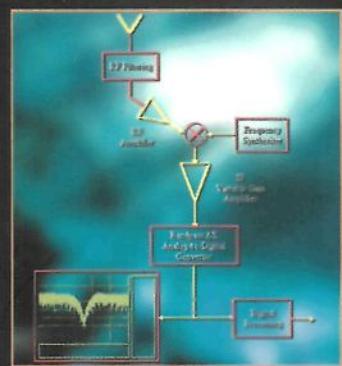


# Delta-Sigma Modulators

Modeling, Design and Applications

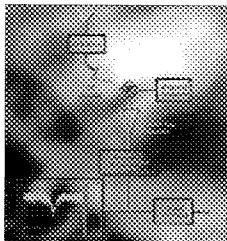


George I Bourdopoulos  
Aristodemos Pnevmatikakis  
Vassilis Anastassopoulos  
Theodore L Deliyannis

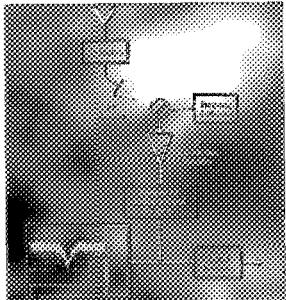
Imperial College Press

Delta-Sigma  
**Modulators**

Modeling, Design and Applications



This page is intentionally left blank



# Delta-Sigma **Modulators**

Modeling, Design and Applications

George I Bourdopoulos

*University of Patras, Greece*

Aristodemos Pnevmatikakis

*Athens Information Technology, Greece*

Vassilis Anastassopoulos

*University of Patras, Greece*

Theodore L Deliyannis

*University of Patras, Greece*

*Published by*

Imperial College Press  
57 Shelton Street  
Covent Garden  
London WC2H 9HE

*Distributed by*

World Scientific Publishing Co. Pte. Ltd.  
5 Toh Tuck Link, Singapore 596224  
*USA office:* 27 Warren Street, Suite 401-402, Hackensack, NJ 07601  
*UK office:* 57 Shelton Street, Covent Garden, London WC2H 9HE

**British Library Cataloguing-in-Publication Data**

A catalogue record for this book is available from the British Library.

First published 2003  
Reprinted 2006

**DELTA-SIGMA MODULATORS: MODELING, DESIGN AND APPLICATIONS**

Copyright © 2003 by Imperial College Press

*All rights reserved. This book, or parts thereof, may not be reproduced in any form or by any means, electronic or mechanical, including photocopying, recording or any information storage and retrieval system now known or to be invented, without written permission from the Publisher.*

For photocopying of material in this volume, please pay a copying fee through the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, USA. In this case permission to photocopy is not required from the publisher.

ISBN 1-86094-369-1

Printed by FuIsland Offset Printing (S) Pte Ltd, Singapore

## Preface

A Delta-Sigma ( $\Delta\Sigma$ ) Modulator is a relatively simple system, which, however, has attracted the interest of many and very important research workers and professionals over the past 40 and more years. It is an academically challenging system, because it involves analog and/or digital issues for its study and implementation. Most important though is the practical aspect of it, because of its robustness and the number of important applications among which its use in the design of Analog-to-Digital Converters and Digital-to-Analog Converters predominates. A great number of research and tutorial papers as well as excellent books have been published. Although the topic of  $\Delta\Sigma$  Modulation has undoubtedly reached a considerably high state of maturity, the literature nevertheless continues to be enriched with new and scientifically sound contributions to this topic.

The extended literature on  $\Delta\Sigma$  modulation has been one of the main reasons behind the decision of these authors to write this book. They felt the need to assist the newcomer professional, the advanced final year First Degree in Electrical Engineering students and early postgraduate students in getting easier involved with this topic. The book covers the most important issues associated with the study, design and implementation as well as some important applications of  $\Delta\Sigma$  modulators.

The content of the book refers nearly exclusively to all issues applied to single-bit, single-stage  $\Delta\Sigma$  modulators for reasons of simplicity in the

presentation and design popularity. Oversampling and noise shaping, noise modeling, system architectures, noise transfer function derivation, stability considerations, design implementation and practical imperfections, stabilization techniques, application of  $\Delta\Sigma$  modulation to the design of data converters, frequency synthesis etc. are the main issues examined to some detail. Moreover the issue of low power systems is explained briefly. Practical design examples are given for those readers who would be interested in reproducing in their laboratory the designed circuits using discrete components. These include low-pass and band-pass single-bit  $\Delta\Sigma$  modulators.

It has been attempted to make the book educationally rather than research oriented. A number of solved examples have been included in the text, where it was felt necessary, while unsolved problems can be found at the end of each chapter. Extended bibliography is also given at the end of each chapter to help the interested reader get a deeper understanding of the various issues described in the text. Finally, a web site has been built relevant to the book with the purpose of bringing the authors and the reader in close contact. The visitor to the site will find information about the book, possible misprint corrections, solutions to the unsolved problems etc. in the following address:

<http://www.ellab.physics.upatras.gr/>

Preparing a manuscript is not an easy task, especially when the authors belong to three different generations as these authors do. However, the fact that the three younger ones had been both undergraduate and postgraduate students of the eldest (T. Deliyannis), coupled with the traditional respect that well brought-up Greek students feel towards their seniors, simplified this task.

