$

Dans ce cas, la solution existe sous forme analytique [70] et elle est donnée par :

$$\begin{array}{rcl}\hat{\theta}_N & = & \arg\min\left\{V_N(\theta, y_N(k), u_N(k))\right\} \\ & = & \left[\frac{1}{N}\sum_{k=1}^{N}\varphi(k)\varphi^\top(k)\right]^{-1}\frac{1}{N}\sum_{k=1}^{N}\varphi(k)y(k)\end{array} \tag{C.14}$$

Ce prédicteur est consistant si [71] :

1. $E\left[\varphi(k)\varphi^\top(k)\right]$ est non singulier,
2. $E\left[\varphi(k)\varepsilon(k)\right] = 0$.

La première condition est souvent vérifiée surtout si l'ordre du modèle est élevé. La deuxième condition signifie que l'erreur de prédiction doit être complètement décorrélée du vecteur de mesure φ(k). C'est une condition nécessaire pour obtenir une solution au sens des moindres carrés. Si ce n'est pas le cas, on fait appel à la méthode de « variables instrumentales » qui introduit une application ζ au vecteur régresseur φ(k) pour le décorréler du bruit ε(k).

Méthode « On Line »

La solution proposée par l'équation (C.14) traite simultanément toutes les N mesures recueillies par le système à identifier. On préfère avoir une méthode récursive (« On Line ») qui traite les mesures successivement et ceci pour deux sortes de raison :

– Les données correspondant aux mesures peuvent être trop nombreuses pour qu'on puisse les stocker simultanément en mémoire vive. On préfère alors les utiliser instantanément et ne conserver en mémoire qu'une quantité limitée d'information, indépendante du nombre des mesures, plutôt que d'avoir à gérer une importante base de données.

– On peut vouloir utiliser les résultats de l'identification pour prendre des décisions immédiates à partir des mesures réalisées, sans devoir attendre de disposer de toutes les données qui vont être recueillies.

Supposons qu'à l'instant N, nous ayons obtenu une estimée $\hat{\theta}_{\mathrm{N}}$ de θ en utilisant la méthode de moindres carrés non récursive (équation (C.14)).

$$\hat{\theta}_N = \left[\sum_{k=1}^{N} \varphi(k)\varphi^\top(k)\right]^{-1} \sum_{k=1}^{N} \varphi(k)y(k) \tag{C.15}$$

À l'instant $\mathrm{N}+1$ arrive une nouvelle mesure $\mathrm{y(N}+1)$, on souhaite modifier $\hat{\theta}_{\mathrm{N}}$ pour tenir compte de cette nouvelle information, sans avoir conservé en mémoire tous les vecteurs φ(k) qui ne font que grandir avec le temps.
Posons :

$$P_N = \left[\sum_{k=1}^{N} \varphi(k)\varphi^\top(k)\right]^{-1}$$

On peut ré-écrire P_N de la façon suivante :

$$P_N = \left[\sum_{k=1}^{N} \varphi(k)\varphi^\top(k)\right]^{-1} = \left[\sum_{k=1}^{N-1} \varphi(k)\varphi^\top(k)\right]^{-1} + \left[\varphi(N)\varphi^\top(N)\right]^{-1} = P_{N-1} + \left[\varphi(N)\varphi^\top(N)\right]^{-1}$$

On déduit l'expression suivante :

$$P_N^{-1} = P_{N-1}^{-1} + \varphi(N)\varphi^\top(N) \tag{C.16}$$

En se basant sur l'expression récursive de $\mathrm{P_N}$, le vecteur des paramètres C.15 s'exprime par :

$$\begin{aligned}\hat{\theta}_N &= P_N \left[\sum_{k=1}^{N} \varphi(k)y(k)\right] = P_N \left[\sum_{k=1}^{N-1} \varphi(k)y(k) + \varphi(N)y(N)\right] = P_N \left[P_{N-1}^{-1}\hat{\theta}_{N-1} + \varphi(N)y(N)\right] \\ &= P_N \left[\left(P_N^{-1} - \varphi(N)\varphi^\top(N)\right)\hat{\theta}_{N-1} + \varphi(N)y(N)\right] = \hat{\theta}_{N-1} + P_N\varphi(N)\left[y(N) - \varphi^\top(N)\hat{\theta}_{N-1}\right]\end{aligned} \tag{C.17}$$

Par conséquent, l'algorithme récursif est donné par l'équation suivante :

$$\hat{\theta}_N = \hat{\theta}_{N-1} + P_N\varphi(N)\varepsilon(N) \tag{C.18}$$

L'inconvénient de ces équations réside dans l'inversion matricielle qui augmente le calcul. Ce problème est résolu en utilisant le lemme d'inversion matricielle suivant :

$$[A + BCD]^{-1} = A^{-1} - A^{-1}B\left[C^{-1} + DA^{-1}B\right]^{-1}DA^{-1}$$

Il suffit de poser dans l'équation (C.16)

$$A = P_{N-1}^{-1}, B = \varphi(N), C = I, D = \varphi^\top(N)$$

pour en déduire

$$P_N = P_{N-1} - \frac{P_{N-1}\varphi(N)\varphi^\top(N)P_{N-1}}{1 + \varphi^\top(N)P_{N-1}\varphi(N)}$$

On a ainsi remplacé l'inversion d'une matrice à chaque pas par la division par le scalaire $1+\varphi^{\top}(\mathrm{N})\mathrm{P}_{\mathrm{N}-1}\varphi(\mathrm{N})$. Finalement, les étapes de calcul de l'algorithme des moindres carrés récursif sont résumées par les trois équations suivantes :

$$K_N = \frac{P_{N-1}\varphi(N)}{1+\varphi^{\top}(N)P_{N-1}\varphi(N)} \tag{C.19}$$

$$P_N = P_{N-1} - K_N\varphi^{\top}(N)P_{N-1} \tag{C.20}$$

$$\hat{\theta}_N = \hat{\theta}_{N-1} + K_N\varepsilon(N) \tag{C.21}$$

La méthode non récursive nécessitait l'inversibilité de la matrice $\left[\sum_{\mathrm{k}=1}^{\mathrm{N}}\varphi(\mathrm{k})\varphi^{\top}(\mathrm{k})\right]$, c'est-à-dire l'identifiabilité du modèle compte tenu des entrées appliquées. Ici, comme il n'y a plus d'inversion de matrice, l'algorithme ne détectera pas les défauts d'identifiabilité, et convergera vers une solution particulière (vers un minimum local et non plus vers le minimum global) dépendant de l'initialisation.

Si les paramètres évoluent lentement dans le temps, il faut pour cela pouvoir oublier les mesures trop anciennes qui correspondent à une situation révolue et viendraient fausser l'estimation. Une façon d'oublier les vieilles mesures est de pondérer l'erreur de prédiction par le coefficient λ avec $0 < \lambda \leq 1$. Dans ce cas le critère à minimiser est :

$$V_N\left(\theta, y_N(k), u_N(k)\right) = \frac{1}{N}\sum_{k=1}^{N}\frac{1}{2}\lambda^{N-k}\left(\varepsilon(k)\right)^2 \tag{C.22}$$

Les étapes de calcul de l'algorithme des Moindres Carrés Récursif (MCR) avec le facteur d'oubli sont les mêmes. L'algorithme MCR avec facteur d'oubli est donné par les équations suivantes :

$$K_N = \frac{P_{N-1}\varphi(N)}{\lambda+\varphi^{\top}(N)P_{N-1}\varphi(N)} \tag{C.23}$$

$$P_N = \frac{1}{\lambda}\left[P_{N-1} - K_N\varphi^{\top}(N)P_{N-1}\right] \tag{C.24}$$

$$\hat{\theta}_N = \hat{\theta}_{N-1} + K_N\varepsilon(N) \tag{C.25}$$

C.3.2 Modèle de structure ARMAX

Méthode « Off Line »

Dans ce paragraphe, on s'intéresse à l'identification des paramètres d'un modèle de structure ARMAX : un système dynamique linéaire discret de fonction de transfert G(z), régi par une équation aux différences (C.8), et soumis à 2 entrées :

- u(k), déterministe (imposable par l'utilisateur)
- e(k), stochastique, traduisant le fait que la sortie brute du système dynamique linéaire discret $\mathrm{G(z)} = \frac{\mathrm{B(z)}}{\mathrm{A(z)}}$, dont on recherche les paramètres, est affectée d'un bruit filtré v(k)

et une sortie y(k).

Le prédicteur du modèle ARMAX est donné par l'équation (C.9). Son équation aux différences s'exprime par :

$$\hat{y}(k) = -\mathrm{a}_1 y(k-1) ... - a_{n_a-1}y(k-n_a+1) - a_{n_a}y(k-n_a) + b_1u(k-1) + ...$$
$$+b_{n_b-1}u(k-n_b+1) + b_{n_b}u(k-n_b) + c_1e(k-1) + ... + c_{n_c-1}e(k-n_c+1) + c_{n_c}e(k-n_c)$$

Le calcul du prédicteur $\hat{y}(k)$ nécessite la connaissance des valeurs précédentes de la variable aléatoire e(k). Comme l'accès à ces valeurs n'est pas possible, elles sont remplacées par les valeurs de l'erreur de prédiction $\varepsilon(k)$. En tenant compte de cette hypothèse, le prédicteur $\hat{y}(k)$ peut se mettre sous forme condensée par :

$$\hat{y}(k) = \varphi^{\top}(k,\theta) \times \theta \tag{C.26}$$

avec :
$\varphi^{\top}(\mathrm{k},\theta) = [-\mathrm{y}(\mathrm{k}-1), ..., -\mathrm{y}(\mathrm{k}-\mathrm{n_a}), \mathrm{u}(\mathrm{k}-\mathrm{n_b}+1), ..., \mathrm{u}(\mathrm{k}-\mathrm{n_b}), \varepsilon(\mathrm{k}-1), ..., \varepsilon(\mathrm{k}-\mathrm{n_c})]$
$\theta = [\mathrm{a}_1, ..., \mathrm{a}_{\mathrm{n_a}-1}, \mathrm{a}_{\mathrm{n_a}}, \mathrm{b}_1, ..., \mathrm{b}_{\mathrm{n_b}-1}, \mathrm{b}_{\mathrm{n_b}}, \mathrm{c}_1, ..., \mathrm{c}_{\mathrm{n_c}-1}, \mathrm{c}_{\mathrm{n_c}}]^{\top}$.
Avec l'hypothèse $\mathrm{e}(\mathrm{k}) = \varepsilon(\mathrm{k})$, le prédicteur $\hat{\mathrm{y}}(\mathrm{k})$ est non linéaire par rapport aux paramètres θ. Ceci est illustré avec l'exemple suivant au premier ordre :

$$\begin{aligned}
\hat{y}(k) &= -a_1 y(k-1) + b_1 u(k-1) + c_1 \varepsilon(k-1) \\
&= -a_1 y(k-1) + b_1 u(k-1) + c_1 \left(y(k-1) - \hat{y}(k-1)\right) \\
&= -a_1 y(k-1) + b_1 u(k-1) + c_1 \left(y(k-1) + a_1 y(k-2) - b_1 u(k-2) - c_1 \varepsilon(k-2)\right) \\
&= \left(c_1 - a_1\right) y(k-1) + a_1 c_1 y(k-2) + b_1 u(k-1) - b_1 c_1 u(k-2) - c_1^2 \left[y(k-2) - \hat{y}(k-2)\right] \\
&= ...
\end{aligned} \tag{C.27}$$

Il ne s'agit malheureusement pas d'une régression linéaire (on parle de régression pseudo-linéaire) et par conséquent, une solution analytique visant à trouver le jeu de paramètres $\hat{\theta}_{\mathrm{N}}$ minimisant $\mathrm{V_N}(\theta, \mathrm{y_N(k)}, \mathrm{u_N(k)})$ n'existe pas. Il faut alors recourir à une solution numérique. Les méthodes numériques sont basées sur le développement limité de la fonction de coût $\mathrm{V_N}$ au voisinage du vecteur de paramètres $\hat{\theta}_{\mathrm{N}}$ afin de construire une direction de recherche et déterminer le vecteur de paramètres qui minimise $\mathrm{V_N}$. L'algorithme de minimisation repose sur le calcul itératif du vecteur de paramètres $\hat{\theta}_{\mathrm{N}}$ et il est donné, d'une façon générale par le système d'équations suivant :

$$\begin{aligned}
\hat{\theta}_N^{i+1} &= \hat{\theta}_N^i - \mu_N^i \left[R_N^i\right]^{-1} V_N^{'}(\hat{\theta}_N^i) \\
V_N^{'}(\hat{\theta}_N^i) &= \left.\frac{\partial V_N(\theta)}{\partial \theta}\right|_{\theta=\hat{\theta}_N^i} = -\frac{1}{N}\sum_{k=1}^{N} \psi(k,\hat{\theta}^{(i)}).\varepsilon(k,\hat{\theta}^{(i)}) \\
\psi(k,\hat{\theta}_N^i) &= \frac{1}{C(z)}\varphi(k,\hat{\theta}_N^i)
\end{aligned} \tag{C.28}$$

On note :
$\theta_{\mathrm{N}}^{\mathrm{i}}$: l'estimation du vecteur de paramètres à l'instant i,
$\mathrm{V}_{\mathrm{N}}^{'}(\hat{\theta}_{\mathrm{N}}^{\mathrm{i}})$: le gradient de la fonction coût $\mathrm{V_N}$ dont le calcul sera détaillé plus loin,
$\mu_{\mathrm{N}}^{\mathrm{i}}$: le pas de l'algorithme,
$\mathrm{R}_{\mathrm{N}}^{\mathrm{i}}$: une matrice carré d'ordre $\mathrm{n_a} + \mathrm{n_b} + \mathrm{n_c}$ et dont le choix dépend de l'ordre du développement limité souhaité par l'algorithme. Suivant la forme de $\mathrm{R}_{\mathrm{N}}^{\mathrm{i}}$, on distingue les algorithmes suivants :

1. **Algorithme du *gradient***
   Cet algorithme est basé sur le développement limité du critère $\mathrm{V_N}$ au premier ordre. L'établissement de cet algorithme se fait par le choix $\mathrm{R}_{\mathrm{N}}^{\mathrm{i}} = \mathrm{I}$ dans l'algorithme général (C.28) où I est la matrice identité. Il est simple à mettre en œuvre, robuste et possède un grand domaine de convergence. Par contre, plus on se rapproche de l'optimum et plus la convergence est lente. Cependant, cet algorithme est bien adapté à la phase initiale des recherches, quand on est loin de l'optimum.

