Low Power Floating Point A/D Converters for Audio Signals

Louisa Grisoni-Busca

THESE PRESENTÉE A LA FACULTE DES SCIENCES
DE L'UNIVERSITE DE NEUCHATEL POUR L'OBTENTION
DU GRADE DE DOCTEUR ES SCIENCE

February 1998

...Recherche et poursuis la paix.

Psaumes 34, 15

IMPRIMATUR POUR LA THÈSE

Low Power Floating Point A/D Converters for Audio Signals

de Mme Louisa Grisoni-Busca

UNIVERSITÉ DE NEUCHÂTEL

FACULTÉ DES SCIENCES

La Faculté des sciences de l'Université de Neuchâtel sur le rapport des membres du jury,

MM. F. Pellandini (directeur de thèse),
N. De Rooij, P. Balsiger, P. Zbinden (Bernafon AG, Berne)
et P.-A. Farine (ASULAB, Marin)

autorise l'impression de la présente thèse.

Neuchâtel, le 20 février 1998

Le doyen:

F. Stoeckli

Summary

Progress in low power micro-electronic technologies and digital signal processing has opened the way for numerous digital portable applications. To improve the overall power consumption and consequently battery life of such devices targeted low power A/D converters are required.

Analog to Digital Converters (ADCs) can be separated into two main classes. The first includes all the devices featuring a linear transfer function while the second one consists of all the devices featuring a non linear transfer function. Within the linear class, four different sub-categories can be identified: absolute, floating point, relative precision and mixed.

Absolute linear converters are the only ones to posses a constant maximum quantization error over the whole dynamic range. Linear floating point, relative precision and mixed converters are non-absolute and their quantization error varies. In particular, in the floating point linear converters, the transfer function can be divided into two or more zones that separately present an absolute characteristic. In other words, each of these zones has a constant maximum quantization error, which value however, is different for each zone.

Absolute linear A/D converters must be used to convert absolute analog signals that will be finely analyzed in the subsequent processing unit. In most of the other situations, and in particular signal measurement, non-absolute linear devices are sufficient.

Compared to those for absolute converters, the design constraints for non-absolute devices are less compelling. As a result less power consuming implementations can be obtained. Hence, non absolute converters are extremely well suited for battery operated applications where a high dynamic range is required and a limited resolution can be tolerated. This is typically the case in many portable audio applications as well as in areas of instrumentation, control, radar etc.

This work is a contribution to the creation of new low power non-absolute converters. In particular, floating-point linear A/D conversion is considered. The idea is to scale (or adapt) the input signal in such a way that it fits well into the fixed range of a coarse quantizer. Feed forwards and feed back adaptation must be distinguished. In the first, the same input sample is used to evaluate the scaling factor and to perform the quantization. In feed back adaptation, the scaling factor is chosen according to a prediction based on the previous sample amplitude. Hence it is not suited for unpredictable signals.

Considering converters for audio signals, different system level solutions as well as realization schemes are discussed. A dedicated audio feed back floating point converter was implemented in a low voltage 2 µm CMOS technology. The measured characteristics are 13 bits dynamic range, 9 bits resolution and 50 µW at 16 kHz and ± 1.3V and the perceived quality is excellent.

The floating point concept is then extended to evaluate the cost of higher dynamic, sampling frequency and/or resolution and to include other types of applications such as instrumentation, radar etc.

This work was supported by the "Commission pour l'Encouragement à la Recherche Scientifique" (grant CTI-2747.1) and the MICROSWISS program (grant TR-IT-005).

Résumé

Grâce aux progrès des nouvelles technologies micro-électronique et du traitement numérique du signal, de nombreuses applications numériques portables sont désormais possibles. Des circuits de conversion analogique digitale dédicacés basse consommation sont toutefois nécessaires pour améliorer la consommation globale et ainsi prolonger la durée de vie de la batterie.

Les Convertisseurs Analogique/Numérique (CAN) peuvent être subdivisés en deux classes principales: la première comprenant les circuits dont la fonction de transfert est linéaire, la seconde comprenant ceux qui possèdent une fonction non linéaire. Parmi le convertisseurs linéaires, on peut encore distinguer quatre sous-catégories: les CAN absolus, à virgule flottante, mixtes et à précision relative.

Les CAN absolus sont les seuls convertisseurs ayant une erreur de quantification maximale constante sur toute la gamme dynamique. Les CAN à virgule flottante, mixtes et à précision relative sont non-absolus et leur erreur de quantification varie. En particulier, dans le cas des CAN à virgule flottante, la fonction de transfert peut être subdivisée en deux ou plusieurs régions qui séparément, présentent une caractéristique absolue. Autrement dit, chacune des régions possède une erreur de quantification maximale constante, dont la valeur est toutefois différente pour chaque région.

Les CAN linéaires absolus sont utilisés pour convertir des signaux absolus qui doivent être finement analysés lors du traitement successif. Dans la plupart des autres cas et en particulier lors de mesure de signaux, les CAN linéaires non-absolus sont suffisants.

Les contraintes de dimensionnement des CAN non-absolus sont moins contraignantes que celles des CAN absolus. Il en résulte des circuits consommant moins d'énergie. Les CAN non-absolus sont donc très bien adaptés pour des applications alimentées par batteries où une grande dynamique est nécessaire mais où une résolution limitée est suffisante. Cela est typiquement le cas dans plusieurs applications portables audios ainsi que pour des applications particulières d'instrumentation, de contrôle, de radar etc.

Ce travail est une contribution à la création de nouveaux CAN non-absolus à basse consommation. Il s'intéresse en particulier aux CAN linéaires à virgule flottante. L'idée est de mettre à l'échelle (d'adapter) un signal d'entrée de telle façon à ce qu'il s'ajuste bien à la gamme d'un CAN grossier. Il faut distinguer l'adaptation « en avant » (sans prévision) et « en arrière » (par prévision). Pour la première, le même échantillon est utilisé pour évaluer le facteur d'échelle, et pour effectuer la quantification. Pour la seconde, le facteur d'échelle est choisi selon une prédiction de l'échantillon à traiter basée sur l'amplitude de l'échantillon précédant. Cette seconde méthode s'applique donc à des signaux qui peuvent être plus ou moins prédits.

Ce travail s'applique à la conversion analogique-numérique de signaux audios et étudie différents types de solutions ainsi que des schémas de réalisation. Un CAN à virgule flottante et à adaptation « en arrière » dédicacé a été réalisé dans une technologie CMOS 2 µm. Les performances mesurées sont 13 bits de gamme dynamique , 9 bits de résolution et 50 µW de consommation pour une fréquence d'échantillonnage de 16 kHz et une alimentation de ± 1.3V. La perception acoustique est excellente.

Pour élargir le domaine d'application de la conversion à virgule, les coûts pour des fréquences d'échantillonnages plus rapides et des gammes dynamiques et/ou résolutions plus élevées sont évalués.

Ce travail a été financé par la Commission pour l'Encouragement à la Recherche Scientifique (subside CTI-2747.1) et le programme MICROSWISS (subside TR-IT-005).

Table of content

1. INTRODUCTION 1

1.1 Motivation 1

1.2 Scope of the research 2

1.3 Main contributions 4

1.4 Structure of report 5

1.5 Previous publications by the author 5

1.6 References 6

2. FUNDAMENTALS OF A/D CONVERERS 7

2.1 A/D converter specification and terminology 8

2.1.1 Transfer function (TF) 8
2.1.2 Dynamic range, quantization error, resolution and static signal to noise ratio 8
2.1.3 Step Over Range (SOR) 11
2.1.4 Gain and offset error 12
2.1.5 SNR, THD, PD and SINAD 12

2.2 Linear A/D converter's types 13

2.2.1 Single and dual slope integrating converter 13
2.2.2 Voltage to frequency converter 15
2.2.3 Oversampling or Sigma-Delta converter 15
2.2.4 Counter ramp converter 16
2.2.5 Successive approximation converter 17
2.2.6 Parallel converter 18
2.2.7 Pipelined converter 18

2.3 Market and academic overview 19

2.4 References 22

2.4.1 References to paragraph 2.1 and 2.2 22
2.4.2 References to paragraph 2.3 23
2.4.3 References to academic non linear converters 24

3. FUNDAMENTALS OF LOW POWER DESIGN 25

3.1 Digital design 26
3.1.1 Technology optimization 27
3.1.2 Physical, circuit and logic level optimization 28
3.1.3 Architecture level optimization 29
3.1.4 Algorithmic level optimization 30
3.1.5 Low power design EDA tools 31
3.2 Analog design 32
3.2.1 MOS transistor 32
3.2.2 OTA 34
3.2.3 Analog Switches 35
3.2.4 Resistor and capacitors 36
3.2.5 Low voltage supply 36
3.2.6 Simulation models 37
3.3 Technologies 38
3.4 Batteries 39
3.4.1 Primary cells 39
3.4.2 Secondary cells 40
3.5 References 42
3.6 Annex: List of symbols 44

4. FLOATING POINT CONVERSION

4.1 Original feed back floating point converter 47
4.2 Enhanced feed back floating point converter 49
4.2.1 Improvement of perceived quality 49
4.2.2 Simplified implementation 52
4.2.3 Ideal characteristics 58
4.3 Feed forward conversion 60
4.3.1 Feed forward conversion for audio signals 60
4.3.2 Ideal characteristics 61
4.4 Floating point conversion for non audio signals 62

4.5 Design considerations 64

4.6 Converters for audio applications: summary 65

4.6.1 Feed back reserve-bit converter 65
4.6.2 Feed back 6 dB converter 66
4.6.3 Feed forward converter 66

4.7 References 66

4.8 Annex 66

5. MICRO POWER RSD CONVERTER 69

5.1 Alexandre Heubi's converter 70

5.1.1 Mixed linear characteristic 71
5.1.2 Implementation and test 72

5.2 14-bit absolute RSD converter 75

5.2.1 Implementation 77
5.2.2 Test 79

5.3 References 81

6. DESIGN OF FLOATING POINT A/D CONVERTERS 83

6.1 Design methodology 84

6.2 Controlled amplifier 85

6.2.1 Limitations 86
6.2.2 Capacitors network schemes 87
6.2.3 Switched capacitor schemes 90
6.2.4 Integrating schemes 91
6.2.5 Design of the controlled amplifier 93
6.2.6 Simulation results 101
6.2.7 Wrap up 102

6.3 Coarse quantizer 103

6.4 Adaptation logic (reserve-bit converter) 103

6.4.1 Task 1: increment/decrement computation 104
6.4.2 Task 2 and 3: accumulation and floor to 6d8 multiples 104
6.4.3 Task 4: Control signals S_1 to S_7 106
6.4.4 Task 5 and 6: Memorizing and shifting 106
6.4.5 Simulation of adaptation logic and shifter 106

6.5 Analog and digital control 108
6.6 Chip presentation 109
6.7 6 dB adaptation strategy 110
6.8 Feed forward converter 111
6.9 To the limits 112
6.10 References 113
6.11 Annex 114

7. IMPLEMENTATION RESULTS 117

7.1 Test equipment 118
7.2 Measurements 118
7.2.1 Informal listening test 118
7.2.2 Noise floor 119
7.2.3 Transfer function 120
7.2.4 Frequency response 123
7.2.5 Controlled gain 123
7.3 Comparison with A. Heubi solution (5.1) 126
7.3.1 Converters for audio applications 126
7.3.2 Converters for non audio applications 127

8. CONCLUSIONS 129

8.1 Main contributions 130
8.1.1 Low power feed back floating point A/D converter for audio applications 130
8.1.2 Low power feed forward floating point A/D converter for audio applications 132
8.1.3 Floating point converters for non audio applications 133
8.2 Future work 134

ACKNOWLEDGMENTS 135

1. Introduction

The research presented in this Ph. D. report addresses the design of micro power floating point A/D converters for audio signals aimed at battery operated micro-systems.

A/D converters are classified according to their static characteristics which strongly influence the application domain. The floating point converter is well suited for signal measurement but not for analysis.

This introduction explains the scope of the research and presents the main contributions. An overview of each chapter's contents is also included as well as a list of related papers written by the author and already published.

1.1 MOTIVATION

Some 35 years ago the first digital systems were designed. A few sceptical engineers though, predicted that the complexity of digital systems would curb any major use. Indeed, to achieve the same functionality of an analog system, the digital equivalent (figure 1.1) required an anti-aliasing filter, two converters i.e. Analog to Digital (A/D) and Digital to Analog (D/A) and a smoothing filter, resulting in higher power consumption and size .

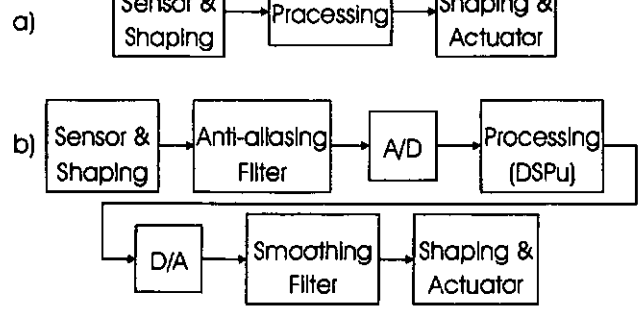


Figure 1.1 : Analog (a) and digital (b) systems

However, these engineers didn't foresee the tremendous evolution and improvement in silicon technologies. Nowadays, digital processing can handle complex functionality combining high performance and tunable parameters which cannot be realized in the analog domain.

Electronic Design Automation (EDA) tools considerably facilitate the design of digital devices and standard A/D and D/A interfaces are now available. As a result, small and low cost systems are obtained and new digital consumer products hit the market every day as for example media control devices, speech processor, navigation systems, portable multimedia etc.

Many of these new products are battery operated and the designers are asked to minimize the power consumption of each implemented component. Some tools focused on that goal are now available and provide great help in optimizing the Digital Signal Processing unit (DSPu). However, small low power A/D and D/A CMOS realizations are uncommon. In particular, converters featuring medium to high dynamic range (12-16 bits), slow to medium sampling frequency (10-50 kHz) and low power consumption (~100 μW), well suited for portable audio, instrumental or telecommunication applications, are typically not commercialized.

The presented work is a contribution to the development of such targeted A/D converters.

1.2 SCOPE OF THE RESEARCH

Analog to Digital Converters (ADCs) can be separated into two main classes. The first includes all the devices featuring a linear transfer function (digital output versus analog input characteristic) while the second one consists of all the devices featuring a non linear transfer function. Non-linear converters are found in instrumentation, communication, nuclear science and control applications [Mahm92]. Since they do not enter the field of this research they are not further considered. Within the linear class four different sub-categories can be identified: absolute, floating point, relative precision and mixed.

Absolute linear converters are the only ones to posses a constant maximum quantization error over the whole dynamic range. In the ideal case, the maximum quantization error is a half Least Significant Bit (LSB) while the dynamic range is defined by the number of converted bits n and amounts to 2^n-1 LSBs. These converters also feature a linear static Signal to Noise Ratio (SNR) characteristic. Floating point, relative precision and mixed converters are non-absolute and their maximum quantization error varies. In floating point linear converters, the transfer function can be divided into two or more zones that

separately present an absolute characteristic. In other words, each of these zones has a constant maximum quantization error, whose value is different for each zone. As a result, the static SNR is made of several shifted linear segments. Relative precision linear converters feature a quantization error that is proportional to the input level. This implies a constant static SNR. Finally, mixed linear converters present a combination of two of the above characteristics.

Absolute linear A/D converters must be used to convert absolute analog signals that will be finely analyzed in the subsequent processing unit. In most of the other situations, the non-absolute linear converters can be used. Indeed, if an analog signal is not absolute in the sense, for example, that its noise level increases with the signal amplitude, a converting device whose quantization error follows the analog signal noise level over the whole dynamic range is sufficient. Another case would be that of processing algorithms that only require a minimal SNR from their converted digital input signal. Signal measurement is a typical case.

Strictly speaking, absolute converters could also be used in both the above situations. However, this would result in a waste of energy since some of the computed bits would be either noisy (and thus worthless) or useless for the considered algorithm. The primary goal of non-absolute converters is thus to achieve reduced power consumption while keeping a conversion quality that is well suited to the considered application.

Non-absolute converters are particularly well suited to certain audio applications. Indeed, the human auditory system is a non-absolute sensor: the perceived noise loudness is not only determined by the noise power but also depends on the main signal power and its distribution along the basilar membrane [Flet40]. Thus, depending on the main signal, a perceived noise loudness can be reduced or even made completely inaudible. This phenomenon is known as auditory masking [Schr79]. Consequently, although 14 bits (13 for the amplitude and 1 for the sign) are required to convert sounds ranging from quiet sleeping room to discotheque or air hammer, a resolution of 6 or 7 bits is sufficient to reach, once the signal is converted to analog again, an excellent perceived quality [Schr79]. The amplitude of the various sounds can thus be quantified as in figure 1.2. Loud (top) and soft (bottom) sounds in both absolute (left) and floating point (right) representations are given. b_i are the relevant bits while n_i are "useless" (the ear won't be able to hear them). In the floating point representation, M stands for mantissa, while E is the exponent or in other words, the number of shift left (divide by two) that must be applied to the mantissa to obtain the actual signal amplitude.

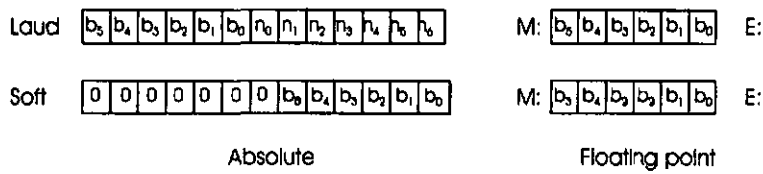


Figure 1.2 : Sound coding using an absolute and floating point code

This work is a contribution to the creation of new nan absolute converters. The development and design of floating-point linear A/D canverters, where a coarse quantizer providing the necessary resolution is preceded by a controlled gain that increases the dynamic range, are explained. Considering converters for audio signals, different system level solutions as well as implementation schemes are discussed. The floating point concept is then extended to evaluate the cost af higher dynamic, sampling frequency and/or resolution and to include other types of applications such as instrumentation, radar etc.

1.3 MAIN CONTRIBUTIONS

In floating point linear converters, the input signal is scaled in such a way that it fits well into the fixed range af a coarse quantizer. Feed forwards and feed back adaptation must be distinguished. In the first, the same input sample is used to evaluate the scaling factor and to perform the quantization. In feed back adaptation, the scaling factor is chosen according to a prediction based on the previous sample amplitude. The main contribution af this reseorch is the development of the enhanced feed back floating point conversian concept. A methodology to successfully scale the input sample is also proposed. The concept was applied to audio signal and a dedicated chip was implemented in a low voltage CMOS technology. The measured characteristics are 13 bits dynamic range, 9 bits resolution and 50 μW at 16 kHz sampling frequency and ± 1.25V supply voltage. The perceived quality was excellent.

The feed forward adaptation is useful for nan predictable signals. However, its use in the case of audio signals is also considered. Chip estimates show that a higher power consumption is nevertheless required.

The limits of the implementation of the floating point conversion concept in a low voltage CMOS technology have also been evaluated. Considering a targeted resolution of 14 bits and a power consumption of less then 1mW, a maximum dynamic range of 18 bits could be realized.

Another contribution is the design and implementation in a low voltage CMOS technology of a 14-bit absolute linear version of A. Heubi's mixed RSD converter [Heub96]. This device is important because it can be used as coarse quantizer in high resolution floating point converters.

1.4 STRUCTURE OF REPORT

Following this short introduction, chapter two overviews A/D converter characteristics and types as well as their typical application domains. A non-exhaustive general survey of recent publications and commercialization is also included. Chapter three gives some background in low power analog and digital design as well as silicon CMOS technologies. A general survey of batteries is also included. These two chapters introduce the reader more deeply into the general context of this work and can be skipped without loss of understanding of the subsequent chapters.

All the remaining chapters, except for the first part of chapter five, form the heart of this work. In chapter four, the floating-point approach and its variants are described from a system level point of view. The discussion is targeted to audio signals. Chapter five first overviews the Redundant Signed Digit (RSD) converter developed by Alexandre Heubi [Heub96]. This device features a mixed linear characteristic and is well suited for comparison with the floating point approach. The second part of chapter five describes an absolute version of the RSD and presents results from chip integration in a 1 µm low voltage CMOS technology. These results are mandatory to extend the floating point concept to non audio signals. Implementation solutions for each component of the floating-point converter are detailed in chapter six which terminates with chip estimates for both audio and non-audio devices. One of the proposed audio chips was integrated in a 2 µm low voltage CMOS technology and was tested in IMT's laboratory. Chapter seven presents the obtained results and compares the floating point approach with the RSD solution. Finally, chapter eight draws the conclusions.

1.5 PREVIOUS PUBLICATIONS BY THE AUTHOR

The work described in this document has already been the subject of some publications. A first scientific paper was presented at the International Conference on Signal Processing Application and Technology (ICSPAT) in Boston in October 95 [Gris95]. The principles of the floating-point converter were discussed as well as implementation ideas. A second paper [Gris96_1], presented at the International Symposium on Low Power Electronics and Design (ISLPED) in Monterey in August

96, gave implementation results. A paper explaining the steps to transform the mixed linear RSD converter into an absolute device was also published at ICSPAT'96 [Gris96_2]. Implementation results were presented at the seventh International Symposium on IC technology systems and applications (ISIC'97) [Gris97] in Singapore.

[Gris95] L. Grisoni, A. Heubi, S. Grassi, P. Balsiger and F. Pellandini, A. Schaub « Micro Power Relative Precision 15-bit A/D Converter », ICSPAT'95, Oct. 24-26 1996, Boston MA, USA, pp 420-424.

[Gris96_1] L. Grisoni, A. Heubi, P. Balsiger and F. Pellandini, « Implementation of a Micro Power 14-bit Floating-Point A/D Converter », ISLPED'96, Aug. 12-14 1996, Monterey CA, USA, pp 247-252.

[Gris96_2] L. Grisoni, A. Heubi, P. Balsiger and F. Pellandini, « Micro Power 14-bit RSD A/D Converter », ICSPAT'96, Oct. 8-10 1996, Boston MA, USA, pp 510-514.

[Gris97] L. Grisoni, A. Heubi, P. Balsiger and F. Pellandini, « Micro Power 14-bit ADC: 45 µW at ±1.3V and 16 ksamples/s », ISIC'97, 10-12 September 97, Singapore.

1.6 REFERENCES

[Heub96] A. Heubi, P. Balsiger and F. Pellandini, « Micro Power 13 bits Cyclic RSD A/D Converter », ISLPED'96, Aug. 12-14 1996, Monterey CA, USA, pp 253-257.

[Mohm92] K. M. Mohmoud, R. F. Wolffenbuttel, « A Non-Linear A/D Converter for Smart Silicon Sensors », Proceedings of ISCAS'92, Vol. 1, pp 605-608, IEEE, 1992.

[Schr79] M. R. Schroeder, B. S. Atal, J. L. Hall, "Optimizing Digital Speech Coders by Exploiting Masking Properties of the Human Ear", J. Acoust. Soc. AM. 66, Dec 1979, pp 1647-1642.

[Flet40] H. Fletcher « Auditory Patterns », Reviews of Modern Physics, Vol. 12, pp 47-65, January 1940.

2. Fundamentals of A/D Converters

This chapter provides a general survey of A/D converter fundamentals.

The various linear converters are explained and the identification of their application domains shows that the floating point conversion is particularly well suited for signal measurement.

The specific terminology used throughout the report is also introduced, followed by an overview of major linear A/D converter types. Their ability to meet the targeted medium to high dynamic range (12-16 bits), slow to medium sampling frequency (10-50 kHz) and micro power consumption (~100 mW) specification is also discussed.

Some of the commercially available converters as well as the academic ones are presented in a comparative table and classified according to an "efficiency" factor that takes into account dynamic range, sampling frequency and power consumption. The targeted A/D converter features a greatly improved efficiency compared to marketed devices featuring the same dynamic range.

The choice of a converter depends on many factors. The most important of these are certainly the characteristics of the analog signal to be converted and the information that must be captured. This usually determines the desired dynamic range, sampling frequency and maximum quantization error. Operating environment will define constraints such as power consumption, size, packaging, etc. as well as reliability needs. Last (but not least !), price has to be considered as well. The final choice is always a trade off between accuracy, speed and simplicity !

The next paragraphs include information harvested from the articles and books listed in section 2.4.1, at the end of this chapter. Nevertheless, should the reader decide to consult only a few of them, [Haes94] and [Teme93] are certainly the most relevant.

2.1 A/D CONVERTER SPECIFICATION AND TERMINOLOGY

According to specialists, several tens of parameters should be used to fully specify A/D converters. Unfortunately, depending on the device manufacturer, the definition of these parameters can differ.

This section only defines parameters that are used throughout this report.

2.1.1 Transfer function (TF):

The transfer function maps the analog input into digital output code. Figure 2.1 shows the ideal transfer function for a 3 bit unipolar (no sign bit) linear converter (bold). The gray curve shows a logarithmic characteristic that is typical of non linear converters used in telecommunications for example. More details about the various linear transfer functions are given in the next section, where other important parameters are defined as well.

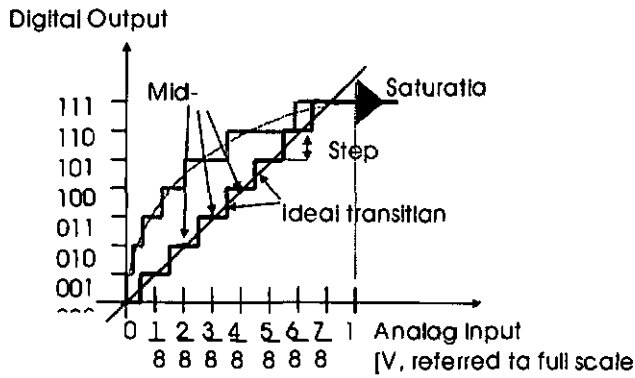


Figure 2.1 : Transfer function for linear and non linear unipolar 3-bit ADC

Note that a range of analog inputs can result in the same output code: for example, analog values from 0 to 1/16 in the absolute linear characteristic, correspond to the digital word 000.

2.1.2 Dynamic range, quantization error, resolution and static signal to noise ratio:

Let us first give a "textual" definition of these parameters. A figure will then help to clarify these words!

The dynamic range (DR) is the ratio of the largest analog input that can be converted to the smallest one. A dynamic range is often expressed in decibels [dB].

The quantization error is the difference, in magnitude, between analog input and resulting digital code. It is normally given in 'minimum step size' units.

The resolution indicates the precision or accuracy of the converter in bits.

DC signals are used to obtain the Static Signal to Noise Ratio (as opposed to dynamic SNR, section 2.1.5) and the quantization noise level versus the DC input amplitude is plotted. Static SNR is useful from a theoretical point of view, to explain the difference between the various linear converters. The dynamic range and maximum resolution can be extracted from the Static SNR plot.

Figure 2.2 shows the transfer function, quantization error and static SNR for two particular linear absolute (left) and floating point (right) converters. Both feature the same dynamic range (DR).

In absolute converters, the steps of the transfer function are all identical. The smallest analog increment that must be converted defines the step size and hence the LSB. Sixteen levels are needed to cover the whole dynamic range: a four bit converter is thus used. The maximum quantization error over the whole dynamic range is a half LSB. The maximum resolution (MR) is 4 bits as shown on the static SNR plot which features a linear characteristic.

In linear floating point converters, the quantization step is enlarged for increased signal amplitude and a "mantissa-exponent" internal representation is used. The smallest analog increment still defines the LSB while the required resolution defines the mantissa's number of bits. To reach the necessary dynamic range the analog signal is first amplified to fit into the mantissa range. The exponent keeps track of the amplification value. Combining the exponent and the mantissa results into the equivalent digital output. Floating point converters can be seen as 'piece-wise' absolute devices: the transfer function can be divided into two or more zones that separately present an absolute characteristic. In other words, each of these zones has a constant maximum quantization error, whose value however, is different for each zone. This translates into a quantization error plot with increasing "saw tooth" amplitudes. The maximum resolution (MR) is given by the mantissa size. In the floating point converter of figure 2.2, the mantissa is two bits and a pre-scaling of 1, 2 or 4 can be applied. As a result, only the following output words can be obtained: 0000, 0001, 0010, 0011, 0100, 0110, 1000, 1100. The static SNR is made of several shifted linear segments. For signal amplitudes of -24 to -12 dB, the characteristic is identical to that of the absolute converter, then the quantization step is changed and the resolution

drops by one bit. Hence, the static signal to noise drops by 6 dB. From -12 to -6 dB the quantization step is again constant and the static SNR is linear. At -6 dB, the step changes again, the resolution decreases and the 6dB drop occurs.

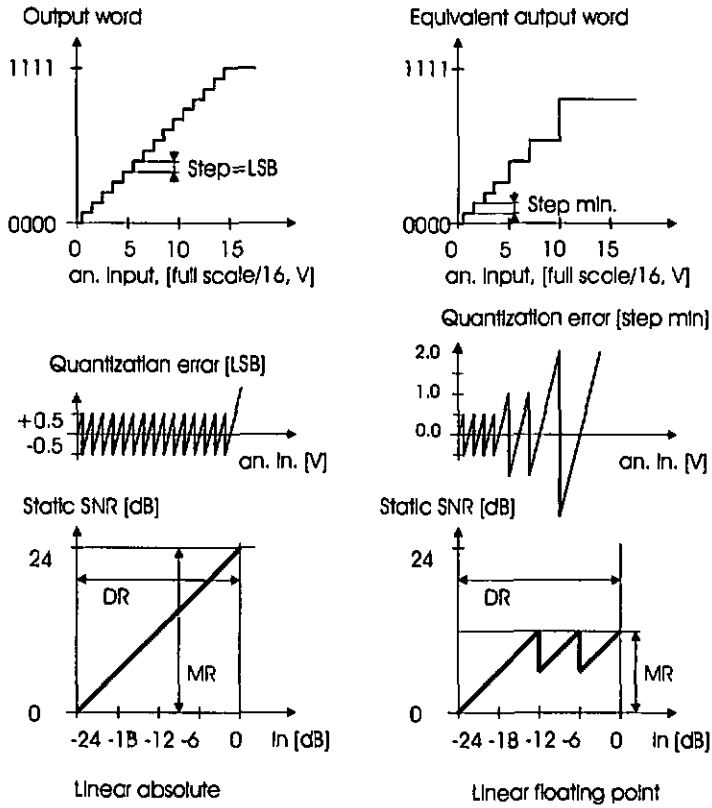


Figure 2.2 : Transfer function, quantization error and static SNR plots

Figure 2.2 does not show the case of a relative linear converter. In an ideal case, this converter would feature a quantization error that is proportional the the amplitude of the signal to be converted. As a result, the static SNR plot would be perfectly horizontal. In reality, because of the finite number of bits and the fact that quantization steps can only have predefined values (multiples of minimum step) such a characteristic is not physically achievable!

When the converted signals must be precisely analyzed, absolute linear converters are mandatory. In all the other situations, and in particular, in the case of signal measurements, an ideal relative device would be sufficient to

ensure the conversion quality while keeping the power consumption low. As mentioned, such devices cannot be realized. However, the floating point converter is a close approximation of that ideal characteristic.

2.1.3 Step Over Range (SOR) :

The Step Over Range measurement is used to determine whether an A/D converter is absolute or not.

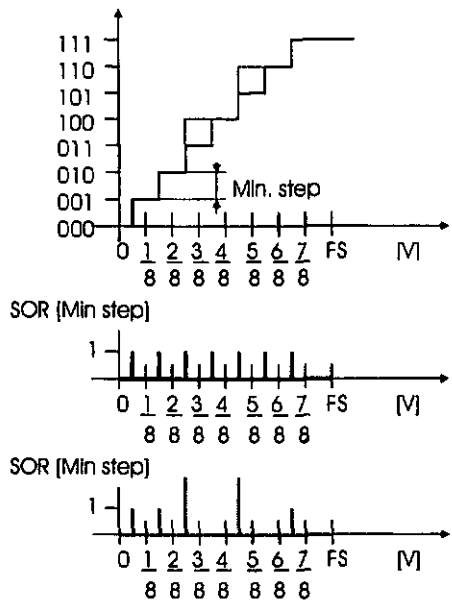


Figure 2.3 : Illustration of SOR

The converter is fed with a ramp signal covering the whole input range and of high accuracy (ramp increments smaller then the minimum converter step size). The difference between adjacent output samples, is plotted versus the analog input as in figure 2.3 which illustrates ideal cases. SOR are given in minimum ideal step size rather than in LSB. This ensures a unique notation for both absolute and non absolute devices (in floating point conversion for example, the LSB is the minimum step size of the coarse quantizer which is different from the minimum step size of the whole converter).

In absolute converters, the ideal maximum SOR is 1. Higher SORs indicate that the converter features a non-absolute behavior.

2.1.4 Gain and offset error :

The gain of a linear A/D converter is equivalent to the slope of the transfer function. The gain error is thus the deviation of the measure slope from the ideal one. The gain error is sometimes called linearity error.

The offset error identifies the deviation of the lowest transition level from the ideal location.

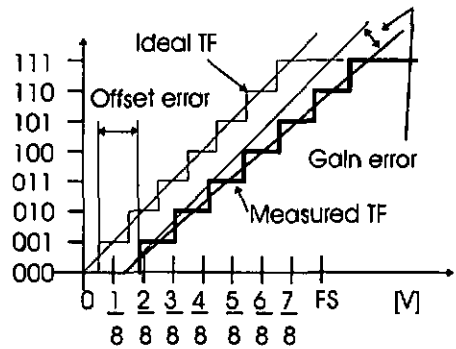


Figure 2.4 : Gain and offset errors

2.1.5 SNR, THD, PD and SINAD :

All the previous parameters refer to static characteristic. Signal to Noise Ratio (SNR), Total Harmonic Distortion (THD), Peak harmonic Distortion (PD) and Signal to Noise And Distortions (SINAD) are dynamic parameters. They are sometimes preferred to static ones for applications dealing with dynamic signals.