Therefore, the completion of the manuscript was achieved in an amicable and co-operative atmosphere. It is with great pleasure that the senior author acknowledges the important contribution of each one of his co-authors in preparing the manuscript and expresses his gratitude to them for offering him the satisfaction of terminating his official university career in coincidence with the publication of this book. He also expresses his thanks to A. Doyle and Yeow-Hwa-Quek for making all necessary arrangements for the publication of the book as well as to Professor A.G. Constantinidis of Imperial College, London for many

useful discussions and his encouragement. Drs N. Kouvaras and D. Lagoyiannis are gratefully acknowledged for their earlier collaboration with this author in Delta Modulation and for their assistance in his research in the subject of the book. Thanks are also due to V. Boile for typing chapter 1 and other minor parts of the manuscript. The married authors (T.D, V.A. and A.P.) express their thanks to their spouses Myriam, Maria and Efi respectively for their patience and understanding, particularly when many professional and personal needs had to be left aside in order to give priority to the preparation of some urgent parts of the manuscript. The other author (G.B.) thanks his parents for their love and understanding.

GB Patras  
AP Athens  
VA Patras  
TLD Patras

This page is intentionally left blank

# Contents

<i>Preface</i>	v
1. Introduction	1
1.1 Modulation – Demodulation . . . . .	1
1.2 $\Delta\Sigma$ Modulation . . . . .	2
1.3 Design and Implementation of $\Delta\Sigma$ Modulators . . . . .	4
1.4 Applications . . . . .	6
1.5 Book Organization. . . . .	7
References . . . . .	10
2. Analog to Digital Conversion	12
2.1 Introduction. . . . .	12
2.2 The Basic Concept of A/D Conversion . . . . .	14
2.3 Uniform Sampling . . . . .	15
2.4 Quantization Error and the Linear Model . . . . .	20
2.5 Sampling of Band-Pass Signals . . . . .	25
2.6 Oversampling Principles . . . . .	27
2.7 The Delta Modulator . . . . .	29
2.8 Performance of the Delta Modulator . . . . .	33
2.9 The Exponential DM. . . . .	34
2.10 The Concept of Noise Shaping . . . . .	36

2.11	Summary . . . . .	37
	Problems . . . . .	38
	References . . . . .	40
3.	$\Delta\Sigma$ Modulators — Architectures	41
3.1	Introduction. . . . .	41
3.2	First-Order $\Delta\Sigma$ Modulators. . . . .	42
3.2.1	Input and output waveforms . . . . .	45
3.2.2	SNR and PSD diagrams . . . . .	46
3.3	Comparison of Delta and $\Delta\Sigma$ Modulators . . . . .	48
3.3.1	Linear DM. . . . .	48
3.3.2	Exponential DM . . . . .	48
3.3.3	$\Delta\Sigma$ Modulator . . . . .	49
3.4	Second-Order $\Delta\Sigma$ Modulators . . . . .	49
3.5	High-Order $\Delta\Sigma$ Modulators . . . . .	51
3.6	Stability of Single-Stage $\Delta\Sigma$ Modulators . . . . .	58
3.7	Multi-Stage $\Delta\Sigma$ Modulators . . . . .	59
3.8	Multi-Bit $\Delta\Sigma$ Modulators. . . . .	64
3.9	Hybrid $\Delta\Sigma$ Modulators . . . . .	68
3.10	Adaptive $\Delta\Sigma$ Modulators. . . . .	69
3.11	Band-Pass $\Delta\Sigma$ Modulators . . . . .	70
3.12	Summary . . . . .	73
	Problems . . . . .	73
	References . . . . .	75
4.	Single-Bit Single-Stage $\Delta\Sigma$ Modulators, Modeling and Design	77
4.1	Introduction. . . . .	77
4.2	Modeling of $\Delta\Sigma$ Modulators . . . . .	78
4.2.1	Linear model . . . . .	79
4.2.2	Quasi-Linear model. . . . .	81
4.3	NTF Characteristics . . . . .	83
4.4	Stability of $\Delta\Sigma$ Modulators . . . . .	85
4.5	Stability Criteria. . . . .	87
4.6	Noise Transfer Function Determination . . . . .	89
4.6.1	Optimizing the position of zeros . . . . .	90
4.6.2	Optimizing the position of poles . . . . .	93