2. **Algorithme de *Newton-Raphson***
   Cet algorithme repose sur un développement limité de $\mathrm{V_N}$ poussé au deuxième ordre. Si l'on prend en compte la dérivée seconde de $\mathrm{V_N}(\theta_\mathrm{N})$, *i.e.* la matrice *Hessienne* de $\mathrm{V_N}$, on peut affiner la direction de recherche en tenant compte de l'évolution du gradient $\mathrm{V_N^{'}}(\hat{\theta}_\mathrm{N})$. Pour appliquer cet algorithme on remplace la matrice $\mathrm{R_N^i}$ par la matrice *Hessienne* exprimée par :

$$V_N^{''}(\hat{\theta}_N^i) = \frac{\partial V_N^{'}(\hat{\theta}_N^i)}{\partial \theta} = \frac{1}{N}\sum_{k=1}^{N}\psi(k,\hat{\theta}^{(i)})\psi^{\top}(k,\hat{\theta}^{(i)}) - \frac{1}{N}\sum_{k=1}^{N}\psi^{'}(k,\hat{\theta}^{(i)}).\varepsilon(k,\hat{\theta}^{(i)}) \qquad \text{(C.29)}$$

   Avec cet algorithme, les calculs requis par chaque itération i sont beaucoup plus lourds qu'avec l'algorithme du gradient.

3. **Algorithme de *Gauss-Newton***
   Cet algorithme consiste à négliger dans l'expression de la matrice *Hessienne* le terme qui dépend de $\psi^{'}(\mathrm{k},\hat{\theta}^{(\mathrm{i})})$ lorsqu'on est proche du minimum de $\mathrm{V_N^{'}}(\hat{\theta}_\mathrm{N}^\mathrm{i})$:

$$V_N^{''}(\hat{\theta}_N^i) \approx \frac{1}{N}\sum_{k=1}^{N}\psi(k,\hat{\theta}^{(i)})\psi^{\top}(k,\hat{\theta}^{(i)}) \qquad \text{(C.30)}$$

   Cette dernière approximation permet d'alléger le calcul. Dans certains cas, la matrice $\mathrm{V_N^{''}}(\hat{\theta}_\mathrm{N}^\mathrm{i})$ peut être singulière. C'est le cas où le modèle est sur-paramétré ou bien la quantité de données recueillies n'est pas suffisante. Le moyen pour surmonter ce problème est la méthode de *Levenberg-Marquardt* qui consiste à ajouter $\alpha\mathrm{I}$ (I est la matrice identité) à l'expression (C.30) pour assurer l'inversion matricielle. Le coefficient α est considéré positif de valeur d'ordre 10^{-6} à 10^{-3}. La valeur de α contrôle la convergence ([68], p. 329).

Le calcul du gradient pour un modèle $ARMAX$ dans le cas d'un critère quadratique :
Le calcul du gradient $\mathrm{V_N^{'}}(\hat{\theta}_\mathrm{N}^\mathrm{i})$ consiste à évaluer l'ensemble des paramètres constituant le vecteur $\mathrm{V_N^{'}}$; $\left\{\frac{\partial \mathrm{V_N}}{\partial \mathrm{a_j}}, \frac{\partial \mathrm{V_N}}{\partial \mathrm{b_j}}, \frac{\partial \mathrm{V_N}}{\partial \mathrm{c_j}}\right\}$. L'évaluation de ces dérivées partielles, dans le cas d'une fonction coût $\mathrm{V_N}$ quadratique (équation (C.12)), donne :

$$\frac{\partial V_N}{\partial a_j} = \frac{1}{2}\overbrace{\frac{\partial V_N}{\partial \varepsilon}}^{2\varepsilon(k,\hat{\theta})}\overbrace{\frac{\partial \varepsilon}{\partial \hat{y}}}^{-1}\frac{\partial \hat{y}}{\partial a_j} = -\frac{1}{N}\sum_{k=0}^{N-1}\frac{\partial \hat{y}}{\partial a_j}\varepsilon(k,\hat{\theta})$$

$$\frac{\partial V_N}{\partial b_j} = \frac{1}{2}\frac{\partial V_N}{\partial \varepsilon}\frac{\partial \varepsilon}{\partial \hat{y}}\frac{\partial \hat{y}}{\partial b_j} = -\frac{1}{N}\sum_{k=0}^{N-1}\frac{\partial \hat{y}}{\partial b_j}\varepsilon(k,\hat{\theta})$$

$$\frac{\partial V_N}{\partial c_j} = \frac{1}{2}\frac{\partial V_N}{\partial \varepsilon}\frac{\partial \varepsilon}{\partial \hat{y}}\frac{\partial \hat{y}}{\partial c_j} = -\frac{1}{N}\sum_{k=0}^{N-1}\frac{\partial \hat{y}}{\partial c_j}\varepsilon(k,\hat{\theta})$$

Afin d'exprimer le gradient, nous avons besoin de connaître $\frac{\partial \hat{\mathrm{y}}}{\partial \mathrm{a_j}}, \frac{\partial \hat{\mathrm{y}}}{\partial \mathrm{b_j}}$ et $\frac{\partial \hat{\mathrm{y}}}{\partial \mathrm{c_j}}$. Le prédicteur $\hat{\mathrm{y}}$ s'exprime dans le cas d'un modèle ARMAX (équation (C.9)) par :

$$C(z)\hat{y}(k) = B(z)u(k) + \left(C(z) - A(z)\right)y(k)$$

De cette équation on peut exprimer les différentes dérivées partielles par :

$$C(z)\frac{\partial \hat{y}}{\partial a_j} = -z^{-j}y(k) = -y(k-j)$$

$$C(z)\frac{\partial \hat{y}}{\partial b_j} = z^{-j}u(k) = u(k-j)$$

$$C(z)\frac{\partial \hat{y}}{\partial c_j} + z^{-j}\hat{y}(k) = y(k-j)$$

De ce fait, les composants du vecteur gradient seront exprimés par :

$$\frac{\partial V_N}{\partial a_j} = \frac{1}{N}\sum_{k=0}^{N-1}\frac{1}{C(z)}y(k-j)\varepsilon(k,\hat{\theta})$$

$$\frac{\partial V_N}{\partial b_j} = \frac{1}{N}\sum_{k=0}^{N-1}\frac{1}{C(z)}u(k-j)\varepsilon(k,\hat{\theta})$$

$$\frac{\partial V_N}{\partial c_j} = \frac{1}{N}\sum_{k=0}^{N-1}\frac{1}{C(z)}\varepsilon(k-j,\hat{\theta})\varepsilon(k,\hat{\theta})$$

En introduisant :

$$\varphi^{\top}(k,\theta) = [-y(k-1),...,-y(k-n_a),u(k-1),...,u(k-n_b),\varepsilon(k-1),...,\varepsilon(k-n_c)]$$

ainsi que

$$\psi^{\top}(k,\theta) = \left[\frac{\partial \hat{y}}{\partial a_1},\frac{\partial \hat{y}}{\partial a_2},...,\frac{\partial \hat{y}}{\partial a_{n_a}},\frac{\partial \hat{y}}{\partial b_1},\frac{\partial \hat{y}}{\partial b_2},...,\frac{\partial \hat{y}}{\partial b_{n_b}},\frac{\partial \hat{y}}{\partial c_1},\frac{\partial \hat{y}}{\partial c_2},...,\frac{\partial \hat{y}}{\partial c_{n_c}}\right]$$

on peut encore écrire :

$$\psi(k,\theta) = \frac{1}{C(z)}\varphi(k,\theta) \tag{C.31}$$

L'expression du vecteur du gradient en tenant compte de l'équation (C.31) est :

$$V_N'(\hat{\theta}_N^i) = -\frac{1}{N}\sum_{k=1}^{N}\psi(k,\hat{\theta}^{(i)}).\varepsilon(k,\hat{\theta}^{(i)}) \tag{C.32}$$

Méthode « On Line »

L'estimation des paramètres du modèle ARMAX d'une façon récursive peut être établi suivant l'une des trois approches suivantes :

1. *RPEM* ***(Recursive Prediction Error identification Method)***
   Cette approche se base sur le calcul des paramètres avec la méthode PEM. Dans le cas où le prédicteur n'est pas linéaire par rapport aux paramètres (modèle ARMAX), le calcul des paramètres s'effectue de façon itérative avec la méthode de *Newton-Raphson*. Cet algorithme annoncé par le système d'équation (C.28) repose sur le calcul itératif en traitant à chaque itération k la totalité des données N comme le montre l'expression suivante :

$$\theta_N^k = \theta_N^{k-1} + \mu_N^k\left[R_N^k\right]^{-1}V_N'\left(\theta_N^{k-1}\right) \tag{C.33}$$

   L'idée de récursivité vient du besoin de recalculer l'expression C.33 à chaque fois que l'on a recueilli une information au lieu d'attendre la collecte de N points de mesure. Ceci se traduit par la modification de l'expression C.33 pour obtenir l'expression récursive suivante :

$$\theta_k^k = \theta_{k-1}^{k-1} + \mu_k^k\left[R_k^k\right]^{-1}V_k'\left(\theta_{k-1}^{k-1}\right) \tag{C.34}$$

Afin d'alléger l'écriture des expressions, on pose $\theta_{\mathrm{k}}^{\mathrm{k}} = \theta(\mathrm{k})$, $\mu_{\mathrm{k}}^{\mathrm{k}} = \mu(\mathrm{k})$ et $\mathrm{R}_{\mathrm{k}}^{\mathrm{k}} = \mathrm{R}(\mathrm{k})$.
Pour établir la récursivité de l'algorithme de *Newton-Raphson*, il faut trouver les relations récursives qui gèrent le calcul du vecteur du gradient $\mathrm{V}_{\mathrm{k}}^{'}$ et de la matrice R(k). Pour élaborer ces relations, nous partons de l'expression globale de la fonction coût à minimiser afin d'obtenir un algorithme général. La fonction coût est donnée par son expression générale :

$$V_N\left(\theta, Z^N\right) = \gamma(N)\frac{1}{2}\sum_{i=1}^{N}\beta(N,i)\varepsilon^2(i,\theta) \tag{C.35}$$

$\mathrm{V_N}$: est l'erreur quadratique moyenne pondérée par les coefficients $\beta(\mathrm{N},\mathrm{i})$.
$\mathrm{Z^N}$: est le vecteur des données accessible jusqu'à l'instant N
$\beta(\mathrm{N},\mathrm{i})$: le coefficient de pondération avec $\begin{cases} \beta(N,i) = \lambda(N)\beta(N-1,i), & 0 \leq \mathrm{i} \leq \mathrm{N}-1 \\ \beta(N,N) = 1 \end{cases}$
λ : est le facteur d'oubli
$\gamma(\mathrm{N})$: est le facteur de normalisation avec $\gamma(\mathrm{N}) = \left[\sum_{\mathrm{i}=1}^{\mathrm{N}}\beta(\mathrm{N},\mathrm{i})\right]^{-1}$
Le facteur $\gamma(\mathrm{N})$ peut être calculé d'une façon récursive comme le montre l'équation suivante :

$$\begin{aligned} [\gamma(N)]^{-1} &= \sum_{i=1}^{N}\beta(N,i) = \sum_{i=1}^{N-1}\beta(N,i) + \beta(N,N) \\ &= \lambda(N)\sum_{i=1}^{N-1}\beta(N-1,i) + 1 = \lambda(N)\left[\gamma(N-1)\right]^{-1} + 1 \end{aligned} \tag{C.36}$$

Dans ce cas, le gradient (voir calcul précèdent) s'exprime par :

$$V_N^{'}\left(\theta, Z^N\right) = -\gamma(N)\sum_{i=1}^{N}\beta(N,i)\psi(i,\theta)\varepsilon(i,\theta)$$

En se basant sur l'expression ci dessus, le gradient peut se mettre sous la forme récursive suivante :

$$\begin{aligned} & V_N^{'}\left(\theta, \mathrm{Z}^N\right) = -\gamma(N)\sum_{i=1}^{N}\beta(N,i)\psi(i,\theta)\varepsilon(i,\theta) \\ & = -\gamma(N)\left[\sum_{i=1}^{N-1}\beta(N,i)\psi(i,\theta)\varepsilon(i,\theta) + \overbrace{\beta(N,N)}^{=1}\psi(N,\theta)\varepsilon(N,\theta)\right] \\ & = \gamma(N)\left[\frac{\lambda(N)}{\gamma(N-1)}\overbrace{\left(-\gamma(N-1)\sum_{i=1}^{N-1}\beta(N-1,i)\psi(i,\theta)\varepsilon(i,\theta)\right)}^{V_N^{'}(\theta,\mathrm{Z}^{N-1})} - \psi(N,\theta)\varepsilon(N,\theta)\right] \\ & = \gamma(N)\left[\left(\frac{1}{\gamma(N)} - 1\right)V_N^{'}\left(\theta, \mathrm{Z}^{N-1}\right) - \psi(N,\theta)\varepsilon(N,\theta)\right] \end{aligned} \tag{C.37}$$

alors $\mathrm{V}_{\mathrm{N}}^{'}\left(\theta, \mathrm{Z}^{\mathrm{N}}\right)$ s'exprime par :

$$V_N^{'}\left(\theta, Z^N\right) = V_{N-1}^{'}\left(\theta, Z^{N-1}\right) + \gamma(N)\left[-\psi(N,\theta)\varepsilon(N,\theta) - V_{N-1}^{'}\left(\theta, Z^{N-1}\right)\right] \tag{C.38}$$

On suppose que le vecteur de paramètres $\hat{\theta}_{\mathrm{N}-1}$ minimise le critère $\mathrm{V}_{\mathrm{N}-1}\left(\theta, \mathrm{Z}^{\mathrm{N}-1}\right)$. Ce qui revient à dire :

$$V_{N-1}^{'}\left(\theta, Z^{N-1}\right) = 0$$

Cette condition est une grande **approximation** parce qu'elle repose sur la minimisation de l'erreur quadratique ε^2 à chaque itération, ce qui n'entraîne pas forcément la minimisation de la fonction coût globale donnée par l'équation (C.35). Avec cette dernière approximation, le vecteur du gradient à l'instant N est égal à :

$$V_N^{'}\left(\theta, Z^N\right) = -\gamma(N)\psi(N,\theta)\varepsilon(N,\theta)$$

Il reste à déterminer la relation récursive qui gère R(k). Comme dans le cas « *Off Line* » on distingue deux types d'algorithmes suivant le choix de R(k).