The dynamic characteristics are measured in the frequency domain: a pure sine wave is converted and a FFT is performed on the collected data.

Signal to Noise Ratio measures the signal power relative to noise power and is computed by equation 2.1 where M_i are the magnitudes of all spectral components except for the DC, fundamental and its harmonics.

$$SNR = -20 \cdot \log \sqrt{\sum (10^{M_i/20})^2} \qquad (dB) \tag{2.1}$$

An equation similar to that of 2.1 can be used to compute the Signal to Noise And Distortion ratio. However, in this case, M_i are the magnitudes of all spectral components except for the DC and the fundamental.

The Total Harmonic Distortion measures the power of the converter's distortion and is computed by 2.2 where M_i are the magnitudes of the harmonic spectral lines. Often, for practical reasons, only the ten first harmonics are considered.

$$SNR = 20 \cdot \log \sqrt{\sum (10^{M_i/20})^2} \qquad (dB) \tag{2.2}$$

Finally, Peak harmonic Distortion is the ratio of the fundamental magnitude to the highest harmonic peak and is given in decibels.

2.2 LINEAR A/D CONVERTER'S TYPES

Linear A/D converters can be classified as « direct-comparison » or « integrating » devices. Direct-comparison ADCs use the analog input signal to directly compute the digital equivalent by, for example, comparison to one or several reference voltages (flash, successive approximations, pipelined, etc.). In integrating architectures, the analog signal is first transformed into an intermediate signal which is then used to obtain the digital output (ramp-integrating, voltage-frequency, Sigma-Delta, etc.).

This section discusses conversion principles but no electrical realization schemes are given. The ability of each converter to meet the medium to high dynamic range (12-16 bits), slow to medium sampling frequency (10-50 kHz) and low power consumption (~100 µW) requirements as defined in section 1.1 is also examined.

2.2.1 Single and dual slope integrating converter :

Single and dual slope integrating converters are described below. Generally speaking, both are well suited for slowly varying signals (1Hz-1kHz). They feature high dynamic range and resolution, no missing code (all digital words correspond to an analog input), monoticity (monotone transfer function) and good high frequency noise rejection. Very simple architectures are possible but they are extremely sensible to voltage reference error, comparator inaccuracy etc. More complex structures can however reach 20 bits of resolution. The main applications are precision instrumentation and telemetry.

Single slope converter (figure 2.5) : a reference signal is integrated and the output of the integrator is compared to the analog input signal. A counter measures the elapsing time. When the output voltage from the integrator equals

the analog input, the counter is stopped. Its final count is used to obtain the digital equivalent of the analog input signal.

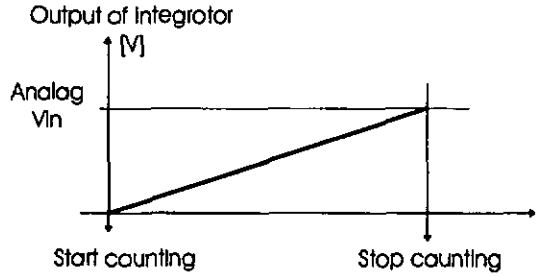


Figure 2.5 : Single slope integrating converter

Dual slope integrating converter (figure 2.6): an up/down integration is performed. A first integration is done on the input signal and lasts a fixed interval of time t_1. The input of the integrating circuit is then switched to a known reference signal and down integration takes place until the integrator reaches a fixed level (0 V in the figure). The down integration time gives a measurement of the analog input signal. The benefit compared to single slope is that absolute errors in the ramp generation are canceled.

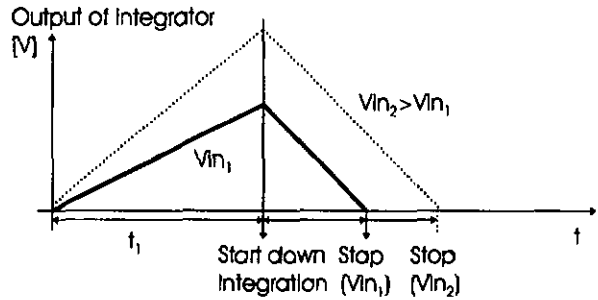


Figure 2.6 : Dual slope integrating converter

Considering the specification of section 1.1, single or dual slope integrating converters are too slow and are not a valid alternative to meet the constraints of section 1.1.

2.2.2 Voltage to frequency converter :

These converters generate digital output pulses at a repetition rate that is linearly proportional to the analog input voltage. The pulses are counted for a fixed conversion time period. Since the pulse rate varies within a conversion period as the input increases or decreases, the resulting digital output is an averaged value.

This type of converter can only be used with slowly varying signals and is thus not suited to meet the constraints of section 1.1.

2.2.3 Oversampling or Sigma-Delta ($\Sigma\Delta$) converter :

These converters have become prevalent for high accuracy (16 to 20 bits) A/D conversion of moderately high varying signals ($\leq$ 0.5 MHz). The first major use was in CD players though nowadays they can be found in modem, multimedia, digital audio and signal processing applications. The great advantage of $\Sigma\Delta$ A/D converters is that they trade greatly reduced analog circuit accuracy for increased digital complexity. This is an advantage when one considers the performance of digital VLSI silicon technologies.

$\Sigma\Delta$ are so called because they integrate (Σ) the difference (Δ) between the input signal and its bit stream equivalent. This is illustrated in figure 2.7. The feedback voltage (output from the D/A) is subtracted from the analog input. The resulting error is integrated and its polarity is detected by the comparator which drives the D/A to 1 if the integrator output is negative or to 0 if the integrator output is positive. For a 1, the D/A outputs a voltage equivalent to the full scale of the input signal. Hence, for small input values, the bit stream will be made of several 0s after a single 1. On the other hand, if the input signal is large, the bit stream will alternate 0s and 1s. The digital filter's function is thus to determine a digital word that is proportional to the number of 1s in the bit stream.

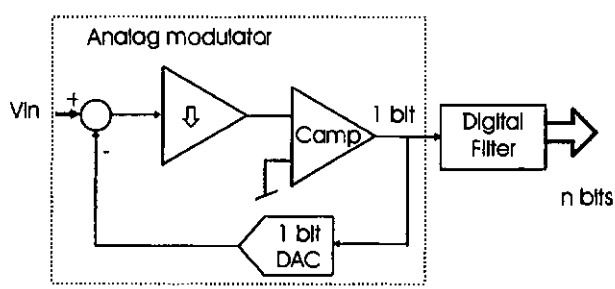


Figure 2.7 : Flow graph of $\Sigma\Delta$ converter

The sampling is performed at a much higher rate then the Nyquist rate (oversampling). Consequently, the requirements on the anti-aliasing filter are greatly reduced. In practice the filter can even be omitted sometimes.

The modulator and its associated decimation filter can have different architectures to match the specific applications. The decimation can also be performed in more then one stage by using cascaded filters working at different sampling frequencies (to avoid a costly abrupt low pass filter working at high sampling frequencies).

A very good article published by J.C. Candy and G.C. Temes [Cand92] discusses the various structures and provides all the necessary information to the interested reader.

$\Sigma\Delta$ converters fit well in the dynamic range and sampling frequency specification of section 1.1. However, the power consumption specification would be difficult to meet. The digital filter indeed requires a fairly high amount of energy.

2.2.4 Counter ramp converter :

In counter ramp converters, a D/A converter is slaved to a counter so that as the count builds up, the output of the decoder increases. The count is stopped when the output of the D/A is slightly higher then the analog input signal and is its digital equivalent.

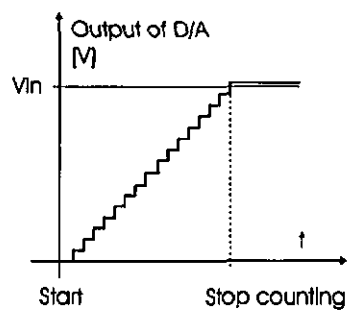


Figure 2.8 : Counter ramp converter

These converters have a low complexity but are very slow : 2^n-1 steps are required to complete one conversion (where n is the number of bits of the output word). To decrease the conversion time, multi-step architectures exist : 2 or 3

different step sizes are used, the coarse one determining the MSBs and the finer one(s) the LSBs.

Counter ramp converters are well suited for extremely slow varying analog signals. They are not a valid alternative to meet the constraints of section 1.1.

2.2.5 Successive approximation converter :

Successive approximation converters are also known as serial. They are very common and feature medium speed (10-1000 kHz), medium resolution (8-12 bits, some even 14 bits) and fairly small sizes at a low price. They are widely used and well suited for telecommunication and signal processing applications as well as for interfacing to µPs.

Serial converters can be considered as n-section counter ramps. Each step is different in size and each bit (from MSB to LSB) is compared in turn to the analog input. In other words, the process consists of successively trying a 1 in each bit of a D/A decoder starting with the MSB. As each bit is tried, the output of the D/A is compared against the analog input signal. If the D/A output is larger, the 1 is removed from that bit as the process continues and a 1 is tried in the next most significant bit. If the analog signal is larger than the D/A output the 1 remains in that bit. At the end of the process, the digital number in the D/A is the digital equivalent of the analog voltage. n steps are required to encode a n-bit binary value.

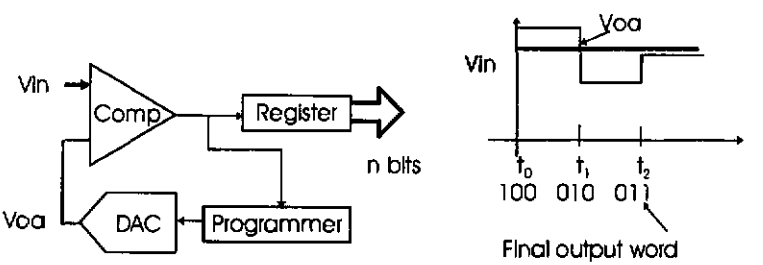


Figure 2.9 : Successive approximation

Algorithmic (or cyclic) successive approximation converters are often used. The conversion process involves n recursive cycles « recirculated » within the same feed-back loop. The corresponding flow graph is given in figure 2.10. These structures have a very small size and do not require a DAC. Nevertheless, a precise doubling amplifier and a comparator with at least 0.5 LSB precision is needed.

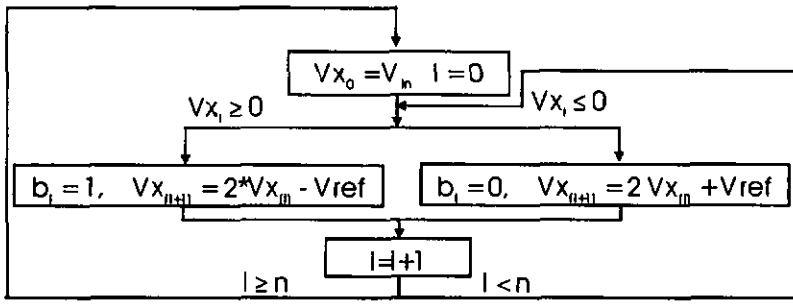


Figure 2.10 : Algorithmic successive approximations converter

Cyclic successive approximation converters are the most serious candidate to meet the requirement of section 1.1.

2.2.6 Parallel converter :

Parallel or flash converters use one analog comparator with fixed reference voltage for every quantization level. They can achieve very fast conversion since only one step is required. The drawback is the complexity of the hardware, whose size is proportional to 2^n (where n is the number of bits). They feature dynamic range of 6-8 bits and sampling frequencies up to a few hundred MHz. Typical applications are fast signal processing such as radar and video.

To increase the dynamic range and decrease the complexity, size and price, half flash (or two step parallel) converters can be used. This is a combination of successive approximation and flash. The converter has two stages in which half of the bits are determined. Two steps are thus required for conversion, but the hardware size is reduced and is proportional to $2^{(1+n/2)}$.

These converters are not suited to reach dynamic range 14-bit while keeping the power consumption low.

2.2.7 Pipelined converter :

All the converters presented so far are real-time i.e. a full A/D conversion is completed during an analog sampling period.

In pipelined converters, the conversion is performed on each sample of analog data during a period of time of two or more samples. A first stage converts the high end bits during the first sample data time and then passes the

residue of the input to the next stage in the pipeline for processing. Then, during the second analog signal sampling period, the first stage converts the high end bits for the next analog sample while the second stage converts the previous sample's residue. The process goes on until the last stage is reached and the conversion is completed. Consequently, the converter's sampling rate is equal to the analog signal sampling rate but the full digital equivalent value is delayed in the pipeline. This delay or latency is often of little importance compared to the conversion speed.

Pipeline converters have two to n stages. The latter can be counter ramp, successive approximation or parallel realizations. These A/D converters can reach very high conversion speeds (GHz) at the price of high complexity, area and power consumption. They are clearly not targeted for the relatively slow sampling frequency of 16 kHz defined in section 1.1.

2.3 MARKET AND ACADEMIC OVERVIEW

The ADC market offers a very wide choice of low dynamic range devices (8-12 bits).16-bit converters are less common but all the major providers offer at least one $\Sigma\Delta$ solution. Higher dynamic ranges are commercialized by specialized companies only. Table 2.1 provides a partial survey of the published or available linear devices at the date of January 1997.

Since all the commercially available devices couldn't be considered (for example Analog Devices has more then 100 different ADCs), choice criteria were defined. Thus the table contains ADCs that fit into the following categories

a) Low dynamic range (≤10 bit) / maximum speed
b) Medium dynamic range (12-16 bits) / medium speed (10-50 kHz)
c) High dynamic range (>16 bits) /maximum speed

Whenever possible, optimized low power devices (according to their providers) were chosen. Category b) is actually of most interest since it corresponds to the target specification given in the introduction (section 1.1). Categories a) and c) represent the boundaries of current technology.

#	Provider	Reference name	Type	Dyn. R. [bit]	Frequ. ksamples/s	Power [mW]	SNR [dB]
1	Acad.	[Fong94]	Algorith.	4	4500	0.5	
2	Harris	HI3304	Flash	4	25000	25	
3	Lin. Tech.	LTC1096	Suc. Ap.	8	33	0.216	
4	ADI	AD775	Pipelin.	8	20000	60	47
5	Harris	HI1175		8	20000	60	
6	Exar	MO8775	Flash	8	20000	85	
7	Burr B.	ILC5540		8	40000	150	
8	Maxim	MAX1154	Flash	8	750000	5500	45
9	Acad.	[Wu94]	Pipelin.	9	33000	180	
10	Acad.	[Kusu93]	Pipelin.	10	20000	30	
11	ADI	AD9050		10	40000	315	53
12	Acad.	[Yotsl93]	Pipelin.	10	50000	900	53
13	ADI	AD7822		12	7.5	0.06	
14	Lin. Tech.	LTC1285		12	7.5	0.48	67
15	Acad.	[Chen95]	Cyclic	12	10	2	
16	Lin. Tech.	LTC1286		12	12.5	1.25	71
17	Harris	HI5813	Suc. Ap.	12	40	3.3	65.1
18	T.I.	TLV2543	Suc. Ap.	12	66	8	
19	Maxim	MAX1241	Suc. Ap.	12	73	3	
20	Acad.	[Wit_93]	Suc. Ap.	12	200	10	70
21	Acad.	[Zhon_94]	Pipelin.	12	1000	60	
22	Nat. Sem.	ADC12062	Flash	12	1000	75	70
23	Com. Lin.	CLC949		12	20000	220	69
24	Nat. Sem.	ADC12L038	Suc. Ap.	13	73	15	73
25	Acad.	[Clin_95]	Pipelin.	13	5000	166	80.1
26	Crystal	CS537_0	ΣΔ	16	20	220	80
27	Acad.	[Mots_87]	ΣΔ	16	48	110	92
28	Motorola	MC145073	ΣΔ	16	48	250	82
29	ADI	AD1380	ΣΔ	16	50	15	
30	ADI	AD976	Suc. Ap.	16	200	100	83
31	Burr B.	ADS7815	Suc. Ap.	16	250	200	84
32	ADI	AD7721	ΣΔ	16	469	150	74
33	Maxim	MAX132	Integrat.	18	0.1	0.3	
34	Burr B.	DDC101	ΣΔ	20	15	170	
35	Crystal	CS5324	ΣΔ	20	32	150	110

Table2.1 : ADC Overview

Figure 2.14 (♦) plots the converter number (first column of table 2.1) versus an « efficiency » factor defined as the product of the sampling frequency and the dynamic range divided by the power consumption [bits*ksamples/s*mW]. To be complete, issues such as normalized chip size, SNR quality (difference between

Ideal and measured values), etc. could also be taken into account (a small size for example would increase the efficiency factor). The resulting graph could lead to completely divergent conclusions. Figure 2.14 however, allows a comparison considering the particular aspects of dynamic range, speed and power consumption.

Generally speaking, as the number of bits increases, ADCs tend to be less « efficient ». This is not a surprise since studies [Dijk94] showed that the power consumption increases with n^2 where n is the dynamic range in bits.

Low dynamic range converters (#1 to 12) are well grouped, with a slight tendency to become less « efficient » as speed increases, except for #1 and #10 that have a considerably higher factor.

The dispersion among the 12-13 bits (#13 to 25) is greater and results are slightly better for faster devices.

Apparently 16-bit ΣΔ converters are more efficient at higher frequencies (#29 and 32) than successive approximation ones (#30 and 31).

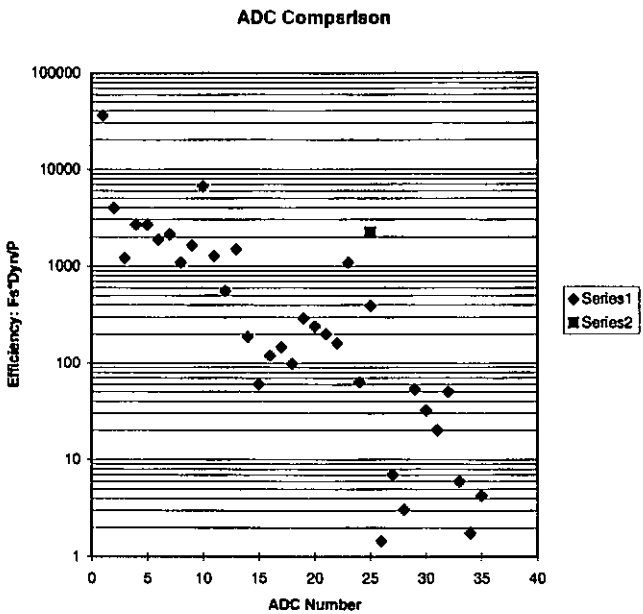


Figure 2.14 : ADC »efficiency » comparison

It is interesting to note that considering the specification of 1.1, the desired efficiency number is 14*16/0.1=2240. Such a device is represented by the black square (■) in figure 2.14. Considering efficiency numbers (triangles) obtained by the commercialized devices, the targeted efficiency is clearly a major improvement!

2.4 REFERENCES

2.4.1 References to paragraphs 2.1 and 2.2.

[Heub96] A. Heubi, P. Balsiger, and F. Pellandini, "Micro Power 13-bits Cyclic RSD A/D Converter", ISLPED'96, Aug. 12-14 1996, Monterey CA, USA, pp 253-257.

[Miel96] J. A. Mielke, « Frequency Domain Testing of ADCs », IEEE Design & Test of Computers, Spring 1996, pp 64-69.

[Teme93] Gabor Temes, « An overview of A/D and D/A Conversion Techniques », lectures notes, Intensive Summer Course on Integrated A/D and D/A Converters, July 5-91993, EPFL, CH.

[Haes94] David F. Hoeschle, « Analog-to-Digital and Digital-to-Analog Conversion Techniques », second edition, Wiley Interscience publication, 1994, ISBN 0-471-57147-4.

[Elec93_1] « l'acquisition de données domine les Géch./s », Electronique Vol.24, Janvier 1993, pp 45-49

[Elec93_2] « Les Convertisseurs A/N et N/A de Précision », Electronique Vol.26, Mars 1993, pp 64-75.

[AD92] Analog Devices, « Data Converter Reference Manual », Vol. II, 1992, pp 2.14-2.23.

[Cand92] James C. Candy and Gabor C. Temes, « Oversampling Delta-Sigma Data Converters : Theory, Design and Simulation », IEEE Circuits and Systems Society, IEEE Press,1992.

[Vale87] Vlodo Valencic, « Convertisseurs A/N à Approximations Successives en Technologie CMOS à Capacités Commutées », Ph. Dissertation #708, 1987, EPFL, CH.

[Anol86] Analog Devices engineering staff, « Analog-Digital Conversion Handbook », Prentice Hall, 1986, ISBN 0-13-032848-0.

[Deck] Michel Decklerck, « Conversion A/D et D/A », unpublished lecture notes, LEG-EPFL, Lousanne, CH.

2.4.2 References to paragraphs 2.3

http://www.analog.com :80/products/index/1.html

http://www.burr-brown.com :80/products/p2.html

http://www.compdist.com :80/comlinear/clc949.html

http://www.semi.harris.com/converters/conva-d.html

http://www.maxim-ic.com

http://www.nsc.com/search.cgi

http://motserv.indirect.com/analog/products1

http://www.exar.com/products/datacq/adcmenu.html

http://www.semiconductors.philips.com :80/philips52.html

http://www.crystal.com :80/pub/dap.html

[Chen95] C.-C. Chen, C.-Y. Wu, J.-J. Cho, « a 1.5V CMOS Current Mode Cyclic Analog-to-Digital Converter with Digital Error Correction », Proceedings of IEEE Int. Symp. on Circuit and Systems, pp 537-540, 1995.

[Clin95] D. W. Cline , P. R. Gray, « A Power Optimized 13-bit 5Msamples/s Pipelined Analog to Dogital Converter in 1.2 um CMOS », Proceeding of CICC, May 1995.

[Zhan94] L. Zhong, T. Sculley, T. Flez, « A 12 8it, 2V Current Mode Pipelined A/D Converter Using a Digital CMOS Process », Proceedings of IEEE Int. Symp. on Circuit and Systems, pp 369-372, London, May1994.

[Dijk94] E. Dijkstra, « Low Power Oversampled A/D Converters », Low-Power/Low-Voltage IC Design, June 27-July 1 1994, EPFL, Lousanne, CH.

[Fong94] K. L. Fong, C. A. T. Salama, « Low Power Current Mode Algorithmic ADC », », Proceedings of IEEE Int. Symp. on Circuist and Systems, pp 473-476, London, May1994.

[Yots93] M. Yotsuyanagi, T. Etoh, K. Hirata, « A 10-bit 50-MHz Pipelined CMAOS A/D Converter with S/H », IEEE Jour. Of Solid-State Circuits, Vol. 28, N0.3, March 1993, pp 292-300.

[Kusu93] K. Kusumata, A. Matsuzawa, K. Murata, « A 10-bit 20 MHz 30 mW Pipelined Inperpolating ADC », IEEE Journal of Solid State, Vol. 28, No. 12, December 1993

[Mats87] Y. Matsuya et al., « A 16-bit Oversampling A/D Conversion Technology Using Triple Integration Noise Shaping », Journ. of Solid-State Circuits, Vol. sc-22, No. 6, December 1987, pp 921-929.

2.4.3 References to academic non linear converters

[Gull95] J. Guilherme and J.E. Franca, « New CMOS Logarithmic A/D Converters Employing Pipeline and Algorithmic Architectures», Proceedings of IEEE Int. Symp. on Circuist and Systems, Vol. 1, pp 529-532, 1995.

[Gull94] J. Guilherme and J.E. Franca, « Digitally Controlled Analog Signal Processing and Conversion Techniques Employing a Logarithmic Building Block », Proceedings of IEEE Int. Symp. on Circuist and Systems, Vol. 5, pp 337-380, London, May1994.

[Kall94] K. Kalliojärvi, J. Kontro, Y. Neuvo, « Novel floating Point A/D and D/A Conversion Methods », Proceedings of IEEE Int. Symp. on Circuist and Systems, pp 1-4, London, May1994.

[Lygo88] J. N. Lygouras, « Nonlinear ADC with Digitally Selectable Quantizing Characteristic », IEEE Trans. on Nuclear Science, Vol. 35, No 5, pp 1088-1091, October 1988.

[Gott78] Gottschalk, « Logarithmic Analog-Digital Converter using Switched Attenuators », Rev. Sci. Instrum. No. 49, pp 200-204, February 1978.

3. Fundamentals of Low Power Design

This chapter provides an overview of some of the finesses used to obtain low power systems.

Digital power consumption is in part due to leakage and short circuit currents though the main contribution comes from the switching power whose value is proportional to the switching frequency, the capacitance load and the square of the power supply. Hence, actions can be taken at either algorithmic, architectural, logic or technological levels to reduce the power consumption by decreasing either the working frequency, the capacitance load or more efficiently the supply voltage. Whatever the action, the performance of the digital module must be maintained. Electronic Design Automation (EDA) tools provide a great help to rapidly achieve valid results.

Things are different in the analog domain: lowering the supply voltage only complicates the design and dedicated structures must be used. Also, the working domain (weak or strong inversion) of the MOS transistors must be carefully chosen. Highly reliable simulation models, such as the EKV (Enz, Krummacher, Vittoz) developed at CSEM and EPFL are mandatory but unfortunately they are only available for a few technologies.

Considering that the targeted chip combines both analog and digital functions, that it is battery supplied and that a Multi-Project-Wafer (MPW) implementation is foreseen, only four technologies are accessible. The final choice goes to EM Microelectronic-Marin SA technologies since EKV models are available.

A short survey of today's primary and secondary cells (batteries) shows that in spite of the progress in capacity (and thus life cycle) and chemical materials within such devices remains an ecological burden and that recycling is still the only solution.

Most of the research and development effort in the area of digital products has been oriented towards increasing the speed and the complexity of a single chip. However, as the density and size increase, difficulty in providing adequate cooling represents a limit and power consumption becomes a main concern. The need for low power devices is aggravated by the increasing demand for portable systems.

Such reduction is not easily achieved and requires optimization at all levels of the design hierarchy. As explained in paragraph 2.1, lowering the power supply voltage has a great impact on power consumption of digital circuits. It has however little effect on the power consumption of analog circuits and rather complicates their implementation since the corresponding reduction of the maximum signal amplitude must be compensated by lowering the noise floor.

Proceedings of Electronic and DSP conferences as well as trade magazines are rich in articles on particular low power optimized library elements, architectures, logic strategies, algorithms, etc. However, the annual « International Symposium on Low Power Electronics and Design » is certainly one of the most focused conferences. The « International Symposium on IC technology systems and applications » also has dedicated sessions. Another good reference is the special issue on low power electronics of the « Proceedings of the IEEE » (volume 83, number 4, April 95).

3.1 DIGITAL DESIGN

There are three major sources of power dissipation in Digital CMOS circuits :

$$P = P_{sw} + P_{sc} + P_l = \alpha_{0\to1} \cdot C_l \cdot V_{dd}^2 \cdot f_{clk} + I_{sc} \cdot V_{dd} + I_l \cdot V_{dd} \qquad (3.1)$$

The first term, P_{sw}, is the switching component where C_l is the load capacitance, f_{clk} the clock frequency and $\alpha_{0->1}$ the node transition activity i.e. the number of power consuming transitions (from 0 to 1) in one clock period. The second term, P_{sc}, is due to the direct path short circuit current I_{sc} which arises when both PMOS and NMOS transistors are active. Finally, P_l, is due to the leakage current I_l which arises from substrate injection and subthreshold effects and is primarily dependent on the silicon technology.

Clearly the best way to reduce the switching power is to lower V_{dd} (quadratic relation !). The next step is to minimize $\alpha_{0->1}$ C_l (also called the effective capacitance C_{eff}) through proper choice of logic function and style, circuit topology, data statistic, sequencing and glitching reduction.

Short circuit power is due to finite and very different rise/fall times at the input of the gates. Ensuring similar rise and fall times will make the short circuit consumption level about 10% of that of the switching power [Chan95]. Note that if the supply voltage was lowered below the sum of the threshold voltages of the transistors, the short circuit current would be eliminated. Other problems such as speed limitations would occur though!

There are two types of leakage current : reverse bias diode and subthreshold. Diode leakage occurs when a transistor is turned off and another transistor charges up or down the drain with respect to the former bulk potential. The resulting current is proportional to the leakage current density set by the technology and the drain area. It is however a small current and results in a fraction of the total power consumption [Raba96], although it can become significant in applications where much of the time is spent in stand-by mode.

Subthreshold leakage occurs when the gate-source voltage is below V_t but has exceeded the weak inversion point and is due to carrier diffusion between source and drain [Raba96]. This current depends on technological parameters. Similar to the reverse bias current it becomes relevant only in stand-by mode.

The sources of power dissipation in CMOS digital circuits have been identified. The next sections describe some of the actions that can be taken at different levels to decrease this power consumption.

3.1.1 Technology optimization

As already mentioned reducing the supply voltage results in a drastic improvement in the switching power consumption. Unfortunately, a speed penalty is paid and the delay significantly increases as V_{dd} approaches the threshold voltage V_t (see figure 3.1 [Chan95]).

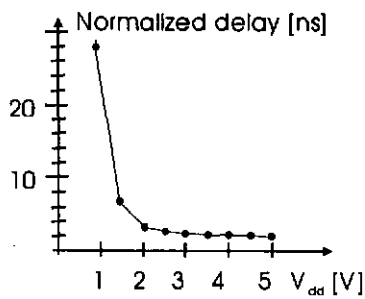


Figure 3.1 : Normalized delay for a typical CMOS gate.

Since the objective is to reduce the power while keeping the throughput of the overall system fixed, the « technical » solution is to reduce the threshold voltage (architecture improvement is another solution and is described in 2.1.3). [Chan95] gives the example of a circuit running at V_{dd} =1.5 V with V_t =1 V which has the same speed performance as another running at V_{dd} =0.9 V with V_t =0.5 V. The limit of V_t reduction is set by the requirement to retain adequate noise margins and reasonable subthreshold currents.

This optimization is performed by the technology engineers of the various foundries. The limit to which the IC designer can optimize V_t is constrained by choosing a technology that will satisfy throughput constraints (considering an eventual particular architecture) while guaranteeing a minimal power supply.

3.1.2 Physical, circuit and logic level optimization

At the layout level, signals with high switching activity (clocks) should be assigned short wires while signals with lower switching could be allowed progressively longer ones. This will result in lower $C_{eff}*f_{clk}$ products.

Transistor sizing should ensure that all delay paths are equalized so that a single critical path doesn't unnecessarily limit the performance of the entire system. The question about uniformly increasing the W/L ratios can be raised. A rule of thumb is that minimum sized devices should be used when the load capacitances is not dominated by interconnect (local busses inside datapath blocks). Buffers with appropriate strength should be employed to drive large capacitance such as those between datapath modules. The IC designer usually has no access to such « deep » sizing since he uses a (hopefully) optimized library which provides all the basic gates and elements. Nevertheless he might be responsible for inserting the proper buffers.

Reducing voltage swing is sometimes used. Because special gates are required (to avoid static power increases), extra parasitic capacitance is obtained and the method is thus only useful on already high capacitance nodes.

Clever logic minimization and technology mapping can reduce the power consumption by orders of 20-25%. The idea is to minimize switching and glitching activity. Logic level power down through gated clocks can be used at the price of some additional control circuitry.

3.1.3 Architecture level optimization

Lowering the supply voltage increases the delay. Architecture techniques must thus be used to compensate for the reduced speed.

Let us assume a reference architecture made of a module M, powered at V_{ref} and working at maximum speed i.e. the clock period $1/CK_{ref}$ is equal to the worst case delay. The voltage cannot be lowered since it would result in loss of throughput.

Using two modules in parallel is an improvement. Indeed, the throughput is maintained but each module works at $CK_{ref}/2$. The supply voltage can thus be lowered until the delay has doubled. The hardware is duplicated and some extra routing is required leading to an effective capacitance that is slightly more than double compare to the reference architecture. The balance is :

$$V_{par} < V_{ref} \quad C_{eff_par} \geq 2 \cdot C_{eff} \quad CK_{par} = 0.5 \cdot CK_{ref} \tag{3.2}$$

Another solution is to pipeline the operations of the module by inserting some registers. Again, the delay of each operation can be lengthened until the most critical one equals the throughput time and consequently the supply voltage can be reduced. Each operation is still performed at the original rate but because of the additional registers the area is slightly increased. Thus :

$$V_{pip} < V_{ref} \quad C_{eff_pip} \geq C_{eff} \quad CK_{pip} = CK_{ref} \tag{3.3}$$

From 3.2 and 3.3 the products C_{eff_par}*CK_{par} and C_{eff_pip}*Ck_{pip} are slightly higher than C_{eff_ref}*CK_{ref}. Nevertheless, the quadratic dependence on V_{dd} still ensure a substantial improvement in the final power consumption.

Pipeline and parallelism can be exploited together. At a certain point though, the overhead circuitry will dominate and no further power consumption decrease will be obtained.

In applications where very tight area specifications are imposed and provided that the throughput allows it, time multiplexed operations are used. The idea is to have modules (for example an ALU) that perform n (n>1) operations within one clock cycle (throughput rate). The approach seems favorable but actually results in higher switching activity than fully parallel solutions. Indeed, in the parallel case, each module switches once per clock cycle but only if the inputs have changed. In the time multiplexed solution however, operations are sequential i.e. operation 1 to n take place within one clock cycle. Thus even when the inputs are not

changing, switching occurs; the results that have not changed (compared to the previous clock cycle) must be computed again.

Another architecture level optimization (though it could be classified as algorithmic level as well) is the choice of number representation. In most DSP applications two's complement representation is used and addition and subtraction are easily performed. However, sign extension causes the MSBs to switch when a signal transitions from positive to negative or vice versa. Depending on the signal statistics (especially when the signal is small compared to the dynamic range) this can result in very high switching activity. One approach is thus to used sign-magnitude representation. It can be demanstrated [Chan95] that this leads to a lower number of transitions. Addition and subtraction are consequently more difficult to handle. Sign-magnitude should thus be used for large buses where the overhead for converting back and forth to 2's complement is not significant compared to the overall switching activity balance.

