HYBRID CONTINUOUS-DISCRETE-TIME MULTI-BIT DELTA-SIGMA A/D CONVERTERS WITH AUTO-RANGING ALGORITHM

THÈSE N^{O} 3912 (2008)

PRÉSENTÉE LE 22 FÉVRIER 2008
À LA FACULTÉ DES SCIENCES ET TECHNIQUES DE L'INGÉNIEUR
Laboratoire d'électronique générale 1
SECTION DE GÉNIE ÉLECTRIQUE ET ÉLECTRONIQUE

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

POUR L'OBTENTION DU GRADE DE DOCTEUR ÈS SCIENCES

PAR

Sergio PESENTI

ingénieur en microtechnique diplômé EPF
et de nationalités italienne et suisse, originaire de Corsier-sur-Vevey (VD)

acceptée sur proposition du jury:

Prof. J. R. Mosig, président du jury
Prof. M. Kayal , directeur de thèse
P. Clement, rapporteur
Prof. W. Redman-White, rapporteur
Prof. J.-Ph. Thiran, rapporteur

EPFL
ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE
Lausanne, EPFL
2007

Je dédie cette thèse à ma tendre
et merveilleuse épouse, Francesca.

Remerciements

Je souhaiterais avant tout remercier **Maher Kayal** et **Patrick Clément** pour m'avoir donné l'opportunité de travailler sur ce projet de thèse et, plus particulièrement, pour m'avoir soutenu et encouragé dans les moments les plus difficiles. Leur confiance m'a permis de donner le meilleur de moi-même dans des conditions parfois défavorables. Je tiens à leur exprimer toute ma reconnaissance quant à leurs efforts de titan - le plus souvent cachés et silencieux - pour faire aboutir ce projet. Leurs compétences techniques ainsi que leurs qualités humaines font de Maher et de Patrick des personnes d'exception.

Je remercie très chaleureusement toutes les personnes avec qui j'ai partagé le bureau 331 du Laboratoire d'Electronique Générale pendant près de cinq longues années:

Danica Stefanovic, avec qui j'ai étroitement collaboré et avec laquelle je continue à travailler à ce jour. Brillante, exigeante et passionnée, elle a su apporter l'optimisme quant tout me semblait converger vers un inéluctable désastre.

Louis Harik, qui a également contribué à la réalisation de ce travail. Il est pour moi un modèle de patience et de modestie.

Nicolas Schlumpf et **Cédric Bassin**, mes premiers collègues de bureau. Ils excellent dans l'électronique autant que dans l'art de l'humour, deux disciplines indispensables dans notre métier.

Jari Curty et **Patrice Beaud**, visiteurs inconditionnels du bureau. Leurs connaissances techniques et philosophiques n'ont cessé de m'impressionner, autant que leur maîtrise de l'outil informatique.

Mes remerciements s'adressent aussi aux incontournables piliers du laboratoire que sont **Joseph Guzzardi**, **Isabelle Buzzi**, **Marie Halm** et **Danielle Gallay**. Je veux également exprimer une immense reconnaissance à tous mes autres amis et collègues. Je m'estime heureux d'avoir croisé sur mon chemin autant de personnes formidables qui, avec leurs paroles d'encouragement et de soutien, m'ont fait oublier les personnes qui sont leur exact opposé.

Pour terminer, j'aimerais exprimer ma plus profonde gratitude à mes parents, **Margherita** et **Daniele Pesenti**, pour leur confiance en moi, sans compromis et sans limites, et à mon épouse **Francesca Catalano** qui, à elle seule, donne un sens à tout ce que j'accomplis.

Abstract

In wireless portable applications, a large part of the signal processing is performed in the digital domain. Digital circuits show many advantages. The power consumption and fabrication costs are low even for high levels of complexity. A well established and highly automated design flow allows one to benefit from the constant progress in CMOS technologies. Moreover, digital circuits offer robust and programmable signal processing means and need no external components. Hence, the trend in consumer electronics is to further reduce the part of analog signal processing in the receiver chain of wireless transceivers. Consequently, analog-to-digital converters with higher resolutions and bandwidths are constantly required. The ultimate goal is the direct digitization of radio frequency signals, where the conversion would be performed immediately after the front -end amplifier.

$\Delta\Sigma$-modulation-based converters have proved to be the most suitable to achieve the required performance. Switched-capacitor implementations have been widely used over the last two decades. However, recent publications and books have shown that continuous-time architectures can achieve the same performance with lower power consumption. Most designs found throughout the literature use a single- or few-bit internal quantizer with a high-order modulation. As a result, in order to achieve the resolutions and bandwidths required today, the sampling frequency must exceed 100MHz. This approach leads to non-negligible power consumption in the clock generation. Moreover, the presence of such fast squared signals is not suitable for a system-on-chip comprising radio frequency receivers.

In this thesis we propose a low-power strategy relying on a large number of internal levels rather than on a high sampling frequency or modulation order. Besides, a hybrid continuous-discrete-time approach is used to take advantage of the accuracy of switched-capacitor circuits and the low power consumption of continuous-time implementation. The sensitivity to clock jitter brought about by the continuous-time stage is reduced by the use of a large number of levels. An auto-ranging algorithm is developed in this thesis to overcome the limitation of a large-size quantizer under low-voltage supply. Finally, the strategy is applied to a design example addressing typical specifications for a Bluetooth receiver with direct conversion.

Keywords: analog-to-digital conversion, delta-sigma modulation, multibit, quantization noise, hybrid architecture, continuous-time, discrete-time, full clock-cycle sampling, double-sampling, digital calibration, autoranging algorithm, tracking quantizer, dynamic element matching, spectral shaping, tree-structured encoder, segmented DAC, Bluetooth, WCDMA, GSM, EDGE, comparator, CMOS technology, fully-differential amplifier, transconductance amplifier.

Résumé

Dans les applications portables de communication sans-fil, une grande partie du traitement du signal est exécutée de manière numérique. Les circuits numériques présentent plusieurs avantages. Même pour un haut degré de complexité, la consommation d'énergie ainsi que les coûts de fabrication sont faibles. Le processus de conception est bien établi avec un haut degré d'automatisation permettant de tirer profit du progrès constant des technologies CMOS. En outre, les circuits numériques offrent un traitement du signal robuste et programmable et ne nécessitent aucun composant externe. La tendance pour le marché grand public est donc de réduire encore plus la partie analogique dans les chaînes de réception de systèmes sans fil. En conséquence, des convertisseurs avec des résolutions et des bandes passantes toujours plus élevées sont nécessaires. Le but ultime est la numérisation directe du signal radio, où la conversion serait faite après amplification du signal fournit par l'antenne.

Les convertisseurs basés sur la modulation $\Delta\Sigma$se trouvent être les plus appropriés pour atteindre les performances requises. Les implémentations à capacités commutées ont été très utilisées ces deux dernières décennies. Cependant, de récentes publications ont montré que les architectures à temps continu pouvaient atteindre les mêmes performances avec une consommation réduite. La plupart des réalisations présentées dans les publications font usage d'un quantificateur interne à deux ou à un nombre restreint de niveaux et une modulation d'ordre élevé. Par conséquent, pour atteindre les résolutions et bandes passantes demandées de nos jours, la fréquence d'échantillonnage dépasse la centaine de mégahertz. Cette approche amène à une consommation non négligeable du générateur d'horloge. En outre, la présence de signaux carrés d'une telle rapidité est à éviter sur un système intégré de grande taille qui comprend des récepteurs radio.

Dans cette thèse, nous proposons une stratégie à faible consommation qui compte sur un grand nombre de niveaux internes plutôt que sur une haute fréquence d'échantillonnage et un ordre de modulation élevé. De plus, une approche hybride à temps continu et discret est employée afin de profiter de la précision des circuits à capacités commutées et de la faible consommation des circuits à temps continu. La sensibilité à la gigue du signal d'horloge, introduite par le traitement à temps continu, se voit diminuée par l'utilisation d'un nombre élevé de niveaux internes. Un algorithme d'ajustement automatique d'amplitude est développé dans cette thèse afin de gommer l'inconvénient d'un quantificateur de grande taille travaillant sous basse tension d'alimentation. Pour terminer, la stratégie est appliquée à un exemple de conception visant les spécifications typiques d'un récepteur Bluetooth à conversion directe.

Mots-clés: conversion analogique-numérique, modulation delta-sigma, multibit, bruit de quantification, architecture hybride, temps continu, temps discret, échantillonnage sur deux demi périodes, double échantillonnage, calibration numérique, algorithme d'ajustement automatique d'amplitude, quantificateur, appariement dynamique, mise en forme spectrale, encodeur à structure en arbre, CAN segmenté, Bluetooth, WCDMA, GSM, EDGE, comparateur, technologie CMOS, amplificateur différentiel symétrique, amplificateur à transconductance.

Contents

Contents xi

1 Introduction 1

1.1 Motivation 1

1.2 Intention of the this work 2

1.3 Hybrid multi-bit modulators 2

1.4 Organization of the dissertation 3

2 Delta-Sigma ADC from the ground 5

2.1 From analog to digital 5

2.1.1 Sampling 5

2.1.2 Quantization 8

2.1.3 Over-sampling 12

2.1.4 Spectral shaping 14

2.1.5 Digital decimation 19

2.2 Performance calculations 23

2.2.1 Linear model 23

2.2.2 Modulator general description 24

2.2.3 Dynamic range 25

2.2.4 Resolution 27

2.2.5 Simulations 31

2.2.6 Aggressive pole placement 34

2.2.7 Optimal zero placement 36

2.3 Circuit imperfections 38

2.3.1 Removal of quantization noise 38

2.3.2 Circuit noise 39
2.3.3 Clock jitter 40
2.3.4 Component mismatch 41
2.3.5 Harmonic distortion 42
2.4 Architecture classification 44
2.4.1 Single- and multi-bit 44
2.4.2 Low-, high- and band-pass 46
2.4.3 Single and multi-feedback path 46
2.4.4 Single and multi-stage 48
2.5 Conclusions 50

3 Low-power strategy 51
3.1 Strategy outline 51
3.1.1 Continuous-time implementations 51
3.1.2 Multi-bit feedback 52
3.1.3 Optimal DEM segmentation 53
3.1.4 Auto-ranging algorithms 55
3.1.5 Hybrid architecture 56
3.1.6 Continuous-to-discrete domain interface 57
3.1.7 Summary 58
3.2 Anti-aliasing property 60
3.2.1 Signal processing of the continuous-time modulator 60
3.2.2 Input signal spectrum folding 62
3.3 Clock jitter issue 64
3.3.1 Discrete- and continuous-time mechanisms 64
3.3.2 Analytical expression of the jitter transfer factor 68
3.3.3 Resolution degradation 71
3.4 Accuracy of coefficients 76
3.5 Sampling capacitor mismatch 79
3.6 Conclusions 84

4 Auto-ranging algorithm 85

4.1 Observations 85

4.1.1 Slow and fast components 85

4.1.2 Exploiting the phenomenon 86

4.1.3 Auto-ranging technique 88

4.2 Principle in detail 88

4.2.1 Last integrator reuse for signal shifting 88

4.2.2 Control function 91

4.2.3 Implementation constraints 92

4.2.4 Number of step changes 94

4.3 Realization and limitations 96

4.3.1 Simulations 96

4.3.2 Input signal constraints 100

4.3.3 Sensitivity to circuit imperfections 102

4.3.4 Hardware implementation 105

4.4 Optimization 108

4.4.1 Reduced number of quantization levels 108

4.4.2 Reduction factor 110

4.4.3 Extendability 110

4.5 Conclusions 114

5 Mismatch shaping optimizations 115

5.1 Dynamic element matching 115

5.2 Tree-structured architectures 117

5.2.1 Spectral shaping 118

5.2.2 Higher-order spectral shaping 121

5.3 Optimization 123

5.3.1 DAC segmentation 123

5.3.2 Synthesis automation 124

5.3.3 Standard segmentation 128

5.4 Conclusions 132

6 Design example 133
- 6.1 Modulator design . . . 133
  - 6.1.1 Targeted specifications and topology choice . . . 133
  - 6.1.2 Circuit description . . . 134
- 6.2 System Design . . . 137
  - 6.2.1 Signal processing of the circuit . . . 137
  - 6.2.2 Design equations . . . 139
  - 6.2.3 Design procedure . . . 142
  - 6.2.4 Noise considerations . . . 144
- 6.3 Design of the amplifiers . . . 147
  - 6.3.1 Behavioral models . . . 147
  - 6.3.2 Continuous-time amplifier . . . 148
  - 6.3.3 Switched capacitor devices . . . 150
- 6.4 Quantizer design . . . 154
  - 6.4.1 Comparator . . . 154
  - 6.4.2 Offset compensation circuit . . . 156
  - 6.4.3 Residual offset . . . 158
- 6.5 Final circuit . . . 161
  - 6.5.1 Benchmarking . . . 164
- 6.6 Conclusions . . . 167

7 Conclusions 169
- 7.1 Thesis outlook . . . 169
- 7.2 Main contributions . . . 171
- 7.3 Future perspectives . . . 172

A Mathematical definitions and proofs 175
- A.1 Definition of the norms . . . 175
- A.2 The 1-norm of a z transfer-function . . . 176
- A.3 The 2-norm of a z transfer-function . . . 177
- A.4 Differential pair model . . . 178

B Verilog-A codes 183

B.1 Transconductance behavioral model 183

C MATHEMATICA codes 185

C.1 Synthesis functions 185

C.2 Extraction and display of the solutions 186

D MATLAB codes 189

D.1 Modulators with auto-ranging 189

D.2 Mismatch shaping encoder 193

Bibliography 195

List of Figures 203

List of Tables 213

List of publications 217

Chapter

1

Introduction

1.1 Motivation

In wireless portable applications, a large part of the signal processing is performed in the digital domain. Digital circuits show many advantages. The power consumption and fabrication costs are low even for high levels of complexity. A well established and highly automated design flow allows one to benefit from the constant progress in CMOS technologies. Moreover, digital circuits offer robust and programmable signal processing means and need no external components. Hence, the trend in consumer electronics is to further reduce the part of analog signal processing in the receiver chain of wireless transceivers. Consequently, Analog-to-Digital Converters (ADC) with higher resolutions and bandwidths are constantly required. According to [Raz96], the ultimate goal is the direct digitization of Radio Frequency (RF) signals, where the conversion would be performed immediately after the front-end Low-Noise Amplifier (LNA).

$\Delta\Sigma$-modulation-based converters have proved to be the most suitable to achieve the required performance [Jes01, GWT02]. Discrete-Time (DT) Switched-Capacitor (SC) implementations have been widely used over the last two decades. However, recent publications [Kap03, YS04, AL02] and books [CS00, BH01, KvR06, OG06, Sho95, Yan02] have shown that Continuous-Time (CT) architectures can achieve the same performance with lower power consumption. Besides, CT modulators present an inherent anti-aliasing property, relaxing the specifications of the analog low pass filter placed in front of the ADC. Most designs found throughout the literature use a single- or few-bit internal quantizer with a high-order modulation. As a result, in order to achieve the resolutions and bandwidths required today, the sampling frequency must exceed 100MHz. This approach leads to non-negligible power consumption in the Phase-Locked Loop (PLL) generating the clock signals. Moreover, the presence of such fast squared signals is not suitable for a System-on-Chip (SoC) comprising RF receivers.

1.2 Intention of the this work

The objective of this thesis is to propose a low-power strategy relying on a large internal Number-of-Levels (NL) rather than on a high sampling frequency (f_s) or modulation order (n). Besides, a hybrid continuous-discrete-time approach is used to take advantage of the accuracy of SC circuits and the low power consumption of continuous-time implementations. The intention of this work is to provide designers with an optimization methodology.

1.3 Hybrid multi-bit modulators

In the general case, a $\Delta\Sigma$modulator consists of an NL-level quantizer and an nth-order analog filter providing spectral shaping of the quantization noise. The analog filter is set up with n successive integration stages. From the second to the last one, each successive stage takes advantage of an additional order of the spectral shaping. As a consequence, the power consumption of the modulator is generally dominated by its first stage. Hence, having the first stage as a CT integrator enables a significant amount of power saving. On the other hand, keeping the upper stages as DT integrators allows one to take advantage of the accuracy of SC circuits coefficients. The modulator still takes partial advantage of inherent anti-aliasing filtering. Figure 1.1 shows an illustrative example of such a hybrid architecture.

Multi-bit hybrid architectures were proposed by [MCL$^+$05, KRSC05] for audio applications and by [RLS$^+$03] for communication systems. In this thesis, we suggest of further increasing the number of levels. An auto-ranging algorithm is developed to overcome the limitation of a large-size quantizer with a low voltage supply. The complexity of the Dynamic Element Matching (DEM) is addressed by an appropriate segmentation

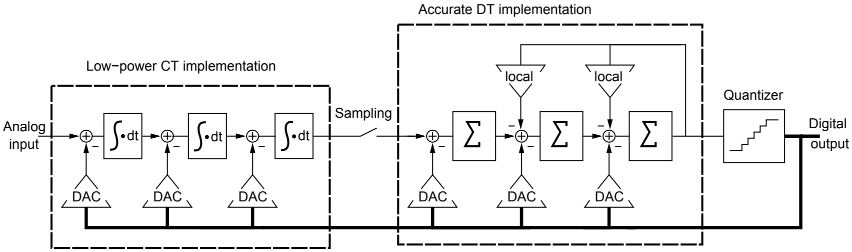


Figure 1.1 Illustrative example of the hybrid architecture.

of the DACs as proposed by [FSW$^+$02, Gal97]. A synthesis algorithm is developed to evaluate all the possible segmentations.

1.4 Organization of the dissertation

The thesis starts by a didactic review of the $\Delta\Sigma$-modulation signal processing, then presents the low-power strategy and its relevant aspects, and ends with a design example.

Chapter 2 introduces the $\Delta\Sigma$modulation for analog-to-digital conversion. The fundamental aspects of the signal processing are covered as well as the methods used for performance estimation. An accurate analytical expression of the expected resolution is derived for the general case of a multi-bit single-stage modulator. The impact of the most relevant circuit imperfections is shown. The chapter ends with an overview of the well-known architectures and variants commonly used nowadays.

Chapter 3 proposes the low-power strategy which consists of an optimal combination of different techniques. Relevant aspects of the techniques used are covered. The sensitivity to clock jitter is studied in detail and used as a main optimization criterion.

Chapter 4 explains the principle of the auto-ranging algorithm. The main limitations of the technique are analyzed, taking into account the implementation constraints. An analytical expression of the efficiency of the algorithm is developed and used for optimization. The extendability to different applications is demonstrated.

Chapter 5 discusses the optimization of the DEM. The segmented tree-structured architecture is introduced. Different shaping algorithms are provided as well as a synthesis method. An analytical expression of the power consumption is given.

Chapter 6 presents a design example addressing the specifications for a BLUETOOTH receiver with direct conversion. The modulator equations are derived, giving a general design procedure. The design of the amplifiers and the quantizer is outlined. The digital calibration of the comparators offset is analyzed. The chapter ends with a comparison with other published works.

Chapter 7 gives the final conclusions and summarizes the main contributions of the thesis. Suggestions for further development and analyses are made.

Chapter
2
Delta-Sigma ADC from the ground

The purpose of this chapter is to cover the fundamental aspects of $\Delta\Sigma$ analog-to-digital conversion. It considers the basics of signal processing up to the main issues known today. It also introduces the definitions commonly used and makes the link between the analytical models and the reality of implementation. The first section enables the non-expert reader to rapidly understand how a low-resolution quantizer can become a high-resolution converter by subsequently over-sampling, grafting a control system and filtering. The detailed calculations of the expected performance are given in the second section. The third section introduces the main circuit imperfections that would potentially reduce the expected performances. Finally, the last section gives a broad view of the most classic architectures known today.

2.1 From analog to digital

2.1.1 Sampling

The first task in analog-to-digital conversion consists in providing samples of the analog signal equally apart in time by the *sampling period* T_s. This process of going from the *continuous-time* to the *discrete-time* domain is called *sampling*. We define the *sampling frequency* as $f_s = 1/T_s$. After sampling, there is no any signal faster than half of the sampling frequency, also referred to as the *Nyquist frequency*.

Figure 2.1 shows a low-pass filter followed by an ideal sampler. Code 2.1 and 2.2 simulate the sampling of a single-tone and the additional filtered resistor white Gaussian noise, whose Power Spectral Density (PSD) is given by:

$$\mathscr{P}_N(f) = 4kTR\,. \tag{2.1}$$

As shown in Figure 2.2(a), when the noise is sampled at a frequency higher than the filter cut-off frequency f_{RC}, no aliasing occurs and the output spectrum contains the tone and the filtered thermal noise. In contrast, Figure 2.2(b) shows that if the bandwidth of the noise is higher than the Nyquist rate, the noise unavoidably folds back and its level increases. As treated in detail in many textbooks [Raz01, JM97], the noise floor is multiplied by $\pi/2$ and the ratio between the cut-off and Nyquist frequencies.

$$\mathscr{P}_N(f) = 4kTR\cdot\frac{\pi}{2}\cdot\frac{2f_{RC}}{f_s}\,. \tag{2.2}$$

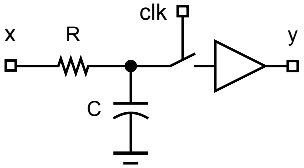


Figure 2.1 Sampling: An RC-filter is placed in front of the sampler providing anti-aliasing filtering.

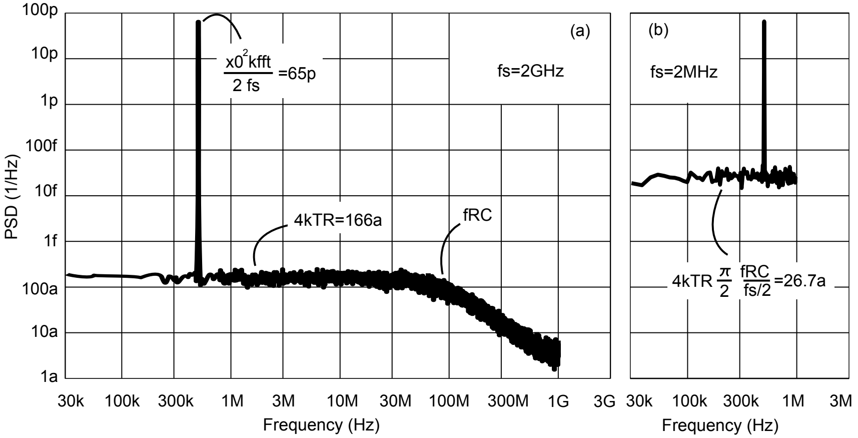


Figure 2.2 Sampling process: Simulations performed with 50 average points and 2^{18} FFT bins. A 1mV tone at 500kHz together with a 10kΩ resistor thermal noise are sampled at a frequency of 2GHz (a) and 2MHz (b). The filter cut-off is set at 100MHz.

Code 2.1 MATLAB code for the sampling process: **avr** and **kfft** control the averaging number and the FFT number of bins. The down-sampling factor is set by **down**, whereas **f0**, **fs** and **fRC** are respectively the tone, sampling and filter cut-off frequencies.

```
%%%%%%%%%%% parameters %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
avr=50;waste=1024;down=1024;
kfft=2^18;k=kfft*avr+2*waste;    %sim. parameters
f0=500e3;fs=2e9;                 %signal/sampling freq.
f0=fs*round(kfft*f0/fs)/(kfft); %integer nbr.of period
fRC=100e6;                       %filter cut-off freq.
v0=1.0e-3;                       %signal amplitude
n0=4.07e-4;                      %sqrt(4kTR fs/2))

%%%%%%%%% generation of signals %%%%%%%%%%%%%%%%%%%%%%%%%%%
noise  = n0*randn(k,1);
signal = v0*(sin(2*pi*f0*(1:k)'/fs));
v=signal+noise;
RCfilter=(1-exp(-2*pi*fRC/fs))*exp(-(2*pi*fRC/fs)*(0:2^10)');
v_filtered =filter(RCfilter,1,v);
v_down  =v_filtered(1:down:k);
```

Code 2.2 MATLAB for power spectral density esitmation.

```
%%%%%% power spectral density of the filtered signal %%%%%%%%
y=v_filtered;
psd = zeros(kfft, 1);
for j=0:avr-1
    psd=psd+...
    abs(fft(y(waste+1+j*kfft:waste+(j+1)*kfft))).^2/kfft/fs*2;
end
psd=psd(1:kfft/2+1)/avr;f=(0:fs/kfft:fs/2);loglog(f,psd);

%%%%%% power spectral density of the down-sampled signal %%%%%
kfft=kfft/down;fs=fs/down;waste=waste/down;
y=v_down;
psd = zeros(kfft, 1);
for j=0:avr-1
    psd=psd+...
    abs(fft(y(waste+1+j*kfft:waste+(j+1)*kfft))).^2/kfft/fs*2;
end
psd=psd(1:kfft/2+1)/avr;f=(0:fs/kfft:fs/2);loglog(f,psd);
```

With the circuit described in Figure 2.1 the cut-off is given as $2\pi RC$. As a well-known result, commonly referred to as the *kT/C-noise*, the noise floor becomes independent of the resistor size:

$$\mathscr{P}_N(f) = \frac{kT}{C} \cdot \frac{2}{f_s} \,. \tag{2.3}$$

All these considerations reveal that the process of sampling itself already degrades the quality of the signal, even before its conversion into a digital representation. Furthermore, it gives good reasons to place an *anti-aliasing* filter in front of an ADC to avoid the folding of any signal or noise present at frequencies higher than the Nyquist rate

The experiment of Code 2.1 introduces a few important features of spectral estimations. First of all, it is important to appropriately round the number of periods of the tone in the simulation in order to avoid spectral spreading. Secondly, in the presence of random signals, averaging is essential to reveal the continuous shapes hidden by important variations. A power-of-two is suitable choice for the number of samples to avoid an artificial zero padding by the *Fast Fourier Transform* (FFT) algorithm. Finally, in order to provide a representation of the single side-band PSD, the absolute value of the FFT is squared, normalized by the number-of-bins k_{FFT} and by the sampling frequency [II02, HvV99]:

$$\mathrm{PSD}[k] = \frac{2\left|\mathrm{FFT}[k]\right|^2}{f_s k_{\mathrm{FFT}}} \;\forall\, k \;\in\; \{0,\ldots,k_{\mathrm{FFT}}/2\} \,. \tag{2.4}$$

2.1.2 Quantization

The second task of the conversion consists in providing samples with an amplitude rounded to the closest value of a finite set of so-called *levels*. This process of going from the continuous-amplitude to the discrete-amplitude is called *quantization*.

This task is performed by a *quantizer*, also called *flash ADC*, which is the core of any analog-to-digital conversion circuit. Figure 2.3 describes a quantizer with NL levels. In some architectures, such as successive approximation, single-bit $\Delta\Sigma$or pipeline converters, NL can go down to two, requiring only one comparator.

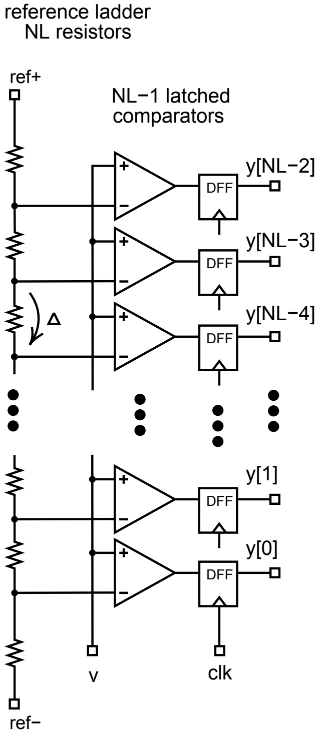


Figure 2.3 A quantizer with NL levels consists of a bank of (NL − 1) comparators, providing a quantization in amplitude, and(NL − 1) Data-Flip-Flops (DFF), providing the quantization in time. The (NL − 1) voltage thresholds, separated by steps of Δ, are usually generated by a resistive ladder with NL resistors. The (NL − 1) outputs of the data-flip-flops form an integer number, represented by a thermometer code.

Figure 2.4 Mid-thread 5-level quantizer transfer characteristic described by Equation (2.5). The vertical parts correspond to the voltage thresholds and the horizontal parts to the possible output levels.

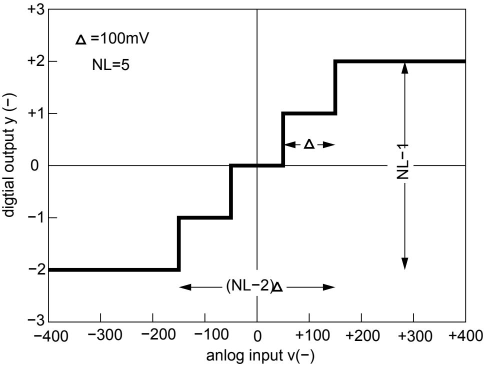


Figure 2.4 shows the transfer characteristic from the analog input v to the digital output y, which in all cases can be described by

$$y = \begin{cases} \mathrm{sign}(v)(\mathrm{NL}-1)/2\ , & \text{if } |v| \geq \mathrm{NL}\Delta/2\ , \\ \mathrm{round}(v/\Delta)\ , & \text{if NL is odd}\ , \\ \mathrm{floor}(v/\Delta)+1/2\ , & \text{otherwise}\ . \end{cases} \tag{2.5}$$

Whether the Number-of-Levels NL is even or odd, the quantizer has a mid-rise or mid-thread transfer characteristic. NL is often chosen as a power-of-two because the quantizer output can be further encoded in a binary representation with the minimum Number-of-bits N, where $\mathrm{NL} = 2^{\mathrm{N}}$.

Alternatively, choosing NL as a power-of-two plus one allows encoding in the so-called *extra-LSB* representation, as is necessary for the use of a tree-structured mismatch shaping encoder, as discussed in Chapter 5. In such a case, NL is odd and the number of comparators is even. In this thesis only quantizers with power-of-two plus one levels are considered.

Code 2.3 MATLAB for the quantization process: The tone amplitude **v0** is set at the maximum allowed, namely at half of **NL**. The quantization error **q** is extracted to determine its statistical distribution. The power spectral density is evaluated as already described in Code 2.1. For the sake of simplicity, an odd-NL quantizer with steps equal to one is considered.

```
%%%%%%%%%%% parameters %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
avr=20;kfft=2^11;k=kfft*avr;    %sim. parameters
NL=2^5+1;                       %Number-of-Levels
v0=NL/2;                        %signal amplitude
f0=200e3;fs=32e6;               %signal/sampling freq.
f0=fs*round(kfft*f0/fs)/(kfft); %integer nbr.of period
%%%%%%%%%%% signals %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
v = v0*(sin(2*pi*f0*(1:k)'/fs));
y = zeros(k,1);
%%%%%%%%%% quantization %%%%%%%%%%%%%%%%%%%%%%%%%%%%%
for i=1:k
    if (abs(v(i)) >=NL/2)
        y(i)=sign(v(i))*(NL-1)/2;
    else
        y(i)=round(v(i));
    end
end
q=y-v;     %quantization noise extraction
```

Code 2.4 Additional MATLAB code for dithering: The tone amplitude is reduced by **extra** to prevent over-loading as the necessary dithering amplitude **d0** is quite high

```
extra=-1.5;                     %input range reduction
v0=NL/2+extra;                  %signal amplitude
d0=0.5; %dither signal amplitude
dither = d0*randn(k,1);
....
in=in+dither; %dithering
....
out=out-dither; %dither removal
```

Code 2.3 is a simulation of the quantization process applied to a single tone. The results in Figure 2.5 show an output PSD that is nothing but

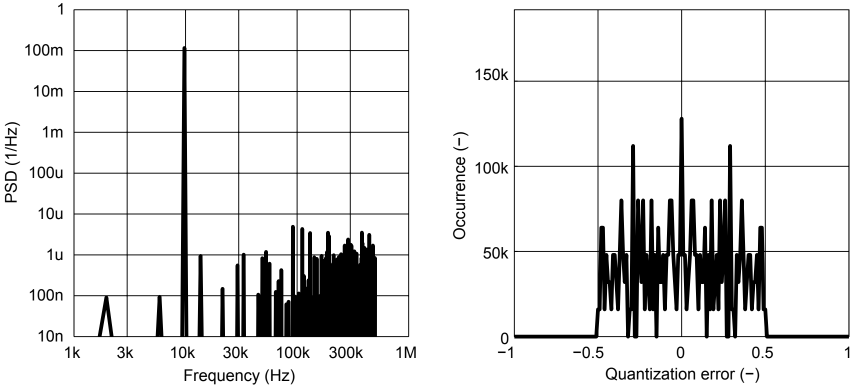


Figure 2.5 Quantization process: Simulation performed with 2000 averages, NL=33, extra=−1.5, single-tone at 10kHz and sampling frequencies f_s=1MHz.

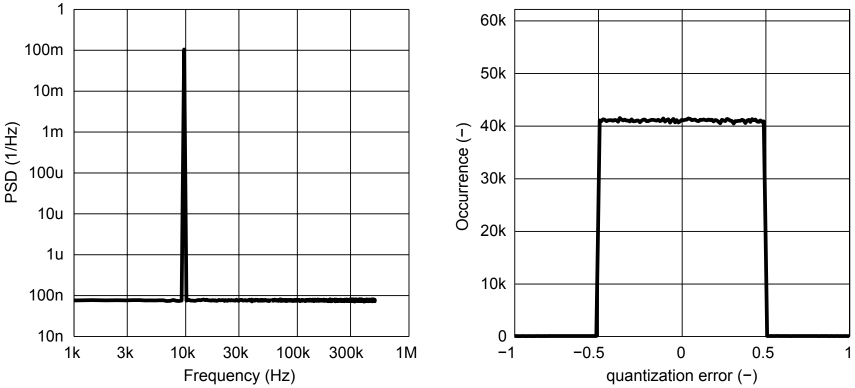


Figure 2.6 Dithered quantization: Simulation performed with 2000 averages, NL=33, extra=−1.5, Gaussian dithering of 0.5, single-tone at 10kHz and sampling frequencies f_s=1MHz.

a sum of harmonics because of the deterministic nature of the quantization process. In the second experiment, with the modification proposed in Code 2.4, a Gaussian *dithering* signal is artificially added at the input and removed at the output. The quantization becomes a random process. According to Figure 2.6, the output PSD is white and the statistical distribution is uniformly distributed between ±1/2. In real cases, the signal to be converted is quite random and no dithering is usually required.

This observation leads to the so-called *linear approximation* where the highly non-linear behavior of Equation (2.5) is modeled as a gain of $1/\Delta$. Any error from that linear characteristic generated by the quantization process is considered an additional random noise. As highlighted in Figure 2.7, this approximation holds as long as the input stays within $\pm NL\Delta/2$, namely as long as errors are bounded and random. If we exceed this range, the quantizer overloads and, as shown in Figure 2.8, with a sinusoidal input, harmonic distortion is generated.

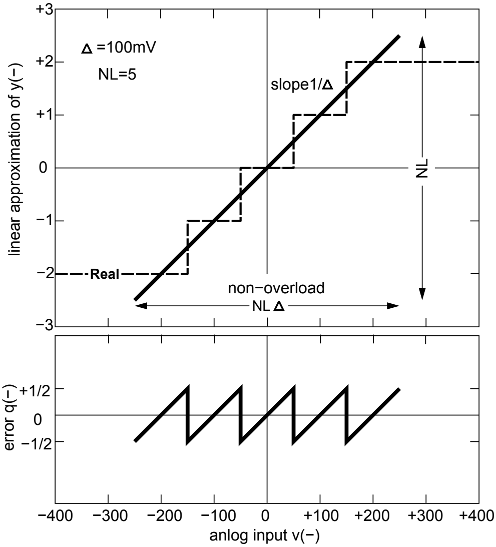


Figure 2.7 A mid-thread 5-level quantizer: The real transfer characteristic is shown as a dashed line. The linear model consists in separating the characteristic in a gain of $1/\Delta$ and a quantization error q bounded by $\pm 1/2$. Beyond the range of $\pm NL\Delta/2$ the quantizer is overloaded and the linear model no longer holds.

The subject of quantization noise is only outlined here to give an insight. The subject is handled analytically at greater length in [Gra90, Gal94, Gal93].

2.1.3 Over-sampling

Today it is hard to find applications requiring quantizers with a number of bits less than 10, meaning 1024 levels. Building a quantizer with such a

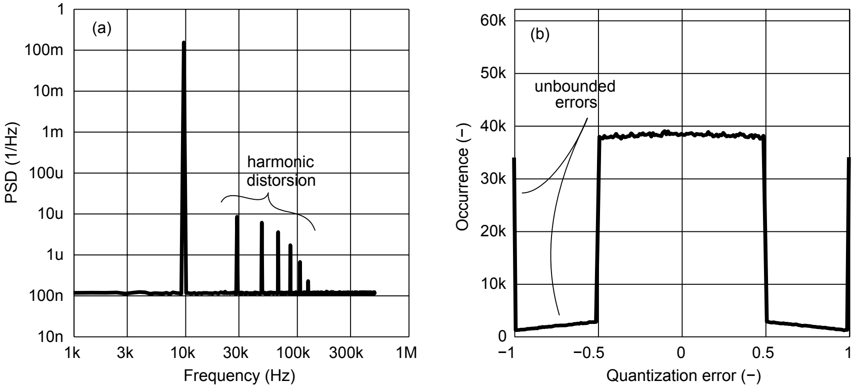


Figure 2.8 Overloaded quantizer: 2000 averages, NL=33, extra=0, Gaussian dithering of 0.5, in-band single-tone at 10kHz, sampling frequencies fs=1MHz.

large NL is impractical. But, as explained in more detail in the next section, the quantization noise level is inversely proportional to the sampling frequency. We can therefore choose to sample the signal faster.

Let us suppose that the input signal frequency never exceeds a certain *band-of-interest* f_b. If f_s equals twice f_b, the quantizer is said to operate at the *Nyquist-rate*, whereas if f_s is higher we talk about *over-sampling.* We commonly define the *Over-Sampling Ratio* OSR as how much faster the signal is sampled with respect to the Nyquist rate.

$$\text{OSR} = \frac{f_s}{2f_b} \,. \tag{2.6}$$

Figure 2.9 highlights how the 33-level example studied previously, over-sampled 32 times, sees its quantization level dropping. In addition to lowering the quantization noise PSD, over-sampling provides a frequency space able to contain unwanted signals that would be removed afterwards. Signals appearing either inside or outside the band-of-interest are usually referred to as the *in-band* and *out-of-band* signals respectively.

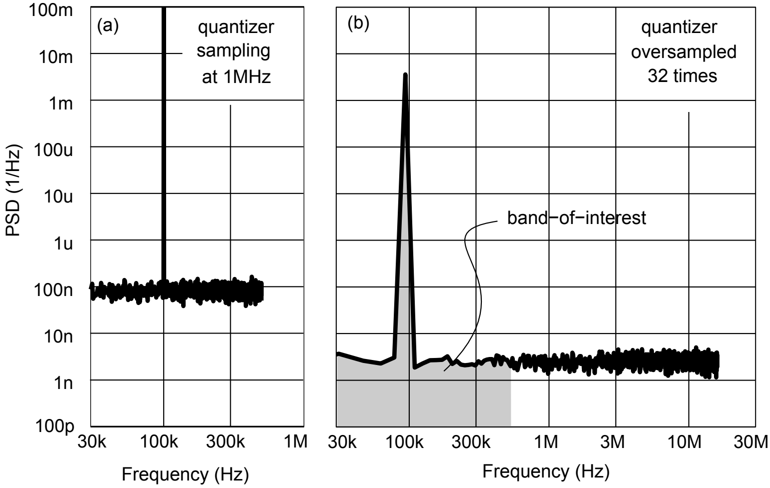


Figure 2.9 Over-sampled quantizer: Simulation results of Code 2.3 with avr=20, NL=33, extra=−1.5, d_0=0.5. As the sampling frequencies goes from (a)1MHz to (b)32 MHz, the quantization noise, integrated over the band of interest, here from 0 to 500kHz, drops by the same factor. Meanwhile, the area under the in-band tone at 100kHz remains unchanged.

2.1.4 Spectral shaping

$\Delta\Sigma$-modulation consists in grafting an analog error control system onto an over-sampled quantizer. The digital quantizer output is converted by DACs and re-injected before the quantizer, so creating one or more feedback loops. Unlike the quantizer, a DAC does not introduce any quantization error and is therefore represented by a linear gain as in Figure 2.10.

A nth-order modulator is a cascade of n integrators whose purpose is to provide a strong negative reaction at low frequencies. This forces the quantizer output to be an accurate replica of the analog input at low frequencies. This process is commonly referred to as *spectral noise shaping* and its mathematical description is analyzed in detail in the next section.

Figure 2.10 describes a conventional nth-order $\Delta\Sigma$-modulator implemented by Code 2.5.

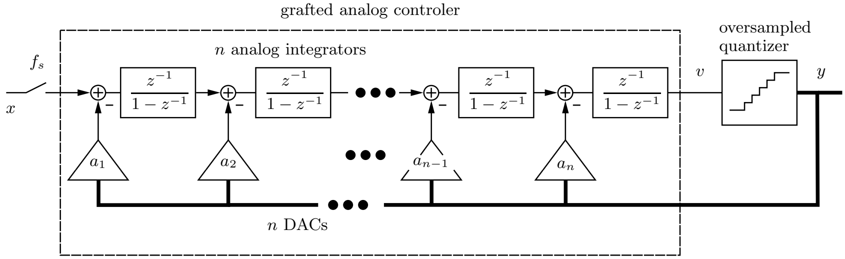


Figure 2.10 ΔΣ-modulation: Conventional nth-order architecture.

Code 2.5 MATLAB code for an nth-order ΔΣ-modulator.

```
%%%%%%%%%%% parameters %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
coeff=[1 3 3];
n=length(coeff);
x0=(NL-2^n+1-2)/2;
%%%%%%%%%%% signals %%%%%%%%%%%%%%%%%%%%%%%%%%%
dit = d0*randn(k+n+1,1);
sig = zeros(k,n+1);
out = zeros(k,1);
x=x0*(sin(2*pi*f0*(1:k)'/fs));
sig(:,1)=x+dit(n+1:k+n);
for i=n+1:k
  %%%% integrators %%%%%%%%%%%%%%%%%%%%%%%%%%%%%
  for r=1:n
    sig(i,r+1)=sig(i-1,r+1)-coeff(r)*out(i-1)+sig(i-1,r);
  end
  %%%%%%% quantizer %%%%%%%%%%%%%%%%%%%%%%%%%%%
  if (abs(sig(i,n+1)) >= NL/2)
    out(i)=sign(sig(i,n+1))*(NL-1)/2;
  else
    out(i)=round(sig(i,n+1));
  end
end
out=out-dit(1:k);          %dither remotion
```

Figure 2.11 shows the quantizer output with a first-order control system. Compared to the over-sampled quantizer alone, the quantization noise level droped around DC. In contrast, the input tone remained unchanged. The control system provided what we call *spectral shaping*. As for the simple over-sampled case, we consider our input signals are within a certain band-of-interest.

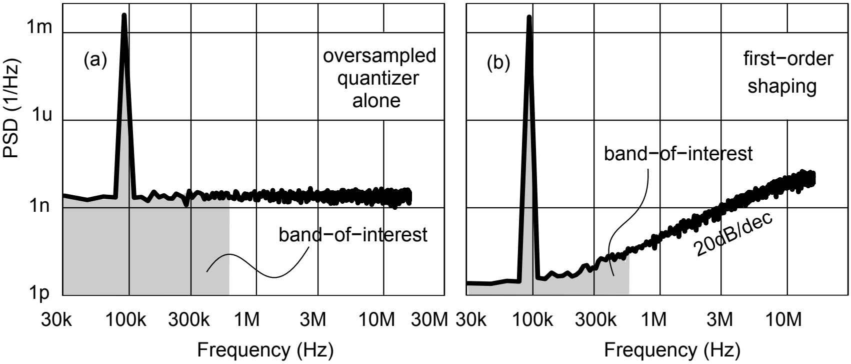


Figure 2.11 Spectral shaping: Simulation results of (a) Code 2.3 and (b) Code 2.5 with avr=20,NL=33,extra=-1.5, d_0=0.5,f_s=32MHz. The shaping provides a 3-decade attenuation of the quantization level at low frequencies. Meanwhile, the area under the in-band tone at 100kHz remains unchanged.

Figure 2.12 shows the quantizer output with a second-order control system. At first sight, the second-order shaping does not bring any benefit with respect to the first-order shown previously. In reality, the quantization noise floor goes further down but this is hidden by a phenomenon called *spectral leakage*.

Fourier analysis supposes that the data go from minus infinity to plus infinity and are perfectly periodic. To this extent what we did is equivalent to multiplying these samples by a finite rectangular window. As a result, the spectrum of the data, periodic and of infinite length, convolves with the spectrum of the window. Hence, the spectrum of the rectangular window is a sinc function and therefore has so-called *side-lobes* which make any point of the data spectrum *leak* on the other frequencies. Well-known textbooks like [OSB99, HvV99], where the subject is studied in detail, propose other windows with reduced side-lobes but increased speed of the spectrum. The *Blackman-Harris* window provides an interesting trade-off. Figure 2.12(b) shows the same second-order modulator output with the proposed modification of Code 2.6 including the windowing of the samples. The tone is now represented by about five dominant points and the side-lobes are low enough to reveal the drop of quantization noise level.

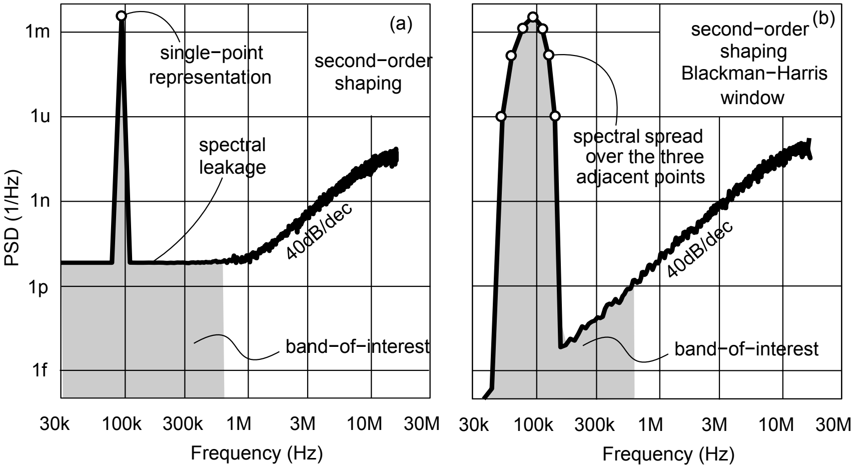


Figure 2.12 Second-order spectral shaping: 20 averages, NL=33, extra=−1.5, Gaussian dithering of 0.5, in-band single-tone at 100kHz, sampling frequencies f_s=1MHz and 32MHz.

Code 2.6 Modification of the MATLAB code for PSD estimation with Blackman-Harris window

```
w=blackmanharris(kfft);
w=w/mean(w);
psd=zeros(kfft, 1);
for j=0:avr-1
    y=w.*out(1+j*kfft:(j+1)*kfft);
    psd=psd+abs(fft(y)).^2/kfft/fs;
end
psd=psd(1:kfft/2+1)/avr;
```

As we increase either n or NL, the tone representation in the PSD moves up and the side-lobes around the tone emerge from the noise. As highlighted by the third-order case in Figure 2.13, it becomes harder to make a clear distinction between the tone and the noise contributions.

Another way of proceeding, practiced in particular by [FSW$^+$02], consists in extracting the tone from the samples using the best root-mean-square method. Then both the tone and the noise PSD are evaluated separately. Distinction between the tone and the noise is now guaranteed and a narrow spectrum window, like the Blackman-Harris, is therefore not

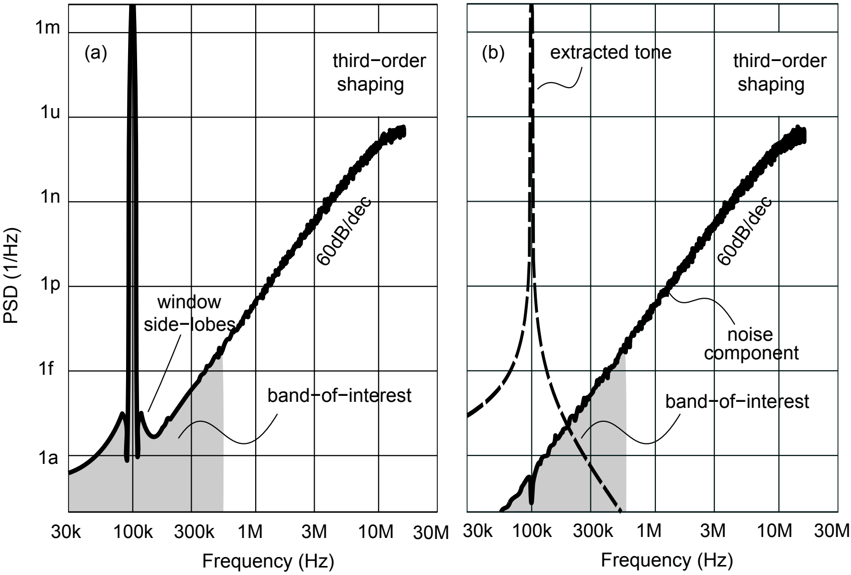


Figure 2.13 Third-order spectral shaping: 20 averages, NL=33, no extra, no dithering, in-band single-tone at 100kHz, sampling frequencies 32 MHz.

necessary. Only the side-lobes are to be considered to avoid leakage. We can also remove the period synchronization, since an integer number of periods is not essential. Code 2.7 encompasses the modifications of PSD estimation.

Code 2.7 Modification of MATLAB code for PSD estimation with signal extraction method.

```
w=hann(kfft);
w=w/mean(w);
s=w.*sin(2*pi*f0*(1:kfft)'/fs);
c=w.*cos(2*pi*f0*(1:kfft)'/fs);
psd_n=zeros(kfft,1);
psd_s=zeros(kfft,1);
for j=0:avr-1
    y=w.*out(1+j*kfft:(j+1)*kfft);
    A=  [+sum(y.*s) +sum(y.*c)]/...
        [+sum(s.^2) -sum(s.*c);...
         -sum(s.*c) +sum(c.^2)];
    y_s=(A(1).*s+A(2).*c);
    y_n=y-y_s;
    psd_n=psd_n+abs(fft(y_n)).^2/kfft/fs;
    psd_s=psd_s+abs(fft(y_s)).^2/kfft/fs;
end
psd_s=psd_s(1:kfft/2+1)/avr;
psd_n=psd_n(1:kfft/2+1)/avr;
```

A Hann window is used. Ideal sine-wave and cosine-wave vectors s_i and c_i are generated. Their amplitudes, s_0 and c_0, are found by minimizing the root-mean-square errors with respect to the sample vector y_i:

$$\frac{\partial}{\partial s_0, c_0} \sum_{i=1}^{k} (y_i - s_0 s_i - c_0 c_i)^2 = 0 \ . \tag{2.7}$$

which lead to the following solution:

$$\begin{bmatrix} s_0 \\ c_0 \end{bmatrix} = \begin{bmatrix} +\sum s_i^2 & +\sum s_i c_i \\ -\sum s_i c_i & +\sum c_i^2 \end{bmatrix}^{-1} \begin{bmatrix} +\sum y_i s_i \\ -\sum y_i c_i \end{bmatrix} \ . \tag{2.8}$$

Note that in this last example no dithering is used. In fact, as the order is higher, the length of any output pattern that would repeat periodically becomes extremely large.

2.1.5 Digital decimation

As mentioned earlier, the digital quantizer output goes through a filter, ideally removing everything outside the band, and is re-sampled at the Nyquist rate. Such a process is called *decimation*. The design of the digital filter depends strongly on the application, more specifically on what we expect as out-of-band unwanted signals. In most cases, the decimation is done in two stages. Each of them consists of an anti-aliasing filter followed by a down-sampler. Because a linear phase response is often required and stability must be guaranteed, Finite Impulse-Response (FIR) filters with symmetric coefficients are used almost exclusively.

The first-stage performs a rough filtering and down-samples at twice the Nyquist rate. It consists of a cascade of n identical m-tap FIR filters. Because all the m coefficients are equal to one, this is called an nth-order *comb-filter*. The transfer function $H_c(z)$ of the filter becomes a geometric series and can therefore be rewritten as follows:

$$H_c(z) = (1 + z^{-1} + z^{-2} + z^{-3} + \ldots + z^{-m})^n = \left(\frac{1 - z^{-m}}{1 - z^{-1}}\right)^n . \quad (2.9)$$

Therefore, as described in Figure 2.14, a comb-filter can be realized with n accumulators, operating at f_s, and n differentiators operating at f_s/m.

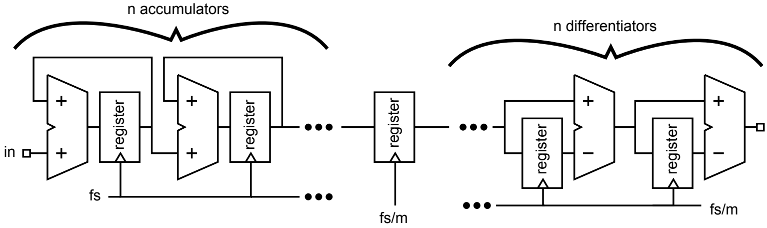


Figure 2.14 First-stage decimation filter made of n cascaded accumulators and n differentiators.

According to [Bra91], it is sufficient to chose a filter order n equal to the order of the modulator plus one. Figure 2.15 show the first-stage decimation process applied to a second-order modulator.

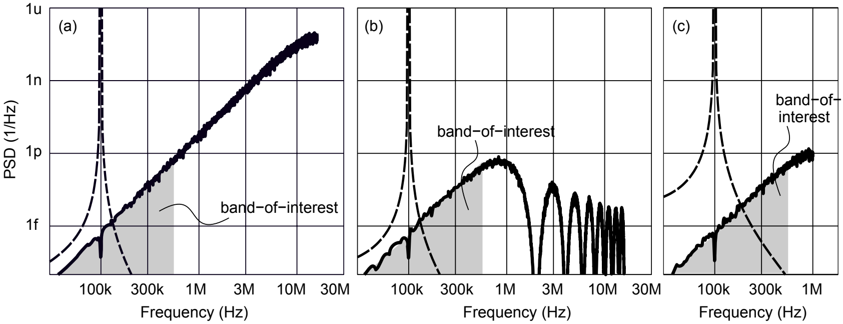


Figure 2.15 First-stage decimation: (a) 33-level second-order modulator output sampled at 32MHz and comb filter output before (b) and after down-sampling (c) at 2MHz. A band of 500KHz is considered.

Code 2.8 provides the necessary additional procedure for this experiment. Figure 2.15(b) highlights the comb filter shaping with its notches at multiples of 2MHz. Figure 2.15(c) shows the PSD after down-sampling at 2MHz. The spectrum around the notches folds back to DC. Nevertheless, the attenuation around the notches prevents the noise floor increasing.