4.7	$\Delta\Sigma$ Modulator Assessment . . . . .	95
4.8	Summary . . . . .	96
	Problems . . . . .	97
	References . . . . .	99
5.	Implementation of $\Delta\Sigma$ Modulators	100
5.1	Introduction . . . . .	100
5.2	Basic Blocks of a $\Delta\Sigma$ Modulator for Analog-to-Digital Conversion . . . . .	100
5.2.1	The loop filter. . . . .	102
5.2.2	Local ADC and DAC. . . . .	106
5.3	Continuous-Time Loop Filter Implementation. . . . .	108
5.4	Circuit Design of a First-Order $\Delta\Sigma$ Modulator. . . . .	113
5.5	Circuit Design of a Second-Order $\Delta\Sigma$ Modulator. . . . .	114
5.6	Circuit Design of a Third-Order $\Delta\Sigma$ Modulator . . . . .	119
5.7	Circuit Design of a Fourth-Order Band-pass $\Delta\Sigma$ Modulator. . . . .	124
5.8	Testing the Operation of $\Delta\Sigma$ Modulators Experimentally . . . . .	127
5.9	Low-Power, Low-Voltage $\Delta\Sigma$ Modulators . . . . .	131
5.10	Summary . . . . .	134
	Problems . . . . .	135
	References . . . . .	137
	Appendix . . . . .	138
6.	Practical Limitations of $\Delta\Sigma$ Modulators	139
6.1	Introduction. . . . .	139
6.2	Practical Circuit Limitations . . . . .	139
6.2.1	Noise sources in the $\Delta\Sigma$ loop . . . . .	140
6.3	Effect of Thermal Noise . . . . .	144
6.3.1	SC circuits. . . . .	144
6.3.2	CT active RC circuits . . . . .	146
6.4	Effects of the Opamp Non-Idealities . . . . .	147
6.4.1	Thermal and flicker noise and DC offset of the opamp . . . . .	147
6.4.2	Finite opamp gain. . . . .	149
6.4.3	Finite bandwidth and slew-rate of the opamp. . . . .	151

6.5	Effect of Jitter . . . . .	152
6.5.1	SC circuits . . . . .	153
6.5.2	CT active RC circuits . . . . .	154
6.6	Effect of Rise and Fall Times of DAC Pulses in CT Circuits . . . . .	156
6.7	A Comparison of SC and CT active RC Circuits . . . . .	158
6.8	DAC Errors. . . . .	159
6.8.1	Single-bit DAC . . . . .	160
6.8.2	Multibit DAC . . . . .	161
6.9	Summary . . . . .	163
	Problems . . . . .	164
	References . . . . .	165
7.	Stabilization and Suppression of Tones for Higher-Order Single-Stage $\Delta\Sigma$ Modulators	167
7.1	Introduction . . . . .	167
7.2	Bounds on Quantizer Input Using the Variable Gain Method . . . . .	168
7.3	Stabilization Methods . . . . .	170
7.3.1	Resetting the integrators . . . . .	171
7.3.2	Clipping the integrators . . . . .	171
7.3.3	Activation of local feedback loops around the integrators . . . . .	172
7.3.4	Reducing the order of the loop filter . . . . .	173
7.4	Comparison of the Stabilization Methods. . . . .	175
7.5	Tones in $\Delta\Sigma$ Modulators . . . . .	182
7.6	Fixed Techniques. . . . .	184
7.6.1	Dithering. . . . .	184
7.6.2	Chaotic $\Delta\Sigma$ modulators . . . . .	185
7.6.3	Bit-flipping . . . . .	188
7.6.4	Comparison of fixed techniques. . . . .	191
7.7	Adaptive Techniques . . . . .	192
7.7.1	Adaptive dithering. . . . .	192
7.7.2	Adaptive bit-flipping . . . . .	193
7.7.3	Comparison of the adaptive techniques . . . . .	193

7.8 Summary . . . . .	193
References . . . . .	195
8. Decimation, Interpolation and Converters . . . . .	196
8.1 Introduction . . . . .	196
8.2 Rate Conversion . . . . .	197
8.2.1 Decimation . . . . .	197
8.2.2 Interpolation. . . . .	198
8.3 Decimation and Interpolation filters. . . . .	201
8.3.1 Multi-stage rate conversion . . . . .	203
8.3.2 Comb filters . . . . .	205
8.3.3 Fractional rate conversion. . . . .	207
8.4 $\Delta\Sigma$ Modulators for ADC. . . . .	208
8.5 $\Delta\Sigma$ Modulators for DAC. . . . .	211
8.6 Summary . . . . .	212
Problems . . . . .	212
References . . . . .	213
9. Applications	214
9.1 Introduction . . . . .	214
9.2 $\Delta\Sigma$ Modulation in Digital Radio . . . . .	214
9.2.1 Conversion bandwidth and resolution. . . . .	217
9.2.2 Sampling rate . . . . .	219
9.3 Frequency Synthesis . . . . .	221
9.3.1 Principles of the PLL operation . . . . .	221
9.3.2 $\Delta\Sigma$ modulation in frequency synthesis . . . . .	224
9.3.3 Analysis of the frequency divider . . . . .	226
9.4 Clock Generators Using $\Delta\Sigma$ Modulation . . . . .	229
9.5 $\Delta\Sigma$ Modulation in Analog-Input Digital Phase-Locked Loops for Frequency and Phase Demodulation . . . . .	232
9.6 $\Delta\Sigma$ Modulation in Analog Oscillators . . . . .	234
9.7 Summary . . . . .	236
References . . . . .	238
<i>Index</i>	241