Finally, glitching reduction is handled at architectural level by balancing all signal paths and reducing logic depth as much as possible.

The designer has to trade off throughput and area constraints, degree of resource sharing (time multiplex), parallel and pipelined architectures and number representation to ensure minimum power supply and switching activity resulting in overall reduced switching power.

3.1.4 Algorithmic level optimization

The choice of an algorithm is a highly leveraged decision in meeting the power consumption constraint. The ability for the algorithm to be implemented in parallel is critical along with the complexity of the computation.

Algorithm transformation such as loop unrolling, algebraic transformations and constant propagation can be used to modify the flow graph so that either pipeline or parallelism can be applied.

The computation complexity can sometimes be reduced by taking into account the correlation between neighboring samples (eg, vector quantization). It is also very common to replace constant multiplication functions by shift and add operations.

Which of high-level algorithm design, architecture or technology optimization has the greater impact on power reduction is still an open issue. Animated panel session held at ISLPE'96 and ICSPAT'96 nevertheless stressed that optimization has

to be performed at each level. To reach this goal targeted EDA tools are mandatary !

3.1.5 Low power design EDA tools

Within the last two years nearly a dazen different commercial tools have arrived an the market to address low power. This section assess the situation at the end of 1996 [Caud96].

Transistor level estimations are most accurate but they require a very precise circuit description and can only be performed close to the final stage of development. These commercial tools are based on a simulation of the transistor netlist that tries to be as close to SPICE as possible, but faster. As a result good accuracy is reached at the price of a long estimation time. Only one tool (Epic's AMPS)is presently available for optimization at this level.

Gate level estimations can be performed earlier in the design cycle, are faster but less accurate. They help determine which gates and wires should be the focus for power reduction. A power optimization tool is available and claims reductions of 10 to 15%.

HDL and RTL estimations are too crude to provide absolute accuracy. Nevertheless, they are very useful as relative indicators to allow evaluation of design tradeoffs. The designer uses these indicators to determine which blocks and signals he/she should focus on, tries different solutions (architectures) and gets quick feedback in measuring the effectiveness of those redesigns.

While algorithmic, behavioral or system level decision have a profound impact on the power consumption of a system no commercial tool at this level is available. Consequently, the algorithmic and behavioral specifications are often frozen early in the design process with no verification means. This reduces the power minimization space in the lower levels where tools are available.

Academic research on targeted low power design tools is also reported. In particular, at IMT, a tool has been developed by Alexandre Heubi to very efficiently implement DSP algorithms on parallel architectures. Applied to a speech processing algorithm, the tools results in a 23x improvement on the MIPS number (in comparison with a TMS32C5x software implementation) and a 21x improvement in power consumption.

3.2 ANALOG DESIGN

Many of the problems and solutions encountered in micro-power analog design circuits are directly related to the properties of the MOS transistor. The next section is thus dedicated to its overview. Active elements (Operational Transconductance Amplifier) and analog switches are then considered as well as passive elements such o resistors and capacitors. Finally effects of power supply decrease are evaluated. A short paragraph about simulation models concludes the topic.

3.2.1 MOS Transistor

The drain current I_D is a function of the drain, gate and source voltages V_D, V_G and V_S (all referred to the local substrate). This function depends on the operating mode of the transistor namely Weak Inversion (WI), Middle Inversion (MI) and Strong Inversion (SI).

In weak inversion the drain current is given by equation 3.4.

$$I_D = I_{D0} \cdot e^{V_G/nU_T} \cdot (e^{-V_S/U_T} - e^{-V_D/U_T}) \tag{3.4}$$

where:

$$I_{D0} \propto W/L \cdot e^{-V_T/nU_T} \quad n = 1 + C_d/C_{ox} \quad V_T = V_{T0} + (n-1)V_S \tag{3.5}$$

Equation 3.4 is thus valid (condition for WI) when:

$$V_G - V_S < V_T - some U_T \quad \text{or} \quad I_D << I_{lim} = \mu C_{ox} \cdot W/L \tag{3.6}$$

Two different modes arise in strong inversion. The linear mode applies when $V_{DS} < (V_{GS}-U_T)/n$ and the drain current is:

$$I_D = \beta \left((V_{GS} - U_T) \cdot V_{DS} - \frac{n}{2} \cdot V_{DS}^2\right) \tag{3.7}$$

When $I_D >> I_{lim}$ or $V_{DS} > (V_{GS}-U_T)/n$ the saturation model applies and the drain current is:

$$I_D = \beta/2n (V_{GS} - U_T)^2 \cdot (1 + V_{DS} - V_{Dsat}/V_{early} \cdot L) \tag{3.8}$$

A small signal model is needed to depict the AC behavior of analog circuits.

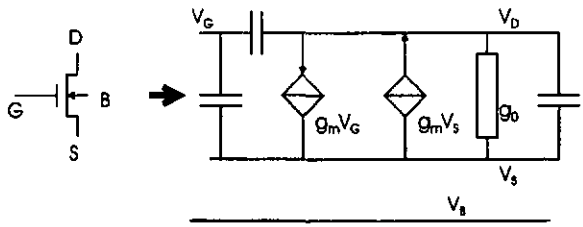


Figure 3.2 : Equivalent AC model.

$$\Delta I_D = \underbrace{\frac{\delta I_D}{\delta V_G}}_{g_m} \cdot \Delta V_G + \underbrace{\frac{\delta I_D}{\delta V_S}}_{g_{ms}} \cdot \Delta V_S + \underbrace{\frac{\delta I_D}{\delta V_D}}_{g_0} \cdot \Delta V_D \qquad (3.9)$$

Applying 3.9 to 3.4 3.7 and 3.8 results in table 3.1

	Weak inversion	Strong inversion	
		Linear	Saturation
g_m	I/nU_t	$\beta(V_D-V_S)$	$(2\beta I/n)^2$
g_{ms}	I/U_t	$n\,g_m$	$n\,g_m$
g_0	$I/V_{Early}L$	$\beta(V_{gs}-V_t)$	$I/V_{Early}L$

Table 3.1: AC model

Noise models are mandatory and the equivalent input noise of the transistor gate is given by

$$Vn^2 = k/W \cdot L \cdot f + 4 \cdot kT/g_m \cdot \alpha \qquad SI{:}\alpha = 2n/3 \quad WI{:}\alpha = n/2 \qquad (3.10)$$

These models suffer from limitations (short channel effects, high frequency, carrier recollection...) but are precise enough to draw interesting remarks. Indeed, compared to SI, WI provides:

- minimum value of drain to source saturation voltage, which helps keep the to peak to peak signal amplitudes close to the power supply rails.
- maximum g_m as well as bandwidth g_m/C and minimum input noise ($1/g_m$) for a given C and limited current.
- maximum voltage gain g_m/I_D, which helps simplify OTA design

A general conclusion from the above remarks is that differential pairs in OTA will preferably be PMOS (kn≅10kp) and work in WI while the transistors used for current mirrors will work in SI.

3.2.2 OTA

Operational Transconductance Amplifiers (OTA's) are needed to provide amplification and to drive the capacitors. Different structures are possible: class A (or simple) OTA's which use fixed currents, class AB (or adaptive) which have current that varies with the input signal, and dynamic ones which are controlled by a clock.

Hae-Seung Lee (Lee94) provided a very interesting comparison between various OTA schemes from all three above types. A class A « 2-stage» OTA has the highest Figure Of Merit (FOM, see Lee94 for definition) and its schematic is given in figure 3.3. FOM indicates how well an OTA performs considering aspects such as noise excess factor, output swing, supply voltage as well as input and output current ratio.

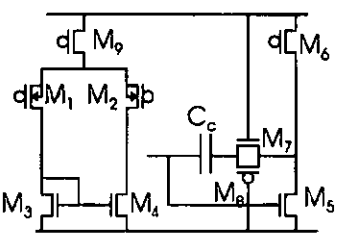


Figure 3.3 : Class A 2-stage OTA

M_1 and M_2, the differential pair, work in weak inversion while all the other transistor work in SI. M_7, M_8 and C_C are necessary to stabilize the OTA (pole splitting). Dc gain is by equation 3.11 and values of 60-80 dB can be obtained.

$$GCD = \frac{g_{m1} \cdot g_{m3}}{(g_{01} + g_{03}) \cdot (g_{05} + g_{05})} \tag{3.11}$$

The equivalent input noise is mostly due to the first stage and thus the spectral noise density is :

$$S_{vn}^{\ 2} = \frac{4KT}{g_{m1}} \left[n_1 + \frac{4}{3} \cdot \frac{n_3 g_{m3}}{g_{m1}} \right] \tag{3.12}$$

The output swing is only limited by the saturation voltages of M_6 and M_5.

3.2.3 Analog Switches

Analog switches are critical in two aspects: their conductance and the charge injection they provide.

A usual implementation is to have two complementary switches which conduct by connecting their gate to the positive and negative power rail. Conduction in SI is ensured when $V_{DS}>(V_{GS}-U_t)/n$. Figure 3.4 shows the conductance g of both N and P transistors as V_A increases.

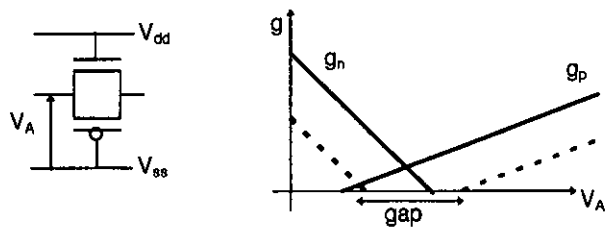


Figure 3.4 : Analog switches

If the supply voltage decreases below a critical value, a gap appears and conduction is lost! A solution to avoid this gap is to use a clock voltage multiplier (Vitt94).

Charge injection (or clock feed-through) occurs each time a switch opens. The relative voltage error $\Delta V/V$ depends on the transistor size and the supply voltage but not on the capacitor value. As the supply decreases, the relative error increases.

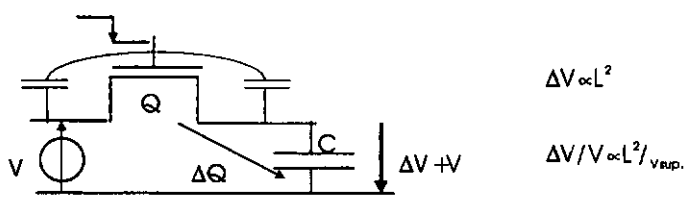


Figure 3.5 : Clock Feed-through

Clock feed-through thus results in offset voltages on the various capacitors. Some solutions (dummy switches for example) reduce the amount of this offset.

The ultimate solution however is to use an analog scheme or a function which is insensitive to offset or where the offset can be easily compensated.

3.2.4 Resistor and capacitors

Generally speaking, resistors are only used in combination with capacitors to obtain a filtering function. Even when the process provides dedicated layers such as high resistance poly, large resistance requires large area.

Capacitors are widely used. Their ultimate size depends on the minimum quantization step size. The noise level on C results from sampling the thermal noise. In a n-bit ADC, this level must be definitely smaller then a LSB and thus equation 3.13 must be satisfied.

$$KT/LSB^2 < C \qquad LSB = 2^{n-1} \cdot V_{dd} \tag{3.13}$$

LSB is proportional to the supply voltage, so as the supply decreases, C increases in a quadratic way. However, bigger capacitors result in longer load times and require higher currents to maintain a fixed slew rate, which means "bigger" OTAs as well. The parasitic capacitors are proportional to C and contribute to further increases in current demand and load time.

3.2.5 Low voltage supply

Eric Vittoz in [Vitt94] analyzes the limitations and effects of lowering the supply voltage in analog design. As mentioned in the former sections, some of the limiting factors are :

- Noise and precision in current (WI)
- Speed of transistor
- Threshold voltage
- Conductivity of analog switches
- Charge injection

The effect on the power consumption can be quantified. Let's assume an analog circuit with a given S/N and frequency and a scaling of the supply voltage V_s to V_s' where $V_s'=V_s/K$. In an ideal case, $S/N=V_s^2/KTC$ and $f \propto g_m/C$. Thus $f*S/N \propto g_m * V_s^2$. Since V_s is scaled, equation 3.14 must be satisfied.

$$g' = K^2 \cdot g \qquad (3.14)$$

In weak inversion, $g_m = I/nU_T$ which results in $I' = K^2 * I$ and thus $P' = K * P$.

In strong inversion, $g_m = (2\beta I/n)^{0.5} = 2I/nV_p$ where V_p is the pinch off voltage. V_p is also scaled down and thus to satisfy 3.14, $I' = K * I$ and $P' = P$.

The maximum frequency is proportional to V_p / L^2. Thus, for a fixed process $f'_{max} = f_{max}/K$.

The conclusion is straightforward ! Reducing the supply voltage does not help in reducing the power consumption of analog circuits !

3.2.6 Simulation models

As stated, problems and solutions in analog design are closely related to the MOS transistor. Furthermore, to obtain power optimized implementations, careful simulation must first be performed. As a consequence, the models used by the simulator must be as close as possible to the reality of weak-inversion, middle inversion and strong inversion.

SPICE1 is not applicable since it doesn't take into account the weak inversion mode. SPICE2 and SPICE3 are better, though they present a discontinuity in the transition between WI and SI. This is also the case with BSIM1.

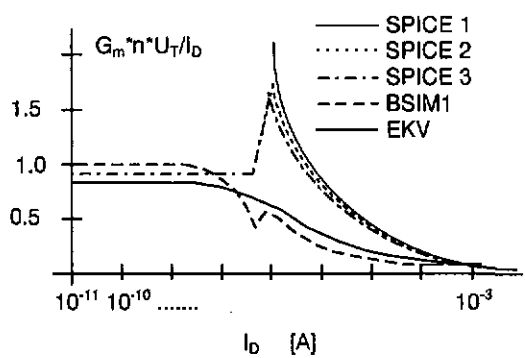


Figure 3.6 : Simulation model comparison

As shown in figure 3.6, the EKV (CSEM) model is continuous and is well suited to simulating circuits combining transistors in WI and SI. BSIM3 (not represented in

figure 3.6) is also an alternative although its complexity is much greater while providing slightly less accurate results [Eny97].

3.3 TECHNOLOGIES

Grossly simplified, the MOS transistor works on the principle of modifying the electric field in the substrate under the gate in such a way as to control the flow of current between the source and the drain electrodes. In a smaller transistor a given electric field pattern can be achieved with reduced voltage (provided the impurity doping concentration has been adjusted). Thus scaling improves the density (smaller devices and wiring), the speed (reduction of the capacitance while keeping the transconductance more or less constant) which, combined with the lower voltage supply, results in an overall power dissipation reduction.

The task of the technology engineer is to optimize all the scaling factors, doping concentration, threshold voltages etc. to get the best trade off between desired characteristics (speed, etc.) and undesired ones (off current, etc.) [Cao95], [Ma96], [Carl96], [Chat96]. The problem is complicated by the fact that analog and digital technologies are diverging with respect to voltage, current and size optimization. Although A/D and D/A only require a small fraction of the transistors needed in a micro system chip, additional technology and circuit design complexity arises from combining analog with digital on the same silicon die.

Silicon On Insulator (SOI) has recently received a great deal of attention because of its high potential for low power applications. Its desirable characteristics are reduced device capacitance, higher transconductance at given voltage, better and denser isolation (thus better analog/digital decoupling), smaller threshold voltage (off current) and finally better scalability. The promise of SOI is impressive, though for widespread use, material quality and cost as well as circuit models, design rules etc. must be greatly improved.

These « philosophical » considerations do not interest the IC designer who has to choose the cheapest and most suited technology. In this case, since a mixed mode chip is foreseen, the technology must provide precise simulation models and altered layers for capacitor design. Since minimum size is on issue, a low lambda (minimum transistor size) technology is preferable. The device will be operated by a 1.3 V battery so a technology which has been characterized for this voltage should be used. Speed limit is not a constraint since the digital part of the A/D will work at the sampling frequency. If an ASIC containing A/D and D/A converters and DSP core had to be integrated, the speed would become a major factor as well.

With the technology specifications in hand, the designer can now browse the available technologies and make his choice. Based on a technology survey issued in June 96 and considering the particular constraints of this work (mixed analog/digital design, Multiple Project Wafer, price etc.) the choice is limited to.... 4 technologies !

AMS 0.8 and 1.2 µm CMOS mixed-mode technologies are characterized for 2.2 to 5.5 V supply voltage. They have two poly layers to form the capacitors and two metal layers. Apparently some customers have performed characterization at lower supply voltages. Typical Vt are 0.85 and -0.75 for the NMOS and PMOS transistor in the 0.8 technology and 0.7 and -0.7 for the 1.2 µm.

EM Microelectronic-Marin SA ALP1 and ALP2 µm Low Voltage CMOS mixed-mode technologies are characterized for a power supply as low as 1V. They also provide two poly and two metal layers. Typical Vt are 0.6 and -0.67 for the NMOS and PMOS transistor in the 2 µm technology and 0.63 and -0.73 for the 1 µm. EKV simulation models are available for both technologies.

Both foundries provide MPW (Multi Project Wafer) passibilities. For geographical reasons EM Microelectronic-Marin SA technologies were chosen.

3.4 BATTERIES

Progress in battery technology is tied to that in portable electronics. Devices with higher energy density, longer shelf life and freedom from leakage have thus been investigated. However, the time scale for improvement is long compared to the « doubling time » for microelectronics. Ultimate size and energy are limited by system chemistry.

Energy and power per unit volume are usually the critical characteristics though the energy delivered by a specific battery depends upon the rate at which the power is withdrawn

Low power portable applications cover consumption ranges of a few micro watt for watches to 10-20 W by notebook computers.

3.4.1 Primary cells

The majority of standard size batteries (D, C, AA etc.) used worldwide are still zinc-carbon. They are cheap and well suited for most consumer applications. The performance of these batteries is not likely to improve. Actually, as environmental

pressures force some of their components to be removed decreased performance can be expected.

In Europe and America alkaline batteries are in fact more common than zinc-air. The elimination of mercury has been a major step in their growing use and success among environmentally sensitive populations. A new manufacturing method allowed a 15-20% increase in capacity in 97. Some minor changes might gain a few extra % but major improvement is not expected.

Typical power densities are 150 mW/L for zinc-air and 250 mW/l for alkaline batteries.

Button cells have seen little performance changes. Table 3.2 [Powe95] gives the characteristics for 11.6 mm diameter by 5.4 mm high button cells realized in various systems. With the exception of zinc-air cells, all these systems will probably give way to lithium batteries.

Zinc-air has the highest energy density of any commercialized system, provided the air access is limited to prevent leakage and/or short life. Densities up to 1200 Wh/l are available in capacities from 50 to 6600 mAh. The main applications are hearing aids and pagers.

System	Operating Volt.	Capacity [mAh]
Zn-Air	1.3-1.2	550
Zn-HgO	1.35	220-275
Zn-Ag2O	1.55	175-190
Zn-MnO2	1.25	145

Table 3.2 : Primary button cells

Primary lithium cells are becoming common place as new consumer products are designed around them (3V operation). These cells have excellent characteristics namely high energy density, very good shelf and operating life. The most widely produced and used lithium battery is the lithium-manganese dioxide system which provide capacities from 10 to 10'000 mAh/cell. It is interesting to note that these cells are exempt from environmental regulation but are considered hazardous material from the standpoint of flammability.

3.4.2 Secondary cells

This market is experiencing a near 20% growth rate fueled by the explosion in cellular phones, portable computers, camcorders and entertainment devices which all require more power than can be provided by primaries.

Sealed lead acid cells offer 1-2Ah in a size close to a pack of chewing gum. Their main advantage is low initial cost, low self discharge and good rate capability.

The mainstay of small secondary batteries is however the Nickel-Cadmium system. Major improvements in energy density resulted in capacities of 800-850 mAh for AA cells and devices with more then 1000 mAh should be available soon. Ni-Cd systems are capable of very high discharge rates and are tolerant to over charge and discharge which makes them easy to use. However they suffer from memory effect. Because of the cadmium, serious effort to collect and recycle used cells had to be undertaken.

Nickel-Metal Hybrids were introduced in the 90's and are winning market share over Ni-Cd. They do not suffer from memory effect and since no cadmium is used, these cells are more "environment-friendly". Furthermore, the actual capacity is already 1200 mAh (1600 mAh devices are foreseen) and the operating voltage is identical to that of Ni-Cd.

Lithium Ion secondary batteries are used in camcorders, computers and cellular phones. They offer high energy density and since they operate at 3.6 V they can replace 3 Ni-Cd or NiMH. Since strict control of charge and discharge is required both for safety and long life cycle, smart charge chips must be used. Energy density up to 360 Wh/l are projected.

Other secondary cell types are Lithium polymer Electrode, which only operate at high temperatures (>60 °C) and Zinc Manganese dioxide, which are low cost and low density and offer no real benefit compared to Ni-Cd and NiMH. Finally, secondary Zinc-air for portable computers have been announced. The energy density is about the same as NiMH though their weight is much higher. Table 3.3 [Powe95] gives nominal characteristics of AA size secondary batteries (C rate is a current equal in amperes to the nominal ampere-hour capacity of the battery).

System	Volts	mAh	Rate	Wh/l	Wh/kg	Cycles	Loss/mo
Ni-Cd	1.2	1000	10C	150	60	1000	15%
Ni-MH	1.2	1200	2C	175	65	500	20%
Li Ion	3.6	500	C	225	90	1200	8%

Table 3.3 : Nominal characteristics of AA secondary cells

Batteries and the environment is an area of much concern. Regulations differ from one country to another though, from an environmental point of view, all used batteries should be collected and recycled. Indeed, even mercury cadmium etc. free batteries contain lead and or other metals which cannot be eliminated easily [Dupg89]. All battery operated devices should thus ensure easy

removal and provide information for proper disposal. Good collecting and recycling programs should also be set out. Unfortunately, this might only become feasible through economical pressures....

3.5 REFERENCES

[Eny97] C. Enz et al., « The EKV MOS transistor Model: from theory to practice », Electronics Laboratories, EPFL-Lausanne, June 25 1997

[Raba96] Jan Rabaey, M. Pedram « Low Power Design Methodologies », Kluwer, NY, 1996

[Ma96] J. Ma, H.-B. Laing et al, « A Graded-Channel MOS (GCMOS) VLSI Technology for Low Power DSP Applications », Proceedings of ISLPE'96, Monterey, CA, August 96, pp 129-132

[Carl96] L. R. Carley, D. F. Guillou, S. Santhanam, « Fabrication and Performance of Mesa Interconnect », Proceedings of ISLPE'96, Monterey, CA, August 96, pp 133-137

[Chat96] A. Chatterjee, M. Nandakumar, I.-C. Chen, « An Investigation of the Impact of Technology Scaling on Power Wasted as Short Circuit Current in Low Voltage Static CMOS Circuits », Proceedings of ISLPE'96, Monterey, CA, August 96, pp 145-150.

[Coud96] O. Coudert, R. Haddad, K. Keutzer, « What is the State of the Art in Commercial EDA Tools for Low Power ? », proceedings of ISLPED'96, August 12-14 1996, Monterey, Ca, USA.

[Chan95] A. P. Chanrakasan, R. W. Brodersen, « Minimizing Power Consumption in Digital CMOS Circuits », Proceedings of the IEEE, Vol. 83, No. 4, pp 498-523, April 1995

[Powe95] R. A. Powers, « Batteries for Low Power Electronics », Proceeding of the IEEE, Vol. 83, No. 4, pp 687693, April 1995.

[Stor95] J. M. C. Stork, « Technology Leverage for Ultra Low Power Information Systems », Proceedings of the IEEE, Vol. 83, No. 4, pp 607-618, April 1995

[Dava95] B. davari, R. H. Dennard, G. G. Shahidi, « CMOS Scaling for High Performance and Low Power : The Next 10 Years », Proceedings of the IEEE, Vol. 83, No. 4, pp 595-605, April 1995

[Cao95] M. Cao, H. Stork, « Optimization of Low Power Quartermicron MOSFET », Proceedings of ISLPE'95, San Jose, CA, October 95, pp 84-85.

[Raba94] Jan Rabaey, « Low Power Circuit Design, Low Power Design at Architectural and System Level, Computer Aided Design for Low Power Systems», Workshop on Low Power/Low-Voltage IC Design, EPFL, Lausanne, CH, 1994

[Pigu94] C. Piguet, « Ultra Low Power Design », Workshop on Low Power/Low-Voltage IC Design, EPFL, Lausanne, CH, 1994

[Vitt94] E. Vittoz, Lecture notes, Workshop on Low Power/Low-Voltoge IC Design, EPFL, Lausanne, CH, 1994.

[Degr94] M. Degrauwe, Low Power Sensor Interface Circuits, Workshop on Low Power/Low-Voltage IC Design, EPFL, Lausanne, CH, 1994.

[Decl94] M. Declercq, M. Degrauwe, Low Power/Low Voltage Ic Design: an Overview, Workshop on Low Power/Low-Voltage IC Design, EPFL, Lousanne, CH, 1994.

[Lee94] H.-S. Lee, CMOS Opeartional Amplifier for Low Power/Low Voltage Circuits, Workshop on Low Power/Low-Voltage IC Design, EPFL, Lousanne, CH, 1994.

[Enz94] C. Enz, « Device Modeling for Low Voltage and Low Current Circuits », Workshop on Low Power/Low-Voltoge IC Design, EPFL, Lousanne, CH, 1994.

[Leac94] Leach, « Fundamentals of Low Noise Analog Circuit Design », IEEE Proceedings, Vol. 82, No. 10, October 1994.

[WWF89] WWF, « Les Piles », Eco-conseil No.5, February 1989.

[Dupg89] « Les Déchets Urbains : Problème Globol » , Comission Nationale Suisse pour l'UNESCO, 1989.

[Tevk88] « Technishe Eigenshaten von Kleinebatterien », Infel-Info No. 1, March 1988.

[Gdd88] « Gestion des Déchets : protection de l'environnement en Suisse », Februory 1988.

[Caho] N. C. Cahoon, G. W. Heise, « The primory battery », Ed J. Wiley, New York.

[Degra] M. Degrauwe, Lecture notes, Design of Analog CMOS IC's, Institute of Microtechnology, University of Neuchâtel.

3.6 ANNEX: LIST OF SYMBOLS

β: Defined as $\beta_s W/L$ where $\beta_s = \mu C_{ox}$

C_d: Drain capacitor

C_{ox}: Oxyde capacitor

g_m: Transconductance

g_0: Output conductance

I_{D0}: Quiescent drain current

K: Boltzmann constant ($1.381\ 10^{-23}$ J/K)

L: Length of transistor

μ: Carriers mobility

n: slope factor

q: Elementary charge ($1.602\ 10^{-19}$ C)

Svin: Noise spectral density

T: Absolute temperature

U_t: Thermal voltage. Defined as KT/q

V_{early}: Early voltage (artifact for output conductance)

V_D: Drain voltage referred to local substrate

V_G: Gate voltage referred to local substrate

V_S: Source voltage referred to local substrate

V_{DSat}: Saturation voltage (if $V_{DS} > V_{DSat}$ then strong inversion)

V_t: Threshold voltage

W: Width of transistor

4. Floating-point conversion

This chapter gets into the heart of the subject and explains, from a signal processing point of view, the effect of non-absolute conversion. The original feed back floating point approach developed by A. Schaub back in 1992 is presented as well as its limitations. The latter are overcome by the enhanced feed back floating point concept.

Both feed forward and enhanced feed back conversion concepts are detailed. The enhanced feed back approach is well suited for predictable input signals while the feed forwards approach is useful for random-type signals. Audio signals are fairly predictable (pitch, spectral density) and two adaptation "strategies" are proposed for enhanced feed back conversion: the first is inspired from Jayant's work on speech coding while the second is based on results coming from recursive simulations using a pool of input signals.

Numerous informal listening tests are performed to define targeted converters for audio applications. Once the system architecture for both enhanced feed back and feed forward converter is established, simulations furnish the ideal characteristics of both devices.

Non audio signals are considered as well and the trade offs between feed forward and enhanced conversion are explained. Finally, some design considerations conclude this chapter.

As mentioned in the first two chapters, absolute and non-absolute converters can be distinguished. The latter present a quantization step that varies along the dynamic range. Normally, the step size increases with the signal amplitude, however, this increase can be linear, as in the ideal relative non-absolute converter, or by step as in the floating point approach, or even by 'burst' (at certain amplitude only), as in the mixed RSD converter described in chapter 5. The benefit of non-absolute devices versus absolute ones is that a reduced power consumption is obtained.

Whatever the particular increase mode, the effect is the same: in small amplitude signals the quantization is fine while in larger amplitude the

quantization gets coarser and the small "details" are lost. Non-absolute devices are thus useful either in the case of non-absolute signals delivered by the sensor or in the case of digital signal processing algorithm that are not sensitive to small amplitude variations occurring at high amplitude.

Let us consider an analog signal composed of two sine waves. The first is at low frequency and has a close to maximum dynamic range amplitude. The second has a higher frequency but its amplitude is much smaller, only a few minimum quantization steps. If the digital processing algorithm must extract the high frequency component of this signal, an absolute converter is mandatory. Indeed, if a non-absolute device is used, the information about the high frequency component is lost at large amplitude. On the other hand, if the processing algorithm is interested in the high frequency component only as long as its amplitude is bigger than a certain level below the slow varying signal, the non-absolute device is sufficient.

Signal quantization results, in the frequency domain, in the adjunction of spectral lines at harmonic and non harmonic frequencies. These lines "translate" the quantization noise and their amplitude increases as the quantization get coarser. Hence, in the case of non-absolute converter, the spectral line "configuration" depends on the input signal amplitude level (since the quantization step depends on the signal amplitude). According to the subsequent algorithm this can be a major drawback.

Assuming that a considered application tolerates a non-absolute converter, the best choice would be to use the relative precision device. As explained in chapter 2 (see 2.1.2), such a device is not physically realizable. However, the ideal relative precision characteristic can be approximated through a floating point approach.

In floating point conversion, the idea is to scale (exponent) the input signal in such a way that it fits well into the fixed conversion range of a coarse quantizer (mantisse). The floating point conversion is so called because the internal exponent/mantisse representation is equivalent to a floating point representation. However, the output of the quantizer is a fixed point format word resulting from the combination of the mantissa and exponent value.

Feed back and feed forward floating point conversion must be distinguished. In feed back adaptation (figure 4.1), the previous coarse quantizer output is used to determine, via the adaptation logic, which scaling factor should be applied to the next input sample. In other words, a « 1 sample prediction » is achieved and the scaling set accordingly. On the other hand, feed forward adaptation (figure 4.2) uses the input sample, on which the scaling is then applied, to evaluate the scaling factor. In both types, the coarse quantizer output is fed to a « formatting »

back-end which reconstructs the "fixed point" final word, taking into account the digital word and the gain which was applied.

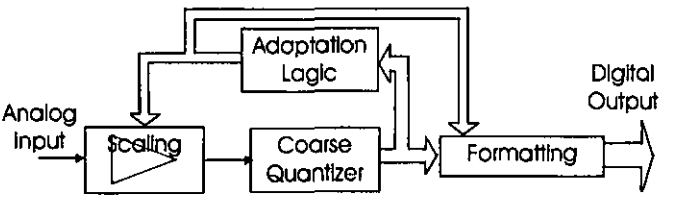


Figure 4.1 : Feed-back floating point ADC

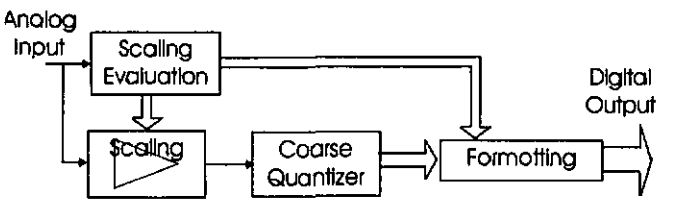


Figure 4.2 : Feed-forwards floating point ADC

The primary goal of this research is to develop floating point converters for audio applications. To cover sound signals ranging from quiet sleeping room to pneumatic hammer, a dynamic of 14 bits is necessary. However, because of the masking effects occurring in the human ear a limited resolution is tolerated. For such audio devices, the ultimate quality criteria is informal listening test.

The first article about feed-back floating point converters was published in 1992 by A. Schaub [Scha92]. As explained in section 4.1, his converter is not suited for demanding audio applications. Nevertheless, the feed back principle is interesting and section 4.2 explains how it can be enhanced to meet the excellent perceived audio quality constraint. Section 4.3 deals with feed forward conversion while section 4.4 extends the discussion to look at other applications. Finally, section 4.5 considers implementation perspectives and helps determine where the design effort is particularly important.

4.1 ORIGINAL FEED BACK FLAOTING POINT CONVERTER

Arthur Schoub [Scho92] proposed a feed back floating point conversion concept back in 1992. His solution was aimed at speech coding and hence the perceived quality constraint was rather loose.

Figure 4.3 shows the front end stage of A. Schaub's converter. The back-end is made of look-up tables and a shifter and is used to format the output into the fixed point word.

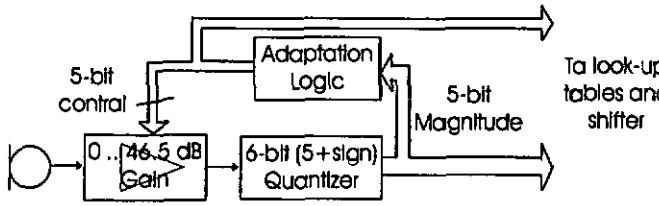


Figure 4.3 : A Schaub's floating point converter

The analog microphone signal is fed to a controlled amplifier that multiplies the analag signal according to its gain setting. This programmable amplifier can have gain values ranging from 0 to 46.5 dB in steps of 1.5 dB, hence 32 different gain vales. The gain value is set by the 5-bit control word which is provided by the adaptation logic. The amplified signal is applied to a 6-bit quantizer. The five bits representing the magnitude of the quantized sample serve as input to the adaptation logic. The later determines, according to table 4.1, the gain increment/decrement that must be applied to update the gain of the programmable amplifier. The 6-bit quantized output as well as the gain setting are loaded into the back end which reconstructs the fixed point 14-bit. This is performed by reference to look-up tables and subsequent shifting operations.