Code 2.8 MATLAB code for the first-stage decimation filter.

```
x=out_mod;
out_comb=zeros(k,1);
acc1=zeros(k,1);acc2=zeros(k,1);acc3=zeros(k,1);
dif1=zeros(k,1);dif2=zeros(k,1);dif3=zeros(k,1);
out=zeros(k,1);
for i=17:k
    acc1(i)=acc1(i-1)+x(i-1);
    acc2(i)=acc2(i-1)+acc1(i-1);
    acc3(i)=acc3(i-1)+acc2(i-1);
    dif1(i)=acc3(i)-acc3(i-16);
    dif2(i)=dif1(i)-dif1(i-16);
    out_comb(i)=dif2(i)-dif2(i-16);
end
out_comb=out_comb/16^3;
out_comb_down=out_comb(1:16:k);
x=out_comb_down;
```

The second-stage of decimation consists of a sharper filter, usually implemented as a *half-band* filter, performing the remaining down-sampling factor of two. A half-band filter is a symmetric $(m/2+1)$-tap FIR filter, where the order m is even and all odd coefficient are chosen to be zero, except the middle one. Its transfer function $H_h(z)$ can be written as

$$\begin{array}{l} H_h(z) = \left\{ \begin{array}{l} c_1 z^{-1} + c_3 z^{-3} + \ldots + c_{m/2} z^{-m/2} + \ldots \\ c_0 z^{-0} + c_2 z^{-2} + \ldots + c_{m/2} z^{-m/2} + \ldots \end{array} \right. \\ \qquad\qquad \begin{array}{ll} \ldots + c_3 z^{-m+3} + c_1 z^{-m+1} \;, & \text{if } \frac{m}{2} \text{ is even} \;, \\ \ldots + c_2 z^{-m+2} + c_0 z^{-m} \;. & \text{if } \frac{m}{2} \text{ is odd} \;. \end{array} \end{array} \tag{2.10}$$

Depending on the application, a high-order FIR filter is required to remove unwanted signals, as well as quantization noise, that would fold back to the band-of-interest when down-sampling. A poly-phase half-band filter allows the number of operations to be drastically reduced compared to conventional FIR filters. Nevertheless, such an architecture can only perform a decimation by two.

Figure 2.16 shows the second-stage decimation process simulated with Code 2.9.

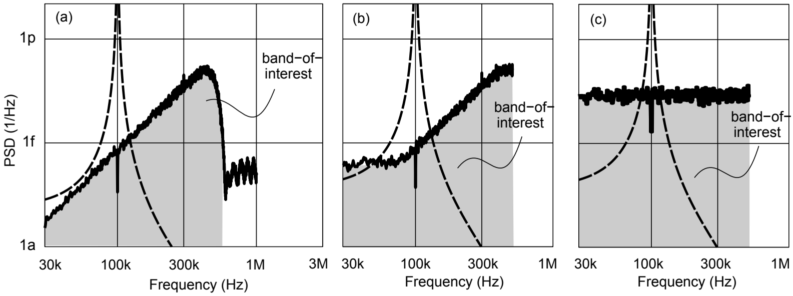


Figure 2.16 Second-stage decimation: Half-band FIR filter output before (a) and after down-sampling (b) at 1MHz. (c) PSD of a Nyquist-rate 32k-level quantizer.

Code 2.9 MATLAB code for the second-stage decimation filter.

```
x=out_comb_down;
out_fir=zeros(length(out_comb_down),1);
for i=23:length(out_comb_down)
    out_fir(i)=(...
        -8*(x(i-22)+x(i-1))...
        +14*(x(i-21)+x(i-3))...
        -26*(x(i-19)+x(i-5))...
        +48*(x(i-17)+x(i-7))...
        -96*(x(i-15)+x(i-9))...
        +315*(x(i-13)+x(i-11))...
        +500*(x(i-12)) )/1000;
end
out_fir=out_fir(200:length(out_fir)-200);
out_fir_down=out_fir(1:2:length(out_fir));
```

Figure 2.16(b) shows the filter output after down-sampling. The PSD highlights the aliasing of the spectrum around $f_s/2$ which folds back to DC. Figure 2.16(c) shows the PSD of a Nyquist-rate quantizer with a large number-of-levels. The areas under the band-of-interest in Figures 2.16 and 2.15 are all the same. The 33-level analog modulator, together with digital filters, is comparable to a 32k-level Nyquist-rate quantizer.

Another filter, operating at the Nyquist rate, is usually added to compensate for the droop of the signal transfer function inside the band-of-interest brought about by the preceding filter stages.

Since the complexity of the stages progressively increseases up as the sampling frequency decreases, this set-up is efficient in terms of area and power consumption. Moreover, filtering and down-sampling are often combined. The digital filter is part of the complete analog-to-digital conversion process. One of the interesting aspects of $\Delta\Sigma$-modulation-based converters is that they are realized with half analog circuitry and half digital hardware. The latter takes benefits from the constant progress in CMOS technologies. Its contribution to power consumption and die area therefore becomes negligible. Furthermore, the digital section is usually synthesized from a formal description, in a Hardware Description Language (HDL), so allowing jumping from one technology to the other without redesigning the whole filter.

2.2 Performance calculations

2.2.1 Linear model

In the previous section we saw that the quantization process described by Equation (2.5), under certain conditions, is equivalent to a linear gain of $1/\Delta$ and an additional random signal q, called the *quantization noise.* The quantizer can therefore be represented as in Figure 2.17.

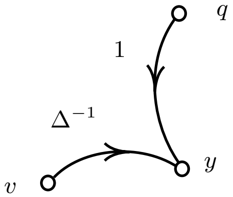


Figure 2.17 Quantization: Linear signal flow graph representation where, q is white noise uniformly distributed between $\pm 1/2$.

According to the observation made in Section 2.1.2, the probability density function of the quantization noise $p_q(x)$ has a uniform distribution bounded by ±1/2:

$$p_q(x) = \begin{cases} 1 \, , & \text{if} - 1/2 < x < +1/2 \, , \\ 0 \, , & \text{otherwise} \, . \end{cases} \tag{2.11}$$

The variance of such a distribution[1] is well-known as:

$$\sigma_q^2 = \int_{-\infty}^{+\infty} p_q(x) x^2 \mathrm{d}x = \frac{1}{12} \, . \tag{2.12}$$

Additionally, the power spectral density of the quantization noise $\mathscr{P}_q(f)$ is constant. With the hypothesis that the quantization noise is stationary and according to Rayleigh's energy theorem we can write:

$$\sigma_q^2 = \int_{-\infty}^{+\infty} \mathscr{P}_q(f) \mathrm{d}f = \mathscr{P}_q(f) f_s \, . \tag{2.13}$$

We can finally claim that the quantization noise power spectrum is of the form

$$\mathscr{P}_q(f) = \begin{cases} \dfrac{1}{12 f_s} \, , & \text{if} - f_s/2 < f < +f_s/2 \, , \\ 0 \, , & \text{otherwise} \, . \end{cases} \tag{2.14}$$

2.2.2 Modulator general description

Any single-stage $\Delta\Sigma$-modulator could be described by the linear *Signal Flow-Graph* (SFG) shown in Figure 2.18. As considered from the beginning of this chapter, nodes x and y represent the modulator analog input and digital output, whereas node v is the input of the quantizer.

We apply the quantizer linear approximation discussed previously. The quantizer is therefore modeled as a gain of Δ^{-1} with an additional white noise source q, uniformly distributed between ±1/2. By applying Mason's gain formula [MZ60], we can calculate the expression of frequency response $Y(z)$ and $V(z)$ as functions of the sources $X(z)$ and $Q(z)$:

---

[1] Many textbooks refer the quantization noise to the quantizer input. In such a case, the random variable q is uniform to within $\pm\Delta/2$ and the variance is equal to $\Delta^2/12$

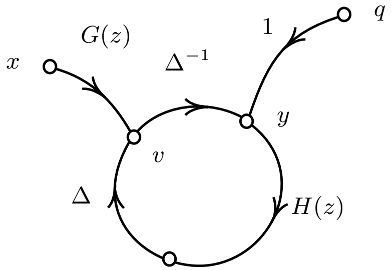


Figure 2.18 General linear Signal Flow Graph representation of the single-stage modulator: The quantizer is represented by a gain of Δ^{-1} and an additional random source q. In contrast, the DACs do not introduce errors and are simply modeled as a gain of Δ.

$$Y(z) = \underbrace{\frac{G(z)\cdot\Delta^{-1}}{1-H(z)}}_{=\mathrm{STF}(z)} X(z) + \underbrace{\frac{1}{1-H(z)}}_{=\mathrm{NTF}(z)} Q(z)\ , \tag{2.15}$$

$$V(z) = \underbrace{\frac{G(z)}{1-H(z)}}_{=\Delta\cdot\mathrm{STF}(z)} X(z) + \underbrace{\frac{H(z)\cdot\Delta}{1-H(z)}}_{=\Delta\cdot(\mathrm{NTF}(z)-1)} Q(z)\ . \tag{2.16}$$

Equation (2.15) provides the common definition of the *Signal Transfer Function* (STF) and the *Noise Transfer Function* (NTF).

2.2.3 Dynamic range

Let us consider as a hypothesis the classic nth-order architecture with a flat $\mathrm{STF}(z)$, which simply consists of an n-delay function, and an $\mathrm{NTF}(z)$ with n poles at the center of the z-plane, and n zeros z_i at frequencies f_i, that are either real or complex conjugated pairs:

$$\mathrm{STF}(z) = \Delta^{-1} z^{-n}\ , \tag{2.17}$$

$$\mathrm{NTF}(z) = \prod_{i=1}^{n}(1 - z_i z^{-1})\ , \tag{2.18}$$

where

$$z = \mathrm{e}^{j2\pi f/f_s}\ , \tag{2.19}$$

$$z_i = \mathrm{e}^{j2\pi f_i/f_s+\sigma_i}. \tag{2.20}$$

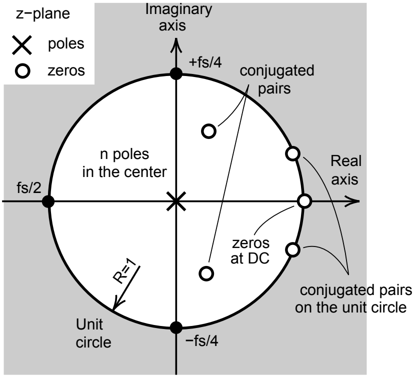


Figure 2.19 Poles and zeros: representation in the complex z-plane. Stability is guaranteed if all the poles are within the unit circle. Here all the n poles are placed at the center providing the maximum stability. The zeros appear either on the real axis or in complex conjugated pairs.

In such conditions, with the additional constraint that $\mathrm{Re}[z_i] \geq 0, \forall z_i$, referring to Equation (2.16), the highest possible value of v is given by:

$$\|y\|_\infty = \underbrace{\|\mathrm{STF}(z)\|_1}_{=k_x \Delta^{-1}} \cdot \|x\|_\infty + \underbrace{\|\mathrm{NTF}(z)\|_1}_{=k_q} \cdot \|q\|_\infty \ , \tag{2.21}$$

$$\|v\|_\infty = \underbrace{\|\mathrm{STF}(z)\|_1 \cdot \Delta}_{=k_x} \cdot \|x\|_\infty + \underbrace{\|\mathrm{NTF}(z) - 1\|_1}_{=k_q - 1} \cdot \Delta \cdot \|q\|_\infty \ . \tag{2.22}$$

Let us call the quantities k_x and k_q signal and noise *range factors* respectively. The initial hypotheses of Equation (2.17) imply that k_x is always unity. According to the definitions and proofs provided in Appendix A.1 and A.2, we can write that:

$$k_x = 1 \ , \tag{2.23}$$

$$k_q = \prod_{i=1}^{n} (1 + z_i) \ . \tag{2.24}$$

For the specific case where all the zeros are placed at DC, these relationships become

$$k_x = 1 \ , \tag{2.25}$$

$$k_q = 2^n \ . \tag{2.26}$$

Let us assume that the modulator is at the limit of overloading. The quantization errors are still uniformly distributed between ±1/2 and we can write:

$$\|q\|_\infty < \tfrac{1}{2} \qquad (2.27)$$

$$\|v\|_\infty < \tfrac{1}{2}\Delta(\mathrm{NL}) \qquad (2.28)$$

The definition of v claims that the quantizer never overloads, which implies that q never exceeds $\pm 1/2$. Once all these conditions are met, we can rewrite Equation (2.22) to optain that

$$\boxed{\|x\|_\infty < \tfrac{1}{2}\Delta\left(\mathrm{NL} - k_q + 1\right)}\ , \qquad (2.29)$$

which provides the maximum input signal the modulator can sustain without overloading, also referred to as the *full-scale* amplitude. It is interesting to note that NL cannot be smaller than $kq - 1$ and that this limitation increases with the modulator order. For this reason, the non-overload dynamic range is limited by the equation:

$$\mathrm{NL} = 2^n - 1\ . \qquad (2.30)$$

With a single comparator quantizer $\mathrm{NL} = 2$ and with any order $n > 1$ the modulator is overloaded for an $\mathrm{NTF} = (1 - z^{-1})^n$.

Figure 2.20 illustrates how a high-order modulator is limited. More room is required for quantization noise than for the signal itself. From Equation (2.21) we find the range of the digital output to be right at the limit of the maximum deliverable by the quantizer:

$$\|y\|_\infty < \tfrac{1}{2}(\mathrm{NL} + 1)\ . \qquad (2.31)$$

2.2.4 Resolution

Let us consider now an nth-order modulator, such as described by Equations (2.18) and (2.17), with all its zeros at DC, namely $z_i = 1, \forall z_i$. A full-scale tone applied at the input appears at the output with a total power

$$P_{x,y} = \frac{\|x\|_\infty^2}{2}\left|\mathrm{STF}(f)\right|^2 = \tfrac{1}{8}\left(\mathrm{NL} - 2^n + 1\right)^2\ . \qquad (2.32)$$

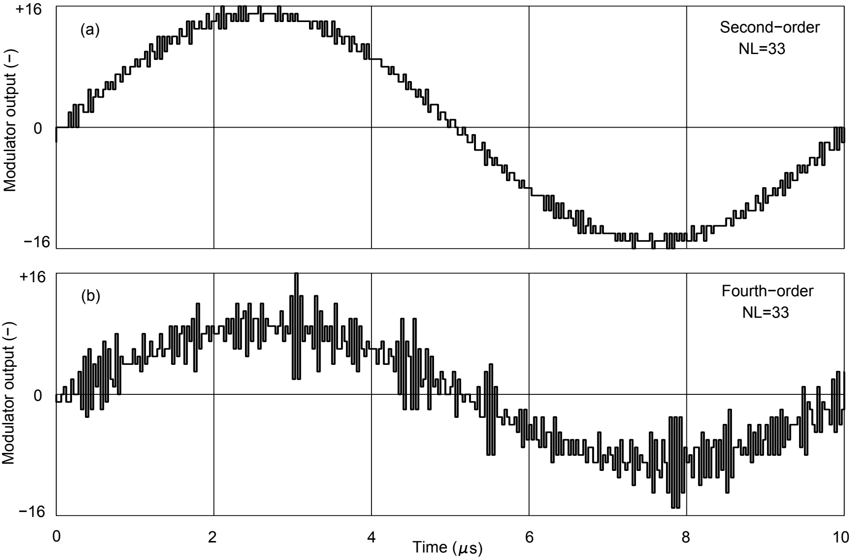


Figure 2.20 Digital output of a second (a) and a fourth-order (b) modulator with a 33-level quantizer. A clear distinction can be made between the fast quantization noise at 32MHz and the slow 100kHz-tone input.

However, the quantization noise q undergoes the NTF before reaching the output. As shown in the previous section, the modulator output passes through different digital filters and is down-sampled to $2f_b$. For this reason, supposing ideal infinitely sharp filtering, only the quantization noise integrated over that band-of-interest is to be considered:

$$P_{q,y} = \int_{-f_b}^{f_b} \mathscr{P}_Q(f)\mathrm{d}f = 2\frac{1}{12f_s}\int_0^{f_b} |\mathrm{NTF}(f)|^2\,\mathrm{d}f\ . \tag{2.33}$$

As mentioned earlier, the *Over-Sampling Ratio* (OSR) is defined as:

$$\mathrm{OSR} = \frac{f_s}{2f_b}\ . \tag{2.34}$$

Assuming the OSR $\gg 1$, namely $f_s \gg f_b$, the NTF can be simplified by removing the high-order terms in its Taylor expansion.

$$|\mathrm{NTF}(f)|^2 = \left|(1-z^{-1})^n\right|^2_{z=\mathrm{e}^{j2\pi f/f_s}} = \cdots \tag{2.35}$$
$$\cdots = \left|\left(-j\frac{2\pi f}{f_s} - \frac{2\pi^2 f^2}{f_s^2} + j\frac{4\pi^3 f^3}{3f_s^3} + \cdots\right)^n\right|^2 \cong \left(\frac{2\pi f}{f_s}\right)^{2n} .$$

The *high-over-sampling* assumption simplifies the calculation of $P_{q,y}$:

$$P_{q,y} \cong \frac{2}{12 f_s}\int_0^{f_b} (2\pi f/f_s)^{2n}\mathrm{d}f = \frac{1}{12\pi}\left(\frac{\pi}{\mathrm{OSR}}\right)^{2n+1} /(2n+1) . \tag{2.36}$$

We finally calculate the *Signal-to-Quantization Noise Ratio* and find the general relationship:

$$\boxed{\mathrm{SQNR}_{\max} = \frac{P_{x,y}}{P_{q,y}} = \frac{3}{2}\pi(\mathrm{NL}-2^n+1)^2(2n+1)\left(\frac{\mathrm{OSR}}{\pi}\right)^{2n+1}} . \tag{2.37}$$

This relationship is evaluated in Figure 2.21(a) for different cases of modulator order. As an alternative, Figure 2.21(b) shows the equation inverted so as to find the quantizer size for a targeted $\mathrm{SQNR}_{\max}$:

$$\mathrm{NL} = 2^n - 1 + \sqrt{\frac{2\,\pi\,\mathrm{SQNR}}{3\,(2n+1)}\left(\frac{\pi}{\mathrm{OSR}}\right)^{2n+1}} . \tag{2.38}$$

Similarly, the expression for the SQNR can be inverted so as to find the required OSR as a function of the other parameters:

$$\mathrm{OSR} = \pi\sqrt[2n+1]{\frac{2\,\mathrm{SQNR}}{3\pi(\mathrm{NL}-2^n+1)^2(2n+1)}} . \tag{2.39}$$

In the case of non-over-sampled quantizer alone, referred to as a *Nyquist-rate* quantizer, n=0, OSR=1 and we get:

$$\mathrm{SQNR}_{\max} = \frac{3}{2}\mathrm{NL}^2 . \tag{2.40}$$

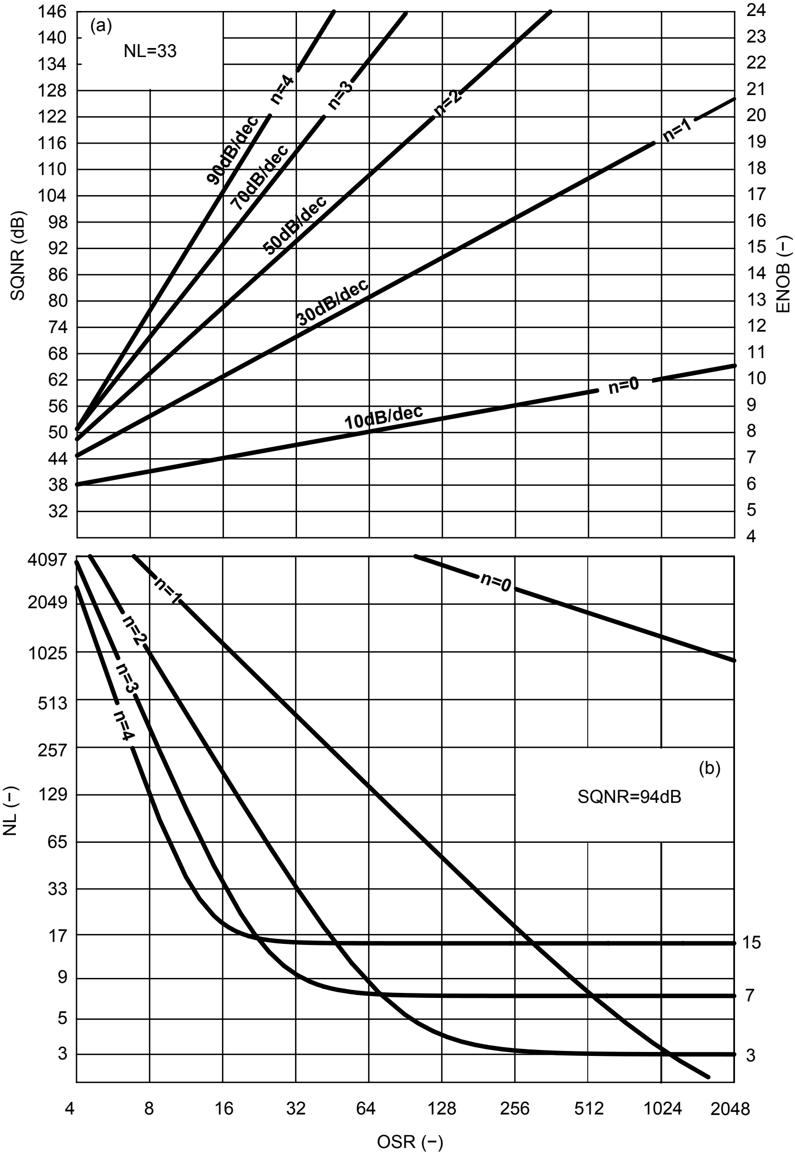


Figure 2.21 Equation (2.37): (a) achievable SQNR for a given NL=33 and (b) required NL for a given SQNR=94dB.

As mentioned in the previous section, we can define the *number-of-bits* N, also referred to as the *resolution*

$$\mathrm{N} = \frac{\log \mathrm{NL}}{\log 2}\,. \tag{2.41}$$

By combining the last two equations we find the relationship between the SQNR and an the equivalent N-bit quantizer resolution, commonly called *Effective Number-of-Bits* ENOB.

$$\boxed{\mathrm{ENOB} = \frac{\mathrm{SQNR_{dB}} - 1.76}{6.02}}\,. \tag{2.42}$$

The left and right scales in Figure 2.21(a) provide this direct relationship between the ENOB and the SQNR. In a real implementation, the distortion generated by the circuit itself and by other sources of noise degrades the SNQR. The latter becomes the *Signal-to-Noise plus Distortion Ratio* SNDR. This last relationship therefore provides a way of evaluating an equivalent resolution in terms of the number of bits.

2.2.5 Simulations

Given the power spectral density, extracted according to the methods described in Section 2.1.4, the expected resolution is determined as the ratio of the total signal and noise power within the band-of-interest. As proposed by the additional Code 2.10, the total powers are evaluated by summing the noise and extracted power spectral densities along the band.

Code 2.10 Additional MATLAB code for SNR calculation with the signal extraction method.

```
f1=0; f2=500e3;
kf1=round(f1/fs*kfft)+1;
kf2=round(f2/fs*kfft)+1;
SNR=10*log10(sum(psd_s(kf1:kf2))/sum(psd_n(kf1:kf2)))
```

Alternatively, the conventional evaluation of the SNR is given by Code 2.11.

The SNR is usually evaluated for different input signal amplitude providing the so-called *Dynamic-Range Plot* (DR-Plot). Figure 2.22 shows

Code 2.11 Additional MATLAB code for SNR calculation with direct evaluation.

```
f1=0; f2=500e3;
kf1=round(f1/fs*kfft)+1;
kf2=round(f2/fs*kfft)+1;
kf0=round(f0/fs*kfft)+1;
SNR=10*log10(sum(psd(kf0-3:kf0+3))/sum(psd([kf1:kf0-4,kf0+4:
    kf2])));
```

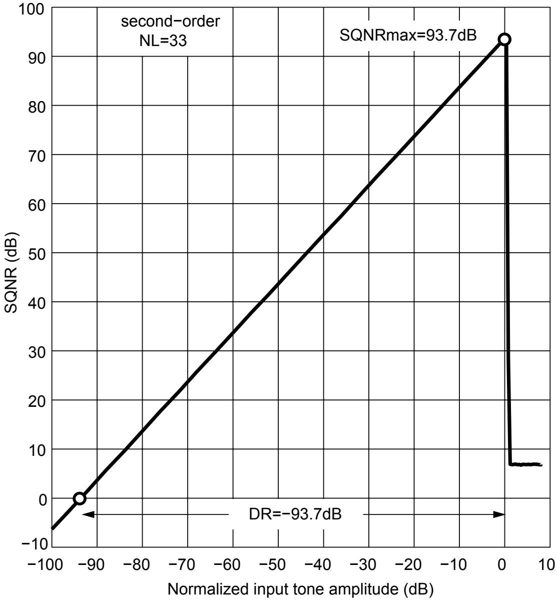


Figure 2.22 Dynamic range plot around the under- and over-load regions for a second- and third-order modulator. To optain such a smooth plot, 1000 averaging points are required and a small 0.05 dithering signal for the second-order modulator.

the DR-Plot for a 33-level second-order modulator. As expected, the SNR grows monotonically from zero up to $\mathrm{SQNR}_{\mathrm{max}}$ with a slope of 10dB per decade. For some architecture or in the presence of circuit imperfections, the DR-Plot may present a deviation from the ideal form. To make things clear, we commonly define the *Dynamic Range* as the input amplitude

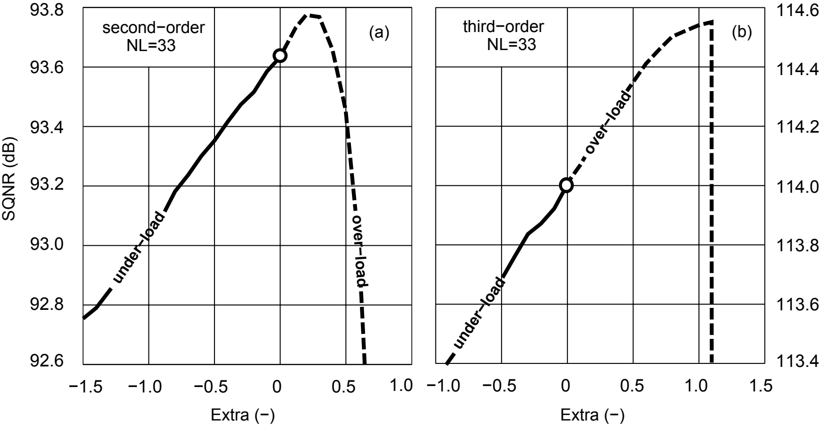


Figure 2.23 Dynamic range plot around under-over-load region for a second- and third-order modulator. To obtain such a smooth plot 1000 averaging points are required and a small 0.05 dithering signal for the second-order modulator.

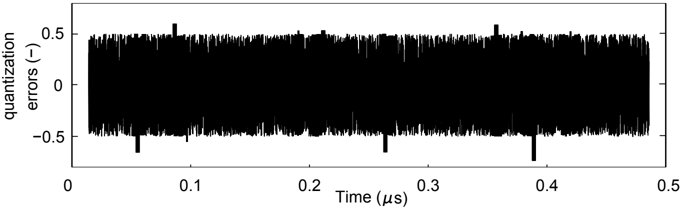


Figure 2.24 Quantization noise in the overload region for a second-order modulator with extra=0.2.

range for which the SNR goes from zero to its maximum, usually referred to as the *Peak*-SNR. In the ideal case DR and SNR are the same.

For input amplitudes higher than the range determined by Equation (2.29) the quantizer overloads. The modulator cannot be considered as a linear system anymore. The feedback loops small-signal gain is periodically deactivated and the system tends to work as a cascade of n integrators in open-loop. Since such a system is unstable, increasing the amplitude leads to further instability. For this reason, the SNR drops abruptly. Figure 2.23 highlights the transition from an *under-load* situation, through an *over-load* region, before reaching a completely unstable condition.

Figure 2.24 shows the quantization errors under over-load conditions. The errors remain within ±0.5 with periodical exceptions when the input signal is either high or low, thus generating harmonics and additional noise. The behavior within this region is hard to predict since the modulator cannot be modeled anymore as a linear system. In a conservative design we try to avoid the overload region. Nevertheless, the third-order case in Figure 2.23(b) reveals a longer overload region because of a larger k_q which can be exploited for higher-order cases.

2.2.6 Aggressive pole placement

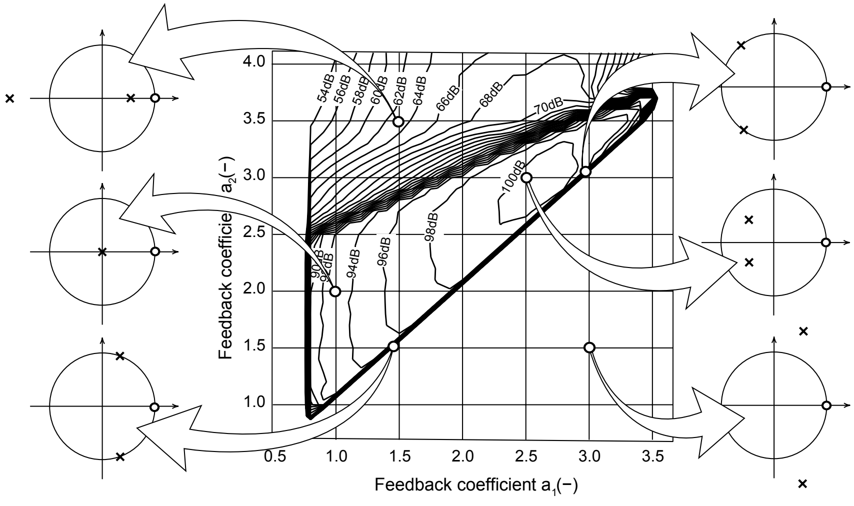


Figure 2.25 Maximum achievable SQNR of a 33-level second-order modulator with an OSR of 32 as a function of the global feedback coefficients a_1 and a_2. The resulting position of the poles and zeros for few relevant case are depicted on the sides. The experiment is performed with at input signal at 100kHz for a sampling frequency of 32MHz.

The poles are responsible [HvV99] for the rising peaks, even to infinity, of the frequency response. Any real pole, or conjugated complex pair of poles, inside the unity circle provides a damped impulse response. In contrast, the poles outside the circle provide an impulse response which grows unbounded. The position of the poles, shared by both the NTF and the STF, is determined by global feedback coefficients a_i. As already mentioned in Section 2.2.3, we chose throughout this thesis to place the poles at the center of the z-plane. This provides the system with the best robustness in terms of stability [Oga95, Lon95], the center of z-plane being the farthest point from the unity circle.

The performance of the spectral shaping in a $\Delta\Sigma$-modulator is essentially determined by the zeros. The poles may nevertheless increase the integrated quantization noise if they are placed close to the band-of-interest. Figure 2.25 show the maximum achievable SQNR as a function of the global feedback coefficients. Paradoxically, an appropriate placement of the poles allows an improvement of the SQNR by 6dB with respect to the conservative case where the poles are at the center.

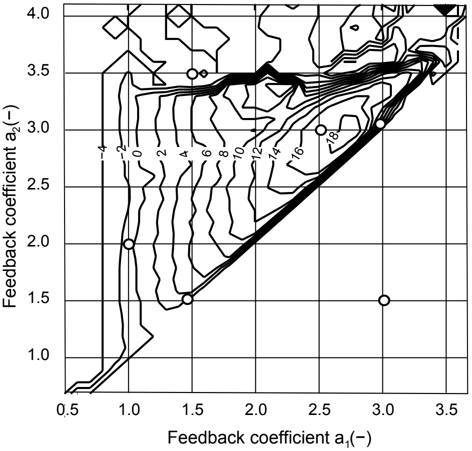


Figure 2.26 Parameter extra as a function of the global feedback coefficients a_1 and a_2 for the case of Figure 2.25.

This improvement is due to an increase in dynamic range. In this experiment, for each set of coefficients, a dynamic range plot is performed to find the maximum input amplitude, sometimes also over-loading the quantizer. Figure 2.26 depicts the parameter extra for each set of feedback coefficients. This parameter is defined as the additional number-of-levels with respect to the over-load limit for the simple case with all the poles in the center.

The Figure 2.25 shows that the maximum of 100dB is reached when $\{a_1, a_2\}$=$\{2.5,3\}$. The case with $\{1.5,3.5\}$ is unstable but provided an

SQNR of 63dB. The saturation of the quantizer limiting the unstable behavior caused the modulator to operate as a highly non-linear system.

Designing an NTF with optimal placed poles is risky and is usually referred to as *aggressive noise shaping* in contrast to a *conservative noise shaping*. Furthermore, to have a flat STF, feed-forward paths are required to compensate for the rise provided by the poles.

2.2.7 Optimal zero placement

So far we have considered the NTF of modulators with zeros placed at DC. The general case studied in Section 2.2.3 started with the hypothesis, in Equation (2.18), that zeros are either real or complex conjugated pairs. To be able to control the zeros, additional *local* feedback paths are necessary. By local feedback, we mean feedback loops not comprising the quantizer. The generic architecture, described in Figure 2.27 and implemented by the additional Code 2.12, forces the zeros to be placed on the unit circle.

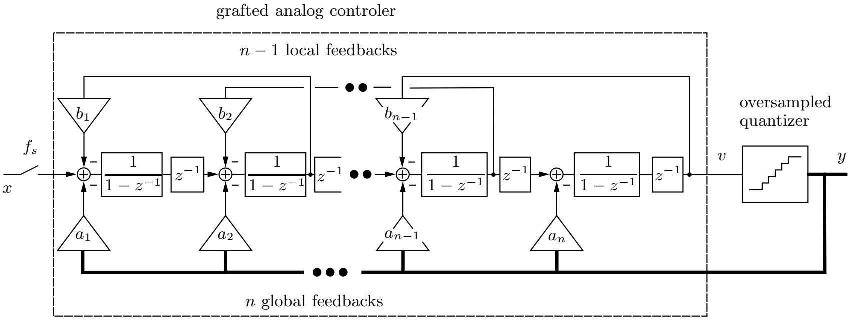


Figure 2.27 ΔΣ-modulation: Additional local feedbacks b_i controlling the zeros of the NTF.

Code 2.12 Additional MATLAB code for local feedback paths. The vectors **coeff** and **local** provide the list a_i and b_i coefficients respectively.

```
coeff=[0.940499  3.9006  5.8801  3.92];
local=[0.01 0 0.07];
...
%%%%%%% local feedbacks %%%%%%%%%%%%%%%%%%%%%
for r=1:n-1
```

```
    sig(i,r+1)=sig(i,r+1)-local(r)*sig(i,r+2);
end
%%%%%%% quantizer %%%%%%%%%%%%%%%%%%%%%%%%%%%
...
```

The zeros are responsible [HvV99] for peaks of the frequency responses falling even to zero. Any real zero, or conjugated complex pair of zeros, on the unity circle provides a notch in the frequency response. The zeros that are not on the circle provide only a light drop. Without local feedback loops, all the n zeros are automatically placed at DC.

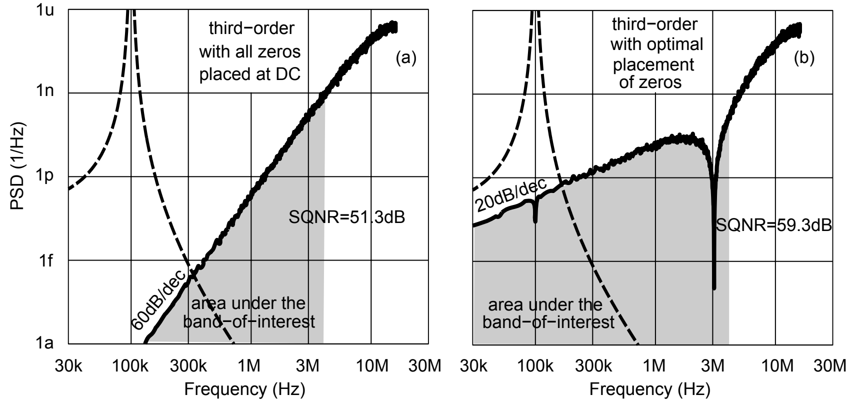


Figure 2.28 33-level third-order modulator sampling at 32 MHz with a band-of-interest from 0 to 4MHz. All the poles are placed at the center of the z-plane. In (a) all the zeros are at DC. In (b) one zero is placed at DC and a complex conjugated pair at 3MHz. The coefficient and local feedback vectors are respectively [1 2.64 2.64] and [0 0.36].

Placing zeros on the unit circle gives the NTF with deep drops at specific frequencies. According to [ST05], an optimal placement can provide important resolution improvements. Table 2.1 gives the frequency positions of the zeros, with respect to the band of interest, allowing calculation of the local coefficients b_i. The global coefficient values a_i are set in a *Pascal triangle* configuration. Because the local coefficients influence the location of the poles, correction terms are added to keep them at the center of the z-plane.

Table 2.1 Feedback coefficients to place the poles at the center of the unity circle, optimal zero placement on the unity circle according to [ST05]. The nominal values of the global coefficients a_i are in bold to highlight the *Pascal triangle* configuration.

n	$(b_1, b_2, \ldots, b_{n-1})^{\mathrm{T}}$	$(a_1, a_2, \ldots, a_n)^{\mathrm{T}}$	impr.(dB)
1	$[-]$	$\begin{bmatrix} \mathbf{1} \end{bmatrix}$	0
2	$\begin{bmatrix} 2-2\cos(\frac{0.77\pi}{\mathrm{OSR}}) \end{bmatrix}$	$\begin{bmatrix} \mathbf{1}-b_1 \\ \mathbf{2}-b_1 \end{bmatrix}$	3.5
3	$\begin{bmatrix} 0 \\ 2-2\cos(\frac{0.58\pi}{\mathrm{OSR}}) \end{bmatrix}$	$\begin{bmatrix} \mathbf{1} \\ \mathbf{3}-b_1 \\ \mathbf{3}-b_1 \end{bmatrix}$	8
4	$\begin{bmatrix} 2-2\cos(\frac{0.34\pi}{\mathrm{OSR}}) \\ 0 \\ 2-2\cos(\frac{0.86\pi}{\mathrm{OSR}}) \end{bmatrix}$	$\begin{bmatrix} \mathbf{1}-b_1^3+5b_1^2-6b_1 \\ \mathbf{4}-b_1^3+6b_1^2-10b_1 \\ \mathbf{6}+b_1^2-5b_1-b_2 \\ \mathbf{4}-b_1-b_2 \end{bmatrix}$	13
5	$\begin{bmatrix} 2-2\cos(\frac{0.54\pi}{\mathrm{OSR}}) \\ 0 \\ 2-2\cos(\frac{0.91\pi}{\mathrm{OSR}}) \\ 0 \end{bmatrix}$	$\begin{bmatrix} \mathbf{1} \\ \mathbf{5}-b_1^3+6b_1^2+b_2b_1-10b_1 \\ \mathbf{10}-b_1^3+7b_1^2+2b_2b_1-15b_1 \\ \mathbf{10}+b_1^2-6b_1-b_2 \\ \mathbf{5}-b_1-b_2 \end{bmatrix}$	18

2.3 Circuit imperfections

2.3.1 Removal of quantization noise

We usually make a clear distinction between the quantization noise and errors provided by circuit imperfections such as thermal noise and component mismatches. The former can be seen as a system limitation imposed by design. The quantization noise often hides these imperfections. The study of their impact on the resolution becomes easier if we selectively remove quantization noise from the modulator output sequence in simulations. As implemented by Code 2.13, the difference between the output and the input of the quantizer gives the quantization errors. These errors go through an imitation of the NTF before being subtracted from the modulator output.

Code 2.13 Additional MATLAB code for quantization noise removal. The **filter** function imitate a second-order with all the poles at the center of the z-plane an all the zeros at DC.

```
...
  q(i)=out(i)-sig(i,n+1); %get quantization errors
end
out=out-filter([1 -2 1],[1],q); %removal
...
```

2.3.2 Circuit noise

The simplest model of circuit noise consists of a white Gaussian variable, accounting for thermal noise, and an additional $1/f$-power law component, accounting for the so-called *flicker*-noise which is particularly important in CMOS transistors. Code 2.14 generates such a noise sequence with an IIR filter. The generation of colored noise is treated at great length in [Kas95].

Code 2.14 MATLAB code for the generation of circuit noise with white and 1/f-power law components.

```
pink=zeros(1,1000);
pink(1)=1;
for m=2:1000
   pink(m)=(m-2.5)*pink(m-1)/(m-1);
end
noise=filter(1,pink,0.1*randn(k,1))+0.5*randn(k,1)
```

Figure 2.29 shows the output of a conventional second-order modulator in the presence of noise. The quantization errors are removed. A colored noise is introduced at different nodes of the system, at the modulator input, after the first integrator and before the quantizer. The results illustrate how the circuit errors, here circuit noise, benefit from the modulator spectral shaping. As a consequence, the errors due to the imperfections of the first components in the processing chain, like the input signal, undergo the STF before reaching the modulator output. It is not possible distinguish these errors form the wanted signal. As shown in

Figure 2.29(a), the colored noise appears at the output without attenuation or spectral shaping. Hence these errors usually limit the resolution of the modulator.

In contrast, the errors introduced after the kth integrator are spectrally shaped with a kth-order high-pass transfer function. In particular, as revealed in Figure 2.29(c), the errors occuring at the input of the quantizer, like the quantization noise, undergo the NTF.

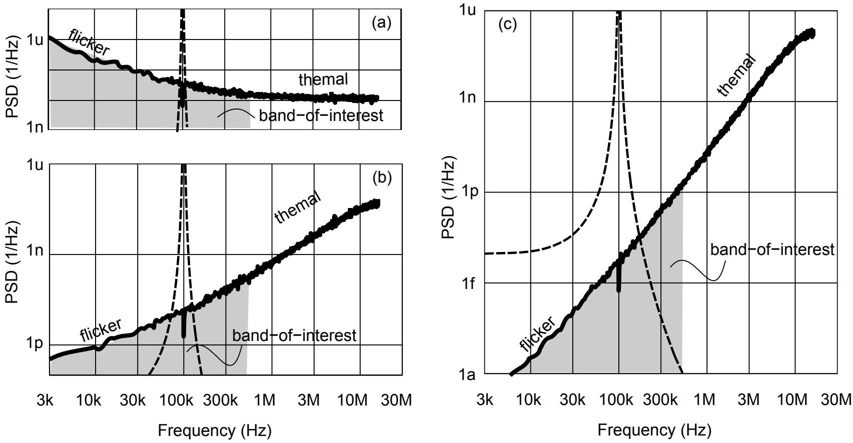


Figure 2.29 Second-order 33-level modulator with circuit noise injected at the input (a), after the first integrator (b) and before the quantizer (c). The quantization noise is removed.

2.3.3 Clock jitter

The modification proposed in Code 2.15 introduces the effect of an imperfect sampling clock signal. The sampling period is now randomly distributed around $1/f_s$ using a Gaussian variable whose spectrum is white.

Code 2.15 Modification of the MATLAB code of the signal generation to introduce the clock jitter.

```
x=x0*(sin(2*pi*f0*(0.1*randn(k,1)+(1:k)')/fs));
```

This imperfection is called *clock jitter*. Figure 2.30 show how the presence of a high-amplitude high-frequency turns the clock jitter into a noise

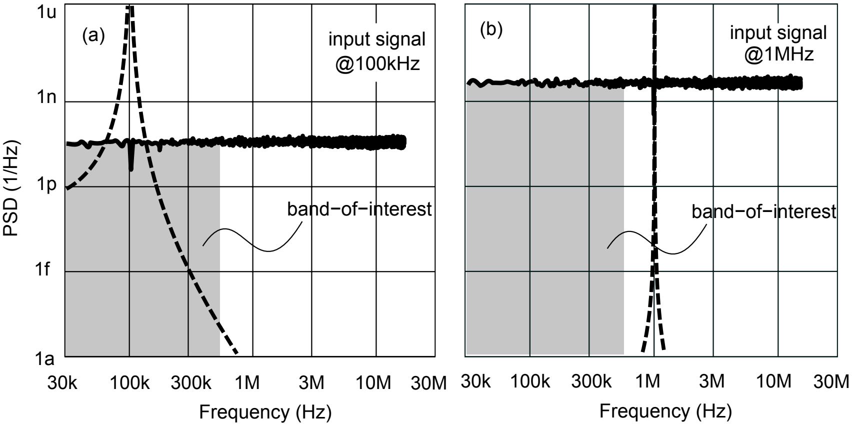


Figure 2.30 Output power spectral density of a second-order 33-level modulator, whose sampling is affected by clock jitter, with a full-scale input tone at 100kHz (a) and 1MHz (b). The quantization noise is removed.

signal. The clock jitter is analyzed in detail in the next chapter, showing the different transfer mechanisms for the continuous- and discrete-time modulators.

2.3.4 Component mismatch

A multi-bit $\Delta\Sigma$-modulator requires an ADC and, for each node fed back from the digital output, a DAC. In a conventional implementation, these blocks are realized as a bank of NL comparators for the ADC, and as a bank of NL one-weight elements for the DAC. Among the major benefits of integrated circuits, we find that identical components close to each other are well matched. Nevertheless, the matching properties are highly dependent on the size of the components. But increasing their size also increases the distance between the components therefore reducing the matching properties. As a result, the matching properties are limited by the technology.

The mismatches between the components of the ADC and DAC affect their input-output transfer characteristic which is, in an ideal situation, a perfect staircase function. The DAC element mismatch changes the height of levels and the comparators offset changes the thresholds. These changes in the characteristic cause harmonic distortion. Figure 2.31 shows the modulator output in the presence of mismatch. The quantization noise is completely removed. Since the harmonic distortion occurs in the $\Delta\Sigma$-modulator loop, the harmonic tones generated also fold parts of the quantization noise.

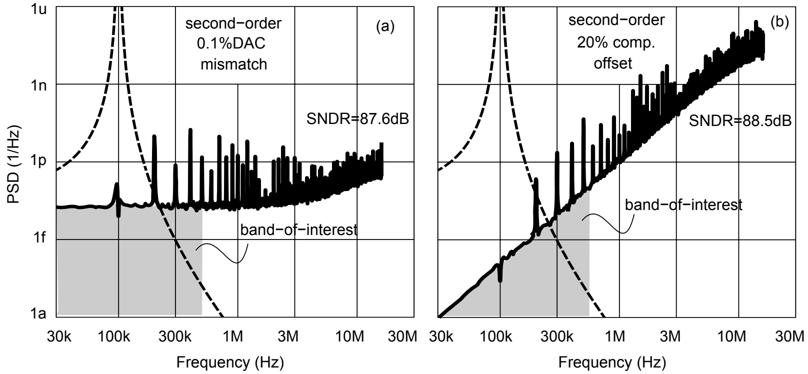


Figure 2.31 Output power spectral density of a second-order 33-level modulator with DAC elements mismatch (a) and ADC comparator offset (b).

As highlighted in Figure 2.31, the errors brought about by the ADC are shaped contrary to the errors of the first DAC. A similar degradation of SNR is produced with 0.1% and 20% of mismatch, in the DAC and ADC respectively. This illustrates once again how the spectral shaping of the modulator itself allows relaxing the circuit imperfection of components placed at the end of the signal processing chain, like the quantizer in this case.

2.3.5 Harmonic distortion

A non-linear behavior in the system gives rise to harmonic distortion. As a didactic model, let us consider the simple function $f(x) = x + kx^3$ and a single-tone input signal. When this function is applied directly at the input of the modulator, an additional sinewave at three times the tone frequency appears with an amplitude of $3k/4$. As highlighted in Figure 2.32(a), the additional sinewave is an unwanted signal and brings about a degradation of the SNR. As already mentionned earlier, in such a case, we use the *Signal-to-Noise plus Distortion Ratio* (SNDR) to determine the resolution. In order to take into account the distortion in simulations, the test tone needs to be set at a rather low frequency with respect to the band-of-interest. Consequently, the number of clock cycles to simulate can be quite large. This might be impractical in complex circuit level simulations.

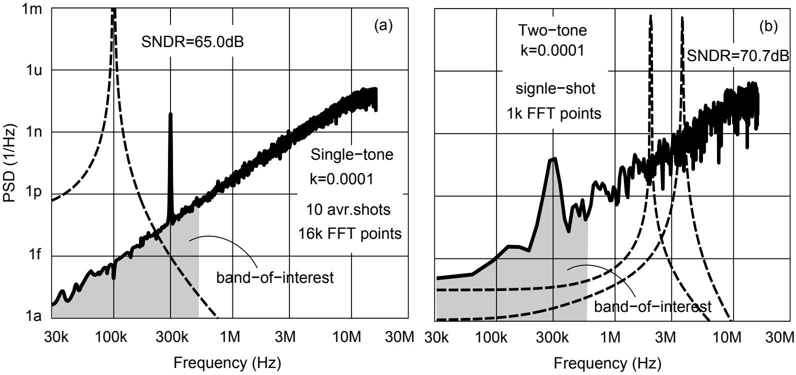


Figure 2.32 Simulation of distorsion in a second-order 33-level modulator with a full-scale low-frequency single-tone (a) and two half-scale high-frequency tones (b). A 500kHz bandwidth is considered.

Alternatively, the two-tone test allows short simulations. Figure 2.32 shows the experimental setup with two half-scale out-of-band tones, set at frequencies f_1 and f_2. In such a case, different harmonics appears at high frequencies $2f_2 + f1$, $2f_1 + f2$, $f_1 - 2f_2$, $3f_1$ and $3f_2$. Since these tones remain out of the band-of-interest, their presence is usually hidden by quantization noise. On the other hand, an unwanted tone is generated at $f_2 - 2f_1$ which falls into the band-of-interest. The tone has an amplitude of $3k/32$, namely hight times lower than the harmonics found with the single-tone test. As a result, the measured degradation is 6dB lower with the two-tone test.

2.4 Architecture classification

2.4.1 Single- and multi-bit

Many designs have been proposed with a number-of-levels brought to the extreme limit of two, while keeping a high resolution by increasing both the order and the over-sampling. Such an architecture is commonly referred to as a *single-bit* modulator in contrast to the *multi-bit* general case treated so far. Because of its simplicity, this architecture has been very popular over the last decades. In a large part of the didactic literature, such as [AH02, JM97, Par93, PVS96, Hay01], the explanations are entirely built on the single-bit first-order topology, considering the multi-bit feedback as a rather particular case.

At first glance, the implementation is drastically simplified since only one comparator and one DAC element are required. Moreover, the mismatch and offset issues disappear. As a drawback, both the circuit speed and the analog signal processing are increased. Furthermore, as shown in Figure 2.33, the signals processed by the modulator constantly jump from one level to another at high speed, putting an important slew-rate constraint on the amplifiers.

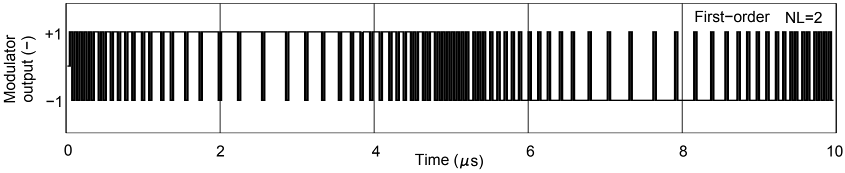


Figure 2.33 First-order single-bit modulator output.

Figure 2.34 highlights an important difference between the single- and multi-bit cases. As predicted by Equation (2.29), if NL goes below $k_q - 1$ the quantizer overloads and the linear model does not hold anymore. The NTF deviates from the predicted shape. A flat region with spurs appears at high frequencies with odd harmonic distortion. In contrast to the multi-bit architecture, the DR-Plot in Figure 2.35 reveals a DR higher than the peak SNR. The slope is not perfectly constant and the curve smoothly drops down at an amplitude which is hard to predict.

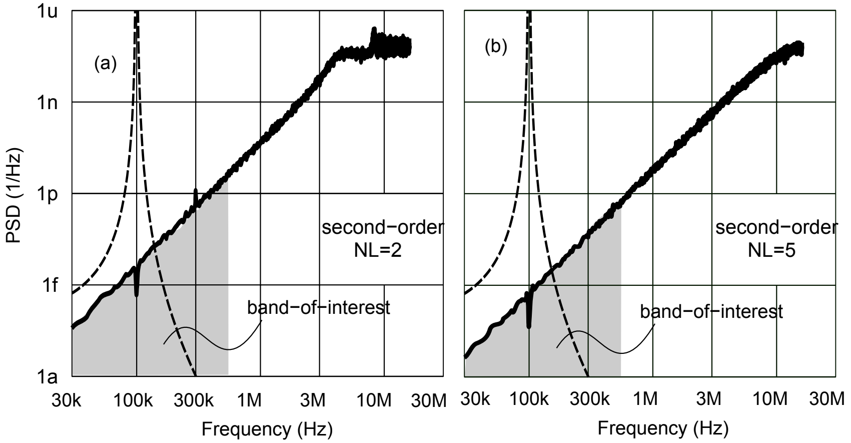


Figure 2.34 Second-order shaping power spectral density with 2-level (a) and 5-level (b) quantizer.

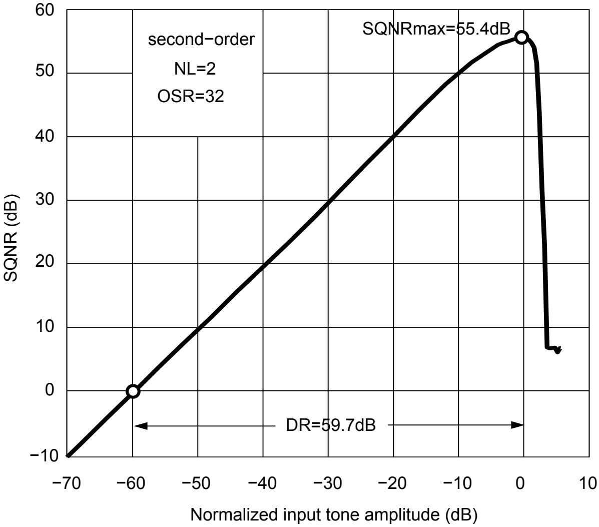


Figure 2.35 DR-Plot of a single-bit second-order modulator.

Important work has been carried out to understand the behavior of such highly non-linear systems. In particular, the contribution of [vEvdP99] brought an interesting modeling of the single-bit quantization along with prediction of unstable conditions allowing robust design. [**?**] provided an extensive analysis of the switched-capacitor implementation and [Bal03] studied the three-level case.

2.4.2 Low-, high- and band-pass

So far we have considered a band of interest from DC to a given frequency. However, in some applications it may turn out that the bandwidth has to start and end at f_1 and f_2 respectively. Considering $f_2 > f_1$, the band-of-interest is the difference $f_b = f_2 - f_1$. If either f_1 or f_2 is set to zero, the modulator is said to be *low-pass* or *high-pass* respectively, the intermediate case being called *band-pass.* Figure 2.36 shows fourth-order examples of the three cases.

2.4.3 Single and multi-feedback path

So far we have considered an architecture with a feedback on each integrator. We should recall that each feedback coming from the modulator output requires its own DAC. In contrast to this so-called *multi-feedback* topology, the *single-feedback* architecture, described in Figure 2.37(b), simplifies the implementation to the limit of only one path from the digital modulator output. The same NTF can be realized by introducing the *interpolation* paths d_i. Unfortunately, in this configuration the *feed-forward* paths c_i are necessary to avoid important ripples in the STF. Moreover, the addition before the quantizer requires another amplifier. The feed-forward path can also be used in a multi-feedback configuration to control the STF independently from the NTF. A combination of both approaches is also possible.

Many designs are proposed with a single-bit single-feedback topology. Their simplicity allows focusing on fewer constraints since there is only one comparator and one DAC element in the whole modulator.