(a) **Algorithme de *Gauss-Newton***
Dans ce cas R(k) est le *Hessien* et il est approximé par l'équation suivante :

$$\begin{aligned}
R(N) &= \gamma(N)\sum_{i=1}^{N}\beta(N,i)\psi(i,\theta)\varepsilon(i,\theta) = \gamma(N)\left[\sum_{i=1}^{N-1}\beta(N,i)\psi(i,\theta)\varepsilon(i,\theta) + \psi(N,\theta)\varepsilon(N,\theta)\right] \\
&= \gamma(N)\left[\frac{\lambda(N)}{\gamma(N-1)}\overbrace{\left(\gamma(N-1)\sum_{i=1}^{N-1}\beta(N-1,i)\psi(i,\theta)\varepsilon(i,\theta)\right)}^{R(N-1)} + \psi(N,\theta)\varepsilon(N,\theta)\right] \\
&= \gamma(N)\left[\left(\frac{1}{\gamma(N)} - 1\right)R(N-1) + \psi(N,\theta)\varepsilon(N,\theta)\right] \\
&= R(N-1) + \gamma(N)\left[\psi(N)\varepsilon(N) - R(N-1)\right]
\end{aligned} \tag{C.39}$$

En se basant sur ces hypothèses et en considérant $\mu(\mathrm{k}) = 1$, l'algorithme récursi RPEM (Recursive Prediction Error Method) peut s'exprimer par le système d'équations suivant [68] :

$$\begin{aligned}
\varepsilon(k) &= y(k) - \hat{y}(k) \\
\hat{\theta}(k) &= \hat{\theta}(k-1) + \gamma(k)\left[R(k)\right]^{-1}\psi(k)\varepsilon(k) \\
R(k) &= R(k-1) + \gamma(k)\left[\psi(k)\psi^{\top}(k) - R(k-1)\right]
\end{aligned} \tag{C.40}$$

Comme dans le cas linéaire, afin d'alléger le calcul, l'inversion matricielle est remplacée par une division scalaire grâce à l'utilisation du lemme d'inversion. En effet, la matrice R(N) peut être mise sous la forme suivante :

$$R(N) = \gamma(N)\sum_{i=1}^{N}\beta(N,i)\psi(i,\theta)\varepsilon(i,\theta) = \gamma(N)\bar{R}(N)$$

avec :

$$\bar{R}(N) = \lambda(N)\bar{R}(N-1) + \psi(N)\varepsilon(N)$$

En posant $\mathrm{P(N)} = \gamma(\mathrm{N})\mathrm{R}^{-1}(\mathrm{N}) = \bar{\mathrm{R}}^{-1}(\mathrm{N})$ et en utilisant le lemme d'inversion donné par l'équation suivante :

$$\left[A + BCD\right]^{-1} = A^{-1} - A^{-1}B\left[DA^{-1}B + C^{-1}\right]^{-1}DA^{-1} \tag{C.41}$$

En prenant $\mathrm{A} = \lambda(\mathrm{N})\bar{\mathrm{R}}(\mathrm{N}-1)$, $\mathrm{B} = \mathrm{D}^{\top} = \psi(\mathrm{N})$ et $\mathrm{C} = 1$ on peut écrire :

$$P(N) = \frac{1}{\lambda(N)}\left[P(N-1) - \frac{P(N-1)\psi(N)\psi^{\top}(N)P(N-1)}{\lambda(N) + \psi^{\top}(N)P(N-1)\psi(N)}\right]$$

Dans le cas d'un facteur d'oubli exponentiel $\beta(\mathrm{N},\mathrm{i}) = \lambda^{\mathrm{N-i}}$ ($\gamma = 1-\lambda$), l'algorithme global est donné par le système d'équation suivant :

$$\begin{aligned}
\varepsilon(k) &= y(k) - \hat{y}(k) \\
\psi(k) &= \frac{1}{C(z,\hat{\theta}_{k-1})}\varphi(k) \\
P(k) &= \frac{1}{\lambda}\left[P(k-1) - \frac{P(k-1)\psi(k)\psi^{\top}(k)P(k-1)}{\lambda + \psi^{\top}(k)P(k-1)\psi(k)}\right] \\
K(k) &= P(k)\psi(k) = \frac{P(k-1)\psi(k)}{\lambda + \psi^{\top}(k)P(k-1)\psi(k)} \\
\hat{\theta}(k) &= \hat{\theta}(k-1) + K(k)\varepsilon(k)
\end{aligned} \tag{C.42}$$

λ : le facteur d'oubli.
$\psi(k)$: l'estimée du gradient.
$K(k)$: le gain de *Kalman*.
Dans le cas d'un modèle linéaire en paramètres *AR*, C(z) = 1 et par conséquent $\psi(\mathrm{k}) = \varphi(\mathrm{k})$. Dans ce cas, l'algorithme *RPEM* est le même que celui des moindres carrés récursif *MCR*. On se contente de choisir comme valeurs initiales $\hat{\theta}(0) = [0\ldots0]^{\top}$ et P(0) symétrique et très grande (par exemple 10^6I ou 10^{12}I), ce qui revient à dire qu'on n'accorde aucune confiance à l'estimée initiale des paramètres $\hat{\theta}(0)$ [70].

(b) **Algorithme de *gradient***
Dans ce cas R(k) = I et l'algorithme récursif se résume par les deux équations suivantes :

$$\begin{aligned}
\varepsilon(k) &= y(k) - \hat{y}(k) \\
\hat{\theta}(k) &= \hat{\theta}(k-1) + \gamma(k)\psi(k)\varepsilon(k)
\end{aligned} \tag{C.43}$$

Le choix entre les deux directions est un compromis entre la vitesse de convergence et puissance de calcul. L'algorithme exige moins d'opérations de calcul mais il possède un temps de convergence très grand vers les vraies valeurs des paramètres.

2. **Filtre de *Kalman***
Une autre approche pour l'identification récursive se base sur l'utilisation du filtre de *Kalman*. L'idée de base est de supposer que le vecteur de paramètres $\hat{\theta}$(k) à l'instant k est égal à celui de l'instant k − 1 plus une petite erreur aléatoire :

$$\hat{\theta}(k) = \hat{\theta}(k-1) + w(k) \tag{C.44}$$

w(k) étant le bruit d'innovation ou bien le bruit sur le modèle
La sortie du système y(k) peut s'exprimer par :

$$y(k) = \varphi^{\top}(k)\theta(k) + v(k) \tag{C.45}$$

v(k) étant le bruit de mesure
Ces deux équations (C.44) et (C.45) définissent le filtre de *Kalman* dont les équations générales sont données par :

$$\begin{cases} x(k+1) = F(k)x(k) + w(k) \\ y(k) = H(k)x(k) + v(k) \end{cases}$$

où v(k) et w(k) sont deux bruits blancs gaussien décorrélés entre eux. On note les matrices de covariance suivantes :

$$\begin{cases} E\left[v(k)v^\top(k)\right] = R_2(k) \\ E\left[w(k)w^\top(k)\right] = R_1(k) \\ E\left[v(k)w^\top(k)\right] = 0 \end{cases}$$

R_2(k) est la variance du bruit de mesure, c'est donc un scalaire.
En se basant sur les équations de base du filtre de *Kalman* [72] qui séparent le filtrage en une phase de prédiction et une phase de correction, l'algorithme récursif du filtre de *Kalman* est donné par le système d'équations suivant ([68], éq 11.67 page 380).

$$\begin{aligned} \varepsilon(k) &= y(k) - \hat{y}(k) \\ P(k) &= P(k-1) - \frac{P(k-1)\Psi(k)\Psi^\top(k)P(k-1)}{R_2(k) + \Psi^\top(k)P(k-1)\Psi(k)} + R_1(k) \\ K(k) &= \frac{P(k-1)\Psi(k)}{R_2(k) + \Psi^\top(k)P(k-1)\Psi(k)} \\ \hat{\theta}(k) &= \hat{\theta}(k-1) + K(k)\varepsilon(k) \end{aligned} \tag{C.46}$$

La matrice R_1 a un rôle similaire au facteur d'oubli λ dans l'algorithme RPEM. Cependant, La matrice de covariance R_1 offre la possibilité de choisir la variance sur chacun des paramètres du modèle. Dans le cas où les paramètres n'évoluent pas dans le temps, le choix de $\lambda = 1$ dans l'algorithme RPEM correspond à une matrice de covariance R_1 proche de 0 (par exemple 10^{-6}I) et $R_2 = 1$.

3. *RPLR* (***Recursive PseudoLinear Regression***)
Le prédicteur du modèle ARMAX donné par :

$$\hat{y}(k) = \varphi^\top(k,\theta) \times \theta$$

n'est pas linéaire par rapport aux paramètres θ. On parle d'un régresseur pseudo-linéaire. Pour rendre le calcul des paramètres $\hat{\theta}$ récursif, on applique l'algorithme *MCR* dans lequel on utilise le régresseur pseudo-linéaire φ(k). Cet algorithme peut être vu comme une version simplifiée de l'algorithme *RPEM* ([71], p. 333) dans lequel on impose l'approximation suivante :

$$\psi(k) = \frac{1}{C(z)}\varphi(k) \approx \varphi(k)$$

Cette simplification consiste à négliger l'effet du filtrage par $\frac{1}{C(z)}$. Elle n'entraîne pas une diminution considérable de la puissance de calcul à mettre en œuvre, par contre elle a un effet important sur le comportement de l'algorithme. La convergence vers les vraies valeurs des paramètres θ^* n'est pas garantie. Une condition suffisante de convergence est que $\Re\left(\frac{1}{C(Z,\theta^*)}\right) \geq \frac{1}{2}$ pour tout ω réel [68]. Comme cette condition dépend de la vraie valeur des paramètres, par nature inconnue, on ne peut pas la vérifier a priori. Il est même possible, pour des modèles de type *ARMAX*, que cette méthode ne convergera jamais ([70], p. 112).

C.4 Validation des modèles identifiés

La validation des modèles identifiés à l'aide des méthodes d'identification consiste à prouver que le modèle obtenu décrit bien le comportement du système réel. Plusieurs méthodes ont été

proposées pour valider un modèle en se basant sur la comparaison : de la réponse impulsionnelle, de la réponse fréquentielle ou du diagramme des pôles et des zéros. Un autre moyen pour valider le processus d'identification est basé sur le calcul de la fonction d'autocorrélation de l'erreur de prédiction. Ce critère comporte deux tests :

- **le test de blanchiment** : ce test suppose que :
  - si la structure « modèle + perturbation » choisie est correcte, c'est à dire représentative de la réalité,
  - si on a utilisé une méthode d'identification approprié pour la structure choisie,
  - si les degrés des polynômes A(z), B(z) et C(z) ont été correctement choisi,

  alors l'erreur de prédiction ε(k) tend asymptotiquement vers un bruit blanc, ce qui implique :

$$\lim_{k\to\infty} E\left[\varepsilon(k)\varepsilon(k-i)\right] = 0 \quad \forall i \neq 0$$

  Soit $\{\varepsilon(\mathrm{k})\}$ la séquence centrée des erreurs de prédiction. Le calcul des coefficients d'autocorrélation est donné par :

$$R(i) = \begin{cases} \frac{1}{N}\sum_{k=1}^{N} \varepsilon^2(k) & avec \quad i{=}0 \\ \frac{1}{N}\sum_{k=1}^{N} \varepsilon(k)\varepsilon(k{-}i) & avec \quad i{\neq}0 \end{cases} \tag{C.47}$$

  Les coefficients d'autocorrélation sont souvent représentés par les valeurs normalisées suivantes :

$$R_N(i) = \begin{cases} 1 & avec \quad i{=}0 \\ \frac{R(i)}{R(0)} & avec \quad i{\neq}0 \end{cases}$$

  Si la séquence des erreurs de prédiction est parfaitement blanche (situation théorique) alors on obtient : $\mathrm{R_N(0) = 1; R_N(i) = 0 \quad i \geq 1}$. Dans les situations réelles, ceci n'est jamais le cas ($\mathrm{R_N(i) \neq 0 \quad i \geq 1}$) car ε(k) contient des erreurs dues à la structure du modèle (ordre, non-linéarité) et que l'horizon de mesure n'est jamais infini. Si la séquence d'erreur de prédiction ε(k) est blanche de distribution gaussienne $\mathrm{N(0, R(0))}$, alors les coefficients d'autocorrélation tendent asymptotiquement vers une gaussienne centrée ([68] § 16.6) avec :

$$\sqrt{N}\frac{R(i)}{R(0)} \rightarrow N(0,1)$$

  Comme la distribution des coefficients d'autocorrélation est gaussienne, on peut définir un intervalle de confiance où se trouvent 99% des valeurs de $\mathrm{R_N(i)}$ et dont la valeur est conditionnée par le nombre de points de mesure N. Pour une loi gaussienne, l'intervalle de confiance couvrant 99% ($\pm 3\sigma$, σ est l'écart type) des valeurs prises par $\mathrm{R_N(i)}$ est donné par la formule :

$$|R_N(i)| \leq \frac{2.58}{\sqrt{N}} \quad i \geq 1 \tag{C.48}$$

  Si l'un des coefficients $\mathrm{R_N(i)}$ dépasse cette limite, on peut considérer que l'erreur de prédiction n'est pas blanche et par la suite le modèle n'est pas valide.