Magnitude Codeword	Gain inc. / decr. (dB)
0..15	+ 1.5
16, 17	-1.5
18, 19	-3
20..23	-4.5
24 .. 27	-6
28, 29	-7.5
30, 31	-9

Table 4.1 : Gain adaptation for 6-bit quantizer

The adaptation table is based on N. S. Jayant's paper [Jaya73] on adaptive quantization of PCM signals. Basically, the problem assessed by Jayant is identical to that of the floating point conversion: based on a previous sample value and to scale the input signal so that it fits a coarse quantizers, the best multipliers must be found. Jayant's multipliers are computed for quantizer of up to 5 bits and range from 0.85 to 2.6. Schaub used Jayant's multipliers but restricted their values to powers of $2^{1/4}$ (corresponding to approximately 1.5 dB steps) and obtained table 4.1. Hence, when the magnitude of the coarse quantizer output is

less than 15, the gain value of the controlled amplifier is increased by 1.5 dB and the new gain value is used for the next sample. On the other hand, if the magnitude of the coarse quantizer output is more than 15, the gain value is decremented according to the table. The higher the magnitude, the bigger the decrement. Conceptually, the adaptation scheme can be considered as a feed back loop whose purpose is to bring back the amplified input signal into half the dynamic range of the coarse quantizer.

Simulations using male and female speech as well as music have been performed and typical SNR values around 26 dB were obtained. Comparative listening tests revealed noticeable differences between original and processed signals. Although the quality is sufficient for applications such as speech coding, A. Schaub's converter cannot be used in more demanding systems such as a portable speech processor.

4.2 ENHANCED FEED-BACK FLOATING POINT CONVERTER

A. Schaub's converter suffers from two critical problems : its perceived quality is insufficient (4.1) and its realization, in particular the 1.5 dB step gain and the look-up table, is not well suited to a low power implementation (4.2.2). An enhanced feed-back floating point converter must thus be developed and the mentioned problems overcome [Gris95].

4.2.1 Improvement of perceived quality

4.2.1.1 Finer quantizers

Improving the signal to quantization noise ratio seems straightforward. In a series of experiments the number of bits in the quantizer, and thus the adaptation table, is increased. Input files are full scale music as well as full scale and -30 dB male speech signals. The latter is most significant since in the foreseen application, normal speech level is about 30 dB below the loudest signal considered (pneumatic hammer or discotheque).

To evaluate the performance, Signal to Noise Ratio (S_SNR) and SEGmental S_SNR (SEG) values are computed according to equation 4.1 are used. x is the original signal and y the converted one. SEG is obtained by applying a regular S_SNR on N segments of 20 ms (stationary time frame for speech) and averaging the obtained values. As a result, small and large amplitude segments contribute in the same way to the final SEG value. On the other hand, in standard S_SNR, large amplitudes mask smaller ones which barely influence the final result.

$$S_SNR = 20 \cdot Log\left[\frac{\sum_i (x_i)^2}{\sum_i (x_i - y_i)^2}\right] \qquad SEG = 1/N \sum_{j=1}^{N} S_SNR_j \tag{4.1}$$

Table 4.2 shows S_SNR, SEG and the amount of overflow (when the adaptation fails to scale the input signal within the coarse quantizer range). Informal listening results are included as well, where N stands for Noisy, F for Fair and E for Excellent.

	speech, -30 dB			speech, full scale			music, full scale		
bits	S_SNR	SEG	Over.	S_SNR	SEG	Over.	S_SNR	SEG	Over.
6	27.4 N	25.9	300	26.2 N	29.2	1439	20.5 N	26.1	2030
7	28.6 N	30.5	321	27 N	32.2	1484	20.7 N	28.1	2078
8	29 N	35.2	321	27 N	34.9	1497	20.7 N	30.5	2111
9	29.1 N	39.2	321	27.1 N	37.1	1504	20.8 N	32.6	2117
10	29.1 N	42.4	317	27.1 N	38.7	2420	20.8 N	33.6	2123

Table 4.2 : Increasing number of bits in A. Schaub's converter

The results clearly show that the number of overflows is considerable and that, as the number of bit increases, S_SNR values barely change. On the other hand, SEG values increase in a more encouraging way. This can be explained as follows : as long as the amplified signal is within the range of the quantizer, improvement is obtained by increasing the number of bits (increasing SEG). However, when the maximum amplitude of the quantizer is exceeded, the additional bits fail to improve the perception (more or less constant S_SNR). Although the number of bits has been increased up to 10, all the processed signals are perceived as noisy (see table 4.2).

4.2.1.2 The « reserve bit » converter

It has been mentioned that in Schaub approach, the adaptation scheme can be considered as a feed back loop whose purpose is to bring back the amplified input signal into half the dynamic range of the coarse quantizer.

To reduce the overflow, the concept of a so called « reserve bit » is introduced. The idea is to modify the feed back loop so that its purpose is to bring back the amplified input signal into a quarter of the dynamic range of the coarse quantizer. In this way a larger margin to protect against overflow is introduced. Hence, the resulting adaptation table for a 7-bit quantizer (6-bit magnitude, 1 sign-bit) is almost identical to table 4.1: the first 6 rows remain the same (i.e. for codewords of up to 29) while the last row reads a -9 dB for magnitude of 30 to 63.

A new set of simulations, with the so called reserve-bit and with increasing quantizer bits was performed. SNR, SEG and overflow values as well as comparative listening are given in table 4.3.

	speech, -30 dB			speech, full scale			music, full scale		
bits	S_SNR	SEG	Over.	S_SNR	SEG	Over.	S_SNR	SEG	Over.
7	33.6 N	27.5	4	33.2 N	32.8	46	29.4 N	30.8	90
8	38.5 F	33.5	4	38.6N	38.7	52	30.6 N	36.6	89
9	44.1 E	39.6	4	42.6 E	44.3	49	31.1 F	41.6	86
10	48.4 E	45.6	4	44.6 E	49.8	54	31.3 E	47.2	86

Table 4.3 : Increasing quantizer bits with the reserve bit concept.

Results show that the reserve bit is very useful in limiting the number of overflows. Obviously the price paid is a reduced resolution. Nevertheless, SEG comparison between table 4.2 and 4.3 indicates that the reduction is less then 6 dB (1 bit). It should be noticed also that as the number of bit increases, the SEG values of table 4.3 for -30 dB speech increase by 6 dB. Excellent processed speech signals i.e. no audible difference from input, are obtained using a 9 or more bit quantizer.

The improvement of the perceived quality can be illustrated with figure 4.3. At the top, a speech segment of 1 second duration is shown. The middle graph shows S_SNR values computed on 20 ms segments in Schaub's converter and for increasing number of bits in the quantizer (6 to 10). Clearly, in some portions (0.3s to 0.45s, 0.85s to 1s), the number of bit increase (and thus the precision) improves the S_SNR values. However, in many other portions (0.2s, 0.55s etc.) overflows occurs and no improvement can be seen. The third graph illustrates the case of the reserve bit converter with increasing number of quantizer's bit. The number of overflow is greatly reduced and exploits the improvement in the perceived quality.

The sound quality in feed-back floating point conversion is drastically improved by applying the reserve bit concept together with an increase of the coarse quantizer size to at least 9 bits.

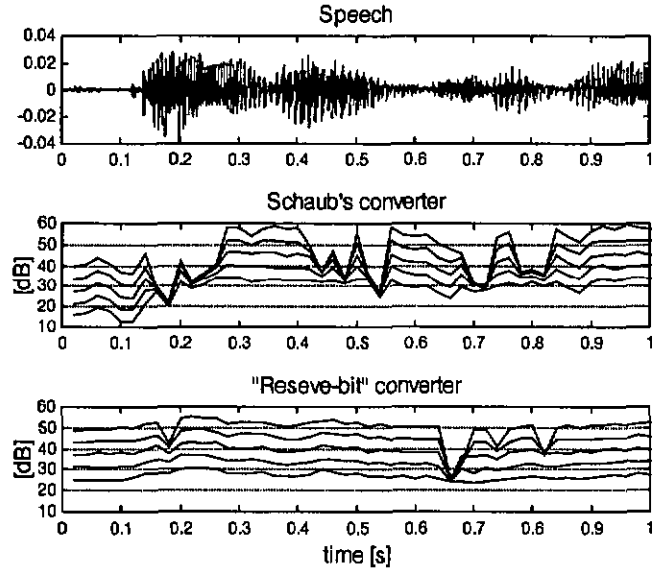


Figure 4.3 : S_SNR on 20 ms segments for A. Schaub (center) and reserved bit converters (bottom), with increasing number of bits.

4.2.2 Simplified implementation

In the former section, the problem of poor perceived quality was solved. Solutions to simplify the implementation of the adaptation strategy must now be investigated.

To counteract the input gain multiples of 1.5 dB, a time varying alignment of the quantizer codeword must be provided (formatting block in figure 4.1). Thus, multiplication with $2^{-1/4}$, $2^{-1/2}$ and $2^{-3/4}$ are required along with shifting operations that compensate 6 dB steps. Integrating a digital multiplier is however not well suited with micro power implementation. A. Schaub proposed to use 3 look-up tables with pre-multiplied codewords. In his design, look-up tables of size 64 were sufficient (6-bit quantizer). Since the size grows exponentially with the number of bits, this approach is not suitable for quantizers with higher numbers of bits.

4.2.2.1 1.5 dB table with 6 dB accumulation :

If the adaptation used coarser minimum steps of 6 dB, the multiplications could completely be avoided and a shifting operation would be sufficient for codeword alignment at the back end.

An elegant solution to reach that goal is to compute the increment/decrement in input gain in 1.5 dB steps as before (using the reserve bit concept) and to compute the accumulated gain. The input amplifier is then set to a "6 dB multiples floored" value of the accumulated gain. This is illustrated in table 4.4 which shows how the gain increment and setting are performed for a given quantizer's output data stream.

As sample #0 is converted, a default gain of 12 dB is applied. Δ gain is computed according to the modified version of table 4.1 (as explained in 4.2.1.3). To set the following input gain, the accumulated gain is floored to a 6 dB multiples value.

T	input gain [dB]	output of quantizer	Δ gain [dB]	accumulated gain [dB]
0	12	000110	+ 1.5	13.5
1	12	000010	+ 1.5	15
2	12	001000	+ 1.5	16.5
3	12	000111	+ 1.5	18
4	18	001000	- 1.5	16.5
5	12	000010	+ 1.5	18
6	18	011100	- 9	9
7	6	010100	-4.5	4.5
8	0	....	....	....

Table 4.4 : 6 dB accumulated values adaptation strategy

Simulation showed that SEG and S_SNR values dropped by only 2 dB compared to the case where the amplifier is actually changed in 1.5 dB steps. Results are shown in table 4.5.

	speech, -30 dB			speech, full scale			music, full scale		
bits	S_SNR	SEG	Over.	S_SNR	SEG	Over.	S_SNR	SEG	Over.
7	31 N	27	14	31.8 N	31.1	75	28.6 N	29.5	152
8	37.1 N	33	10	37.2 N	36.8	86	30.8 N	35	155
9	42.7 E	38.9	16	41.4 F	42.2	87	31.5 N	40	161
10	47.2 E	44.8	18	43.6 E	47.4	88	31.6 F	44.6	154

Table 4.5 : Results from using both the reserve bit and a « 6 dB accumulation » adaptation strategy.

All the simulations results can be summarized in figure 4.4 where SEG values for -30 dB speech of table 4.2 (+), 4.3 (x) and 4.5 (o) as well as perceived quality are presented. Note that perceived quality and SEG (or S_SNR) are sometimes contradictory. The 10-bit quantizer that uses A. Schaub's adaptation (10_+) is

perceived as noisy though has a higher SEG then the 9-bit « reserve bit » converter (9_x and 9_o). The latter, however, are perceived as excellent.

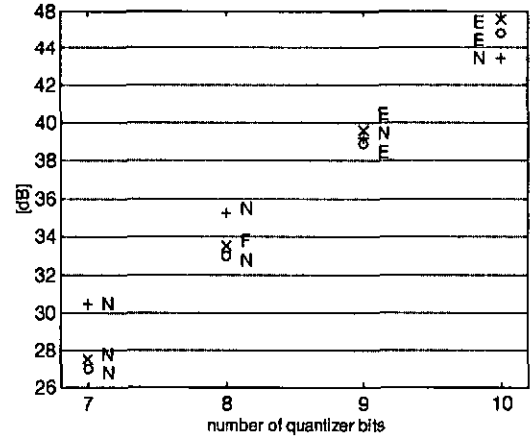


Figure 4.4 : SEG values for different adaptation strategies.

By modifying the adaptation strategy, an input controlled gain with gain values in 6 dB multiples, can thus be used. A question arises : which combination of maximum gain and quantizer size (number of bits) is most suitable ? Further simulation results, with different combinations using male speech are given in table 4.5 and 4.6

-30 dB male speech									
	0-42 dB gain values			0-36 dB gain values			0-30 dB gain values		
bits	S_SNR	SEG	Over.	S_SNR	SEG	Over.	S_SNR	SEG	Over.
8	31.2 F	33.1	7	36.2 N	30.5	2	32.9 N	26	0
9	43 E	39.1	18	42 E	36.4	4	39.2 F	32.4	0
10	48 E	45	29	48.3 E	42.4	5	45.2 E	38.3	0
11	51.5 E	51	40	52 E	48.3	7	51.2 E	44.2	0

Table 4.5 : Gain and quantizer combinations with -30 dB male speech

One sees that three solutions can be considered :

a) 9-bit quantizer and 0 to 42 dB input gain (-> 16 bits dynamic range)

b) 9-bit quantizer and 0 to 36 dB input gain (-> 15 bits dynamic range)

c) 10-bit quantizer and 0 to 30 dB input gain (-> 15 bits dynamic range)

full scale male speech									
	0-42 dB gain values			0-36 dB gain values			0-30 dB gain values		
bits	S_SNR	SEG	Over.	S_SNR	SEG	Over.	S_SNR	SEG	Over.
8	38.4 N	37.7	168	38.4 N	37.3	153	38.5 N	37.5	133
9	41.2 F	43.1	337	41.3 F	43.3	304	41.4 F	43.4	266
10	42.3 E	48	520	42.4 E	48.3	466	42.6 E	48.2	410
11	42.4 E	52.3	702	42.5 E	53.1	627	42.5 E	53.3	555

Table 4.6 : Gain and quantizer combinations with full scale male speech

Numeraus supplementary simulatians invalving different male and female speech signals have been perfarmed. It is interesting to note that sametimes, far a given canfiguration, pracessed male speech files were judged excellent while some pracessed female ones were still anly fair. This is probably due to the pitch (and resulting masking pattern) differences in male and female vaices.

Using a 9-bit quantizer seems more efficient fram a power cansumptian point af view. The first solutian far an audio feed-back floating point converter is thus to use a 9-bit quantizer in cambinatian with a 7 value controlled amplifier and an adaptation lagic camputed according to the reserve bit with 6 dB multiple gain accumulatian strategy. A simple shifter provides the necessary farmatting at the back end.

4.2.2.2 6dB table

Multipliers	4-bit	5-bit	4-bit exten. to 5-bit
M1 to M4	0.8	0.85	0.8
M5	1.2	0.85	0.8
M6	1.6	0.85	0.8
M7	2.0	0.85	0.8
M8	2.4	0.85	0.8
M9	-	1.2	1.2
M10	-	1.4	1.2
M11	-	1.6	1.6
M12	-	1.8	1.6
M13	-	2.0	2.0
M14	-	2.2	2.0
M15	-	2.4	2.4
M16	-	2.6	2.4

Table 4.7 : Jayant's optimal multipliers.

The adaptation table used by A. Schaub and later modified to include the reserve bit and to fit finer quantizers is issued from N. S. Jayant work.

In his work on adaptive quantization with one word memories, Jayant used speech signals sampled at 8 kHz. In order to find the best multipliers (gain increment/decrement) he performed numerous simulations and table 4.7 shows the multipliers he obtained for 4 and 5 bit quantizers (first and second column). The last column results from extending the 4 bit multipliers to fit a 5-bit quantizer.

In this application, a 9 (sign +8) bit quantizer and speech signal sampled at 16 kHz are used. As shown in table 4.7, simply extending the coarse quantizer multiplier's value does not results in optimal values for finer devices. The question of whether the sampling frequency affects the table also arises. Furthermore, would Jayant's table still be useful if signals different from speech were to be converted ?

Because of these questions, it was decided to establish a new methodology to define the multiplier tables. The idea is to use a pool of input signals possessing the same statistical characteristics of the ones to be converted, and to successively simulate different multiplier combinations. The combination resulting in the converted signal of « best quality » is retained.

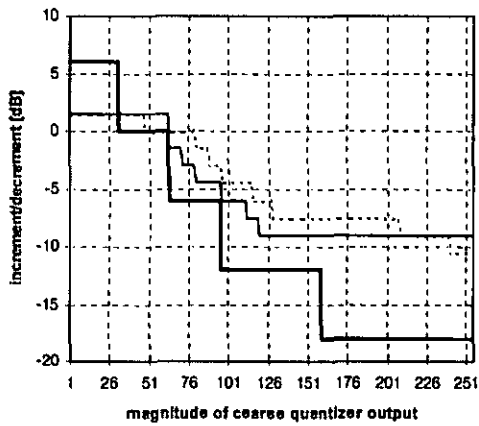


Figure 4.5 : Multipliers for a 9 bit quantizer (8-bit magnitude)

For speech input signals, three different tables have been computed with values restricted to 1.5, 3 and 6 dB multiples respectively. The above figure plots the new 1.5 and 6 dB multiplier tables obtained using a pool of female and male

speech files sampled at 16 kHz as well as the 1.5 dB reserve bit table proposed earlier. For more clarity, the new 3 dB table is not shown.

Figure 4.6 illustrates a few steps of the procedure to obtain the 6 dB table for a 8 bit quantizer. First, the 256 multipliers M are set to 1 (a). The upper half is then successively set to different values (up to G_{max}) and the value G1 resulting in the best SEG (quality criteria) is kept (b). The second half is then scrolled from G_1 to G_{min} and again the best one, G_2, kept (c). The process goes on as illustrated, by successively dividing the working interval until each multiplier value has been set individually. Table computations have been performed in Matlab and the program's flow diagram is given in the annex at the end of this chapter.

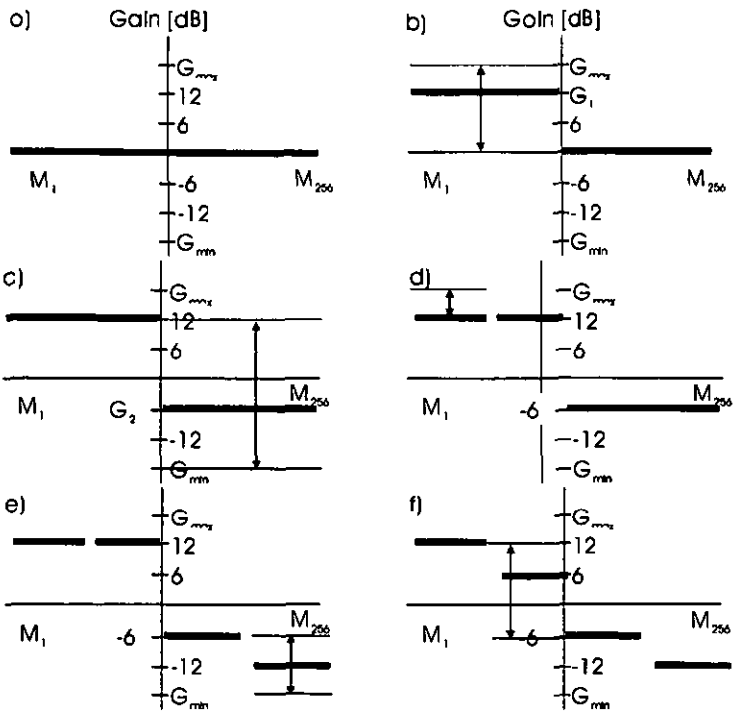


Figure 4.6 : First steps of search of best multipliers combination for 8 bit quantizer

Test signals have been converted. As in the former section, 1.5 dB and 3 dB multiple step values are accumulated and changes brought to effect once 6 dB multiples are obtained. When the 6 dB table is used, no accumulation is performed and changes occur at each iteration. Compared to former SEG

results, the new 1.5 dB table provides slightly better values (≈ +2 dB) while lower ones (≈ -2 dB) are obtained with the 3 and 6 dB tables. These SEG differences are not significant and informal listening tests showed no perceived difference ! This is a fundamental result since it implies that the 6 dB table can be used achieving drastically simplified implementation without loss of perceived quality !

A second solution to convert audio signals is thus proposed. It is similar to the first one (end of 4.2.2.1) though the adaptation strategy is much simplified.

4.2.3 Ideal characteristics

The previous sections defined two feed-back floating point converters well suited for audio applications which differ by their adaptation strategy. Both have a 9-bit quantizer, a 7 gain value (0 to 36 dB) controlled amplifier and a shifter at the back end. The first uses a 1.5 dB reserve bit adaptation with accumulation, where changes are brought to effect when multiples of 6 dB are obtained. The second directly uses a 6 dB table.

Ideal transfer function as well as frequency response have been simulated for both converters. Results are very similar and only the ones obtained with the first solution (reserve bit and accumulation) are shown below.

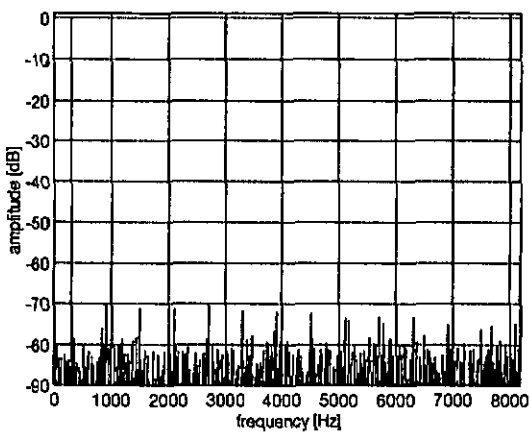


Figure 4.7 : Frequency response to a full scale 300 Hz sine wave.

The frequency response to a full scale 300 Hz sinusoidal input sampled at 16.4 kHz is shown in figure 4.8. The spectral lines at 900, 1500, 2100, etc. Hz (even

harmonics) are particularly well marked. The following ideal characteristics (see 2.1.8 for definition) are obtained :

THD=-61.1 dB SNR=56.2 dB

PHD=-73.3 dB SINAD=55 dB

Plotting the transfer and error function over the whole input range does not properly illustrate the floating point converter characteristics. Partial plots, which have been zoomed are used. Small input signals are used on the left side of figure 4.8. The highest gain is thus applied and the quantization error is 2^{-15} i.e. ½ LSB of the maximum resolution. On the right side, large signals are used. The applied gain is 1 and the maximum error corresponds to 1/2 LSB of a 9-bit quantizer.

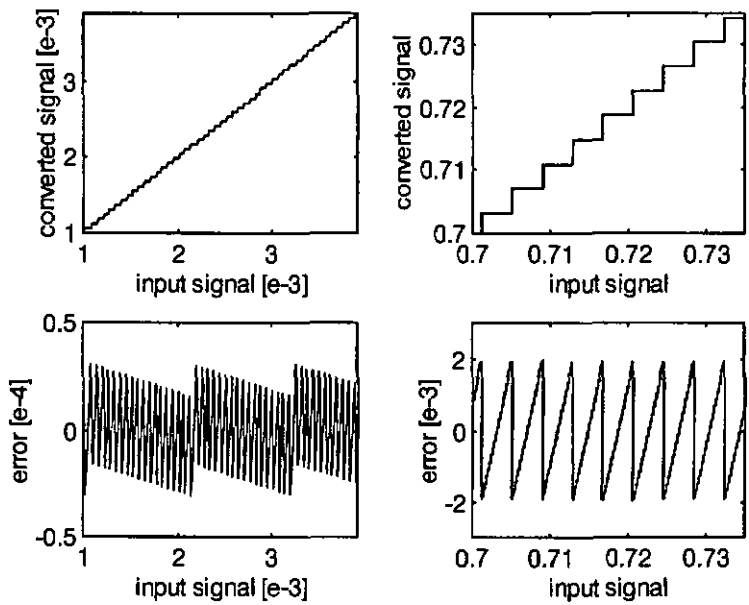


Figure 4.8 : ideal transfer function and quantization error.

From now on, for the sake of simplicity, enhanced feed back conversion (or converters) will be referred as feed back. Enhanced will be added only when confusion with A. Schaub original concept can arise.

4.3 FEED FORWARD CONVERSION

As mentioned at the beginning of the chapter, feed forward adaptation uses the input sample, on which the scaling is then applied, to evaluate the scaling factor. The main advantage of feed forward conversion is that it can be applied to non predictable signals. One possible implementation is to use two coarse quantizers as shown in figure 4.9.

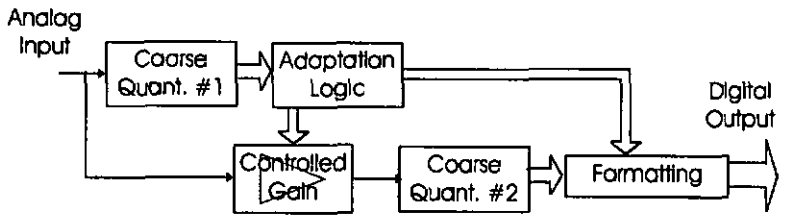


Figure 4.9 : Feed forward floating point ADC implementation

In feed forward conversion, the gain value is computed so that the amplified signal fits exactly the second coarse quantizer. Thus its whole resolution is used and no overflows occur. As a result, better SEG and S_SNR values are obtained. This means that the size of the second quantizer can be reduced and the number of gain values as well as the size of the first quantizer accordingly increased (to keep the required dynamic range).

Issuing conclusions about power consumption is not straightforward. Power consumption in quantizer is proportional to the square of the number of converted bits. Clearly, comparing n bits and m gain feed forward and feed back solutions results in considerable less power for the latter (one quantizer instead of two). Nevertheless, as mentioned, feed forward uses smaller quantizers. Thus, in particular situations (see section 4.4), the feed forward implementation can be more efficient from a power consumption point of view.

4.3.1 Feed forward conversion for audio signals

The feed back converter designed in the previous chapter required a 9-bit quantizer. The adaptation tried to fit the input signal into the seven LSB's while the two MSB's were provided to avoid overflows. In feed forward adaptation, a 7 bit quantizer should thus be sufficient. As shown in table 4.8, this is true for full scale signals. One more bit is required to reach excellent quality with -30 dB speech signals.

	-30 dB		Full scale	
bits in #2	SEG	S SNR	SEG	S SNR
6	21.0	28.1 N	36.4	37.6 N
7	26.9	43.0 N	42.3	43.6 E
8	33.3	40.3 E	48.3	49.6 E
9	39.9	46.3 E	54.0	55.57 E

Table 4.8 : Feed forward measurements

The feed forward converter for audio applications uses a 7-bit #1 quantizer, a 7 value controlled gain and a 8-bit #2 quantizer, resulting in a dynamic range of 14 bits.

4.3.2 Ideal characteristics :

The frequency response of the above converter to a full scale 300 Hz sinusoidal input sampled at 16.4 kHz was simulated.

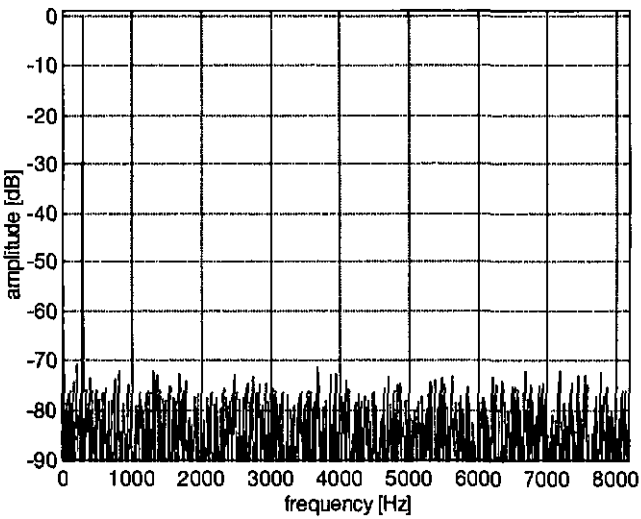


Figure 4.10 : Frequency response to a full scale 300 Hz sine wave

The following ideal characteristics were obtained :

THD=-72.1 dB SNR=55.9 dB

PHD=-70.9 dB SINAD=54.4 dB

Compared to figure 4.7, the spectral distribution is more uniform and no harmonics are clearly visible. As a result higher THD and PHD values are obtained. On the other hand, the mean magnitude of all the spectral lines but the fundamental are higher. This is due to the 8-bit coarse quantizer (9-bit for the feed-back floating point solution) and results in lower SNR and SINAD values.

Partial plots, which have been zoomed, of the transfer function for the audio feed forward converter are given in figure 4.11. For small signals (left), the error is 2^{-14} while for bigger ones it is a half LSB of a 8-bit quantizer.

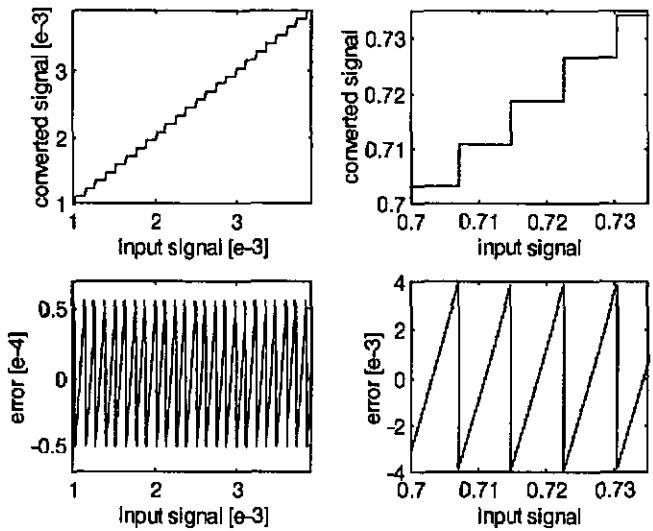


Figure 4.11 : Ideal transfer function and error

4.4 FLOATING POINT CONVERSION FOR NON AUDIO SIGNALS

The use of feed forward or feed back conversion entirely depends on the data to be converted. Indeed, since feed back conversion works by predicting the next input sample, a certain « correlation » must exist between adjacent samples. More precisely, the difference between adjacent samples must be reasonably predictable.

One interesting measure is to compute the probability that the difference between two samples falls within a given interval. A combination of high

probability and small interval is desired for any signal to be feed-back converted. Such a measurement has been performed on various signals and results are given in table 4.9.

type	sampling frequency	difference interval	probability
male speech	16 KHz	± 1/64	0.94
classic music	16 KHz	± 1/64	0.5
pop music	16 KHz	± 1/64	0.41
pseudo-random	16 KHz	± 1/64	0.035
180 Hz sine	100 KHz	± 1/100	0.7

Table 4.9 : Difference statistic for various signals

These results can drastically change depending on the sampling frequency and/or difference interval. The inter-sample statistic must thus be thoroughly investigated.

type	quantizer #1 [bit] [μW]	quantizer #2 [bit] [μW]	control. gain [# of gain] [μW]	total dissipation [μW]
feed forward	7 15	9 24	7 20	59
feed back p1	-	10 30	7 20	50
feed back p1>p2	-	13 50	7 20	70

Table 4.10 : Estimated power consumption comparison

There is no clear limit on the minimum probability and maximum interval suitable for feed back conversion. It depends entirely on the application and its « quality criteria ». Nevertheless, as the probability decreases and/or the interval increases, to ensure a minimum given quality, the size of the coarse quantizer must increase. At a certain point, it will be more efficient, from a power consumption point of view, to use feed forward conversion. This is shown in table 4.10 for a hypothetical converter with a required resolution of 9 bits and a minimum dynamic range of 15 bits. First a feed forward solution is considered. A 9-bit #2 quantizer and a 7 value controlled gain are used. A 7-bit quantizer #1 is thus required. The second solution is suited for signals with extremely good inter-sample correlation (p1) thus only 10 bits are sufficient for the coarse quantizer. In the third solution, a poor inter-sample correlation is assumed (p2<<p1), thus a very large quantizer is required to compensate the mediocre "predicability". Power consumption are rough estimations.

Feed forward converter specification is simple : the size of the second quantizer is defined by the desired resolution and the first quantizer, as well as the number of gain values in the controlled gain, is then set to satisfy the dynamic range constraints.

For feed back conversion the major problem is to elaborate on appropriate adaptation strategy. This includes the adaptation table as well as the number of gain values in the controlled gain and number of bits in the coarse quantizer. The methodology proposed in section 4.2.2.1 can be applied. The « better quality » criteria however must be redefined according to the considered application.

From a theoretical point of view, the floating point conversion could be extended indefinitely. Implementation issues, such as minimum noise level for a given technology, will however limit the maximum resolution, or for a given resolution, the maximum dynamic range.

4.5 DESIGN CONSIDERATIONS

The feed back floating point converter is made of four modules : control logic, adaptation logic, coarse quantizer and controlled gain. The feed forward converter has an extra coarse quantizer.

Since control and adaptation logic are implemented in the digital domain, a perfect « translation » of the adaptation strategy can be obtained. The matter is different with the analog blocks. In particular, the controlled gain is critical to the overall performance of the system. If the gain values are not exact 6 dB multiples, reconstruction at the back end is disturbed. The shifter provides exact division by 2, 4, 8 etc. and distortions will occur.