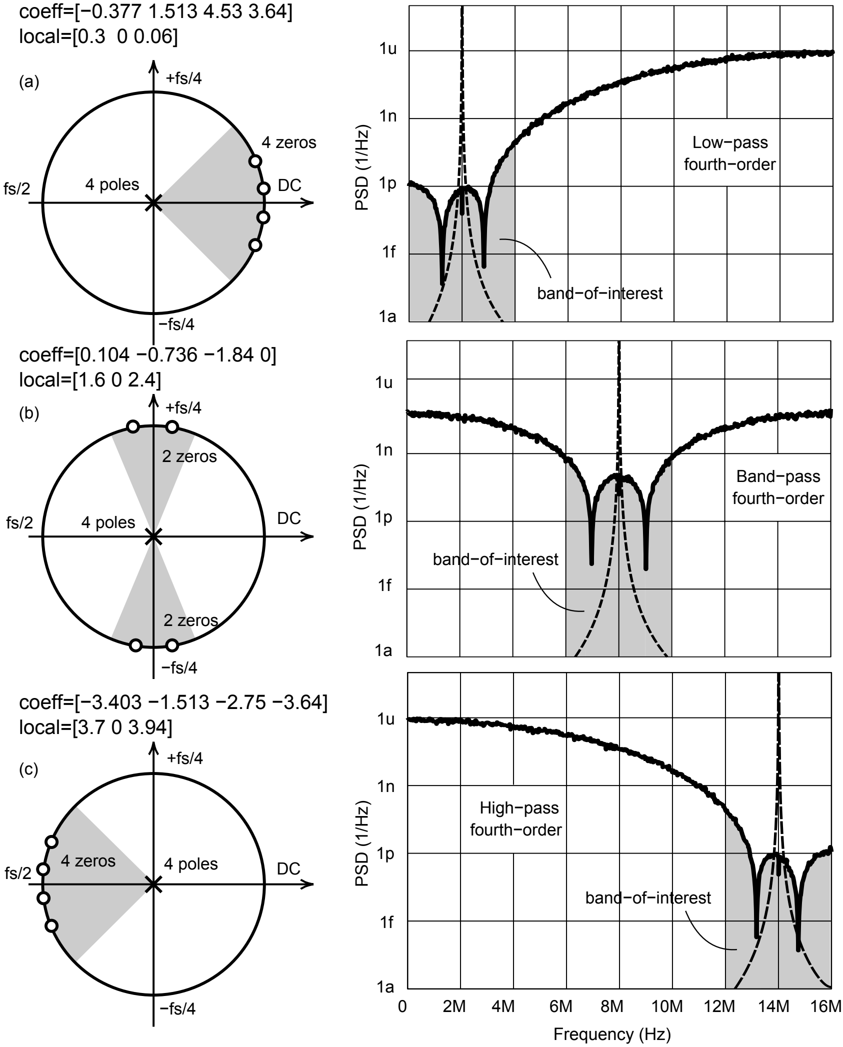


Figure 2.36 Output power spectral density for a fourth-order, low-pass (a), band-pass (b) and high-pass (c) modulator. The frequency axis is linear to clearly present all three cases. The simulation coefficients are provided on the left together with the z-plane representation of the resulting poles and zeros.

In most of the designs published, like [BH01], the circuit consists of a continuous-time RC integrator followed by gm-C cells ending with a comparator. Nevertheless, these choices force the modulator to work at high sampling frequencies and with high-order shaping.

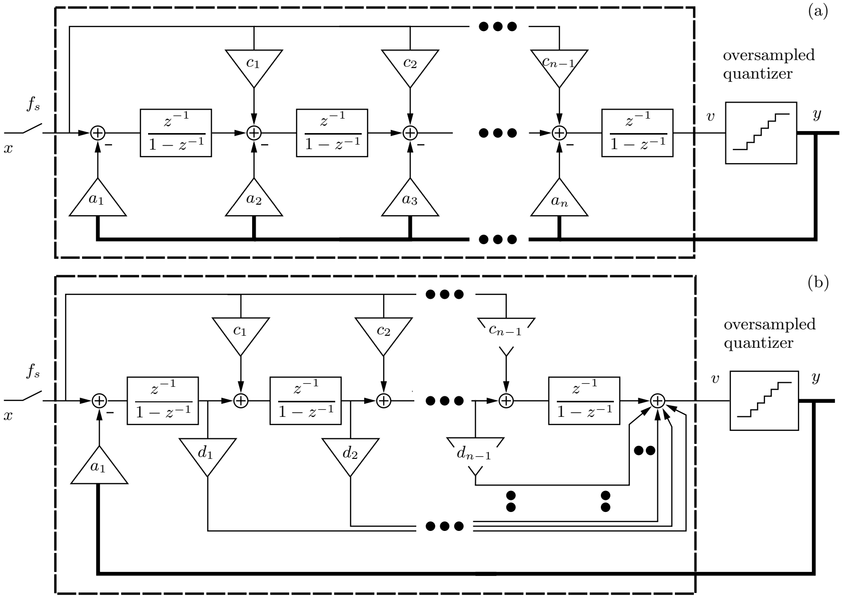


Figure 2.37 General (a) multi-feedback (b) single-feedback and configurations.

2.4.4 Single and multi-stage

Because of the stability issues mentioned previously, building high-order modulators with a single-bit quantizer calls for a careful design. Alternatively, a cascade of low-order modulators can give the performance of a single high-order modulator without the stability issue. Each modulator is referred to as a *stage* and such a topology is commonly called *multi-stage* or *MASH* architecture. In the general case described in Figure 2.38, low-pass $\Delta\Sigma$-modulators are considered with the following characteristics:

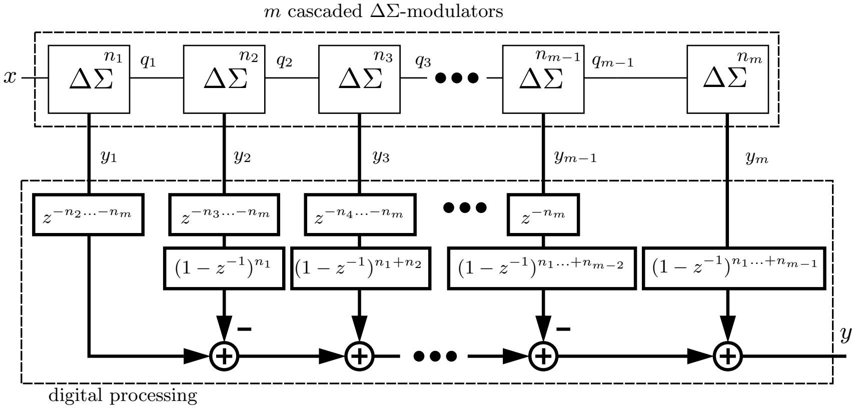


Figure 2.38 m-stage architecture: Each stage consists of a low-pass n_ith-order $\Delta\Sigma$-modulator with all poles at the center and zeros at DC. Only the quantization noise q_i is transferred from one stage to the other.

$$\mathrm{STF}_i(z) = z^{-n_i} \,, \tag{2.43}$$

$$\mathrm{NTF}_i(z) = (1 - z^{-1})^{n_i} \,. \tag{2.44}$$

Each stage feeds its own generated quantization noise q_i to the next stage. Each q_i is obtained by sampling the quantizers input and subtracting the their output. A post-processing combines the m digital outputs y_i to cancel out the quantization noise of the first $m-1$ stages. Provided that the digital and analog processing match perfectly, the processed output y is the same as a*single-stage* modulator with an order equal to the sum of the individual stage orders:

$$Y(z) = X(z)z^{-n_e} + Q_m(z)(1 - z^{-1})^{n_e} \,, \tag{2.45}$$

$$n_e = n_1 + n_2 + \ldots + n_{m-1} + n_m \,. \tag{2.46}$$

An architecture with, for example, a second-order and two first-order modulators is referred to as a *2-1-1* MASH modulator. A good matching of the analog and digital processing requires almost exclusively an SC implementation.

2.5 Conclusions

We have introduced the well-known linear model of the quantization process in a didactic way through computer experiments. Different aspects of spectral estimation issues were addressed. Over-sampling and spectral shaping were presented as a means of increasing the resolution of a multi-bit quantizer. In contrast to other authors in the field, the single-bit modulator is considered here as a particular case of the multi-bit $\Delta\Sigma$-modulation, operating as a non-linear system. The analytical expression of the resolution, often given in a less accurate form in the literature, was derived and exploited to guide a preliminary design. The chapter ended with a brief review of the different architectures known today.

Chapter

3

Low-power strategy

The first section of this chapter briefly introduces a design strategy to lower the power consumption of single-stage modulators. The strategy consists in combining different techniques, each of them addressing a specific issue. The designer is naturally led to multi-bit hybrid continuous-discrete-time architectures. The subsequent sections analyze in detail particular aspects of the techniques used, highlighting the advantages and drawbacks. In this way, the chapter gives designers the tools to make the appropriate trade-offs. For the sake of readability, two important subjects of the strategy, the auto-ranging technique and the optimization of the DEM, are covered in dedicated chapters.

3.1 Strategy outline

3.1.1 Continuous-time implementations

Many recent publications [Kap03, YS04, AL02, Abo02, PNR$^+$04, GM01], books [CS00, BH01, KvR06, OG06] and thesis dissertations [Sho95, Yan02] have demonstrated the low power consumption of continuous-time $\Delta\Sigma$-modulators. As a matter of fact, continuous-time implementations require amplifiers with less current and bandwidth than their discrete-time counterparts. While a continuous-time filter operates smoothly without interruption, in a discrete-time circuit half a clock period is used to sample the input and the remaining half to perform the integration. The integration usually requires a charge transfer determined at each clock cycle by a decaying exponential which demands a high initial slewing current.

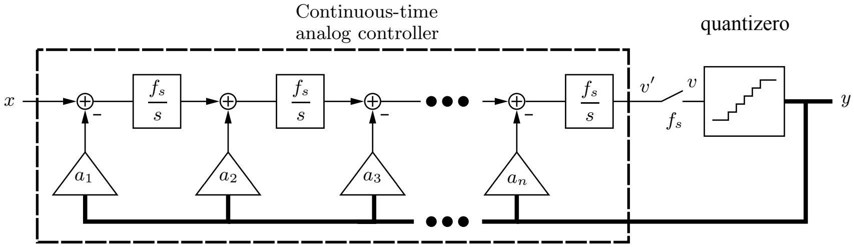


Figure 3.1 Continuous-time implementation of a multi-feedback single-loop.

Figure 3.1 shows a single-loop multi-feedback modulator implemented with a continuous-time analog control system. The block diagram shows that the position of the sampling process, with respect to a discrete-time implementation, has moved from the modulator input x to the quantizer input v, after the nth-order integration chain. As shown further on, this inherently provides anti-alias filtering because the STF is no longer the result of the compensation of the NTF by the cascade of discrete-time integrators.

Despite the benefit of anti-aliasing and low-power processing, two important issues arise, an increased sensitivity to clock jitter, as highlighted in [Oli01, Oli99, CS99] and an inaccurate control of the feedback coefficients. The next section presents a detailed analysis of these two aspects. In particular, it is shown under which conditions the multi-bit feedback path can significantly reduce the jitter sensitivity. Next, it is shown that the tolerance on the last feedback coefficients is more stringent than on the first one.

3.1.2 Multi-bit feedback

Table 3.1 gives the required over-sampling ratios for a targeted SQNR, using the analytical model developed in Chapter 2. The NL-OSR couples are to be seen as solutions for a chosen modulator order. We find the large-NL-low-OSR solutions on the right side and the small-NL-high-OSR solutions on the left.

Often designers choose the high-order single-bit solutions with a moderate OSR. While the single-bit solutions are attractive because of their simplicity, their stability can easily be compromised. Furthermore, the

Table 3.1 Required OSR for an NL-level nth-order modulator with SQNR=94dB according to Equation (2.39).

	Number-of-levels NL								
n	3	5	9	17	33	65	129	257	513
1	1113	701	442	278	175	110	70	44	28
2	-	96	62	44	32	24	18	14	10
3	-	-	34	22	17	13	11	9	7
4	-	-	-	20	12	10	8	7	6
5	-	-	-	-	14	8	7	6	5

next section shows that continuous-time modulators are more sensitivity to clock jitter for high shaping order. It is also shown that among the possible solutions of Table 3.1, for a given order, an optimal case with moderate sets of NL/OSR provides the lowest jitter sensitivity.

Besides the aspect of stability and sensitivity to jitter, increasing NL while decreasing n and OSR relaxes the specifications of all the analog circuitry. In fact, the OSR is related to the sampling frequency and therefore to the amplifier bandwidth and slewing capability, whereas the order roughly corresponds to the number of amplifiers. To some extent, using a large number-of-levels significantly decreases the signal voltage steps, further alleviating the slewing specifications of the amplifiers.

3.1.3 Optimal DEM segmentation

The DAC is generally implemented as a bank of $\mathrm{NL}-1$ one-bit elements. As already mentioned at the end of Chapter 2, the mismatch between the elements introduces harmonic distortion. A digital Dynamic Element Matching (DEM) algorithm is therefore necessary. The hardware complexity of the DEM grows in proportion to NL. Furthermore, the low OSR, brought about by the increase of NL, reduces the efficiency of the mismatch shaping algorithm, forcing the use of higher-order algorithms, which further increase the size and consumption of the DEM.

Among the different DEM algorithms summarized in [GS02], the tree structured architecture proposed by [FSW^{+}02] is extendable to any shaping order. This architecture takes advantage of extra-LSB coding to re-

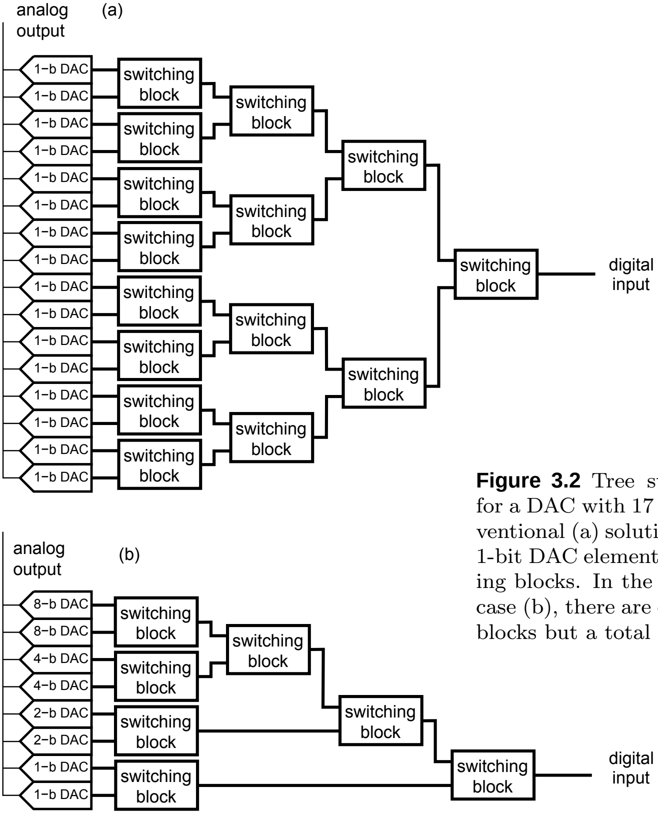


Figure 3.2 Tree structured DEM for a DAC with 17 levels. The conventional (a) solution consists of 16 1-bit DAC elements and 15 switching blocks. In the fully segmented case (b), there are only 7 switching blocks but a total of 31 elements.

duce the logic circuitry provided that NL is a power-of-two-plus-one. As illustrated by Figure 3.2(a), a conventional tree consists of $\mathrm{NL}-1$ one-bit DAC elements and $\mathrm{NL}-2$ switching blocks. Each switching block contains its own mismatch shaper controlling a switching network.

A segmentation of DAC into elements with different weights, proposed by [FSW$^+$02] allows reducing the number of switching blocks. Consequently, for a fully segmented solution, as in the example in Figure 3.2(b), $2\log_2(\mathrm{NL}-1)-1$ switching blocks are needed. At the same time, the total number of 1-bit elements is then $2\mathrm{NL}-3$. In other words, segmentation drastically reduces the digital circuitry but at the same time increases the total number of 1-bit DAC elements. Chapter 5 covers this subject in detail and reveals the existence of a partial segmentation solution that is optimal in terms of power consumption.

3.1.4 Auto-ranging algorithms

The most limiting factor for large-NL-low-OSR solutions remains the number of comparators. In a DAC bank, each one-bit element uses the full range of the reference voltage. In contrast, in a quantizer the reference voltage splits into NL − 1 quantization steps. With a large NL, the quantization steps become too small to be resolved by a bank of comparators strongly handicapped by a statistical offset. A large NL may simply not be achievable, especially with low voltage supplies.

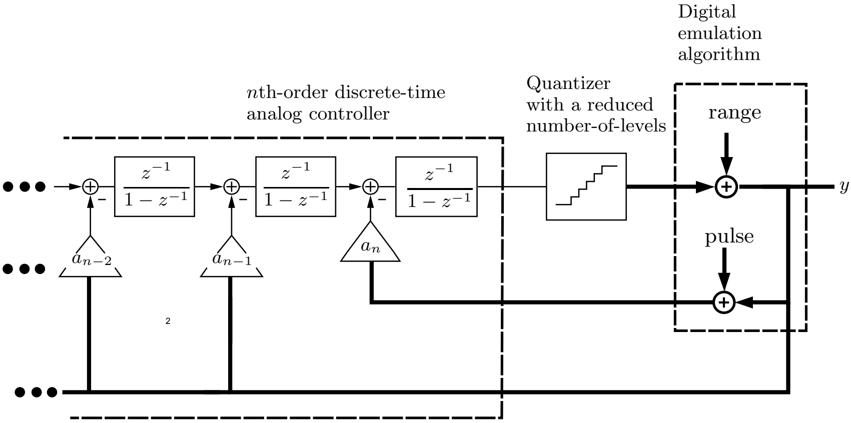

Digital emulation algorithm
nth-order discrete-time analog controller
Quantizer with a reduced number-of-levels
range
pulse
y
$\frac{z^{-1}}{1-z^{-1}}$
$\frac{z^{-1}}{1-z^{-1}}$
$\frac{z^{-1}}{1-z^{-1}}$
a_{n-2}
a_{n-1}
a_n
2

Figure 3.3 Digital algorithm to emulate a large NL with a reduced quantizer number-of-levels.

Moreover, the power consumption and chip area increase significantly with the number of comparators. At the same time, the bank of comparators provides the last integrator with a rather large capacitive load, which further raises the consumption of the modulator.

Nevertheless, the number of comparators can be reduced while keeping the NL unchanged. Different methods have been presented by [Zie00, Lu04, DKG^{+}05] to allow the emulation of NL with a small number of comparators. A novel technique is proposed here which consists of shifting both the input and output of the reduced size quantizer.

As illustrated in Figure 3.3, the shift of the digital output is provided by a *range* signal, whereas the analog shift of the input is provided by a digital *pulse* integrated by the last feedback path. The reuse of an existing path circumvents the need for additional analog circuitry. At the same time, the quantization steps can be enlarged to take advantage of the full voltage swing of the last integrator amplifier.

Chapter 4 treats in detail the implementation aspects, the optimization and sensitivity to imperfections. In particular, it is shown that the accuracy of the reused feedback path is crucial.

3.1.5 Hybrid architecture

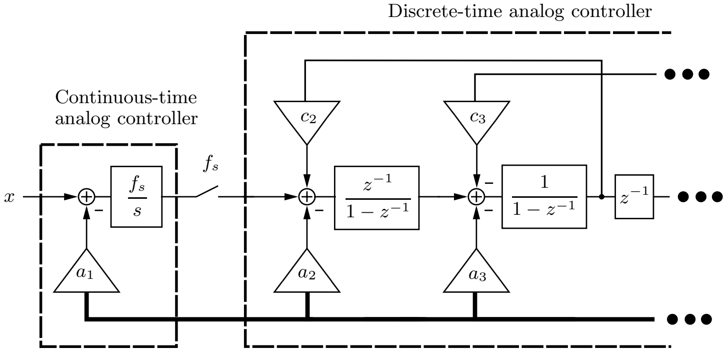

Discrete-time analog controller
Continuous-time analog controller
x
$\frac{f_s}{s}$
f_s
c_2
c_3
$\frac{z^{-1}}{1-z^{-1}}$
$\frac{1}{1-z^{-1}}$
z^{-1}
a_1
a_2
a_3
−
●●●

Figure 3.4 Multi-bit hybrid continuous-discrete-time $\Delta\Sigma$-modulator.

These considerations naturally lead to a multi-bit hybrid architecture such as described by Figure 3.4. In the proposed architecture only the first integrator is continuous-time. As demonstrated at the end of Chapter 2, the first integrator imperfections do not benefit from any spectral shaping. Hence, the modulator power consumption is often dominated by the first amplifier. Turning this stage into a continuous-time version has a large impact on the overall consumption. Keeping the rest of the modulator in the discrete-time domain allows a more aggressive spectral shaping, with local feedback paths, since switched-capacitor circuits can rely on accurate coefficients depending on the matching of capacitors. In turn, this allows a reduction of the over-sampling ratio for a given performance. An increase in the number of quantization levels allows the jitter issue brought about by the continuous-time first-stage to be circumvented. At the same time, more levels bring a reduction of power dissipation in the discrete-time controller since the step size is small with respect to the single-bit case.

3.1.6 Continuous-to-discrete domain interface

In the case of a pure discrete-time $\Delta\Sigma$-modulator, the continuous-to-discrete-time interface is located right at the input. The amplifier of the sampling device does not take advantage of any spectral shaping. Therefore, in order to meet the linearity requirement, its specifications are stringent, resulting in a large power dissipation.

In contrast, in the case of a hybrid implementation, the sampling process takes place after at least one integration stage, depending on the architecture chosen. Consequently, at least a first-order shaping mechanism can be exploited.

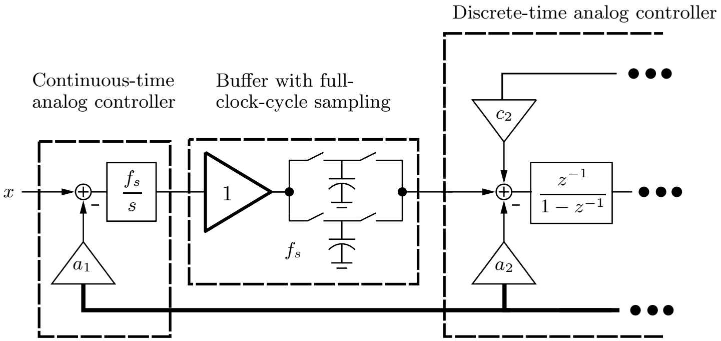

Discrete-time analog controller
Continuous-time analog controller
Buffer with full-clock-cycle sampling
x
$\frac{f_s}{s}$
-
1
f_s
a_1
c_2
$\frac{z^{-1}}{1-z^{-1}}$
-
a_2
●●●
●●●
●●●

Figure 3.5 Continuous-to-discrete-time transition with an isolation buffer and a full-clock-cycle sampling scheme.

As shown in Figure 3.5, the process of sampling requires a unity gain buffer to guarantee a proper operation of the continuous-time integrator. Without this buffer, the output of the continuous-time integrator is periodically connected to the sampling capacitors, formerly emptied during the integration phase.

The buffer provides the continuous-time integrator with a small and constant capacitive load. On the other hand, the buffer has the task of filling the sampling capacitor at each clock cycle. As described conceptually in Figure 3.5, a full-clock-cycle sampling scheme can be used to alleviate the sampler specification. As shown further in this chapter, thanks to spectral shaping the errors introduced by the capacitors mismatch have a limited impact on the resolution.

3.1.7 Summary

Figure 3.6 summarizes the low-power strategy proposed in this chapter. Each technique used brings, directly or indirectly, a reduction of the dissipated power. At the same time, each technique introduces a drawback while addressing another one [PCSK07]. In reality, the strategy starts with the task of converting an analog signal into a digital sequence with a targeted resolution and bandwidth. The choice of a $\Delta\Sigma$-modulation-based converter allows the use of an over-sampled single-comparator quantizer while providing about 12’000 level. The price is an important increase in sampling frequency. The overall consumption is then decreased by relying on a continuous-time implementation. The sensitivity to clock jitter and the overall modulator consumption are further reduced by finding a quantizer number-of-levels between the two extremes of 12000 and 2. The lower sensitivity to clock jitter has an indirect impact on the circuit consumption, since in a large System-on-Chip (SoC) the clock signal is provided by a PLL, generally not considered in converter power consumption summary. Then, two adapted digital solutions, a segmented DEM tree and an auto-ranging algorithm, are used to circumvent the drawbacks of a multi-bit implementation. Because the auto-ranging requires an accurate last feedback path, the modulator upper stages must be realized with discrete-time integrators. Such a multi-bit hybrid architecture allows one to take advantage of the benefits of the discrete- and continuous-time. The continuous-to-discrete-time interface has therefore moved inside the modulator loop to take advantage of spectral shaping.

Analog-to-digital conversion consists in combining sampling and quantization. In a multi-bit hybrid $\Delta\Sigma$ADC, the two processes are placed in-

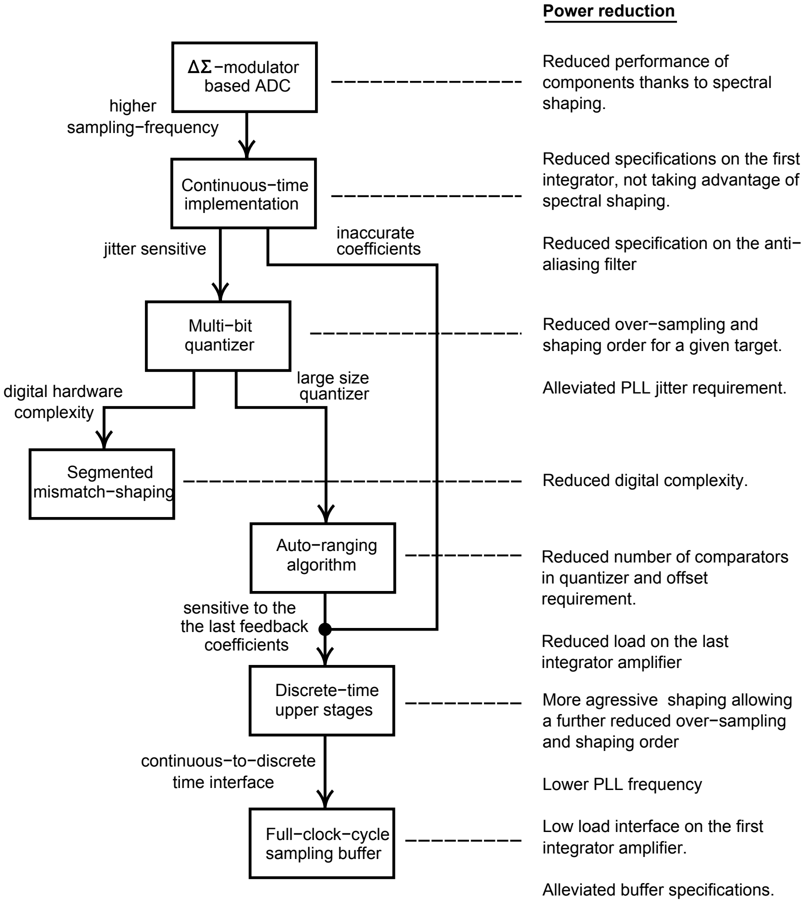

Power reduction
ΔΣ−modulator based ADC
higher sampling−frequency
Continuous−time implementation
jitter sensitive
inaccurate coefficients
Multi−bit quantizer
digital hardware complexity
large size quantizer
Segmented mismatch−shaping
Auto−ranging algorithm
sensitive to the the last feedback coefficients
Discrete−time upper stages
continuous−to−discrete time interface
Full−clock−cycle sampling buffer
Reduced performance of components thanks to spectral shaping.
Reduced specifications on the first integrator, not taking advantage of spectral shaping.
Reduced specification on the anti−aliasing filter
Reduced over−sampling and shaping order for a given target.
Alleviated PLL jitter requirement.
Reduced digital complexity.
Reduced number of comparators in quantizer and offset requirement.
Reduced load on the last integrator amplifier
More agressive shaping allowing a further reduced over−sampling and shaping order
Lower PLL frequency
Low load interface on the first integrator amplifier.
Alleviated buffer specifications.

Figure 3.6 Low-power design strategy.

side the control loop thus taking the best advantage of spectral shaping. Furthermore, the signal representation progressively goes from an infinite to a moderate number-of-levels before reaching the expected high resolution. In contrast, with a single-bit modulator, the representation goes down to a low number-of-levels with a high over-sampling frequency.

3.2 Anti-aliasing property

3.2.1 Signal processing of the continuous-time modulator

The signal processing of the continuous-time modulator in Figure 3.1 is described in detail in the block diagram in Figure 3.7.

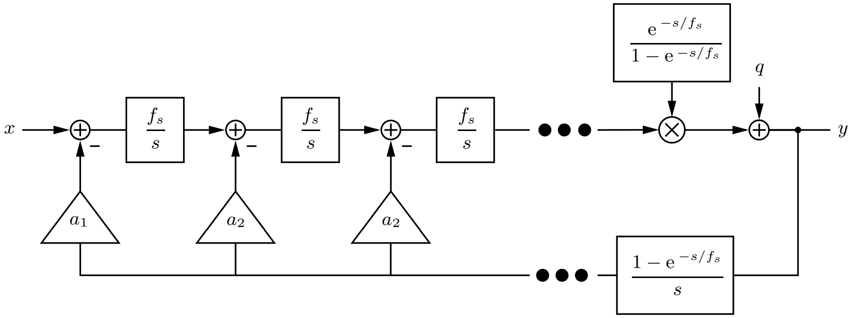


Figure 3.7 Continuous-time ΔΣ-modulator block-diagram.

The sampling process, which converts a continuous-time signal into a discrete-time representation, is mathematically expressed as a multiplication by a train of pulses in the time domain. In the frequency domain, the multiplication becomes a convolution by the Laplace transform of the train of pulses which is given by

$$\mathscr{L}\left\{\sum_{k=1}^{\infty}\delta(t-k/f_s)\right\} = \frac{e^{-s/f_s}}{1-e^{-s/f_s}}\,. \tag{3.1}$$

In contrast, the conversion of a discrete-time signal into a continuous-time representation by a No-Return-to-Zero (NRZ) DAC, is mathematically expressed as convolution, in the time domain, by a rectangular pulse whose duration is $1/f_s$. In the frequency domain, the convolution becomes a multiplication by the Laplace-transform of the rectangular pulse which is given by

$$\mathscr{L}\left\{u(t)-u(t-1/f_s)\right\} = \frac{1-e^{-s/f_s}}{s}\,. \tag{3.2}$$

As already explained in detail in Chapter 2, the quantization process can be modeled as an additional discrete-time random sequence q with a uniform distribution and a white power spectral density.

A multiplication in the time domain is not a linear operation. Nevertheless, the sampling of the sum of two signals can be expressed here as the sum of the two sampled signals. We can therefore redraw the block diagram as in Figure 3.8.

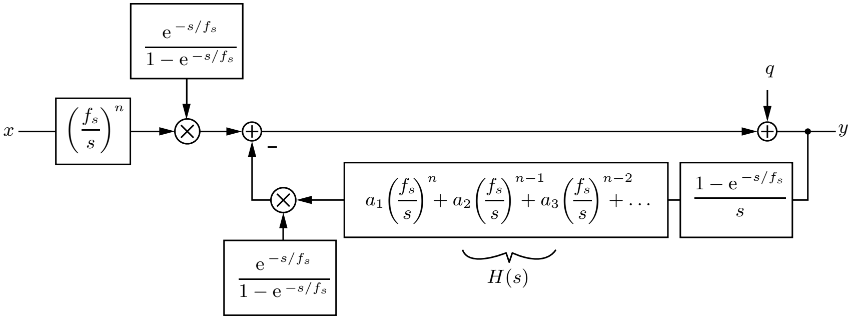

x
$\left(\frac{f_s}{s}\right)^n$
$\frac{e^{-s/f_s}}{1-e^{-s/f_s}}$
q
y
–
$a_1\left(\frac{f_s}{s}\right)^n + a_2\left(\frac{f_s}{s}\right)^{n-1} + a_3\left(\frac{f_s}{s}\right)^{n-2} + \ldots$
$\frac{1-e^{-s/f_s}}{s}$
$\frac{e^{-s/f_s}}{1-e^{-s/f_s}}$
$H(s)$

Figure 3.8 Block diagram in Figure 3.7 redrawn to distinguish a continuous-time forward path and the discrete-time internal loop.

The processing chain involving sampling, the continuous-time filter $H(s)$ and holding can be described as a pure discrete-time transfer function using the so-called *step-invariant transform*, which is stated in [Lon95, Oga95] as

$$H(z) = (1 - z^{-1})\mathscr{Z}\left\{\mathscr{L}^{-1}\left\{\frac{H(s)}{s}\right\}\right\} . \tag{3.3}$$

The coefficients a_i are designed such as to provide a discrete-time transfer function $H(z) = 1 - 1/\mathrm{NTF}(z)$ as shown in Figure 3.9. In this way, the quantization error sequence goes through the NTF before reaching the output y.

As a result, the input signal x goes through an nth-order integration before sampling and then through the NTF.

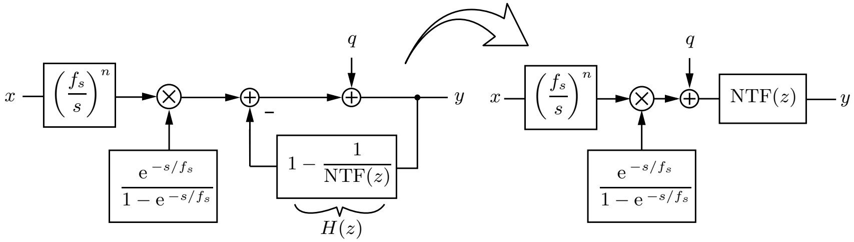

x
$\left(\frac{f_s}{s}\right)^n$
$\frac{e^{-s/f_s}}{1-e^{-s/f_s}}$
q
−
$1-\frac{1}{\mathrm{NTF}(z)}$
$H(z)$
y
x
$\left(\frac{f_s}{s}\right)^n$
$\frac{e^{-s/f_s}}{1-e^{-s/f_s}}$
q
NTF(z)
y

Figure 3.9 Block diagram in Figure 3.8 redrawn to highlight the STF and NTF.

3.2.2 Input signal spectrum folding

According to the block diagram in Figure 3.9, and assuming the NTF has all the poles at the center of the z-plane and all its zeros at DC, the STF can be expressed, for signal frequencies lower than the Nyquist rate $f_s/2$, as

$$\mathrm{STF}(f) = \left(\frac{f_s}{s}\right)^n \cdot (1 - z^{-1})^n \,. \tag{3.4}$$

The transfer function is plotted in Figure 3.10 for different modulator orders n. The second-order case is plotted as a solid line. For in-band input signals the continuous-time integrators and the discrete-time high-pass function cancel each other. The STF is therefore equal to unity. For a signal at the Nyquist rate $f_s/2$, the terms do not cancel, giving an attenuation of $(2/\pi)^n$.

A signal at f_s is filtered by the cascade of integrators. But the discrete-time high-pass function, thanks to its periodicity, completely removes the signal. As highlighted by the grey areas in Figure 3.10, the attenuated parts of the spectrum close to fold back to the band-of-interest.

For a case with an OSR of 32, we find 72dB of attenuation of the out-of-band interferers around the sampling frequency. This contribution to anti-aliasing alleviates the specifications of the filter placed in front of the modulator. Since in a discrete-time implementation the sampling process occurs at the modulator input, no anti-aliasing can be provided by the modulator itself. The parts of the input spectrum around multiples of $\pm f_s$ directly fold back in the band-of-interest at sampling, without attenuation.

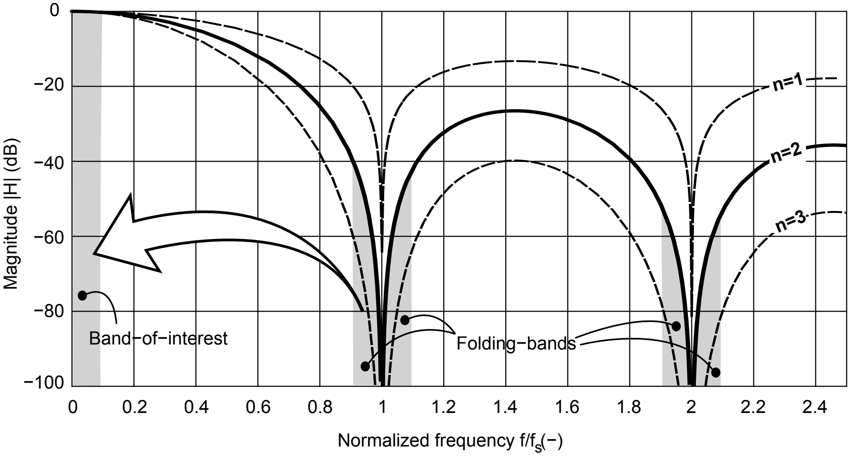


Figure 3.10 Continuous-time transfer function for the modulator input to the sampling node for a first-, second- and third-order modulator, according to Equation (3.4). The second-order case is shown in as solid line and the gray areas represent the parts of the spectrum that fold back to the band-of-interest when the OSR is equal to 5.

In a hybrid case the loop transfer function is the same. Assuming the c first integrators are continuous-time and the remaining $n-c$ integrators are discrete-time, the forward path become partially continuous-time. The STF becomes

$$\mathrm{STF}(f) = \left(\frac{f_s}{s}\right)^c \cdot (1 - z^{-1})^c \cdot z^{c-n} \ . \tag{3.5}$$

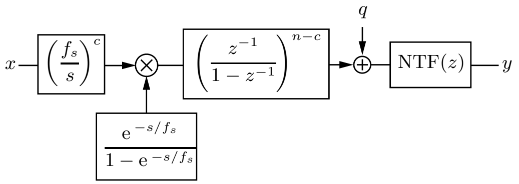

x
$\left(\frac{f_s}{s}\right)^c$
$\left(\frac{z^{-1}}{1-z^{-1}}\right)^{n-c}$
q
NTF(z)
y
$\frac{\mathrm{e}^{-s/f_s}}{1-\mathrm{e}^{-s/f_s}}$

Figure 3.11 Block diagram of an nth-order hybrid continuous-discrete-time modulator with c continuous-time first integrators.

As a result, the inherent anti-aliasing filter is the same as in the case of a cth-order purely continuous-time modulator. For a case with only a first continuous-time integrator and an OSR of 32, an interferer around f_s is attenuated by at least 36dB.

3.3 Clock jitter issue

3.3.1 Discrete- and continuous-time mechanisms

The timing uncertainty on the clock edges, also referred to as *clock jitter*, turns into an error signal similar to thermal noise. The transfer mechanism from time to amplitude is different for continuous- and discrete-time ΔΣ-modulators. Figure 3.12 shows the two transfer processes, revealing that the clock signal controls the feedback holder in the continuous-time case and the input sampler in the discrete-time case.

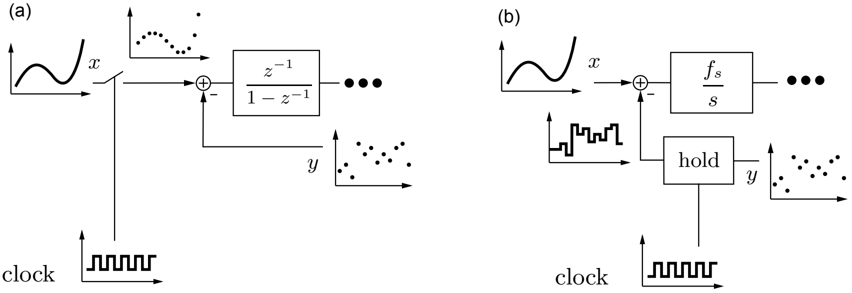


Figure 3.12 ΔΣ-modulator input stage with discrete- (a) and continuous-time (b) implementations.

In the discrete-time modulator, the input signal x is sampled before processing. The samples are further considered by the analog controller as being equally apart in time by the period T_s. In reality, in the presence of jitter, this period is not constant, which provides the controller with a randomly modified signal. We refer to these generated errors as *sampling errors.*

In a continuous-time modulator, the signal is first processed and then sampled at the quantizer input, thus taking full advantage of spectral shaping. Hence, *sampling errors* are negligible. On the other hand, the feedback digital signal y is held for a period T_s considered constant by the controller. In reality, the first stage performs the integral of y over a varying period. We refer to these errors as *holding errors*.

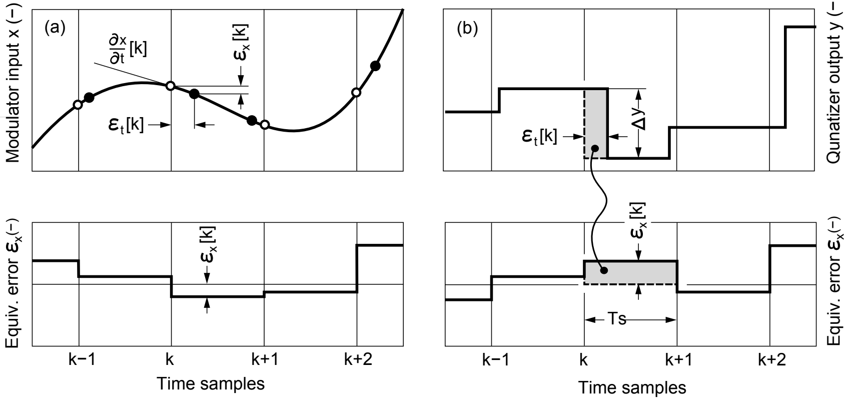


Figure 3.13 Discrete- (a) and continuous-time (b) clock-jitter transfer mechanism to an equivalent input error sequence.

Figure 3.13 presents the two mechanisms in detail. In both cases, an equivalent input error $\varepsilon_x[k]$ can be calculated, related to the clock timing error ε_t.

$$\varepsilon_x[k] = \varepsilon_t[k] \cdot \left\{ \begin{array}{ll} \dfrac{\partial x}{\partial t}[k] & \text{Discrete-time} \\ \dfrac{\Delta y}{T_s}[k] & \text{Continuous-time} \end{array} \right. . \tag{3.6}$$

The amount ε_t is commonly referred to as *cycle-to-average jitter* in literature [SSS05]. As described in Figure 3.13, the sampling errors depend on the derivative of x whereas the holding errors are linked to the variations from sample to sample of y. Assuming a stochastic timing error sequence, we can write

$$\sigma_x^2 = \sigma_t^2 \cdot \left\{ \begin{array}{ll} \sigma^2_{\partial x/\partial t} & \text{Discrete-time} \\ \sigma^2_{\Delta y/T_s} & \text{Continuous-time} \end{array} \right. . \tag{3.7}$$

Considering a clock-jitter with a Gaussian distribution and a white spectrum, the PSD of the equivalent discrete-time input noise keeps the same properties

$$\mathscr{P}_x(f) = \frac{2}{f_s} \cdot \sigma_x^2 \,. \tag{3.8}$$

Let us define here the *Signal-to-Jitter Noise Ratio* SJNR as the signal power P_s to P_j, the *jitter-induced* noise power, integrated over the band-of-interest f_b as

$$\mathrm{SJNR} = \frac{P_s}{P_j} = \frac{\|x\|_\infty^2 /2}{\int_0^{f_b} \mathscr{P}_x(f)\mathrm{d}f} = \frac{1}{\mathrm{OSR}(2 f_b \sigma_t \xi)^2} \,, \tag{3.9}$$

or as function of the the *relative* jitter standard deviation $\sigma_r = \sigma_t f_s$

$$\mathrm{SJNR} = \frac{P_s}{P_j} = \frac{\mathrm{OSR}}{(\sigma_r \xi)^2} \,. \tag{3.10}$$

We introduce here the *jitter transfer factor* ξ as the jitter to equivalent amplitude error gain, normalized to the maximum input amplitude $\|x\|_\infty$.

$$\xi = \frac{\sqrt{2}}{\|x\|_\infty} \cdot \begin{cases} \sigma_{dx} & \text{Discrete-time} \\ \sigma_{\Delta y} & \text{Continuous-time} \end{cases} \,. \tag{3.11}$$

Figure 3.14 shows the simulated jitter transfer factor for different modulator orders and number-of-levels. These curves are determined with the MATLAB scripts provided in Chapter 2 and the additional Code 3.1.

Code 3.1 Additional MATLAB code to determine the jitter transfer factor.

```
x =x0*(sin(2*pi*f0*(1:k)'/fs));
dx=2*pi*f0/fs*x0*(cos(2*pi*f0*(1:k)'/fs));
...
xi_discretetime=sqrt(2)*std(dx)/x0;
xi_continuoustime=sqrt(2)*std(diff(out))/x0;
```

This simulation is purely discrete-time and its purpose is to consider the statistical nature of Δy and $\mathrm{d}x$. As a consequence, the attenuation for a tone close to the Nyquist frequency $f_s/2$ provided by the continuous-time integrators was not taken into account. Hence the continuous-time

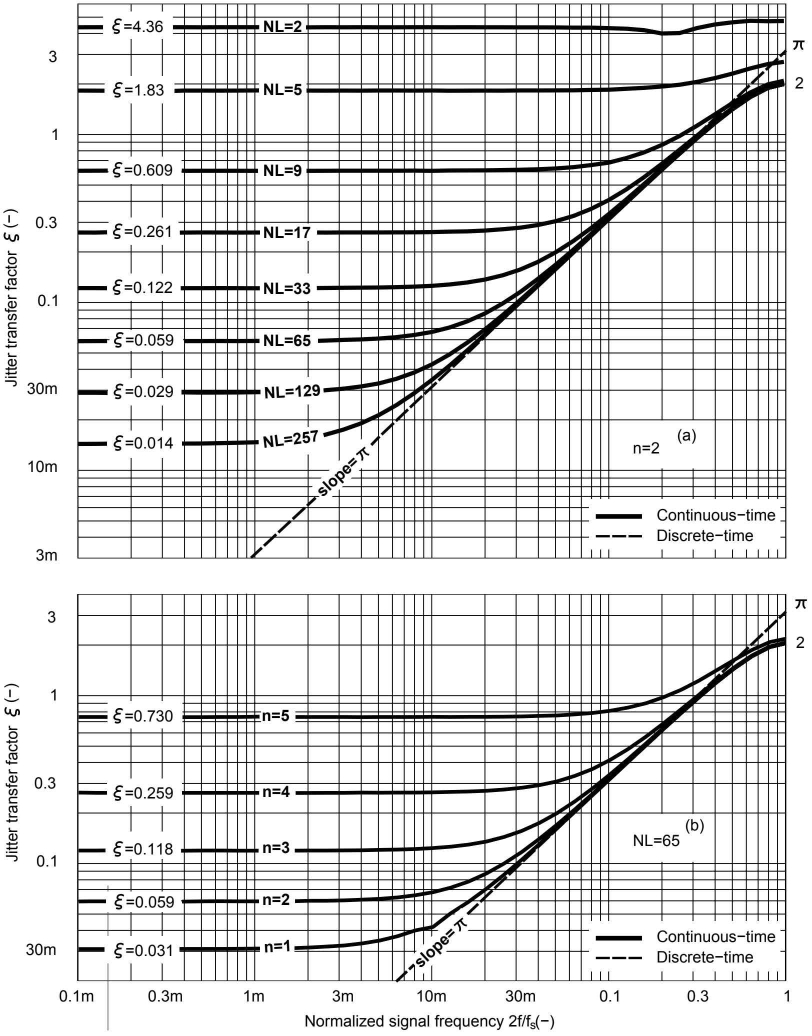


Figure 3.14 Simulated *jitter transfer factor* ξ for (a) a given modulator order n and for (b) a given number-of-levels NL. The continuous- and discrete-time factors are plotted as solid and dashed lines respectively.

transfer factors at Nyquist frequency reach a value of 2 independently of the modulator order. The analytical development in the next section introduces the effect of this attenuation. The factor at $f_s/2$ is found to be slightly less than two.

The chart shows that with a large NL, the continuous- and discrete-time ξ factors are the same for a certain frequency range. Whereas, close to Nyquist, the discrete-time factor is higher, at low frequencies the continuous-time factor reaches the baseline. This happens because even for slow input signals, a minimum of activity is brought abount by sequence of quantization errors. This activity was already highlighted by the *quantization factor* k_q defined and calculated in Section 2.2.3. As confirmed in Figure 3.14(b), the higher the shaping order n the large this activity.

As a consequence, the sensitivity to clock-jitter of a continuous-time $\Delta\Sigma$-modulator can be reduced by increasing NL while keeping a low shaping order n. The single-bit case, NL=2, presents the highest jitter transfer factor. In contrast, if NL is large enough the continuous-time case reaches the limit set by the discrete-time case. According to Equations (3.9) and (3.10), provided ξ constant, by increasing the OSR the sensitivity to the relative jitter σ_r is improved but the sensitivity to time jitter σ_t is degraded.

This limit can be overcome with Return-to-Zero (RZ) techniques where the pulse duration is constant and its integral independent of the clock edges. This can be done with either an RC-shape or rectangular pulse. Recent publications such as [OG06, DKG$^+$05, vV03] demonstrated the reduction of sensitivity of switched-capacitor feedback in continuous-time. Nevertheless, the return to zero feedback demands an rather larger amplifier bandwidth. The benefits of the continuous-time implementations are significantly reduced.

3.3.2 Analytical expression of the jitter transfer factor

In the discrete-time case, the transfer factor depends on dx. For a full-scale sinusoidal input $x(t) = \|x\|_\infty \sin(2\pi f)$ we find the standard deviation of dx

$$\sigma_{\mathrm{d}x} = \frac{1}{f_s} \cdot \frac{2\pi f \, \|x\|_\infty}{\sqrt{2}} \, . \tag{3.12}$$

As a consequence, the jitter transfer factor is π times the tone oversampling ratio

$$\xi = \pi \cdot \frac{2f}{f_s} \, . \tag{3.13}$$

In the continuous-time case, the transfer factor depends on Δy. According to the modulator equation (2.15), the output y is a superposition of the signal x and the quantization noise q. We can therefore write

$$\Delta Y(z) = (1 - z^{-1}) \Big(\mathrm{STF}(z) X(z) + \mathrm{NTF}(z) Q(z) \Big) \, . \tag{3.14}$$

Let us restrict our development to a multi-feedback nth-order modulator with all the poles at the center of the z-plane and all the zeros at DC. On the other hand, we consider the general case where the c first integrators are continuous-time. Under these conditions we can express the *Noise* and *Signal Transfer Functions*, NTF and STF as

$$\mathrm{NTF} = (1 - z^{-1})^n \, , \tag{3.15}$$

$$\mathrm{STF} = \left(\frac{1 - z^{-1}}{2\pi j f / f_s} \right)^c \cdot z^{-n+c} \, . \tag{3.16}$$

Replacing these expressions in Equation (3.14) and further extracting its statistical variance we find

$$\sigma^2_{\Delta y} = \left| \frac{(1 - z^{-1})^{c+1}}{(j 2\pi f / f_s)^c} \right|^2 \cdot \sigma^2_x + \left\| (1 - z^{-1})^{n+1} \right\|^2_2 \cdot \sigma^2_q \, . \tag{3.17}$$

The 2-norm $\|.\|_2$ is defined in Appendix A.1.

The variance of the input signal σ_x^2 is equal to the signal power $P_s = \frac{1}{8}(\mathrm{NL} - 2^n + 1)^2$ and the variance of the quantization sequence σ_q^2, calculated in Chapter 2, equal to 1/12. The development of 2-norm $\|.\|_2$ is treated in detail in Appendix A.3. As a result, once normalized according to its definition (3.11), the *jitter transfer factor* becomes

$$\boxed{\xi = \sqrt{\left|\frac{(1-\exp[-j2\pi f/f_s])^{c+1}}{(2\pi f/f_s)^c}\right|^2 + \frac{2}{3}\cdot\frac{(2n+2)!/(n+1)!^2}{(\mathrm{NL}-2^n+1)^2}}}\,. \tag{3.18}$$

The jitter transfer factor is made up of a frequency-dependent term related to the input signal, and a constant related to the quantization error sequence. The latter is responsible for the flat behavior at low input frequencies. The flat level can therefore be found at zero frequency, when the first term is negligible. Therefore

$$\xi(f=0) = \frac{\sqrt{\frac{2}{3}(2n+2)!}}{(n+1)!(\mathrm{NL}-2^n+1)}\,. \tag{3.19}$$

Table 3.2 Continuous-time jitter transfer factor at low frequency according to Equation (3.19).

		Number-of-levels NL							
n	$\xi(0)$	2	5	9	17	33	65	127	257
1	$\frac{2}{\mathrm{NL}-1}$	2	0.50	0.25	0.125	0.0625	0.0313	0.0156	0.00781
2	$\frac{2\sqrt{10/3}}{\mathrm{NL}-3}$	-	1.83	0.609	0.261	0.122	0.0589	0.029	0.0144
3	$\frac{2\sqrt{35/3}}{\mathrm{NL}-7}$	-	-	3.42	0.683	0.263	0.118	0.056	0.0273
4	$\frac{2\sqrt{42}}{\mathrm{NL}-15}$	-	-	-	6.48	0.720	0.259	0.114	0.0536
5	$\frac{2\sqrt{154}}{\mathrm{NL}-31}$	-	-	-	-	12.4	0.730	0.253	0.110

Equation (3.19) is evaluated for different modulator orders and number-of-levels and summarized in Table 3.2.

For moderate frequencies lower than $f_s/2$, the first term dominates and we can approximate the exponential term and find that

$$\xi(f < f_s/2) \cong \left| \frac{(1-\exp[-j2\pi f/f_s])^{c+1}}{(2\pi f/f_s)^c} \right| \cong \pi \cdot \frac{2f}{f_s}\,. \tag{3.20}$$

Thus, for moderate frequencies the transfer factors of the discrete- and continuous-time case are almost identical. Finally, at the Nyquist frequency, the continuous-time factor saturates and the STF further attenuates it. Therefore

$$\xi\,(f = f_s/2) = \frac{2^{c+1}}{\pi^c}\,. \tag{3.21}$$

According to Equation (3.13) in the discrete-time case $\xi = \pi$ at the Nyquist rate. The fact that $\pi > 2$ and so $\pi^{c+1} > 2^{c+1}$ implies that $\pi > 2^{c+1}/\pi^c$. The continuous-time jitter transfer factor at the Nyquist rate is therefore always smaller than the discrete-time case. Notice that $c = 0$ gives $\xi = 2$ as simulated in the chart in Figure 3.14. In such a case the STF attenuation is simply not considered.

3.3.3 Resolution degradation

In the presence of jitter, the resulting SNR degradation can be evaluated as

$$\mathrm{SNR} = \frac{P_s}{P_q + P_j} = \frac{1}{\dfrac{1}{\mathrm{SQNR}} + \dfrac{1}{\mathrm{SJNR}}}\,. \tag{3.22}$$

When the contributions of the quantization errors and the jitter are the same SQNR=SJNR, bringing the SNR down by 3dB. We can therefore find the relative jitter standard deviation under these conditions

$$\sigma_{3\mathrm{dB}} = \frac{1}{\xi}\sqrt{\frac{\mathrm{OSR}}{\mathrm{SQNR}}}\,. \tag{3.23}$$

This relationship, along with the chart in Figure 3.14, shows that the sensitivity to clock jitter can be improved by either increasing the OSR or the number-of-levels. On the other hand, increasing the shaping order is not recommended.