– **le test de d'indépendance** : un bon modèle a une erreur de prédiction décorrélé du signal d'entrée u(k). Ceci signifie que l'erreur de prédiction ne contient aucune information utile. Si ce n'est pas le cas, un coefficient d'intercorrélation $R_{u\varepsilon}(k)$ dépassant l'intervalle de confiance signifie que le signal en sortie s(t) généré par le signal d'entrée u(t − k) n'est pas proprement décrit par le modèle.

Pour des modèles non-linéaires, la validation nécessite le test de blanchiment et le test d'indépendance. Pour un modèle ARMA où le signal est nul en entrée on se contente du test de blanchiment pour valider le modèle.

C.5 Identification de la NTF du modulateur ΣΔ à temps continu

La linéarisation du modulateur ΣΔ suivant les conditions de *Bennett* [28] (voir figure C.4) permet d'exprimer le signal de sortie du modulateur dans le domaine en z par :

$$S(z) = STF(z)X(z) + NTF(z)Q(z) \quad \text{avec} \quad \begin{cases} STF(z) = \frac{S(z)}{X(z)} = \frac{F(z)}{1+F(z)} \\ NTF(z) = \frac{S(z)}{Q(z)} = \frac{1}{1+F(z)} \end{cases} \quad \text{(C.49)}$$

avec q(n) le bruit de quantification (bruit blanc).

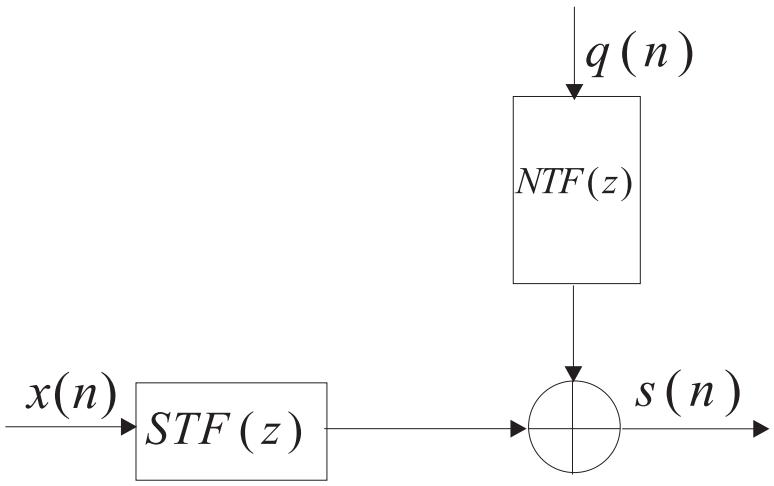

Fig. C.4 – Modèle linéaire du modulateur ΣΔ.

En se basant sur ce modèle linéaire du modulateur ΣΔ, le modèle mathématique qui correspond à l'identification du modulateur est le modèle *ARMAX*(Auto Regressive Moving Average with eXternal inputs) dont la représentation générale est donnée par la figure C.5.

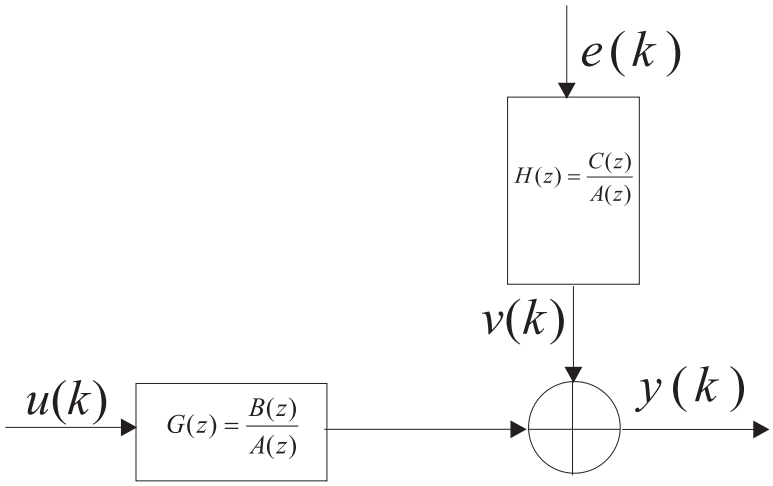

Fig. C.5 – Modèle de structure ARMAX, u(k) le signal d'entrée du modèle, e(k) bruit blanc gaussien $N(0, \sigma^2)$.

Notre but est de déterminer les trois zéros de la $NTF^k(z)$. De ce fait, il est plus judicieux de mettre à zéro l'entrée x(n) du modulateur. Dans ce cas, la sortie s(n) n'est que le bruit de quantification q(n) filtré par NTF(z) (figure C.4). D'une part, cette hypothèse permet de réduire le modèle mathématique du modulateur à un modèle *ARMA* (sans entrée exogène u(k)) et par conséquent de diminuer la quantité de calculs nécessaires pour l'identification. D'autre part, elle facilite l'implantation en reliant l'entrée du modulateur à la masse au lieu de générer un signal spécial pour l'identification.

C.5.1 Résultat de simulation avec les algorithmes *Off Line*

Le modulateur ΣΔ est un système à boucle fermée. Le modèle linéaire représenté à la figure C.4 n'est qu'une approximation sachant que les conditions de *Bennett* sont respectées. Pour cela, nous allons tester l'efficacité de chaque algorithme de calcul sur une fonction de transfert en boucle ouverte avant de la tester avec le vrai signal en sortie du modulateur ΣΔ.

1. **Identification d'une fonction de transfert** NTF(z)
   En exemple d'illustration, nous considérons la fonction de transfert par rapport au bruit $NTF^5(z)$ du cinquième modulateur. En effet, le modulateur ΣΔ génère des harmoniques dans le spectre du signal en sortie si son entrée est un signal continu ou un signal sinusoïdal de fréquence rationnelle ou de très faible amplitude voire nulle. Ce phénomène s'appelle phénomène de cycle limite [73, 74, 75, 76] et se manifeste également par un bruit de quantification coloré. Nous avons constaté également que ce problème de bruit de quantification coloré est très visible pour les modulateurs du centre de l'architecture FBD (dans notre cas le cinquième) dont les fréquences centrales des résonateurs sont proches de la fréquence normalisée 1/4. Ceci est dû au fait que le rapport entre la fréquence d'échantillonnage et la fréquence centrale des résonateurs est très proche d'un entier [77]. Ainsi le choix de cet exemple d'illustration nous permet de considérer un pire-cas. Nous imposons à l'entrée de cette fonction de transfert un bruit blanc gaussien $N(0, 1)$ (figure C.6) du fait de sa densité spectrale constante facilitant ainsi l'identification.

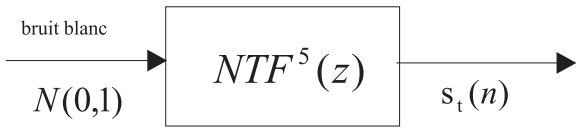

Fig. C.6 – Exemple d'illustration avec NTF^5(z).

La fonction de transfert NTF^5(z) d'un modulateur d'ordre 6 avec un facteur de qualité infini est donnée par :

$$\mathrm{NTF}^5(z) = \frac{1 - 0.1376z^{-1} + 3.003z^{-2} - 0.2751z^{-3} + 3.003z^{-4} - 0.1376z^{-5} + z^{-6}}{1 - 0.1032z^{-1} + 1.501z^{-2} - 0.1032z^{-3} + 0.7507z^{-4} - 0.0258z^{-5} + 0.125z^{-6}} \tag{C.50}$$

Le choix d'un facteur de qualité infini permet de mettre en évidence les trois zéros de la NTF. Le diagramme des pôles et zéros de la NTF^5(z) est représenté par la figure C.7.

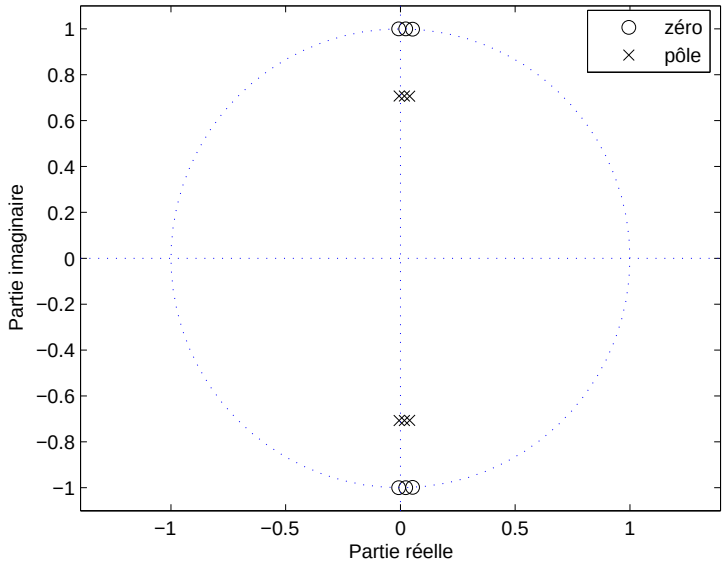

Fig. C.7 – Pôles et zéros de la fonction NTF^5(z).

L'identification de la fonction NTF^5(z) est établi en utilisant un modèle ARMA d'ordre ($\mathrm{n_a} = 6$, $\mathrm{n_c} = 6$) (u(k) est nul, il suffit d'identifier $\mathrm{H(z)} = \frac{\mathrm{C(z)}}{\mathrm{A(z)}}$ figure C.5). Le calcul des paramètres du modèle s'effectue à l'aide de la fonction armax de *Matlab*. La fonction armax utilise en premier lieu l'algorithme de *Gauss-Newton*. Si la convergence vers le minimum de la fonction coût n'est pas assurée, elle utilise l'algorithme de *Levenberg-Marquardt*.

En appliquant la fonction armax aux signaux de sortie de l'exemple d'illustration $\mathrm{s_t(n)}$, nous pouvons constater que la convergence vers le minimum global de la fonction coût $\mathrm{V_N}$ est atteinte avec un modèle ARMA($\mathrm{n_a} = 6, \mathrm{n_c} = 6$) au bout de 25 itérations au maximum et ceci à partir d'un nombre de points de mesure N = 15000. Nous avons également constaté que si la convergence n'est pas assurée, il suffit de faire tourner l'algorithme une deuxième fois. Le module de la réponse fréquentielle de la NTF^5(z) et de la fonction du modèle H(z) sont représentées par la figure C.8.

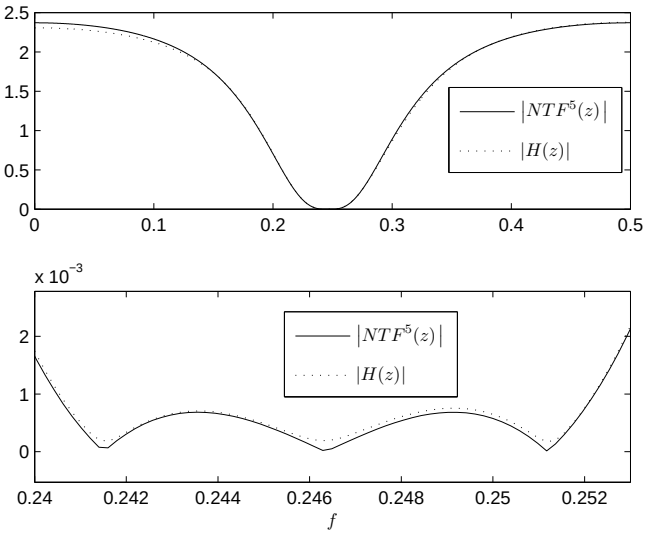

Fig. C.8 – Module de la $NTF^5(z)$ et de son estimée avec un modèle ARMA(6, 6).

Nous constatons une bonne superposition des deux réponses autour des zéros de la NTF. Ceci peut être également constaté par le tracé des pôles et zéros de H(z) qui montre que les zéros coïncident bien avec les zéros de la $NTF^5(z)$ (figure C.9).