In the case of audio feed back converters, simulations of non-ideal controlled amplifiers were performed. First, a linearity error was introduced. In other words, all the gain values were uniformly multiplied by an error factor. The perceived quality was not degraded even with gain errors of up to 50%. However, the loudness perception was modified. A second set of simulations, with non-linear errors (some gain values are multiplied by the error factor while others are divided) was performed. A degradation of perceived quality could be heard for error factors higher then 2%. The frequency response as well as computed characteristics of such a non ideal converter, with an error factor of 1% are given in figure 4.12 and below. As expected, compared to the ideal solution (section 4.2.3) a degradation of the characteristic is obtained.

THD=-59.7 dB SNR=55.9 dB

PHD=-66.6 dB SINAD=54.4 dB

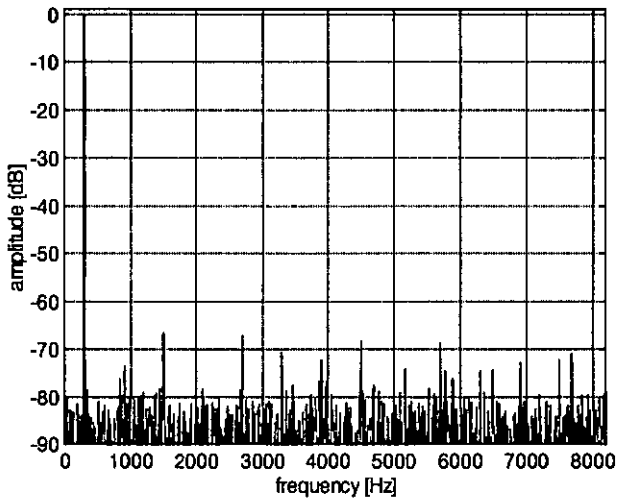


Figure 4.12 : Frequency response of non ideal feed back floating point converter

4.6 CONVERTERS FOR AUDIO APPLICATIONS : SUMMARY

In this chapter three floating point converters for audio applications have been defined. They all feature the same perceived quality.

The overall performance of the converters is given by two critical elements : the coarse quantizers and the controlled amplifiers. The linearity error of the latter is not so significant, though the gain value dispersion is critical and should be smaller then 1%.

4.6.1 Feed back reserve-bit converter :

This converter is based on the feed back conversion principle. It uses a 9-bit coarse quantizer and a 7 gain value controlled amplifier. The adaptation logic is performed according to a modified 1.5 dB increment / decrement step table

(providing margin against overflow). The gain settings are brought to effect only when they have accumulated to multiples of 6 dB.

4.6.2 Feed back 6 dB converter :

As in the previous converter a 9-bit coarse quantizer and a 7 gain value controlled amplifier are used. The feed back adaptation logic is based on a 6 dB increment / decrement table obtained through simulations with a pool of input signals.

4.6.3 Feed forward converter :

This converter is based on the feed forward conversion principle. It uses an 8-bit coarse quantizer, a 7 gain value controlled amplifier and thus a 7-bit quantizer for the gain adaptation.

4.7 REFERENCES

[Gris95] L. Grisoni, A. Heubi, S. Grassi, P. Balsiger and F. Pellandini, A. Schaub « Micro Power Relative Precision 15-bit A/D Converter », ICSPAT'95. Oct. 24-26 1996, Boston MA, USA, pp 420-424.

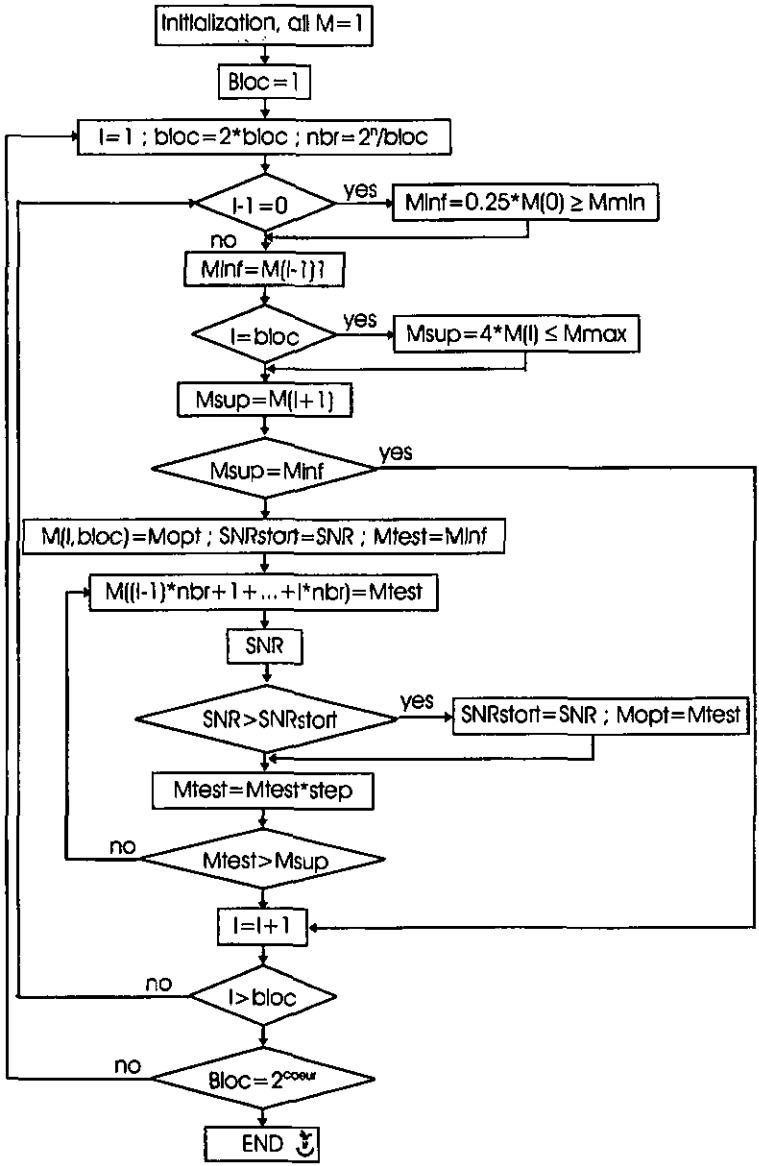
[Scha92] A. Schaub, « Micro Power A/D Converter for Audio Signals », Digital Signal Processing, Vol. 2, pp 110-114, 1992.

[Jaya84] N. S. Jayant, P. Noll, « Digital Coding of Waveforms : principles and application to speech and video », Prentice-Hall, 1984, ISBN 0-13-211213-7 01.

[Jaya73] N. S. Jayant, « Adaptive Quantization with One Word Memory », Bell Syst. Tech Journ., pp 1119-1144, September 1973.

4.8 ANNEX

Flow diagram of Matlab program for multipliers research (see section 4.2.2.2).


Initialization, all M=1
Bloc=1
I=1 ; bloc=2*bloc ; nbr=2^n/bloc
I-1=0
yes
Minf=0.25*M(0) ≥ Mmin
no
Minf=M(I-1)
I=bloc
yes
Msup=4*M(I) ≤ Mmax
Msup=M(I+1)
Msup=Minf
yes
M(I,bloc)=Mopt ; SNRstort=SNR ; Mtest=Minf
M((I-1)*nbr+1+...+I*nbr)=Mtest
SNR
SNR>SNRstort
yes
SNRstort=SNR ; Mopt=Mtest
Mtest=Mtest*step
no
Mtest>Msup
I=I+1
no
I>bloc
no
Bloc=2^coeur
END

5. Micro power RSD converters

The Redundant Signed Digit (RSD) non-absolute linear converter has been developed by Alexandre Heubi from IMT. A close look to the quantization steps along the dynamic range shows that this converter features a mixed characteristic. Indeed, the characteristic is absolute except at certain input levels where a « burst » occurs and the quantization step gets bigger.

This converter is extremely well suited to audio applications : it has a small size and features an ultra low power consumption. It is thus a perfect for comparison with the floating point devices.

For dynamic ranges smaller or equal to 10 bits, the A. Heubi converter actually features an absolute characteristic. The concept can thus be used to implement the coarse quantizer(s) in the floating point realizations defined in the previous chapter.

The mixed RSD converter presents the advantage that it can be digitally corrected to result in absolute devices. In connection with this PhD, the design, implementation and test of a 14-bit absolute RSD converter was thus performed. Compared to the non-absolute device, the absolute converter features a bigger die size (because of the correction algorithm), however, by optimizing the digital supply voltage, the power consumption is kept low.

The 14-bit absolute RSD converter is useful to draw conclusions about extended (higher dynamic range and/or resolution) floating point realizations as discussed at the end of chapter 6.

The finesse of the converter implemented by A. Heubi, which results in micro power consumption, is covered by the patent entitled "Dispositif de traitement numerique d'un signal analogique devant etre restitue sous forme analogique", registered at the "Institut National de la Propriete Industrielle" (Paris, France) with number 95 10 174 and dated 29 August 1995. The patent belongs to the University of Neuchatel.

No electrical schemes are discussed here, however, the RSD algorithm is described and A. Heubi implementation results presented. The interested reader might want to consult [Heub96]

Section 5.1 deals with the mixed RSD converter while section 5.2 presents its 14-bit absolute version.

5.1 A. HEUBI'S CONVERTER [HEUB96]

As explained in 2.2.5, conventional cyclic converters require extremely precise comparators (inaccuracy smaller then a half LSB). This is to ensure that the analog input is coded into the correct digital word. If the comparator is not precise enough, the comparison result is wrong and all the subsequent bits are erroneous as well.

Ginetti [Gin88] proposed to use the Redundant Signed Digit (RSD) cyclic successive approximation algorithm shown in figure 5.1. In this case, the intermediate voltage Vx is not compared to 0 but to both a negative (-Vth) and positive (Vth) threshold value. Three situations can occur: the intermediate voltage is either smaller than -Vth (right branch), bigger than Vth (left branch) or between -Vth and Vth (middle). In all three branches, to obtain the next intermediate voltage, the previous value is first doubled. Then, in the external branches, the reference voltage is either added or subtracted. This is to ensure that the resulting new intermediate voltage value remains within the ±Vref range. Note that the bits are set to either 0, 1 or -1, hence a « ternary » representation.

To reconstruct the decimal value from the ternary RSD word, each weight (2^0, 2^{-1}, 2^{-2}, etc.) is multiplied by the bit value and summed. Because a bit can be negative, a single decimal value can be expressed by various RSD words. For example, an analog input value of 0.265625 can be digitally encoded by [0 1 0 0 0 1] or [1 -1 0 0 0 1] or [0 1 0 0 1 -1] etc... Hence, even if a comparison error occurs and a bit is wrongly set, the following bits can "compensate" the error and the digital word still be correct. The design constraints on the comparators are thus drastically simplified and inaccuracies of up to half V_{ref} are tolerated,

regardless of the number of bits. As a result, non active architectures can be used which clearly benefits the power consumption balance.

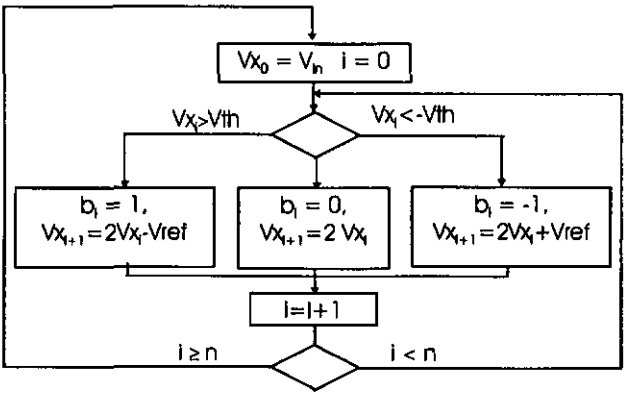


Figure 5.1 : Cyclic RSD conversion algorithm [Gin92]

Alexandre Heubi proposed a very efficient switched capacitor implementation of the above algorithm. The input signal is sampled at the beginning of the conversion cycle and thus no sample & hold is required. The large tolerance comparators are realized by simple strobed cross-coupled inverters. The RSD output is converted into a two's complement representation. The proposed schematic requires one active element only and 3 identical capacitors. Its advantages are :

- active element offset gives only a digital offset (which can be easily compensated if needed).
- switch charge injection has the same effect as above.
- active element saturation causes digital saturation (no distortion for low level input signals).

5.1.1 Mixed linear characteristic

In fact, because of technological limitations, the proposed implementation does not perfectly translate the algorithm of figure 5.1: because of capacitor mismatch the doubling as well as the voltage reference (Vref) addition/subtraction is not perfect. This is worsened by the fact that the active element only features a limited DC gain. As a result, the converter presents a non-absolute characteristic.

Static measurements have never been performed, however A. Heubi was able to determine that at some particular input amplitude only, the quantization error was more than a half LSB. One can thus classify this converter in the mixed linear category.

It should be stressed that this mixed behavior only appears in devices featuring more than 10-bit dynamic range. Indeed for a given supply voltage, the LSB level is much higher in a 10-bit device than in a 14-bit one and hence bigger errors (due to non ideal doubling, voltage reference source, voltage reference addition/subtraction etc.) can be tolerated without affecting any of the converter bits. This means that the schematic proposed by A. Heubi can be reused for the 9 bit coarse quantizer of the floating point converter defined in the previous chapter. A redesign is nevertheless mandatory to optimize the capacitor sizes as well as the transistors and current source to ensure minimal power consumption.

5.1.2 Implementation and test

The circuit was implemented by A. Heubi in the ALP2 LV double metal, double poly, 2 µm CMOS technology from EM Microelectronic Marin SA in Switzerland. The analog part is full custom layout while the digital part is realized using CSEL_LIB, a low power standard cell library developed at CSEM SA, Neuchâtel, Switzerland. Additional components, in particular a D/A with the same relative precision as the A/D and a track and hold, have been added to the test circuit. Two signals namely « ck » and « sync » are used to control the sampling frequency and the number of bits as shown in figure 5.2. This circuit is thoroughly presented in [Heu96].

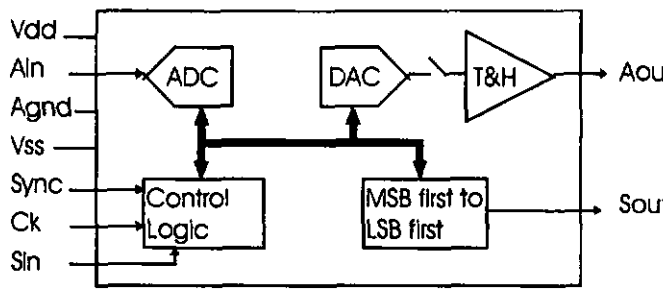


Figure 5.2 : Block diagram of A. Heubi's chip.]

The whole circuit has an area of 4 mm² (including pads) and a core of about 0.8 mm². The analog part of the A/D (top left of figure 5.3) is only 0,06 mm² including the polarization circuit. The top part is the A/D, D/A and Track & Hold.

The very dense part, in the lower section of the chip, groups the digital functions i.e. control logic and output register.

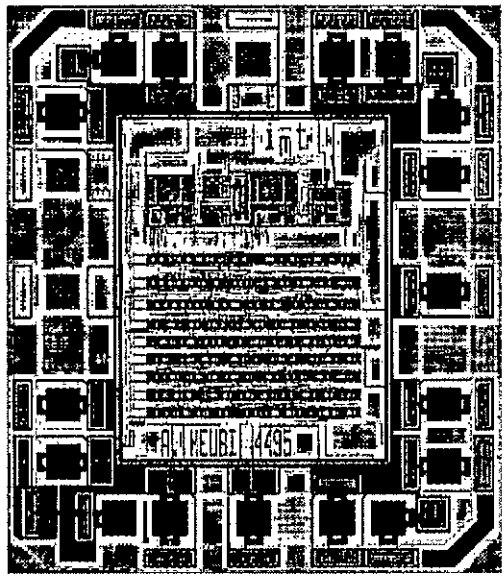


Figure 5.3 : Layout of A. Heubi's converter

The circuit has been tested on a PC based environment equipped with a data acquisition card from National Instruments (AT-MIO-16E-1) driven with the LabVIEW software. The measured power consumption at ± 1.2 V power supply and 16 kHz sampling frequency was:

13 μA for the A/D converter (31 μW)

32 μA for the D/A converter and the track & hold circuit (75 μW)

7 μA for the logic part (17 μW)

125 μW total power dissipation

< 1 μA in stand-by mode

The frequency response to a -25 dB below full scale 1 kHz sinusoidal input sampled at 16 kHz and with ± 1.2 V power supply is shown in figure 5.4. A DC component is clearly visible as well as a few harmonics.

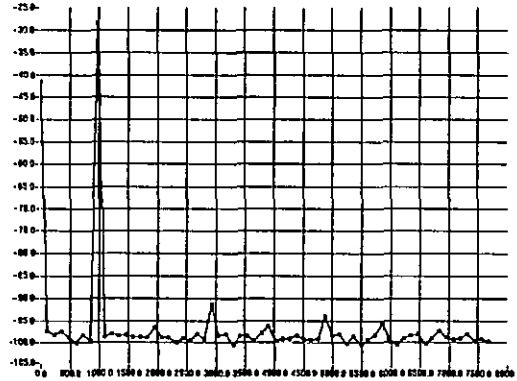


Figure 5.4 : Frequency response of A. Heubi's converter

The measurements of the Total Harmonic Distortion (THD) as well as THD + noise which is equivalent in magnitude to the dynamic SNR (see equations 2.1 and 2.2, section 2.1.5) is particularly interesting. Figure 5.5 show the THD (lower) and THD+noise (higher) values for input 1 kHz sine waves with amplitude of 0 to -80 dB below full scale, hence covering the most of the ideal dynamic range The curves feature a unity slope up to about -30/-25 dB. This means that the converter features an absolute behavior over approximately 60 dB of dynamic range. A saturation at about 60 dB SNR occurs for larger input levels and the behavior moves to a non-absolute characteristic.

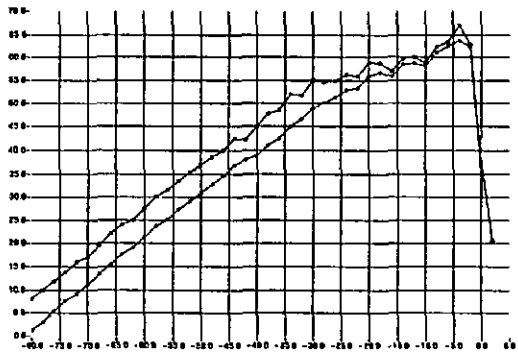


Figure 5.5 : THD and THD+noise versus input signal level

Noise floor measurements have also been performed. They showed that a dynamic range of 82.5 dB (13.75 bits) is reached at ± 1.25 V while 86.1 dB (14.35 bits) can be obtained at ±2.25 V.

5.2 14-BIT ABSOLUTE RSD CONVERTER

As mentioned in the previous section A. Heubi's converter presents a non-absolute characteristic due to technology sensitivity. In fact, the « real » (or implemented) algorithm can be expressed as in figure 5.6. Compared to figure 5.1, the perfect doubling is replaced by 2+ε, where ε expresses the error coming both from capacitor matching and finite amplifier gain. The addition/subtraction of the reference voltage now becomes ±(1+β)Vref and again β expresses the error coming both from capacitor matching and finite amplifier gain. Figure 5.6 assumes that the reference voltage is stable (precision greater then that of the converter). If it wasn't the case, the left and right branch should be modified to show ±(1+β)(Vref+δ). Finally, an offset voltage φ is added to the input.

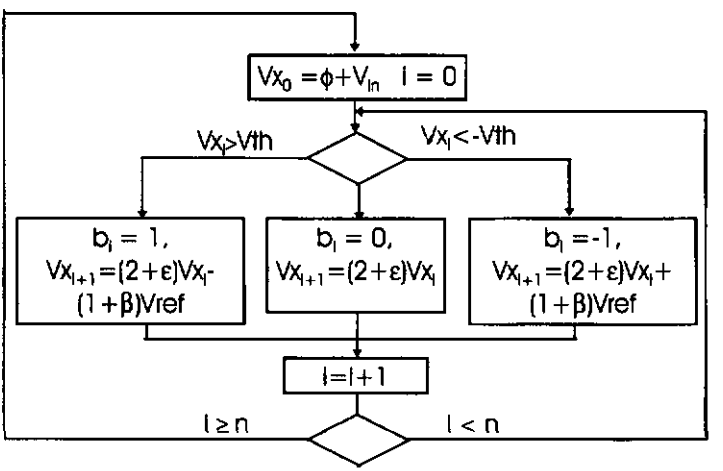


Figure 5.6 : Implemented cyclic RSD algorithm

From figure 5.6, one sees that the doubling error affects all three branches, while the addition/subtraction of the reference voltage affects only the external one. This means that if the input value is such that the algorithm spends most of the time in the middle branch, the addition/subtraction will only slightly affect the results. On the contrary, if most of the time is spent in the external branches, this error really becomes significant.

According to figure 5.6, to compute the correct digital output magnitude equation 5.1 must be used:

$$DIG_out = (1+\beta)\cdot\sum_{i=0}^{n-1} b_i(2+\varepsilon)^{-i} \qquad 5.1$$

Indeed, because of the non ideal doubling factor, each bit computed by the implemented RSD algorithm has a real weight of $(2+\varepsilon)^{-1}$ [Gine88]. In the ideal case, this weight would be 2^{-1} and a « base translation » has occurred. On the other hand, the addition/subtraction error results in a $(1+\beta)$ scaling of the bits.

Since ε is fairly small (<0.4 %), $(2+\varepsilon)^{-1}$ can be expressed by the two first terms of its Taylor development, as in equation 5.2.

$$(2+\varepsilon)^{-i} = 2^{-i} + \frac{i\cdot\varepsilon}{2^{i+1}} \qquad 5.2$$

Thus :

$$\begin{aligned}
&(2+\varepsilon)^{0} = 1\\
&(2+\varepsilon)^{-1} = 1/2 - \varepsilon/4\\
&(2+\varepsilon)^{-2} = 1/4 - 2\varepsilon/8 = 1/2\cdot(1/2 - \varepsilon/4) - \varepsilon/8\\
&\ldots\\
&(2+\varepsilon)^{-i} = \frac{(2+\varepsilon)^{-(i-1)}}{2} - \frac{\varepsilon}{2^{i+1}}
\end{aligned} \qquad 5.3$$

From equations 5.1 and 5.3, doubling and addition/subtraction errors are corrected by the algorithm of figure 5.7 where b_i are the bits computed by the non absolute RDS converter (A. Heubi converter). Register #1 is initialized (at each conversion beginning) at $1+\beta$ and register #2 at $\varepsilon/2$. If necessary, register #3 can be initialized at the proper value to ensure offset compensation. The algorithm directly performs the RSD (ternary) to two's complement transformation.

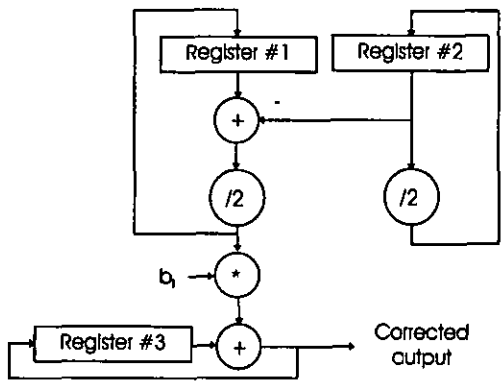


Figure 5.7 : Correction algorithm

Using β and ε values of 0.4 % (maximum mismatch for ALP1 or ALP2), simulation showed that a 17 bit decimal datapath is necessary and results in an average absolute error of 0.4 LSB. The maximum error is 1 LSB with a probability of 0.05 %.

The major remaining problem is to properly measure ε, β and ϕ values. Assuming these values are available, the algorithm of figure 5.7 successfully transforms the mixed linear RSD converter into an absolute device.

5.2.1 Implementation

Before starting the design process, the calibration strategy must be defined. One solution is to measure ε, β and ϕ during foundry test and store the values in a non volatile memory. It is nevertheless simpler to perform the calibration at each power up and to locally store the measured values on the device (5.2.1.1).

The converter has been designed in ALP1, the low voltage, 1μm, 2 poly, 2 metal CMOS technology from EM Microelectronic-Marin SA.

5.2.1.1 Analog design :

The analog part of the absolute RSD converter is very similar to that of the mixed one.

Lee in [Lee93] explains how the ε, β and ϕ values can be measured. Basically, three supplementary switches and a slightly modified switching control are required in the calibration phase.

The targeted performances were 14 bits at ± 1,25 V and 16 kHz and analog design was done accordingly. However, a final post layout simulation showed that at this supply voltage a dynamic range of only 13 bits is obtained. This is due to the OTA internal noise which is a little higher than in the first transistor level simulation. However, at ± 1.5V, according to the post layout simulation, the dynamic reaches 14 bits and the power consumption increases from 36.5 μW to 51 μW.

5.2.1.2 Digital design

The digital design includes 3 main modules (figure 5.8) :

- correction algorithm and the initialization registers (CORR)

- digital and analog control (AH_CONT, A_TO_L, CORR_CONT)

- output interface (OUTPUT)

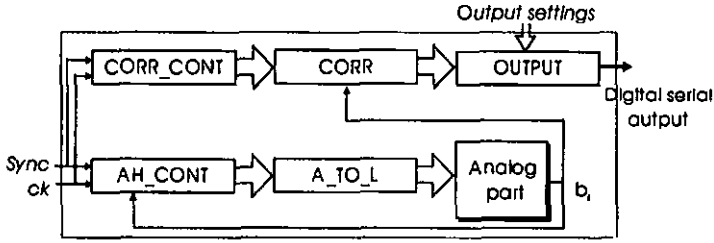


Figure 5.8 : block diagram of linear RSD converter

Regarding the correction algorithm (equation 5.3), VHDL code was written, simulated and synthesized within COMPASS. To optimize the area and power consumption, each register's size was minimized and subsequent sign extension used whenever possible. The main components are 6 registers (2 x 20 bits, 2 x 12 bits and 2 x 19 bits), 3 multiplexers and some gates.

The control module is made of two main blocks : the first, A_TO_L, uses the control signals generated by A. Heubi, AH_CONT, for his mixed converter and modifies them to allow proper measurement of ε, β and ϕ. It is made of a 3-bit counter, 2 D flip-flops and several gates. The second, CORR_CONT generates the correction algorithm control signals (to store/load ε, β and ϕ, clock the registers, etc.).

The output block allows the users to chose between a serial LSB first or MSB first output. Furthermore, to be compatible with commercial DSP processors, the number of output bits can be set to 8, 12, 14 or 16. This function must be used in combination with the « ck » and « sync » signals. It must be stressed that when the output is 16 bits, the converter actually produces 14 true bits and two noisy ones (LSB and LSB-1).

A level shifter between the analog and digital parts made it possible to run the digital part with half the supply voltage. The simulated power consumption, for the digital part, is thus 8.5 µW at 1.25 V.

5.2.1.3 Final layout :

The final chip layout is given in figure 5.9. The die size is 1.03 by 1.14 mm (1.29 by 1.45 with pads). The expected total power consumption at 16 kHz is 45 µW at ± 1.2 V for a 13 bits dynamic and 64 µW at ± 1.5 V for a 14 bits dynamic.

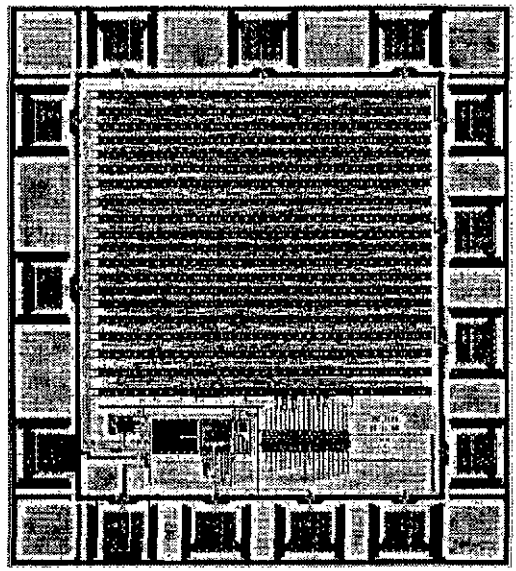


Figure 5.9 : Layout of 14-bit absolute RSD converter

5.2.2 Test

As explained later in 7.1, the measurement of a 14 bits resolution device is difficult because of the environmental noise. Furthermore, the supply voltage, which is also used as reference voltage, is only accurate to 12 bits. This will limit the measured performances.

Power consumption measurements are given in table 5.1 and are in agreement with the predictions. These somehow « strange » sampling frequencies result from the pattern generator used to obtain the sync and ck signals.

Sampling frequency	Power consumption [µW]	
	at ± 1.25 V	at ± 1.5 V
8.9 kHz	26.1	36.75
17.8 kHz	47.5	65.4

Table 5.1: Measured power consumption

A second set of tests showed that the internal reset of the converter capacitors (performed at the beginning of each conversion cycle) is a little short to allow full

discharge at 16 kHz sampling frequency. As a result all the following measurements will be performed at a lower sampling frequency. This reset problem will be easily corrected in a future redesign by modifying the control signals and will imply no power consumption or die area increase.

The noise floor is given in figure 5.10 for supply voltage of ± 1V to ± 1.53 V.

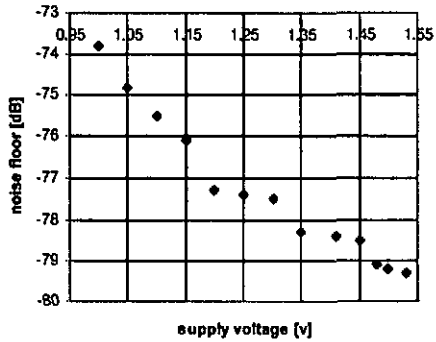


Figure 5.10: Noise floor measurement

The noise floor value at ± 1.25 V shows a dynamic of 12.9 bits. At ± 1.5 V, only 13.2 dB are reached. These values are slightly lower then the prediction of section 4.3 (based on post-layout simulations) and are certainly due to the reference voltage inaccuracy as well as other environmental perturbations (measurement probes, analog/digital coupling etc.).

The transfer function of the converter was measured and the Step Over Range (SOR) is plotted in figure 5.11 where the values are given in LSB's. For each input value several conversions were performed. The SOR values are not integers since they are computed using the average output. Because of the environmental noise, SOR of about ±1 LSB can be considered as within the measurement error. However, one sees that for high input levels the SOR values can get as high, in magnitude, as 4. This means that a perfect linearity of 13 bits is not reached. This is due to the imprecise reference voltage. The higher the input level, the sooner a reference addition/subtraction must take place. An error occurring during this operation is then multiplied in all the successive iterations.

From the above test results one can conclude that the power consumption estimations have been confirmed by chip measurements. The dynamic range could not be satisfactorily tested because of environmental noise. However, it was possible to bring to light that the calibration phase must be improved in order to ensure an absolute characteristic over 14 bits.

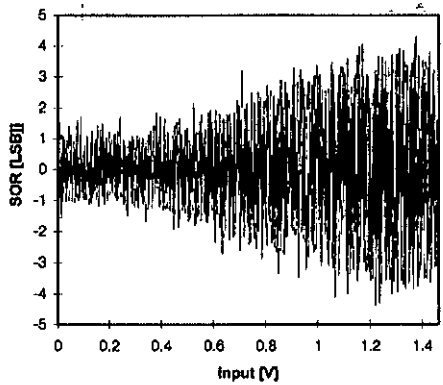


Figure 5.11: SOR characteristic

In spite of a few youthful indiscretions that could be easily corrected, the presented ADC remains a valuable contribution to the development of ultra low power devices to be used in battery operated and portable applications.

5.3 REFERENCES

[Gris97] L. Grisoni, A. Heubi, P. Balsiger and F. Pellandini, « Micro Power 14-bit ADC : 45µW at ±1.3V and 16 ksamples/s », Proceedings of ISIC'97, 10-12 September 1997, Singapore.

[Heub96] A. Heubi, P. Balsiger and F. Pellandini, « Micro Power 13 bits Cyclic RSD A/D Converter », ISLPED'96, Aug. 12-14 1996, Monterey CA, USA, pp 253-257

[Gris96_2] L. Grisoni, A. Heubi, P. Balsiger and F. Pellandini, « Micro Power 14-bit RSD A/D Converter », ICSPAT'96, Oct. 8-10 1996, Boston MA, USA, pp 510-514.

[Gin92] B. Ginetti, P. Jespers, and A. Vandemeulebroecke, « A CMOS 13-bit Cyclic A/D Converter », IEEE Journal of Solid-State Circuits, vol. 27, No 7, July 1992, pp. 957-965

[Lee93] Hae-Seung Lee, « A 12 Bit 600kS/s Digitally Self-Calibrated Pipeline Algorithmic ADC », 1993 Symposium on VLSI Circuits, Kyoto, May 19-21, 1993

[Gin88] B. Ginetti, A. Vandemeulebroecke, and P. Jespers, « RSD Cyclic Analog-to-Digital Converter » , Symp. VLSI Circuits Dig. Tech. Papers, 1988, pp. 125-126

6. Design of floating point analog to digital converters

This chapter describes the design and implementation, in a low voltage 2 µm CMOS technology, of floating point linear A/D converters for audio signals. Each constituent block i.e. controlled gain, coarse quantizer, adaptation logic and control logic, are considered separately.

The advantage and disadvantage of various electrical schemes for the controlled gain are studied. A simple inverting amplifier, which insures the required performance while keeping the power consumption low, is chosen and the transistor sizing thoroughly explained.

The coarse quantizer is a 9-bit absolute version of Alexandre Heubi device though its design is not given.

The implementation of the adaptation logic is relatively simple: the adaptation behavior is first described in VHDL code, then compiled and synthesised using a standard cell library from CSEM.

Finally, the control logic is simply described by a block diagram that is compiled and synthesized.

Based on the above designs, chip estimations for both the feed back and feed forward audio A/D converter are presented. The discussion is also extended to converters featuring higher dynamic range and resolution.

To verify the predictions of chapter 4 and complete the study of floating point converters, hardware implementations were performed.