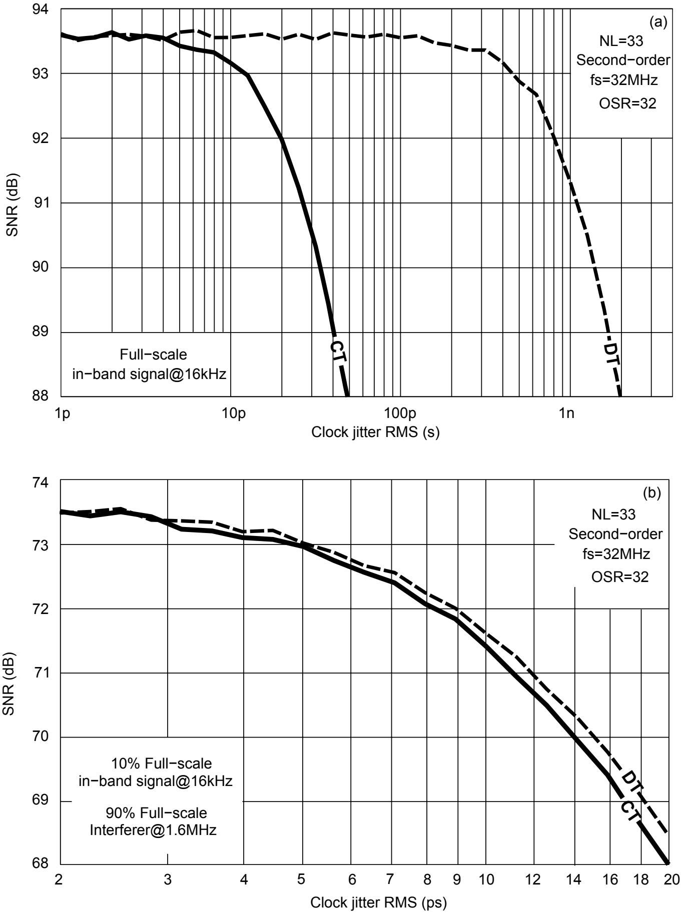


Figure 3.15 Jitter sensitivity plot for a 33-level second-order case with an OSR of 32. The input signal is (a) a full-scale in-band tone and (b) a −20dB in-band tone with an additional −0.9dB out-of-band signal.

As an example, a 33-level second-order modulator, sampled at 32MHz, with a full-scale input signal at 16kHz presents a transfer factor of 0.122 and $\pi/1000$ for the continuous- and discrete-time cases respectively. The −3dB SNR drop is therefore expected at 30ps and 1.2ns, which highlights the higher tolerance to clock jitter in the discrete-time case. The simulation results in Figure 3.15 reveals the accuracy with which the charts in Figure 3.14 can predict the impact of jitter on the modulator resolution. In contrast, in the presence of a full-scale out-of-band signal at 1.6MHz, the transfer factors are close to each other, namely 0.28 and 0.30 once scaled by 90%. The SQNR in the formula is kept unchanged, therefore providing the drop at 12ps and 13ps for the continuous- and discrete-time cases.

Figure 3.16 shows the calculated relative jitter sensitivity $\sigma_{3\mathrm{dB}}$ with respect to the design parameters SQNR, NL and OSR. For that purpose the analytical expressions of the jitter transfer factor (3.13) and (3.18) are used. The chart is intended to help to choose at an early design stage how to address a targeted resolution. Sets of OSR and NL providing a given SQNR are found by using the inverted form of the resolution equation (2.39). The curves confirm that, for a large enough NL, a discrete- and a continuous-time modulator have the same sensitivity.

The solutions with large NL and small OSR are shown on the right side of Figure 3.16, where the impact of jitter is less attenuated by over-sampling as predicted by Equation (3.23). As shown previously, close to the Nyquist rate the continuous-time sensitivity slightly outperforms the discrete-time. Hence, these solutions may present a better sensitivity, especially when considering signals that are far out-of-band.

In Figure 3.16, the solutions with small NL and high OSR are shown on the left side, where a clear separation is made between the discrete- and continuous-time implementations. These solutions present a deceptively low sensitivity in the discrete-time case. For high-bandwidth applications, the resulting clock frequency would not be appropriate for a low power implementation. Besides, the continuous-time case fails to take advantage of the attenuation due to over-sampling.

Somewhere in between, the continuous-time curves reveal an optimal solution where the relative sensitivity is minimal. In the presence of a full-scale out-of-band interferer at twice the bandwidth, the optimal is found for the 33-level solution with an OSR of 32.

Figure 3.17 presents the same analysis for different modulator orders n. The curves reveal an increased relative sensitivity for high modulator orders. Nevertheless, it should be noted that as the order increases, for

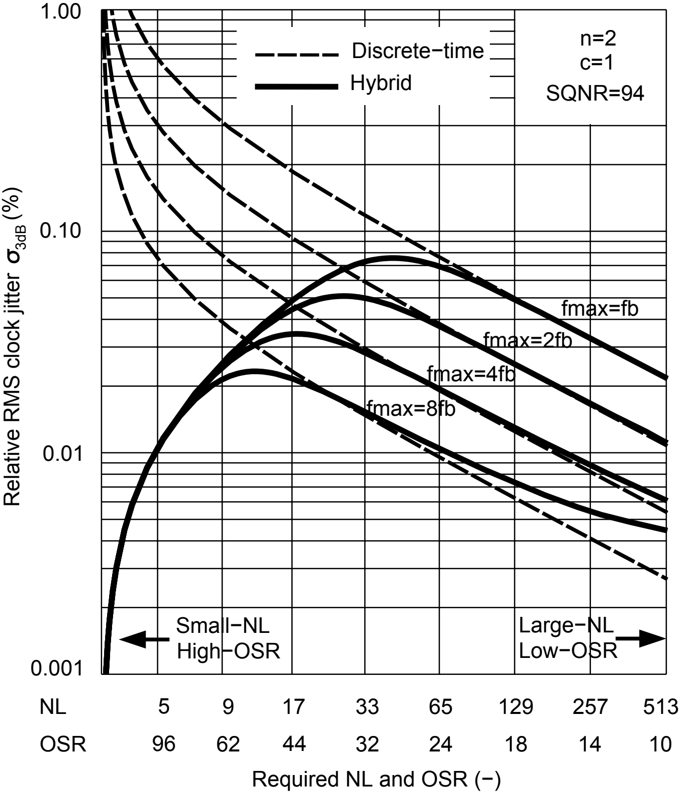


Figure 3.16 Design chart for a second-order modulator with an SQNR of 94dB using Equations (3.18) and (3.13). The relative jitter sensitivity is plotted considering different maximum frequencies $f_{\max}$. The discrete- and hybrid cases are represented by dashed and solid lines respectively.

a targeted SQNR, the OSR decreases. As a result, the absolute jitter sensitivity remains almost constant. Again, the optimal solutions are found for moderate NL's between 33 and 65.

Finally, we recall that we restricted the analytical development to single-stage multi-bit non-overloaded cases with an NTF whose poles and zeros are at the center of the z-plane and at DC respectively. The single-bit case with NL=2, being constantly overloaded, is therefore not

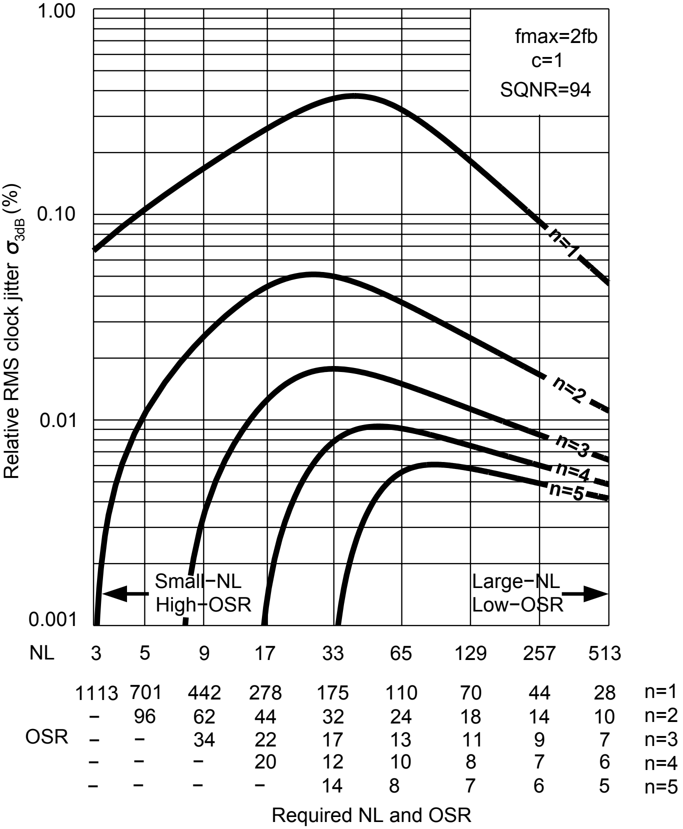


Figure 3.17 Design chart for modulator with an SQNR of 94dB and different orders n using Equations (3.18) and (3.13). The relative jitter sensitivity is plotted considering a maximum frequency of twice the bandwidth. Only the hybrid case with a continuous-time first integrator is represented.

accurately predicted by the chart where the sensitivity is expected to drop down to zero. In particular, according to the dynamic range equation (2.30) developed in Chapter 2, the 3-level case is the limit of overload for a second-order modulator with an NTF $= (1 - z^{-1})^2$. For that solution, Equation (3.18) provides an infinite jitter transfer factor. In reality, the allowed input signal amplitude is not zero and therefore the normalization does not change the transfer factor to infinity. As a matter of fact, the simulated jitter factor in Figure 3.14 is equal to 2.34 and is almost constant with respect to signal frequency.

3.4 Accuracy of coefficients

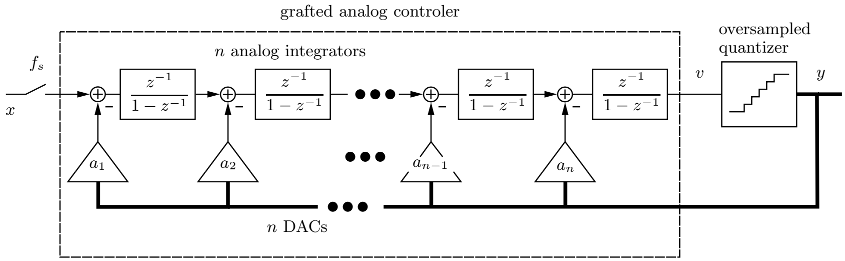


Figure 3.18 $\Delta\Sigma$-modulation: Conventional nth-order architecture.

Any single-stage n-th order modulator consists of a control system containing n loops. Each loop is a integration path weighted by a coefficient a_i. The order of integration goes from one to n. For example, a third-order modulator is made up of a single, double and triple integration loop. From the conventional architecture in Figure 3.18 we can write the NTF as

$$\mathrm{NTF}(z) = \frac{1}{1+\sum_{i=1}^{n} a_i \left(\frac{z^{-1}}{1-z^{-1}}\right)^{n-i+1}} = \frac{(z-1)^n}{(z-1)^n + \sum_{i=1}^{n} a_i (z-1)^{i-1}} \,. \tag{3.24}$$

We choose the coefficient a_i such as to cancel out the terms of the expansion of $(z-1)^n$ leaving only z^n. Under this condition, all the poles are located at the center of the z-plane. Now we consider that the coefficient a_k presents a certain deviation Δa_k around its nominal value. Therefore we can write

$$\mathrm{NTF}(z) = \frac{(z-1)^n}{z^n + \Delta a_k (z-1)^{k-1}} \,. \tag{3.25}$$

The location of the poles is therefore provided by the solution of the following equation

$$z^n + \Delta a_k (z-1)^{k-1} = 0 \,. \tag{3.26}$$

For each coefficient, taken individually, we can find the deviation Δa_k such that at least one pole reaches the unity circle of the complex z-plane. Numerical solutions of Equation (3.26) are provided in Table 3.3. Each coefficient has a nominal value when poles are placed at the center of the z-plane. The poles remain within the unity circle as long as the coefficients do not go beyond the maximum and minimum deviations. Beyond these ranges, stability is compromised.

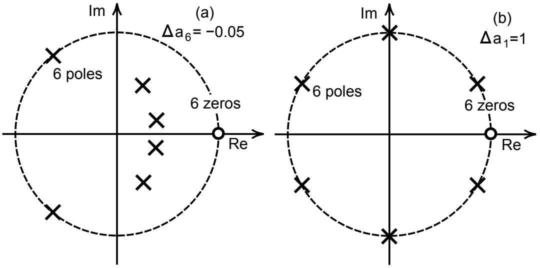


Figure 3.19 NTF zeros and poles location for a sixth-order modulator. The six zeros are located at DC and the six poles at the center of the complex z-plane. As the sixth (a) and first (b) coefficients deviate from their nominal value, the poles moves toward the unity circle. The poles are represented when the variation brings the modulator at its limit of stability.

Table 3.3 Feedback coefficients with their respective tolerance to guarantee stability.

n	a_1	a_2	a_3	a_4	a_5	a_6
1	$1^{+1.00}_{-1.00}$					
2	$1^{+1.00}_{-1.00}$	$2^{+0.50}_{-1.00}$				
3	$1^{+1.00}_{-1.00}$	$3^{+0.62}_{-0.50}$	$3^{+0.25}_{-0.50}$			
4	$1^{+1.00}_{-1.00}$	$4^{+0.50}_{-0.56}$	$6^{+0.33}_{-0.25}$	$4^{+0.13}_{-0.24}$		
5	$1^{+1.00}_{-1.00}$	$5^{+0.53}_{-0.50}$	$10^{+0.25}_{-0.29}$	$10^{+0.17}_{-0.13}$	$5^{+0.06}_{-0.11}$	
6	$1^{+1.00}_{-1.00}$	$6^{+0.50}_{-0.52}$	$15^{+0.28}_{-0.25}$	$20^{+0.13}_{-0.15}$	$15^{+0.08}_{-0.06}$	$6^{+0.03}_{-0.05}$

Since the coefficients need to compensate for the terms of the polynomial $(z-1)^n$, the table takes the shape of the *Pascal triangle*. The allowed deviation ranges drastically decrease with the coefficient index. As a result, high-order modulators see their stability compromised by the accuracy of the last loop coefficient. In other words, the signal paths with single and double integration are the most sensitive ones. Besides, the n-integration loop coefficients a_1 can theoretically sustain variations of ±100% for all values of n if the system is really linear.

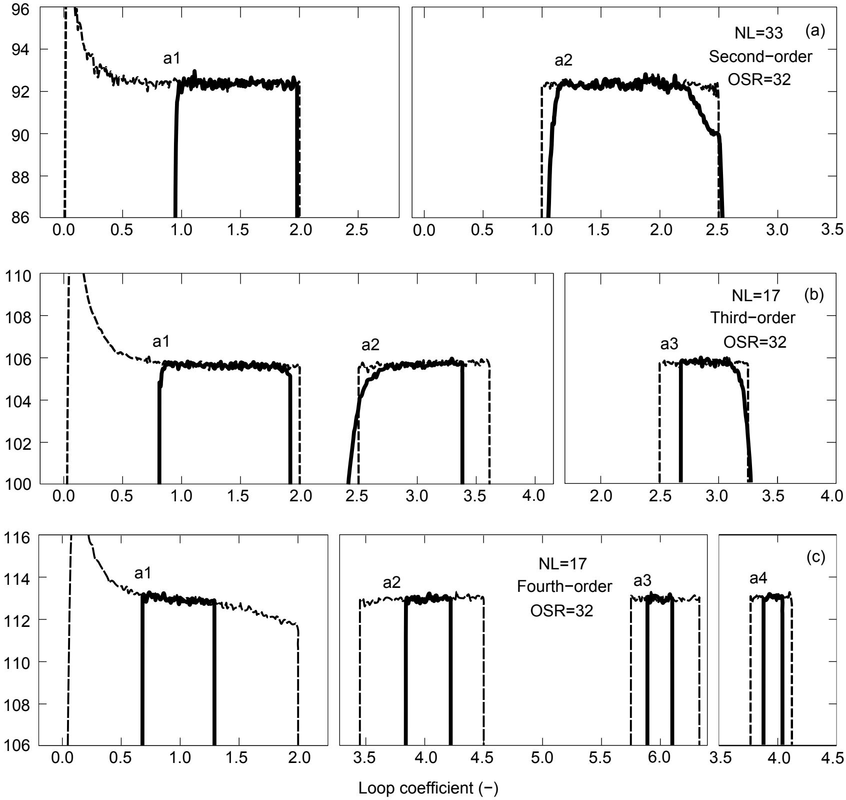


Figure 3.20 Feedback coefficient sensitivity for a second- (a), third- (b) and fourth-order (c) modulator. Results with and without the quantizer saturation are displayed as solid and dashed lines respectively.

The simulation results in Figure 3.20 reveal a further reduction of the allowed range. Removing the quantizer saturation gives the results shown with the dashed lines. In such a case, the number-of-levels is infinite, the modulator remains a linear system and the tolerances of Table 3.4 perfectly predict the range of operability. The solid lines show the quantizer with a limited number-of-levels NL. The deviation of coefficients brings the modulator to a non-linear region, causing an early drop of the resolution.

3.5 Sampling capacitor mismatch

We already mentioned, in the description of the strategy, that sampling the signal over the whole clock-cycle implies alternating between two sets of capacitors at a frequency of $f_s/2$. Therefore, the gain of the sampling device also alternates between two slightly different values, G_1 and G_2. These can be expressed as a constant, the mean value $\overline{G} = (G_1 + G_2)/2$, and a mismatch term $\eta = (G_2 - G_1)/\overline{G}$ such that

$$G_{1,2} = \overline{G}\left(1 \pm \frac{\eta}{2}\right) \ . \tag{3.27}$$

Consequently, as illustrated in Figure 3.21, the gain splits into a linear gain $\overline{G}$ and a modulating part $(-\eta/2)^k$. The modulating part takes the signal and the quantization sequence at the sampling node and shifts their spectrum by $f_s/2$.

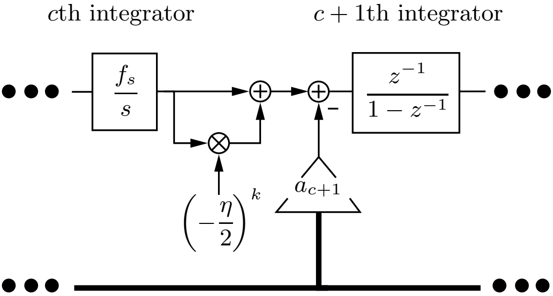


Figure 3.21 Equivalent block diagram of the full-clock-cycle sampling.

In the multi-feedback modulators, as considered until now with an NTF with all poles located at the center of the z-plane, the coefficients a_i

are binomial. The output of the cth integrator $C(z)$ is given by

$$C(z) = -Q(z)(1 - z^{-1})^n \sum_{i=1}^{c} \binom{i-1}{n} \left(\frac{z^{-1}}{1 - z^{-1}} \right)^{c-i+1} . \tag{3.28}$$

The modulation shifts the spectrum of $C(z)$ by $f_s/2$ and weights it by η. Once reintroduced, the sequence goes through the first-order spectral shaping and we find at the modulator output a contribution $M(z)$ equal to

$$M(z) = -Q(z) \cdot \frac{\eta}{2} \cdot \underbrace{(1 + z^{-1})^n \sum_{i=1}^{c} \binom{i-1}{n} \left(\frac{-\,z^{-1}}{1 + z^{-1}} \right)^{c-i+1}}_{\alpha} . \tag{3.29}$$

Considering moderate to high over-sampling ratios allows us to approximate $z^{-1} \cong 1$ within the band of interest. The factor α can therefore be calculated as

$$\alpha = 2^n \sum_{i=1}^{c} \binom{i-1}{n} \left(-\frac{1}{2} \right)^{c-i+1} . \tag{3.30}$$

The power spectral density $\mathscr{P}_M$ referred to the modulator output can therefore be calculated as

$$\mathscr{P}_M = \frac{\eta^2}{4} \cdot \alpha^2 \cdot \left| (1 - z^{-1})^c \right|^2 \cdot \mathscr{P}_Q . \tag{3.31}$$

In the presence of mismatch, the resulting SNR degradation can be evaluated as

$$\mathrm{SNR} = \frac{P_s}{P_Q + P_M} = \frac{\mathrm{SQNR}}{1 + P_M/P_Q} . \tag{3.32}$$

The power ratio is evaluated over the band-of-interest and gives

$$\frac{P_M}{P_Q} \cong \frac{\eta^2}{4} \alpha^2 \frac{\int_0^{f_b} (2\pi f/f_s)^{2c} \mathrm{d}f}{\int_0^{f_b} (2\pi f/f_s)^{2n} \mathrm{d}f} . \tag{3.33}$$

Again considering high over-sampling ratios allows us to approximate the integrals and we find

$$\frac{P_M}{P_Q} \cong \frac{\eta^2}{4} \cdot \alpha^2 \cdot \frac{(2n+1)(2\pi f_b/f_s)^{2c+1}}{(2c+1)(2\pi f_b/f_s)^{2n+1}} = \frac{\eta^2}{4} \cdot \alpha^2 \cdot \frac{2n+1}{2c+1} \cdot \left(\frac{\text{OSR}}{\pi}\right)^{2(n-c)} . \tag{3.34}$$

When the contributions of the quantization errors and the mismatch are the same $P_M = P_Q$, bringing the SNR down by 3dB. The mismatch at that point is therefore calculated as

$$\boxed{\eta_{3\text{dB}} = \sqrt{\frac{2c+1}{2n+1}} \cdot \frac{2}{\alpha} \cdot \left(\frac{\pi}{\text{OSR}}\right)^{n-c}} . \tag{3.35}$$

In the case of a pure continuous-time modulator, $n = c$ and the sensitivity to the capacitor mismatch $\eta_{3\text{dB}} = 2/\alpha$ and is therefore independent of the order and the over-sampling ratio. In contrast, for a hybrid case where $c = 1$, the relationship reveals an important increase in sensitivity with regard to the OSR and the modulator order. Consequently, as for the jitter sensitivity issue, it is preferable to choose, for a targeted SQNR, solutions with low OSR and small n to reduce the sensitivity to mismatch in the sampling device.

Table 3.4 Calculated fifth-order cases with OSR=32.

c	1	2	3	4	5
α	16	72	124	98	31
$\eta_{3\text{dB}}$	0.00061%	0.0018%	0.012%	0.18%	6.5%

The simulated sensitivities in Figure 3.22(a) show that with a full-scale input signal, the SNR drops prematurely, due to the presence of the modulated signal close to $f_s/2$. Figure 3.23 shows the output of the modulator with a capacitor mismatch of 10%. The modulated replica of the full-scale signal appears around $f_s/2$. With such a large mismatch the power of this unwanted signal is great and overloads the quantizer giving rise to an important harmonic distortion. The system is no longer linear and degradation deviates from the prediction of Equation (3.32). Instead,

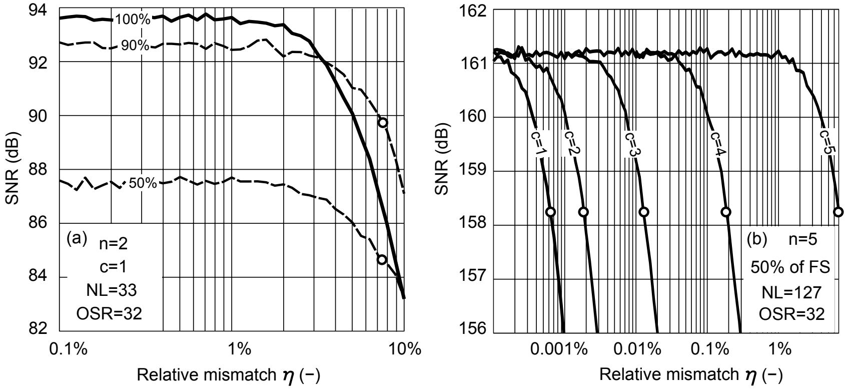


Figure 3.22 Simulated sampling capacitor mismatch sensitivity. Second- (a) and fifth-order (b) modulators. The −3dB drops are highlighted with a dot.

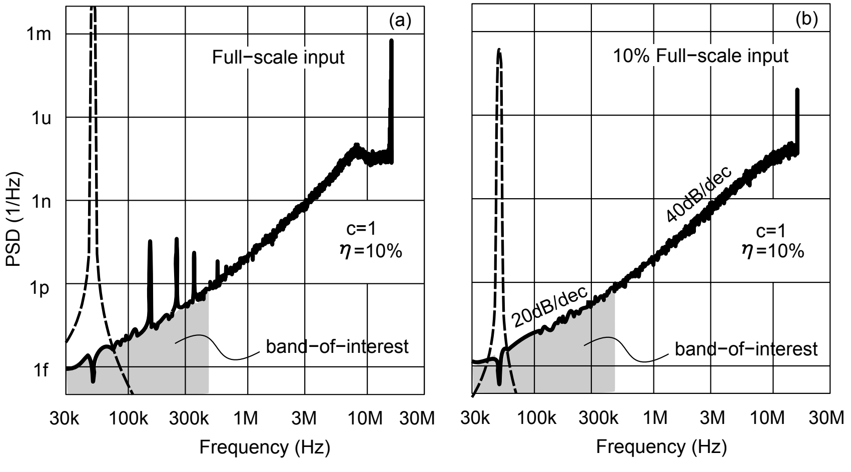


Figure 3.23 Modulator output with a capacitor mismatch of 10%.

with a reduced amplitude, the degradation is dominated by the quantization noise modulation. For this second-order case, the −3dB drop is expected and occurs at 7.6%. Such capacitor matching is easily achieved. Figure 3.22(b) together with Table 3.4 allows us to compare the simulated and calculated sensitivities, showing how accurately Equation (3.35) can predict the SNR degradation.

The mismatch sensitivity η of Equation (3.35) is plotted in Figure 3.24 for different NL providing a targeted SQNR. The corresponding OSR is determined by Equation (2.39) for the different orders n.

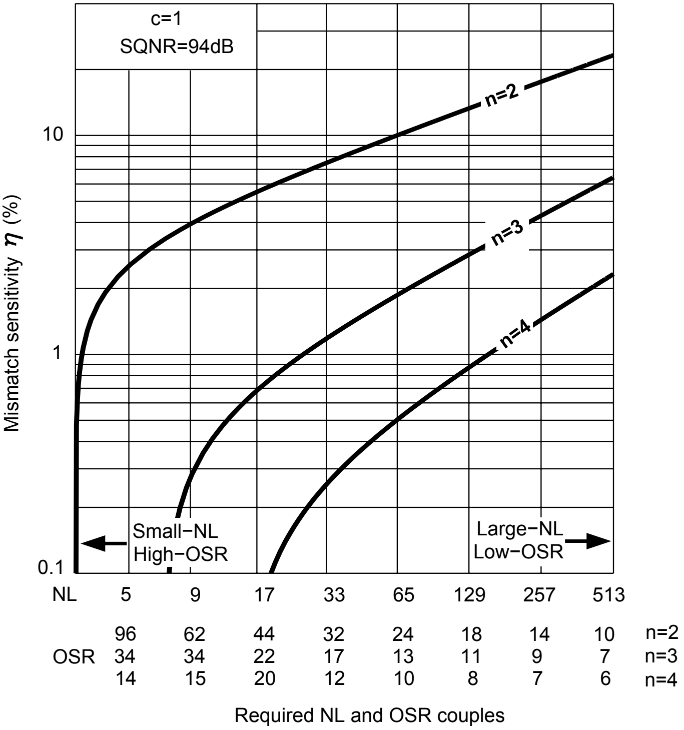


Figure 3.24 Sampling capacitor mismatch sensitivity for 94dB of SQNR.

The curves highlight the loss of efficiency of a first continuous-time integrator case, where $c = 1$, as we choose higher-order modulators.

3.6 Conclusions

A low-power design strategy was presented. The limitations were addressed while other specific aspects will be analyzed in further dedicated chapters. Without losses of generality, examples were provided in view of the case study of Chapter 6.

The chapter allows us to conclude that hybrid continuous-discrete-time architectures are well suited for low-power applications. Increasing the internal number-of-bits can significantly reduce the sensitivity to clock jitter. The analytical model developed in this chapter revealed the existence of an optimal solution for a moderate number-of-levels and oversampling ratios. The analysis showed that the higher the shaping order, the easier the timing errors transfer to an equivalent noise, degrading the modulator resolution.

With the hybrid architecture the sampling process has moved inside the modulator, thus partially benefiting from the spectral shaping. A full-clock-cycle scheme therefore becomes possible. The modulation of quantization noise introduced by the capacitor mismatch was analyzed and its impact on performance estimated.

Chapter

4

Auto-ranging algorithm

This chapter starts with the observation that in low-pass multi-bit modulators a large number of comparators remain periodically unused. The different tracking techniques proposed throughout the literature to reduce the quantizer size and consumption are compared. A novel technique, called auto-ranging, is presented and analyzed in detail in the second section. The third section studies all the setup aspects together with the limitations of the technique. The algorithm principle is applied to a conventional first-, second- and third-order modulator. The fourth section provides the design and optimization method based on an analytical model. A realistic second-order case, whose realization is the subject of the next chapter, is analyzed together with its extendibility to different applications.

4.1 Observations

4.1.1 Slow and fast components

The typical digital output signal of a low-pass multi-bit modulator displayed in Figure 4.1 leads to an interesting observation. As already shown throughout Section 2.2.3, the output is a superposition of two components: the *slow large-swing* input signal and the *fast small-swing* quantization noise.

This observation is quite representative of the quantizer activity. In fact, from sample-to-sample, only a small group of comparators, called here *active group*, senses variations. All the comparators below this group stay at logical one and the comparators above stay at logical zero. This

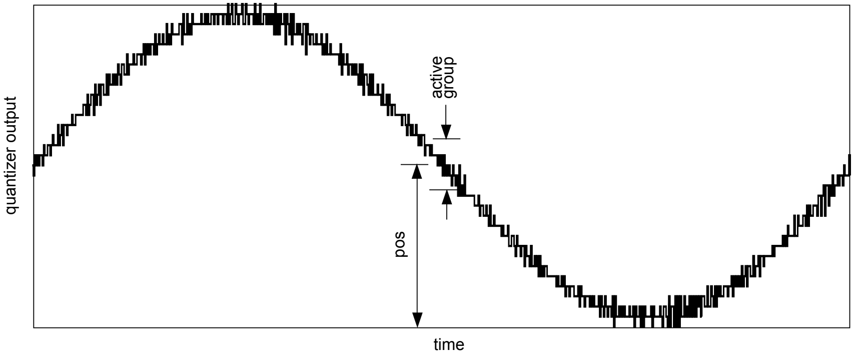


Figure 4.1 Typical output signal of a low-pass multi-bit ΔΣ-modulator. This example is a second-order 33-level modulator. The input is a sine wave 640 times slower than the clock signal.

situation remains unchanged for many clock-cycles. Therefore, the *active group* is sufficient to obtain the unpredictable quantization error sequence.

In contrast, the *position* of this group moves slowly with the input signal and can therefore be predicted. The output of the *active group* also carries information about the tendency of the slow component. In fact, as the signal either rises or falls, either the upper or the lower edge of the *active group* window starts changing. This information can therefore be used to predict the *position* of the group. However, let us recall that this superposition is correct as long as the linear approximation holds.

4.1.2 Exploiting the phenomenon

Figure 4.2 illustrates three ways of exploiting this interesting quantizer output composition. The simplest technique consists in turning off all the comparators that are not part of the active group. Given the predicted future position of the group, the comparators in the future active group progressively turn on as the others turn off. This first approach enables a reduction of the quantizer consumption to the strictly necessary amount. Nevertheless, the whole bank of comparators still needs to be implemented and ready for activation. Therefore, our wish to increase NL in the modulator is still limited.

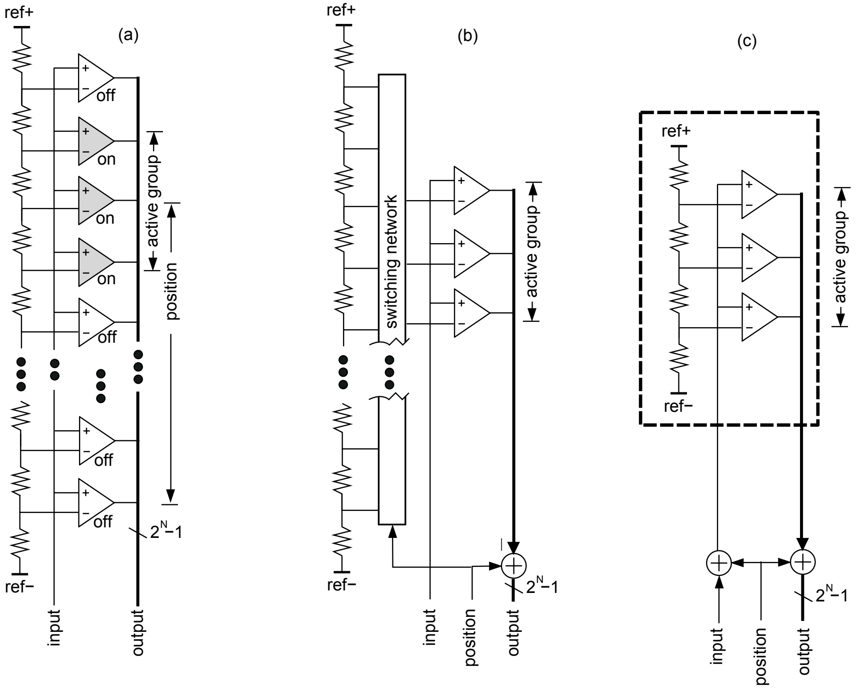


Figure 4.2 Different methods: (a) conventional ADC with only the active group turned on, (b) switching references proposed by [Lu04] and [DKG$^+$05] and (c) signal shifting proposed here as an extension of [Zie00].

However, we should notice that at any time the whole bank of comparators can be turned on to allow the modulator to recover from a transient unstable condition or at start-up. With this flexibility, we can keep a robust quantizer and start turning the comparators on and off when the application is battery-powered.

The second approach, illustrated in Figure 4.2(b), was presented in [Lu04] and [DKG$^+$05]. It consists in building only the *active group* of comparators. The complete bank of references is necessary. A switching network selects the references corresponding to the *position* of the *active group.* A large part of the chip area on silicon might be spared. Nevertheless, the presence of the entire NL-resistor reference ladder and a complex switching network are limiting factors. Using a large number-of-levels is impractical, especially if we consider a voltage supply as low as 1.2V. In fact, the reference voltage, represented here by **ref+** and **ref–**, is divided by NL. The random comparator offset is likely to become dominant over these small quantization steps. In such a case, the resolution of the quantizer is determined by the offset of the comparators and no longer by the number-of-levels.

This technique is commonly referred to as *tracking quantizer* since the principle consists in following the slow component signal, namely the group position, with the comparator references.

4.1.3 Auto-ranging technique

In this work, we propose a different approach, similar to the solution published by [Zie00] which uses a single-bit first-order modulator. As illustrated in Figure 4.2(c), again only the *active group* of comparators is really implemented. But this time only the references of the *active group* are necessary. The analog switching network is no longer required.

The technique consists in shifting both the analog input and the digital output of the quantizer by exactly the same amount. The few comparators implemented constantly sense the fast unpredictable signal component and, as the group *position* changes, the *range* of the quantizer is adapted by shifting. The quantizer itself is a conventional flash ADC exploiting the whole voltage swing between **ref+** and **ref–**.

The quantizer output provides a digital algorithm with enough information to automatically shift the signal range. Because of this, we refer to this technique as the *auto-ranging* (AR) algorithm.

4.2 Principle in detail

4.2.1 Last integrator reuse for signal shifting

As shown in Figure 4.3, the quantizer input and output will be shifted by the current *range* value. That amount can therefore be stored and updated to track the slow input signal. This interesting observation leads to the idea of reusing the analog integrator placed in front of the quantizer. By adding digital pulses at the input of the last feedback DAC, we can

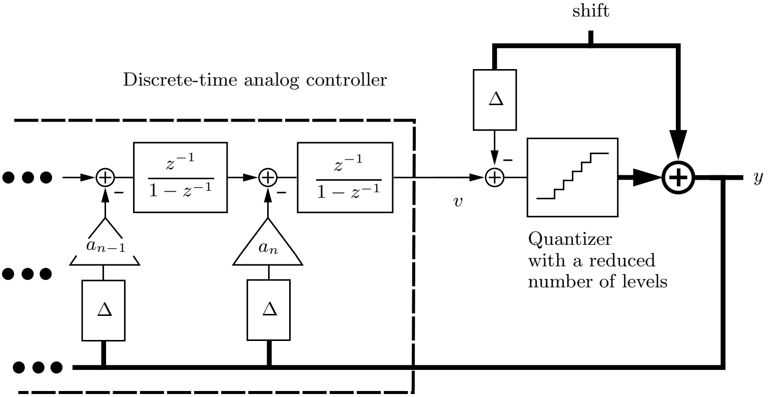


Figure 4.3 Signal shifting process. The bold and light lines represent digital and analog signal processing respectively.

generate and apply an analog shift at the quantizer input. The same operation can be performed in the digital domain at the quantizer output. This process is illustrated in Figure 4.4.

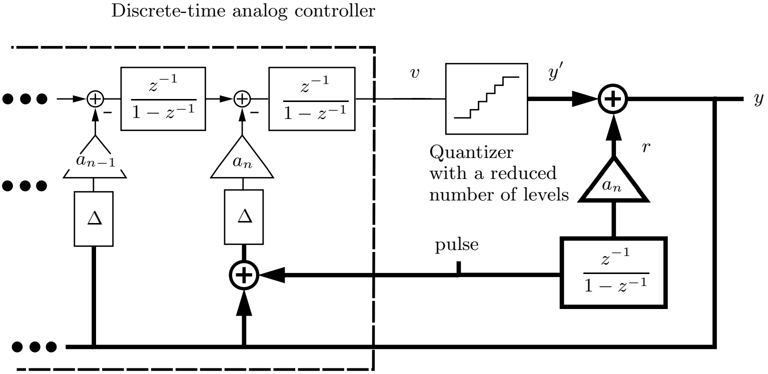

Discrete-time analog controller
$\frac{z^{-1}}{1-z^{-1}}$
$\frac{z^{-1}}{1-z^{-1}}$
a_{n-1}
a_n
Δ
Δ
v
y'
y
r
Quantizer
with a reduced
number of levels
a_n
pulse
$\frac{z^{-1}}{1-z^{-1}}$

Figure 4.4 Feedback path reuse. The bold and light lines represent digital and analog signal processing respectively. The dashed area shows the existing processing being reused.

As shown in Figure 4.5, a digital algorithm is added to control the shifts. Consequently, the AR technique does not require any additional

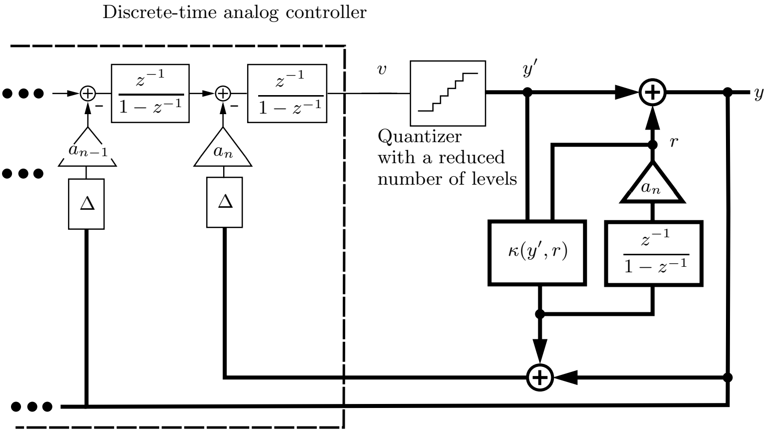


Figure 4.5 Generation of the pulses based on the quantizer output. The bold and light lines represent digital and analog signal processing respectively. The dashed area shows the existing processing being reused.

analog circuitry. Given a $\Delta\Sigma$-modulator with a high number-of-levels, the comparator and reference bank can be reduced. Additionally, a separate mismatch shaping digital encoder is necessary since the signal of the last feedback is no longer identical to the others. Moreover, the quantizer input range can be extended to the limits given by the power supply.In other words, the size of the quantization steps Δ is increased, which alleviates the offset constraints on both the reference ladder and the comparators themselves.

The modulator internal number-of-levels remains unchanged and the DAC still has to reproduce them. Unlike the quantizer, the DAC consists of a bank of NL elements all taking advantage of the full voltage swing. As a consequence, the AR technique allows a large number-of-levels without the usual voltage supply restriction and increase in quantizer size and power consumption.

By using existing analog blocks and only adding limited digital processing, this AR technique is better suited than that of [DKG+05] to deep sub-micron processes.

4.2.2 Control function

We now need to determine the control sequence s based on knowledge of the quantizer output. This link is represented in Figure 4.6 by the *control function* $\kappa(y')$. A simple and efficient method of preventing the quantizer from overloading consists in generating s equal to the quantizer output. In this way, at each clock-cycle the signal, processed by the quantizer, moves into the center of the quantization window.

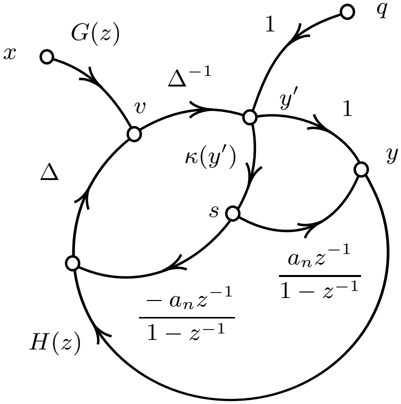


Figure 4.6 General linear Signal Flow Graph representation of the single-stage modulator of Figure 2.18 with the auto-ranging algorithm.

This corresponds in the signal flow graph in Figure 4.6 to $\kappa(y') = y'/a_n$. In such an ideal case we can determine the new expressions of $Y(z)$ and $V(z)$ as

$$Y(z) = \underbrace{\frac{G(z) \cdot \Delta^{-1}}{1 - H(z)}}_{=\mathrm{STF}(z)} X(z) + \underbrace{\frac{1}{1 - H(z)}}_{=\mathrm{NTF}(z)} Q(z)\ , \tag{4.1}$$

$$V(z) = \underbrace{\frac{G(z) - G(z)z^{-1}}{(1 - H(z))(1 + z^{-1})}}_{=\Delta \cdot \mathrm{STF}(z)\frac{1-z^{-1}}{1+z^{-1}}} X(z) + \underbrace{\frac{(H(z) + H(z)z^{-1} - 2z^{-1}) \cdot \Delta}{(1 - H(z))(1 + z^{-1})}}_{=\Delta \cdot \mathrm{NTF}(z)\frac{1-z^{-1}}{1+z^{-1}} - 1} Q(z)\ . \tag{4.2}$$

The overall transfer functions NTF(z) and STF(z) remain unchanged, but the transfer functions to node v now have a pole at $f_s/2$. To prevent the system from becoming unstable, we should introduce a *dead-zone* in the control function therefore deactivating the auto-ranging loops for small amplitude signals. In such a case, neither the quantization noise from node q nor small signals from x will be *seen* by the algorithm. Besides, we intuitively understand that if signals are small in amplitude, the algorithm is no longer necessary.

Because we are reusing the last integrator path, we are forced to make shifts that are multiples of a_n. Rather than describing this mathematically, the control function is represented graphically. Figure 4.7 shows three examples for a first-, second- and third-order modulator. The assumption is made that all the NTF poles are at the center of the z-plane. As shown in Table 2.1, the feedback coefficients are in a Pascal triangle configuration. Therefore the last coefficient is equal to the modulator order n. The functions κ_1, κ_2 and κ_3 are therefore configured to bring the quantizer input into the center of the quantization window. The functions are displayed only for quantizer outputs up to ± 8. The choice of the number of implemented levels causes either an extension or a reduction of the range.

It is interesting to note that the *dead-zone* required for stability is naturally present in the control function. As mentioned earlier, the slope has to be $1/a_n$ and the function provides integer control. As a result, the *dead-zone* is $2a_n$ wide, which is sufficient for the three cases studied here. According to the discussion in Chapter 2, the quantization noise remains within the total range of 2^n. The *dead-zone* would be artificially enlarged with high-order modulators to prevent instability.

As the output y reaches either its maximum or minimum, determined by the number-of-levels NL, the control function is overridden. The shift control s has to be limited to prevent y from going beyond $\pm(\text{NL}-1)/2$. The next section gives more insight into this particular feature, especially when describing the hardware implementation of the algorithm.

4.2.3 Implementation constraints

There are several constraints limiting the choice for the auto-ranging parameters. First of all, the last feedback coefficient a_n has to be an integer

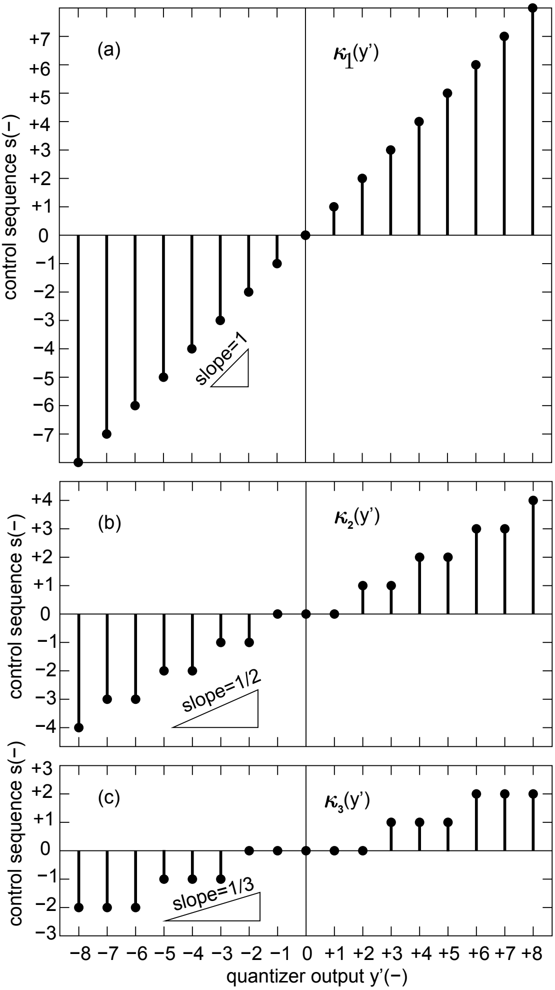


Figure 4.7 Control functions $\kappa(y')$ for a first- (a), second- (b) and third-order (c) modulator. The slope shown corresponds to the inverse of the last modulator feedback coefficient.

value. This is mandatory if we want to be able to reproduce exactly the same signal shift on both the analog and digital side. Secondly, as the internal number-of-levels NL is odd, the reduced number-of-levels NR has to be odd too. Now, since the modulator output y is always equal to the sum of the quantizer output y' and the range r provided by the algorithm, the maximum of r can be calculated:

$$\underbrace{\frac{\mathrm{NL}-1}{2}}_{\|y\|_\infty} = \underbrace{\frac{\mathrm{NR}-1}{2}}_{\|y'\|_\infty} + \|r\|_\infty \Rightarrow \|r\|_\infty = \frac{\mathrm{NL}-\mathrm{NR}}{2}\,. \tag{4.3}$$

Since both NL and NR are odd integers $\|r\|_\infty$ is an integer. This number needs to be a multiple of a_n in order to allow emulation of all the NL levels. As an example, a second-order modulator with $a_n = 2$ and $\mathrm{NL} = 33$ can work with any even $2^n - 1 < \mathrm{NR} < \mathrm{NL}$. But considering high-order modulators, these constraints put together drastically restrict the degrees of freedom. The same example with $n = 3$ and $a_n = 3$ leaves only the possibilities $\mathrm{NR} = 15, 21, 27$.

4.2.4 Number of step changes

As we choose to reduce the number of comparators, the question arises of how many levels should be kept . To give an answer, we start our analysis by calculating the number of steps γ taken from one clock cycle to the other at the quantizer input v. In a way similar to the low-pass modulator output in Figure 4.1, Figure 4.8 illustrates the signal swing at the quantizer input predicted in Section 2.2.3 by Equation (2.22).

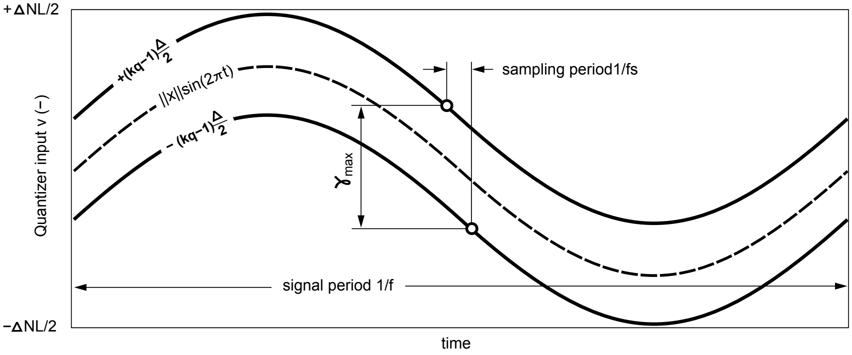


Figure 4.8 Quantizer input signal boundaries.

In a non-overloading situation, the quantizer input should remain within two sine waves $\|x\|\sin(2\,\pi\,ft) \pm \Delta(k_q-1)/2$. We can show that the maximum γ occurs at the zero-crossing of the middle sine wave. It follows that

$$\gamma = \|x\| \sin(\theta + \tfrac{\pi f}{f_s}) - \|x\| \sin(\theta - \tfrac{\pi f}{f_s}) + \Delta(k_q - 1)\ . \tag{4.4}$$

$$\frac{\partial \gamma}{\partial \theta} = 0 : \quad \sin(\theta) \cos(\tfrac{\pi f}{f_s}) = 0\ . \tag{4.5}$$

We can therefore write the expression

$$\gamma_{\mathrm{max}} = \|x\|\, 2 \sin\left(\tfrac{\pi f}{f_s}\right) + \Delta(k_q - 1)\ . \tag{4.6}$$

If we substitute Equation (2.29) in this last relationship, normalize to Δ and further round to the next integer by the ceiling function $\lceil . \rceil$, we obtain an absolute limit:

$$\boxed{\gamma_{\mathrm{max}} = \left\lceil (\mathrm{NL} - k_q + 1) \sin\left(\frac{\pi}{2\ \mathrm{OSR}}\right) + k_q - 1 \right\rceil}\ . \tag{4.7}$$

Figure 4.9 compares the simulation results and the predictions of Equation (4.7). The predictions give an absolute limit that is sometimes not reached. As the order becomes higher the probability of generating an alternative quantization error sequence of ±1/2 becomes very small.

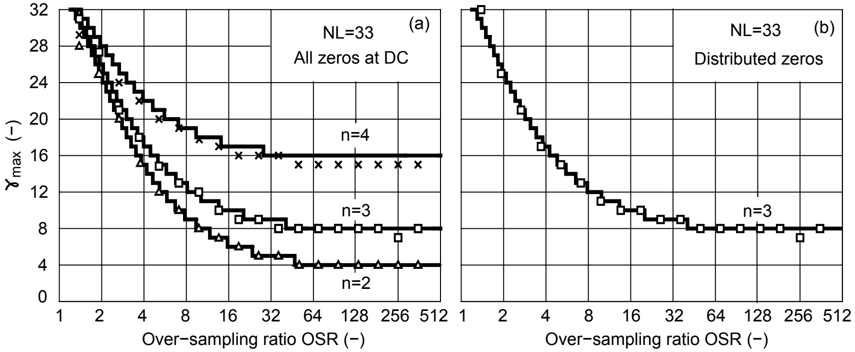


Figure 4.9 Maximum number of steps seen by a 33-level ADC: (a) all the NTF zeros are located at DC, (b) one zero located at DC and two conjugated zeros at 0.77 times the bandwidth.

In Figure 4.9(b), the same equation is plotted for an NTF with its zeros optimally distributed on the unity circle. For a third-order modulator, this means, according to [SRNT97], one zero at DC and a conjugated pair at 0.77 times the bandwidth f_b. The difference, compared to the results with the zeros at DC, is almost imperceptible, which shows that the position of the zeros has little impact on $\gamma_{\max}$. The superimposed points are simulated with MATLAB. They show how accurately Equation (4.7) predicts the maximum number of steps even at low OSRs.

Note that if the OSR is higher than approximately NL, the highest possible step change between two samples is $\pm 2^n$. This baseline in the curve is reached when the quantization noise dominates the variations seen by the quantizer. As already highlighted by Figure 2.20, the higher the order, the more quantization room is required.

4.3 Realization and limitations

4.3.1 Simulations

We chose here to analyze a 33-level modulator with first-, second- and third-order spectral shaping. The simulation codes for these three cases are listed in Appendix D.1. They implement the control functions described in Figure 4.7. An excerpt of the second-order case is reproduced in Code 4.1. The modulator output `outr` is reconstructed as the addition of the quantizer output `out` and the range register `rng`. As explained previously, the last feedback signal `outp` consists of the modulator output with an additional pulse sequence `ctrl` which indirectly generates the shifts at the quantizer input. The same shifting is reproduced in the range register with a factor of two, corresponding to the last feedback coefficient.

Code 4.1 Excerpt of the MATLAB simulation code of the second-order modulator with auto-ranging.

```
...
outr(i)=rng(i) +out(i);
outp(i)=outr(i)+ctrl(i);
rng(i+1)=rng(i)+ctrl(i)*2;
...
```

Figures 4.10, 4.11 and 4.12 compare the simulation results for the three cases with a 15-level quantizer in a 33-level modulator.

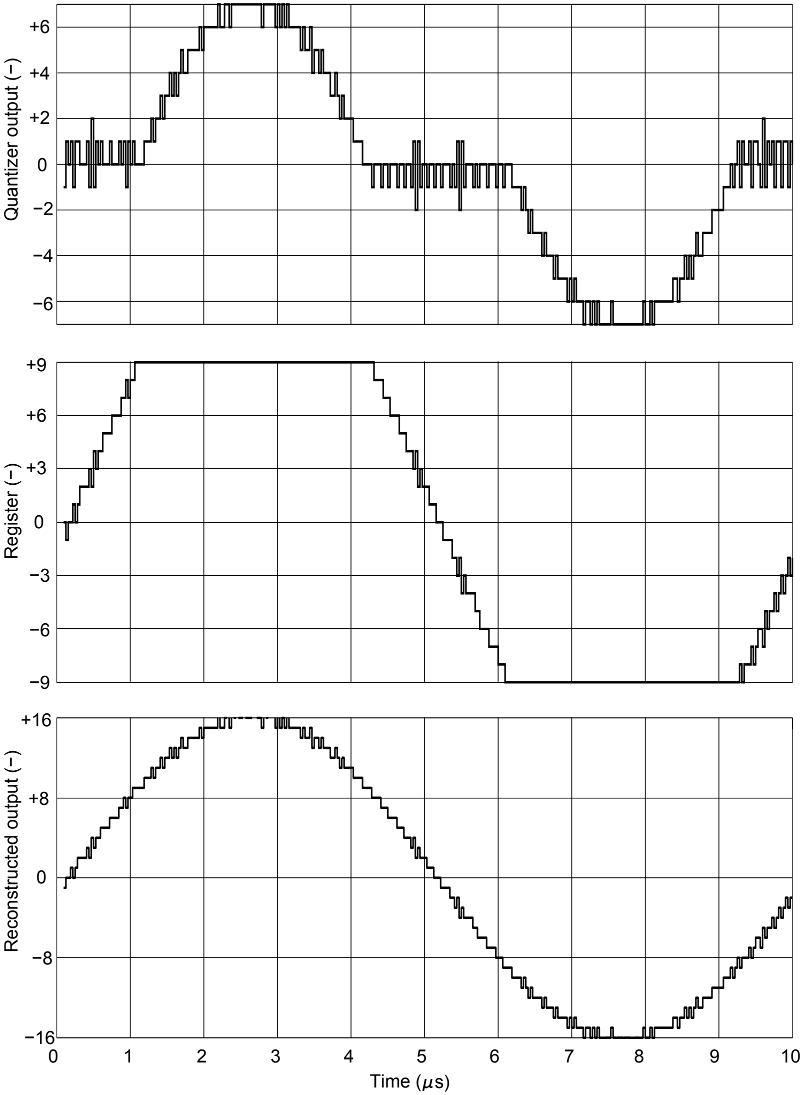


Figure 4.10 Simulation results for the first-order 33-level modulator with a 15-level quantizer.