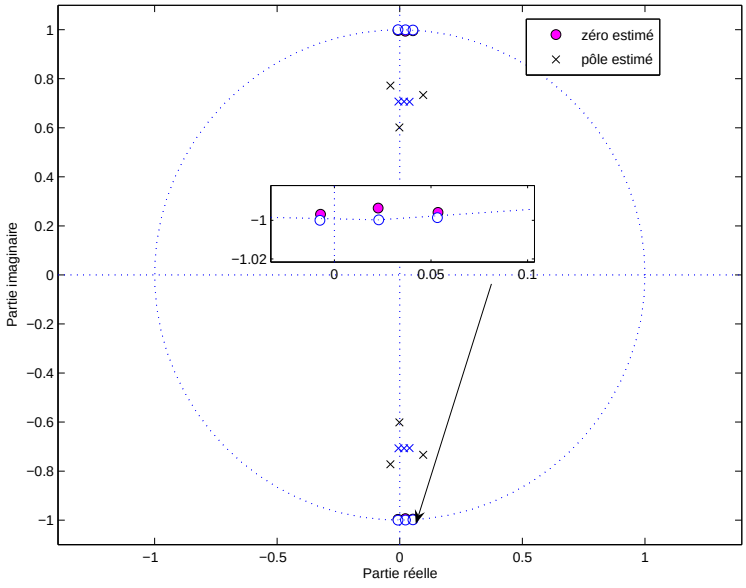

Fig. C.9 – Pôles et zéros de l'estimé de $NTF^5(z)$.

Un troisième critère de performance est le test de blanchiment de l'erreur de prédiction ε(k) (voir § C.4). La figure C.10 montre les coefficients d'autocorrélation R_N(i) de l'erreur de prédiction, i est le retard.

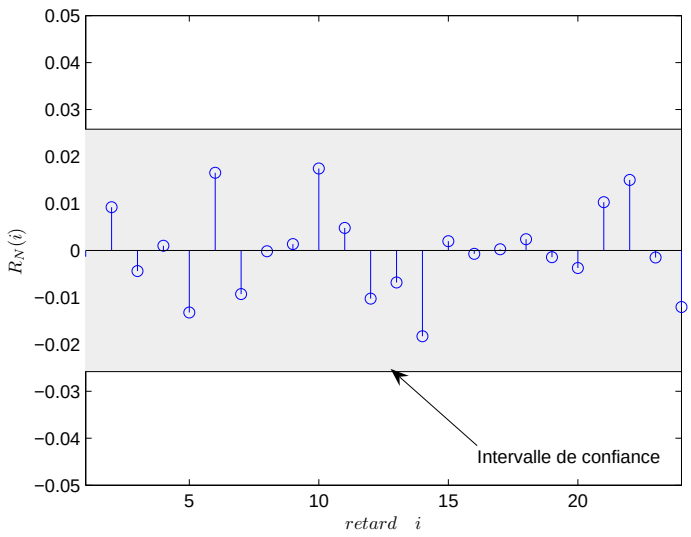

Fig. C.10 – Coefficients d'autocorrélation $R_N(i)$ de l'erreur de prédiction.

Nous constatons que les coefficients d'autocorrélation restent dans l'intervalle de confiance défini par $|R_N(i)| \leq \frac{2.58}{\sqrt{N}} \quad i \geq 1$ (Annexe C.4). Donc, l'erreur de prédiction est blanche et la convergence de la fonction coût vers un minimum global est établi.

2. **Identification de la** NTF(z) **d'un modulateur** ΣΔ
   Après avoir identifier la NTF(z), nous proposons d'appliquer l'algorithme d'identification au vrai signal du modulateur ΣΔ en imposant son entrée à zéro. Nous avons constaté que la méthode d'identification réussit à identifier, avec un modèle ARMA$(6, 6)$ les zéros de la NTF à condition que le bruit de quantification soit blanc (*i.e.* si les conditions de *Bennett* sont satisfaites).

   Une technique pour supprimer les harmoniques est le « *dithering* » [78]. Il consiste à ajouter un bruit blanc à l'entrée du CAN dans la boucle du modulateur pour blanchir le bruit de quantification (figure C.11).

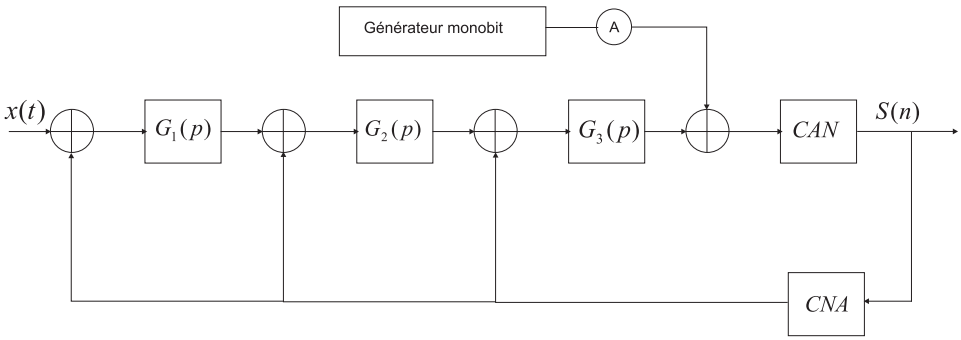

Fig. C.11 – Modulateur ΣΔ d'ordre 6 avec un dither monobit. A est le gain contrôlant la puissance de bruit injectée.

Cette technique permet de blanchir le bruit de quantification mais en même temps ajoute du bruit dans la bande utile du signal. Pour certains modulateurs, si le niveau de bruit

injecté pour blanchir le bruit de quantification est élevé, on peut aboutir à la déformation de la fonction de mise en forme de bruit en supprimant un de ses minimums. Pour contourner ce problème, l'idée est de chercher une source de bruit permettant de blanchir le bruit de quantification sans augmenter le niveau de bruit dans la bande utile du signal. La première idée qui vient à l'esprit est d'utiliser une source de bruit qui a la même forme que la NTF du modulateur idéal. Cette source sera générée par un modulateur ΣΔ monobit ayant la même fonction $NTF^k(z)$ implanté sur un calculateur numérique. La figure C.11 montre la structure du modulateur ΣΔ à temps continu avec le générateur de bruit monobit (« dither »). Le générateur de bruit monobit est un modulateur ΣΔ d'entrée nulle avec un filtre de boucle numérique F(z) d'ordre 6. Il est présenté sur la figure C.12. Il est constitué d'un filtre de boucle F(z) séparé en deux filtres $\frac{1}{F_{den}(z)}$ et $F_{num}(z)$ pour faciliter l'implémentation et augmenter la vitesse en récupérant une sortie à chaque valeur de retour de boucle.

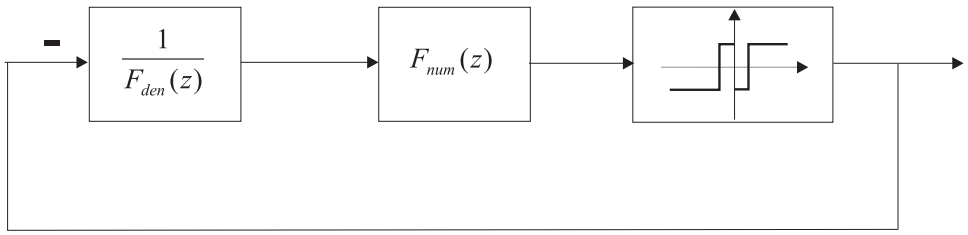

Fig. C.12 – Générateur monobit avec *bit flipping*.

Comme le signal d'entrée de ce générateur est nul, le modulateur ΣΔ entre dans des cycles limites. La solution à ce problème, dans le cas d'un modulateur monobit, consiste à casser la périodicité du signal en sortie et supprimer par la suite les harmoniques en changeant le fonctionnement du comparateur autour de zéro comme le montre la figure C.13. Cette technique est nommée la technique de *Bit Flipping* [79, 80, 81].

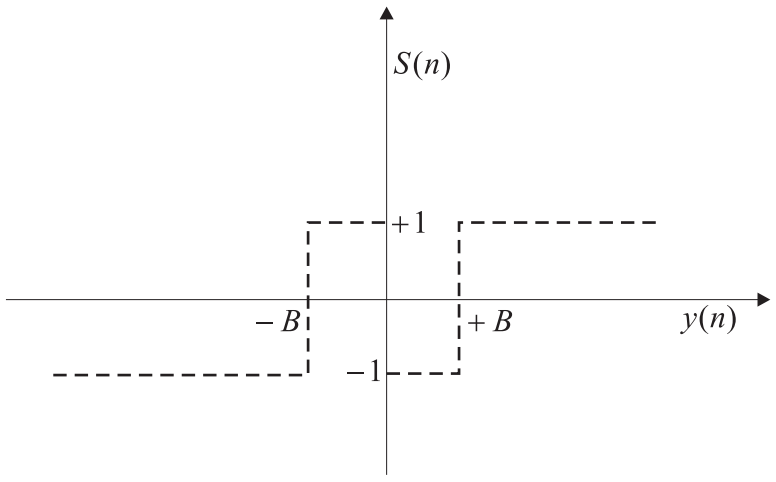

Fig. C.13 – *CAN* monobit avec *bit flipping* . B : est le seuil de comparaison.

Nous avons constaté, à partir des simulations, que les valeurs $B = 1/4$ et $A = 1/4$ présentent la meilleure performance au niveau de blanchiment du bruit de quantification. La figure C.14 présente la densité spectrale de bruit injectée à l'entrée du CAN dans la boucle de modulateur.

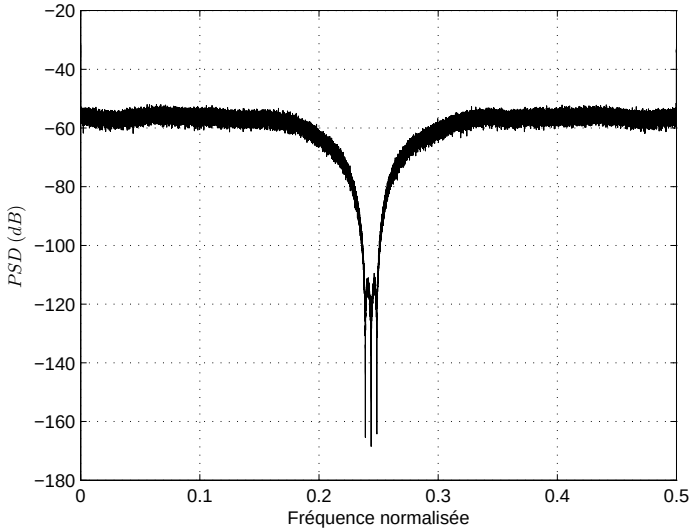

Fig. C.14 – Densité spectrale de puissance (PSD) du bruit injecté pour le cinquième modulateur.

La figure C.15 illustre la densité spectrale de puissance (PSD) du bruit de quantification avant et après l'ajout du bruit monobit « *dither* » pour deux facteurs de qualités ($Q = \infty$ et $Q = 50$).

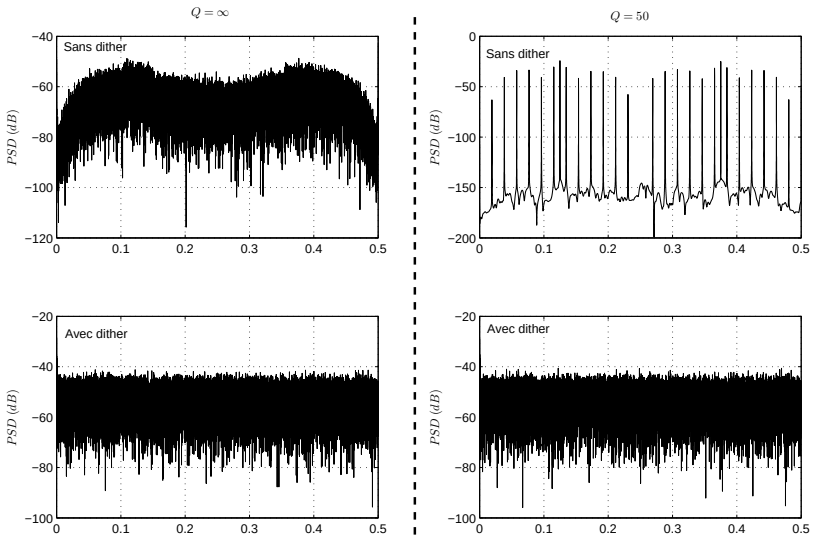

Fig. C.15 – Densité spectrale de puissance (PSD) avant et après « *dithering* »pour $Q = \infty$ et $Q = 50$.

Nous constatons que plus le facteur de qualité est faible, plus le phénomène de cycle limite est visible. Avec le *dithering* avec un générateur de bruit monobit de densité spectrale de la même forme que la NTF(z) idéale de chaque modulateur, le bruit de quantification est blanc et par suite le modèle linéaire de *Bennett* est valable.

Après le blanchiment du bruit de quantification par l'ajout du dither approprié (de densité

spectrale en forme de NTF), la méthode d'identification « *Off Line* » a été testée avec le signal en sortie d'un modulateur ΣΔ dont les fréquences centrales sont proches de $^1/_4$ et ceci pour deux facteurs de qualités ($Q = \infty$ et $Q = 50$). La figure C.16 présente le tracé des pôles et des zéros du modèle estimé.

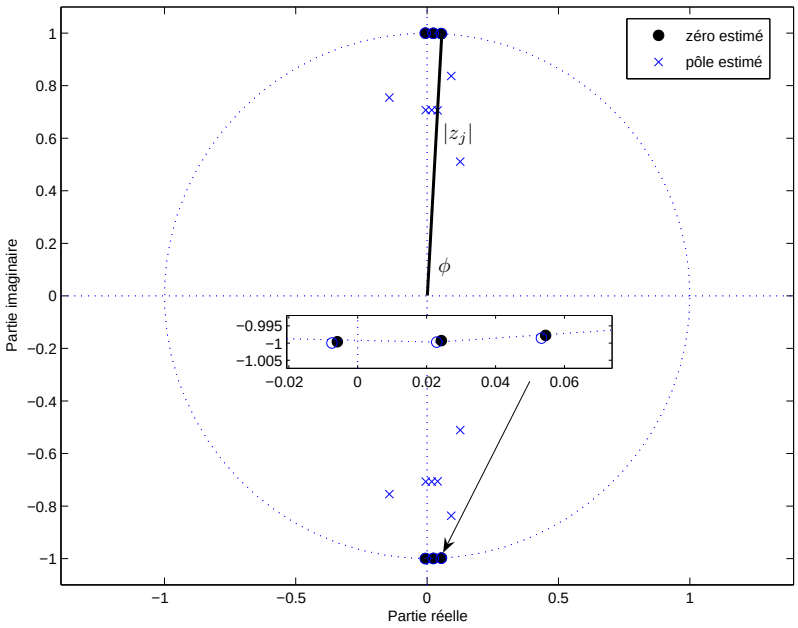

Fig. C.16 – Les pôles et les zéros estimés du cinquième modulateur avec $Q = \infty$.