The feed back reserve bit converter (4.5.1) was implemented in ALP2LV (low voltage 2µm technology from EM Microelectronic-Marin SA). The integrated version is enhanced with special features to facilitate test and to allow adaptation modifications. Hence, to obtain a feed back 6 dB converter, the adaptation logic is simply switched off and an FPGA programmed with the 6 dB table is connected. No feed forward realization was performed. Nevertheless, the results

obtained from the feed back implementation allow a good estimation of power consumption and chip size.

This chapter is organized as follows. Section 6.1 introduces the design methodology. The design of the four basic modules (i.e. controlled amplifier, coarse quantizer, adaptation logic, analog and digital control) for the reserve bit converter is detailed in sections 6.2 to 6.5 and the resulting chip presented in 6.6. Realization of the 6 dB adaptation logic is explained in section 6.7 while feed forward chip size and power consumption are estimated in 6.8. Finally, 6.9 discusses the problems and solutions for the design of a maximum dynamic range device, which could be used for non-audio applications.

6.1 DESIGN METHODOLOGY

At IMT, mixed-mode ASICs are designed using the below methodology:

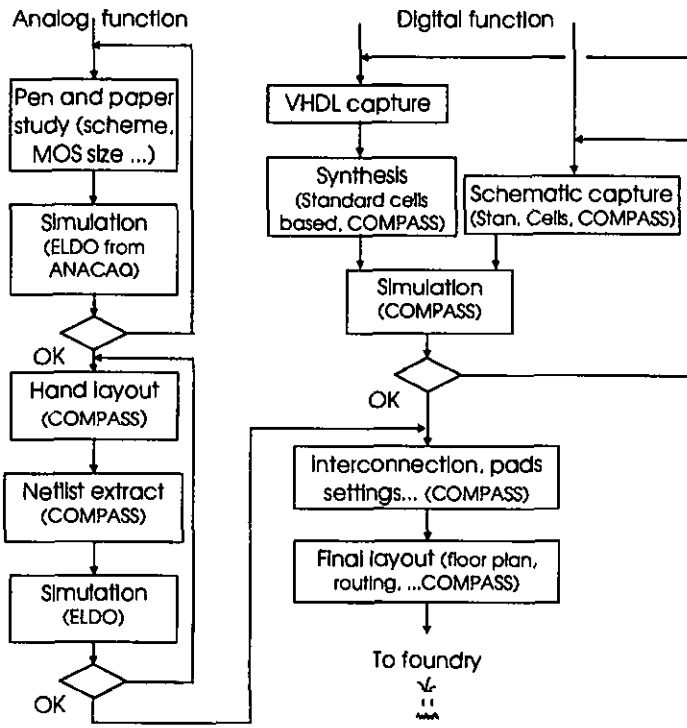


Figure 6.1: Design methodology

The analog blocks or functions are first described with equations whose solutions give the various transistors sizes. The functions are then simulated using a SPICE type simulator named ELDO from ANACAD. As mentioned, these simulations imperatively require accurate models such as EKV ones. Once satisfactory schematics are obtained, hand layout (COMPASS) can take place. Netlists are extracted and a last SPICE simulation ensures that the analog functions (or blocks) are ready.

The digital functions are either described in VHDL and compiled (COMPASS) using a standard cells library (typically CSEL_LIB) or directly captured in a block diagram (COMPASS) of standard gates and basic blocks. Proper behavior is verified through simulation.

The digital and analog blocks are then interconnected, floor plan established, pads defined, etc. to result in the final layout. Normally at this point, a final mixed mode simulation should take place. Unfortunately, this is not possible with IMT's present infrastructure and consequently the chip cannot be thoroughly verified.

As far as the FPGA is concerned, a VHDL description was first performed. Simulation and synthesis took place within SYNOPSYS and this produced a netlist. The latter was fed to dedicated software, which implements the digital function on a targeted FPGA.

6.2 CONTROLLED AMPLIFIER

To obtain the design constraints for the controlled amplifier, the following aspects must be considered.

i) Gain values : as explained in chapter 4, the controlled amplifier must provide a selectable amplification factor of 1, 2, 4, 8, 16, 32, 64.

ii) Input noise level : the input noise level must ensure that for any gain value, the output noise level is sufficiently small. A value of one half LSB is targeted.

$$noise_{out_mx} = V_{dd} \cdot 0.5 \cdot 2^{-8} \qquad noise_{in_mx} = noise_{out_mx} / 64 \tag{6.1}$$

iii) Scheduling : depending on the chosen method (with or without pre-sampling), the time frame to perform the amplification is 1/16000 or 5/(16000*2*8) (see section 6.2.5 and figure 6.14).

iv) Power supply and dynamic range : typical supply voltage is ±1.3 V. Input signals have a dynamic range close to rail-to-rail and a similar characteristic is desired for the output signals.

The various considered schemes can be classified in 3 main categories : capacitor network, switched capacitor and integrating. Their respective advantages and disadvantages are discussed, taking into account some intrinsic limitations.

6.2.1 Limitations

All the presented analog schemes are made of OTAs, capacitors and switches which all introduce some limitations

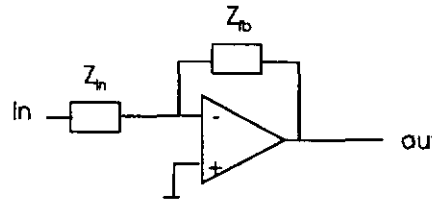


Figure 6.2: Inverting scheme

Let us consider a simple inverting scheme with feed back impedance Z_{fb} and another impedance connected to the inverting OTA input Z_{in}. In the ideal case, the OTA is assumed to have an infinite gain and the transfer function is given by equation 6.2.a. However, when a finite gain A is considered, the function becomes 6.2.b. An error is thus introduced and one should make sure its effect does not impair the final desired results.

$$a)\ out/in = -Z_{fb}/Z_{in} \qquad b)\ out/in = -Z_{fb}/Z_{in} \cdot \frac{1}{1+1/A+Z_{fb}/(Z_{in} \cdot A)} \quad (6.2)$$

Because of the way they are implemented ideal capacitors are always surrounded by parasitic capacitors. An extremely simplified view of a poly1-poly2 capacitor is given in figure 6.3. As illustrated, the ideal capacitor is defined by the two poly electrodes while poly1 to substrate and poly2 to substrate form the parasitic ones. C_{p1} and C_{p2} are about one twentieth of C_{id}.

As far as switches are concerned, one must be careful that the "on" resistance (Ron) and clock feed through (see 3.2.3) do not causes undesirable effects.

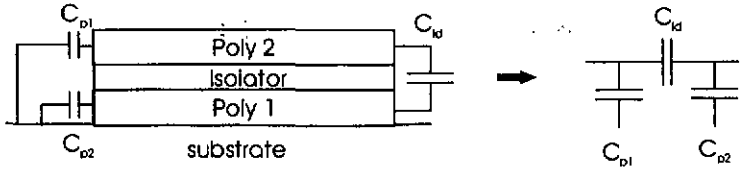


Figure 6.3: Simplified view of poly-poly capacitor

For each proposed scheme, the effects of finite gain, parasitic capacitors and Ron will be evaluated. The clock feed through effect os well as other non ideal behavior will only be discussed for the implemented solution.

6.2.2 Capacitors network schemes

6.2.2.1 Parallel capacitors I

This scheme is one of the simplest. It consists of a simple inverting amplifier scheme where the total amplification is set by properly switching S_1 to S_7. The S_{rs} switch is used to reset the system. Such a structure uses a single active element, which is clearly a benefit from a power consumption point of view.

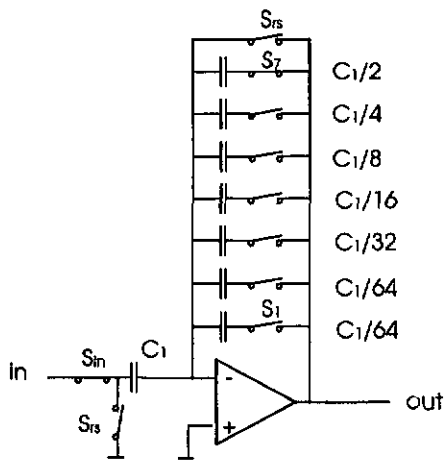


Figure 6.4: Inverting amplifier with parallel capacitors network

The particular switch configuration of figure 6.4 results in amplification by 2. To obtain a unity gain, all the feed back capacitors must be used in parallel.

The effect of limited OTA gain is not dramatic. Indeed, with a finite gain of 60 dB, equation 6.2.b for amplification (out/in) smaller or equal to 64 results in a close to linear characteristic with a 93.8% slope. The limited gain thus produces a linearity error of 6.2%. As explained in section 4.4, this is tolerable for the audio converters.

Let us consider that all the switches have the same size. The effect of the on resistances R_{on} is to slightly modify the transfer function as in equation 6.3, where the OTA gain is ideal (infinite). The worst case is when ν=8000 (input signal highest frequency) and to be disregarded, R_{on} must be definitely smaller than 6x10^6.

$$\frac{out}{in} = \frac{2/(C_1 \cdot 2\pi \cdot \upsilon) + R_{on}}{1/(C_1 \cdot 2\pi \cdot \upsilon) + R_{on}} \cong \frac{2/(C_1 \cdot 2\pi \cdot \upsilon)}{1/(C_1 \cdot 2\pi \cdot \upsilon)} = 2 \qquad (6.3)$$

A modified scheme, including the parasitic capacitors, is given in figure 6.5. In the ideal case (infinite gain, no OTA offset), the transfer function is not modified and is still two. Nevertheless, C_{p4} and C_{p1} are extra loads, which slow down the charge on C_1 and $C_1/2$.

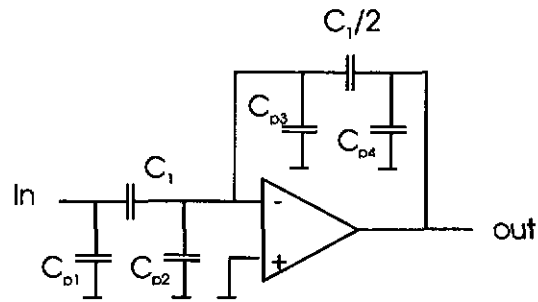


Figure 6.5: Parasitic capacitors in inverting scheme

The above parallel capacitor scheme relies on capacitor ratios and is thus particularly sensitive to capacitor matching. Normally, for two identical capacitors, matching values of about 0.2% are obtained, although the quality does drop down when different capacitors values are used. In the considered structure, the worst case arises for the ideal gain value of 64 and the capacity mismatch provokes a maximum gain value error of 1.8%. This is within the tolerance of section 4.4, however it is worthwhile analyzing other schemes that might result in less sensitivity to matching problems.

6.2.2.2 Serial capacitors

To reduce the matching problem identical capacitors should be used. One solution is presented in figure 6.6 where the feed back capacitors are serial and the desired gain value is obtained by proper switching of S_i.

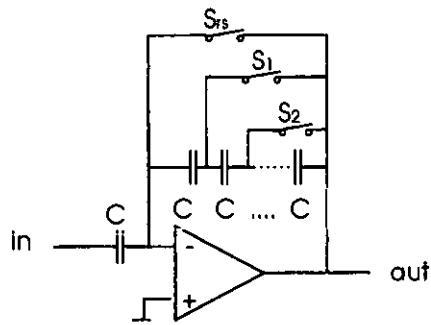


Figure 6.6: Serial capacitor scheme

Let us assume a gain of two is desired and S_2 is closed. Figure 6.7 shows the ideal (top) and real (bottom) schemes and their equivalent single capacitor circuits. The real equivalent capacitor is thus 1.909. Since a unity capacitor is used at the front end a real amplification gain value of 1.909 is reached.

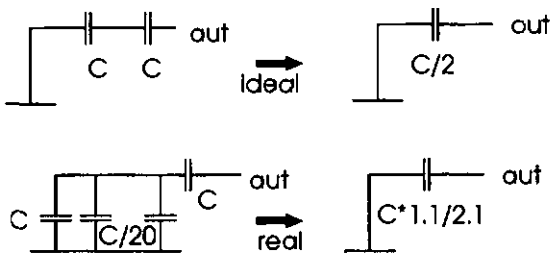


Figure 6.7: Parasitic capacitors in serial scheme

For a desired gain of 4, parasitic capacitors actually limit the real value to 3.065. The situation gets worse with a desired amplification value of 64: the real gain value is only 3.7011 The effect of the parasitic capacitors is thus dramatic and schemes using serial capacitors should thus be avoided.

Another disadvantage of an identical capacitor scheme is the huge resulting area (65 capacitors).

6.2.2.3 Parallel capacitors II

A way to lighten the matching problem of 6.2.2.1 while keeping a reasonable area is to have a parallel network at both the input and feed back points. As a result the ratio between the biggest and smallest capacitor is reduced to 16 (instead of 64).

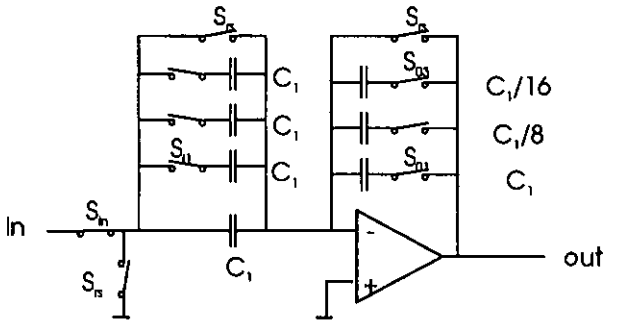


Figure 6.8: Inverting amplifier with 2 parallel capacitors networks

The effects of limited OTA gain and an resistance at the switches are identical to those explained in 6.2.2.1. Again, the parasitic capacitors only contribute to slowing down the load characteristic. The effect might be more pronounced because of the bigger parasitic input and/or feed back capacitance.

The main draw back of this scheme is that the OTA load is larger (compared to the scheme of 6.2.2.1). This means that to maintain the slew-rate performance, a higher current must be used in the OTA, which results into higher power consumption.

6.2.3 Switched capacitor schemes

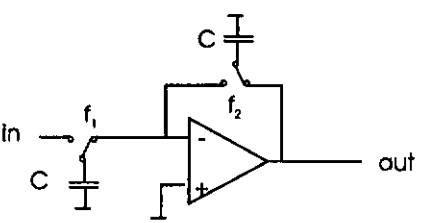


Figure 6.9: Switched capacitors solution

The idea is to use an inverting amplifier with switched capacitors instead of resistors. The amplification gain is then set by the switching ratio t_2/t_1. This seems

particularly interesting since precise clock division can be performed in the digital domain and only two identical capacitors are required.

Such a scheme is nevertheless not feasible ! Indeed, the instantaneous behavior of the switched capacitor is not equivalent to that of a resistor. As a result, because of the switched capacitor in the feed back loop, the above circuit is completely unstable!

6.2.4 Integrating schemes

6.2.4.1 Simple integrator

The simple integrator scheme uses a time duration variation (interval t during which the input switch is closed) to control the amplification gain.

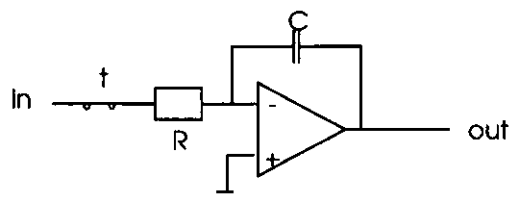


Figure 6.10: Simple integrating solution

Assuming the input is stable, the transfer function is given by equation 6.4.

$$\frac{out}{in} = -\frac{t}{RC} \qquad (6.4)$$

To obtain the proper gain values equation 6.5 must be satisfied. Ideally "Cst" should be 1. If another value, for example a multiple of 2, was used, proper handling at the back end would be required. The main limitation of such a scheme is clearly the poor absolute precision of such an RC product (typically 40% !). To overcome the problem, a switched capacitor can be used instead of the resistor.

$$t = \tau, 2\tau, 4\tau, 8\tau, \cdots 64\tau \qquad \tau/RC = Cst \qquad (6.5)$$

Assuming the maximum integration time (for a 64 gain value) is as long as 1/fs (fs=16000), τ becomes 1/fs/64, that is about 970 ns. This is also the minimum interval during which the input switch can be closed (for a 1 gain value). The input capacitor must thus be switched at a frequency a lot higher than $1/\tau$ (typically 20-

50 MHz). Such a solution results in delicate design constraints on the OTA and analog switches. Furthermore, the parasitic capacitors tend to degrade the overall performance [Degr94]. More complex integrator schemes, insensitive to the parasitic capacitor, exist though they require an extra active element and thus result in higher power consumption.

6.2.4.2 Integration by charge transfer

In the scheme of figure 6.11, each time the switch performs one full cycle, the input voltage is transferred to the feed back capacitor. Hence, the number of switch cycles defines the global amplification gain between input and output.

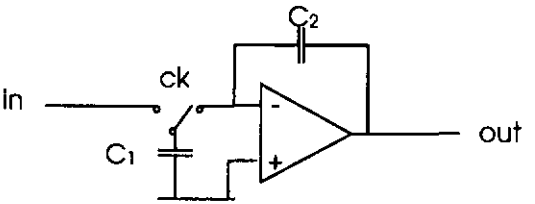


Figure 6.11: Charge transfer scheme

Assuming a symmetrical clock ck, the same amount of time is provided to load the input on C_1 and to transfer the charge to C_2. If C_1 is identical to C_2, 64 cycles are necessary in the worst case amplification and thus the minimum clock frequency is a little more than 1 MHz. Another solution is to use $C_1=2C_2$ and provide a direct path for the unity gain. This decreases the minimal clock frequency by a factor of two and results in easier design constraints for the OTA. They are also facilitated by the fact that a smaller capacitor can be used. Indeed, accumulation results in a lower noise level on the capacitors as shown by equation 6.6 where N is the number of accumulation cycles and $noise_1$ the noise level after one cycle.

$$noise_N = \frac{noise_1}{\sqrt{N}} \qquad (6.6)$$

The above scheme is sensitive to offset and clock feed-through.

This scheme seems a valid alternative to the simple inverting circuit. The gain values are certainly more precisely controlled. However, because of the required sample and hold stage (see next section) the power consumption is much bigger.

6.2.4.3 Sample and hold

Both integrating schemes of sections 6.2.4.1 and 6.2.4. require that the input is stable. A "pre-sampling" must thus be performed and the level maintained during the complete conversion time. This is achieved by the following sample and hold scheme :

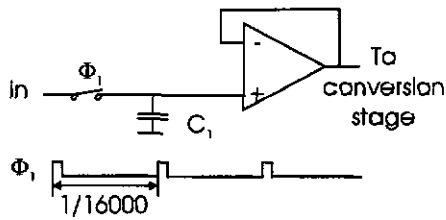


Figure 6.12: Sample and hold for pre-sampling

To avoid error propagation, the follower's offset must be compensated. Also, a double differential pair (nMOS and pMOS) structure must be used to maintain close to rail-to-rail input and output. The latter can be realized by using two OTAs. Integrating solutions have therefore a minimum number of three OTA (two for pre-sampling, at least one for conversion), and as a result higher power consumption is obtained.

6.2.5 Design of the controlled amplifier

In the previous sections, different schemes to implement the controlled gain were discussed. Two circuits seemed particularly well suited : the parallel capacitor I and the integration by charge transfer. The latter is more power demanding since three high slew rate OTAs are required. The first is simpler but relies on capacitor matching.

Based on the power consumption argument, it was decided to implement the parallel capacitor scheme of figure 6.13. S_{is} has a connection to ground and so NMOS switches are sufficient. On the other hand, complementary switches [Degra94] must be used for S_{in}, S_1, ..., S_7. The resistor R reduces the band pass and consequently limits the high frequency noise.

The scheduling of the controlled amplifier is presented in figure 6.14. 'Bit #' indicates the binary bit computed in the coarse quantizer (a modified A. Heubi RSD A/D converter, see 5.1) whose basic clock Ck is also plotted. The sixth bit is known at 'A' and as explained in 6.4 it is also the moment when the adaptation logic can start computing the control signals for S_1 to S_7. The latter are then set at

'B'. 'C' is the beginning of the conversion cycle ; the analog signal is sampled on the input capacitor of the coarse quantizer.

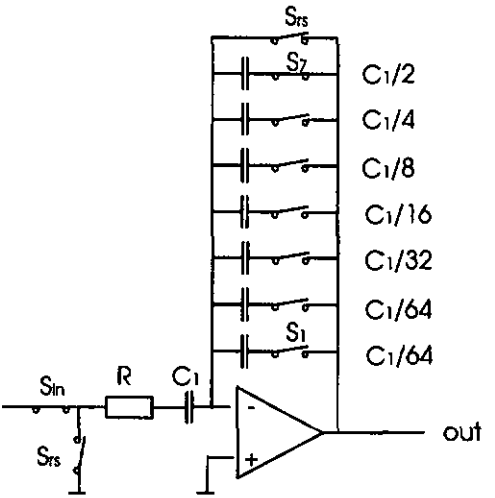


Figure 6.13: Implemented scheme

Switches are assumed to be closed when signals are high. Two phases can be acknowledged : first, to ensure proper reset, S_{in} is open while all the other switches are closed ; second, amplification takes place and S_{in} is closed while S_1 to S_7 are set according to the adaptation logic.

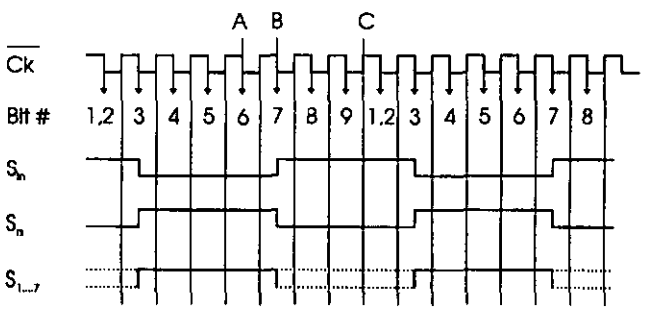


Figure 6.14: Scheduling of controlled gain

The design of the controlled gain according to the overall specifications and scheduling is explained below. The parameters have the following values (ALP2) :

V_{dd}=1.3 [V] V_{ss}=-1.3 [V]

$n_n = 1.44$ $n_p = 1.78$

$\beta_{sn} = 53^{e-6}$ [A/V²] $\beta_{sp} = 18^{e-6}$ [A/V²]

T = 300 [K]

6.2.5.1 Capacitor C_1

Thermal noise is sampled on C_1. The capacitor must be sized to ensure a noise level smaller then a half LSB.

$$\sqrt{\frac{KT}{C_1}} < \frac{1}{2} LSB = 0.5 \cdot 1.3 \cdot 2^{-14} \cong 39\mu V \tag{6.7}$$

$$C_{1min} = \frac{KT}{(39 \cdot 10^{-6})^2} \cong 2.8pF \tag{6.8}$$

Taking into account the nominal error (about 15% in ALP2), a minimal capacitor of 3.4pF must be used.

At a layout level, to decrease matching errors, all the capacitors are realized from basic element at 1/64. Furthermore, classical "gravity center" techniques are used.

6.2.5.2 Resistor R

The resistor R slows down the load characteristic on C_1. It must be chosen so that the load error remains smaller then a half LSB.

$$error = V_{dd} \cdot e^{\frac{-t}{RC}} < 0.5 \cdot LSB \qquad t = \frac{5}{16000 \cdot 2 \cdot 8} \tag{6.9}$$

$$R_{mx} = \frac{-t}{\ln\left[\frac{error}{V_{dd}}\right] \cdot C_1} \Rightarrow R_{mx} \cong 500k\Omega \tag{6.10}$$

Equation 6.10 gives the maximum value. Using a smaller one leads to a lower load error though the resulting band pass is broader and consequently the noise level higher. Simulation with the OTA showed that a value of 400 kΩ was sufficient to meet all the requirements.

6.2.5.3 OTA

A two stage OTA shown in figure 6.15 and described in 3.2.2 is used. The specifications are the equivalent input noise and DC gain. For the latter, a minimum value of 60 dB is targeted. A higher value will nevertheless result in better linearity. As for the equivalent input noise, the level is set to 50 µV. This value derives from a parallel project where the converter is used in combination with a microphone and where the target noise increase due to the OTA was set to 15%.

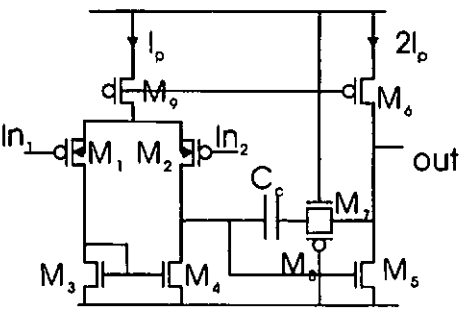


Figure 6.15 : Class A 2-stage OTA

Preliminary studies showed that a polarization current of 1.4 µA is required. The following equations detail the determination of the various transistor sizes according to the methodology presented in [Degr94]

The differential pair works in weak inversion and a I_{M1}/I_{lim} factor of 1/5 is used. From 6.11 and to facilitate the layout, L_1 is set to 2µm (minimum size) while W_1 is set to 168 µm.

$$I_{M1} = I_D/2 \qquad I_{lim} = 2 \cdot \beta_{sp} \cdot \frac{W_1}{L_1} \cdot n_p \cdot U_T^2 \tag{6.11}$$

$$\frac{W_1}{L_1} \geq 81.4 \qquad W_1 = W_2 \qquad L_1 = L_2 \tag{6.12}$$

M3 and M4 work in strong inversion which implies a maximum W/L ratio of 1.56 as computed in 6.12 (using a I_{M3}/I_{lim} factor of 5). This ratio also influences the final equivalent input noise and since the resistor value is used in the calculation, two extreme cases are considered : a maximum resistance of 550KΩ and a minimum one of 100KΩ.

$$\frac{W_3}{L_3} = \frac{I_p}{20 \cdot \beta_{sn} \cdot n_n \cdot U_T^2} = 1.56 \tag{6.13}$$

$$Svin_1 = \sqrt{V^2/BP} = \sqrt{(50\mu)^2 \cdot 2 \cdot \pi \cdot 550k \cdot 3.4p} = 1.71 \cdot 10^{-7} \quad \left[V/\sqrt{Hz}\right] \tag{6.14}$$

$$Svin_2 = \sqrt{(50\mu)^2 \cdot 2 \cdot \pi \cdot 100k \cdot 3.4p} = 7.31 \cdot 10^{-8} \quad \left[V/\sqrt{Hz}\right] \tag{6.15}$$

$$Svin = \frac{4KT}{g_{m1}}\left[n_p + \frac{4}{3} n_n \cdot \frac{g_{m3}}{g_{m1}}\right] \tag{6.16}$$

$$\left.\frac{W_3}{L_3}\right|_{mx2} = \frac{9 \cdot g_{m1}^2}{16 \cdot \beta_{sn} \cdot n_n \cdot I_p}\left[\frac{g_{m1} \cdot Svin_1^2}{4KT} - n_p\right] \cong 800 \tag{6.17}$$

$$\left.\frac{W_3}{L_3}\right|_{mx3} = \frac{9 \cdot g_{m1}^2}{16 \cdot \beta_{sn} \cdot n_n \cdot I_p}\left[\frac{g_{m1} \cdot Svin_2^2}{4KT} - n_p\right] \cong 12 \tag{6.18}$$

Strong inversion is thus the limiting constraint and the ratio value is set to 1. M4's ratio is identical to that of M3.

The output current is chosen to be twice I_p and both M_5 and M_6 work in strong inversion. The W/L ratio of M_5 is thus 4 while that of M_6 is obtained from 6.19.

$$\left.\frac{W_6}{L_6}\right|_{mx} = \frac{I_p}{5 \cdot \beta_{sp} \cdot n_p \cdot U_T^2} = 13.03 \quad \Rightarrow \quad \frac{W_6}{L_6} = 10 \tag{6.19}$$

To go on with the design, the lengths of all the transistors but M_1 and M_2 should be computed. Even assuming $L_3 = L_4 = L_5 = L_6$ the problem remains complex since they strongly influence both the DC gain and stability. The latter also depends on M_7, M_8 and C_c which still haven't been considered. At this point the easiest way to proceed is to perform several simulations, changing the lengths and widths of M_3, M_4, M_5 and M_6 according to the above ratios while tuning M_7, M_8 and C_c. The use of equations as in [Saus97] is not beneficial since at the end several simulations are still necessary to finely tune the OTA.

The circuit providing the highest DC gain and phase margin was selected and is given in table 6.1, where I is the theoretical current in µA and Op the operation mode (WI :Weak inversion, SI :Strong inversion). C_c is 0.3 pF.

The simulated performance (obtain with XPERTSIM, CSEM) is given in table 6.2. A high DC gain is obtained as well as close to rail-to-rail output swing. The phase margin is rather small but simulations showed that no impairing effects should result.

M	Type	W[µm]	L[µm]	i[µA]	Op
M1	P	168	2	0.7	WI
M2	P	168	2	0.7	WI
M3	N	5	5	0.7	SI
M_4	N	5	5	0.7	SI
M_5	N	20	5	2.8	SI
M_6	P	50	5	2.8	SI
M_7	N	5	20	0	linear
M_8	P	5	20	0	linear
M_9	N	25	5	1.4	SI

Table 6.1 : OTA transistor sizes

Open loop gain	66.8 dB
gain bandwidth	4.6 MHz
phase margin	30°
offset	0.0003 V
output swing low	-1.247 V
output swing high	1.29 V
input noise	4.6 10^{-6} V/Hz$^{0.5}$
slew-rate	3 10^{6} V/s
power dissipation (@±1.3 V)	15.5 µW

Table 6.2 : OTA characteristics

6.2.5.4 Current source

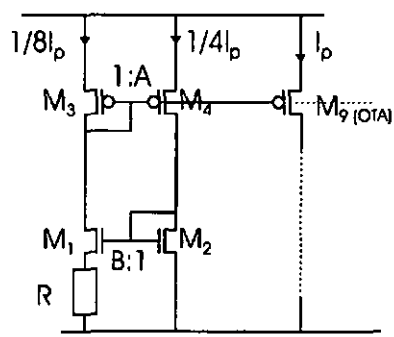


Figure 6.16 : Current source

The current source presented in figure 6.16 is used to polarize the OTA. M_3 and M_4 work in strong inversion while M_1 and M_2 work in weak inversion. As explained in [Degr94], I_p is proportional to U_t and thus $gm_{1,2}$ (trans-conductance of the OTA) becomes independent of the temperature. Furthermore, if a poly resistor is used, the current becomes independent, in first approximation of the supply voltage as

well. The main drawback is the poor precision of the integrated resistor value and consequently of the current value ($\Delta I/I \approx 10\%$).

$$I_p = \frac{8 \cdot \ln(A \cdot B) \cdot U_T}{R} \qquad g_{mi} = \frac{I_p}{2 \cdot n_p \cdot U_T} \tag{6.20}$$

M	Type	W[µm]	L[µm]	I[µA]	Op
M_1	N	80	2	$I_p/8$	WI
M_2	N	40	2	$I_p/4$	WI
M_3	P	25	40	$I_p/8$	SI
M_4	P	25	20	$I_p/4$	SI

Table 6.3 : Current source transistor sizes

The sizes of the transistors, computed according to the WI and SI conditions and using A=B=2, are given in table 6.3. A resistor value of 400 kΩ was used.

6.2.5.5 Switches

The only elements left are the analog switches which don't need careful design as long as their on resistance is sufficiently low. Their length is set to 2 µm and their width to 10µm, resulting in a maximum on resistance of 35 kΩ.

6.2.5.6 Errors and parasitic effects

Finite DC gain and on resistance have already been discussed. Their effect does not impact the final result. Other non ideal characteristic such as OTA offset, leakage current, charge injection etc. must still be considered. One of their possible resulting effects is to contribute to the input offset. This is not a major problem since digital correction can be done after conversion. In all other cases however, compensation is more complex.

The offset of the OTA simply results in a DC offset at the output of the floating point A/D.

Leakage current, due to the 'parasitic' diodes between bulk and drain (source) are present in any switch. The model of figure 6.17 can be used and each source's effect considered separately.

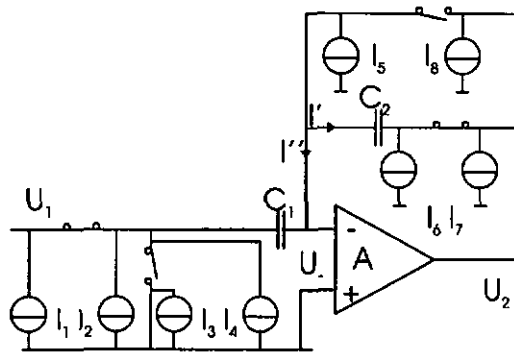


Figure 6.17: Model for leakage current

I_1, I_2, I_3 and I_4 have no effect since they have a direct path to ground. The current I_5 is divided into I' and I''.

$$U_2 = U_- - \frac{1}{C_2}\int I'dt \qquad U_2 = -AU_- \tag{6.21}$$

$$U_- = \frac{1}{C_1}\int I''dt \tag{6.22}$$

Combining 6.21 and 6.22 and assuming A>>1 equation 6.23 is obtained.

$$I'' = \frac{C_1}{C_2}\cdot\frac{1}{A}I' \qquad I_5 = I''+I' = I'\left(1+\frac{C_1}{C_2A}\right) \tag{6.23}$$

U_2 is thus given by 6.24 and its equivalent input voltage (in other words, the voltage that should be applied at the input to result in U_2) by 6.25.

$$U_2 = \frac{1}{C_1}(1+A)\int I''dt \cong \frac{1}{C_1}A\int I''dt = \frac{1}{C_2}\int I'dt \tag{6.24}$$

$$U_{equ} = U_2\frac{C_2}{C_1} = \frac{1}{C1}\int I'dt \xrightarrow[\frac{C_1}{C_2}<<A]{} U_{equ} = \frac{1}{C1}\int I_5dt \tag{6.25}$$

If the finite OTA gain is sufficiently high, the equivalent voltage is independent of C_2 and its effect can be considered as a supplementary offset.