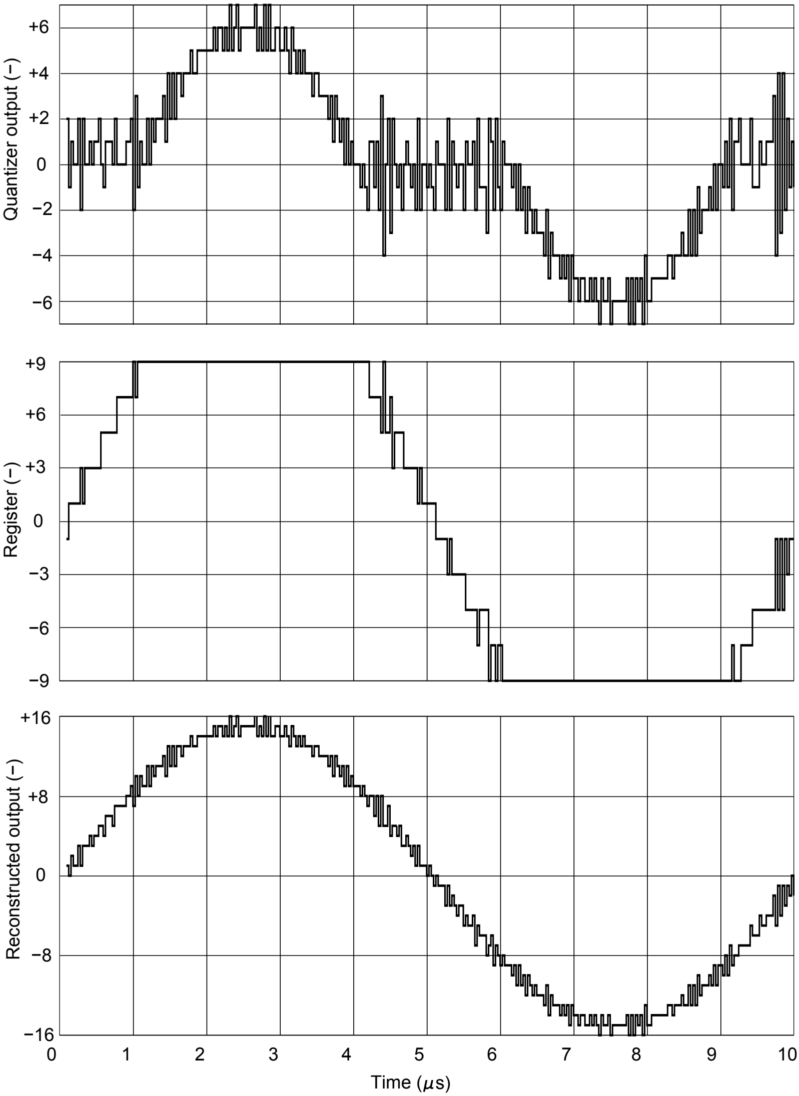


Figure 4.11 Simulation results for the second-order 33-level modulator with a 15-level quantizer.

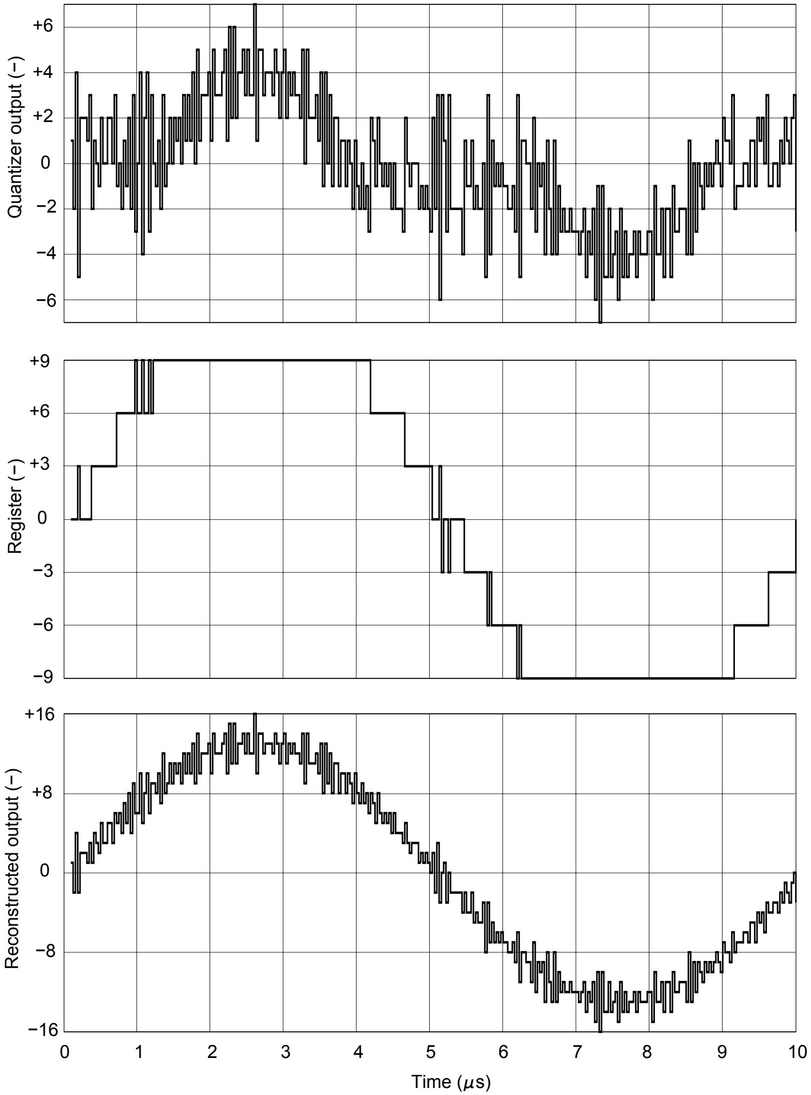


Figure 4.12 Simulation results for the third-order 33-level modulator with a 15-level quantizer.

The case NR=15 is in fact realizable, according to the constraints summarized in Section 4.2.3, for the three cases.

The input signal is a full-scale 100kHz sine wave sampled at 32MHz, where the *full-scale* is chosen according to the definition given by Equation (2.29). The comparison highlights the increased activity in the quantizer with the shaping order n. In the case of first-order modulation, the reduced number-of-levels is still conservative for this input signal. The number-of-levels could be further reduced to reach the situation of the third-order case where the comparators change their state more often.

In all three cases, we recognized the shaping order by the number of steps taken by the range register of one, two and three, respectively. Let us recall that this constraint is imposed by the last feedback coefficient being equal to n for the case studied here.

The reconstructed outputs also highlight the limitation of in input signal dynamic range versus the modulator order for a given NL. The higher the order, the more room necessary to accommodate the quantization noise. Auto-ranging can be seen as a technique that splits the modulator output into a rough and a fine representation. The algorithm task is to guarantee that together they represent the modulator output without errors.

4.3.2 Input signal constraints

The simulation codes presented previously allow one to determine the performance of the algorithm with different parameters. In particular, let us consider the following experiment. The sampling frequency is set at 32MHz and the band-of-interest is 500kHz. Again three cases with a first-, second- and third-order shaping are analyzed. For each case, an in-band tone at 100kHz is applied with an amplitude 40dB lower than the full-scale. An additional full-scale out-of-band tone is applied at different frequencies. This situation is fairly representative of a realistic situation in a communication system application. The additional tone simulates the presence of a large interferer which would not be enough attenuated.

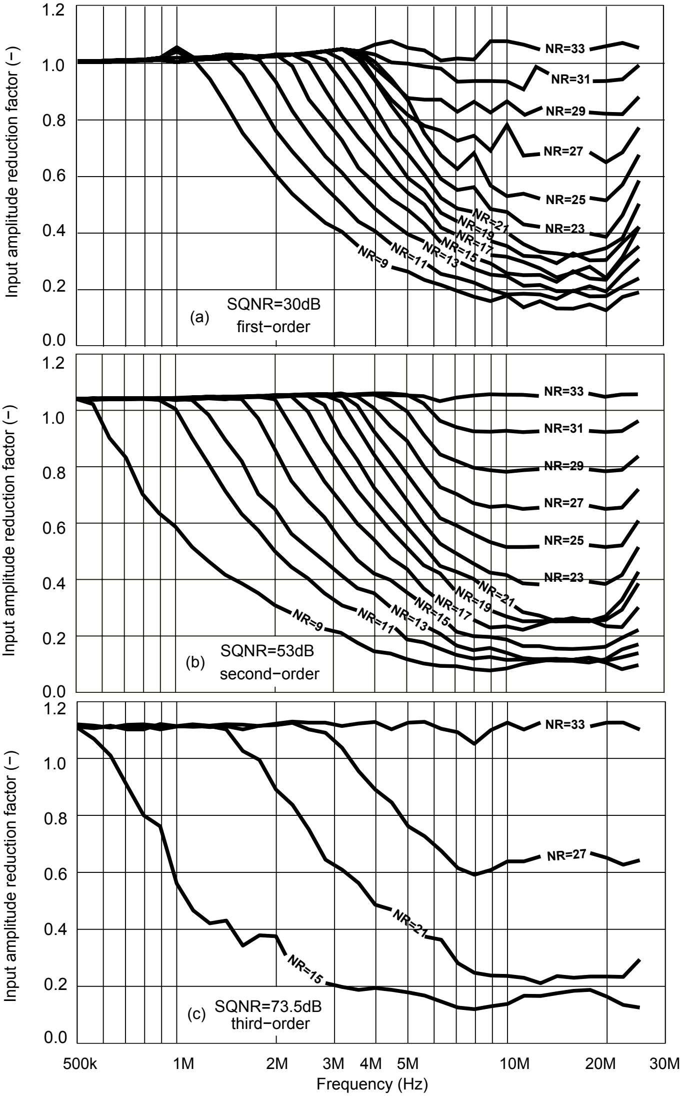


Figure 4.13 Frequency limitation of a 33-level (a) first-, (b) second- and (c) third-order modulator for different reduced number-of-levels NR. The sampling frequency is set at 32MHz and a band-of-interest of 500kHz is considered.

On the other side, the in-band tone is the signal of interest to be further processed in the system. The presence of the interferer may compromise the algorithm operations. In fact, as mentioned earlier, if the amplitude-frequency product is too high, the reduced size quantizer overloads and the algorithm does not work properly anymore.

Figure 4.13 shows the necessary interferer amplitude reduction to keep the SQNR unchanged. As expected, the SQNR's are about 40dB lower than the maximum predicted by Equation (2.37). The code presented previously is placed into a search loop to determine accurately the attenuation at each frequency. Constraints on the search algorithm convergence forced us to aim at resolutions slightly lower than the maximum SQNR of –40dB. For this reason, in some cases the reduction factors are slightly higher than unity for low frequencies. The case with NR=33 is equivalent to not having any auto-ranging at all. As expected, this case is not limited by the full-scale interferer frequency.

These curves are meant to help the designer in choosing the appropriate reduced number-of-levels according to the overall system requirements. For instance, the second-order case with NR=11 allows accommodation full-scale out-of-band signals up to twice the band-of-interest. Interferers with frequencies higher than 1MHz should be progressively attenuated by the anti-alias filter in front of the modulator.

Figure 4.14(a) shows the dynamic range plot performed with a low frequency single tone. The dynamic range plot is perfectly linear and, as expected with multi-bit modulators, the dynamic range is equal to the maximum SQNR. Figure 4.14(b) shows the output PSD with two half-scale out-of-band tones. These results do not reveal any harmonic distortion.

4.3.3 Sensitivity to circuit imperfections

The auto-ranging principle relies on the ability to reproduce accurately the same shift in both the analog and digital side of the quantizer. For the class of modulators studied here, the last feedback coefficient a_n is an integer equal to the modulator order n.

The digital shifts are guaranteed to be exact multiples of a_n. In contrast, the accuracy of the analog shifts depends on the precision of this

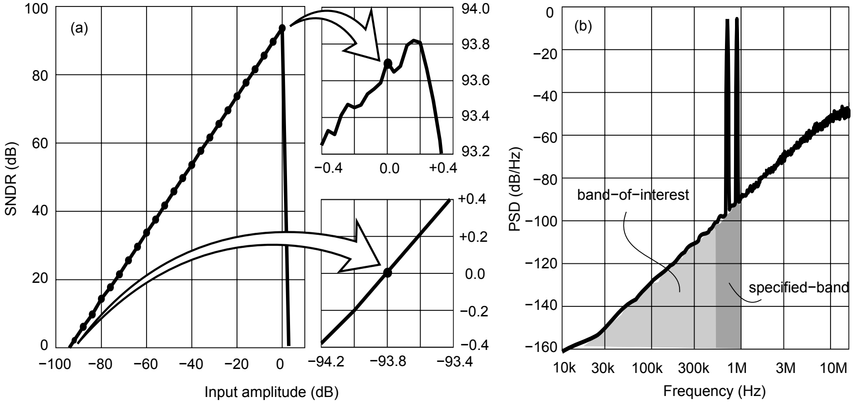


Figure 4.14 33-level second-order modulator with a 11-level quantizer. Simulated (a) dynamic range plot and (b) output PSD with two half-scale input tones at 700kHz and 900kHz. The tones are place within the band, from 500kHz and 1MHz, specified by the plot of Figure 4.13. The dynamic range plot is performed with a 50kHz single tone.

last feedback processing chain, including the DAC, the integrator and the quantizer.

Figure 4.15(a) shows the sensitivity to the overall coefficient for a second-order case. The maximum SQNR of the example analyzed is 94dB. The error on the overall coefficient should therefore stay below 2% to maintain 93dB of resolution. Such a stringent tolerance requires this feedback path to be implemented with a switched-capacitor circuit. Furthermore, special care has to be taken in order to match the DAC and quantizer slopes of Δ and $1/\Delta$, respectively.

The accurate reproduction of the shift also relies on having the same integrator behavior in the analog and digital side of the quantizer. Once again, the digital accumulator is guaranteed to perform the exact operation. In contrast, the analog integrator has a finite DC gain. Figure 4.15(b) shows the sensitivity to the integrator DC gain. To achieve 93.5dB of resolution, more than 60dB of gain is necessary for the case with an 11-level quantizer.

The last imperfection studied here is the offset of the comparators . Figure 4.16 shows the offset sensitivities with and without auto-ranging.

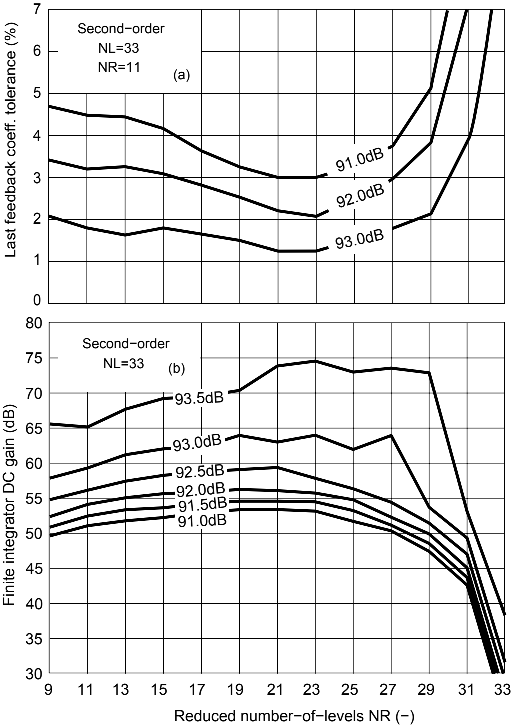


Figure 4.15 Last feedback path sensitivity to imperfections in a 33-level second-order modulator with different reduced number-of-levels NR. The tolerance of the overall coefficient (a) and finite DC gain (b) of the integrator are shown. The simulations are performed with an input signal at 100kHz sampled at 32MHz and considering an OSR of 32.

The case without reduction is slightly more sensitive since it involves more comparators. On the contrary, as mentioned earlier, the reduced size quantization window is enlarged to cover the full output swing of the last amplifier. Therefore, with auto-ranging, the limiting offset of a few percents applies to larger quantization steps Δ. Considering quantization steps of 80mV, offsets inferior to 2mV are necessary. Such a precision can be achieved by enlarging the input transistor pair, or with the capacitive coupling compensations presented in [Raz01, JM97], or with the digital compensation techniques proposed by [PK06]. This subject is further analyzed in the frame of a design example in Chapter 6.

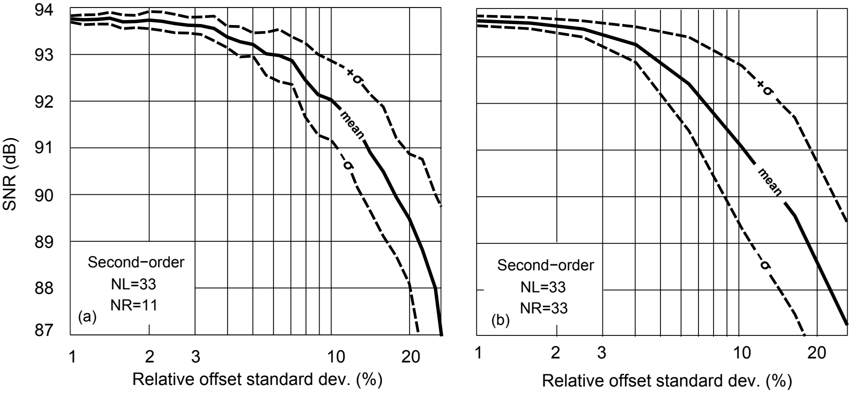


Figure 4.16 Tolerance on the statistic offset of the comparators for the second-order case with (a) and without (b) auto-ranging.

4.3.4 Hardware implementation

Figure 4.17 shows the low-level hardware description of the auto-ranging algorithm. According to the block diagram in Figure 4.4, the algorithm requires three signed adders, a register and the control function. The debubbler and the binary encoder are necessary even without auto-ranging. Nevertheless, they are now smaller in size. At the output of the debubbler, all the bits are equal to zero except the one representing the position of the thermometer code. This simplifies the logic of the encoder and the control function. Meanwhile, the positive and negative quantizer saturation signals, called here **max** and **min**, are provided by the last debubbler output and the NOR applied to all of them.

Table 4.1 shows the decision scheme for the 33-level second-order design as analyzed in Chapter 6. This scheme implements the control function κ_2 described earlier by Figure 4.7 for 11 reduced levels. As explained previously, the control function is overridden if the shift chosen causes the emulated 33-level quantizer to overload. The control function without this saturation mechanism is highlighted by bold numbers. This function slightly differs from κ_2 in that the edge controls are ± 3 instead of ± 2. This change provides a small improvement in the input frequency limitation. In fact, this level is activated only when tracking the fastest

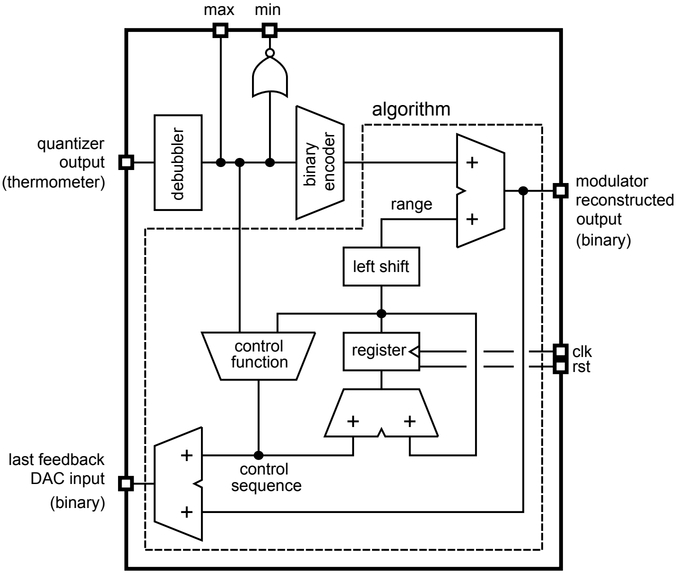


Figure 4.17 Hardware implementation of the algorithm.

full-scale signal.

The **min** and **max** bits control the *auto-reset* mechanism which is shown in Figure 4.18. Two shift registers store the history of **min** and **max**. The reset **rst** is automatically activated if the quantizer goes from its lowest to its highest level within one, two or three clock-cycles, and vice-versa. The mechanism also detects if either **max** or **min** is high for at least four clock-cycles. If activated, the reset bit stays high for two clock-cycles.

Code 4.2 Additional MATLAB code to simulate the auto-reset circuit.

```
...
mx=(NR-1)/2;
if ((i>5) & (...
((out(i)==+mx)&(out(i-1)==+mx)&...
         (out(i-2)==+mx)&(out(i-3)==+mx))|...
((out(i)==-mx)&(out(i-1)==-mx)&...
         (out(i-2)==-mx)&(out(i-3)==-mx))|...
((out(i)==+mx)&(out(i-2)==-mx))|...
((out(i)==-mx)&(out(i-2)==+mx))   ))
```

Table 4.1 Shift control signal.

12-level	11-level quantizer output										
register	0	1	2	3	4	5	6	7	8	9	10
11	–3	–2	–1	–1	0	0	0	0	0	0	0
10	–3	–2	–1	–1	0	0	0	+1	+1	+1	+1
9	–3	–2	–1	–1	0	0	0	+1	+1	+2	+2
3...8	**–3**	**–2**	**–1**	**–1**	**0**	**0**	**0**	**+1**	**+1**	**+2**	**+3**
2	–2	–2	–1	–1	0	0	0	+1	+1	+2	+3
1	–1	–1	–1	–1	0	0	0	+1	+1	+2	+3
0	0	0	0	0	0	0	0	+1	+1	+2	+3

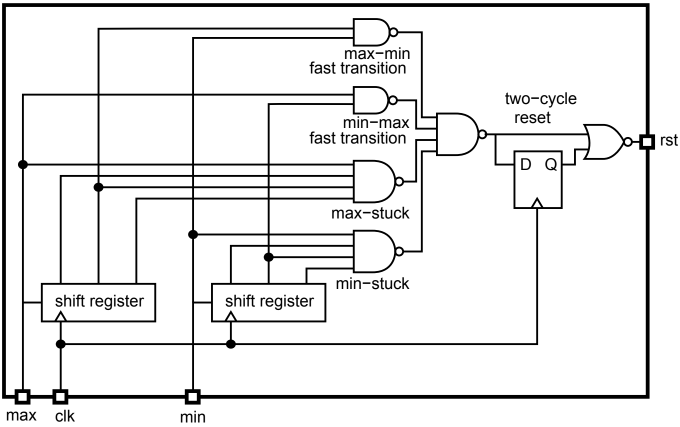


Figure 4.18 Automatic reset mechanism for recovery.

```
  sig(i,:)=zeros(1,3);
   out(i)=0;rng(i+1)=1;
  outp(i)=0;outr(i)=0;
  ctrl(i)=0;
end
...
```

Figure 4.19 shows the recovery process simulated with the additional Code 4.2. The simulation started with an input signal at its maximum level and the integrators output at their minimum value. The range register is set at its highest level. The mechanism described in Figure 4.18 resets the integrators and the algorithm memory many times until stable conditions are reached. The recovery is short and requires less than a quarter of the signal period. This mechanism is activated only when necessary to recover from an unstable situation.

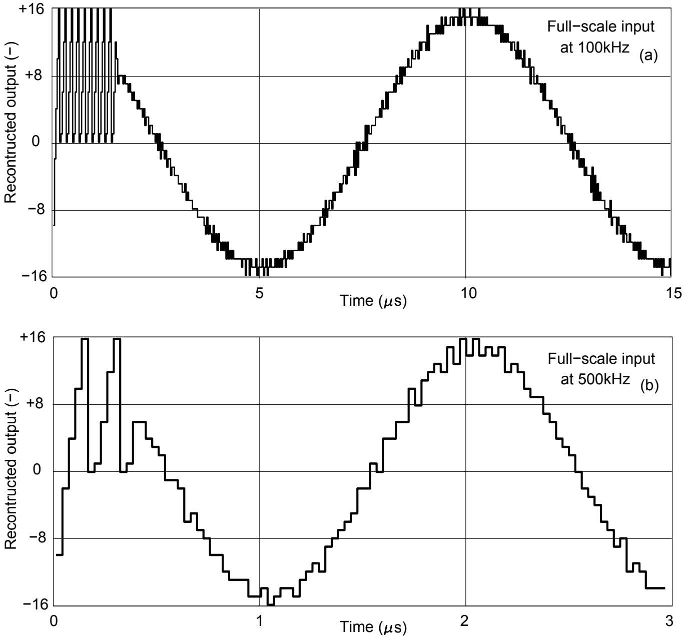


Figure 4.19 33-level modulator output showing the automatic recovery process with a full-scale input signal at 100kHz and 500kHz.

4.4 Optimization

4.4.1 Reduced number of quantization levels

The reduced number-of-levels should prevent the quantizer from loosing any information. The worst situation occurs with a full-scale input signal at the highest frequency. In particular, if the quantizer reaches either edge levels of the quantization window, the algorithm decreases the quantizer input by the maximum possible shift, which in this case is higher than NR/2. According to Table 4.1, if a rising input signal causes the quantizer to reach its highest level, a control shift of +3 is applied, which corresponds to a down-shift by 6 levels, as explained in Section 4.3.4. This shift brings the quantizer input at level 4, half a level below the quantizer window center.

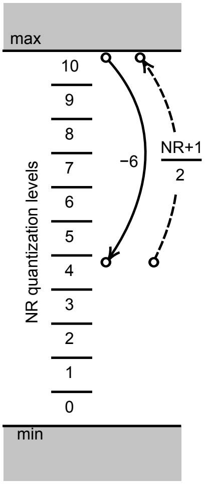


Figure 4.20 Illustration of the worst case situation with a large rising input signal at the maximum frequency. The algorithm performs consecutively maximum shifts of (NR − 1)/2. The quantizer overloads whenever its input goes beyond the maximum and minimum levels. The numbers represent the NR levels and the short lines the NR − 1 comparator thresholds.

The situation is further worsened if $\gamma_{\max}$ steps are taken more than once consecutively. As illustrated by Figure 4.20, the quantizer has (NR+1)/2 levels available to sense the next $\gamma_{\max}$ step. If the signal goes beyond the maximum, the quantizer overloads and part of the information is lost. The situation is the same for the falling signal case. We can therefore calculate the minimum reduced number-of-levels $\mathrm{NR}_{\min}$ as:

$$\mathrm{NR}_{\min} = 2\gamma_{\max} - 1 \ . \tag{4.8}$$

The efficiency of the auto-ranging technique can be evaluated by a reduction factor, defined here as the emulated-to-reduced number-of-levels ratio $\mathrm{NL}/\mathrm{NR}_{\min}$. We first substitute the expression $\gamma_{\max}$ from Equation (4.7) into Equation (4.8) and then, by inverting the resolution equation (2.37), we find the OSR as a function of SQNR, n and NL. As in the example studied, we consider a specified band of twice the modulator bandwidth f_b, which halves the OSR in Equation (4.7), and we find the reduction factor:

$$\frac{\mathrm{NL}}{\mathrm{NR}_{\min}} = \frac{\mathrm{NL}/2}{(\mathrm{NL} - 2^n + 1)\sin\left(\sqrt[2n+1]{\frac{3\pi(2n+1)(\mathrm{NL}-2^n+1)^2}{2\mathrm{SQNR}}}\right) + 2^n - 1.5}\,. \tag{4.9}$$

The reduction factor is plotted in Figure 4.21 for a second-order modulator as a function of the possible sets $\{\mathrm{NL}, \mathrm{OSR}\}$ providing the targeted SQNR of 94dB, according to Equation (2.37). The solid curve reveals the presence of a maximum in the auto-ranging efficiency with a reduction factor of 3, in the case of a specified bandwidth of $2f_b$. The chosen set of $\{33{,}32\}$ for this example is close to the optimum. In contrast, the dashed curve reveals a higher efficiency without the presence of out-of-band interferers. In this case, the appropriate choice would be NL=65, NR=13 with an OSR of 24.

4.4.2 Reduction factor

The parameter $\gamma_{\max}$ is related to the minimum number-of-levels the quantizer should keep. Let us define the *reduction factor* as the ratio between the number-of-levels, determined by Equation (2.38) and $\gamma_{\max}$. Figure 4.21 reveals, for a given SQNR, an optimal over-sampling ratio providing the maximum possible reduction of the quantizer size. The number-of-levels is also plotted on the same figure since the OSR determines NL at the same time, for a given SQNR.

The figure also highlights that as the order is increased, the potential for reduction decreases. In fact, the range of the quantization error is wider and more levels are necessary to sense that unpredictable signal.

4.4.3 Extendability

According to the closed-form expressions developped so far, both $\mathrm{NR}_{\min}$ and SQNR are functions of the modulator parameters n, OSR and NL. If any three of these variables are known, the remaining two are set. In this analysis of extendability SQNR is determined by the specifications

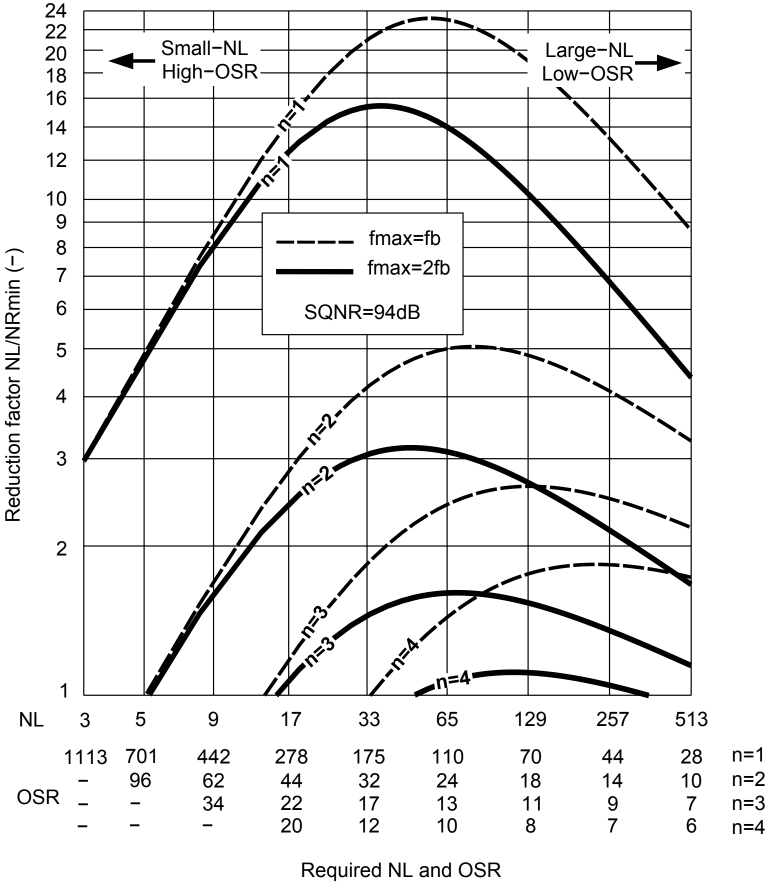


Figure 4.21 Auto-ranging design chart: *reduction factor* $\mathrm{NL}/\mathrm{NR}_{\min}$ for an SNQR of 94dB as functions of the required NL and OSR for the first-, second-, third- and fourth-order cases. The solid and dashed lines correspond to the cases where a maximum full-scale signal frequency is respectively twice or equal to the bandwidth.

of the targeted application, whereas $\mathrm{NR}_{\min}$ and n are determined by the topology choice. Under these considerations, OSR and NL are directly determined by the application.

Let us keep the low-pass second-order architecture with an 11-level ADC, which requires 2 amplifiers and 10 comparators. With voltage supplies from 1.8V down to 1.2V, this reduced number-of-levels is reasonable. By combining Equations (4.7) and (2.37) we can determine the required NL and OSR to reach a targeted resolution. Considering, as in this example, a specified band of twice f_b and choosing $\mathrm{NR} = 2\gamma_{\mathrm{max}} + 1$ we can write:

$$\mathrm{NL} = \frac{\frac{\mathrm{NR}+1}{2} - 2^n + 1}{\sin\left(\frac{\pi}{\mathrm{OSR}}\right)} + 2^n - 1\ . \tag{4.10}$$

It is common practice to design the modulator with an SQNR 10dB higher than the targeted SNR. This margin allows for circuit noise and distortion losses. Therefore, introducing Equation (4.10) in Equation (2.37) yields:

$$\mathrm{SNR} = \frac{3\pi(2n+1)}{20}\left(\frac{\mathrm{OSR}}{\pi}\right)^{2n+1}\left(\frac{\frac{\mathrm{NR}+1}{2} - 2^n + 1}{\sin\left(\frac{\pi}{\mathrm{OSR}}\right)}\right)^2\ . \tag{4.11}$$

By inverting the equation we find the required OSR. As a result, the sampling frequency is OSR times twice the band-of-interest f_b, and a minimum of $(\mathrm{NL} - 1)$ DAC elements are necessary. These two last relationships are plotted in Figure 4.22 for different targets.

Table 4.2 Extension to typical applications for a second-order modulator with NR=11.

	Performance		Requirement		Efficiency
Application	SNR (dB)	f_b (Hz)	NL	f_s (Hz)	NL/NR
Audio	98	22k	82	1.8M	7.5
GSM/EDGE	84	200k	33	13M	3
Bluetooth	84	500k	33	32M	3
WCDMA	74	2M	25	92M	2.3

For the dashed curves NL′ and OSR′ in Figure 4.22, a specified band equal to the band-of-interest is considered. These should be used for applications, such as audio conversion, where no large out-of-band signals are expected. Since the number of comparators is constant, the higher the NL the more efficient the auto-ranging.

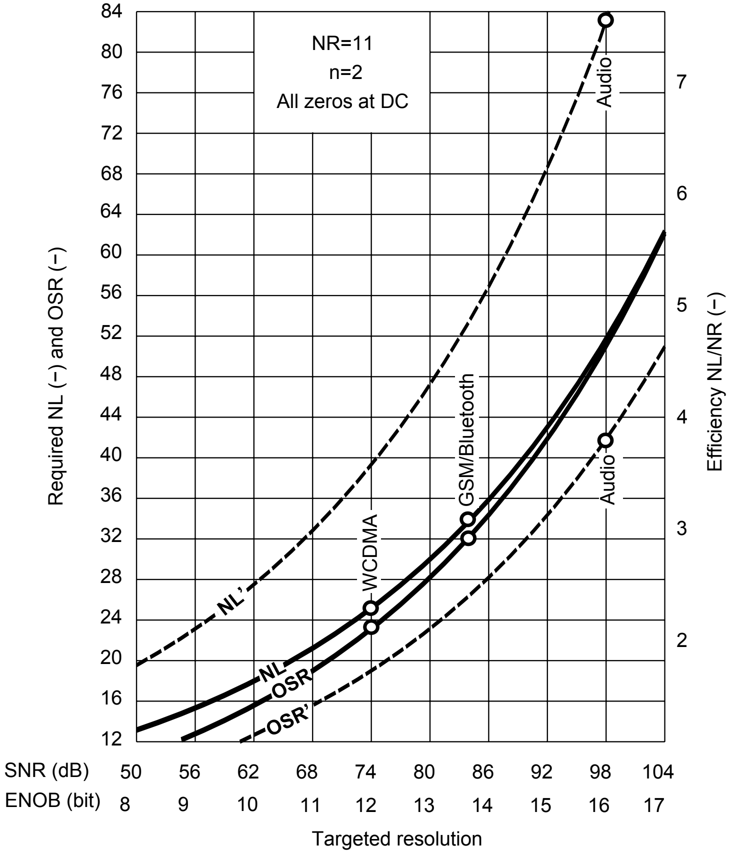


Figure 4.22 Required NL and OSR for a targeted SNR and a given NR. The specified band is twice the bandwidth and the SQNR is 10dB higher than the target. Identical bandwidth and specified band are considered for the dashed curves NL′ and OSR′. The targeted applications shown in Table 4.2 are labeled and highlighted with dots. The design example considered in a subsequent chapter corresponds to the Bluetooth case addressing 84dB, which requires an OSR of 32 and 33 levels.

The curves NL and OSR show that a 25-level modulator sampled at 90MHz can cover the WCDMA standard. In contrast, NL′ and OSR′ lead a typical 16-bit audio converter to 82 levels, emulated from an 11-level quantizer. Table 4.2 summarizes some of the most common application standards. The auto-ranging efficiency is defined here as the ratio NL/NR.

In this approximation trend we considered a conservative non-overloaded modulator with a second-order NTF with all its zeros located at DC. In reality, each case has to be optimized taking into account all the aspects of a specific application. Nevertheless, it shows interesting efficiency with high to medium OSRs. For this reason, high-bandwidth low-resolution applications are not good candidates for the auto-ranging technique.

4.5 Conclusions

The principle of the auto-ranging algorithm has been explained in detail and compared to other published techniques exploiting the same phenomenon. A thorough analysis of the limitations was carried out. The analytical model developed revealed the existence of optimum values for moderate number-of-levels and over-sampling ratios. Thus, for the 94dB modulator studied in the design example, presented in a subsequent chapter, the choice of NL=33 and OSR=32 is the most appropriate. Besides, it was shown that high-order modulators have a limited efficiency in contrast to the first- and second-order case.

The auto-ranging efficiency presents some similarities with the clock jitter sensitivity. They are both related to the step size between two samples at either the input or the output of the quantizer. However, the mechanism linking them to the resolution is different.

Chapter 5

Mismatch shaping optimizations

This chapter starts by explaining the principle of the spectral shaping of DAC element mismatch. The well-known tree-structured architecture is described and the most relevant features are highlighted by simulation experiments. The second section introduces the segmentation of the architecture. A synthesis method using functional programming is presented. The last section addresses the power consumption optimization.

5.1 Dynamic element matching

An NL-level DAC usually consists of a bank of *one-bit* $(NL-1)$ elements. The analog output nodes are all connected together, resulting in a summation of all their individual contributions.

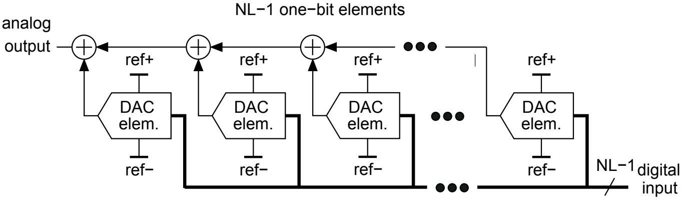


Figure 5.1 Conventional DAC implementation as a bank of $(NL-1)$ elements.

As illustrated in Figure 5.1, each element uses the differential reference voltage with nodes **ref+** and **ref–**. In contrast, as shown in Figure 4.2, each comparator of a quantizer bank uses a fraction of the differential reference voltage. A large NL is therefore less of a limitation factor for the DAC than for the internal ADC in a multi-bit ΔΣ-modulator.

As shown in Chapter 2, the mismatch between the elements of the first DAC bank significantly degrades the resolution. In the presence of mismatch, the transfer characteristic of the DAC is not linear and harmonic distortion is generated if a thermometer coding is used. The purpose of the Dynamic Element Matching (DEM) algorithm is to provide an *intelligent* scrambling such that the transfer characteristic appears linear on average. The algorithm is implemented in a digital circuit commonly referred to as a *mismatch shaping encoder*.

According to the comparisons made by [WG02] and more extensively in [GS02], among the DEM algorithms proposed in the literature, we can distinguish three main classes of encoders: the vector feedback, the Data Weighted Averaging (DWA) and the tree-structured encoder.

As conceptually described in Figure 5.2, the three techniques differ from each other in the way they choose which elements are to be set high or low. At a sample time i, for a given input $x[i]$, the vector feedback encoder employs a *vector quantizer* to choose the $x[i]$ elements less used. The DWA employs a *barrel shifter* to select the $x[i]$ elements consecutive to the ones used at the previous sample time $i - 1$. The tree-structured encoder employs a network of *switching blocks* to split $x[i]$ into the $(NL - 1)$ connections to the elements.

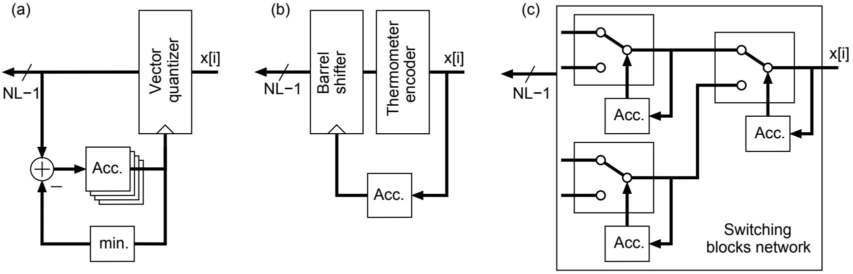


Figure 5.2 DEM algorithm implementations: conceptual representation of (a) the vector feedback, (b) the DWA and (c) the tree-structured encoders. The implementations shown here provide a first-order spectral shaping.

The first-order spectral shaping is provided by the accumulators. By altering the structures and adding accumulators, each implementation can provide arbitrary spectral shaping of mismatch errors. However, the analysis of [WG02] points out that tree-structured encoders can easily avoid the generation of tones. Moreover, according to [FSW$^+$02], these encoders can be used with *segmented* DACs which allows a significant reduction in the hardware complexity significantly. Hence, the tree-structured architecture is chosen here.

5.2 Tree-structured architectures

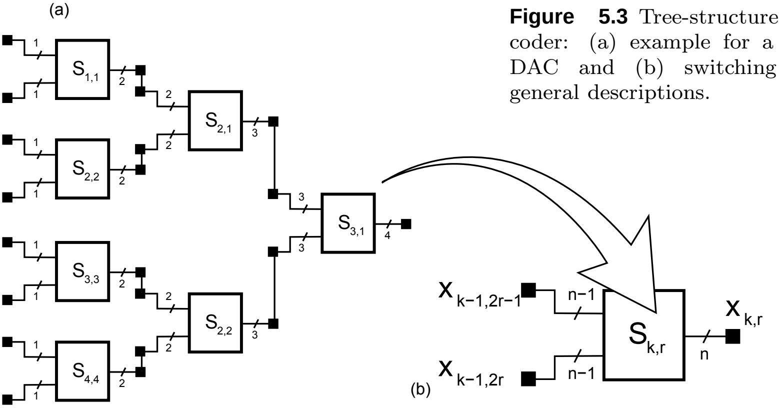


Figure 5.3 Tree-structured encoder: (a) example for a 9-level DAC and (b) switching block general descriptions.

A tree-structured encoder for a 9-level DAC is shown in Figure 5.3 with the definitions set by [Gal97]. The tree is composed of 7 switching blocks placed on 3 layers. Each block $S_{k,r}$ is set on the layer k and the row r. Its input $x_{k,r}$ splits into two parts

$$x_{k-1,2r-1} = \frac{x_{k,r} \pm s_{k,r}}{2} \,, \tag{5.1}$$

$$x_{k-1,2r} = \frac{x_{k,r} \mp s_{k,r}}{2} \,, \tag{5.2}$$

where

$$s_{k,r} = \begin{cases} \pm 1 \,, & \text{if } x_{k,r} \text{ is odd} \,, \\ 0 \,, & \text{if } x_{k,r} \text{ is even} \,. \end{cases} \tag{5.3}$$

If $x_{k,r}$ is odd $s_{k,r}$ is chosen randomly. A *pseudo-random* generator is there necessary to provide each switching blocks with a sequence of ±1 values uncorrelated. The mathematical aspect about the pseudo-random sequences is treated in deepth by [GG05]. The practical aspects reguarding the hardware implementation are explained in [FGHJ00].

As an example, let us consider that at the input of any switching block $x_{k,r}[n] = 17$. Since 17 is odd, it splits into either 8 and 9 or 9 and 8. The sequence $s_{k,r}[n]$ causes these two possibilities to alternate at any time an odd number will be represented. On the other hand, any even number splits perfectly without the opportunity to alternate. The author of [Gal97] showed with a recursive demonstration that, under these conditions, the analog output $y[n]$ of the DAC can be expressed as:

$$y[n] = \alpha x[n] + \beta + e[n]. \tag{5.4}$$

where α and β are respectively the global DAC gain and offset, which are constant and depend on the statistic mismatch of each DAC elements. Besides, the sequence $e[n]$ is equals the sum of each $s_{k,r}[n]$ sequences weighted by their respective local gain $\Delta_{k,r}$:

$$e[n] = \sum_{k=1}^{b} \sum_{r=1}^{2^{b-k}} \Delta_{k,r} s_{k,r}. \tag{5.5}$$

The DAC output can therefore be seen as a linear gain device with an additional offset and noise whose spectrum is determined by the sequences $s_{k,r}[n]$. Each column k needs its own pseudo-random sequence. The sequences should be uncorrelated to each other. [FGHJ00] shows how to transform a conventional b-bit *Linear Feedback Shift Register* (LFSR) such as to generated k sequences whose cross-correlation has a period of $2^b/k$. As shown in [WGF01], this splitting can be implemented with a simplified hardware by taking advantage of the *extra-LSB* coding.

5.2.1 Spectral shaping

Each switching block has its own *shaper* generating the sequences $s_{k,r}[n]$. Figure 5.4 describes the signal processing of a first-order shaper. The shaper works like a single-bit $\Delta\Sigma$-modulator. A 2-bit accumulator, followed by a hard limiter, is placed in a feedback configuration. As explained

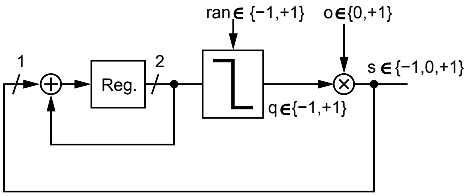


Figure 5.4 Signal processing of the first-order shaping. The output of the hard limiter **q** is either ±1. Whether the switching block input is odd or even, the shaper output **s** is forced to zero or kept unchanged.

at a greater length in [Gal97], the hard limiter generates the sequences **q** such as to force the sum to be bounded. As the input of the switching block is odd, node **o** becomes zero as well as the output sequence **s**. The dithering is introduced by the random variable **ran** when the accumulator output is zero and the hard limiter has the choice of providing either ±1. In this way, the spectrum of each sequence $s_{k,r}$ has a first-order spectral shape as well as the mismatch errors $e[n]$.

The general tree-structured encoder for an NL-level DAC described above is implemented in Code 5.1. The two arrays **y** and **acc** correspond the switching block output and accumulator state respectively.

Code 5.1 MATLAB code for the first-order tree encoder. The parameter **N** is the encoder depth equal to $\log_2(\mathrm{NL}-1)$.

```
y(1,1)=out(i)+(NL-1)/2;
for ii = 2:N+1,
      ran=sign(randn(1));  %random +/-1;
      for jj = 1:2^(ii-2),
      in=y(ii-1,jj);

      %%%%% first-order shaper %%%%
      if (mod(in,2)~=0)
        if (acc(ii,jj)==0) s=ran;
        else s=-sign(acc(ii,jj));end
        else s=0;
      end
      %%%%%%%%%%%%%%%%%%

      acc(ii,jj)=acc(ii,jj)+s;
      y(ii,2*jj-1)=(in+s)/2;
      y(ii,2*jj)=(in-s)/2;
      end
end
out(i)  =sum((y(3,:)*1-0.5).*(dac+mismatch));
```

Figure 5.5 shows the block diagram of the simulation testbench we used to evaluate the spectral shaping properties of the encoder. Any type of signal is fed into the encoder. The output goes through a DAC with mismatches. In accordance to Equation (5.4), the mismatch errors e are extracted by removing the input signal multiplied by α and an offset β from the output.

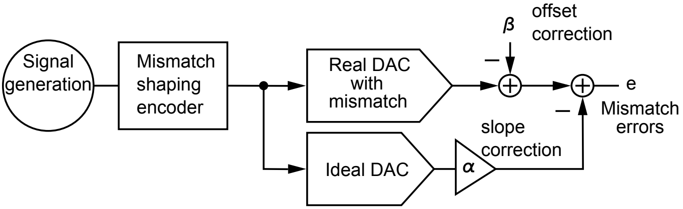


Figure 5.5 DEM testbench to extract the mismatch errors e. α and β are the DAC gain and offset of equation (5.4).

Figure 5.6 shows the simulations with different full-scale input signals. The spectra highlight the signal dependence of the mismatch shaping.

Figure 5.7 shows the simulations results with the $\Delta\Sigma$-modulation used as an input. In the case shown in Figure 5.7(a), the variable **s** in Code 5.1 is kept equal to one if the switching block input is odd. In this way the data are not scrambled at all. This situation is equivalent to a direct connection from the quantizer output to the DAC input. In Figure 5.7(c) the simulation is performed with a $\Delta\Sigma$-modulator output signal. The spectral shaping corresponds exactly to a first-order modulator output. This experiment shows that such an encoder is well suited to $\Delta\Sigma$-modulation-based DACs and ADCs.

As pointed out in Chapter 2, due to the mismatch between its elements, the DAC has a non-linear transfer characteristic and harmonic distortion is generated. In the case depicted in Figure 5.7(b), the variable is set randomly at each clock cycle. As a result, the encoder scrambles the data and the PSD of the mismatch errors is white. The surface below the PSD over the band-of-interest is similar for both cases, bringing about an important degradation of the resolution. In contrast, in the case shown in Figure 5.7(c), the variable is set as described previously. The spectral

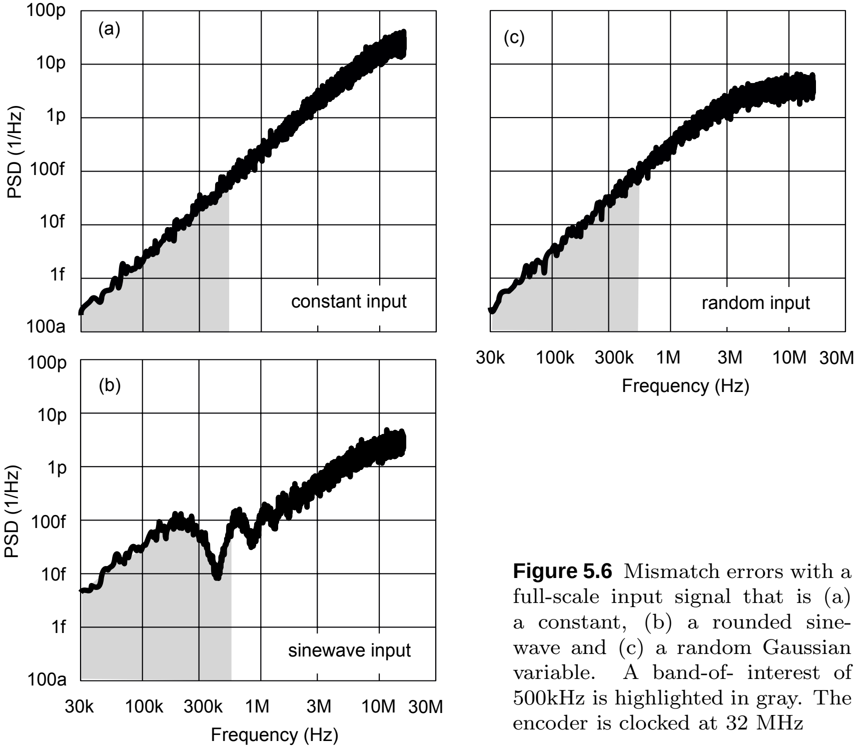


Figure 5.6 Mismatch errors with a full-scale input signal that is (a) a constant, (b) a rounded sinewave and (c) a random Gaussian variable. A band-of- interest of 500kHz is highlighted in gray. The encoder is clocked at 32 MHz

shaping was therefore applied and the PSD presents a prefect first-order characteristic.

5.2.2 Higher-order spectral shaping

The mismatch spectral shaping can be improved by changing the shaper processing. Figure 5.8 shows the shaper block diagram for a second- and third-order modulator. Codes D.4 and D.5 implement these sequence generators. The hard limiter output **q**, which is either + or -1, is chosen such as to keep the accumulator outputs bounded. The first accumulator have priority over the others such as to guaranty the first-order shaping.

Both shapers maintained the same first stage with a 2-bit accumulator.

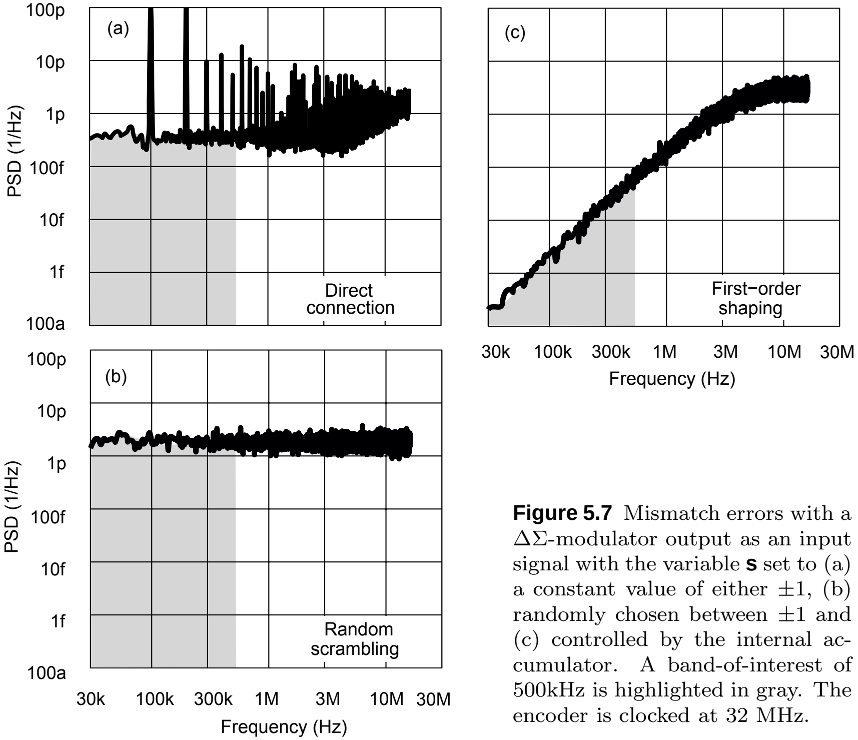


Figure 5.7 Mismatch errors with a $\Delta\Sigma$-modulator output as an input signal with the variable **s** set to (a) a constant value of either ±1, (b) randomly chosen between ±1 and (c) controlled by the internal accumulator. A band-of-interest of 500kHz is highlighted in gray. The encoder is clocked at 32 MHz.