## **Chapter 1**

# **Introduction**

### **1.1 Modulation – Demodulation**

Information is transferred from its source to the receiver through a channel. For this link to be successful certain rules have to be obeyed and requirements to be satisfied. In telecommunications the information is carried by a signal, voltage, current or electromagnetic wave. The initially produced signal, in order to deliver the information to the receiver accurately and safely, when passing through the channel, should not be contaminated by any type of unwanted signals, noise for instance. To avoid this possibility, before entering the channel, the signal is processed in order to take a form that will guarantee the accurate and safe delivery of the information it carries to the receiver [1]. The processing includes amplification but, most important, a type of transformation called modulation, under one of the schemes referred to as Amplitude Modulation (AM), Frequency Modulation (FM), Phase Modulation (PM), Pulse-Code Modulation (PCM), Differential PCM (DPCM), Delta Modulation (DM), Delta-Sigma Modulation ( $\Delta\Sigma$ ), etc.

In AM, FM and PM a carrier (high-frequency sine wave) has its amplitude, frequency or phase respectively modulated i.e. changed by the information carrying signal in the modulating circuit or system.

In PCM the initial signal is converted to pulse trains representing digital words, which correspond to samples of the signal usually obtained at equally spaced time intervals. In DM the difference in the signal from

sample to sample is quantized and transmitted through the channel. Finally, in  $\Delta\Sigma$  Modulation it is the integral of the difference between the information carrying signal and the modulator output signal that is quantized and then transmitted through the channel.

To convert a continuous-time signal to PCM a circuit called Analog-to-Digital Converter (ADC) is required. There is a number of various methods in use to build an ADC. One of them is using the  $\Delta\Sigma$  modulator, which has been proved very successful in this application.

The reverse process of modulation is to extract the information from the modulated signal. This process takes place in the receiver, is called demodulation, and is performed by the demodulator. In the case of a PCM signal, the process of extracting the information in the form of an analog signal is called Digital-to-Analog Conversion and requires the use of a Digital-to-Analog Converter (DAC). One method for building a DAC involves the use of  $\Delta\Sigma$  Modulation and this proves once more the usefulness of this type of modulation.

Studying the various parameters related to  $\Delta\Sigma$  Modulation and some of its numerous applications is the objective of this book.

## 1.2 $\Delta\Sigma$ Modulation

The Delta modulator was proposed in late forties [2] as the simple feedback loop shown in Fig. 1.1, for converting a low frequency analog signal into a bit stream which could be easily transferred through noisy channels. In early fifties the modulator was extensively studied and analyzed [3-6], whereas much work was done in the following years [7].

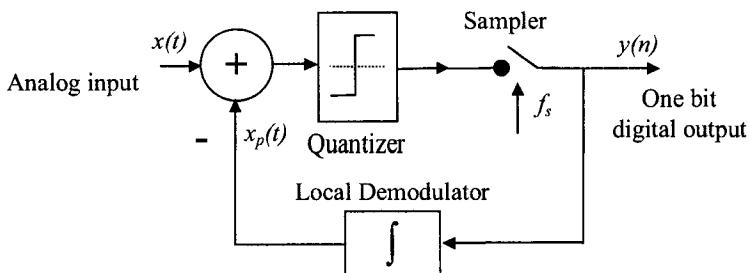
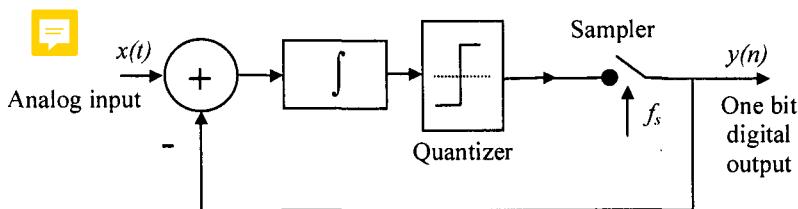


Fig. 1.1 The Delta Modulator

The binary pulses  $y(n)$  produced by the Delta modulator represent the sign of the difference between the input and feedback signals, hence the prefix delta. The integrator plays the role of the decoder so that the feedback signal  $x_p(t)$  would be always approximating the input analog signal  $x(t)$ . Consequently, the decoder circuit at the receiver should be exactly the same. The number of positive or negative pulses at the output bit stream depends on the slope of the input signal hence the name of the Delta-Modulator. The linear Delta-Modulator is difficult to be analyzed since the quantizer is a non-linear device. Furthermore, slope overload [7] is one of the drawbacks of Delta modulator, as this encoder cannot respond to fast changes of the input signal. To overcome this problem an integrator can be used in front of the Delta modulator to limit the amplitude of high frequencies. The use of the integrator results in encoding the integral of the input signal (summation –  $\Sigma$ ) and thus the resulting modulator is called Delta-Sigma or  $\Delta\Sigma$  modulator. The output bit stream is now related to the amplitude of the signal itself. The circuit of the  $\Delta\Sigma$  modulator can be simplified if the two integrators are moved inside the loop as shown in Fig. 1.2. The decoder for the  $\Delta\Sigma$  modulator is just a low pass filter.