If they weren't absorbed by the OTA, I_6, I_7, and I_8, would raise the output level as well as U_- (by loading C_1). However, this would be in contradiction to the

fundamental working of the OTA (if U_2 raises, U_- must decrease). These currents are thus absorbed and have no influence on the output voltage.

To sum up, although most of the leakage currents have no effect, the final contribution is equivalent to a supplementary input offset.

Another important parasitic effect is the charge injection due to the opening of the switches. Fortunately, the OTA's negative node capacitance load is constant, independent of the configuration of the switches. As a result, the overall effect is again a constant input offset.

Some of the parasitic capacitors as well as matching errors have already been considered (6.2.2.1). However there are certainly other parasitic capacitors which have not been considered yet (such as interconnection etc.). Their effects are complex to evaluate but they will contribute to the input offset and/or modify the amplification value. In the latter case the size of the feed back capacitors can be slightly modified to compensate for the errors and to obtain the proper amplification factor. This is achieved by repeatedly modifying the layout and simulating the amplification value.

6.2.6 Simulation results

Figure 6.18 shows the output of the controlled gain (o) for a fixed amplification of 16 and a 0.01V step input signal (+). The simulation lasts a little more than two amplification phases.

The first amplification occurs after 40 µs and since at that moment the input signal is null, the output actually corresponds to the input offset multiplied by 16. The input offset is thus about 3 mV (0.0479 V / 16).

At about 70 µs, all the switches are closed and the reset phase takes place. Oscillations due to the relatively low phase margin of the OTA are present, though the reset is long enough to ensure stabilization.

The second amplification starts at 0.1 ms and the output level falls to -0.112V. The real amplification value is thus given by equation 6.26.

$$\frac{-0.0479 - 0.112}{0.01} = 15.99 \qquad (6.26)$$

The non horizontal slope at the end of the amplification phase is due to leakage currents. Their total amount can be estimated by 6.25.

$$\frac{dU}{dt}C_{16} = I_{leak} \qquad 1.9\cdot 10^{2}\frac{3.4\cdot 10^{-12}}{16} \cong 4\cdot 10^{-11} \quad (A) \tag{6.27}$$

A noise simulation has also been performed and showed a noise level of 4 mV for a 64 amplification value. This value is slightly smaller than a LSB and thus greater then the targeted half LSB. One can expect that the dynamic range will be reduced to less then 15 bits.

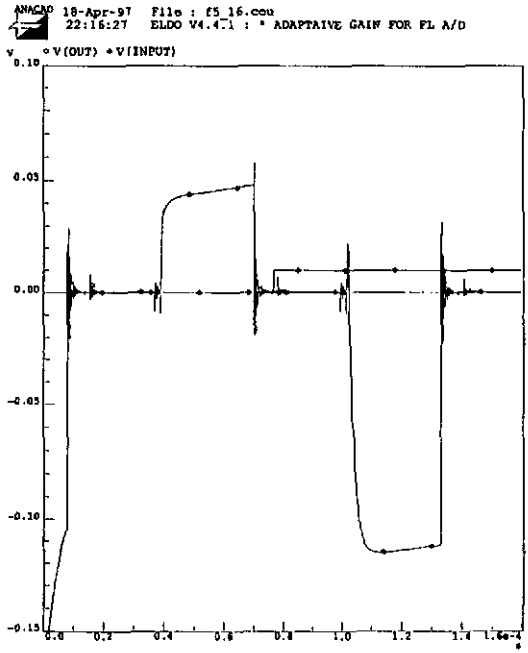


Figure 6.18 : Simulation of the controlled gain

6.2.7 Wrap up

Various solutions have been considered to implement the controlled gain. However, only two seemed to fulfill all the requirements : the inverting amplifier with parallel switchable feed back capacitors and the integrating by charge transfer. The final choice was to use the first: this simpler scheme relies on capacitor matching and results in a ultra low power implementation. The design was performed for ALP2 LV and the predicted power consumption is about 20 µW

@ ± 1.3V and 16 kHz. Because of a slightly higher internal OTA noise, the dynamic range is limited to about 14 bits.

6.3 COARSE QUANTIZER

The RSD conversion algorithm discussed in section 5.1 is used to realize the coarse quantizer. In particular, the scheme proposed by A. Heubi is considered although a redesign, following a methodology similar to that of section 6.2.5 was performed to take into account that only 9 bits are required. The simulated power consumption is 16 µW @ ± 1.3V and 16 kHz. The design is not detailed since it is covered by a patent (see section 5.1 for detail).

6.4 ADAPTATION LOGIC (RESERVE BIT CONVERTER)

The purpose of the adaptation logic is to provide the control signals to the switches S_1 to S_7 in the controlled gain (see figure 6.13) and to the output shifter. This function can be decomposed into successive tasks as defined in figure 6.19. To be precise, task 6 should not be considered as part of the adaptation logic but rather as the back end of the conversion. It will nevertheless be discussed here.

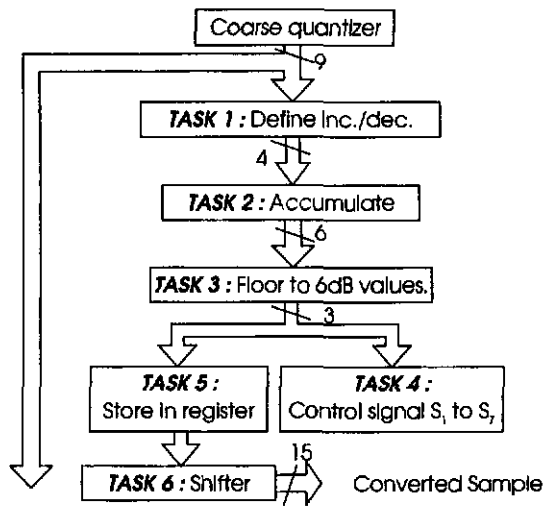


Figure 6.19: adaptation logic tasks

6.4.1 Task 1 : Increment/decrement computation

The increments/decrements are defined by table 6.4 which is obtained by extending Jayant's table and adding a reserve bit (see 4.2.1.2). The table also shows the equivalent coarse quantizer and inc./dec. codes. The latter are coded using a two's complement representation to facilitate the accumulation (task 2). Note that to determine the inc./dec. value, only the 6 most significant bits of the coarse quantizer output must be known. This explains the scheduling proposed in figure 6.14.

magnitude of coarse quantizer	coarse quantizer equivalent code	inc./dec. [dB]	inc./dec. equivalent code
	$b_8\ b_7$ ..... b_0		$b_3\ b_2\ b_1\ b_0$
0 .. 63	000XXXXXX or 111XXXXXX	+ 1.5	0001
64 .. 71	001000XXX or 110111XXX	- 1.5	1111
72 .. 79	001001XXX or 110110XXX	- 3.0	1110
80 .. 95	00101XXXX or 11010XXXX	- 4.5	1101
96 .. 111	00110XXXX or 11001XXXX	- 6.0	1100
112 .. 119	001110XXX or 110001XXX	- 7.5	1011
120 .. 127	001111XXX or 110000XXX	- 9.0	1010
128 .. 255	01XXXXXXX or 10XXXXXXX	- 9.0	1010

Table 6.4 : Increment/decrement table

Task 1 is thus simply realized by combinational logic that maps the second column entries into the 4th column outputs.

6.4.2 Task 2 and 3 :accumulation and floor to 6dB multiples

The accumulated gain values range from 0 to 36 dB with 1.5 dB steps. This means that 5 bits are needed for their coding. However, if a sign bit is provided, accumulation to negative values can easily be detected. To simplify the adaptation logic a code consistent with the last column of the previous table must be used. As a result, the 6 dB multiple gain values are coded as in table 6.5. Only b_4, b_3 and b_2 must be transferred to the next task.

Gain [dB]	equivalent code $b_5\ b_4\ .....\ b_0$
<0	1XXXXX
0	000000
6	000100
12	001000
18	001100
24	010000
30	010100
36	011000

Table 6.5 : Gain coding

The scheme of figure 6.20 performs the accumulation, checks the limit conditions (gain not smaller then 0 and not higher then 36 dB) and floors the result to 6 dB values. To facilitate the test and/or to use an external adaptation logic, multiplexers are provided.

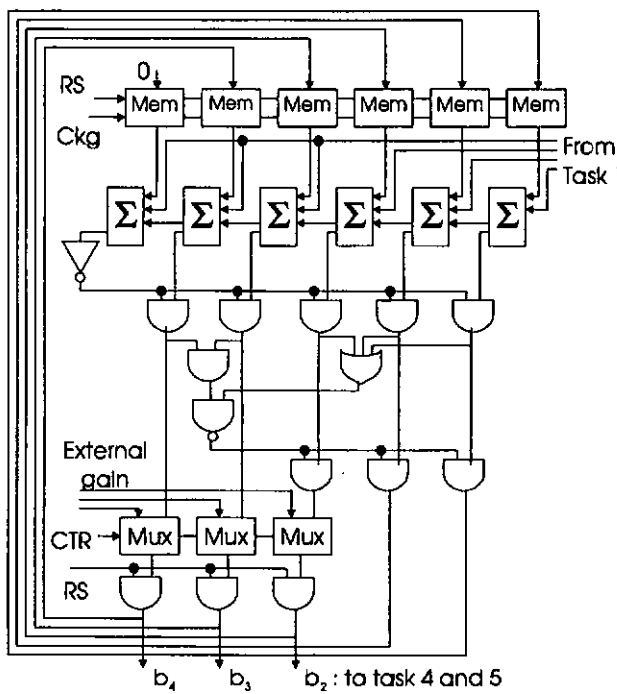


Figure 6.20: Detail of task 2 and 3

6.4.3 Task 4 : Control signal S_1 to S_7

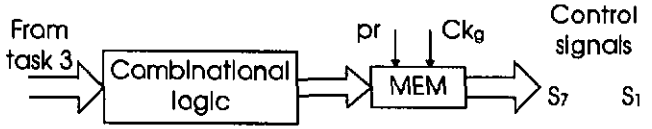


Figure 6.21: Detail of task 4

Computing the control signals for the switches is straightforward and can be achieved with combinational logic. Table 6.6 gives the relation between the N MOS switches S_1 to S_7 and the gain bits b_2 b_1 b_0 resulting from task 3. The control signals are stored in MEM (figure 6.21) to ensure they remain constant over the whole amplification cycle. The « pr » control signal is used to close all the switches during the reset phase.

From task 3 b_4 b_3 b_2	S_1	S_2	S_3	S_4	S_5	S_6	S_7
000	1	1	1	1	1	1	1
001	0	0	0	0	0	0	1
010	0	0	0	0	0	1	0
011	0	0	0	0	1	0	0
100	0	0	0	1	0	0	0
101	0	0	1	0	0	0	0
110	0	1	0	0	0	0	0

Table 6.6 : Combinational logic for switches control

6.4.4 Task 5 and 6 : Memorizing and shifting

Because of the 'one sample' delay between the computed value in task 3 and the one to be applied to the shifter, the gain must be memorized. This is performed by task 5 in figure 6.19.

The shifter at the back end (task 6) is simply realized by cascaded multiplexers which are controlled by b_4, b_3 and b_2 of task 3.

6.4.5 Simulation of adaptation logic and shifter

Figure 6.22 shows the adaptation logic simulation results, where « sync » and « nrsg » are external control signals (see 6.5). Since the simulation file only includes the adaptation logic, the « ob » signal, which should be the coarse quantizer

output, is also specified externally. This is done in a such a way to ensure that during the 17 first « sync » cycles a small value i.e. 000101 is fed to the adaptation logic while from the 18th onwards, a bigger one i.e. 011011 is applied. From table 6.4, these values respectively result in a +1.5 dB and -9 dB gain increment.

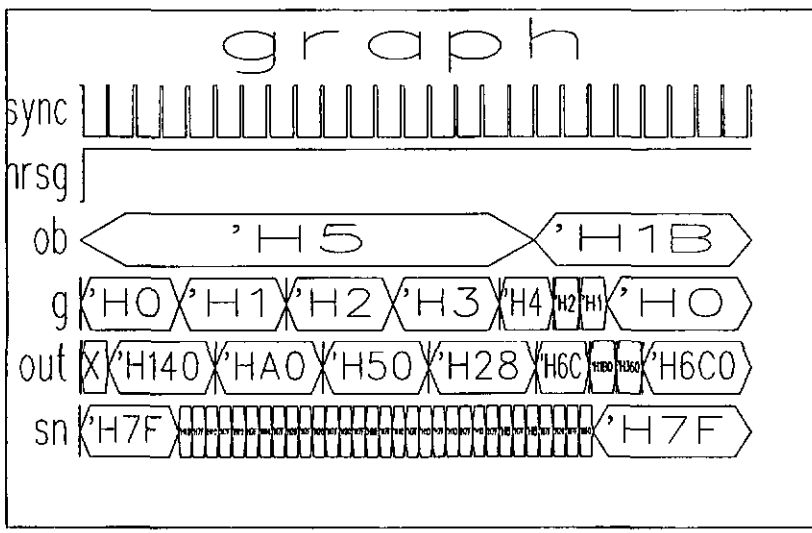


Figure 6.22 : Simulation of adaptation logic and shifter

The 6 dB gain value is given by « g ». As expected, while « ab » is small, the gain value increases steadily each 4th « sync » from 0 dB (H0) to 24 dB (H4). As « ob » gets bigger, the gain value drops down to 12 dB (H2), then to 6 dB (H1) and finally to 0 dB (H0). This is indeed the expected behavior since before dropping the actual accumulated value was 25.5 dB (thus floored to 24 dB). After the first decrement, this value became 16.5 (floored to 12), and 7.5 (floored to 6) after the second one. At the third decrement the limit value of 0 dB is obtained.

The output of the shifter is given by the « out » signal. Table 6.7 computes its expected value considering « ob » and the applied gain. The good news is that « out » and the last column are identical!

The last signal « sn » is the NMOS switches command. First the signal must force the reset phase and thus all switches S_1 to S_7 must be closed. As a result « sn » is 1111111 (H7F). To ensure the proper amplification value « sn » must then be set

according to table 6.6 i.e. H7F, H40, H20, H10, H8, H4 and H2 for 0, 6, 12, ... 36 dB respectively. In the simulation, « sn » and « g » are coherent.

« ab » (binary)	applied gain	shifter output (b)	shifter out (H)
000000101	0 dB	000000101000000	140
000000101	6 dB	000000010100000	A0
000000101	12 dB	000000001010000	50
....	....	....	....
000011011	24 dB	000000001101100	6C
000011011	12 dB	000000110110000	180
....	....	....	....

Table 6.7 : Expected shifter's output

To make a long story short, the simulation confirms the proper working of the adaptation logic and shifter!

6.5 ANALOG AND DIGITAL CONTROL

All the control signals for the adaptation as well as those for the coarse quantizer are generated from the 3 external signals "sync", "ck" and "nrsg". The timing diagram of the A/D chip is given in figure 6.23. As sync rises, sampling takes place. Eight clock cycles are then necessary to compute the 9 binary bits (or 8 RSD bits) in the coarse quantizer. The frequency of the clock is thus 8x16kHz=108 kHz. Just before the first bit of the next sample is obtained, the output is updated with the new converted value. A reset can be performed using nrsg which is active low (not shown in 6.23).

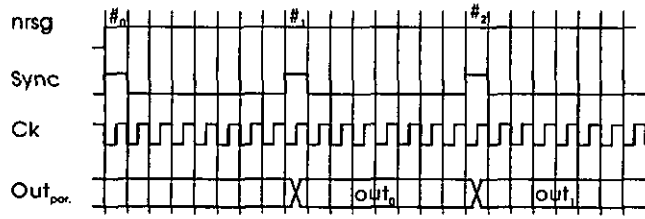


Figure 6.23: Timing diagram of A/D chip

There is no major difficulty in generating the control signals. One must only be careful to ensure non overlapping clocks for the switched capacitor coarse quantizer. Combining these signals with the three external ones and taking into account the working sequence of each block, the control signals can be

generated. To avoid loud snoring, they will not be described in detail. Nevertheless, they can be found in Annex 1 at the end of the chapter.

6.6 CHIP PRESENTATION

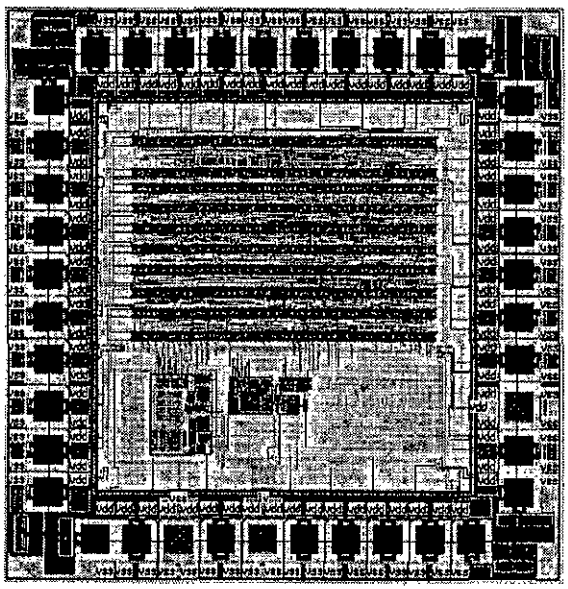


Figure 6.24 : Floating point ADC chip

Figure 6.24 shows the 15 bits floating point A/D chip in a low voltage 2 µm CMOS technology. The digital part (very dense, on top of layout) is realized using CSEL_LIB, the low power library from CSEM SA while the analog one (lower part of layout) is full custom.

The chip has 40 pads though only 22 are actually necessary for basic operating: 15 pads for the ADC output, 3 for power supply and analog ground, 3 more for external control (ck, sync and nrsg), one for analog input. All the remaining pads are useful for testing purposes as well as to modify the adaptation strategy.

The total die size (without pads) is close to 1.4mm by 1.4 mm (less then 2 mm^2) and the simulated power consumption is about 47 µW @ 16 kHz and ± 1.3 V. Percentage of total die size (Serie 1) and total power consumption (Serie 2) for each main functions are given in figure 6.25.

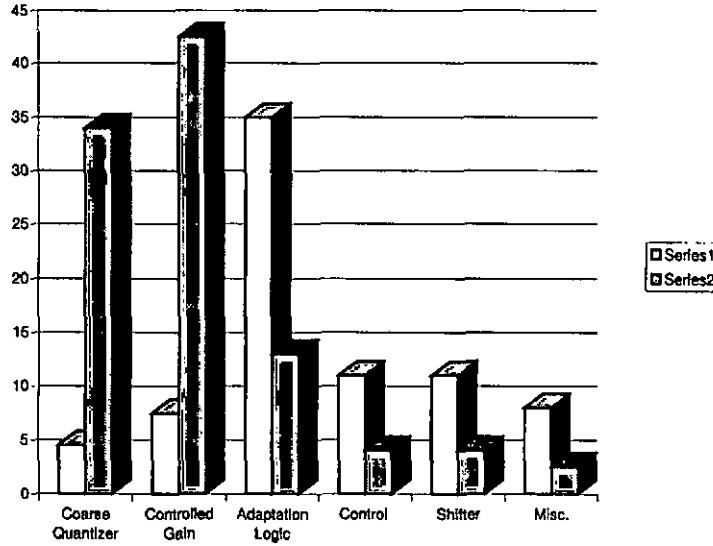


Figure 6.25 : Power and die balance of floating point ADC chip

Summing up, these numbers results in the following remark : 12 % of the die size is used by analog parts which however consume 76.5 % of the power ! This also means that by changing the adaptation strategy and thus modifying the digital part, the area might decrease but little improvement can be expected from the power consumption perspective.

6.7 6 dB ADAPTATION STRATEGY

As explained in 4.2.2.2, a simpler 6 dB adaptation strategy can be used and corresponds to table 6.8.

magnitude of coarse quantizer	inc./dec. [dB]
0 .. 31	+ 6
32 .. 63	0
64 .. 95	- 6
96 .. 159	- 12
160 .. 255	- 18

Table 6.8 : 6 dB increment/decrement table

The testing features of the floating point ADC chip allow the user to output the 6 MSBs of the coarse quantizer and externally specify the gain value by entering right after task 3 (see figure 6.19). All the tasks in between are removed. Thus to obtain a « 6 dB version » of the floating point converter, the new adaptation strategy was implemented on a FPGA and successfully connected to the chip.

The 6 dB adaptation logic can be realized by the simple structure of figure 6.26. In an ASIC implementation it would replace task 1, 2 and 3 of figure 6.19. The complexity of this solution is clearly lower then the previous one and less hardware is required. As a result, if this solution was implemented in the ALP2 technology, improvements of 12 % in die size and 4 % in power consumption would be obtained.

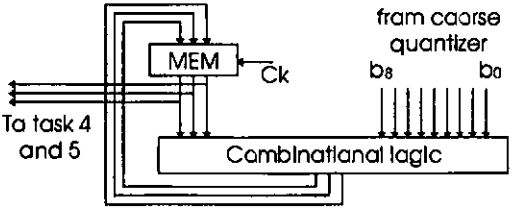


Figure 6.26: new adaptation implementation

6.8 FEED FORWARD CONVERTER

No feed forward implementation has been performed. Nevertheless, by extrapolation and using the numbers of figure 6.25 a fair estimate of the power consumption and area required by such a converter can be obtained.

The feed forward ADC is made of one 7 gain value controlled gain, two coarse quantizers (8 and 7 bits), a shifter, some control logic and extremely simple combinational logic to set the controlled gain switches according to the output of the 7-bit quantizer.

The controlled gain and shifter are identical to that of the feed-back solution and thus consume about 20 and 3 µW respectively, for an area of 0.15 and 0.3 mm^2. The 8 and 7 bits coarse quantizers are slightly smaller than the 9-bit one used in the feed back solution and also consumes less power. One can estimate that 13 and 10 µW are required and that surfaces of less then 0.1 mm^2 are needed. The control and combinational logic only uses 3 µW but has a total area of about 0.3 mm^2.

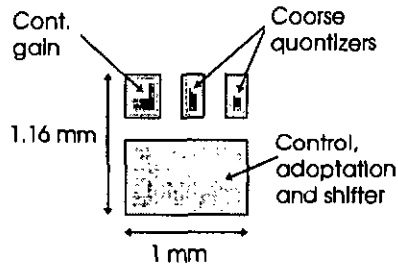


Figure 6.27: Floor plan for feed forwards converter

Using the floor plan of figure 6.27 results in a total die size of less then 1.2 mm^2 for an expected total power consumption of 51 µW.

Compared to the feed back solution (6.6) smaller die size and higher power consumption are obtained. This can be explained by the fact that the feed forward solution presents an extremely simplified adaptation logic, thus reducing the digital part. The increase in power consumption is due to the supplementary coarse quantizer. However, power consumption in such elements is more or less proportional to n^2 (n is the number of bits) [Vitt94]. Since the feed forward solution uses smaller quantizers the final power consumption is only slightly higher.

6.9 TO THE LIMITS

As shown in 5.2, a 14-bit linear converter should be available. Combined with a controlled amplifier, could a floating point solution of 16 or more bits of dynamic range be obtained ? The answer is closely related to the minimum OTA noise level that can be achieved in the considered technology.

Assuming a parallel capacitor scheme (figure 6.13), the first design steps of the controlled amplifier OTA have been performed for floating point solutions with 15 to 20 bits dynamic range. The results are given in table 6.9 where C is the minimum capacitor to be used and HLSB is the value of a half LSB. The ratio of output to polarization current is set to one. This ratio, HLSB, C and the slew rate condition (from the 16 kHz sampling frequency) define the required current I_p. W/L ratio for the differential pair (M1, M2 in figure 6.15) and the mirror below (M3, M4) as well as the resulting equivalent input noise level (NI) and power consumption (with a current source as in figure 6.16) are given for different RW ratios. RW is defined as I_p/I_{lm} for strong inversion and as I_{lm}/I_p for weak inversion. Some designers work with RWs of nine, though five and three are mostly used. A ratio of one is interesting since it results in smaller W/L_{M1} values. However, it cannot be

guaranteed that the transistors will be working in the strong and weak inversion modes required.

# bits		15	16	17	18	19	20
C [pF]		3.0	12.1	48.4	193	774	3099
HLSB [μV]		39.7	19.8	9.92	4.96	2.48	1.24
I_D [μA]		0.7	1.69	5.9	23	97	408
Power [W]		18 μ	41 μ	144 μ	560 μ	2.4 m	10 m
RW=5	W/L_{M1}	29	69	239	951	3.9 k	16.5 k
	W/L_{M3}	0.33	0.5	3	12	51	216
	NI [μV]	21	13.4	7.77	4.00	2.02	1.01
RW=3	W/L_1	17	42	143	571	2.3 k	9.9 k
	W/L_3	0.5	1	5	20	86	361
	NI [μV]	22	14.2	8.17	4.21	2.13	1.06
RW=1	W/L_1	6	14	48	191	791	3.3 k
	W/L_3	1	4	15	62	258	1.1 k
	NI [μV]	23.7	16.7	9.37	4.85	2.44	1.22

Table 6.9: Controlled amplifier for converters with higher dynamics

From a power consumption point of view 17 bits seems to be a limit since higher dynamics will result in controlled gain requiring more than a half mW. They also necessitate fairly big W/L_{M1} ratios and capacitors, resulting in considerable increases in area (for example, 774 pF is more than a half square millimeter).

The integrating by charge transfer controlled gain (6.2.4.2) might seem more suitable to these high dynamic range devices (lower noise level resulting from accumulation → smaller capacitors → smaller current → lower power consumption). However the pre-sampling stage will lead to results similar to those of table 6.9 and hence this solution is also limited.

One can thus say that to maintain reasonable size and power consumption, a maximum converter featuring a dynamic range of 17 bits and a resolution of 14 bits could be implemented in ALP1. The limit to what is considered reasonable is obviously extremely application dependent.

6.10 REFERENCES

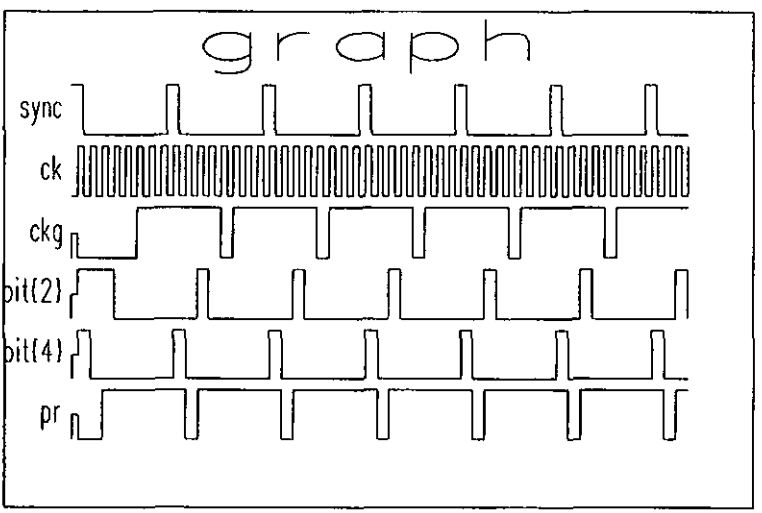
[Saus97] W. Sauser, Lecture notes, "Special Techniques for High Performance Op-Amp IC Design", EPFL, Lausonne, CH, Feb. 1997

[Degr94] M. Degrauwe, Lecture notes, Design of Analog CMOS IC's, Institute of Microtechnology, University of Neuchâtel.

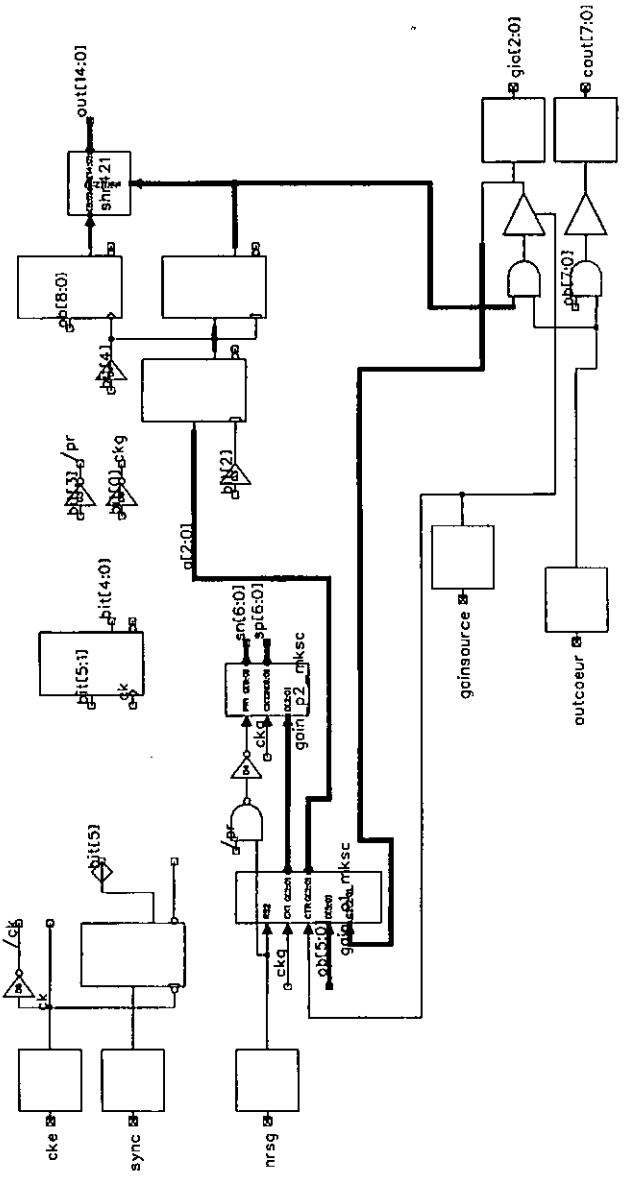
[Vitt94] E. Vittoz, Lecture notes, Workshop on Low Power/Low Voltage IC Design, EPFL, Lausanne, CH, 1994.

6.11 ANNEX

The first figure plots the internal control signal for the adaptation logic and obtained from « sync » and « ck ». The second one presents the adaptation logic schematic. Tasks 1 to 3 (see figure 6.19) are performed in « gain_p1_mskc » (middle-left) while task 4 is achieved in « gain_p2_mksc » (center). The shifter is located at the top right of the figure.



Annex 1: Internal control signal



Annex 2: Schematic of odaptation logic

7. Implementation results

The test of the feed back floating point converter involves four different measurements. The first consists of an informal listening test and power consumption measurement. Its result is critical to establish the success of the research (the primary goal is to develop a low power A/D converter for audio applications). Then, more deterministic measurements, i.e. noise floor, frequency response and transfer function aimed at device characterization, are carried out.

Environmental noise is a major problem and affects all the above measurements.

As mentioned, the controlled gain is critical to the conversion quality. From offset measurement for each gain value, it is possible to estimate that the error, for non unity gain values, is sometimes slightly higher than the 2% tolerance defined in section 4.4. The unity gain error, on the other hand, is much bigger. This provokes a deviation from the ideal transfer function. This doesn't significantly impair the perceived audio quality however, but for more demanding applications, a solution to improve the gain values would be mandatory.

Based on the measurement results, a comparison between both floating point approaches and A. Heubi mixed solution can be drawn. For audio application, A. Heubi solution is better since, for similar performances, it features a reduced size. For other faster and/or lower dynamic range applications, the floating point approach will result in lower power consumption.

Twelve weeks after sending the layout to the foundry, the devices were delivered and testing could start.

The first start up was almost fatal. Indeed, during the fabrication process, the foundry somehow forgot one layer and the devices worked as a short-circuit! Fortunately, a few weeks later, devices from another batch, possessing all the necessary layers, were delivered... and the circuit worked!

The next section describes the testing equipment available at IMT and explains some of the problems encountered during the measurements. The latter are presented in section 2 while section 3 compares the floating point implementation with Alexandre Heubi's mixed linear converter described in section 5.1.

7.1 TEST EQUIPEMENT

IMT does not posses professional testers. Nevertheless, chips returning from the foundry can be at least partially characterized thanks to the equipment described in figure 7.1.

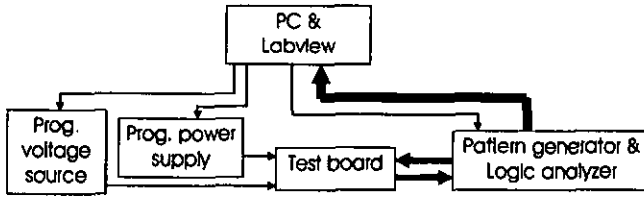


Figure 7.1 : Testing equipment at IMT

Generally speaking a Labview program controls the measurement by communicating with various instruments.

Environmental noise is a major problem. Indeed as levels smaller then -80 dB must be measured (≥14 bits), any perturbation becomes dramatic. For example, measurements of the noise level can be improved by 1 bit by simply removing all the probes. As a result, extra care must be taken when creating the test board (analog/digital separation, capacitors on supply and analog pads etc.) as well as during the measurements (avoid ground loop, braided supply wires etc.). The quality of the supply voltage, which is also used internally as a reference voltage is also critical.

7.2 MEASUREMENTS

7.2.1 Informal listening test

Analog speech (male and female) and music (classical and pop) signals are input to the floating point A/D converter. The obtained digital outputs are then fed

to a suitable D/A converter. Informal listening using both loud speakers and headsets were excellent.

The current furnished by the power supply (± 1.3V) is 19.2 μA and as a result the power consumption at 16 kHz is 50 μW.

7.2.2 Noise floor

Noise floor measurements are interesting since they provide information about the dynamic range of the device. In a practical way, the input of the ADC is connected to the analog ground and the RMS value of the digital output results in the noise floor. Measurements are usually performed at different supply voltages.

The noise floor (NF) measurements of the feed back floating point are reported in table 7.1 (absolute values).