The additional accumulators offer a degree of freedom in the choice of register size. The simulation results in Figure 5.9 reveal that for a low OSR, in the example equal to 32, the second-order shaping with a 3-bit accumulator behaves as a first-order at low frequencies. However, the mismatch noise PSD integrated over the band-of-interest is similar to the case with an infinite number of bits.

The PSD of Figure 5.9(c) shows that a third-order shaping is possible with resonance. Nevertheless, at least 13-bit accumulators are necessary for the second and third stages. The size of the registers and adders drastically increase the complexity of the encoder. Such an architecture should be used only with deep sub-micron CMOS technology with transistor sizes below 90nm.

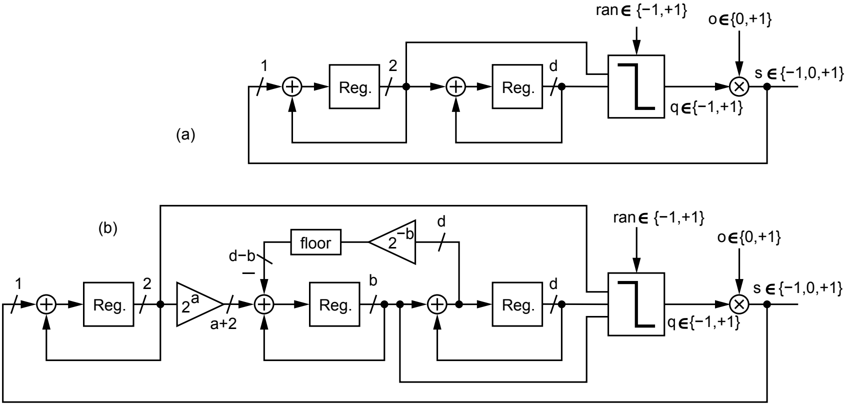


Figure 5.8 Signal processing block diagram of the second- and third-order mismatch shaper.

5.3 Optimization

5.3.1 DAC segmentation

In [FSW$^+$02] it is shown that the processing of the switching block described by Equations (5.1) and (5.2) can be altered thus reducing significantly the number of switching blocks.

However, as the number of switching block decreases, the number of DAC elements increases. The first aspect to consider is that both the switching blocks and the DAC elements contribute to the power consumption of the DAC. An optimization is therefore necessary.

Table 5.1 summarizes the switching block proposed by [FSW$^+$02]. The conventional block here refers to the F block, which stands for Full. The other blocks are respectively the Half (H) and the Quarter (Q) switching blocks. These names are chosen with respect to their functionality. For instance, the H block splits its input x into the next multiple of 2 and its remainder. In this way, the output with the multiple of 2 has half the possibilities. Similarly, the Q block splits its input x into the next multiple of 4 and its remainde. In this way, each output has a quarter the possibilities. A detailed demonstration is provided in [FSW$^+$02].

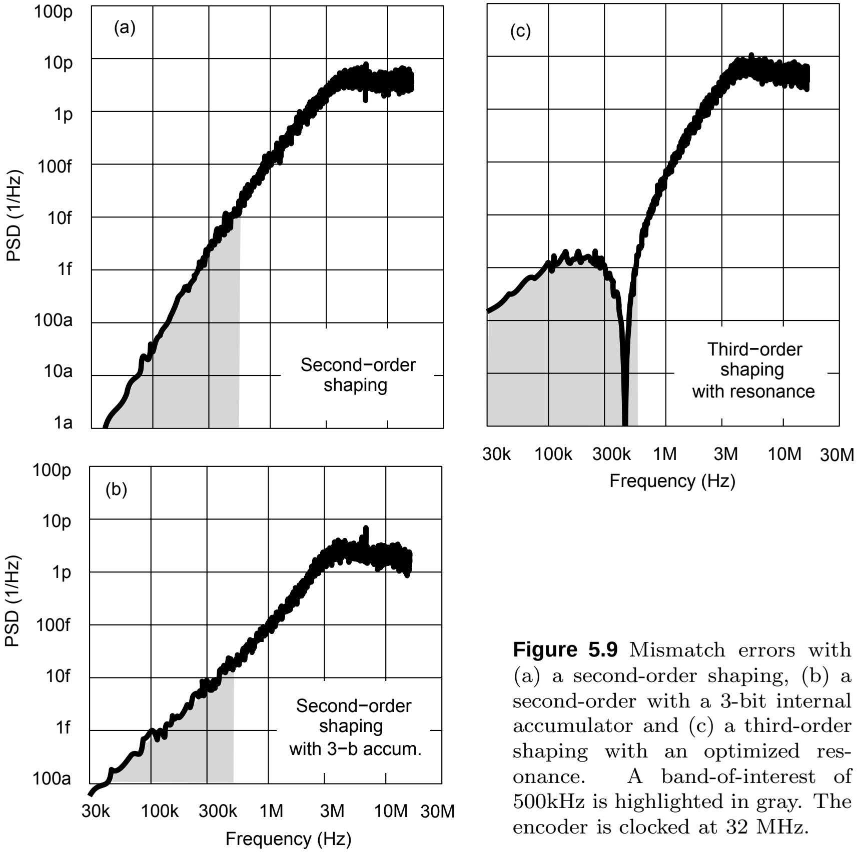


Figure 5.9 Mismatch errors with (a) a second-order shaping, (b) a second-order with a 3-bit internal accumulator and (c) a third-order shaping with an optimized resonance. A band-of-interest of 500kHz is highlighted in gray. The encoder is clocked at 32 MHz.

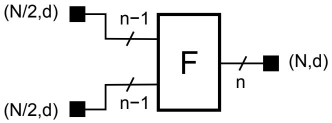
5.3.2 Synthesis automation

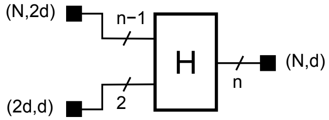
The synthesis process can be automated. We restrict the process to the switching blocks given in Table 5.1. The process is implemented with the functional programming of the MATHEMATICA software. MATHEMATICA is a powerful tool based on the manipulation of lists, which is particularly adapted to the synthesis of trees. Codes C.1, C.2 and C.3 used for the synthesis are provided in appendix.

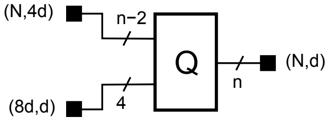
The synthesis is performed in three steps. The first step consists

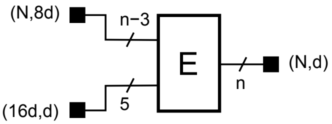
Table 5.1 Switching blocks description

Output range	Switching block	Input range

up={−N/4, ... ,+N/4}
down={−N/4, ... ,+N/4}
		in={−N/2, ... ,+N/2}

up={−N/2, ...,+N/2}
down={−1, ... ,+1}
		in={−N/2, ... ,+N/2}

up={−N/2, ... ,+N/2}
down={−3d, ... ,+3d}
		in={−N/2, ... ,+N/2}

up={−N/2, ... ,+N/2}
down={−7d, ... ,+7d}
		in={−N/2, ... ,+N/2}

in recursively generating the tree of solutions, which is not yet a list of encoders. The recursive function **BuildTree** is used for that purpose. Figure 5.10 gives a partial representation of the result generated at the first step. The starting point is the list of possible switching blocks Q,H and F. Each of these blocks has two outputs. Each output can be followed by either block in the list Q, H and F. The process goes on recursively until the terminating conditions are satisfied. In such a case the branch ends

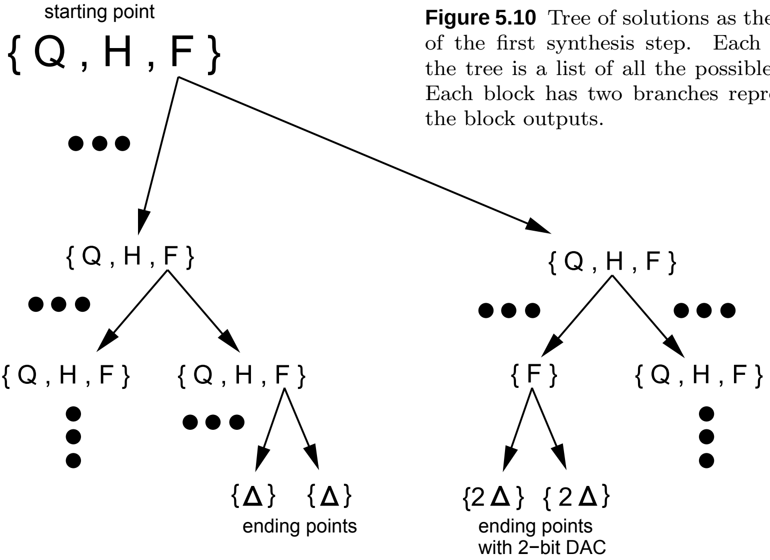


Figure 5.10 Tree of solutions as the output of the first synthesis step. Each node of the tree is a list of all the possible blocks. Each block has two branches representing the block outputs.

with a Δ which corresponds to a one-bit element.

The recursive function propagates the increase in weight brought about by the blocks H and Q. Hence, the branches end in some places with multiples of Δ. The multiples correspond to the weight of the DAC element at that point.

The second step consists in removing redundant solutions with the help of the function **RemoveSymmetry**. Figure 5.11 shows a part of the solution tree displayed in Figure 5.10. The two lists of possible blocks on both sides would generate redundant solutions in the last synthesis step. For instance, solutions with an F block connected to either H and F or F and H are redundant. As illustrated in Figure 5.11, this step starts the extraction of the encoder structures and the combinations of two different blocks only appear once.

In the third step, the function **ExtracTree** operates in the same way on the entire solution tree without taking care of redundant solutions. The process transforms the tree of solutions into a list of tree structures as shown in Figure 5.12.

Code C.5 in the appendix shows how to use the three recursive func-

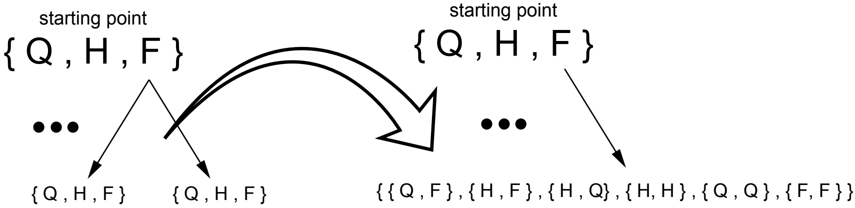


Figure 5.11 Removal of redundant solutions in the solution tree performed during the second synthesis step.

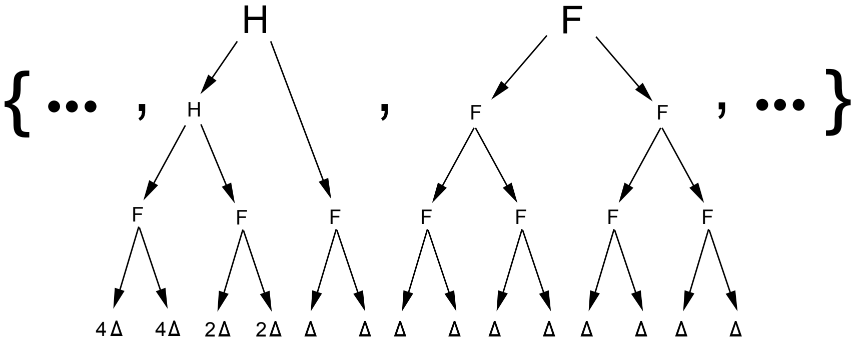


Figure 5.12 Output after the last step of the synthesis.

tions. The synthesis process is applied to the 33-level DAC using F, H and Q blocks. The algorithm finds 628 distinct solutions. Code C.4 provides a function to display the trees. As an example, Figure 5.13 shows the last solution taken from the list of trees. The input is a 6-bit bus represented with an extra-LSB code. The total number of one-bit elements is equal to 70 and there are 11 switching blocks. This solution is unlikely to be optimal. However, since the first splitting device is a Q-block, the encoder depth is reduced by one level with respect to a conventional tree. The critical path is shorter and may allow a longer delay for the logic synthesis.

Figure 5.14 shows the synthesis results for the 33-level using F, H and Q blocks. The last of these solutions corresponds to the structure shown in Figure 5.13. The solutions are classified with respect to the total number of DAC elements and the current consumption of both the DAC itself and

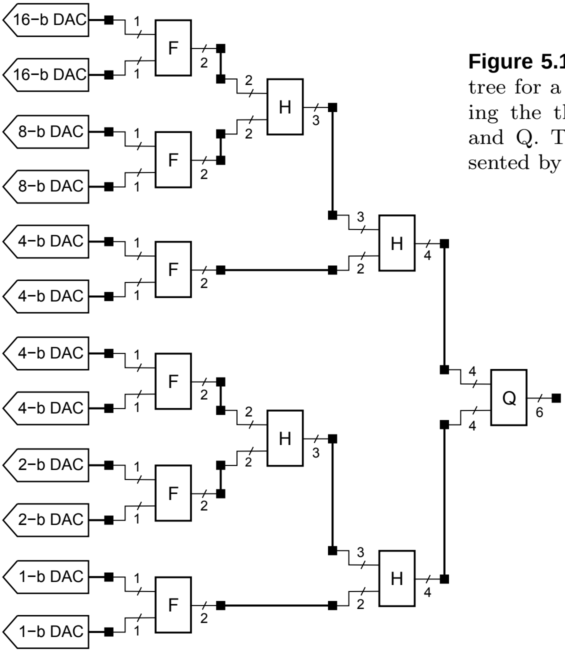


Figure 5.13 Example of a synthesized tree for a 33-level segmented DAC using the three switching blocks, F, H and Q. The digital signals are represented by an extra-LSB code.

the digital encoder.

We are considering here the situation of the design example of Chapter 6. A second-order shaper with a 3-bit accumulator is used. The approximate gate-count as well as the power consumptions of each block are given in Table 5.2. The simulation of standards cell for the 0.18μm CMOS technology used gives a mean current consumption of 0.3μA/gate. We take into account the consumption of the first DAC elements for the Δ's.

5.3.3 Standard segmentation

The solutions giving the best performance in terms of power consumption with respect to the number of elements are highlighted in Figure 5.14 with a dashed curve. It turns out that these solutions have only H and F switching blocks. Moreover, the H blocks are all cascaded at the beginning

Table 5.2 Gate count for switching blocks with a second-order shaper with a 3-bit accumulator.

Block	Sequence generator	Switching logic	Consumption (μA)
F	27	8	10.5
H	27	10	11.1
Q	81	12	27.9
Δ	-	-	2.7

of the tree. Figure 5.15 shows one of these solutions, the number 589, as a general example. We refer to these structures as a *standard segmented* tree. We define the segmentation depth m, as shown in Figure 5.15, as the number of cascaded H blocks. Code D.6, provided in appendix, allows to simulate such a standard segmended tree with all the possible encoder and segmentation depths.

From the above description of the standard segmentation, we can express the current consumption analytically as

$$I = \underbrace{\Delta \cdot 2(2^{\mathrm{N}-1} + 2^m - 1)}_{\text{analog}} + \underbrace{\mathrm{F} \cdot (2^{\mathrm{N}-m} + m - 1) + \mathrm{H} \cdot (m)}_{\text{digital}} , \tag{5.6}$$

where N and m are the encoder and segmentation depths respectively. If only H and F blocks are used, the encoder depth is given by

$$\mathrm{N} = \log_2(\mathrm{NL} - 1) . \tag{5.7}$$

The parameters F, H and Δ are the individual current consumption of the switching blocks and a one-bit DAC element. This equation corresponds to the dashed curve of Figure 5.14. The optimal segmentation depth can be found as

$$m_{\mathrm{opt}} = \log_2 \left(\frac{\sqrt{(\mathrm{H}+\mathrm{F})^2 + 2^{\mathrm{N}+3}\mathrm{F}\Delta \ln^2 2} - (\mathrm{H}+\mathrm{F})}{4\Delta \ln 2} \right) . \tag{5.8}$$

In most cases, the complexity of the switching blocks H and F is dominated by the shaper. The H block has only a few extra gates for the

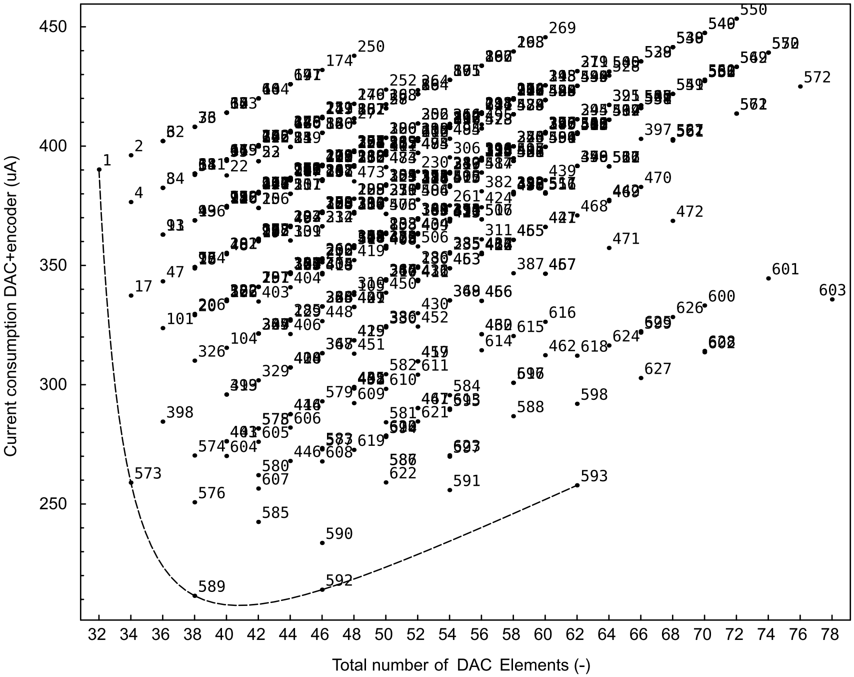


Figure 5.14 Classification of the 628 solutions found by the synthesis process for a 33-level DAC using F, H and Q blocks.

multiplexer with respect to the F block. Consequently, the consumption parameters H and F can be considered equal. Equation (5.8) can therefore be simplified as

$$m_{\mathrm{opt}} = \log_2 \left(\frac{\sqrt{1 + 2^{\mathrm{N}+1} \alpha \ln^2 2} - 1}{2\alpha \ln 2} \right) \tag{5.9}$$

where $\alpha = \Delta/\mathrm{F}$. In the case of a high-order shaping, the ratio α may become small. In such a case, the expression of the optimal segmentation depth can be further approximated as

$$m_{\mathrm{opt}} = \mathrm{N} - 1.53\ . \tag{5.10}$$

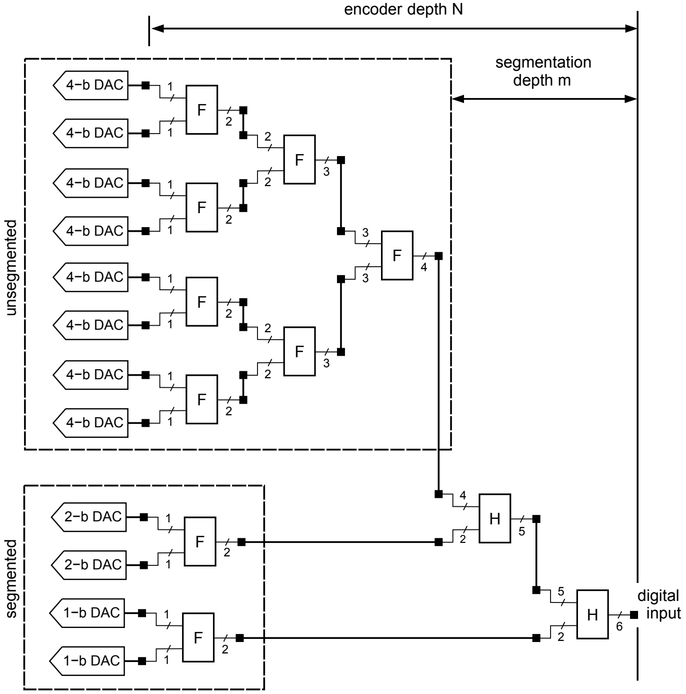


Figure 5.15 Standard segmentation. Solution number 589 is a standard segmented tree with a segmentation depth of 2.

Therefore, if the switching block complexity is high, as in the case of third-order shaping algorithm for instance, a large segmentation is suitable. In contrast, if the switching blocks are either simple, as in the case of a first-order shaping, or complex but benefiting greatly from deep sub-micron devices provided by the technology, a lower segmentation depth is recommended. The other aspect to be pointed out by the classification proposed in Figure 5.14, is that the higher the segmentation depth, the more noise sources there are in the DAC bank.

5.4 Conclusions

The auto-ranging technique described previously allows us to increase the internal number-of-levels thus circumventing the problem of implementing a large quantizer bank. However, the DAC does not benefit from that. On the contrary, as we are now allowed to increase the number-of-levels, the DAC bank size, as well as the complexity of the mismatch shaping encoder, may become an issue.

To that extent, we developed in this chapter an analytical method as well as a synthesis algorithm. This provides designers with a means for estimating and optimizing the consumption. Besides, the classification proposed allows consideration of other constraints, such as the number of extra noise sources in the DAC bank or the encoder depth.

Finally, it should be pointed out that segmentation makes the layout of the DACs and their routing to the digital section easier. The bus size carrying the feedback digital signals is reduced, in the example studies, from 32-bit to 12-bit. The DAC elements with weights of 2 and 4 can be inter-digited such as to guarantee better matching properties.

Chapter

6

Design example

The low-power strategy and the techniques presented previously are applied in a design example in this chapter. A BLUETOOTH receiver with direct conversion architecture was chosen as a typical target application. The examples shown throughout the previous chapters serve as a support to the second-order modulator studied here. The chapter starts by describing the modulator architecture and its analysis at the signal processing level. It then gives details of circuit implementation. Finally, the expected performance and consumption of the designed circuit are summarized and compared with equivalent state-of-the-art realizations.

6.1 Modulator design

6.1.1 Targeted specifications and topology choice

This design example addresses possible BLUETOOTH specifications with 83dB of resolution over a bandwidth of 500kHz. These specifications could be set for a direct conversion receiver where limited selectivity and anti-aliasing filtering is achieved before the ADC.

An SQNR of 94dB was chosen, allocating about 10dB of margin for circuit noise and distortion. Among all the possible choices for n, NL and OSR, we chose a 33-level second-order modulator with an over-sampling ratio of 32. In line with the low-power strategy elaborated in Chapter 3, the architecture is hybrid with a continuous-time first integrator and a discrete-time second integrator. Sampling is performed with a full-clock-cycle scheme to reduce the power consumption of the interfacing buffer.

This solution is optimal in terms of jitter sensitivity as shown in Chapter 3. The design charts in Figure 3.17 showed that modulators with higher shaping orders are more sensitive. With an SQNR of 94dB the sensitivity reaches its optimum for a moderate NL around 33. Moreover, the analysis of Chapter 4 concludes that this moderate NL of 33 also gives the best efficiency for the auto-ranging technique, allowing the number of comparators to be reduced from 32 to 10.

The main purpose of this example is to demonstrate the application of the low power strategy proposed in Chapter 3. The NTF is designed with all its poles in center of the z-plane and all its zeros at DC. A more aggressive pole and zero placement, such as described in Chapter 2, would increase the performance at the cost of introducing complexity in the design. Furthermore, no feed-forward paths are added, providing a flat STF with the anti-aliasing filtering as calculated in Chapter 3.

Table 6.1 summarizes the characteristics of the modulator. The sampling frequency corresponds to the order of a clock signal available on a typical System-on-Chip (SoC).

Table 6.1 System characteristics.

Targeted resolution (13.5bits)	SNR	83 dB
Quantization resolution	SQNR	94 dB
Band-of-interest	f_b	500 kHz
Full-scale interferer band	$f_{\max}$	1 MHz
Sampling frequency	f_s	32 MHz
Over-sampling ratio	OSR	32
Shaping order	n	2
Emulated quantizer	NL	33 levels
Internal quantizer	NR	11 levels

6.1.2 Circuit description

The schematic of the chosen 33-level hybrid continuous-discrete-time modulator is depicted in Figure 6.2. The analog part is composed of fully differential amplifiers for the **CT-integrator** and **DT-integrator**, a fully differential difference amplifier for the **Sampler**, a current steering **i-DAC**, a switched-capacitor **sc-DAC**, an 11-level **Quantizer** and its reference ladder. The **Digital** section comprises the two mismatch shaping encoders, the auto-ranging and a comparator offset calibration algorithms.

The quantizer reference ladder and the switched-capacitor DAC elements are biased with the same voltage references **ref+** , **gnd** and **ref—**, equal to 1.22V, 900mV and 580mV, respectively. This gives a differential reference voltage of 640mV. The resistors ladder provides 10 differential references **ref9** to **ref0** with ±40mV, ±120mV, ±200mV, ±280mV and ±360mV. Consequently, the quantizer has a mid-thread transfer characteristic with steps Δ =80mV.

An active-RC integrator is used for the first integrator to prevent distortion in this first-stage which does not take advantage of any spectral shaping. The current steering of that first-stage generates Non-Return-to-Zero (NRZ) current pulses. To reduce the voltage swing of the control signals, **vh** and **vl** are set to 1.5V and 300mV respectively.

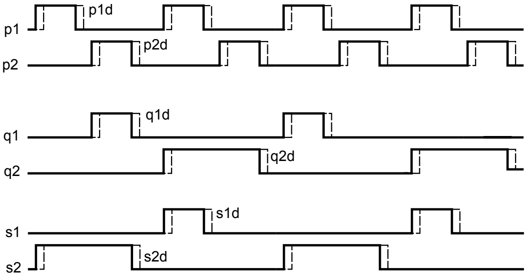


Figure 6.1 System clocking diagram.

Figure 6.1 shows the clocking diagram. The non-overlapping phases **p1** and **p2**, as well as their delayed versions **p1d** and **p2d**, are used in the switched-capacitor DAC elements. The clock generator is designed so as to provide non-overlapping time and delays of 1ns, with rise and fall times of 0.3ns from a 32MHz signal. Capacitors are used to generate such delays instead of a large series of inverters. The alternating phase sets, **s1,s1d,s2,s2d** and **q1,q1d,q2,q2d**, are necessary for the full-clock-cycle scheme described in Chapter 3. The integration process takes place during phase one. Both the sampling and the DAC capacitors pre-charge are performed during phase two.

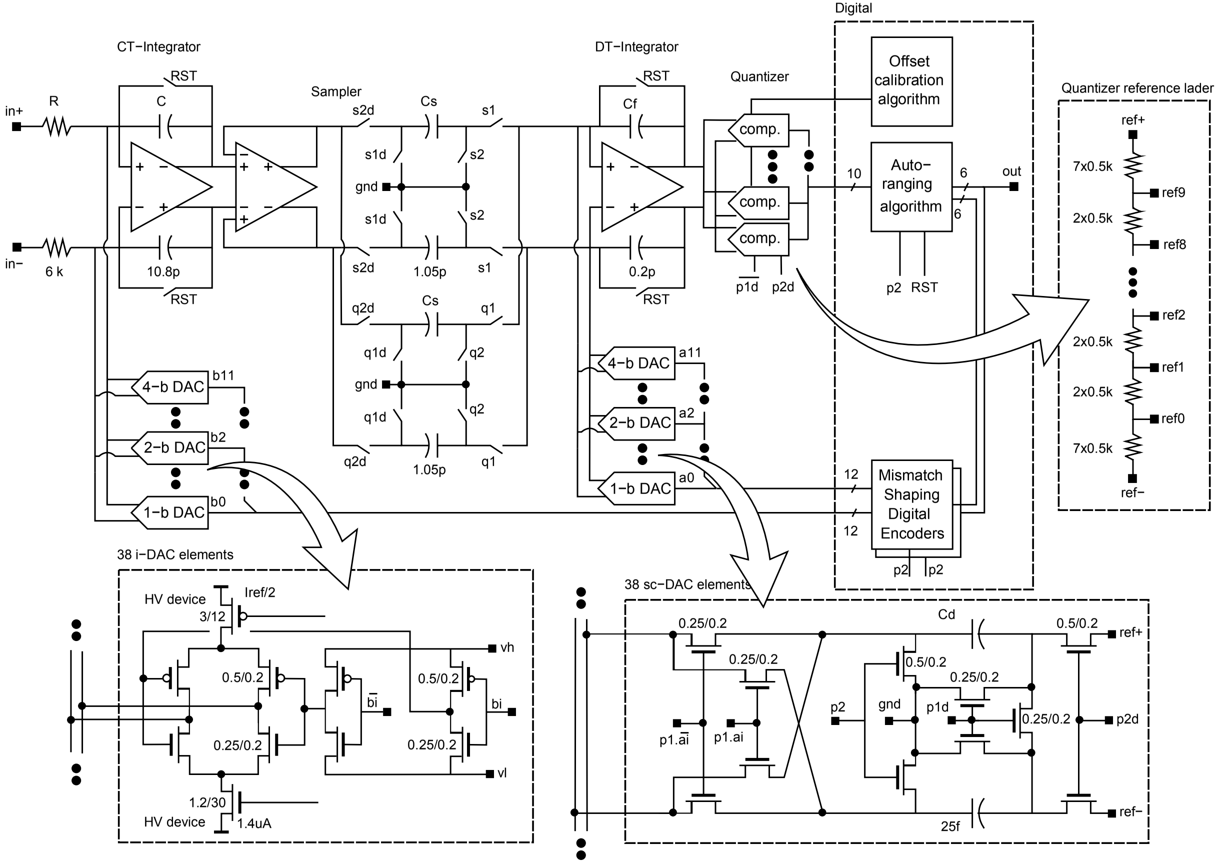


Figure 6.2 Schematic of the 33-level second-order $\Delta\Sigma$-modulator.

Table 6.2 CMOS technology features used.

Feature	Advantage
TaN thin film resistors	High linearity Low temperature coefficient Good matching properties
Metal-Insulator-Metal (MIM) capacitors	High linearity Excellent matching properties
High Voltage (HV) transistors with thick gate oxide	Allows charge pumping for driving switches Provides low noise current sources
Five metal layers	Higher routing density
Triple-well technology	Insulation of sensitive analog blocks Reduction of body effect

According to the developments of Chapter 5, the DACs are partially segmented following a standard segmentation with a depth of 2. Because of this, the digital bus feeding the DACs is 12-bit wide instead of 32. Also, the two outputs of the auto-ranging block are 6-bit wide to allow representation of the 33 levels with an extra-LSB code. The total number of elements ND is 38 for both DACs.

The 0.18μm CMOS technology used in this design is provided by Freescale semiconductor. The special features available are listed in Table 6.2. Freescale also provided its own SPICE-like simulator MICA along with special analysis tools, developed internally, allowing complex statistical simulations for mismatch and process variation analysis.

6.2 System Design

6.2.1 Signal processing of the circuit

The block diagram in Figure 6.3 is a direct translation of each signal processing function of the circuit shown in Figure 6.2. As analyzed in Section 3.2.1, the continuous-time DAC convolves each digital impulse

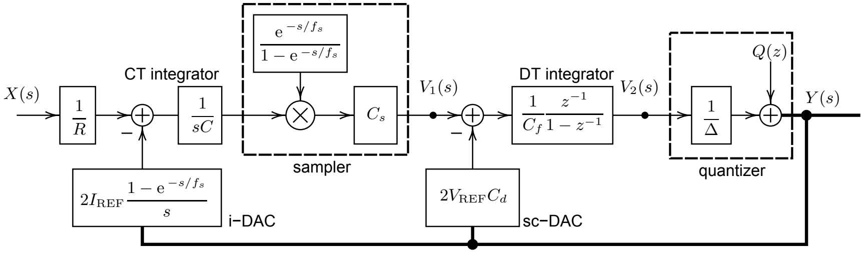


Figure 6.3 Modulator block diagram. The processing blocks are labeled as in the circuit description in Figure 6.2.

with a $1/f_s$-duration rectangular pulse. This becomes a multiplication in the frequency domain. On the other hand, the sampling process is a multiplication by a train of impulses in the time domain and becomes a convolution in the frequency domain.

As seen in Section 3.2.1, the path involving the hold function and the sampling process can be replaced by an equivalent z-transform. The block diagram can therefore be simplified and described by the linear Signal-Flow Graph (SFG) shown in Figure 6.4.

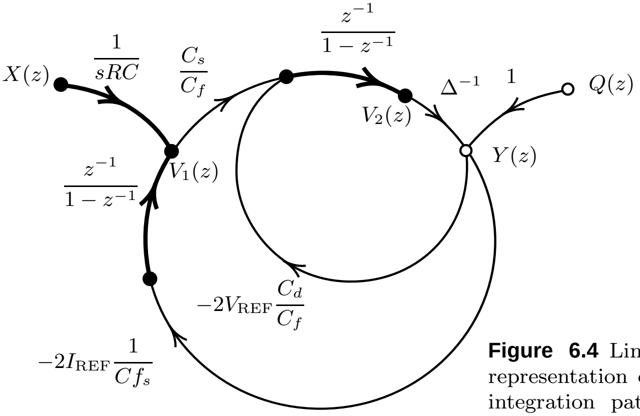

$X(z)$
$\frac{1}{sRC}$
$\frac{C_s}{C_f}$
$\frac{z^{-1}}{1-z^{-1}}$
$V_2(z)$
Δ^{-1}
1
$Q(z)$
$Y(z)$
$V_1(z)$
$\frac{z^{-1}}{1-z^{-1}}$
$-2V_{\mathrm{REF}}\frac{C_d}{C_f}$
$-2I_{\mathrm{REF}}\frac{1}{Cf_s}$

Figure 6.4 Linear signal-flow graph representation of the modulator. The integration paths are highlighted in bold.

We apply Mason's gain formula to the SFG considering $Q(z)$ and $X(z)$

as source nodes, and $Y(z)$ as a sink node and find

$$\mathrm{NTF}(z)=\frac{Y(z)}{Q(z)}=\frac{1}{1+\frac{2V_{\mathrm{REF}}C_d}{\Delta C_f}\cdot\left(\frac{z^{-1}}{1-z^{-1}}\right)+\frac{2I_{\mathrm{REF}}C_s}{\Delta C f_s C_f}\cdot\left(\frac{z^{-1}}{1-z^{-1}}\right)^2}\,, \qquad (6.1)$$

$$\mathrm{STF}(z)=\frac{Y(z)}{X(z)}=\frac{\frac{1}{\Delta RCf_s}\cdot\frac{C_s}{C_f}\cdot\frac{z^{-1}}{1-z^{-1}}\cdot\frac{f_s}{s}}{1+\frac{2V_{\mathrm{REF}}C_d}{\Delta C_f}\cdot\left(\frac{z^{-1}}{1-z^{-1}}\right)+\frac{2I_{\mathrm{REF}}C_s}{\Delta C f_s C_f}\cdot\left(\frac{z^{-1}}{1-z^{-1}}\right)^2}\,. \qquad (6.2)$$

6.2.2 Design equations

For a systematic design of the modulator, the equations linking the swings of the integrators and the components parameters are needed. We start the analysis by introducing the constraints on the STF and NTF. For the best stability, all their poles should be placed at the center of the z-plane. The NTF should therefore be equal to $(1-z^{-1})^2$. As already explained in Chapter 2 and shown in Table 2.1, this means choosing feedback coefficients of 1 and 2. We can therefore write:

$$\mathrm{NTF}(z)=(1-z^{-1})^2 \Rightarrow \begin{cases} \dfrac{2I_{\mathrm{REF}}}{\Delta C f_s}\cdot\dfrac{C_s}{C_f} & =1\,, \\ \dfrac{2V_{\mathrm{REF}}}{\Delta}\cdot\dfrac{C_d}{C_f} & =2\,. \end{cases} \qquad (6.3)$$

In such a case, for low to moderate frequencies the STF can be approximated to a constant.

$$\mathrm{STF}(z)=\frac{C_s}{C_f}\cdot\frac{1}{\Delta f_s RC}\cdot z^{-1}(1-z^{-1})\frac{f_s}{s}\cong\frac{C_s z^{-1}}{\Delta RC f_s C_f}\,. \qquad (6.4)$$

Then we calculate the integrator outputs $V_1(z)$ and $V_2(z)$ by applying

Mason's gain formula, with $Q(z)$ and $X(z)$ as source nodes, and find that

$$V_1(z) = \frac{X(z)}{RCf_s}\left(z^{-1} + (\underbrace{\frac{2V_{\text{REF}}C_d}{\Delta C_f}}_{2} - 1)z^{-2}\right) - Q(z)\left(\underbrace{\frac{2I_{\text{REF}}}{Cf_s}}_{\Delta\frac{C_f}{C_s}}(z^{-1} - z^{-2})\right),$$

$$V_2(z) = \frac{X(z)}{RCf_s}\left(\frac{C_s}{C_f}z^{-2}\right) + Q(z)\left(\underbrace{\frac{2V_{\text{REF}}C_d}{C_f}}_{2\Delta}(z^{-1} - z^{-2}) + \underbrace{\frac{2I_{\text{REF}}C_s}{C_fCf_s}}_{\Delta}z^{-2}\right).$$

If the equation set (6.3) is satisfied, the equations become simpler. We determine the outputs as a function of time by applying the inverse z-transform and find

$$v_1(k) = \frac{(x(k-1) + x(k-2))}{RCf_s} - (q(k-1) - q(k-2))\Delta\frac{C_f}{C_s}, \quad (6.5)$$

$$v_2(k) = \frac{x(k-2)}{RCf_s} \cdot \frac{C_s}{C_f} - q(k-1)2\Delta + q(k-2)\Delta\,. \quad (6.6)$$

Finally, the output swings are calculated by taking the infinite norm $\|.\|_\infty$ as defined in Appendix A.1

$$\|v_1\|_\infty = \|x\|_\infty \frac{2}{RCf_s} + \|q\|_\infty 2\Delta\frac{C_f}{C_s}, \quad (6.7)$$

$$\|v_2\|_\infty = \|x\|_\infty \frac{1}{RCf_s}\frac{C_s}{C_f} + \|q\|_\infty 3\Delta\,. \quad (6.8)$$

Setting the quantizer input so as it never goes beyond its overload limits determines the maximum modulator input swing and the first integrator output.

$$\|v_2(k)\|_\infty = \frac{\Delta\text{NL}}{2} \quad \Rightarrow \quad \begin{cases} \|v_1(k)\|_\infty & = \Delta \cdot \frac{C_f}{C_s} \cdot (\text{NL} - 2)\,, \\ \|\,x(k)\,\|_\infty & = \Delta \cdot \frac{C_f}{C_s} \cdot \frac{\text{NL} - 3}{2} \cdot RCf_s\,. \end{cases} \quad (6.9)$$

Thanks to the auto-ranging algorithm, the quantizer number-of-levels is reduced to NR. Consequently, the real output swing $\|v_{2,R}\|_\infty$ of the last integrator is

$$\|v_{2,R}\|_\infty = \frac{\Delta \mathrm{NR}}{2}\,. \tag{6.10}$$

The parameter $\|v_2\|_\infty$ becomes the *emulated voltage swing* which can go beyond the voltage supply.

The reference ladder determines the relationship between Δ and V_{REF}. The relationship can be altered by adjusting the ratio between the edge and inside resistors. With the configuration shown in Figure 6.2, it follows that

$$\Delta = \frac{V_{\mathrm{REF}}}{8}\,. \tag{6.11}$$

By introducing this last relationship in Equations (6.9) and (6.3), we can express the modulator input swing as

$$\|x\|_\infty = RI_{\mathrm{REF}}(\mathrm{NL}-3)\,. \tag{6.12}$$

The maximum resolution is determined with the maximum input power $P_{x,\max}$ which goes through the STF before reaching the modulator output.

$$P_{X,\max} = \frac{\|x(k)\|_\infty^2}{2}|\mathrm{STF}(z)|^2 = \frac{(\mathrm{NL}-3)^2}{8}\,. \tag{6.13}$$

The quantization noise goes through the NTF before reaching the modulator output. The total power P_Q must be evaluated over the band of interest f_b. Assuming that OSR$\gg 1$, namely $f_s \gg f_b$, the NTF can be simplified by removing the high-order terms in its Taylor expansion.

$$P_Q = 2\int_0^{f_b} \frac{1}{12 f_s}\cdot \left|\mathrm{NTF}(f)\right|^2 \mathrm{d}f \cong \frac{1}{4\cdot 15\pi}\left(\frac{\pi}{\mathrm{OSR}}\right)^5\,. \tag{6.14}$$

We finally calculate the Signal-to-Quantization Noise Ratio

$$\mathrm{SQNR}_{\max} = \frac{P_{X,\max}}{P_Q} = \frac{15}{2}\pi(\mathrm{NL}-3)^2\left(\frac{\mathrm{OSR}}{\pi}\right)^5\,, \tag{6.15}$$

which is in accordance to the general formula (2.37) provided in Chapter 2. Taking 10 times the logarithm on both sides gives the expression

$$\mathrm{SNR}_{\mathrm{max,dB}} = 20\log(\mathrm{NL}-3) + 50\log(\mathrm{OSR}) - 11.1\ . \tag{6.16}$$

6.2.3 Design procedure

The equations derived above are presented in Table 6.3[1] in a compact form so as to provide a three-step design procedure. Firstly, the reference voltage is chosen. As a result, both the voltage swing of the last integrator output and the quantization step are set. Next, the capacitors of the sampling network and the switched-capacitor DAC elements are chosen. Consequently, the voltage swing of the first integrator output and the feedback capacitor of the last integrator are determined. Finally, the input resistance and voltage swing are chosen, thus determining the continuous-time integrator capacitor and the current steering DAC reference.

Table 6.3 Modulator design procedure.

Step	Parameter choice	Consequence
1	V_{REF} =640mV	$\begin{cases} \|v_{2,R}\|_\infty = V_{\mathrm{REF}}\cdot\frac{\mathrm{NR}}{16} = 440\mathrm{mV} \\ \Delta = V_{\mathrm{REF}}\cdot\frac{1}{8} = 80\mathrm{mV}\end{cases}$
2	C_d=25fF, C_s=1.05pF	$\begin{cases} \|v_1\|_\infty = V_{\mathrm{REF}}(\mathrm{NL}-2)\frac{C_d}{C_s} = 472\mathrm{mV} \\ C_f = 8\cdot C_d = 200\mathrm{fF}\end{cases}$
3	$\|x\|_\infty$=500mV, R=6kΩ	$\begin{cases} I_{\mathrm{REF}} = \frac{\|x\|_\infty}{(\mathrm{NL}-3)R} = 2.7\mu\mathrm{A} \\ C = \frac{I_{\mathrm{REF}}}{V_{\mathrm{REF}}}\cdot\frac{2}{f_s}\cdot\frac{C_s}{C_d} = 10.7\mathrm{pF}\end{cases}$

The reference voltage. At the starting point of the procedure, as NL and NR are already set, the reference voltage V_{REF} determines, without any

---

[1] By combining these equations and using the value chosen for the design example, we find that RC$\cong 2/f_s$.

degree of freedom, the voltage swing of the last integrator output $\|v_{2,R}\|_\infty$, the quantization step Δ, as well as the quantizer reference voltages. In a complete System-on-Chip (SoC) solution, the reference voltages **ref+** and **ref–** are provided internally by dedicated buffers. It is therefore suitable to chose V_{REF} between one-half and one-third of the voltage supply so as not to increase the constraints on the circuit design in terms of swing. This is true especially when taking into account the worst case process corners over a wide temperature range. Thus V_{REF} is set at 640mV.

The switched capacitor elements. The second parameter to be chosen is the capacitance C_d of the DAC elements. In the current architecture, the switched capacitor DAC bank comprises 76 capacitors. Reference buffers need to charge them at each clock-cycle. It is therefore desirable to choose the smallest DAC elementary capacitor since it has an important impact of the die size and consumption of the DAC bank. Moreover, the values of the other capacitors are in a direct relationship which has a further impact on the die size and power consumption. Nevertheless, the minimum size is limited by technology. First of all this is because the parasitic capacitances may become comparable to C_d, and secondly because the matching properties are inversely proportional to the capacitor area. Noise considerations further require us not to reduce the elementary DAC capacitor. Fortunately, errors induced by both noise and mismatches benefit from a first-order spectral shaping which allowed us to choose $C_d = 25\mathrm{fF}$. The sampling capacitor C_s sets the first integrator output swing $\|v_1\|_\infty$. This swing is also seen by the sampling amplifier as a common-mode by its differential pairs. In order to relax this design constraint, a sampling capacitor of $C_s = 1.05\mathrm{pF}$ is chosen so as to give a swing lower than 500mV.

Input resistance and voltage swing. The third component to be chosen is the input resistance R. As for C_d, the resistance choice is imposed by noise considerations. The input voltage swing $\|x\|_\infty$ is also related to the noise contribution. The next section addresses the issue with a simple model and determines $R = 6\mathrm{k}\Omega$ and $\|x\|_\infty = 500\mathrm{mV}$. Consequently, according to the design equations, the reference current becomes $I_{\mathrm{REF}} = 2.7\mu\mathrm{A}$ and the first integration capacitor $C = 10.8\mathrm{pF}$.

6.2.4 Noise considerations

The choices in steps two and three of the design procedure are imposed by noise considerations which put another constraint on the choice of components.

First stage. Let us start with the thermal noise contribution of the first stage. The noise power spectral density, referred to the modulator input $V^2_{n,\mathrm{IN},1}(f)$, of two input resistors R and two ND current steering DAC sources is given by

$$V^2_{n,\mathrm{IN},1}(f) = 8kTR + 8kT\gamma g_m \cdot \mathrm{ND} \cdot R^2 \; . \tag{6.17}$$

where g_m is the gate transconductance and ND the total number of DAC elements. To reduce the contribution of the current sources, they have to be sized so as to provide the lowest transconductance. The sourcing and sinking transistors are therefore placed in very strong inversion. In such a case, the factor $\gamma = 2n/3$, where n is the *slope factor*, and the transconductance is given by

$$g_m = \frac{2I_d}{nV_{DS,sat}} \; . \tag{6.18}$$

The transistor drain current I_d is equal to half the reference current I_{REF}. Moreover, the drain-source voltage can be set to the maximum allowed. According to the circuit description in Figure 6.2 and neglecting the voltage drop through the switches, it follows that $V_{DS} = V_{DD}/2$. Assuming that these transistors are at the limit of saturation $V_{DS} \cong V_{DS,sat}$ and

$$V^2_{n,\mathrm{IN},1}(f) = 8kTR\left(1 + \mathrm{ND} \cdot \frac{4I_{\mathrm{REF}}R}{3V_{DD}}\right) \; . \tag{6.19}$$

Introducing Equation (6.12) in this relationship gives

$$V^2_{n,\mathrm{IN},1}(f) = 8kTR\left(1 + \frac{\mathrm{ND}}{\mathrm{NL}-3} \cdot \frac{4\|x\|_\infty}{3V_{DD}}\right) \; . \tag{6.20}$$

Second stage. The second stage is made up of switched-capacitors. When referred to the modulator input, the contribution of the two sampling networks (ND DAC elements) is given by

$$V_{n,\mathrm{IN},2}^2(f) = \left[\frac{4kT}{C_s}\cdot\frac{2}{f_s}\cdot\left(\frac{C_s}{C_f}\right)^2 + \mathrm{ND}\cdot\frac{4kT}{C_d}\cdot\frac{2}{f_s}\cdot\left(\frac{C_d}{C_f}\right)^2\right]\left(\frac{C_f}{C_s}RCf_s\right)^2 |1-z^{-1}|^2\ . \tag{6.21}$$

This is further simplified as

$$V_{n,\mathrm{IN},2}^2(f) = \frac{4kT}{C_s}\cdot\left[1+\mathrm{ND}\cdot\frac{C_d}{C_s}\right]\cdot\frac{2}{f_s}\cdot(RCf_s)^2\,|1-z^{-1}|^2\ . \tag{6.22}$$

The contribution of the first- and the second-stages with reference to the modulator input can thus be summed and integrated over the band of interest f_b. The Signal-to-Thermal-Noise Ratio is finally

$$\mathrm{STNR} = \frac{\|x\|_\infty^2/(8kT)}{\underbrace{2Rf_b\left[1+\frac{\mathrm{ND}}{\mathrm{NL}-3}\cdot\frac{4\|x\|_\infty}{3V_{DD}}\right]}_{\text{CT-stage } 1.1\cdot 10^{10}} + \underbrace{\frac{1}{C_s}\cdot\left[1+\mathrm{ND}\frac{C_d}{C_s}\right]\cdot\frac{(\pi RCf_s)^2}{3\ \mathrm{OSR}^3}}_{\text{DT-stage } 7.8\cdot 10^{8}}}\ . \tag{6.23}$$

As shown with numerical values in Equation (6.23), the continuous-time stage contribution is more than a decade higher than the discrete-time stage contribution. In fact, the errors made in the second-stage benefit from a first-order shaping. The STNR is dominated by the parameters of the first-stage although a small DAC element capacitor C_d is chosen.

The second term in the expression can therefore be neglected and the constraint on R given as

$$R = \frac{\|x\|_\infty^2/(8kT)}{2 \cdot \mathrm{STNR} \cdot f_b \underbrace{\left[1 + \frac{\mathrm{ND}}{\mathrm{NL}-3} \cdot \frac{4\|x\|_\infty}{3V_{DD}}\right]}_{\cong 2}} . \tag{6.24}$$

This equation gives the relationship between the input resistance and input swing. Further simplifying the expression leaves only the fundamental parameters such as kT, the SNR and the band-of-interest f_b.

If we choose an STNR 6dB above the targeted SNR of 83dB, we find a resistance $R = 6.5\mathrm{k}\Omega$. The 6dB margin allows for thermal and flicker noise of the first amplifier and the flicker noise of the current steering DAC, which dominates.

As a consequence of the operating conditions of the current steering DAC, we can find the size ratio for both the sources and sinks

$$\frac{W}{L} = \frac{4I_{\mathrm{REF}}}{KP_{p,n} n V_{DD}^2} . \tag{6.25}$$

Thick oxide transistors are used for the sources and sinks. Both their flicker noise and transconductance coefficients are lower than for core devices. Transconductance coefficients of $KP_n = 230\mu\mathrm{V}/\mathrm{A}^2$ and $KP_n = 60\mu\mathrm{V}/\mathrm{A}^2$ are estimated for the N and P-MOS transistors respectively. The slope factor n tends to unity in strong inversion and we find size ratios of 1/18 and 1/5. Different adjustments are made based on simulation before reaching the final size found in Figure 6.2. Their area is increased to reduce the flicker noise contribution. The resulting transistors are therefore narrow and long, providing a large output impedance g_{ds}.

6.3 Design of the amplifiers

6.3.1 Behavioral models

The design of the amplifiers at the transistor level was carried out by D. Stefanovic in the frame of procedural analog design methodology described in [Ste07, SPPK07, SKP05, SKPK04a, SKPK04b].

The methodology proposed is based on the transistor model developed in [EKV95, Vit01], called the EKV model, referring to its inventors' names. Prior to the design flow, we substituted each of the three amplifiers shown in Figure 6.2 by an equivalent model with a limited set of parameters.

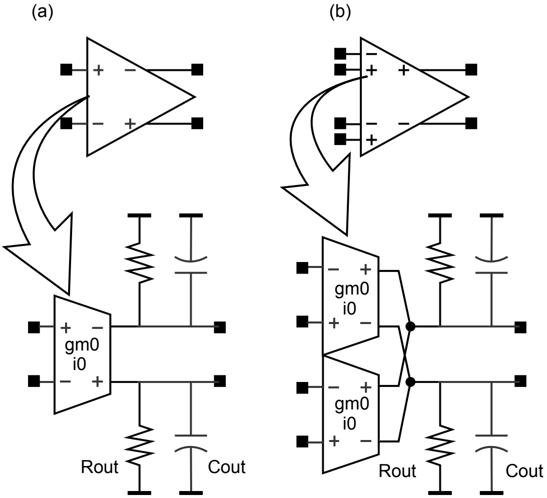


Figure 6.5 Behavioral modeling of a differential (a) and differential difference (b) amplifier.

The last two amplifiers are Operational Transconductance Amplifiers (OTA) whose behavior is well modeled by the transfer characteristic of a differential pair. The behavior of the differential pair, based on the EKV analytical model, in all regions of operation from weak to strong inversion, is provided in Appendix A.4 and implemented in the Verilog-A codes of Appendix B. As shown in Figure 6.5, the model of the differential pair, represented as a transconductance stage, is followed by an output resistance and capacitance, **Rout** and **Cout**. An ideal common-mode feedback is added, which is not shown here.

The differential pair quiescent transconductance g_{m0} and current i_0 and the output impedance are the parameters to adjust. From the EKV analytical model, we know that the highest efficiency, considering the ratio g_{m0}/i_0, is found in the weak inversion region. Nevertheless, the weak inversion implies large transistor sizes. In practice, it turns out that moderate inversion gives the best compromise for a differential pair. Under these conditions, a ratio g_{m0}/i_0 of 20 is systematically used and the parameter to be optimized is the current, along with the output impedance.

As shown in Figure 6.5(b), the differential difference amplifier, used for the sampler, is modeled by two transconductance stages combining their currents through the same output impedance. The continuous-time integrator is implemented with a two-stage amplifier. The simplified model given by Code B.2 can be used. In contrast to the model of Code B.1, the behavior is controlled by the quiescent transconductance and current without restriction on their ratio. In such a case, i_0 models the slewing current available, and g_{m0} the input pair transconductance amplified by the second stage.

These behavioral models help in optimizing the current consumption, related to the parameter i_0. Many iterations are necessary to find a good combination between a realizable output impedance and the transconductance stage.

6.3.2 Continuous-time amplifier

As shown in [Yan02] and confirmed by behavioral simulations, the continuous-time integrator requires less bandwidth and slewing capability. Nevertheless, a large DC gain is still necessary, especially for the first integrator where the non-linearity directly impacts the resolution without any spectral shaping. Therefore the two-stage topology shown in Figure 6.6 was chosen. The topology consists of a folded cascode first-stage followed by a common source amplifier with cascode-Miller compensation. More details are provided in [Ste07, HH96] about how to adequately damp the resonance introduced by such a compensation.

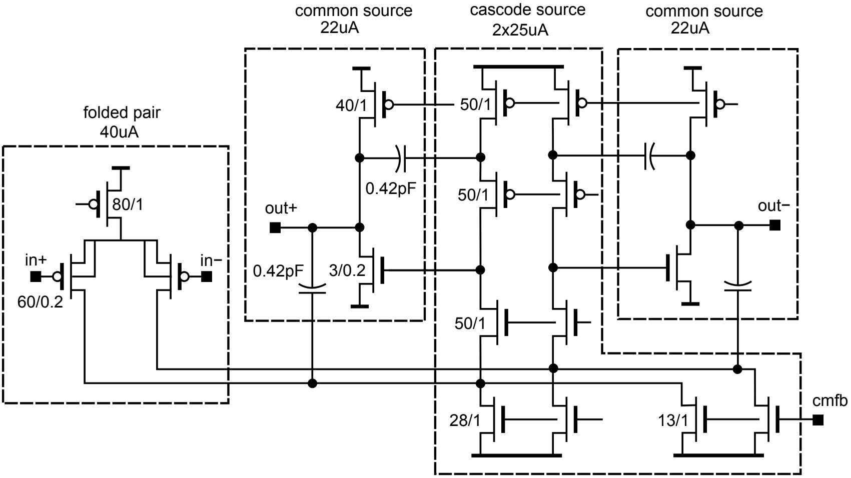


Figure 6.6 Continuous-time integrator amplifier.