On note que même si les pôles du modèle trouvé ne correspondent pas aux pôles de la NTF, les zéros du modèle sont très proches de ceux de la fonction $NTF^5(z)$. La figure C.17 montre les coefficients d'autocorrélation de l'erreur de prédiction.

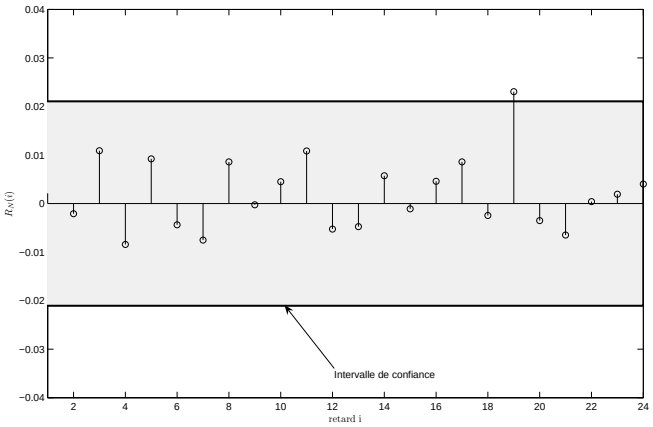

Fig. C.17 – Fonction d'autocorrélation de l'erreur de prédiction avec un modulateur ΣΔ de $Q = \infty$.

Ces coefficients se situent à l'intérieur de l'intervalle de confiance pour $i \neq 0$ et par conséquent l'erreur de prédiction est blanche.

Après la détermination des zéros z_j de la NTF, le calcul des fréquences centrales des résonateurs se fait simplement à partir de l'argument de chaque zéro par $f_{cr_j} = \frac{\arg(z_j)}{2\pi}$. Avec un facteur de qualité infini, les fréquences centrales sont obtenues avec une précision de l'ordre de 10^{-4}. Cette précision convient à notre architecture EFBD qui exige une précision de 10^{-3}. Un autre facteur, mesurable par l'identification des zéros, est le facteur de qualité de chaque résonateur. En effet, en se basant sur l'expression de la fonction de transfert de chaque résonateur donnée par (3.5) et sur le fait que les zéros de la NTF sont des valeurs complexes conjuguées, on peut écrire :

$$z_{j,1} \times z_{j,2} = |z_j|^2 = 1 - \frac{2\pi f_{cr_j}}{Q_j}$$

Alors le facteur de qualité Q_j de chaque résonateur s'exprime par :

$$Q_j = \frac{2\pi f_{cr_j}}{1 - |z_j|^2} \tag{C.51}$$

Nous avons également testé cette méthode d'identification avec des modulateurs réels dont les résonateurs ont un facteur de qualité de 50. Le tracé des pôles et zéros du modèle obtenu (figure C.18) montre que les zéros du modèle obtenu sont très proches de ceux de la NTF du modulateur à estimer.

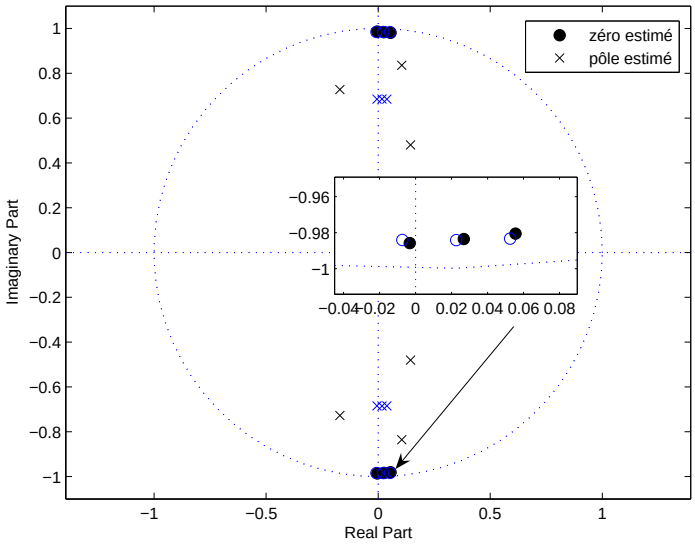

Fig. C.18 – Pôles et zéros estimés du cinquième modulateur avec $Q = 50$.

L'erreur de prédiction vérifie le test de blanchiment (figure C.19) et par conséquent la validité du modèle obtenu.

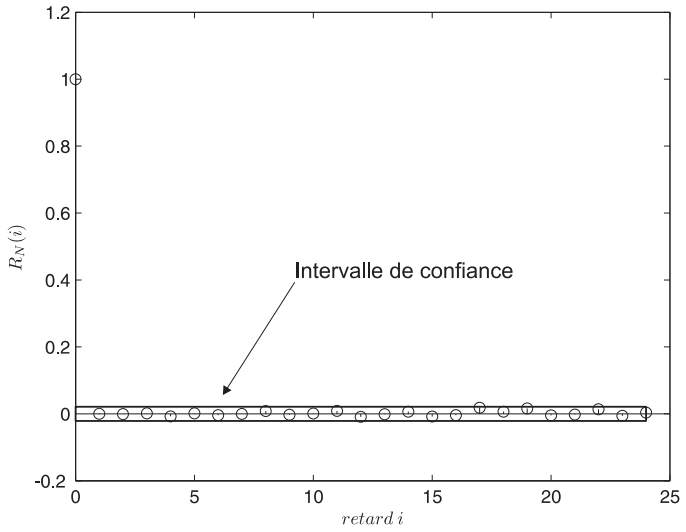

Fig. C.19 – Fonction d'autocorrélation de l'erreur de prédiction avec un modulateur ΣΔ de Q = 50.

La précision sur les fréquences centrales, avec des résonateurs réels, est plus faible que celle obtenue avec des résonateurs idéaux mais elle reste en dessous de 10^{-3}. Les facteurs de qualité sont estimés avec une précision relative entre -0.09% et $+0.01\%$.

C.5.2 Résultat de simulation avec les algorithmes *On Line*

Comme dans le cas « *Off Line* », nous avons considéré l'exemple de la figure C.6 pour tester l'aptitude des algorithmes « *On Line* » à fonctionner avec un cas simple (fonction de transfert + un bruit blanc gaussien en entrée) avant de tester leurs performances avec le signal en sortie du modulateur ΣΔ. En règle générale si la méthode fonctionne avec l'exemple typique, il suffit d'assurer que le bruit de quantification du modulateur est blanc, en injectant un *dither* convenable, pour valider l'algorithme avec le vrai signal en sortie du modulateur.

Identification d'une NTF(z)

Nous avons appliqué les algorithmes d'identification à l'exemple d'illustration (figure C.6) afin de valider leur fonctionnement dans un cas simple. Nous avons considéré, comme critère de validation, la convergence de l'algorithme vers les vraies valeurs des paramètres. La vitesse de convergence et la précision sur les paramètres viennent en deuxième lieu. Nous avons testé ces algorithmes en considérant des fonctions de transfert NTF d'ordre de plus en plus élevé (modulateur ΣΔ passe-bande d'ordre 1, 2 et 3). Le but de l'augmentation de l'ordre de la NTF est de tester l'aptitude de convergence des algorithmes avec l'augmentation du nombre de paramètres à estimer. Les résonateurs utilisés ont un facteur de qualité Q égal à 50 et un décalage en fréquences centrales de $+20\%$. Dans la suite, nous présentons les résultats obtenus avec les fonctions de transfert de différents ordres.

– NTF **d'ordre 1**
La fonction de transfert de la NTF dans le cas d'un modulateur d'ordre 1 avec un facteur de qualité Q = 50 est donnée par l'équation suivante :

$$NTF(z) = \frac{1 - 0.04516z^{-1} + 0.969z^{-2}}{1 - 0.03369z^{-1} + 0.469z^{-2}} \tag{C.52}$$

L'identification de cette fonction de transfert s'effectue avec un modèle ARMA$(2, 2)$ en utilisant les deux algorithmes suivants :

1. algorithme RPEM avec $\lambda = 1$: Ceci revient à dire qu'on n'a pas considéré un facteur d'oubli dans l'algorithme. En effet les paramètres de la NTF ne changent pas de valeur pendant la période de l'exécution de l'algorithme d'identification. Nous constatons que la convergence vers les vraies valeurs des paramètres (représenté en trait pointillé sur la figure C.20) est assurée au bout de 3 000 échantillons.

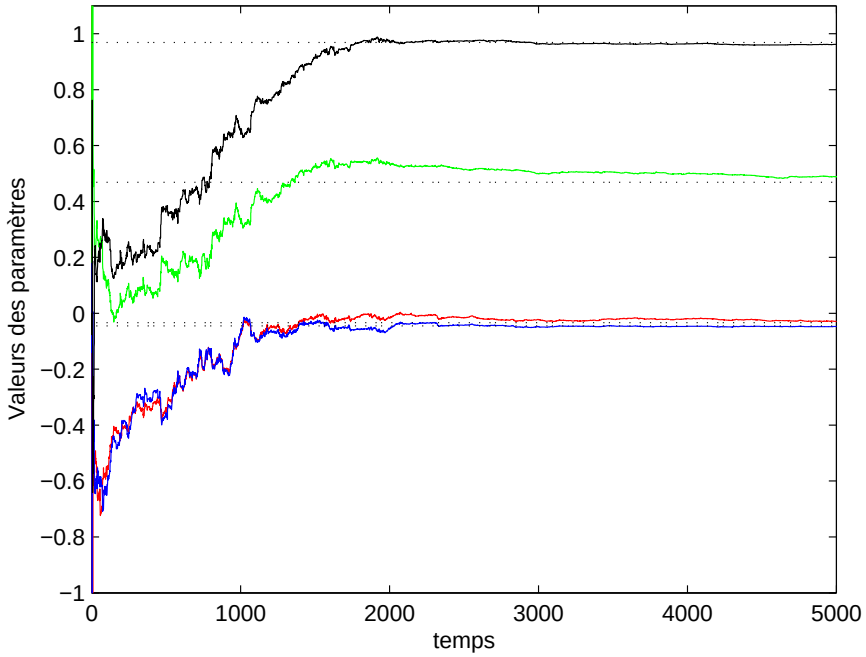

Fig. C.20 – Paramètres estimés pour un modulateur ΣΔ d'ordre 1 avec facteur d'oubli $\lambda = 1$.

2. filtre de *Kalman* avec une matrice de covariance $R_1 = 10^{-8}I$ et $R_2 = 1$: l'évolution des paramètres estimés en fonction du temps est représentée sur la figure C.21.

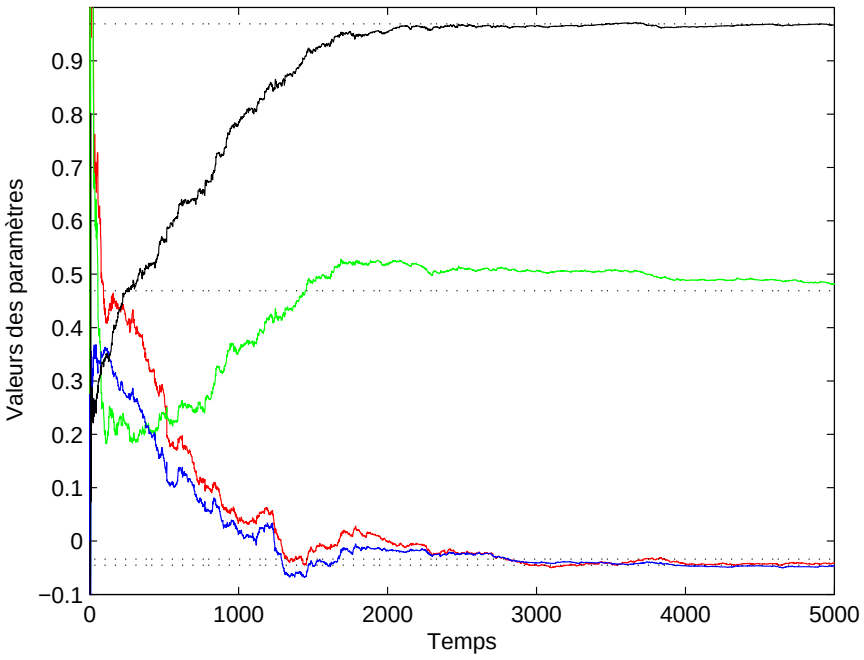

Fig. C.21 – Paramètres estimés pour un modulateur ΣΔ d'ordre 1 avec le filtre de *Kalman* $R_1 = 10^{-8}I$ et $R_2 = 1$.

On note que le filtre de *Kalman* assure une convergence, comme dans le cas RPEM avec $\lambda = 1$, à partir de 4 000 échantillons.