Supply [V]	±1.1	±1.2	±1.25	±1.3	±1.4	±1.5	±1.6	±1.7	±1.8	±1.9
N.F. [dB]	71.1	72.1	72.6	72.6	72.7	72.8	72.1	70.0	67.7	66.6

Table 7.1: Noise floor measurement with EQUI1 and EQUI2

At ±1.3 V, the noise floor is 72.6 dB which corresponds to a dynamic range of only 12.1 bits! This is disappointingly low and is in contradiction with a previous measurement performed with different testing equipment and at ±1.3 V. It resulted into a 77.5 dB noise floor, equaling a close to 13-bit dynamic range. Note also that in table 7.1 the noise level increases for supply voltage higher than ±1.5 V. This is not the expected behavior and a measurement noise problem is suspected. To understand the behavior of table 7.1, the noise sources must be identified as in figure 7.2.

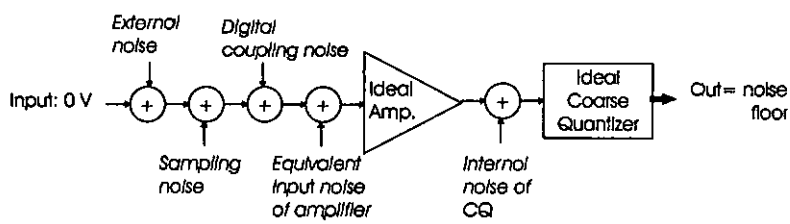


Figure 7.2 : Noise sources while testing the floating point converter

The sampling noise (see section 3.4.2) as well as the equivalent input noise (see equations 6.16 and section 6.2.5.4) are independent of the supply voltage.

The internal noise of the coarse quantizer is composed of sampling and equivalent input noise and it is thus also independent of the supply voltage. The actual level of these noises (sampling noise, amplifier equivalent input noise and coarse quantizer noise) does not depend on the measurement environment.

The switching activity of the digital parts provokes perturbations that propagate through the bulk. One can suppose that as the voltage increases, the perturbations increases as well (higher amplitude of the switching level). Nevertheless, the layout of the chip was made in such a way to provide at least some shielding between the analog and digital part and thus limit the coupling. Unfortunately, this phenomena cannot be simulated with IMT's software tools. The perturbation level is however independent of the testing environment.

The external noise results from antenna and caupling effects occurring at the test board level and is certainly influenced by the testing environment. One can reasonably think that it also depends on the voltage supply.

In fact, the main lesson from this measurement is that no clear conclusion can be drawn about the converter's dynamic range. In 6.2.6, the post-layout simulation of the controlled gain amplifier for a gain value of 64, showed a value only slightly smaller then the LSB of the coarse quantizer. As a result, at ±1.3 V, a dynamic range smaller than 15 bits can be expected. However, the 12.1 bits of table 7.1 is most probably not the real dynamic range since the environmental noise impaired the measurement.

Even though the exact dynamic range value cannot be given, one can say that it is limited from the "small signals side". Indeed, the coarse quantizer has been measured to be absolute and the noise level at the controlled gain amplifier is such that when small gain values are applied, it is well below the quantizer's LSB. On the other hand, when large gain values (64 and maybe also 32) are applied, the noise at the output of the controlled gain is higher then the coarse quantizer LSB. As a result the minimum analog signal that can be converter is not LSB/64 anymore but rather a few LSB/64.

7.2.3 Transfer function

The transfer function as defined in 2.1.1 is given in figure 7.3. The plot shows negative numbers. This is due to the fact that the analog input source was programmed to scan negative voltages. Power supply was ± 1.3V and the sampling frequency 16 kHz. For each input value, 256 conversions were performed and the mean converted value is plotted in 7.3.

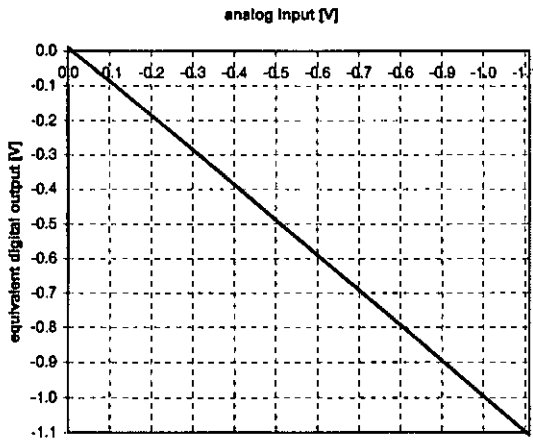


Figure 7.3 : Transfer function

The error function obtained from the above transfer function (which exploins the negotive input voltoge volue) is given in figure 7.4. To onolyze the figure, the following remarks must first be considered:

- only analog signols entering the coarse quantizer with an omplitude smoller thon 0.32 V ($1.3*64*2^{-8}$, see section 6.4.1) lead to o possible increment of contralled goin value. Furthermore, if the signol omplitude is greater than 0.4875 V ($1.3*96*2^{-8}$, see section 6.4.1), the controlled goin is decreased by ot least 6 dB.

- the noise floor, whose measured volue is ot obout 73 dB (12 bits) implies thot the minimol 'saw tooth height' could not be smoller than 8 LSBs. The maximol height on the other hand, should be 64 LSBs (the resolution of coorse quantizer is 10 bits).

- because the error plot is obtained from averoged output values, the octuol 'saw toath height' is reduced.

Clearly, the flooting point converter presents on offset, it's volue, measured with o forced maximum goin, is about 120 LSB or 0.0095 V ($1.3*120*2-14$). Assuming that the coarse quantizer offset is 25 mV (or in other words obout 5 coarse bits i.e. $1.3*5*2-8$) and that the omplification value is ideol, the controlled goin offset is 8.7 mV (computed according to equation 7.1 in next section). This volue is about

three time higher then the simulated one (3 mV, see 6.2.6). Rough simulations showed that this could be caused by a mismatch in the differential pair.

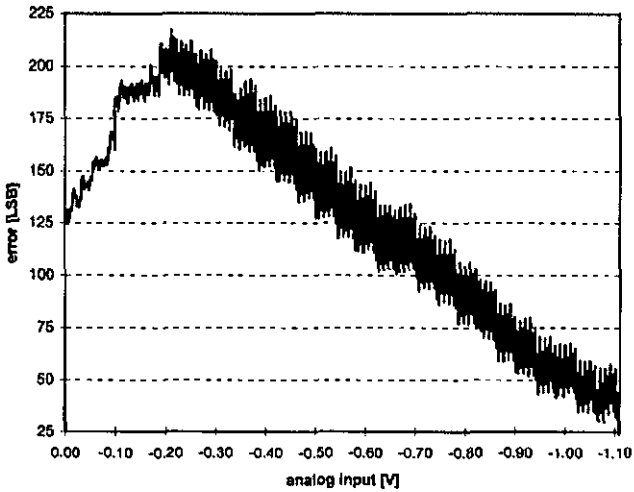


Figure 7.4 : Error function

Three main zones can be defined in figure 7.4. The first, for analog input amplitudes smaller than about 0.1 V, has a positive slope, the second, for analog input amplitudes between 0.1 and 0.25 V as a more or less horizontal slope while the third, for amplitudes bigger than 0.25 V has a negative slope. This can be related to zones in the transfer function plot: where the error slope is null, the transfer function has no gain (or linearity) error. On the contrary, where the error plot has non-null slope, the transfer function has a linearity error.

For very small amplitudes (a few mV) the 36 dB amplification is always applied and the amplitude of the saw teeth is 4 LSB. This concurs with the noise floor measurement, taking into account the effect of the averaging. For very high amplitudes the unity gain is always applied and the saw teeth are about 35 LSB high which again concurs with the expected behavior. It is difficult to predict the quantization step in this "middle" zone (neither 1 nor 64 are always applied). Indeed, for a given input signal and since several conversions take place, the gain "jumps" from one value to the other. This is due to the fact that the adaptation table (based on that of section 4.1) never proposes a null increment. Hence, even if at a given time the amplification gain is "perfect" in fitting the amplified signal into the coarse quantizer, after a maximum of 4 conversion (4*1.5 dB=6dB) the gain will be modified by at least 6 dB.

7.2.4 Frequency response

The ideal frequency response has been discussed in 4.2.3 and was plotted in figure 4.7. The characteristics, for a full scale 300 Hz sine wave were:

THD=-61.1 dB SNR=56.2 dB

PHD=-73.3 dB SINAD=55dB

The measurement of the frequency response to a 500 Hz sine wave at -9 dB below full scale is given in figure 7.4. Power supply is ± 1.3V and the sampling frequency 16 kHz.

Compared to figure 4.7, the spectral lines have a higher magnitude. This results in lower THD and PHD values as well as SNR and SINAD. This is due to the higher noise level as measured in section 7.2.2 (measured 72.6 dB noise floor instead of theoretical 90 dB).

THD=-56.3 dB SNR=52.9 dB

PHD=-49 dB SINAD=50.6 dB

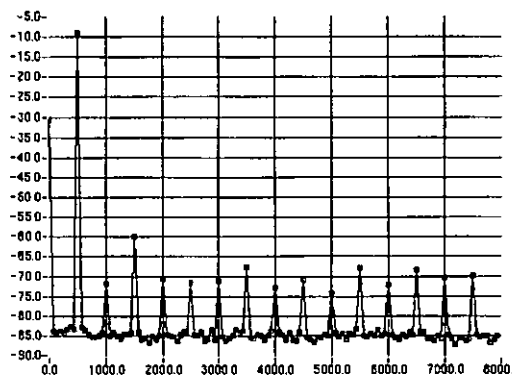


Figure 7.4: Frequency response

7.2.5 Controlled amplifier

The gain values of the controlled amplifier are closely related to capacitor matching and finite DC gain of the OTA. While designing the layout of the controlled gain, numerous simulations were performed. According to the results,

controlled gain, numerous simulations were performed. According to the results, the size of the capacitors were slightly modified in an attempt to optimize the gain values.

To have an idea of the actual behavior of the controlled gain, the output to a 0 input signal can be measured for each amplification value. Equation 7.1 applies where off_{CA} is the offset of the controlled amplifier and off_{CQ} that of the coarse quantizer.

$$out = \frac{off_{CA} \cdot G_{real} + off_{CQ}}{G_{ideal}} \tag{7.1}$$

$$out \cdot G_{ideal} = G_{real} \cdot off_{CA} + off_{CQ} \tag{7.2}$$

To compute G_{real} one must first determine off_{CA} and Off_{CQ} : the points obtained by multiplying out by G_{ideal} (equation 7.2) are plotted versus G_{ideal}. This result in a points constellation that should be more or less aligned. Let us now define a function F(G), that is the best linear approximation of that points constellation. The parameters a and b can be found graphically.

$$F(G) = G \cdot a + b \tag{7.3}$$

Equalizing equation 7.2 and 7.3 results in equation 7.4 and the value of G_{real} can be computed. These values are given in table 7.2.

$$G_{real} \cong \frac{out \cdot G_{ideal} - b}{a} \tag{7.4}$$

From 7.1 and 7.2, G_{real} can be computed as in 7.3. Results are given in table 7.2 for 5 different chips.

	Ideal gain values						
	1	2	4	8	16	32	64
Chip #	Real gain values						
1	1.19	1.95	3.46	6.48	12.53	24.63	48.83
2	0.68	2.02	3.83	7.48	14.39	29.95	58.95
3	0.67	2.01	3.81	7.22	13.83	28.95	57.38
4	0.91	2.00	3.46	6.39	11.89	24.47	48.17
5	0.7	2.00	3.69	7.13	13.59	28.41	56.06

Table 7.2: Real gain values

Once the real gain values are known, the linearity and distortion of the controlled gain can be obtained as in table 7.3.

		Ideal gain values						
		1	2	4	8	16	32	64
Chip #	linearity	Gain error (considering the linearity of column 2) [%]						
1	0.76	23	3	0.2	1.5	4.1	1.1	0.6
2	0.92	37	1.7	0.1	0.25	3.1	1.4	0.04
3	0.89	38	2.0	1.5	1.1	4.1	1.0	0.5
4	0.74	25	1.9	0.4	0.5	3.9	1.0	0.4
5	0.87	35	2.6	0.0	0.6	3.8	1.3	0.3

Table 7.3: Controlled gain characteristics

It should be stressed that the problem of determining the real gain values is not deterministic. Indeed, the best fitting approximation of equation 7.2 can be chosen according to various criteria. For example, if only root mean square error is considered, the computed gain values present a linearity extremely close to one though have distortion for non unity gain of up to 30%. To obtain the presented results, the best fitting approximation was computed taking into account the following facts:

- each feed back capacitor is impaired by an error, resulting from parasitic capacitors and mismatch. Since the unity gain is obtained by using all the feed back capacitors (see figure 6.13) in parallel, one can assume that the error of each capacitor is summed, resulting in a poor amplification value precision. The unity gain is thus not considered to obtain the best approximation.

- the best gain value should be 2 since in this configuration the two biggest capacitors are used. Matching and parasitic errors should thus be less impairing than when a smaller capacitor is used. As a consequence, to compute the best approximation, a "weight" is attributed to the errors: the bigger the gain value, the smaller its error weight.

- the relative error is more relevant then the absolute one. Indeed, an approximation resulting in an absolute error $|G_{real}-G_{ideal}|$ of 0.2 for ideal gain values of 2 and 64 actually means a gain error $|G_{real}-G_{ideal}|/G_{ideal}$ of respectively 10% and 0.3%.

From table 7.3, the linearity (or the slope) of the controlled amplifier is smaller than one. This is not impairing as explained in 4.4. The distortions (for non unity gain values) are slightly higher than the targeted 2%. On the other hand, the unity gain error is extremely high! Because the informal listening tests have been

performed with normal speech level of about -30 dB below full scale, this unity gain is very seldom used and thus doesn't impair the informal listening test.

This short study of the controlled gain clearly shows that the implemented capacitor network structure, relying on capacitor matching is not reliable enough. A redesign, using a different layout configurations would certainly improve the above results, though, the problem of the unity gain would subsist. A simple solution would be to avoid the error accumulation by using a single capacitor for the unity gain as well. The resulting size increase (compared to the whole chip size) would not be significant.

7.3 COMPARISON WITH A. HEUBI'S SOLUTION (5.1)

7.3.1 Converters for audio applications

As explained in chapter 5, Alexandre Heubi developed a low power mixed linear converter that is well suited for audio applications. Its measured performance at ±1.25V and 16 kHz is:

Excellent perceived quality	
Maximum Resolution:	10 bits
Dynamic range:	13.5 bits
Power consumption:	50 μW
Die size: (estimation)	0.85 mm^2

In similar conditions, the feed back floating point reserve bit A/D converter performance is (section 7.2):

Excellent perceived quality	
Maximum Resolution:	9 bits
Dynamic range:	13 bits
Power consumption:	50 μW (48 μW if using the 6dB table)
Die size:	1.96 mm^2 (1.72 mm^2 if using the 6dB table)

For the feed forward floating point converter, no implementation was performed though the theoretical (simulated) characteristics are (4.5.3):

Excellent perceived quality	
Maximum Resolution:	8 bits
Dynamic range:	14 bits
Power consumption:	51 μW

Die size: 1.2 mm²

Comparing the above numbers results in the following remarks: the three audio solutions have similar perceived quality, power consumption and dynamic range. However they are very different in their die size. In particular the feed back floating point converter is twice the size of A. Heubi chip.

7.3.2 Converters for non audio applications

In the context of non audio applications, the critical elements in designing non-absolute converters can be evaluated. Figure 7.5 shows the main 'blocks' for all considered solutions i.e. mixed, feed back and feed forward floating point. Control blocks are not shown and coarse quantizers are assumed to be absolute RSD implementations.

In Alexandre Heubi's converter a single block provides both the dynamic range and resolution. The latter depends on the OTA_1 DC gain, capacitor matching and possibly algorithmic correction, while the former is related to the size of the capacitors C_1 and OTA_1 noise. The working frequency is proportional to the product f_s*n_1, where f_s is the sampling frequency and n_1 the number of bits.

In the feed back floating point approach, three blocks are used. The dynamic range is provided by the controlled gain whose critical elements are the capacitors (same as C_1) and the OTA_1 noise. The working frequency of this block is proportional to f_s. The coarse quantizer provides the resolution and its critical elements are the capacitors ($C_2<C_1$), OTA_2 DC gain and noise and C_2 matching. Its working frequency is f_s*n_2. The digital block providing the adaptation is only critical in the sense that it can result in large area.

In addition to the blocks presented in feed back converters, feed forward floating point implementations require a second coarse quantizer whose critical element are again the capacitors (C3), OTA3 DC gain and noise and C3 matching. Its working frequency is fs*n3.(n2+n3=n1).

From the power consumption point of view, to define the preferential domains of each solution, one must remember that the capacitors are designed according to the thermal noise condition. Doubling the precision thus results in quadrupling the capacitors. Furthermore, the OTA output current is determined either by the internal noise or by the product of capacitor and slew rate. The latter depends on the working frequency. As a result, Alexandre Heubi's principle is well suited for any situation where the OTA current is determined by the internal noise. This is typically the case in low speed and/or high dynamic range applications. On the other hand, when the slew rate becomes the determining factor, as in

high speed and/or low resolution applications, floating point structures should be used. Indeed, the controlled gain, where the biggest capacitors are used, works slower and consequently the slew rate condition becomes easier. As explained in 4.4 the choice of whether to use the feed back or feed forward approach is data driven. The former is only suited for predictable signals conversion.

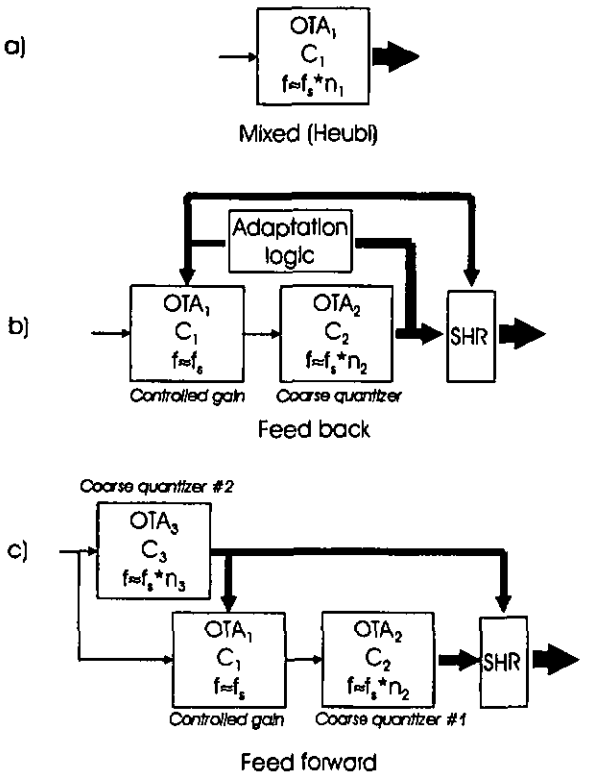


Figure 7.5 : Main blocks for relative precision A/D: Mixed (a), feed back floating point (b) and feed forward floating point (c)

To conclude this comparison, it must be stressed that the floating point converter can use any kind of coarse quantizers. This is an important advantage since cyclic RSD solutions are limited in speed (algorithmic conversion).

8. Conclusions

In this final chapter, general conclusions about the utility of the floating point conversion concept are first drawn. The main contributions and achievements of the PhD research are then detailed. Finally, tracks for future work are outlined.

Linear A/D converters can be classified in absolute and non-absolute (1.2, 2.1.2) devices. Absolute converters feature a constant quantization step and are useful for applications dealing with signal analysis. On the other hand, non-absolute converters have a quantization step that increases along the dynamic range while ensuring that at least an appropriate resolution is achieved. These devices are useful in applications that necessitate signal measurement (control, audio, etc.).

Structures to realize non absolute converters are less complex and as a result less power consuming implementations are obtained. They are thus extremely well suited for battery operated consumer products.

Non-absolute « floating point », « relative precision » and « mixed » converters can be distinguished (1.2, 2.1.2). Floating point conversion is achieved by scaling (or adapting) the input signal in such a way that it fits well into the fixed range of a coarse quantizer. Feed forward and feed back adaptation "strategies" must be distinguished. The latter is only efficient for predictable signals while the former applies for poorly or non predictable ones.

Due to the masking effects occurring in the human hear, converters featuring a reduced resolution are sufficient for certain audio applications. Hence, non absolute devices can be used. Furthermore, audio signals are fairly predictable (pitch, spectral density) and are a perfect case study for feed back floating point conversion.

The implemented feed back floating point A/D converter features a 9-bit mantissa and a 7 value controlled gain. Because of the controlled gain internal noise, the dynamic range is limited, on the small signal side, to 13 bits. This limitation does not impair the perceived audio quality and since the power consumption at 16 kHz and ±1.3 V is 50 µW, the main goal of the project is reached.

Based on the feed back floating point results as well as on simulations, estimates for a dedicated audio feed forward realization can be drawn. For an identical simulated audio quality a slightly higher (+8.5%) power consumption is obtained while the area is significantly reduced (-40%). This is due to the fact that the feed forward approach does not take into account the predictability of audio signals. As a result an extra coarse quantizer (small but power consuming) replaces most of the adaptation logic (big but less power consuming).

The feed back floating point solution is entirely satisfactory for audio signal conversion. However, A. Heubi Redundant Signed Digit (RSD) mixed linear solution features a similar audio quality and power consumption but the area is reduced by half. This is clearly an advantage in chip size sensitive applications such as, for example in The Canal (ITC), hearing aid.

There are other, non audio, applications where the floating point conversion might be more efficient, from a power consumption point of view, than mixed RSD. Considering implementation schemes similar to that of the audio converter, i.e. using absolute RSD converters as coarse quantizers, the design constraints lead to the following conclusions: applications requiring high speed and/or low resolution will benefit from the fact that in floating point converters dynamic range and resolution are provided by separated hardware (controlled gain for the first and coarse quantizer for the latter). On the contrary, when high resolution and/or low speed are required, mixed RSD conversion results in lower power consumption.

Finally, it must be stressed that the floating point approach can use any kind of coarse quantizer. Existing converters can thus be enhanced with increased dynamic range. For example, a 8 bit half-flash converter could be used in combination with a 4 gains controlled amplifier to result in a fast 12 bit device. Because mixed RSD converters are based on a cyclic successive approximation algorithm, their speed is intrinsically limited.

8.1 MAIN CONTRIBUTIONS

8.1.1 Low power feed back floating point A/D converter for audio applications

The idea of feed back floating point conversion was first published by A. Schaub [Scha92]. However it was not suited to low power implementation and the perceived quality was sufficient for speech coding application but still required improvement for the targeted application (speech processor). The main

contribution of this work was thus to modify and refine the feed back floating point approach, resulting in an enhanced concept and finally in a low power chip

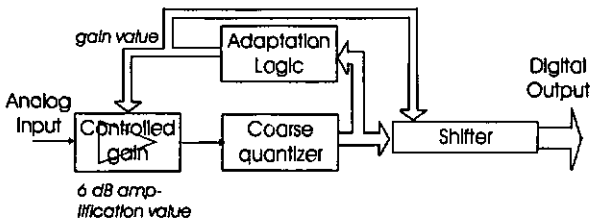


Figure 8.1 : Enhanced feed back A/D conversion

In enhanced feed back floating point A/D conversion, the input sample is fed to a controlled amplifier whose gain values are multiples of 6 dB. The amplified signal is applied to o coarse quantizer whose digital output is fed to a shifter as well as to the adaptation logic. The latter keeps track of the applied gain values to ensure that the shifter properly formats the quantizer output. It also performs o one ward prediction and decides which gain value must be applied to the incoming sample. The use of feed back floating point conversion is thus limited to applications where the signal to be converted can be reasonably well predicted.

For audio signals two solutions can be identified: the first has a 9-bit coarse quantizer and 7 gain values for the controlled gain while the second has a 10-bit bit coarse quantizer and 6 gain values for the controlled gain. From a power consumption point of view the first solution is best. In both case, two adaptation strategies can be applied. The first is based on Jayant's work though is only realizable thanks to our improvements i.e. accumulation to 6 dB and reserve-bit addition. The second directly applies a 6 dB table obtained through a new methodology that uses a pool of input signals and optimizes simulated conversions.

An audio A/D converter, featuring the first adaptation strategy has been implemented in ALP2, a low voltage 2µm CMOS technology from EM Microelectronic-Marin SA. All the analog parts have been carefully designed and dedicated low power techniques applied. Different controlled gain structures have been investigated. The final choice was an inverting amplifier whose feed back capacitor can be selected. This extremely simple scheme is well suited for low power implementation while keeping the area reasonably small. Its main drawback is the sensitivity to capacitor matching. The coarse quantizer is an absolute cyclic RSD converter based on an A. Heubi structure. The digital parts have been realized using a low power standard cell library.

The chip die size is 1.96 mm^2. Only 12% of this area is used by analog parts which however consume 76.5% of the power. The latter was measured at ±1.3 V and 16 kHz sampling frequency and amounts to 50 µW. Other measured characteristics are 13 bits dynamic range and, more importantly, excellent perceived quality. The device fully meets the requirements for a low power audio A/D converter.

Tests of the chips showed that the controlled gain is critical to the overall performance of the converter. In particular, its internal noise must be such that even with the highest gain value, the noise level at the output of the controlled gain must be smaller then a half LSB of the subsequent coarse quantizer. This is not the case in the present realization and as a result the dynamic range is limited to 13 bits (instead of the 15 theoretical ones). The noise level can be decreased by increasing the current of the amplifier as well as modifying some transistor sizes. A rough estimation shows that increasing the OTA polarization current by 20% is sufficient which results in a total (whole chip) power consumption of 55 µW

8.1.2 Low power feed forward floating point A/D converter for audio applications

In feed back conversion, when the signal to be converted cannot be reasonably well predicted, the coarse quantizer size must be increased. However, power consumption in coarse quantizers varies with n^2 (n is the number of bits). Hence, at a certain point, from a power consumption point of view, a feed forward implementation can lead to a more efficient implementation. As a matter of fact, in random type signals, feed forward conversion is the only solution (as opposed to feed back). Another contribution was thus to evaluate the costs of a feed forward implementation in the case of audio signals.

Feed forward conversion uses the input sample, on which the scaling is then applied, to evaluate the controlled gain value.

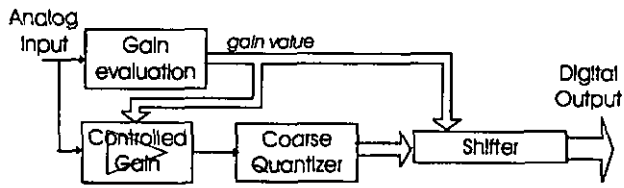


Figure 8.2 : Feed forwards A/D conversion

For audio applications, a feed forward realization requires 7 gain values and 8 bits in the coarse quantizer. Compared to the feed back solution, only one bit is saved.

No feed forward converter chip was realized. Nevertheless, based on the feed back implementation results as well as on simulations, some power consumption and size estimates can be made. The controlled gain and coarse quantizer could be realized similarly as in the feed back implementation and the gain evaluation could be performed by a second coarse quantizer. In this case, compared to the feed back implementation, the resulting chip die size would be about 40% smaller while consuming slightly less then 10% more. This solution would also meet the requirements for a low power audio A/D converter and if the power consumption increase is tolerable, the reduced size is clearly an advantage.

8.1.3 Floating point converters for non audio applications

Floating point converters are well suited for audio applications and are a valid alternative to obtain non-absolute devices. They feature similar performance to A. Heubi's solution though require a bigger die size. If dynamic range and/or resolution were modified, which structure would be best suited for non-audio domains and what would be the upper limit?

From a power consumption point of view, considering mixed RSD converters and floating point implementations using absolute RSD coarse quantizers, the following principles apply: if the controlled gain OTA current is determined by the slew rate constraint, floating point structures are well suited. On the other hand, when the determining factor is the internal OTA noise, floating point implementations should be avoided. The first case typically arises in high speed and/or low resolution applications while the second occurs in low speed and/or high resolution cases. When a floating point solution is foreseen the signal predictability determines which of feed back or feed forward method should be used.

The limit of absolute linear cyclic RSD converter is 14-bit. To reach higher dynamic range and resolution it is better to combine cyclic RSD with some oversampling. The absolute 14 bits RSD converter was realized through digital correction of A. Heubi's mixed non-absolute solution. This device can be used as a coarse quantizer in a floating point implementation and the upper dynamic range limit established. The maximum dynamic range depends on the minimum OTA noise level that can be reached while keeping the consumption low. This is technology related and in the case of ALP1 LV, 18 bits seems the reasonable

limit. The consumption would be below 600 µW at ±1.3 V and 16 kHz and the device would feature 14 bits of resolution.

8.2 FUTURE WORK

This PhD dealt with floating point conversion for audio signals. The main effort was to implement such a feed back A/D converter while solutions for non audio signals were outlined. Some issues are still open and could be further investigated.

At a system level point a view, the feed back adaptation strategy could be thoroughly investigated. It has been mentioned that the difference between adjacent samples is the determining criteria. It could be interesting to consider a wide range of non audio applications and study their signals' statistics. The applicability limit of feed back floating point conversion could then be estimated.

At an architectural level, feed forward solutions could be improved by designing a dedicated gain evaluation block. In the power consumption and size estimations, a RSD coarse quantizer was used. This solution doesn't benefit from the fact that only logarithmic levels must be determined (1, 2, 4, 8, 16, 32, 64). Using some sort of logarithmic comparator could result in more efficient power consumption and such structures should thus be investigated.

The implemented chip could be redesigned to include offset compensation as well as controlled amplifier non linearity correction. This could be realized digitally, resulting in a slight increase of power consumption. However, the resulting die size would probably be much greater.

Finally, since the floating point approach is more efficient for high speed/low resolution applications, and since any coarse quantizer can be used, new targeted non audio devices could be investigated.

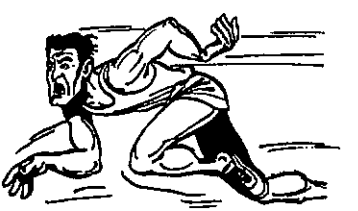
Acknowledgments

Although a single name appears on the cover page, many people contributed to the realization of this PhD.

First, I'd like to thank Professor Fausto Pellandini who started this whole adventure way back in 1975 by creating the Microtechnology Institute (and who sacrified his summer holiday to proofread this report). My thanks also to the reviewers Professor Nico de Rooij, Dr. Paul Zbinden and Dr Pierre-Andre Farine.

I'm extremely grateful to Peter « P'tit Boss » Balsiger for his trust, support and legendary enthusiasm. He successfully created a « family » where each member can bloom and develop his/her « wings ».

I'd also like to thank the people who directly participated in this research. « Guru » Alexandre Heubi was a main contributor : he managed to teach me some analog concepts and was always available for discussions (and debugging !). Steve Tanner, Christian Robert and young Christophe Calame (who form the guru's disciples together with myself) participated in animated discussions and helped in setting up the measurement laboratory, capturing Labview programs, testing the devices and providing some company in the lab. Hein Burri's rôle, although discrete, was certainly mandatory : every day he miraculously maintains the computer network and CAD tools in perfect working order in spite of our craziest demands and mis-manipulations.



Many thanks to my « support comity », Alain Dufaux, Eric Meurville, Vincent Moser and Dequn Sun (the best tenor singer of the university choir) as well as to all my other colleagues.

The world being larger than Rue A. L. Breguet 2 alle est, I'd like to thank my friends from all around the world (don't worry for the bill, we communicate mostly through e_mail). I was also lucky to have the assistance of the great English -oups excuse me- Scottish rugby player David Stewart. His editing and "high level" insight contributed with no doubt to the clarity of the finished work.

I'd like to thank especially my sister, her son in law for all attentions,

my family, and parents, my and my mother those small which summed up result in a lot of care and love.

Finally, my gratitude goes to my friend and husband Laurent. The value of his support, encaurogement, assistance and love is not measurable.

Ch. du Moulin 4b
2022 Bevaix

Phone +41 32 718 34 18
Fax +41 32 718 34 02
E-mail louisa.grisoni@imt.unine.ch

Louisa Grisoni-Busca

Personal Information

- Marital status: Married
- Nationality: Swiss, Canadian and Italian
- Age: 29

Education

1998 Institut de Microtechnique, Uni NE Neuchatel
Dr. es Sciencee Techniques

1966 - 1991 Swiss Institute of Technology (EPFL) Lausanne
Dipl. Eng. Microtechnique EPFL

Languages

French, English, Italien, German

Work experience

Since 04.1992 Institut de Microtechnique, Uni NE Neuchatel
Researcher

- Design and implementation of low power ADC
- System Level DSP design
- Low power DSP systems implementation
- Top-Down design methodology for DSP applications
- External consultant for ALTA GROUP or Cadence Design Systems

03.92-03.91 Swiss Institute of Technology (EPFL) Lausanne
Researcher

- Design, realization and test of a non invasive esophagus probe for optical property analysis

Hobbies

Sports (tennis, scuba-diving, swimming, hiking), music (choir, ukulele), socializing

Patents and publications

- «A 14-bit A/D Converter Featuring 45 mW at ±1.25 V and 16 ksamples/s », L. Grisoni et al, ISIC'97, Sept. 1997, Singapore.
- « Micro Power 14-bit RSD A/D Converter », L. Grisoni et al, ICSPAT'96, Oct. 8-10 1996, Boston MA, USA.
- « Implementation of a Micro Power 14-bit Floating-Point A/D Converter », L. Grisoni et al, ISLPED'96, Aug. 12-14 1996, Monterey CA, USA.
- « Micro Power Relative Precision 15-bit A/D Converter », L. Grisoni et al, ICSPAT'95, Oct. 24-28 1996, Boston MA, USA.
- « Design of Lattice Wave Digital Filters in SPW Environment », L. Grisoni et al, ICSPAT'93, Sept. 28-Oct. 1 1993, Santa Clara, CA, USA.
- « Short-time Spectral Analysis Based On Kalman Filters », L. Grisoni et al, ICSPAT'92, Nov. 2-5 1992, Cambridge MA, USA.