The differential gain reaches 78dB with a bandwidth of 71MHz and a phase margin of 80°. The common-mode gain reaches 64dB with a bandwidth of 7.5MHz and a phase margin of 72°. Time-domain simulations of the system with this amplifier, the sampling amplifier and the i-DAC at transistor level, gave an SNDR of 94.3dB.

The common-mode feedback is realized with an additional current source controlled by the circuit in Figure 6.7. The same circuit is used for all the three amplifiers. The topology comprises two differential pairs. Their output currents are combined such as to provide a current proportional to the difference between the common-mode of **out+** and **out−** and the reference node **ref**. This current is therefore mirrored and either added or used directly in the active load.

Transistor level simulations revealed an important leakage from the differential mode to the output common-mode. common-mode output voltage variations of about 100mV peak-to-peak were observed. Such large common-mode variations drastically reduce the amplifier swings. Since the common-mode Rejection Ratio (CMRR) is limited, the variations propagate through the system with an additional contribution at each amplifier output.

This leakage is mainly due to the highly non-linear nature of the differential pairs. A large differential signal at nodes **out+** and **out–** saturates the pairs. Consequently, the output common-mode tracking is compromised. The degeneration of the pairs with transistors, as proposed by [Vit01] and [Joh92], linearizes the transfer characteristic and partially solves the problem. Besides, while one pair is fed with **out+** and **ref**, the other is fed with **out-** and **ref**. As the differential output voltage increases, the two pairs experience different input common-modes. Because of the Early effect, the pair's current sources are different and the calculation of the output common-mode error is again compromised. Cascoding the pair sources reduces this effect. The degeneration of the pairs together with the cascoding of the current sources allowed the common-mode variations to be reduced by a factor of 30.

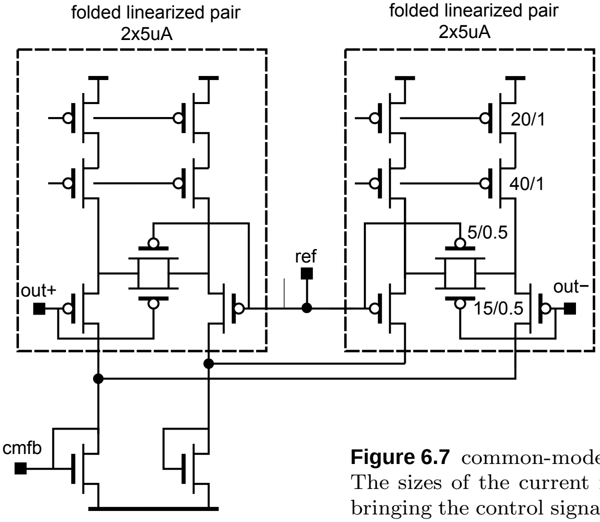


Figure 6.7 common-mode feedback circuit. The sizes of the current mirror transistors bringing the control signals **cmfb** are different for each amplifier and are therefore not displayed.

6.3.3 Switched capacitor devices

Switches. All the switches in the sampling stage are controlled by the voltage elevator shown in Figure 6.8. The topology is based on the bootstrapping switch used by [SG04]. In a bootstrapping configuration, the **shift** port is connected to either switch terminals **a** or **b**. Simulations showed better results with the shift port at VDD, providing a higher driving voltage, whereas the bootstrapping tries to lower the resistance dependence on the terminal voltage. The charging phase switches of the discrete-time DAC, shown in Figure 6.2, were also controlled by the same circuit with smaller sizes.

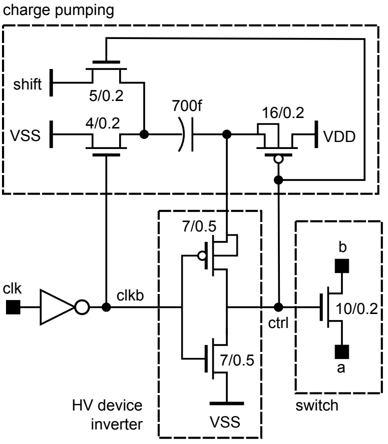


Figure 6.8 Sampling switches and control circuit.

Sampler. The sampling device is composed of a differential difference transconductance amplifier in a unity gain configuration, as shown in Figure 6.2. Figure 6.9 shows the topology chosen. The task of this amplifier is to charge the sampling capacitors which are fully discharged during the integration phase. Thanks to the clocking scheme used, two half clock-cycles are allowed for the signal to settle.

Two complementary differential pairs are folded into a high-impedance load. This wide-swing capability is required since, in a voltage follower configuration, the input nodes experience the same large signal as the outputs. The differential pair sources are cascoded to reject as much as possible the input common-mode. The high impedance is provided by cascode current sources. Gain boosting devices, made of simple common-source amplifiers, further enhance the impedance at low frequencies to provide a large DC gain. The common-mode feedback circuit in Figure 6.7

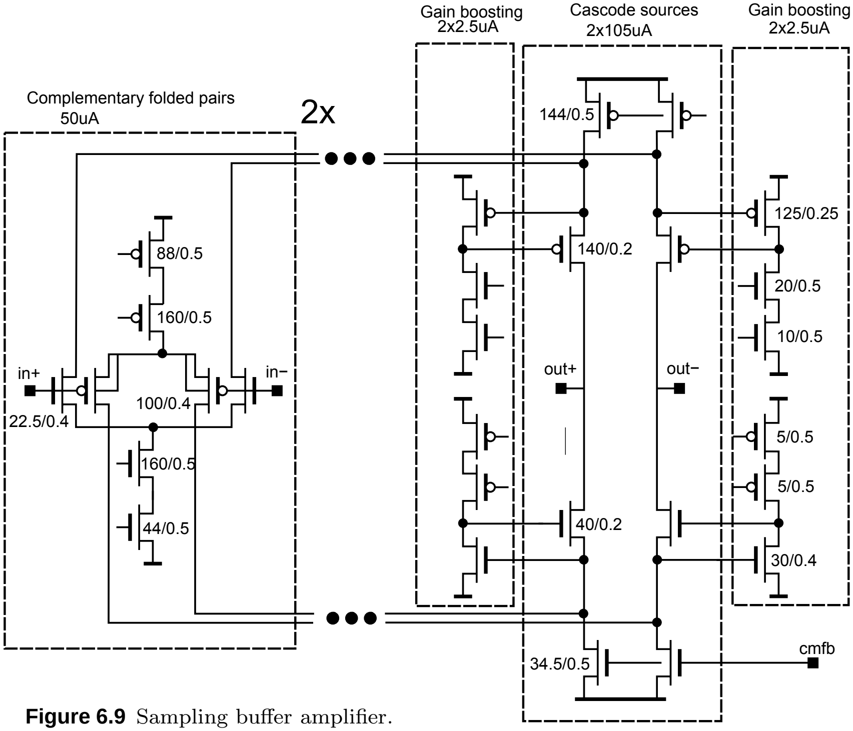


Figure 6.9 Sampling buffer amplifier.

is connected to the node **cmfb**. It directly controls the N-MOS current sources of the output stage. The common-mode loop gain is low enough to keep a good phase margin and ensure its stability.

The differential-difference gain reaches 74dB with a bandwidth of 104MHz and a phase margin of 76°. When the 1pF sampling capacitors are disconnected, the bandwidth goes up to 300MHz and the phase margin down to 52°. This situation occurs only during a short period of time between the end of phase **q2** and the beginning of phase **s2** and vice versa. The common-mode gain reaches 46dB with a bandwidth of 47MHz and a phase margin of 90°. Time-domain simulations of the system with the sampling amplifier and switched network at transistor level gave an SNDR of 92.9dB.

DT-integrator. The amplifier in the DT integrator is shown in Figure 6.10. The chosen topology is a folded cascode structure with a gain enhancement. The same common-mode circuit described previously in Figure 6.7 is used. Nevertheless, to ensure the stability of the common-mode loop, the N-MOS current sources of the output stage are split, as in the case of the CT integrator.

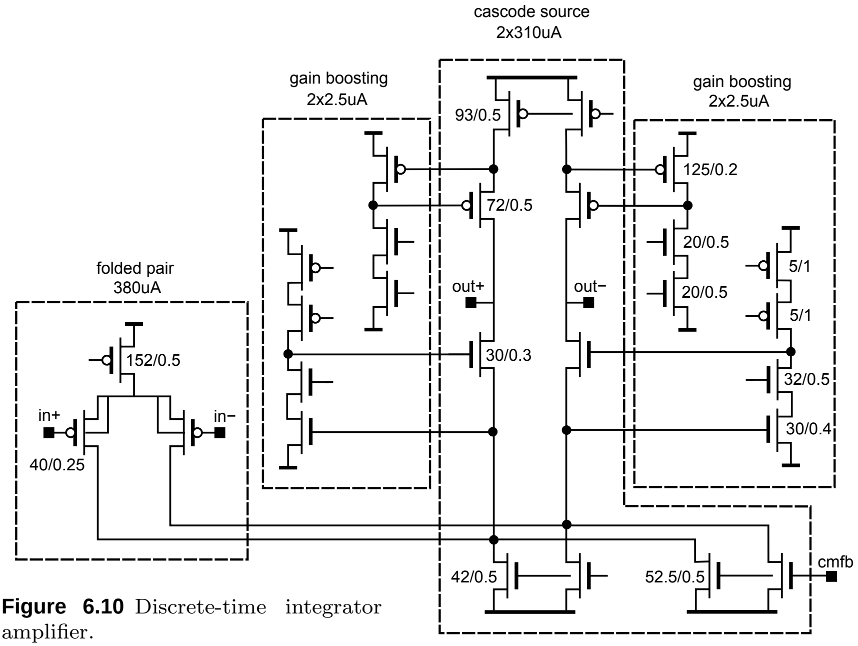


Figure 6.10 Discrete-time integrator amplifier.

The operating conditions are similar to those of the sampling amplifier. The load is capacitive and the output impedance is determined by the amplifier output stage. Nevertheless, the constraints are higher here. In contrast to the sampling case, this amplifier has less that half a clock-cycle to perform the charge transfer. Secondly, as shown in Chapter 4, the auto-ranging algorithm relies on the accuracy of this integration path to provide well-matched analog and digital shifts. Furthermore, the signal fed back through the sc-DAC has increased activity with respect to the i-DAC. In fact, this integrator has to perform two different tasks: the integration of the quantization errors and the shifting of the quantizer input.

Because of this, the slewing capability and the transconductance are more than doubled. These higher specifications require a larger current in the output stage. Consequently, the output impedance drops. The cascoded sources are sized accordingly and gain boosting devices enhanced with cascode transistors.

The differential gain reaches 79dB with a bandwidth of 660MHz and a phase margin of 75°. The common-mode gain reaches 23dB with a bandwidth of 110MHz and a phase margin of 62°. Time-domain simulations of the system with this amplifier, switched network and the sc-DAC at transistor level gave an SNDR of 90.8dB. A degradation of 2dB is brought by the DAC switches.

6.4 Quantizer design

6.4.1 Comparator

The quantizer consists of a bank of 10 comparators, such as described in Figure 6.11. The **ref–** and **ref+** ports are connected to the differential reference ladder shown in Figure 6.2, and **in+** and **in–** to the last amplifier output.

The structure is composed of a preamplifier and a clocked latch followed by flip-flops. The two preamplifier input pairs perform the difference between the input and the reference. The differential pairs are cascoded in order to keep the same characteristics over the whole range of threshold references, from $- -$ 360mV to $+$360mV. The active load is made up of current mirrors in a latched configuration. The one-by-one current ratio gives no hysteresis [AH02, JM97]. Such an active load has the advantage of not requiring a common-mode feedback while providing high impedance.

Minimum size transistors are used to improve the speed and to reduce the capacitive load seen by the last amplifier. Time-domain simulations of the system with the quantizer and the last amplifier at transistor level gave an SNDR of 93.3dB.

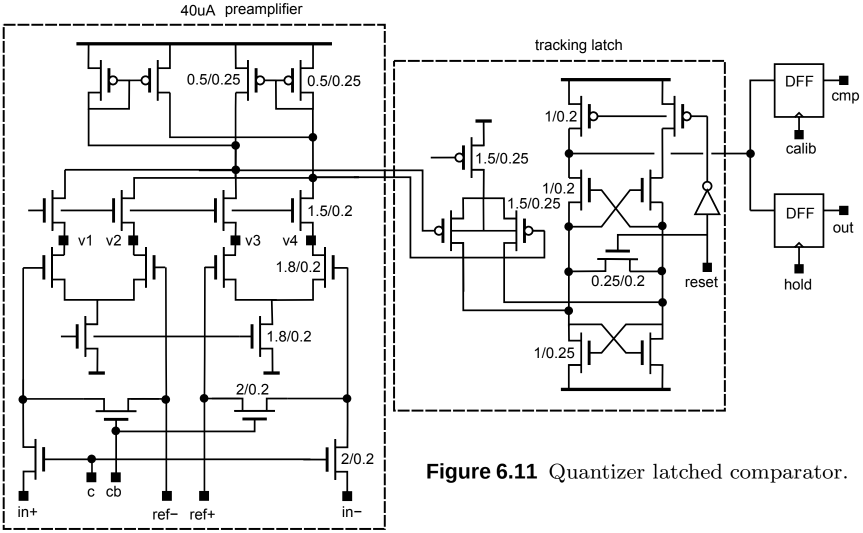


Figure 6.11 Quantizer latched comparator.

As a result, the statistical offset is large. A transient simulation with a successive approximation algorithm is used to determine the offset of the entire comparator chain. The simulation is preformed 2000 times in a Monte-Carlo analysis. The resulting offset distributions displayed in Figure 6.12 show a standard deviation of about $\sigma = 12\text{mV}$.

According to the offset sensitivity determined in Chapter 4, the offset should remain lower than 3mV. The digital compensation proposed by [PK06] is used here. For that purpose, access to the comparator inputs are controlled by the **c** and **cb** ports. The compensation consists in shorting the comparators input over six clock-cycles. A successive approximation algorithm determines sequentially the 6 bits of a current DAC. The current DAC feeds differential currents at nodes **v1**, **v2**, **v3**, **v4**. The compensation cycles are carried out during the half clock cycles where the comparator is not in use. Figure 6.13 provides the clocking diagram. The 10 comparators of the bank are calibrated sequentially. Therefore 60 clock cycles are necessary for a complete calibration of the quantizer.

This compensation method has a few advantages, the first being that the calibration requires no interruption of the modulator. Nevertheless, to avoid any perturbation, the compensation process can be applied oc-

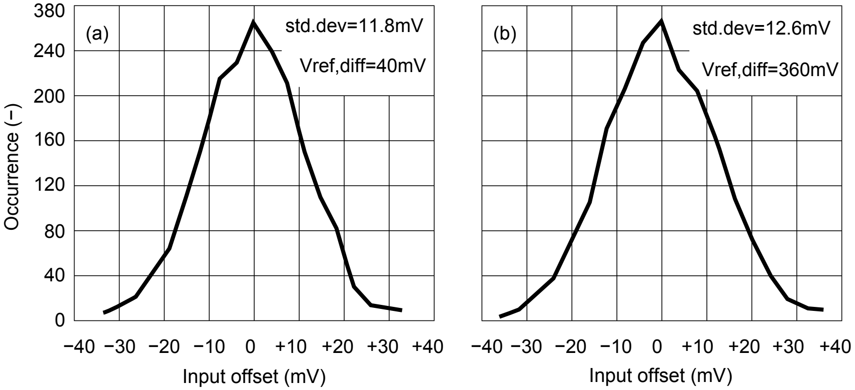


Figure 6.12 Simulated offset distribution with Monte-Carlo analysis for a differential reference voltage of (a) 40mV and (b) 360mV.

casionally. In contrast to the conventional compensation with coupling capacitors, here the entire comparator chain, including the latch, is corrected. In this design the compensation registers and the algorithm are placed in the digital section. An alternative solution consists in keeping the registers within the comparator cell to avoid routing the 60 bits.

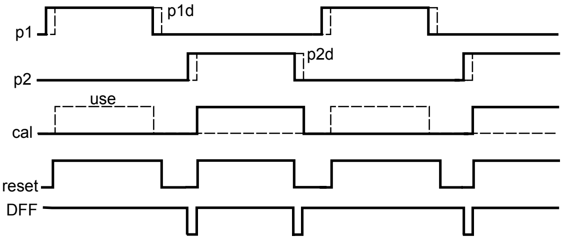


Figure 6.13 Clocking diagram for comparators.

6.4.2 Offset compensation circuit

The compensation is applied via the 6-bit M-3M current DAC described in Figure 6.14 as proposed by [Pas05, PK06]. The DAC works like a digital differential pair placed in parallel to the analog pairs of the preamplifier. A one-to-three transistor network splits the current sink into 6 current branches with different weights. The transistors controlled by the digital ports **x0** to **x5** also participate in the splitting. At the same time, they allow the choice of sinking each weighted current into either the positive or the negative branch of the current collector.

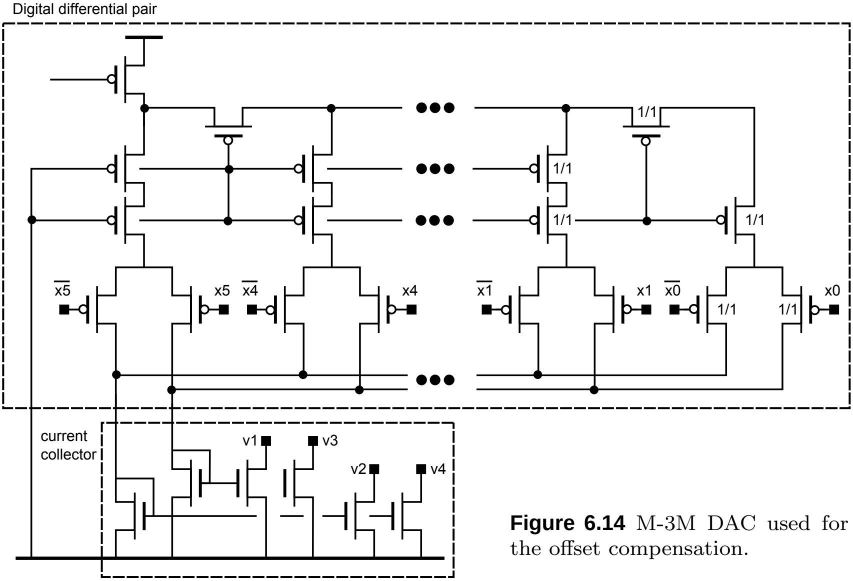


Figure 6.14 M-3M DAC used for the offset compensation.

[PK06] studied the M-2M, M-2.5M and M-3M DAC topologies. An ideal M-2M topology splits its base current into power-of-two weighted currents. Nevertheless, the transistors in this ladder topology does not have the advantage of good matching properties. According to [PK06], the mismatch tends to reduce the effective resolution of the DAC. Figure 6.15 shows the effective achievable resolution found with Monte-Carlo simulations with different topologies. In this transistor level DC simulations, all 64 levels are tested. The smallest step size is extracted and related to the full swing of the DAC.

The M-2M topology can achieve up to 5 bits of resolution but runs the risk of reaching 3.4 bits. Such a wide-spread resolution is not acceptable. In contrast, the M-2.5M topology, such as described in [PK06], achieves lower resolutions but has a narrower distribution. Finally, the M-3M presents a well-controlled resolution of 4.6 bits within ±1% of tolerance.

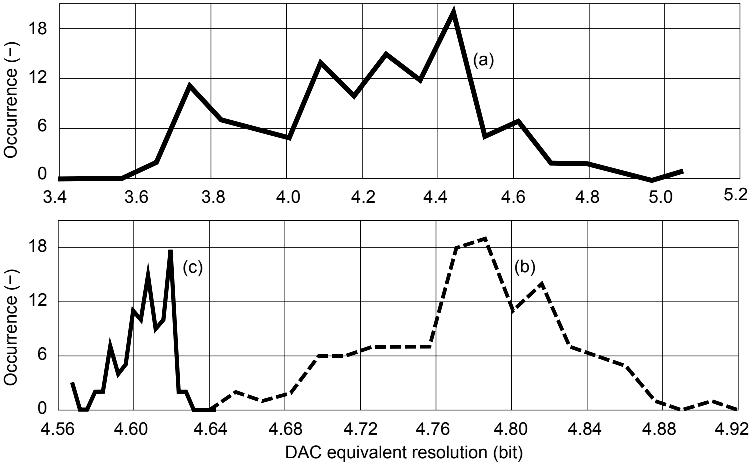


Figure 6.15 Simulated DAC effective resolution for the M2M (a), M2.5M(b) and M3M (c) architectures.

Since the process data used in Monte-Carlo were extracted for matched transistors, we expect worse matching properties in reality. Consequently, the M-3M topology insures an acceptable compensation at the cost of a reduced resolution.

6.4.3 Residual offset

By choosing the M-3M topology we set the number of levels to 24, here referred to the variable n. We now need to choose the DAC steps size Δ which is related to the compensation range $\Delta \cdot n$.

Let us consider the offset has a Normal distribution $p(x)$ with a standard deviation of σ.

$$p(x) = \exp\left(\frac{-x^2}{2\sigma^2}\right) / (\sqrt{2\pi}\sigma) \ . \tag{6.26}$$

According to the illustration in Figure 6.16, each segment folds back to the center of the distribution. As a result, the new probability density

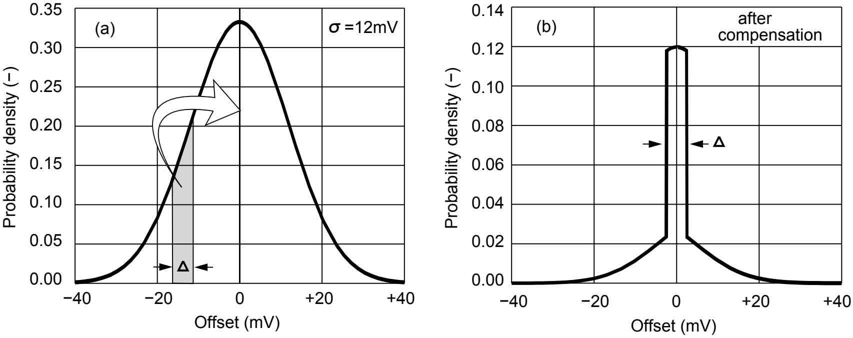


Figure 6.16 Statistical distribution of the offset before (a) and after (b) compensation with the analytical model.

function is

$$q(x) = \sum_{i=0}^{n-1} f(x,i) \ , \tag{6.27}$$

where

$$f(x,i) = \begin{cases} 0, & \begin{array}{l} \text{if } (i=0) \text{ and } (x < -\frac{\Delta}{2}) \ , \\ \text{or} \quad \text{if } (0 < i < n-1) \text{ and } (|x| < \frac{\Delta}{2}) \ , \\ \text{or} \quad \text{if } (i = n-1) \text{ and } (x < \frac{\Delta}{2}) \ , \end{array} \\ p(x - \Delta\left(i - \frac{n-1}{2}\right)) \ , & \text{otherwise} \ . \end{cases} \tag{6.28}$$

Let us recall that the standard deviation σ is defined as

$$\sigma = \int_{-\infty}^{+\infty} x^2 q(x) \mathrm{d}x \ . \tag{6.29}$$

Meanwhile, we define here another deviation σ' for which there is a probability of 68% of finding x to within $\pm\sigma'$. The definition is therefore given by the equation

$$\int_{-\sigma'}^{+\sigma'} q(x) \mathrm{d}x = \mathrm{Erf}\left(\frac{1}{\sqrt{2}}\right) \cong 0.683 \ . \tag{6.30}$$

The standard deviation σ provides a Root Mean Square (RMS) value of the errors. In contrast, the deviation σ' links the distributions $q(x)$

and $p(x)$ according to their most likely maximum error. On applying the Normal distribution $p(x)$, both deviations give the same result and we find $\sigma = \sigma'$.

Figure 6.17 displays these two deviations with respect to Δ. As expected, for a Δ of zero the residual offset is equal to the original value of 12mV. No compensation is provided in this case. On the other hand, if a too large value of Δ is chosen, the comparator offset is over-compensated. The 24 levels used here are therefore wasted. The curves reveal the presence of two different optimal values considering either σ or σ'.

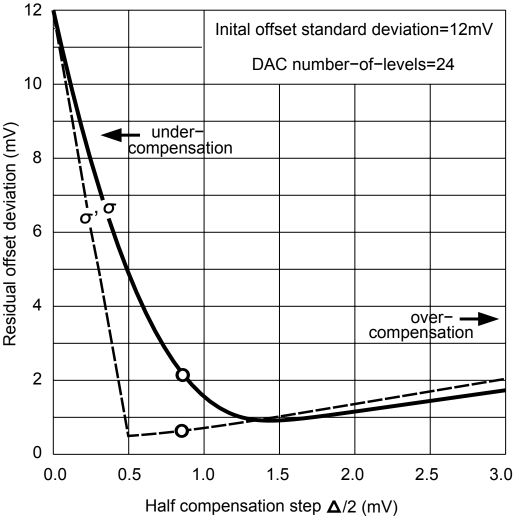


Figure 6.17 Statistical deviation expected after compensation with a 24-level DAC and different step size Δ. An initial offset of 12mW is considered. The dots show the simulated compensation of Figure 6.18.

Figure 6.18 shows the offset after compensation. These are the results of a transient simulation at the transistor level. A dedicated test bench was built for that purpose. During the simulation, the compensation cycle is applied followed by another successive approximation cycle to determine the residual offset. The same simulation is performed 800 times in the Monte-Carlo loop. The residual offset has approximately the characteristics predicted by the chart in Figure 6.16. The distribution width and the standard deviation are roughly equal to 1.7mV and 2.2mV respectively. In fact, the model differs from the simulation in that an ideal DAC is considered here with n constant steps of Δ.

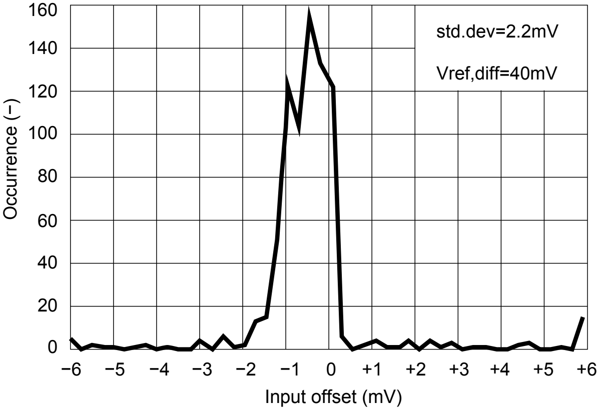


Figure 6.18 Offset statistical distribution after compensation. The simulation is performed at transistor level with 800 Monte-Carlo points.

6.5 Final circuit

The final circuit was designed and brought to layout. Figure 6.19 shows the output power spectral density of a simulation at the transistor level with an achieved SNDR of 90.4dB. The physical layout of the circuit is displayed in Figure 6.20, whereas Table 6.4 summarizes the die area and consumption of the main sections. The digital section consists of two second-order mismatch shaping encoders, the auto-ranging algorithm and the comparator offset compensation. The HDL description of the circuit was synthesized, placed and routed automatically using conventional CAD tools. The area is small with respect to the rest of the circuit. This clearly demonstrates that the digital solutions chosen to solve analog issues, like the ADC and DAC element mismatch, are cost effective.

The auto-ranging algorithm allows a large internal number-of-levels and, consequently, low sampling frequency and shaping order. Meanwhile, the sensitivity to clock jitter, brought about by the continuous-

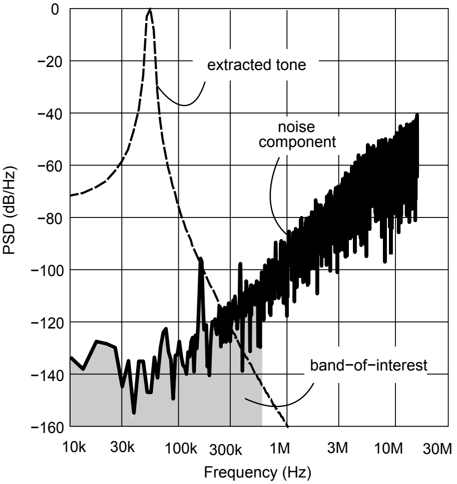


Figure 6.19 Power spectral density of the modulator output simulated at transistor level.

Table 6.4 Consumption and chip area summary.

Component	Current, μA	Area, mm^2
CT integrator amp.	205	0.03
SC integrator amp.	1150	0.04
Sampler amplifier	425	0.04
i-DAC	100	0.04
SC-DAC	140	0.04
Quantizer	545	0.05
Digital circuits	500	0.04
Interconnect and 10pF	-	0.07
Sampling capacitors	-	0.10
Total	3200	0.54

time integrator, is reduced. The expected increased size of the quantizer is circumvented. A quantizer with three times the size, consumption and capacitive load, would have been needed without the algorithm. In contrast, the size of the i-DAC and sc-DAC banks is reasonable, the large number-of-levels notwithstanding. Furthermore, the space occupied by the digital section, by its register and three adders is negligible.

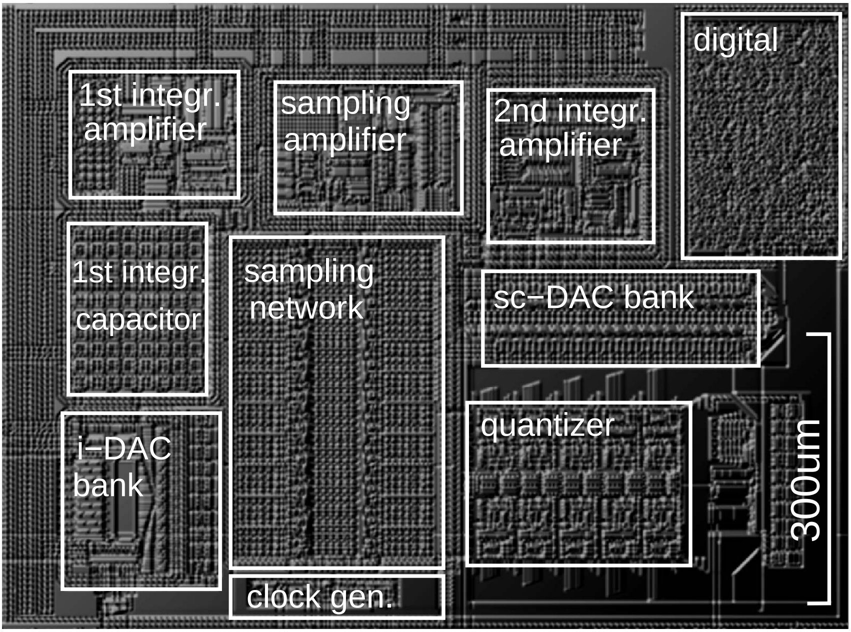


Figure 6.20 Layout of the circuit in a 0.18μm CMOS technology.

The sampling network has been doubled to halve the slew-rate requirement of the sampling amplifier. This roughly corresponds to a 10% increase in area against a 15% improvement in power consumption.

In a pure switched-capacitor implementation, the first-stage amplifier would burn more power than the second-stage amplifier. In contrast, the summary of Table 6.4 shows that the consumption of the first integrator amplifier represents only a fifth of the discrete-time integrator amplifier.

Chapter 4 showed that the auto-ranging requires the last feedback path to be accurately controlled, putting a higher constraint on this amplifier. However, the excessive consumption of the discrete-time integrator is mainly due to the increase by a factor of three of its gain. Though a benefit for the sensitivity to the comparator offsets, this increases the *effective* capacitive load C_{eff} of the amplifier. According to [Bur02] we can write for a simple model that

$$C_{eff} = C_o + C_s + C_i + C_o \frac{C_s + C_i}{C_f}\,, \tag{6.31}$$

where C_o and C_i are the amplifier input and output parasitic capacitances, respectively. In this relationship, C_o is multiplied by a ratio that is close to C_s/C_f. Thus, on the one hand, the auto-ranging algorithm allowed to reduce by a factor of three C_o by reducing the number of comparators. But on the other hand, to alleviate the offset sensitivity, the ratio C_s/C_f was increased by the same factor. This drawback could be addressed by relying more on the offset calibration and reducing the gain.

The sampling amplifier consumes here barely half of the current dissipated in the last stage. In a pure switched-capacitor implementation, the sampling process would be in front of the modulator and would require an amplifier consumption of several milliamps.

Finally, it was shown in Section 2.2.6 that with an aggressive pole placement the modulator could reach an SQNR of 100dB. Choosing to keep 94dB instead would allow the sampling frequency to be decreased to about 24MHz. The switched-capacitor amplifier consumption would have been reduced accordingly by about 30%, bringing the total consumption down to 2.8mA.

6.5.1 Benchmarking

One of the most commonly used *Figure-Of-Merits* in the field of ΔΣADC converters, providing a mean of comparison, is given by

$$\mathrm{FOM} = \frac{P}{2^{\mathrm{ENOB}} f_b}\,, \tag{6.32}$$

where P is the total power consumption of the modulator alone and f_b the band-of-interest. The Effective Number Of Bits (ENOB) is estimated from the Signal-to-Noise-plus-Distortion Ratio (SNDR) as

$$\mathrm{ENOB} = \frac{\mathrm{SNDR_{dB}} - 1.76}{6.02}\,. \tag{6.33}$$

Table 6.5 summarizes some of the most relevant recent publications. The first three citations are multi-bit hybrid architectures followed by four pure Continuous-Time (CT) implementations.

Table 6.5 Performance comparison.

Reference	Architecture	Supply V	Power mW	f_b Hz	f_s Hz	SNDR dB	FOM pJ
1) [MCL$^+$05]	Hybrid 2nd-order 4-bit (RC-SC)	3.3	37	20k	6.1M	95	40
2) [KRSC05]	Hybrid 2nd-order 4-bit (RC-SC)	3.3	18	20k	3.0M	99	12
3) [RLS$^+$03]	Hybrid 6th-order 3-bit (LC^2-RC^2-SC^2)	3.0	21	333k	32M	90	2.4
4) [DKG$^+$05]	CT 3rd-order 3-bit (RC^3) + tracking ADC	1.5	3.0	2M	104M	70	0.7
5) [Kap03]	CT 2nd-order 4-bit (RC^2)	1.8	2.2	1M	48M	61	2.5
6) [Phi03]	CT 5th-order 1-bit (RC-gmC^4)	1.8	2.2	500k	64M	75	0.96
7) [vV03]	CT 5th-order 1-bit (RC-gmC^4)	1.8	3.8 4.1 4.5	200k 600k 1.9M	26M 77M 153M	87 83 72	1.0 0.59 0.73
8) This work	Hybrid 2nd-order 5-bit (RC-SC)	1.8	5.8	500k	32M	83	1.0

Citation number 4 used a multi-bit tracking quantizer. Numbers 4 and 7 used a single switched-capacitor feedback in order to reduce the sensitivity to clock jitter. These two last examples achieved by far the best performance with FOMs lower than 1pJ.

Nevertheless, we should consider that Numbers 6 and 7 used a high sampling frequency while addressing a specification similar to the one chosen in this design example. In both cases, the strategy consisted of achieving the required resolution and bandwidth with a single-bit quantizer and a high order of spectral shaping. We know from the resolution equation (2.37) that increasing the order has a limited effect on the resolution when the over-sampling is low. As a result, the sampling frequency is in both cases at least twice the one used in this work.

The recourse to a high sampling frequency has a great impact on the consumption of the Phase Locked-Loop (PLL). The FOM does not take into account the consumption of such an external device. The increased consumption when using high sampling frequencies is therefore hidden. Moreover, on a large System on Chip (SoC), including an RF front-end, the presence of such high-frequency components could affect the receiver sensitivity. The clock signal over-tones corrupt the RF signals through substrate coupling and the power supply lines. In contrast, choosing to address the same specifications with a large number of levels allows the sampling frequency to be low and circumvents this issue.

According to the trend given in Figure 4.22, a second-order modulator with 11-level quantizer would address a target of 70dB over 2MHz with a sampling frequency of 80MHz. In contrast, Citation Number 4 of Table 6.5, which used a tracking quantizer, a method similar to the auto-ranging algorithm, required a clock at 104MHz. As in conclusion, Citation Number 4, together with this work, showed that continuous-time and hybrid $\Delta\Sigma$-modulators with multi-bit emulation are promising architectures.

As for Citations 1 and 2, which address audio applications with high resolution and small bandwidth, the achieved FOM is rather low. According to the trend in Figure 4.22, less than 2MHz of sampling frequency would be required by a second-order modulator with the auto-ranging algorithm. The algorithm efficiency would be extremely high with more than 80 emulated levels from a 10-comparator quantizer.

6.6 Conclusions

A complete design was carried out from the high-level specification down to the circuit description and layout in a CMOS technology. The low-power design strategy of Chapter 3 was applied. The resulting performance, power consumption and die area, are comparable to the best published solutions. The comparison highlighted that, today, two different strategies have proved to achieve high-performance and low power consumptions. On one side, we find the high-order single-bit continuous-time and, on the opposite side, the low-order multi-bit hybrid implementations. However, we have demonstrated that the strategy proposed here offers better extendability to various specifications and applications. In particular, it would allow for a significant reduction of the sampling frequency for wide bandwidth applications, which becomes mandatory in large multi-standard SoCs addressing the consumer market today.

Chapter
7
Conclusions

7.1 Thesis outlook

This thesis proposes a low-power methodology for Analog-to-Digital Converters (ADC) based on multi-bit $\Delta\Sigma$-modulation. At the system level, the design of a single-stage modulator has three degrees of freedom: the internal Number-of-Levels (NL), the Over-Sampling Ratio (OSR) and the spectral shaping order (n). The equation of the Signal-to-Quantization-Noise Ratio (SQNR) gives the expected ADC resolution as a function of this set of parameters. It shows that, for a targeted resolution, different sets of n, NL and OSR are possible. It is common to find throughout the literature of the past two decades ADCs realized with single-bit modulators. In such cases, NL is set to its minimum value of 2. The modulator has only one comparator and one element in the feedback DAC. Moderate resolutions up to 12 bits are achieved thanks to a large OSR. Single-bit architectures have been highly praised for the simplicity with which such a resolution can be achieved.

However, today's portable communication applications demand the same moderate resolution but with bandwidths of up to several megahertz. Relying on large OSR only would result in very high sampling frequencies. Recent publications have demonstrated the low power consumption of single-bit modulators with continuous-time implementations. The required resolution is reached thanks to an increased spectral shaping order and a moderate OSR. However, the sampling frequency can still exceed 100MHz. Moreover, part of the consumption has simply moved from the modulator to the Phase Locked Loop (PLL) generating the clock signals. Additionally, the presence of such a fast clock signal on a large

RF System-on-Chip (SoC) is not suitable because of the risk of unwanted mixing with RF signals.

In this thesis we propose a low-power strategy relying on a large NL rather than on a high OSR or shaping order. The sampling frequency remains low. However, as highlighted in [Le 05], as the available voltage supply is reduced in modern deep sub-micron technologies, the internal number-of-levels becomes limited. Moreover, increasing NL results in a large capacitive load for the last integrator, which could increase the power consumption in that stage. To circumvent these issues, an auto-ranging algorithm was developed in this thesis. The algorithm emulates the internal levels from a reduced-size quantizer. The technique reuses the feedback path of the modulator to shift the quantizer analog input. No extra analog circuitry is necessary. The same shift is applied on the digital part.

In contrast to single-bit modulators, multi-bit architectures require a Dynamic Element Matching (DEM) algorithm. Choosing to increase the NL therefore has an impact on the digital complexity. This issue is addressed with an appropriate segmentation of the DACs. A synthesis algorithm is proposed in this thesis, allowing all the possible tree structures to be found and the most efficient one in terms of power consumption to be chosen.

As part of the strategy to further reduce the power consumption, the first-stage of the modulator is realized as a continuous-time integrator. The choice of a large NL significantly alleviates the sensitivity to clock-jitter which is the main drawback of continuous-time implementations. The continuous-to-discrete-time interface issue is addressed with a double sampling scheme which allows twice as much time to charge the sampling capacitors of the second stage.

The auto-ranging algorithm requires an accurate last feedback integration path. To that extent, it is necessary to realize the last integrator with a switched-capacitor circuit, therefore relying on the good matching properties of MIM capacitors or metal fringe capacitors available in modern CMOS technologies. We can also claim that keeping the upper stages of the modulator in the discrete-time domain allows a design with more aggressive quantization noise shaping. As a result, with such a hybrid architecture, a further reduction of the sampling frequency is possible.

The methodology presented in this dissertation is sustained by a design example addressing the specifications for a typical BLUETOOTH receiver with direct conversion. A 33-level second-order modulator is designed from these specifications down to the physical layout description. Other constraints, like the presence of large out-of-band interferers, were taken into account. The architecture parameters NL, OSR and n were chosen such as to optimize both the sensitivity to clock jitter and the auto-ranging efficiency.

The proposed strategy appears to be an optimal combination of different techniques in view of reducing the power consumption, the voltage supply and the sampling frequency. These three criteria are essential when developing low-cost large SoCs for future portable applications.

7.2 Main contributions

An auto-ranging algorithm which allows a reduction in the number of comparators in a multi-bit $\Delta\Sigma$-modulator has been developed. With respect to similar techniques proposed by other authors, this algorithm does not require any extra analog circuitry. Instead, an existing part of the modulator is reused. Moreover, the proposed implementation can be used with low voltage supplies. A patent application has been filed in collaboration with Freescale Semiconductor.

The limitation of the algorithm due to large out-of-band interferers, was studied analytically. The development revealed the optimal modulator parameter set that gives the highest efficiency in reducing the number of comparators. Additionally, the extendability to different applications was studied. Behavioral simulations showed that the sensitivity to the offset of the comparators remains unchanged, but the required accuracy on the last feedback path is increased.

An in-depth analytical development of the sensitivity to clock jitter was carried out, also revealing an optimal modulator parameter set. Since the auto-ranging efficiency and the jitter sensitivity mechanism are related to the same phenomenon, the same optimal solution was found.

A synthesis algorithm for segmented tree-structured DEMs was developed. The synthesis revealed an optimal segmentation providing a significant power saving. An analytical expression of the power consumption has been derived for the standard segmentation.

7.3 Future perspectives

The auto-ranging algorithm is a really promising technique. However, only part of its potential has been exploited here. First of all, it was shown that the efficiency is limited by large out-of-band interferers. Hence, a minimum OSR with respect the fastest concerned input frequency is required. In depth investigations should help to find out how to circumvent this important restriction. For instance, the re-centering decision scheme used here was rather conservative. An anticipating scheme, shifting the signal back to a level beyond the center of the quantization window, may allow interferers to be tracked at higher frequencies. Secondly, in this thesis only integer values for the last feedback coefficient were considered. This limits the degrees of freedom when designing a modulator with an aggressive noise shaping. One possible way of removing this restriction is to use a separate feedback for the shifts.

The basic principles of the auto-ranging have been analysed in detail. Nevertheless, an analytical model to predict the sequence of shifting pulses is necessary. Such a model would help to predict the extra activity processed by the last integrator. By this mean, an estimation of the additional power consumption would by possible. Similarly, the sensitivity to circuit imperfections, such as the amplifier DC gain, the comparator offset and the last coefficient accuracy could be analyzed. All these aspects, if studied in detail, would help to determine the achievable performance when addressing different applications.

The synthesis algorithm developed for the segmented DEM can be made more efficient and implemented in a compiled code to reduce the computing time. This is essential if a very large number of levels is to be considered.

More investigations are also necessary at the circuit level. An analytical model of the amplifiers would allow the influence of the auto-ranging on the consumption to be understood and also the exact role of the interfacing buffer. Besides, a general comparison of the continuous- and discrete-time integrators may reveal that with a very large NL their contributions to the power consumption are of the same order of magnitude. This would also help when revisiting the amplifier topology choice made in the design example. In particular, it should be noted that [Bur02] concludes that two-stage amplifiers are better suited for switched-capacitor circuits with a large gain, such as the last integrator of the example.

Finally, increasing the order of the modulator is an option to be considered to further increase the noise shaping and reduce the sampling frequency over very large signal bandwidths. The first-stage would, of course, remain a continuous-time integrator and any further stages should be discrete-time integrators.

Appendix A

Mathematical definitions and proofs

A.1 Definition of the norms

The different norms in use throughout this document follow the definitions provided by the dictionary of mathematics [Uni97], where the p-norm, in the general case, is defined as:

$$\|x\|_p = \left(\sum_{i=0}^{n} |x_i|^p \right)^{1/p} \qquad , \forall p \in \mathbb{R}_+ \ . \tag{A.1}$$

In particular, the 1-norm and 2-norm of a finite impulse response z-transfer-function, such as

$$A(z) = \sum_{i=0}^{n} a_i z^{-i}, \tag{A.2}$$

where the coefficients $a_i \in \mathbb{R}$ are defined as the sum of the absolute value of the coefficients and their square:

$$\|A(z)\|_1 = \sum_{i=0}^{n} |a_i| \ , \tag{A.3}$$

$$\|A(z)\|_2 = \sqrt{\sum_{i=0}^{n} |a_i|^2} \ . \tag{A.4}$$

Besides, the ∞-norm of a sequence x_i becomes the highest absolute value of the sequence:

$$\|x\|_\infty = \max(|x_i|) \ . \tag{A.5}$$

A.2 The 1-norm of a z transfer-function

The 1-norm is used in Chapter 2 to calculate *noise range factor* k_q defined as

$$k_q = \left\| \prod_{i=1}^{n} (1 - z_i z^{-1}) \right\|_1 . \tag{A.6}$$

By applying the so-called Vieta's formulas [Kos81], we can write

$$k_q = \left\| \sum_{i=0}^{n} z^{-i} \underbrace{(-1)^i \sum_{q1<q2<\ldots<q_i} z_{q1} z_{q1} \ldots z_{qi}}_{a_i} \right\|_1 . \tag{A.7}$$

Considering that the complex values of $z_i \in \mathbb{C}$ always appear in conjugated pairs and that $Re[z_i] \geq 0,\ \forall z_i$, the sum of all the products $z_{q1} z_{q1} \ldots z_{qi} \in \mathbb{R}_+$. The sum of the products is always positive, so the terms $z^{-i}(-1)^i$ are removed by the norm and we can write

$$k_q = \sum_{i=0}^{n} \sum_{q1<q2<\ldots<q_i} z_{q1} z_{q1} \ldots z_{qi} . \tag{A.8}$$

The same result can be obtained by applying Vieta's formulas to the product of $(1 + z_i)$:

$$\prod_{i=0}^{n} (1 + z_i) = \left(\sum_{i=0}^{n} \sum_{q1<q2<\ldots<q_i} z_{q1} z_{q1} \ldots z_{qi} \right) . \tag{A.9}$$

We can therefore conclude that

$$k_q = \prod_{i=0}^{n} (1 + z_i) . \tag{A.10}$$

A.3 The 2-norm of a z transfer-function

The 2-norm is used in Chapter 3 to calculate the standard deviation of a random sequence that goes through a z-transfer-function

$$\Delta Y(z) = Q(z) \cdot (1 - z^{-1})^{n+1} \ . \tag{A.11}$$

Each i-delayed version of the sequence is multiplied by its factor a_i providing the variance

$$\sigma^2_{\Delta y} = \sigma^2_q \cdot \sum_{i=0}^{n+1} a_i^2 \ . \tag{A.12}$$

As a result we can write

$$\sigma_{\Delta y} = \sigma_q \left\| (1 - z^{-1})^{n+1} \right\|_2 \ . \tag{A.13}$$

The norm is determined by the sum of the squares of the binomial coefficients

$$\left\| (1 - z^{-1})^{n+1} \right\|_2 = \sqrt{\sum_{k=0}^{n+1} \binom{k}{n+1}^2} \ . \tag{A.14}$$

To further develop (A.14) let us consider the expression

$$(1+x)^{2n} = \left[1 + \binom{1}{n} x + \cdots + x^n\right] \left[x^n + \binom{1}{n} x^{n-1} + \cdots + 1\right] \ .$$

Expanding both sides and highlighting the term in x^n reveals another representation of the sum in Equation (A.14). It follows that

$$1 + \cdots + \binom{n}{2n} x^n + \cdots + x^{2n} \ = \ 1 + \cdots + \sum_{k=0}^{n} \binom{k}{n}^2 x^n + \cdots + x^{2n} \ .$$

We can therefore write

$$\sum_{k=0}^{n} \binom{k}{n}^2 = \binom{n}{2n} \ . \tag{A.15}$$

The equivalence applied to our specific case yields

$$\left\| (1 - z^{-1})^{n+1} \right\|_2 = \sqrt{\binom{n+1}{2n+2}} = \frac{\sqrt{(2n+2)!}}{(n+1)!} \ . \tag{A.16}$$

A.4 Differential pair model

Figure A.1 is a general description of the differential pair used in a transconductance stage. The degeneration resistances $R/2$ are used for the linearization of the characteristic in the gm-C filter or in the common-mode feedback loops.

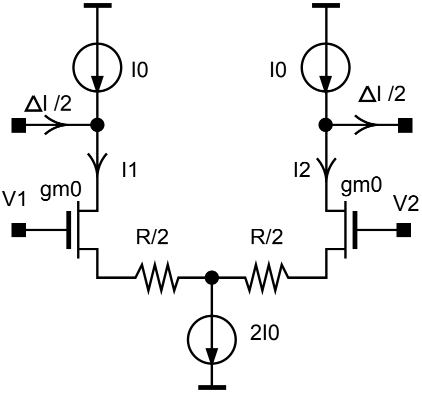


Figure A.1 General description of the differential pair with resistance degeneration.

fAccording to [Vit01], the *inversion-factor*, here referring to the variable x, is defined as the drain current I_d normalized by the *specific current* I_s, which is given by:

$$I_s = 2\mu C_{ox} \frac{W}{L} n U_T^2 \ . \tag{A.17}$$

The slope factor n is considered here as a constant as are the other parameters. The *rest inversion-factor* x_0 of the transistors when the differential pair is in equilibrium is therefore

$$x_0 = I_0 / I_s \ . \tag{A.18}$$

The analytical transistor model presented in [Vit01] considers the transistor terminal voltages V_g, V_s and V_d measured from the substrate. The pinch-off voltage V_p is defined as

$$V_p = (V_g - V_{T0})/n \ , \tag{A.19}$$

where V_{T0} is the threshold voltage. The model provides continuous functions linking the normalized transconductance y and driving voltage z to the *inversion-factor* x given by

$$z = \frac{V_p - V_s}{U_T} = \sqrt{1+4x} + \ln\left(\sqrt{1+4x} - 1\right) - 1.365 \,, \tag{A.20}$$

$$y = \frac{g_m n U_T}{I_d} = \frac{2}{1+\sqrt{1+4x}} \,. \tag{A.21}$$

From the circuit description of the differential pair in Figure A.1, we determine the transistor inversion-factors x_1 and x_2 from

$$\left\{ \begin{array}{rl} \Delta I = & I_1 - I_2 \\ I_0 = & I_1 + I_2 \\ I_{1,2} = & I_0\,(1 \pm \Delta I/2) \\ \Delta V = & V_1 - V_2 \end{array} \right. \Rightarrow x_{1,2} = x_0 \left(1 \pm \frac{\Delta I}{2I_0}\right) \,. \tag{A.22}$$

As a consequence, the normalized driving voltage can be expressed as

$$\Delta z = \frac{\Delta V}{nU_T} - \frac{\Delta I}{I_0}\rho = Q_1 - Q_2 + \ln\left(\frac{Q_1 - 1}{Q_2 - 1}\right) \,, \tag{A.23}$$

where

$$Q_{1,2} = \sqrt{1 + 4x_0\left(1 \pm \frac{\Delta I}{2I_0}\right)} \,. \tag{A.24}$$

The parameter ρ is the normalized degeneration resistance defined as

$$\rho = \frac{RI_0}{2U_T} \,. \tag{A.25}$$

Equation (A.23) is used for the behavioral model in spice simulation described in Appendix B. This implicit equation provides the differential current ΔI given a differential input voltage ΔV. The other variables are fixed parameters. The relationship reproduces the transfer characteristic of the differential pair in all regions from *weak* to *strong* inversion in a continuous way. For better convergence of the numerical solver, the equation is rewritten as:

$$\frac{\Delta I}{2I_0} = \frac{\left[\exp\left(\frac{\Delta V}{nU_T} - \frac{\Delta I}{I_0}\rho - Q_1 + Q_2\right)(Q_2 - 1) + 1\right]^2 - 1}{4x_0} - 1\,. \quad \text{(A.26)}$$

We note that, without degeneration resistors, for extremely weak inversion-factors Equation (A.23) becomes

$$\lim_{x_0 \to 0} \frac{\Delta V}{nU_T} = \ln \frac{1 + \frac{\Delta I}{2I_0}}{1 - \frac{\Delta I}{2I_0}}\,. \quad \text{(A.27)}$$

By rewriting the expression we find the well-known transfer characteristic of the differential pair in weak inversion:

$$\Delta I = 2I_0 \tanh\left[\frac{\Delta V}{2nU_T}\right]\,. \quad \text{(A.28)}$$

By taking the derivative of the implicit equation (A.23) we find an expression with the equilibrium transconductance g_{m0}:

$$\frac{I_0}{g_{m0} n U_T} - \rho = \frac{1 + \sqrt{1 + 4x_0}}{2}\,. \quad \text{(A.29)}$$

We note that the expression without degeneration resistors is identical to the definition of the normalized transconductance (A.21). From this expression, we calculate the inversion-factor for a targeted g_{m0} as well as a degeneration normalized resistance ρ:

$$x_0 = \frac{\left(\frac{2I_0}{g_{m0} n U_T} - 2\rho - 1\right)^2 - 1}{4}\,. \quad \text{(A.30)}$$

As a result, we are able to find $I_{1,2}$ as a function of the input differential voltage ΔV for a desired g_{m0} and I_0. This model is implemented by the Verilog-A Code B.1. The transconductance is given as a normalized parameter $\gamma = g_{m0}/I_0$. For typical values of $n = 1.2$ and $U_T = 26mV$, the ratio γ goes from 0, in strong inversion, to a maximum of 32 in extremely weak inversion. For an inversion-factor of 1 we find $\gamma = 20$.

Equation (A.28) is used for the simplified model implemented by Code B.2. By taking the derivative of the expression we find the quiescent normalized transconductance $\gamma = 1/nU_T$. Substituting γ in Equation (A.28) gives an expression controlled by the quiescent transconductance and current:

$$\Delta I = 2I_0 \tanh\left[\gamma \Delta V/2\right] \ . \tag{A.31}$$

As expected, Equations (A.31) and (A.28) perfectly match for values of γ close to 30 without degeneration resistance.