Fig. 1.2 The  $\Delta\Sigma$  modulator

The Delta as well as the  $\Delta\Sigma$  modulators are oversampling converters in the sense that the sampling frequency is higher than twice the Nyquist frequency. As a result, the quantization error power is spread over a wider area of frequencies, compared to the signal power that remains within the signal band. Clearly, the circuit of  $\Delta\Sigma$  modulator behaves differently for the quantization error and the signal. This fact may require the realization of complicated transfer functions instead of the single integration, so that the signal would pass unaltered through the

modulator, while the quantization noise was high-pass filtered. In this way the quantization noise is separated from the signal in the frequency domain and can be removed by a high precision digital filter [9]. Using continuous-time or discrete-time circuitry  $\Delta\Sigma$  modulators of various orders and characteristics can be built [10]. These modulators accompanied with high-precision digital filters for the quantization noise removal can lead to Analog-to-Digital conversion circuits of high resolution [11].

The main advantage of  $\Delta\Sigma$  modulators is that they offer a very good separation of the input signal from the quantization noise due to oversampling and noise shaping. Thus, employing high precision digital circuitry a high quality digital signal can be obtained. With contemporary technology 24-bit Analog-to-Digital converters can be built for audio frequencies based on  $\Delta\Sigma$  converters. Signal bandwidths of 20 MHz can be encoded with 8 bit accuracy. Nevertheless, component non-idealities and other sources contribute to an increased noise floor in the digital signal. Various techniques can be employed for circuit matching and optimizing modulator performance [10].

### 1.3 Design and Implementation of $\Delta\Sigma$ Modulators

Each  $\Delta\Sigma$  modulator is characterized by its order. The one shown in Fig. 1.2 is a first-order in accordance with the order of the integrator. As it was mentioned in the previous section, to improve the design more noise power from the signal band has to be shaped, i.e. to be pushed out of the signal frequency band. To achieve this the integrator is replaced by a suitable filter, low-pass or band-pass, which has higher selectivity than the simple integrator. The order of this filter will determine the  $\Delta\Sigma$  modulator order, the two orders being the same. Thus a third-order  $\Delta\Sigma$  modulator will employ a low-pass filter of third-order. The transfer function of this filter should be such that it will shape the noise spectrum, however leaving the signal spectrum unaltered. This function is the so-called Noise Transfer Function (NTF) and can be obtained from the existing lists of such functions, as Butterworth, Chebychev and Inverse Chebychev etc. or by optimization.

The next step in the design of a  $\Delta\Sigma$  modulator would be to realize this NTF using one of the well-known methods for filter design. Here the technology has to be chosen on the basis of the requirements. Thus the filter can be realized as a continuous-time or as discrete-time circuit. There is an abundance of such circuits cited in the literature [12-14] and the designer should select the one they consider more suitable for their application. Depending on the signal frequency range and the oversampling frequency as well as on other criteria, the circuit of the filter can be implemented as an integrated circuit or using discrete components.

One characteristic of the  $\Delta\Sigma$  modulator of main importance is the ratio of the signal power over the power of the quantization noise lying inside the signal band. This Signal-to-Noise Ratio (SNR) depends on both the signal power and the effectiveness of the filter in reducing the noise power inside the signal band. Unfortunately the magnitude of the signal cannot be unlimited. Assuming the signal to be a pure sinusoid for example, its amplitude cannot be increased beyond a certain value, because then instability will occur and the modulator will become useless. So a figure of merit in a  $\Delta\Sigma$  modulator design will be the highest amplitude of the input signal before instability occurs, coupled with the corresponding maximum value of the achieved SNR.

The first-order  $\Delta\Sigma$  modulator shown in Fig. 1.2 is always stable. The problem of instability occurs in single-bit, higher-order modulator circuits, which are implemented in one-stage. To avoid the occurrence of instability it is possible to implement the high-order single-bit  $\Delta\Sigma$  modulator by properly cascading first-order stages. Although the problem of instability is solved this way, matching of the cascaded stages becomes necessary if the SNR is to be high. On the other hand, various stabilization methods have been developed [10] through which the amplitude of the input signal can be increased to some extent.

Apart from single-bit, multi-bit  $\Delta\Sigma$  modulators have been suggested offering certain advantages. However, the single-bit ones are more popular, mainly because of the simplicity of their implementation. Hybrid as well as adaptive  $\Delta\Sigma$  modulators have also been proposed, which are considered advantageous in certain applications. Regarding the

requirements in voltage and consumed power, low-power, low-voltage  $\Delta\Sigma$  modulators have been proved practical.