– NTF **d'ordre 2**
La fonction de transfert de la NTF dans le cas d'un modulateur d'ordre 1 est donnée par l'équation suivante :

$$NTF(z) = \frac{1 - 0.0903z^{-1} + 1.939z^{-2} - 0.08749z^{-3} + 0.939z^{-4}}{1 - 0.06738z^{-1} + 0.9386z^{-2} - 0.03159z^{-3} + 0.22z^{-4}} \tag{C.53}$$

L'identification de cette fonction de transfert s'effectue avec un modèle ARMA$(4, 4)$ en utilisant les deux algorithmes suivants :

1. RPEM avec $\lambda = 1$: Cet algorithme converge à partir de 30 000 échantillons (figure C.22). On note que plus on augmente le nombre de paramètres à estimer, plus le temps de convergence est grand.

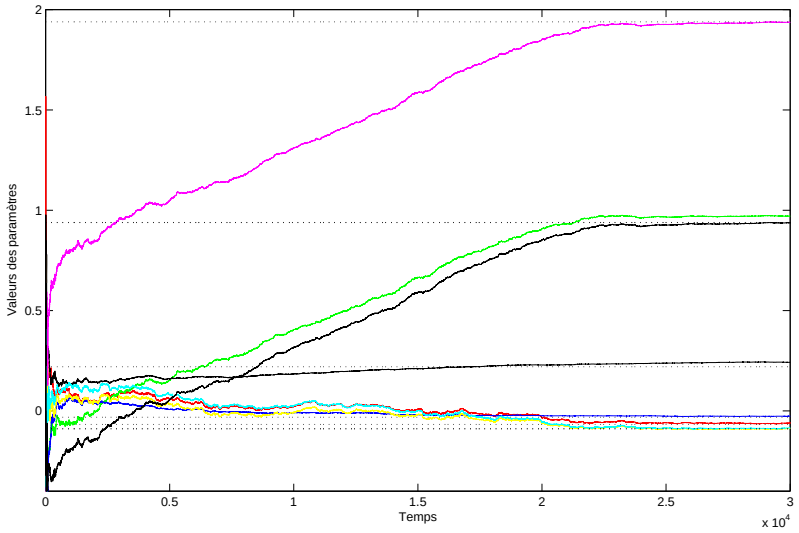

Fig. C.22 – Paramètres estimés pour un modulateur ΣΔ d'ordre 2 avec facteur d'oubli $\lambda = 1$.

2. filtre de *Kalman* avec une matrice de covariance $R_1 = 10^{-8}I$ et $R_2 = 1$: cet algorithme a le même comportement que celui de RPEM avec $\lambda = 1$ au niveau de la vitesse de convergence (30 000 échantillons) même si la précision sur les paramètres est plus faible.

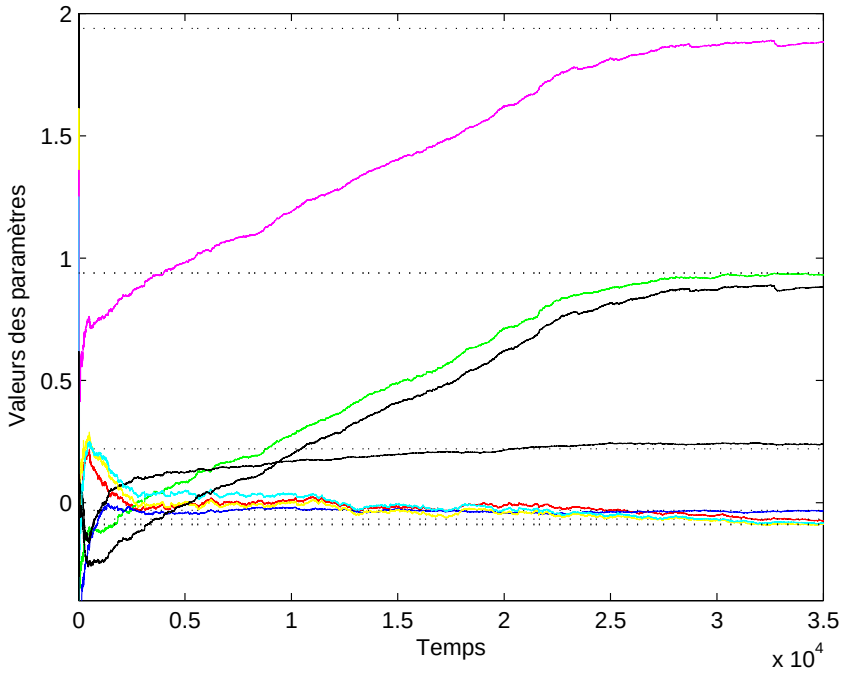

Fig. C.23 – Paramètres estimés pour un modulateur ΣΔ d'ordre 2 avec le filtre de *Kalman* $R_1 = 10^{-8}I$ et $R_2 = 1$.

– NTF **d'ordre 3** : la fonction de transfert de la NTF dans le cas d'un modulateur d'ordre 6 est donnée par l'équation suivante :

$$NTF(z) = \frac{1 - 0.1355z^{-1} + 2.91z^{-2} - 0.2624z^{-3} + 2.82z^{-4} - 0.1271z^{-5} + 0.91z^{-6}}{1 - 0.1011z^{-1} + 1.409z^{-2} - 0.09473z^{-3} + 0.6607z^{-4} - 0.02221z^{-5} + 0.1032z^{-6}} \tag{C.54}$$

L'identification de cette fonction de transfert s'effectue avec un modèle ARMA$(6, 6)$ en utilisant les deux algorithmes suivants :

– RPEM avec $\lambda = 1$: avec un modèle ARMA$(6, 6)$ la convergence de l'algorithme n'est pas assurée (figure C.24) même avec 250 000 échantillons.

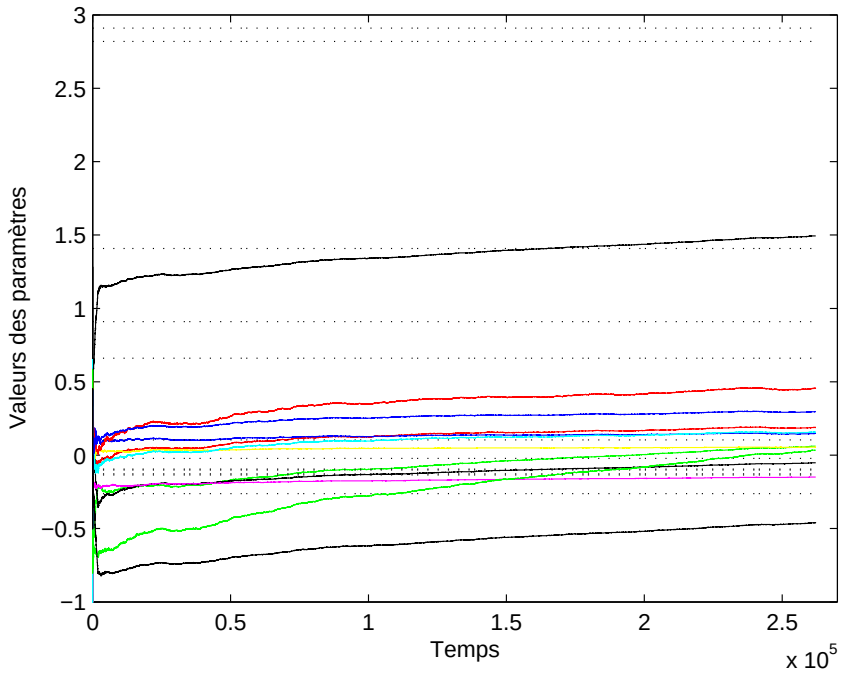

Fig. C.24 – Paramètres estimés pour un modulateur ΣΔ d'ordre 3 avec facteur d'oubli $\lambda = 1$.

– filtre de *Kalman* avec une matrice de covariance $R_1 = 10^{-8}I$ et $R_2 = 1$: le filtre de *Kalman* possède le même type de problème, expliqué ci-dessous, avec le même nombre d'échantillons que l'algorithme RPEM avec facteur d'oubli (figure C.25).

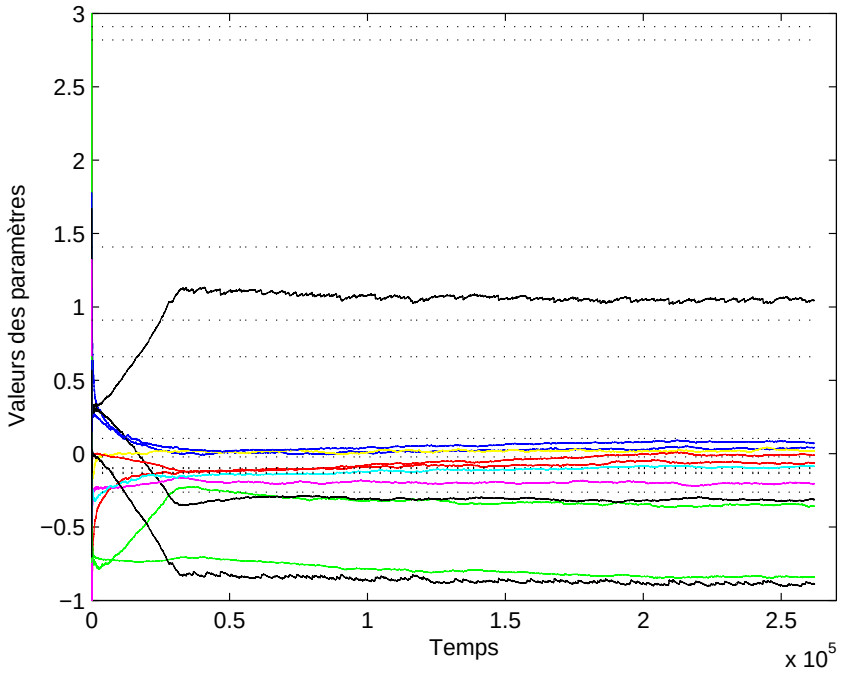

Fig. C.25 – Paramètres estimés pour un modulateur ΣΔ d'ordre 3 avec le filtre de *Kalman* $R_1 = 10^{-8}I$ et $R_2 = 1$.

Ce problème de convergence est dû à la minimisation de la fonction coût V_N pour la recherche des vrais paramètres dans un espace à 12 dimensions (12 paramètres recherchés). En effet, nous pouvons noter d'après les simulations précédentes, que la convergence vers le minimum est de plus en plus difficile lorsque le nombre de paramètres augmente (temps de convergence plus grand). De plus, dans notre cas, les pôles et les zéros de la fonction NTF(z) sont très proches. De ce fait, les vraies valeurs des paramètres recherchés sont très proches. Ce qui augmente la difficulté de l'algorithme à minimiser la fonction coût dans un espace à 12 dimensions dans une zone très petite.

Compte tenu de ces faits, les algorithmes *On Line* ne peuvent pas être appliqués au signal en sortie du modulateur ΣΔ pour identifier sa NTF. En effet, nous ne pouvons pas envisager d'appliquer cette méthode avec des modulateurs d'ordre 6 pour adapter le traitement numérique de l'architecture EFBD.

Références bibliographiques

[1] I. Galton and H. T. Jensen, "Delta-sigma modulator based A/D conversion without oversampling," *IEEE Trans. Circuit and Sys.II*, vol. 42, pp. 773–784, December 1995.

[2] H. S. P. Aziz and J. V. der Spiegel, "Multiband sigma-delta modulation," *Electronics Letters*, pp. 760–762, April 1993.

[3] A. Eshraghi and T. Fiez, "A time-interleaved parallel $\Delta\Sigma$ A/D converter," *IEEE Trans. Circuit and Sys.II*, vol. 50, pp. 118–129, March 2003.

[4] A. Eshraghi and T. Fiez, "A comparison of three parallel $\Delta\Sigma$ A/D converters," *ISCAS*, vol. 1, pp. 517–520, May 1996.

[5] A. Eshraghi and T. Fiez, "A comparative analysis of parallel delta-sigma ADC architectures," *Circuits and Systems I : Regular Papers, IEEE Transactions on [see also Circuits and Systems I : Fundamental Theory and Applications, IEEE Transactions on]*, vol. 51, no. 3, pp. 450–458, 2004.

[6] V. Nguyen, P. Loumeau, and J. Naviner, "Avantages des modulateurs delta-sigma passe haut pour le convertisseur sigma-delta combiné avec l'entrelacement temporel," *Proceeding of TAISA*, vol. 2002, 2002.

[7] R. Batten, A. Eshraghi, and T. Fiez, "Calibration of Parallel $\Sigma\Delta$ ADCs," *IEEE Trans. Circuits Syst. II*, vol. 49, no. 6, pp. 390–399, 2002.

[8] K. Grati, *Architecture d'un récepteur radio multistandard à sélection numérique des canaux*. PhD thesis, ENST - COMELEC Communication et Electronique, Juin 2005.

[9] B. Razavi, *RF microelectronics*. Prentice-Hall, Inc. Upper Saddle River, NJ, USA, 1998.

[10] J. Mitola, "The software radio architecture," *Communications Magazine, IEEE*, vol. 33, no. 5, pp. 26–38, 1995.

[11] J. Mitola and Z. Zvonar, *Software Radio Technologies : Selected Readings*. IEEE Press, 2001.

[12] L. Mitola III, "Technical challenges in the globalization of software radio," *Communications Magazine, IEEE*, vol. 37, no. 2, pp. 84–89, 1999.

[13] W. Tuttlebee, *Software Defined Radio : Origins, Drivers & International Perspectives*. John Wiley and Sons, 2002.

[14] F. Rivet, Y. Deval, J. Begueret, D. Dallet, P. Cathelin, and D. Belot, "A Disruptive Receiver Architecture Dedicated to Software-Defined Radio," *Circuits and Systems II : Express Briefs, IEEE Transactions on [see also Circuits and Systems II : Analog and Digital Signal Processing, IEEE Transactions on]*, vol. 55, no. 4, pp. 344–348, 2008.