Appendix

B

Verilog-A codes

B.1 Transconductance behavioral model

Code B.1 Transconductor stage behavioral model based on EKV transistor model.

```
`include "std.va"
`include "const.va"

module gmcell( inp, inn, outp, outn);
        inout      inp, inn, outp, outn ;
        electrical inp, inn, outp, outn ;

        parameter real i0=1u;// diff pair current source
        parameter real gamma=10;// norm. transc. gm0/i0
        parameter real rho=0;// norm. resistor R*io/2UT
        parameter real n=1.2;// MOSFET slope factor
        parameter real UT=26m;// Thermodynamic voltage
        parameter real offset=0; // Input offset
        parameter real ic0min=1e-10;//Minimum inv. factor

        real v,deltai,ic0,Qp,Qn;

        analog begin
        v=(V(inp,inn)-offset)/(n*UT);
        ic0=0.25*(pow(2/(gamma*n*UT)-2*rho-1,2)-1);
        if (ic0<ic0min) begin ic0=ic0min;
          $strobe("inversion-factor required too small!\n");
        end;

        deltai=2*I(outp);
        Qp=sqrt(abs(1+4*ic0*(1+deltai/i0/2)));
        Qn=sqrt(abs(1+4*ic0*(1-deltai/i0/2)));
        I(outp) <+ i0*(0.25*(pow(1+(Qn-1)*exp(v-rho*deltai/i0-
            Qp+Qn),2)-1)/ic0-1);
        I(outn) <+ -I(outp);

        I(inp)  <+0;
        I(inn)  <+0;
        end
endmodule
```

Code B.2 Transconductance stage simplified model.

```
`include "std.va"
`include "const.va"

module gmcell_simple( inp, inn, outp, outn);
        inout      inp, inn, outp, outn ;
        electrical inp, inn, outp, outn ;

        parameter real i0=1u;   // diff pair current source
        parameter real gamma=10;// norm. transc. (DI/DV)/i0
        parameter real offset=0;        // Input  offset
        real v;

        analog begin
                v=V(inp,inn)-offset;

                I(outp) <+ i0*tanh(gamma*v/2);
                I(outn) <+ -I(outp);

                I(inp)  <+0;
                I(inn)  <+0;
        end

endmodule
```

Appendix

C

MATHEMATICA codes

C.1 Synthesis functions

Code C.1 Recursive synthesis function.

```
BuildTree[N_, d_] := Module[{sollist = {}},
  If[N == 2d, {f, {d \[CapitalDelta], d \[CapitalDelta]}},
  If[N > 8 d, sollist = Join[sollist,
    {{q, {BuildTree[N, 4d], BuildTree[8 d, d]}}}]];
  If[N > 2 d, sollist = Join[sollist,
    {{h, {BuildTree[N, 2d], BuildTree[2 d, d]}},
    {f, {BuildTree[N/2, d], BuildTree[N/2, d]}}}]];
 {sollist}]];
```

Code C.2 Removal of symmetric solutions.

```
RemoveSymmetry[sol_] := Module[
 {poslist =
   Sort[Position[sol,
    {_, {p : {{{_, _} ..}}, p : {{{_, _} ..}}}} ]],
     symmetricblock, blocklist, pos},
  If[poslist == {}, sol, pos = First[poslist];
  symmetricblock = Part[sol, Sequence @@ pos];
  blocklist =
   Union[Sort[#] & /@ Tuples[
        Last[Last[Last[symmetricblock]]], 2]];
  If[MatchQ[Part[sol,
        Sequence @@ (Most[Most[pos]])], {_, {_, _}}],
     ReplacePart[sol,
        {{symmetricblock[[1]], #} & /@ blocklist}, pos],
     ReplacePart[sol,
       Unevaluated[
            Sequence @@ ({symmetricblock[[1]], #} & /@ blocklist
                )],
        pos]]]
];
```

C.2 Extraction and display of the solutions

Code C.3 Extraction of the solutions.

```
ExtractTrees[sol_] := Module[
  {poslist = Sort[Position[sol, {{{_, _}, {_, _} ..}}]],
      sollist, pos},
  If[poslist == {}, sol,
    pos = First[poslist];
    sollist = First[Part[sol, Sequence @@ pos]];
  Join[
    First[Take[ReplacePart[sol, #, pos],
     {First[pos]}]] & /@ sollist,
    Drop[sol, {First[pos]}]]]
  ];
```

Code C.4 Displaying function.

```
TreeShow[tree_, n_] := Module[{},
 TableForm[{tree[[n]] /. {
  q -> DisplayForm[FrameBox[StyleForm["Q"]]],
  h -> DisplayForm[FrameBox[StyleForm["H"]]],
  f -> DisplayForm[FrameBox[StyleForm["F"]]],
  a_Integer \[CapitalDelta] ->DisplayForm[FrameBox[ StyleForm[
      a"\[CapitalDelta]"]]],
  \[CapitalDelta] -> DisplayForm[FrameBox[ StyleForm["\[
      CapitalDelta]"]]]}}]
 ];
```

Code C.5 Example of Code usage.

```
tree = {BuildTree[32, 1]};
treeNoSymmetric = FixedPoint[RemoveSymmetry, tree];
sol=FixedPoint[ExtractTrees, treeNoSymmetric];
TreeShow[sol, 347];
Length[sol]

powers = {\[CapitalDelta] -> P\[CapitalDelta], f -> Pf, h ->
    Ph, q -> Pq};
```

```
deltas = {\[CapitalDelta] -> 1, f -> 0, h -> 0, q -> 0};
points ={deltas, #\/1000 /.
                powers} &\) /@ \((\(Total[Flatten[#]] &\) /@ sol
                     )\);\)\)
```

Appendix

D

MATLAB codes

D.1 Modulators with auto-ranging

Code D.1 First-order modulator with auto-ranging algorithm.

```
for i=2:k

    %first integrator
    sig(i,2)= sig(i-1,2) ...
    + 1.0*x0*(sin(w0*time(i-1)/fs)) ...
    + 1.0*xi*(sin(wi*time(i-1)/fs)) ...
    - 1.0*outp(i-1)...
    + dither(i+1);

    %quantization

    if abs(sig(i,2)) >= NR/2
        out(i)=sign(sig(i,2))*(NR-1)/2;
    else
        out(i)=round(sig(i,2));
    end

    %auto-ranging algorithm

    outsign=sign(out(i));
    rngmax=(NL-NR)/2;
    ctrlmax=(rngmax-outsign*rng(i));
    switch abs(out(i))
        case 12,    ctrl(i)=outsign*min(12,ctrlmax);
        case 11,    ctrl(i)=outsign*min(11,ctrlmax);
        case 10,    ctrl(i)=outsign*min(10,ctrlmax);
        case 9,     ctrl(i)=outsign*min(9,ctrlmax);
        case 8,     ctrl(i)=outsign*min(8,ctrlmax);
        case 7,      ctrl(i)=outsign*min(7,ctrlmax);
        case 6,      ctrl(i)=outsign*min(6,ctrlmax);
        case 5,      ctrl(i)=outsign*min(5,ctrlmax);
        case 4,      ctrl(i)=outsign*min(4,ctrlmax);
        case 3,      ctrl(i)=outsign*min(3,ctrlmax);
        case 2,      ctrl(i)=outsign*min(2,ctrlmax);
        case 1,      ctrl(i)=outsign*min(1,ctrlmax);
        otherwise,   ctrl(i)=0;
    end

    %quantizer reconstruction
    outr(i)      =rng(i) +out(i);

    %pulse for next clock cycle
    outp(i)      =outr(i)+ctrl(i);  %acc2

    %range update for next clock cycle
    rng(i+1)     =rng(i) +ctrl(i);  %acc3

end

outr=outr(1:k)-dither(1:k);
```

Code D.2 Second-order modulator with auto-ranging algorithm.

```
for i=2:k

    %first integrator
    sig(i,2)= sig(i-1,2) ...
    + 1.0*x0*(sin(w0*time(i-1)/fs)) ...
    + 1.0*xi*(sin(wi*time(i-1)/fs)) ...
    - 1.0*outr(i-1)...
    + dither(i+1);

    %second integrator
    sig(i,3)= sig(i-1,3) ...
    - 2.0*outp(i-1) ...
    + 1.0*sig(i-1,2);

    %quantization

    if abs(sig(i,3)) >= NR/2
        out(i)=sign(sig(i,3))*(NR-1)/2;
    else
        out(i)=round(sig(i,3));
    end
    %auto-ranging algorithm

    outsign=sign(out(i));

    ctrlmax=floor((rngmax-outsign*rng(i))/2);
    switch abs(out(i))
        case 16,    ctrl(i)=outsign*min(8,ctrlmax);
        case 15,    ctrl(i)=outsign*min(8,ctrlmax);
        case 14,    ctrl(i)=outsign*min(7,ctrlmax);
        case 13,    ctrl(i)=outsign*min(7,ctrlmax);
        case 12,    ctrl(i)=outsign*min(6,ctrlmax);
        case 11,    ctrl(i)=outsign*min(6,ctrlmax);
        case 10,    ctrl(i)=outsign*min(5,ctrlmax);
        case 9,     ctrl(i)=outsign*min(5,ctrlmax);
        case 8,     ctrl(i)=outsign*min(4,ctrlmax);
        case 7,     ctrl(i)=outsign*min(4,ctrlmax);
        case 6,     ctrl(i)=outsign*min(3,ctrlmax);
        case 5,     ctrl(i)=outsign*min(3,ctrlmax);
        case 4,     ctrl(i)=outsign*min(2,ctrlmax);
        case 3,     ctrl(i)=outsign*min(2,ctrlmax);
        case 2,     ctrl(i)=outsign*min(1,ctrlmax);
        otherwise,  ctrl(i)=0;
    end

    %quantizer reconstruction
    outr(i)     =rng(i) +out(i);

    %pulse for next clock cycle
    outp(i)     =outr(i)+ctrl(i);  %acc2

    %range update for next clock cycle
    rng(i+1)    =rng(i) +ctrl(i)*2;  %acc3

end

outr=outr(1:k)-dither(1:k);
```

Code D.3 Third-order modulator with auto-ranging algorithm.

```
for i=2:k
    %first integrator
    sig(i,2)= sig(i-1,2) ...
    + 1.0*x0*(sin(w0*time(i-1)/fs)) ...
    + 1.0*xi*(sin(wi*time(i-1)/fs)) ...
    - 1.0*outr(i-1);
    %second integrator
    sig(i,3)= sig(i-1,3) ...
    - 3.0*outr(i-1) ...
    + 1.0*sig(i-1,2);
    %third integrator
    sig(i,4)= sig(i-1,4) ...
    - 3.0*outp(i-1) ...
    + 1.0*sig(i-1,3);
    %quantization
    if abs(sig(i,4)) >= (NR)/2
        out(i)=sign(sig(i,4))*(NR-1)/2;
    else
        out(i)=round(sig(i,4));
    end

    %auto-ranging algorithm

    outsign=sign(out(i));

    ctrlmax=floor((rngmax-outsign*rng(i))/3);
    switch abs(out(i))
        case 16,    ctrl(i)=outsign*min(5,ctrlmax);
        case 15,    ctrl(i)=outsign*min(5,ctrlmax);
        case 14,    ctrl(i)=outsign*min(4,ctrlmax);
        case 13,    ctrl(i)=outsign*min(4,ctrlmax);
        case 12,    ctrl(i)=outsign*min(4,ctrlmax);
        case 11,    ctrl(i)=outsign*min(3,ctrlmax);
        case 10,    ctrl(i)=outsign*min(3,ctrlmax);
        case 9,     ctrl(i)=outsign*min(3,ctrlmax);
        case 8,     ctrl(i)=outsign*min(2,ctrlmax);
        case 7,     ctrl(i)=outsign*min(2,ctrlmax);
        case 6,     ctrl(i)=outsign*min(2,ctrlmax);
        case 5,     ctrl(i)=outsign*min(1,ctrlmax);
        case 4,     ctrl(i)=outsign*min(1,ctrlmax);
        case 3,     ctrl(i)=outsign*min(1,ctrlmax);
        otherwise,  ctrl(i)=0;
    end

    %quantizer reconstruction
    outr(i)     =rng(i) +out(i);

    %pulse for next clock cycle
    outp(i)     =outr(i)+ctrl(i);  %acc2

    %range update for next clock cycle
    rng(i+1)    =rng(i) +ctrl(i)*3;  %acc3

end
outr=outr(1:k);
```

D.2 Mismatch shaping encoder

Code D.4 Second-order mismatch shaper.

```
if (mod(in,2)~=0)
  if  acc(ii,jj)==0)&(acc2(ii,jj)==0)  q=ran;
  elseif  (acc(ii,jj)==0) q=-sign(acc2(ii,jj));
  else  q=-sign(acc(ii,jj)); end
else
 q=0;
end
if (abs(acc2(ii,jj)+acc(ii,jj))<=acc2max)
  acc2(ii,jj)=acc2(ii,jj)+acc(ii,jj); end
if (mod(in,2)~=0)
  acc(ii,jj)=acc(ii,jj)+q; end
```

Code D.5 Third-order mismatch shaper.

```
if (mod(in,2)~=0)
if  (acc(ii,jj)==0)&(acc2(ii,jj)==0)&(acc3(ii,jj)==0)
 q=ran;
elseif  (acc(ii,jj)==0)&(acc2(ii,jj)==0)
 q=-sign(acc3(ii,jj));
elseif  (acc(ii,jj)==0)
 q=-sign(acc2(ii,jj));
else
 q=-sign(acc(ii,jj));
end
else
q=0;
end
%%%%% accumulators %%%%%%%%%%%
acc3(ii,jj)=acc3(ii,jj)+acc2(ii,jj);
acc2(ii,jj)=acc2(ii,jj)+(2^acr)*acc(ii,jj)-floor(acc3(ii,jj)
    /128);
acc(ii,jj)=acc(ii,jj)+q;
```

Code D.6 Standard segmented tree structure with first-order shaping. The parameters **N** and **depth** are the encoder and segmentation depths respectively.

```
treesize=2^(N-depth-1);

for ii = 2:N+1,
    if     (ii<=depth+1)  jpairrange=1;
    elseif (ii< N+1)  jpairrange=2^(ii-depth-2);
    else                  jpairrange=treesize+depth;
    end
    r=sign(randn(1));  %+1 or -1;

    for jj = 1:jpairrange,
        %%%%%%% wireing %%%%%%%%%
        if (jj<=treesize)
            in=y(ii-1,jj);      %%unsegmented
        else
            in=y( (depth+2)-(jj-treesize),2);%%segmented
        end
        %%%%%%% first-ordre shaper %%%%%%%
           if (mod(in,2)~=0)
                if (acc(ii,jj)==0)
                    q=ran;
                else
                    q=-sign(acc(ii,jj));
                end

            else
                q=0;
            end
            if (mod(in,2)~=0) acc(ii,jj)=acc(ii,jj)+q; end

        %%%%%%%  switching blocks %%%%%%%%%
        if (ii>depth+1), %F blocks
            y(ii,2*jj-1)=(in+q)/2;
            y(ii,2*jj)  =(in-q)/2;
        else  %H blocks
            y(ii,2*jj-1)=(in+q)/2;
            y(ii,2*jj)  =-q+1;  %%+1signed to unsigned
        end

   end
end
```

Bibliography

[Abo02] H. Aboushady. *Conception en Vue de la Réutilisation de Converstisseurs Analogique-Numérique Sigma-Delta Temps-Continu Mode-Courant.* PhD thesis, Université de Paris VI, 2002. 3.1.1

[AH02] P. E. Allen and D. R. Holberg. *CMOS Analog Circuit Design.* Oxford University Press, New-York, second edition, 2002. 2.4.1, 6.4.1

[AL02] H. Aboushady and M.-M. Louerat. Systematic approach for discrete-time to continuous-time transformation of Sigma-Delta modulators. *Journal of Analog Integrated Circuits and Signal Processing*, pages 229 – 232, 2002. 1.1, 3.1.1

[Bal03] P. Balmelli. *Broadband Sigma-Delta A/D Converters.* PhD thesis, Swiss Federal Intitute of Technology Zurich, Zurich, 2003. 2.4.1

[BH01] L. Breems and J. H. Huijsing. *Continuous-Time Sigma-Delta Modulation for A/D Conversion in Readio Receivers.* Kluwer Academic Publishers, Boston, 2001. 1.1, 2.4.3, 3.1.1

[Bra91] B. P. Brandt. *Oversampled Anlog-to-Digital Conversion.* PhD thesis, Stanford University, 1991. 2.1.5

[Bur02] T. Burger. *Optimal Design of Operational Transconductance Amplifiers with Application for Low Power $\Delta\Sigma$ Modulators.* Hartung-Gorre Verlag Konstanz, Germany, 2002. 6.5, 7.3

[CS99] J. A. Cherry and W. M. Snelgrove. Clock jitter and quantizer metastability in continuous-time Delta-Sigma modulators. *IEEE Transactions on Circuits and Systems II*, 46(6):661–676, June 1999. 3.1.1

[CS00] J. A. Cherry and W. M. Snelgrove. *Continuous-Time Delta-Sigma Modulators for High-Speed A/D Convertion, Theory,* *Practice and Fundamental Performance Limits.* Kluwer Academic Publishers, Boston, 2000. 1.1, 3.1.1

[DKG$^+$05] L. Dorrer, F. Kuttner, P. Greco, P. Torta, and T. Hartig. A 3mW 74dB-SNR 2MHz continuous-time Delta-Sigma ADC with a tracking ADC quantizer in 0.13-um CMOS. *IEEE Journal of Solid-State Circuits*, 40:2416 – 2426, December 2005. 3.1.4, 3.3.1, 4.2, 4.1.2, 4.2.1, 6.5, D.2

[EKV95] C. Enz, F. Krummenacher, and E. A. Vittoz. An analytical MOS transistor model valid in all regions of operation and dedicated to low-voltage and low-current applications. *Journal of Analog Integrated Circuits and Signal Processing*, 8:408 – 415, July 1995. 6.3.1

[FGHJ00] E. Fogleman, I. Galton, W. Huff, and H. Jensen. A 3.3V single-poly CMOS audio ADC Delta-Sigma modulator with 98dB peak SINAD and 105dB peak SFDR. *IEEE Journal of Solid-State Circuits*, 35(3):297–307, March 2000. 5.2, 5.2

[FSW$^+$02] A. Fishov, E. Siracusa, J. Welz, E. Fogleman, and I. Galton. Segmented mismatch-shaping D/A conversion. In *Proc. of the IEEE International Symposium on Circuits and Systems*, volume 4, pages 679–682, May 2002. 1.3, 2.1.4, 3.1.3, 5.1, 5.3.1, 5.3.1

[Gal93] I. Galton. Granular quantization noise in the first-order Delta-Sigma modulator. *IEEE Transactions on information theory*, 39(6):1944–1956, November 1993. 2.1.2

[Gal94] I. Galton. Granular quantization noise in a class of Delta-Sigma modulators. *IEEE Transactions on information theory*, 40(3):848–859, May 1994. 2.1.2

[Gal97] I. Galton. Spectral shaping of circuit errors in digital-to-analog converters. *IEEE Transactions on Circuits and Systems II*, 44(10):808–817, October 1997. 1.3, 5.2, 5.2, 5.2.1

[GG05] S. W. Golomb and G. Gong. *Signal Design for Good Correltation.* Cambridge University Press, New-York, 2005. 5.2

[GM01] F. Gerfers and Y. Manoli. A design strategy for low-voltage low-power continuous-time Sigma-Delta A/D converters. In *Proc. of the Conference on Design, automation and test in Europe*, pages 361–369. IEEE, 2001. 3.1.1

[Gra90] R. M. Gray. Quantization noise spectra. *IEEE Transactions on information theory*, 36(6):1220–1244, November 1990. 2.1.2

[GS02] Y. Geerts and M. Steyaert. *Design of Multi-bit Delta-Sigma A/D Converters.* Kluwer Academic Publishers, Boston, 2002. 3.1.3, 5.1

[GWT02] M. Gustavsson, J. J. Wikner, and N. N. Tan. *CMOS Data Converters for Communications.* Kluwer Academic Publishers, Boston, 2002. 1.1

[Hay01] S. Haykin. *Communication Systems.* John Wiley & Sons Inc., NewYork, fourth edition, 2001. 2.4.1

[HH96] R. Hogervorst and J. H. Huijsing. *Design of Low-Voltage, Low-Power Operational Amplifier Cells.* Kluwer Academic Publishers, Boston, 1996. 6.3.2

[HvV99] S. Haykin and B. van Veen. *Signals and Systems.* John Wiley & Sons Inc, New-York, 1999. 2.1.1, 2.1.4, 2.2.6, 2.2.7

[II02] L. W. Couch II. *Digital and Analog Communication Systems.* Prentice Hall, New Jersey, 2002. 2.1.1

[Jes01] P. G. A. Jespers. *Integrated Converters, D to A and A to D architectures, analysis and simulation.* Oxford University Press, New-York, 2001. 1.1

[JM97] D. A. Johns and K. Martin. *Analog Integrated Circuit Design.* John Wiley & Sons Inc., New-York, 1997. 2.1.1, 2.4.1, 4.3.3, 6.4.1

[Joh92] N. Johl. *Conception de filtres a transconductances et capacités en technologie CMOS pour applications haute fréquence.* PhD thesis, Ecole Polytechnique Fédérale de Lausanne, Lausanne, 1992. 6.3.2

[Kap03] M. Kappes. A 2.2mW CMOS band-pass continuous-time multi-bit Delta-Sigma ADC with 68dB of dynamic range and 1MHz bandwidth for wireless applications. *IEEE Journal of Solid-State Circuits*, 38:1098–1104, July 2003. 1.1, 3.1.1, 6.5

[Kas95] N. J. Kasdin. Discrete simualtion of colored noise and stochastic processes and $1/f^{\alpha}$ power law noise generation. In *Proc. of the IEEE*, volume 83, pages 802–827, May 1995. 2.3.2

[Kos81] A. Kostrikin. *Introduction à l'algèbre*. Éditions MIR, Moscow, 1981. A.2

[KRSC05] K.Nguyen, R.Adams, K. Sweetland, and H. Chen. A 106dB-SNR hybrid oversampling analog-to-digital converter for digital audio. *IEEE Journal of Solid-State Circuits*, 40(12):2408–2415, December 2005. 1.3, 6.5

[KvR06] K. and A.H.M. van Roermund. $\Delta\Sigma$ *Conversion for Signal Conditioning*. Springer, The Netherlands, 2006. 1.1, 3.1.1

[Le 05] Y. Le Guillou. Analyzing sigma-delta ADCs in deep-submicron CMOS technologies, February 2005. http://www.rfdesign.com. 7.1

[Lon95] R. Longchamp. *Commande numérique de systèmes dynamiques*. Presses polytechniques et universitaires romandes, Lausanne, 1995. 2.2.6, 3.2.1

[Lu04] X. Lu. A novel signal-predicting multibit Delta-Sigma modulator. In *Proc. of the IEEE International Conference on Electronics, Circuits and Systems*, pages 105 – 108, December 2004. 3.1.4, 4.2, 4.1.2, D.2

[MCL$^+$05] P. Morrow, M. Chamarro, C. Lyden, P. Ventura, A. Abo, A. Matamura, M. Keane, R. O'Brien, P. Minogue, J. Mansson, N. McGuinness, M. McGranaghan, and I. Ryan. A 0.18um 102dB-SNR mixed CT SC audio-band Delta-Sigma ADC. In *Proc. of the IEEE International Solid-State Circuit Conference*, pages 178–179, 2005. 1.3, 6.5

[MZ60] S. J. Mason and H. J. Zimmermann. *Electronic circuits, signals, and systems*. John Wiley, New-York, 1960. 2.2.2

[OG06] M. Ortmanns and F. Gerfers. *Continuous-Time Sigma-Delta A/D Converters, Fundamentals, Performance Limits and Robust Implementations.* Springer, Berlin, 2006. 1.1, 3.1.1, 3.3.1

[Oga95] K. Ogata. *Discrete-Time Control Systems.* Prentice Hall, New Jersey, second edition, 1995. 2.2.6, 3.2.1

[Oli99] O. Oliaei. Clock jitter noise spectra in continuous-time delta-sigma modulators. In *Proc. of the International Symposium on Circuits and Systems ISCAS*, pages 192–195. IEEE, 1999. 3.1.1

[Oli01] O. Oliaei. Clock jitter effect in continuous-time oversampling converters. In *Proc. of the International Symposium on Circuits and Systems ISCAS*, pages 288–291. IEEE, 2001. 3.1.1

[OSB99] A. V. Oppenheim, R. W. Schafer, and J. R. Buck. *Discrete-Time Signal Processing.* Prentice Hall, New-York, second edition, 1999. 2.1.4

[Par93] S. Park. *Motorola Digital Signal Processors, Principles of Sigma-Delta Modulation for Analog-to-Digital Converters.* Motorola APR8/D Rev. 1, 1993. 2.4.1

[Pas05] M. Pastre. *Methodology for the Digital Calibration of Analog Circuits and Systems.* PhD thesis, Ecole Polytechnique Fédérale de Lausanne, Lausanne, 2005. 6.4.2

[PCSK07] S. Pesenti, P. Clement, D. Stefanovic, and M. Kayal. A low-power strategy for Delta-Sigma modulators. In *Proc. of the International Conference on Mixed Design of Integrated Circuits and System, MIXDES*, pages 203–208, June 2007. 3.1.7

[Phi03] K. Philips. A 4.4mW 76dB complex $\Delta\Sigma$ ADC for bluetooth receivers. In *Proc. of the IEEE International Solid-State Circuits Conference*, February 2003. 6.5

[PK06] M. Pastre and M. Kayal. *Methodology for the digital calibration of analog circuits and systems with case studies.* Springer, Berlin, 2006. 4.3.3, 6.4.1, 6.4.2, 6.4.2

[PNR$^+$04] K. Philips, P.A.C.M. Nuijten, R. Roovers, F. Munoz, M. Tejero, and A.Torralba. A 2mW 89dB DR continuous-time $\Delta\Sigma$ ADC with increased immunity to wide-band interferers. In *Proc. of the IEEE International Solid-State Circuits Conference*, February 2004. 3.1.1

[PVS96] P.M.Aziz, Henrik V.Sorensen, and J.Van Der Spiegel. An overview of Sigma-Delta converters: How a 1-bit ADC achieves more than 16-bit resolution. *IEEE Signal Processing Magazine*, pages 61–84, January 1996. 2.4.1

[Raz96] B. Razavi. Challenges in portable RF transceiver design. *IEEE Circuits and Devices Magazine*, pages 12–25, Septembre 1996. 1.1

[Raz01] B. Razavi. *Design of Analog CMOS Integrated Circtuits.* Mc Graw Hill International Edition, New-York, 2001. 2.1.1, 4.3.3

[RLS$^+$03] R.Schreier, J. Lloyd, L. Singer, D. Paterson, M.Timko, M. Hensley, G. Patterson, K. Behel, and J.Zhou. A 10-300-MHz IF-digitizing IC with 90-105dB dynamic range and 15-33-kHz bandwidth. *IEEE Journal of Solid-State Circuits*, 38:1098–1104, December 2003. 1.3, 6.5

[SG04] E. Siracusa and I. Galton. A digitally enhanced 1.8-V 15-bit 40-MSamples/s CMOS piplined ADC. *IEEE Journal of Solid-State Circuits*, 39(12):2126–2138, December 2004. 6.3.3

[Sho95] O. Shoaei. *Continuous-Time Delta-Sigma A/D Converters for High Speed Applications.* PhD thesis, Carlton University, 1995. 1.1, 3.1.1

[SKP05] D. Stefanovic, M. Kayal, and M. Pastre. Pad: A new interactive knowledge-based analog design approach. *Analog Integrated Circuits and Signal Processing Journal*, 42:291–299, March 2005. 6.3.1

[SKPK04a] D. Stefanovic, F. Krummenacher, M. Paste, and M. Kayal. BSIM2EKV: BSIM3.3 to EKV 2.6 model library file automatic conversion. In *Proc. of the EKV Users' Meeting/Workshop*, November 2004. 6.3.1

[SKPK04b] D. Stefanovic, F. Krummenacher, M. Pastre, and M. Kayal. BSIM2EKV: un outil pour la conversion automatique des paramètres du modèle BSIM aux paramètres du modèle EKV. In *Proc. of TAISA04, 5ème Colloque sur le Traitement Analogique de l'Information, du Signal et ses Applications*, pages 85–88, Septembre 2004. 6.3.1

[SPPK07] D. Stefanovic, S. Pesenti, M. Pastre, and M. Kayal. Structured design based on the inversion parameter: Case study of DS modulator system. In *Proc. of the International Conference on Mixed Design of Integrated Circuits and System, MIXDES*, pages 95–102, June 2007. 6.3.1

[SRNT97] R. Schreier S. R. Norsworthy and G. C. Temes. *Delta-Sigma Data Converters, Theory, Design, and Simulation.* IEEE Press, New-York, 1997. 4.2.4

[SSS05] K. Shu and E. Sanchez-Sinencio. *CMOS PLL Synthesizers: Analysis and Design.* Springer, Boston, 2005. 3.3.1

[ST05] R. Schreier and G. C. Temes. *Understanding Delta-Sigma Data Converters.* IEEE Press, New-York, 2005. 2.2.7, 2.1, D.2

[Ste07] D. Stefanovic. *Structured Analog CMOS Design Based on the Device Inversion Level.* PhD thesis, Ecole Polytechnique Fédérale de Lausanne, Lausanne, 2007. N°3857. 6.3.1, 6.3.2

[Uni97] Encyclopaedia Universalis. *Dictionnaire des mathématiques: algèbre, analyse, géométrie.* Encyclopaedia Universalis France, Paris, 1997. A.1

[vEvdP99] J. van Engelen and R. van de Plassche. *Bandpass Sigma-Delta Modulator, Stability Analysis, Performance and Design Aspects.* Kluwer Academic Publishers, Boston, 1999. 2.4.1

[Vit01] E.A. Vittoz. MOS transistor. In *Proc. of Summer Courses: Advanced Analog IC Design,*, Lausanne, 2001. Mead. 6.3.1, 6.3.2, A.4, A.4

[vV03] R. H. M. van Veldhoven. A triple-mode continuous-time $\Delta\Sigma$ modulator with switched-capacitor feedback DAC for GSM-EDGE/CDMA2000/UMTS receiver. *IEEE Journal of Solid-State Circuits*, 38:2069–2076, December 2003. 3.3.1, 6.5

[WG02] J. Welz and I. Galton. Necessary and sufficient conditions for mismatch shaping in a general class of multibit DACs. *IEEE Transactions on Circuits and Systems II*, 49(12):748–759, December 2002. 5.1, 5.1

[WGF01] J. Welz, I. Galton, and E. Fogleman. Simplified logic for first-order and second-order mismatch-shaping digital-to-analog converters. *IEEE Transactions on Circuits and Systems II*, 48(11):1014–1027, November 2001. 5.2

[Yan02] S. Yan. *Baseband Continuous-Time Sigma-Delta Analog-to-Digital Conversion for ADSL Applications.* PhD thesis, Texas A&M University, 2002. 1.1, 3.1.1, 6.3.2

[YS04] S. Yan and S. Sanchez. A continuous-time Sigma-Delta modulator with 88dB dynamic range and 1.1MHz signal bandwidth. *IEEE Journal of Solid-State Circuits*, 39:75–86, Januray 2004. 1.1, 3.1.1

[Zie00] C. M. Zierhofer. Adaptive Sigma-Delta modulation with one-bit quantization. *IEEE Transactions on circuits and systems II*, pages 408 – 415, May 2000. 3.1.4, 4.2, 4.1.3, D.2

List of Figures

1.1 Illustrative example of the hybrid architecture. 3

2.1 Sampling: An RC-filter is placed in front of the sampler providing anti-aliasing filtering. 6

2.2 Sampling process: Simulations performed with 50 average points and 2^{18} FFT bins. A 1mV tone at 500kHz together with a 10kΩ resistor thermal noise are sampled at a frequency of 2GHz (a) and 2MHz (b). The filter cut-off is set at 100MHz. 6

2.3 A quantizer with NL levels consists of a bank of $(\mathrm{NL}-1)$ comparators, providing a quantization in amplitude, and$(\mathrm{NL}-1)$ Data-Flip-Flops (DFF), providing the quantization in time. The $(\mathrm{NL}-1)$ voltage thresholds, separated by steps of Δ, are usually generated by a resistive ladder with NL resistors. The $(\mathrm{NL}-1)$ outputs of the data-flip-flops form an integer number, represented by a thermometer code. 9

2.4 Mid-thread 5-level quantizer transfer characteristic described by Equation (2.5). The vertical parts correspond to the voltage thresholds and the horizontal parts to the possible output levels. 9

2.5 Quantization process: Simulation performed with 2000 averages, NL=33, extra=-1.5, single-tone at 10kHz and sampling frequencies f_s=1MHz. 11

2.6 Dithered quantization: Simulation performed with 2000 averages, NL=33, extra=-1.5, Gaussian dithering of 0.5, single-tone at 10kHz and sampling frequencies f_s=1MHz. 11

2.7 A mid-thread 5-level quantizer: The real transfer characteristic is shown as a dashed line. The linear model consists in separating the characteristic in a gain of $1/\Delta$ and a quantization error q bounded by $\pm 1/2$. Beyond the range of $\pm \mathrm{NL}\Delta/2$ the quantizer is overloaded and the linear model no longer holds. . 12

2.8 Overloaded quantizer: 2000 averages, NL=33, extra=0, Gaussian dithering of 0.5, in-band single-tone at 10kHz, sampling frequencies fs=1MHz. 13

2.9 Over-sampled quantizer: Simulation results of Code 2.3 with avr=20, NL=33, extra=-1.5, d_0=0.5. As the sampling frequencies goes from (a)1MHz to (b)32 MHz, the quantization noise, integrated over the band of interest, here from 0 to 500kHz, drops by the same factor. Meanwhile, the area under the in-band tone at 100kHz remains unchanged. 14

2.10 $\Delta\Sigma$-modulation: Conventional nth-order architecture. 15

2.11 Spectral shaping: Simulation results of (a) Code 2.3 and (b) Code 2.5 with avr=20,NL=33,extra=-1.5, d_0=0.5,f_s=32MHz. The shaping provides a 3-decade attenuation of the quantization level at low frequencies. Meanwhile, the area under the in-band tone at 100kHz remains unchanged. 16

2.12 Second-order spectral shaping: 20 averages, NL=33, extra=-1.5, Gaussian dithering of 0.5, in-band single-tone at 100kHz, sampling frequencies f_s=1MHz and 32MHz. 17

2.13 Third-order spectral shaping: 20 averages, NL=33, no extra, no dithering, in-band single-tone at 100kHz, sampling frequencies 32 MHz. 18

2.14 First-stage decimation filter made of n cascaded accumulators and n differentiators. 20

2.15 First-stage decimation: (a) 33-level second-order modulator output sampled at 32MHz and comb filter output before (b) and after down-sampling (c) at 2MHz. A band of 500KHz is considered. 20

2.16 Second-stage decimation: Half-band FIR filter output before (a) and after down-sampling (b) at 1MHz. (c) PSD of a Nyquist-rate 32k-level quantizer. 22

2.17 Quantization: Linear signal flow graph representation where, q is white noise uniformly distributed between $\pm 1/2$. 23

2.18 General linear Signal Flow Graph representation of the single-stage modulator: The quantizer is represented by a gain of Δ^{-1} and an additional random source q. In contrast, the DACs do not introduce errors and are simply modeled as a gain of Δ. . . 25

2.19 Poles and zeros: representation in the complex z-plane. Stability is guaranteed if all the poles are within the unit circle. Here all the n poles are placed at the center providing the maximum stability. The zeros appear either on the real axis or in complex conjugated pairs. 26

2.20 Digital output of a second (a) and a fourth-order (b) modulator with a 33-level quantizer. A clear distinction can be made between the fast quantization noise at 32MHz and the slow 100kHz-tone input. 28

2.21 Equation (2.37): (a) achievable SQNR for a given NL=33 and (b) required NL for a given SQNR=94dB. 30

2.22 Dynamic range plot around the under- and over-load regions for a second- and third-order modulator. To optain such a smooth plot, 1000 averaging points are required and a small 0.05 dithering signal for the second-order modulator. 32

2.23 Dynamic range plot around under-over-load region for a second- and third-order modulator. To obtain such a smooth plot 1000 averaging points are required and a small 0.05 dithering signal for the second-order modulator. 33

2.24 Quantization noise in the overload region for a second-order modulator with extra=0.2. 33

2.25 Maximum achievable SQNR of a 33-level second-order modulator with an OSR of 32 as a function of the global feedback coefficients a_1 and a_2. The resulting position of the poles and zeros for few relevant case are depicted on the sides. The experiment is performed with at input signal at 100kHz for a sampling frequency of 32MHz. 34

2.26 Parameter extra as a function of the global feedback coefficients a_1 and a_2 for the case of Figure 2.25. 35

2.27 $\Delta\Sigma$-modulation: Additional local feedbacks b_i controlling the zeros of the NTF. 36

2.28 33-level third-order modulator sampling at 32 MHz with a band-of-interest from 0 to 4MHz. All the poles are placed at the center of the z-plane. In (a) all the zeros are at DC. In (b) one zero is placed at DC and a complex conjugated pair at 3MHz. The coefficient and local feedback vectors are respectively [1 2.64 2.64] and [0 0.36]. 37

2.29 Second-order 33-level modulator with circuit noise injected at the input (a), after the first integrator (b) and before the quantizer (c). The quantization noise is removed. 40

2.30 Output power spectral density of a second-order 33-level modulator, whose sampling is affected by clock jitter, with a full-scale input tone at 100kHz (a) and 1MHz (b). The quantization noise is removed. 41

2.31 Output power spectral density of a second-order 33-level modulator with DAC elements mismatch (a) and ADC comparator offset (b). 42

2.32 Simulation of distorsion in a second-order 33-level modulator with a full-scale low-frequency single-tone (a) and two half-scale high-frequency tones (b). A 500kHz bandwidth is considered. . 43

2.33 First-order single-bit modulator output. 44

2.34 Second-order shaping power spectral density with 2-level (a) and 5-level (b) quantizer. 45

2.35 DR-Plot of a single-bit second-order modulator. 45

2.36 Output power spectral density for a fourth-order, low-pass (a), band-pass (b) and high-pass (c) modulator. The frequency axis is linear to clearly present all three cases. The simulation coefficients are provided on the left together with the z-plane representation of the resulting poles and zeros. 47

2.37 General (a) multi-feedback (b) single-feedback and configurations. 48

2.38 m-stage architecture: Each stage consists of a low-pass n_ith-order $\Delta\Sigma$-modulator with all poles at the center and zeros at DC. Only the quantization noise q_i is transferred from one stage to the other. 49

3.1 Continuous-time implementation of a multi-feedback single-loop. 52

3.2 Tree structured DEM for a DAC with 17 levels. The conventional (a) solution consists of 16 1-bit DAC elements and 15 switching blocks. In the fully segmented case (b), there are only 7 switching blocks but a total of 31 elements. 54

3.3 Digital algorithm to emulate a large NL with a reduced quantizer number-of-levels. 55

3.4 Multi-bit hybrid continuous-discrete-time $\Delta\Sigma$-modulator. . . . 56

3.5 Continuous-to-discrete-time transition with an isolation buffer and a full-clock-cycle sampling scheme. 57

3.6 Low-power design strategy. 59

3.7 Continuous-time $\Delta\Sigma$-modulator block-diagram. 60

3.8 Block diagram in Figure 3.7 redrawn to distinguish a continuous-time forward path and the discrete-time internal loop. 61

3.9 Block diagram in Figure 3.8 redrawn to highlight the STF and NTF. 62

3.10 Continuous-time transfer function for the modulator input to the sampling node for a first-, second- and third-order modulator, according to Equation (3.4). The second-order case is shown in as solid line and the gray areas represent the parts of the spectrum that fold back to the band-of-interest when the OSR is equal to 5. 63

3.11 Block diagram of an nth-order hybrid continuous-discrete-time modulator with c continuous-time first integrators. 63

3.12 $\Delta\Sigma$-modulator input stage with discrete- (a) and continuous-time (b) implementations. 64

3.13 Discrete- (a) and continuous-time (b) clock-jitter transfer mechanism to an equivalent input error sequence. 65

3.14 Simulated *jitter transfer factor* ξ for (a) a given modulator order n and for (b) a given number-of-levels NL. The continuous- and discrete-time factors are plotted as solid and dashed lines respectively. 67

3.15 Jitter sensitivity plot for a 33-level second-order case with an OSR of 32. The input signal is (a) a full-scale in-band tone and (b) a -20dB in-band tone with an additional -0.9dB out-of-band signal. 72

3.16 Design chart for a second-order modulator with an SQNR of 94dB using Equations (3.18) and (3.13). The relative jitter sensitivity is plotted considering different maximum frequencies $f_{\max}$. The discrete- and hybrid cases are represented by dashed and solid lines respectively. 74

3.17 Design chart for modulator with an SQNR of 94dB and different orders n using Equations (3.18) and (3.13). The relative jitter sensitivity is plotted considering a maximum frequency of twice the bandwidth. Only the hybrid case with a continuous-time first integrator is represented. 75

3.18 $\Delta\Sigma$-modulation: Conventional nth-order architecture. 76

3.19 NTF zeros and poles location for a sixth-order modulator. The six zeros are located at DC and the six poles at the center of the complex z-plane. As the sixth (a) and first (b) coefficients deviate from their nominal value, the poles moves toward the unity circle. The poles are represented when the variation brings the modulator at its limit of stability. 77

3.20 Feedback coefficient sensitivity for a second- (a), third- (b) and fourth-order (c) modulator. Results with and without the quantizer saturation are displayed as solid and dashed lines respectively. 78

3.21 Equivalent block diagram of the full-clock-cycle sampling. . . . 79

3.22 Simulated sampling capacitor mismatch sensitivity. Second- (a) and fifth-order (b) modulators. The -3dB drops are highlighted with a dot. 82

3.23 Modulator output with a capacitor mismatch of 10%. 82

3.24 Sampling capacitor mismatch sensitivity for 94dB of SQNR. . . 83

4.1 Typical output signal of a low-pass multi-bit $\Delta\Sigma$-modulator. This example is a second-order 33-level modulator. The input is a sine wave 640 times slower than the clock signal. 86

4.2 Different methods: (a) conventional ADC with only the active group turned on, (b) switching references proposed by [Lu04] and [DKG$^+$05] and (c) signal shifting proposed here as an extension of [Zie00]. 87

4.3 Signal shifting process. The bold and light lines represent digital and analog signal processing respectively. 89

4.4 Feedback path reuse. The bold and light lines represent digital and analog signal processing respectively. The dashed area shows the existing processing being reused. 89

4.5 Generation of the pulses based on the quantizer output. The bold and light lines represent digital and analog signal processing respectively. The dashed area shows the existing processing being reused. 90

4.6 General linear Signal Flow Graph representation of the single-stage modulator of Figure 2.18 with the auto-ranging algorithm. 91

4.7 Control functions $\kappa(y')$ for a first- (a), second- (b) and third-order (c) modulator. The slope shown corresponds to the inverse of the last modulator feedback coefficient. 93

4.8 Quantizer input signal boundaries. 94

4.9 Maximum number of steps seen by a 33-level ADC: (a) all the NTF zeros are located at DC, (b) one zero located at DC and two conjugated zeros at 0.77 times the bandwidth. 95

4.10 Simulation results for the first-order 33-level modulator with a 15-level quantizer. 97

4.11 Simulation results for the second-order 33-level modulator with a 15-level quantizer. 98

4.12 Simulation results for the third-order 33-level modulator with a 15-level quantizer. 99

4.13 Frequency limitation of a 33-level (a) first-, (b) second- and (c) third-order modulator for different reduced number-of-levels NR. The sampling frequency is set at 32MHz and a band-of-interest of 500kHz is considered. 101

4.14 33-level second-order modulator with a 11-level quantizer. Simulated (a) dynamic range plot and (b) output PSD with two half-scale input tones at 700kHz and 900kHz. The tones are place within the band, from 500kHz and 1MHz, specified by the plot of Figure 4.13. The dynamic range plot is performed with a 50kHz single tone. 103

4.15 Last feedback path sensitivity to imperfections in a 33-level second-order modulator with different reduced number-of-levels NR. The tolerance of the overall coefficient (a) and finite DC gain (b) of the integrator are shown. The simulations are performed with an input signal at 100kHz sampled at 32MHz and considering an OSR of 32. 104

4.16 Tolerance on the statistic offset of the comparators for the second-order case with (a) and without (b) auto-ranging. 105

4.17 Hardware implementation of the algorithm. 106

4.18 Automatic reset mechanism for recovery. 107

4.19 33-level modulator output showing the automatic recovery process with a full-scale input signal at 100kHz and 500kHz. 108

4.20 Illustration of the worst case situation with a large rising input signal at the maximum frequency. The algorithm performs consecutively maximum shifts of $(\mathrm{NR}-1)/2$. The quantizer overloads whenever its input goes beyond the maximum and minimum levels. The numbers represent the NR levels and the short lines the $\mathrm{NR}-1$ comparator thresholds. 109

4.21 Auto-ranging design chart: *reduction factor* $\mathrm{NL}/\mathrm{NR}_{\mathrm{min}}$ for an SNQR of 94dB as functions of the required NL and OSR for the first-, second-, third- and fourth-order cases. The solid and dashed lines correspond to the cases where a maximum full-scale signal frequency is respectively twice or equal to the bandwidth. 111

4.22 Required NL and OSR for a targeted SNR and a given NR. The specified band is twice the bandwidth and the SQNR is 10dB higher than the target. Identical bandwidth and specified band are considered for the dashed curves NL′ and OSR′. The targeted applications shown in Table 4.2 are labeled and highlighted with dots. The design example considered in a subsequent chapter corresponds to the Bluetooth case addressing 84dB, which requires an OSR of 32 and 33 levels. 113

5.1 Conventional DAC implementation as a bank of $(NL-1)$ elements. 115

5.2 DEM algorithm implementations: conceptual representation of (a) the vector feedback, (b) the DWA and (c) the tree-structured encoders. The implementations shown here provide a first-order spectral shaping. 116

5.3 Tree-structured encoder: (a) example for a 9-level DAC and (b) switching block general descriptions. 117

5.4 Signal processing of the first-order shaping. The output of the hard limiter **q** is either ±1. Whether the switching block input is odd or even, the shaper output **s** is forced to zero or kept unchanged. 119

5.5 DEM testbench to extract the mismatch errors e. α and β are the DAC gain and offset of equation (5.4). 120

5.6 Mismatch errors with a full-scale input signal that is (a) a constant, (b) a rounded sine-wave and (c) a random Gaussian variable. A band-of- interest of 500kHz is highlighted in gray. The encoder is clocked at 32 MHz 121

5.7 Mismatch errors with a $\Delta\Sigma$-modulator output as an input signal with the variable **s** set to (a) a constant value of either ±1, (b) randomly chosen between ±1 and (c) controlled by the internal accumulator. A band-of-interest of 500kHz is highlighted in gray. The encoder is clocked at 32 MHz. 122

5.8 Signal processing block diagram of the second- and third-order mismatch shaper. 123

5.9 Mismatch errors with (a) a second-order shaping, (b) a second-order with a 3-bit internal accumulator and (c) a third-order shaping with an optimized resonance. A band-of-interest of 500kHz is highlighted in gray. The encoder is clocked at 32 MHz. 124

5.10 Tree of solutions as the output of the first synthesis step. Each node of the tree is a list of all the possible blocks. Each block has two branches representing the block outputs. 126

5.11 Removal of redundant solutions in the solution tree performed during the second synthesis step. 127

5.12 Output after the last step of the synthesis. 127

5.13 Example of a synthesized tree for a 33-level segmented DAC using the three switching blocks, F, H and Q. The digital signals are represented by an extra-LSB code. 128

5.14 Classification of the 628 solutions found by the synthesis process for a 33-level DAC using F, H and Q blocks. 130

5.15 Standard segmentation. Solution number 589 is a standard segmented tree with a segmentation depth of 2. 131

6.1 System clocking diagram. 135

6.2 Schematic of the 33-level second-order ΔΣ-modulator. 136

6.3 Modulator block diagram. The processing blocks are labeled as in the circuit description in Figure 6.2. 138

6.4 Linear signal-flow graph representation of the modulator. The integration paths are highlighted in bold. 138

6.5 Behavioral modeling of a differential (a) and differential difference (b) amplifier. 147

6.6 Continuous-time integrator amplifier. 149

6.7 common-mode feedback circuit. The sizes of the current mirror transistors bringing the control signals **cmfb** are different for each amplifier and are therefore not displayed. 150

6.8 Sampling switches and control circuit. 151

6.9 Sampling buffer amplifier. 152

6.10 Discrete-time integrator amplifier. 153

6.11 Quantizer latched comparator. 155

6.12 Simulated offset distribution with Monte-Carlo analysis for a differential reference voltage of (a) 40mV and (b) 360mV. . . . 156

6.13 Clocking diagram for comparators. 156

6.14 M-3M DAC used for the offset compensation. 157

6.15 Simulated DAC effective resolution for the M2M (a), M2.5M(b) and M3M (c) architectures. 158

6.16 Statistical distribution of the offset before (a) and after (b) compensation with the analytical model. 159

6.17 Statistical deviation expected after compensation with a 24-level DAC and different step size Δ. An initial offset of 12mW is considered. The dots show the simulated compensation of Figure 6.18. 160

6.18 Offset statistical distribution after compensation. The simulation is performed at transistor level with 800 Monte-Carlo points. 161

6.19 Power spectral density of the modulator output simulated at transistor level. 162

6.20 Layout of the circuit in a 0.18μm CMOS technology. 163

A.1 General description of the differential pair with resistance degeneration. 178

List of Tables

2.1 Feedback coefficients to place the poles at the center of the unity circle, optimal zero placement on the unity circle according to [ST05]. The nominal values of the global coefficients a_i are in bold to highlight the *Pascal triangle* configuration. 38

3.1 Required OSR for an NL-level nth-order modulator with SQNR=94dB according to Equation (2.39). 53
3.2 Continuous-time jitter transfer factor at low frequency according to Equation (3.19). 70
3.3 Feedback coefficients with their respective tolerance to guarantee stability. 77
3.4 Calculated fifth-order cases with OSR=32. 81

4.1 Shift control signal. 107
4.2 Extension to typical applications for a second-order modulator with NR=11. 112

5.1 Switching blocks description 125
5.2 Gate count for switching blocks with a second-order shaper with a 3-bit accumulator. 129

6.1 System characteristics. 134
6.2 CMOS technology features used. 137
6.3 Modulator design procedure. 142
6.4 Consumption and chip area summary. 162
6.5 Performance comparison. 165

List of symbols

NTF	Noise Transfer Function
STF	Signal Transfer Function
NL	Number-of-Levels
NR	Reduced Number-of-Levels
N	Number-of-bits
DR	Dynamic Range
ENOB	Effective Number-of-Bits
SNR	Signal to Noise Ratio
SQNR	Signal to Quantization Noise Ratio
SNDR	Signal to Noise plus Distorsion Ratio
SJNR	Signal to Jitter Noise Ratio
STNR	Signal to Thermal Noise Ratio
OSR	Over-Sampling Ratio
FS	Full-Scale
DC	Zero Frequency
RZ	Return-to-Zero
NRZ	No-Return-to-Zero
RMS	Root Mean Square value
PSD	Power Spectral Density
$P_{a,b}$	Total Power of signal a referred to b
$\mathscr{P}_a(f)$	Power spectral density of signal a
$\mathscr{L}\{.\}$	Laplace Transform
$\mathscr{Z}\{.\}$	Z-Transform
$\binom{a}{b} = \frac{b!}{(b-a)!a!}$	Binomial coefficient
κ	Control function
η	Normalized gain mismatch
ξ	Jitter transfer factor
γ	Number-of-step change
σ_a	Standard deviation of a
Δ	Quanitzation and DAC elements steps
k_x	Signal range factor
k_q	Noise range factor

T_s	Sampling period
f_s	Sampling frequency
f_b	Band-of-interest
$f_{\max}$	Highest frequency for full-scale input signals
c	Number of continuous-time integrators
n	Modulator order
m	Segmentation depth
$a_1, a_2, \ldots, a_n$	Global feedback coefficients
$b_1, b_2, \ldots, b_n$	Local feedback coefficients
$c_1, c_2, \ldots, c_n$	Feedforward coefficients
$d_1, d_2, \ldots, d_n$	Interpolation coefficients
x, $X(z)$	Modulator input
v, $V(z)$	Quantizer input
y, $Y(z)$	Quantizer output
q, $Q(z)$	Quantization error
R	Modulator input resistance
C	First integrator capacitor
C_s	Sampling capacitor
C_d	DAC element capacitor
C_f	Feedback capacitor
I_{REF}	Current steering DAC reference current
V_{REF}	Switched-capacitor DAC reference voltage

List of publications

[1] D. Stefanovic, S. Pesenti, M. Pastre, M. Kayal, Structured design based on the inversion parameter: Case study of DS modulator system, In *Proceedings of the International Conference on Mixed Design of Integrated Circuits and System,MIXDES*, June 2007, pages 95-102.

[2] S. Pesenti, P. Clement, M. Kayal, Reducing the number of comparators in multi-bit $\Delta\Sigma$ modulators, IEEE Transactions on circuits and systems I, 2007, (submitted third revision).

[3] S. Pesenti, P. Clement, D. Stefanovic, M. Kayal, A low-power strategy for Delta-Sigma modulators, In *Proceedings of the International Conference on Mixed Design of Integrated Circuits and System, MIXDES*, June 2007, pages 203-208. Received the Outstanding Paper Award.

[4] M. Demierre, S. Pesenti, J. Frounchi, P.-A. Besse and R. S. Popovic, Reference magnetic actuator for self-calibration of a very small Hall sensor array, Sensors & Actuators A: Physical, Vol 97, pp. 39-46, April 2002.

[5] M. Demierre, S. Pesenti, and R. S. Popovic, Self calibration of a CMOS twin Hall microsystem using an integrated coil, In *Proc. 16th Eurosensors Conference*, Prague, Czech Republic, pp. 573-574., 15-18 Sept. 2002.

[6] M. Demierre, S. Pesenti, J. Frounchi, P.A. Besse and R.S. Popovic, Reference magnetic actuator for self calibration of a very small Hall sensor array, *TRANSDUCERS'01, Eurosensors XV*, Munich, Germany, June 2001.

[7] S. Pesenti, P. Clement, M. Kayal, EPO patten Application PCT/EP2006/060979, ”Electronic device and integrated circuit comprising a Delta-Sigma converter and method therefor”, (pending).

.

Curriculum vitae

Nom:	Pesenti
Prénom:	Sergio
Date de naissance:	10 juin 1973
Lieu de naissance:	Vevey, Suisse
Lieu d'origine:	Corsier-sur-Vevey (VD)
Etat civil:	Marié
Adresse:	Rue des deux Marchés 15, CH-1800 Vevey

Formation

1997–2001	**Diplôme d'Inégnieur Microtechnicien EPFL** Ecole Polytechnique Fédérale de Lausanne
1993–1997	**Diplôme d'Inégnieur ETS en méchanique** Ecole d'Ingénieurs de l'Etat de Vaud à Yverdon
1989–1993	**Certificat Fédéral de Capacité de Dessinateur de machines A** Ateliers de constructions méchaniques de Vevey S.A. **Bacalauréat Technique** Ecole Profesionnelle de la S.I.C de Lausanne

Activités professionelles

2006–	Mixed-Signals IC Designer Marvell Switzerland, CH-1163 Etoy
2001–2006	Assistant de recherche au Laboratoire d'Electronique Générale Ecole Polytechnique Fédérale de Lausanne, CH-1015 Lausanne
1993–1999	Dessinateur de machine Bombardier Transport S.A., CH-1844 Villeneuve