#### 1.4 Applications

$\Delta\Sigma$  modulators have been employed in numerous fields of signal acquisition and processing, with dominating the one of ADC and DAC converters [15]. ADCs and DACs based on  $\Delta\Sigma$  modulators have been devised for conversion in various frequency bands and resolutions. Accordingly, for low frequency signals, such as those met in medicine (ECG), resolution of 24 bits is achieved. For audio frequencies 20 bits is nowadays a typical resolution, while 14 bit ADCs are available for encoding signals with over 1 MHz bandwidths. Higher bandwidths necessary for video signals can be achieved by means of  $\Delta\Sigma$  modulators employing lower oversampling ratios.

An ADC is employed in a digital radio receiver at the IF stage. Band-pass  $\Delta\Sigma$  modulators have been extensively used to efficiently digitize the IF signal, exhibiting interesting characteristics for the suppression of out-of-band noise and channel selection [16,17]. An interesting application of  $\Delta\Sigma$  modulation is found in frequency synthesis as well as in frequency and phase demodulation [18,19]. In these cases, it is combined with a Phase-Locked Loop (PLL). Furthermore,  $\Delta\Sigma$  encoders are employed for implementing digitally programmable analog oscillators thus minimizing both the analog and the digital circuitry [20].

Many companies worldwide are involved in high-speed and/or high-resolution  $\Delta\Sigma$  modulators design and implementation [11,21]. Various types of integrated  $\Delta\Sigma$  encoders are available on the market, including low-voltage low-power 16-bit or 24-bit multi-channel ADCs, 24-bit signal conditioning ADCs, low noise, high dynamic range as well as ICs incorporating seventh-order  $\Delta\Sigma$  encoders [10]. As the CMOS technology feature size is continuously decreasing,  $\Delta\Sigma$  ADCs will operate at higher frequencies and finer resolution. In the meantime new application areas for  $\Delta\Sigma$  modulation are emerging some of which are driven by the wireless (mobile telephony and GPS facilities) and the Internet market [21].

## 1.5 Book Organization

The book is organized in 9 chapters including the present one which is introductory. Each chapter starts with an introduction and finishes with a summary of comments and conclusions, followed, in some chapters, by unsolved problems and bibliography. Solved problems are also included in each chapter.

Chapter 2 contains mostly background material on Analog-to-Digital Conversion and should be useful to the reader who is not an expert in digital signal processing. Basic concepts, like sampling and the sampling theorem, aliasing and the antialiasing filter, spectral images, quantization and the resulting quantization noise, characteristics under consideration when one studies A/D conversion, are briefly reviewed. Sampling of band-pass signals is also looked at. Then the effect of oversampling on the power spectrum is explained and the Delta Modulator, Linear and Exponential, is introduced as a system taking advantage of oversampling and prediction. The advantageous use of oversampling for the reduction in quantization noise is further improved by applying the principle of noise shaping to the oversampled signal. The latter serves as preliminary for the introduction to the concept of  $\Delta\Sigma$  modulation, which is the object of the following chapter 3.

Chapter 3 is of primary importance in this book, because it contains the architectures of the various  $\Delta\Sigma$  modulators. It starts with the introduction of the first-order  $\Delta\Sigma$  Modulator, which is analyzed, and its performance compared with that of the Linear and the Exponential Delta modulator, mainly for educational purposes. Then it extends this modulator to the second- and higher-order single-bit, single-stage modulators. The stability problems, which incur in the latter, are revealed and as alternative the multistage high-order  $\Delta\Sigma$  modulators are considered. Various other type  $\Delta\Sigma$  modulators, like multi-bit, hybrid and adaptive, are briefly introduced. Finally, the Band-pass  $\Delta\Sigma$  modulator, which is of high importance for the development of digital radio, is examined.

The objective of Chapter 4 is to design Noise Transfer Functions (NTF), which achieve an optimum trade off between maximum Signal-to-Noise Ratio,  $SNR_{max}$ , and maximum amplitude  $x_{max}$  of the input signal

before the modulator becomes unstable. For this purpose, first the well known models for the quantization noise of the single-bit, single-stage  $\Delta\Sigma$  modulators, namely, the Linear and Quasi-Linear models, are reviewed. Then the characteristics of a “good” NTF are stated. The stability of  $\Delta\Sigma$  Modulators is also examined and relevant criteria are given. Based on this knowledge and following an optimization procedure, useful NTFs are derived, which are proved advantageous with regard to the  $SNR_{max}$  –  $x_{max}$  trade off when compared to other NTF families.