[15] P. Gray and R. Meyer, "Future directions in silicon ICs for RF personal communications," *Custom Integrated Circuits Conference, 1995., Proceedings of the IEEE 1995*, pp. 83–90, 1995.

[16] J. Rudell, J. Ou, T. Cho, G. Chien, F. Brianti, J. Weldon, and P. Gray, "A 1.9-GHz wide-band IF double conversion CMOS receiver for cordless telephone applications," *Solid-State Circuits, IEEE Journal of*, vol. 32, no. 12, pp. 2071–2088, 1997.

[17] T. View, "Design of a multiband internal antenna for third generation mobile phone handsets," *Antennas and Propagation, IEEE Transactions on*, vol. 51, no. 7, pp. 1452–1461, 2003.

[18] T. View, "Compact internal multiband antennas for mobile handsets," *Antennas and Wireless Propagation Letters*, vol. 2, 2003.

[19] P. Ciais, R. Staraj, G. Kossiavas, and C. Luxey, "Compact internal multiband antenna for mobile phone and WLAN standards," *Electronics Letters*, vol. 40, no. 15, pp. 920–921, 2004.

[20] J. Rode, A. Swaminathan, I. Galton, and P. Asbeck, "Fractional-N Direct Digital Frequency Synthesis with a 1-Bit Output," *Microwave Symposium Digest, 2006. IEEE MTT-S International*, pp. 415–418, 2006.

[21] J. Rogers, F. Dai, M. Cavin, and D. Rahn, "A Multiband $\Delta\Sigma$ Fractional N Frequency Synthesizer for a MIMO WLAN Transceiver RFIC," *Solid-State Circuits, IEEE Journal of*, vol. 40, no. 3, pp. 678–689, 2005.

[22] S. P. V. T. Chalvatzis, "A low noise 40-GS/s Continuous-Time Bandpass Delta-Sigma ADC Centred at 2 GHz," *IEEE Symposium on Radio Frequency Integrated Circuits*, june 2006.

[23] S. Norsworthy, R. Schreier, and G. Temes, *Delta-sigma data converters : theory, design, and simulation*. IEEE Press, 1996.

[24] E. Avignon, *Contribution à la conception d'un modulateur sigma-delta passe-bande à temps continu pour la conversion directe de signaux radiofréquences*. PhD thesis, Dept.SSE, SUPELEC, Université Paris VI, Décembre 2007.

[25] S. Benabid, *Architecture et conception d'un modulateur sigma delta passe-bande à filtre LC adaptés à la numérisation des signaux rapides*. PhD thesis, Dept.SSE, SUPELEC, Université Paris XI, 2005.

[26] W. Black Jr and D. Hodges, "Time interleaved converter arrays," *Solid-State Circuits Conference. Digest of Technical Papers. 1980 IEEE International*, vol. 23, 1980.

[27] P. Vaidyanathan, *Multirate systems and filter banks*. Prentice-Hall, Inc. Upper Saddle River, NJ, USA, 1993.

[28] W. Bennett, "Spectra of quantized signals," *Bell syst. tech. J.*, vol. 27, pp. 446–472, 1948.

[29] V. Nguyen, P. Loumeau, and J. Naviner, "Analysis of time-interleaved delta-sigma analog to digital converter," *IEEE Vehicular Technology Conference*, vol. 4, pp. 1594–1597, 2002.

[30] R. Khoini-Poorfard and D. Johns, "Mismatch effects in time-interleaved oversampling converters," *Circuits and Systems, 1994. ISCAS'94., 1994 IEEE International Symposium on*, vol. 5.

[31] A. Petraglia and S. Mitra, "Analysis of mismatch effects among A/D converters in a time-interleaved waveform digitizer," *Instrumentation and Measurement, IEEE Transactions on*, vol. 40, no. 5, pp. 831–835, 1991.

[32] A. Eshraghi, *High-Speed Parallel Delta-Sigma Analog-To-Digital Converters*. PhD thesis, Washinton State University, May 1999.

[33] M. Sarhang-Nejad and G. Temes, "A high-resolution multibit $\Sigma\Delta$ ADC with digital correction and relaxed amplifier requirements," *Solid-State Circuits, IEEE Journal of*, vol. 28, no. 6, pp. 648–660, 1993.

[34] A. Eshraghi and T. Fiez, "An area efficient time-interleaved parallel delta-sigma A/D converter," *Circuits and Systems, 1998. ISCAS'98. Proceedings of the 1998 IEEE International Symposium on*, vol. 2.

[35] Y. Chang, C. Wu, and T. Yu, "Chopper-stabilized sigma-delta modulator," *Circuits and Systems, 1993., ISCAS'93, 1993 IEEE International Symposium on*, pp. 1286–1289, 1993.

[36] E. King, A. Eshraghi, I. Galton, and T. Fiez, "A Nyquist-rate delta-sigma A/D converter," *Solid-State Circuits, IEEE Journal of*, vol. 33, no. 1, pp. 45–52, 1998.

[37] H. S. P. Aziz and J. V. der Spiegel, "Multiband sigma-delta analog to digital conversion," *ICASSP*, vol. 3, pp. 249–252, April 1994.

[38] R. Cormier Jr, T. Sculley, and R. Bamberger, "Combining subband decomposition and sigma delta modulation for wideband A/D conversion," *Circuits and Systems, 1994. ISCAS'94, 1994 IEEE International Symposium on*, vol. 5.

[39] M. K. P. Benabes and R. Kielbasa, "A methodology for designing continuous-time sigma-delta modulators," *Proceedings of European design and test conference*, pp. 46–50, March 1997.

[40] H. Aboushady, *Conception en vue de la réutilisation de convertisseur analogique-numerique* $\Delta\Sigma$ *temps-continu mode courant*. PhD thesis, Dept. ELECTRONICS, COMMUNICATIONS AND COMPUTER SCIENCE, UNIVERSITY OF PARIS VI, 2002.

[41] O. Shoaei, *Continuous-Time Delta-Sigma A/D Converters for High Speed Applications*. PhD thesis, Carleton University, November 1995.

[42] A. G. P. Benabes and R. Kielbasa, "A multistage closed-loop sigma-delta modulator (MSCL)," *Journal of Analog Integrated Circuits and Signal Processing*, vol. 11, pp. 195–204, novembre 1996.

[43] E. N. Aghdam, *Nouvelles techniques d'appariement dynamique dans un CNA multibit pour les convertisseurs* $\Sigma\Delta$. PhD thesis, Dept.SSE, SUPELEC, Université Paris XI, 1994.

[44] P. Benabes, *Etude de nouvelles structures de convertisseurs Sigma-Delta passe-bande*. PhD thesis, Dept.SSE, SUPELEC, Université Paris XI, 1994.

[45] R. Schreier and G. Temes, *Understanding Delta-sigma data converters*. New Jersy : Wiley, 2005.

[46] T. Saramaki and H. Tenhunen, "Efficient VLSI-realizable decimators for sigma-delta analog-to-digital converters," *Circuits and Systems, 1988., IEEE International Symposium on*, pp. 1525–1528, 1988.

[47] S. Chu and C. Burrus, "Multirate filter designs using comb filters," *Circuits and Systems, IEEE Transactions on*, vol. 31, no. 11, pp. 913–924, 1984.

[48] G. Bourdopoulos, *Delta-SIGMA Modulators : Modeling, Design and Applications*. Imperial College Press, 2003.

[49] Y. Dupret, *Modélisation des dispersions des performances des cellules analogiques intégrées en fonction des dispersions du processus de fabrication*. PhD thesis, Dept.SSE, SUPELEC, Université Paris-Sud, 2005.

[50] C. Toumazou and D. Haigh, "Integrated microwave continuous-time active filters using fully tunable GaAs transconductors," *Circuits and Systems, 1991., IEEE International Sympoisum on*, pp. 1765–1768, 1991.

[51] Y. Tsividis, "Integrated continuous-time filter design-an overview," *Solid-State Circuits, IEEE Journal of*, vol. 29, no. 3, pp. 166–176, 1994.

[52] W. Kuhn, F. Stephenson, and A. Elshabini-Riad, "A 200 MHz CMOS Q-enhanced LC bandpass filter," *Solid-State Circuits, IEEE Journal of*, vol. 31, no. 8, pp. 1112–1122, 1996.

[53] W. Gao and W. Snelgrove, "A 950-MHz IF second-order integrated LC bandpass delta-sigma modulator," *Solid-State Circuits, IEEE Journal of*, vol. 33, no. 5, pp. 723–732, 1998.

[54] L. Thourel, *Initiation aux techniques modernes des radars*. CEPADUES, 1982.

[55] S. Salous and P. Green, "A novel digital chirp generator using a dual clock field programmable gate array architecture," *HF Radio Systems and Techniques, 1994., Sixth International Conference on*, pp. 391–395, 1994.

[56] J. Vankka and K. Halonen, *Direct Digital Synthesizers : Theory, Design and Applications*. Kluwer Academic Publishers, 2001.

[57] S. Parkes, "Advanced digital techniques for high bandwidth chirp generation," *Radar 92. International Conference*, pp. 501–505, 1992.

[58] J. Candy, "Decimation for Sigma Delta Modulation," *Communications, IEEE Transactions on [legacy, pre-1988]*, vol. 34, no. 1, pp. 72–76, 1986.

[59] E. Hogenauer, "An economical class of digital filters for decimation and interpolation,"

[60] D. Chan and L. Rabiner, "Analysis of quantization errors in the direct form for finite impulse response digital filters," *Audio and Electroacoustics, IEEE Transactions on*, vol. 21, no. 4, pp. 354–366, 1973.

[61] P. Wong, "Quantization and roundoff noises in fixed-point FIR digital filters," *Signal Processing, IEEE Transactions on [see also Acoustics, Speech, and Signal Processing, IEEE Transactions on]*, vol. 39, no. 7, pp. 1552–1563, 1991.

[62] C. Barnes, B. Tran, and S. Leung, "On the statistics of fixed-point roundoff error,"

[63] T. Hayashi, Y. Inabe, K. Uchimura, and T. Kimura, "A multistage delta-sigma modulator without double integration loop," *Solid-State Circuits Conference. Digest of Technical Papers. 1986 IEEE International*, vol. 29, 1986.

[64] J. Cherry and W. Snelgrove, *Continuous-Time Delta-Sigma Modulators for High-Speed A/D Conversion : Theory, Practice and Fundamental Performance Limits*. Kluwer A.P., 1999.

[65] O. Shoaei and W. Snelgrove, "A multi-feedback design for LC bandpass delta-sigma modulators," *Proceedings Inter Symposium on Circuits and Systems*, vol. 1, pp. 171–174, May 1995.

[66] R. Schreier and M. Snelgrove, "Bandpass sigma-delta modulation," *Electronic letters*, vol. 25, pp. 1560–61, Nov. 89.

[67] A. Yahya, *Architecture et conception d'un modulateur sigma delta passe-bande à filtre LC adaptés à la numérisation des signaux rapides*. PhD thesis, Dept.SSE, SUPELEC, Université Paris XI, Dec. 2002.

[68] L. Ljung, *System Identification, Theory For The User Second Edition*. Prentice Hall PTR, 1999.

[69] L. Ljung, "System Identification Toolbox for use with Matlab," *The Mathworks Inc., Mass., USA*, 1991.

[70] E.Walter and L. Pronzato, *Identification de Modèles Paramétriques à partir de données Expérimentales*. Masson, 1994.

[71] L. Söderström and P. Stoica, *System Identification*. Prentice Hall PTR, 1989.

[72] G. Fleury and J. Oksman, *Analyse spectrale méthodes non-paramétriques et paramétriques Les Cours de l'École supérieure d'électricité*. Ellipses, 2001.

[73] V. Friedman, "The structure of the limit cycles in sigma delta modulation," *Communications, IEEE Transactions on*, vol. 36, no. 8, pp. 972–979, 1988.

[74] J. Iwersen, "Comments on The structure of the limit cycles in sigma-delta modulation'," *Communications, IEEE Transactions on*, vol. 38, no. 8, 1990.

[75] D. Hyun and G. Fischer, "Limit cycles and pattern noise in single-stage single-bit delta-sigma modulators," *Circuits and Systems I : Fundamental Theory and Applications, IEEE Transactions on [see also Circuits and Systems I : Regular Papers, IEEE Transactions on]*, vol. 49, no. 5, pp. 646–656, 2002.

[76] S. Hein and A. Zakhor, "On the stability of sigma delta modulators," *Signal Processing, IEEE Transactions on [see also Acoustics, Speech, and Signal Processing, IEEE Transactions on]*, vol. 41, no. 7, pp. 2322–2348, 1993.

[77] P. Perez-Alcazar and A. Santos, "Relationship between sampling rate and quantization noise," *Digital Signal Processing, 2002. DSP 2002. 2002 14th International Conference on*, vol. 2, 2002.

[78] S. Norsworthy, R. Schreier, and G. Temes, *Delta-sigma data converters, Theory, design and simulation*. NJ : IEEE Press, 97.

[79] A. Magrath and M. Sandler, "Non-linear deterministic dithering of sigma-delta modulators," *Oversampling and Sigma-Delta Strategies for DSP, IEE Colloquium on*, p. 2, 1995.

[80] A. Magrath and M. Sandler, "Performance enhancement of sigma-delta modulator D/A convertersusing non-linear techniques," *Circuits and Systems, 1996. ISCAS'96.,'Connecting the World'., 1996 IEEE International Symposium on*, vol. 2, 1996.

[81] A. Magrath and M. Sandler, "Resolution enhancement and dither of sigma-delta modulator digital-to-analogue converters," *Electronics Letters*, vol. 31, no. 18, pp. 1540–1542, 1995.