Chapter 5 deals with the design and implementation of  $\Delta\Sigma$  modulators used in Analog-to-Digital Conversion. First the various stages in the  $\Delta\Sigma$  Modulator block diagram are identified. Since integrators are the basic stages in the implementation of the loop-filter, their realization as continuous-time and discrete-time circuits is presented. Besides, the implementation of the required local ADC and DAC is explained. A section is devoted to explaining how a continuous-time filter can be used to implement the discrete-time function required in the design of a  $\Delta\Sigma$  modulator. Then we proceed with the implementation of  $\Delta\Sigma$  modulators of orders 1 up to 4 giving the detailed circuit diagrams for both discrete-time and continuous-time realization. For some of these circuits we present experimental and simulated plots of the SNR (against the amplitude of input signal). The relevant experimental procedure is explained for those readers who would like to repeat the experiments or to test their own circuits. Finally, the case of low-power, low-voltage  $\Delta\Sigma$  modulators is briefly reviewed.

In chapter 6 various undesirable effects on the performance of the  $\Delta\Sigma$  Modulator are examined. These effects, which are caused by the non-ideality of the components, concern the Switch Capacitor as well as the Continuous Time implementations. The component imperfections do not cause the same problems and of the same severity to both implementations. Thermal and flicker noise, pulse jitter, opamp imperfections (namely, finite open-loop gain, slew-rate, settling-time) as well as the local DAC non-linearity (mainly of concern for multi-bit  $\Delta\Sigma$  modulators) are main factors that affect the performance of a  $\Delta\Sigma$  modulator. In some cases certain remedies are suggested.

The stability and tonal problems of higher-order, single-stage  $\Delta\Sigma$  modulators encountered throughout this book are considered in detail in Chapter 7. These modulators tend to become unstable as the magnitude of the input signal approaches a certain value, which is characteristic of each modulator circuit. In multi-bit  $\Delta\Sigma$  Modulators this critical magnitude is increased compared to that for the single-bit modulators of same order. Additionally, the noise can drive the modulator to instability even if the magnitude of the input signal is constrained. To overcome these problems, stabilization methods are presented and compared. The appearance of tones at the output sequence is then considered. These tones are troublesome even if the Signal-to-Noise Ratio is high. This problem is also examined in chapter 7, where some methods to suppress the tones are presented.

One of the most important applications of  $\Delta\Sigma$  modulators is in the design of Analog-to-Digital Converters (ADC) and Digital-to-Analog Converters (DAC). In this use the  $\Delta\Sigma$  modulator is followed by the necessary circuitry to reduce the sampling rate, since the modulator signals are oversampled. This process is called decimation. On the other hand the input signal to the DAC is sampled at low rate, which will have to be increased if it is to be processed by the  $\Delta\Sigma$  modulator. The process is called interpolation and together with decimation is referred to as rate-conversion. Various types of rate-conversion are examined in chapter 8 and suitable filters for achieving this are discussed. Finally, full ADC and DAC systems based on  $\Delta\Sigma$  modulation are presented.

The last chapter of the book is chapter 9, where some important applications of  $\Delta\Sigma$  modulation are presented, apart from the ADC and DAC, which are discussed in chapter 8. Digital radio today makes use of  $\Delta\Sigma$  modulators and this application is presented first. Frequency synthesis is another application, where the  $\Delta\Sigma$  modulator operates in combination with a Phase-Locked-Loop. Digitally programmable clocks with reduced jitter and oscillators with high accuracy in the generated frequency have highly benefited by the inclusion of  $\Delta\Sigma$  modulators in their circuitry. All these applications are reviewed in this chapter 9.

## References

- [1] J. Proakis and M. Salehi, *Communications Systems Engineering*, Prentice-Hall Inc., New Jersey, 2002.
- [2] E.M. Deloraine, S. Van Miero and B. Derjavitch, French Pat. 932140.
- [3] C.C. Cutler, U.S. Pat. 2605361, (29.7.52).
- [4] J.F. Schouten, F. de Jager and J.A. Greefkes, ‘Delta Modulation – a new modulation system for telecommunications’, *Communication News*, pp. 115-124, 1952.
- [5] F. De Jager, ‘Delta modulation – a new method of PCM transmission using 1 unit code’, *Philips Research Report*, Vol. 1, pp. 442-466, December 1952.
- [6] L.H. Zetterberg, ‘A comparison between delta modulation and Pulse Code Modulation’, *Ericsson Technics*, Vol. I, pp. 95-154, 1955.
- [7] R. Steele, *Delta Modulation System*, Pentech Press, London, 1975.
- [8] H. Inose, Y. Yasuda and J. Murskani, ‘A Communications System by Code Modulation Delta-Sigma’, *J. Inst. Elect. Eng. Japan*, 44, No. 11, pp. 1775-1780, November 1961.
- [9] J. Candy, ‘Decimation for Sigma-Delta Modulation’, *IEEE Trans. Commun.*, Vol. 34, pp. 72-76, 1986.
- [10] S. Norsworthy, R. Schreier, and G. Temes, *Delta-Sigma Data Converters, Theory, Design and Simulation*, New York, IEEE Press, 1997.
- [11] Analog Devices Product Index, <http://www.analog.com>.
- [12] T. Deliyannis, Y. Sun and J.K. Fidler, *Continuous-Time Active Filter Design*, CRC Press, London, 1999.
- [13] R. Schaumann, M.S. Ghausi and K.R. Laker, *Design of Analog Filters: Passive, Active RC and Switched-Capacitor*, Prentice-Hall Intern, New Jersey, 1990.
- [14] R. Gregorian and G.C. Temes, *Analog MOS Integrated Circuits for Signal Processing*, John Wiley and Sons, New York, 1986.
- [15] R.W. Stewart and E. Pfann, ‘Oversampling and sigma-delta strategies for data conversion’, *Electronics & Communication Engineering Journal*, pp. 37-47, February 1998.
- [16] S.A. Jantzi, K. W. Martin and A.S. Sedra, ‘Quadratic Bandpass  $\Delta\Sigma$  Modulation for Digital Radio’, *IEEE Journal of Solid-State Circuits*, Vol. 32, No. 12, pp.1935-1950, December 1997.
- [17] L. Louis, J. Abcarius and G.W. Roberts, ‘An Eight-Order Band-pass  $\Delta\Sigma$  Modulator for A/D Conversion in Digital Radio’, *IEEE Journal of Solid-State Circuits*, Vol. 34, No. 4, pp. 423-431, April 1999.
- [18] T.A.D. Riley, M.A. Copeland and T.A. Kwasniewski, ‘Delta-Sigma Modulation in Fractional-n Frequency Synthesis’, *IEEE Journal of Solid-State Circuits*, Vol. 28, No. 5, pp. 553-559, May 1993.
- [19] I. Galton, ‘Analog-Input Digital Phase-Locked Loops for Precise Frequency and Phase Demodulation’, *IEEE Trans. on Circuits and Systems-II: Analog and Digital Signal Processing*, Vol. 42, No. 2, pp. 621-630, October 1995.

- [20] A.K. Lu, G.W. Roberts and D.A. Johns, 'A High-Quality Analog Oscillator Using Oversampling D/A Conversion Techniques', *IEEE Trans. on Circuits and Systems-II: Analog and Digital Signal Processing*, Vol. 41, No. 7, pp. 437-443, July 1994.
- [21] S. Yan and E.S. Sinencio, 'Broad band Continuous-time Sigma-Delta Analog-to-Digital Conversion Using MOSIS AMI 0.5 um CMOS Technology', *A Research/Educational Report*, Department of Electrical Engineering, Analog and Mixed-Signal Center, Texas A&M University, June, 2001.

## **Chapter 2**

# **Analog to Digital Conversion**

### **2.1 Introduction**

Most signals in nature are of analog form representing the continuous variation of physical quantities in time or in space [1,2]. Representative examples are the sound, the temperature, and various types of images, medical signals and seismograms. The processing of these signals is more effectively performed in the digital domain. Digital Signal Processing (DSP) techniques [3] present serious advantages over analog processing. Among them the following are prominent:

- a. High accuracy
- b. Loss-less and high-density storage with perfect reproducibility
- c. Flexibility and high performance in realizing various functions
- d. Low cost
- e. Small size, low power consumption and high throughput rate
- f. High reliability.

Analog-to-Digital Converters as well as Digital-to-Analog Converters (DACs) [5,6] are very essential electronic components, since they act as interface between the digital system and the real world as shown in Fig. 2.1. The Analog-to-Digital Conversion process aims at the representation of an analog signal by a sequence of binary numbers (digital signal). The digital signal can be easily processed (digital signal processing), transmitted (digital transmission), stored or converted back to analog signal, so that it could be properly detected by the real world sensors.

Analog-to-Digital Conversion can be carried out using the conventional Nyquist converters or those, which are based on other techniques. Conventional high-resolution A/D converters, such as successive approximation and flash type converters, operating at sampling frequency approximately equal to twice the maximum frequency in the input signal (Nyquist rate), often do not make use of the exceptionally high speeds achieved with VLSI technology. Additionally, the analog circuitry required in conventional A/D converters limits their accuracy in representing the digital signal. High precision successive approximation techniques achieve up to 12 or 14 bits accuracy for the digital samples, while in the case of flash type converters this accuracy is smaller [7]. In various applications, such as high fidelity audio systems, the above accuracy is inadequate, resulting in poor performance of the system.

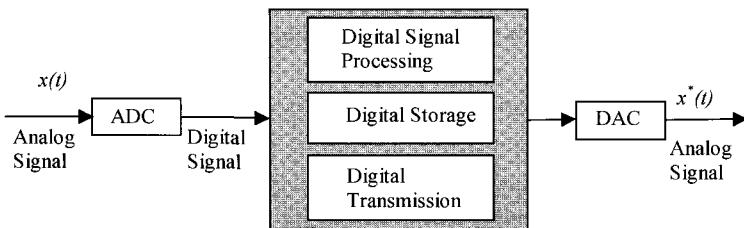


Fig. 2.1 ADC and DAC is the interface between digital systems and real world.

Delta-Sigma ( $\Delta\Sigma$ ) modulation based analog-to-digital (A/D) conversion technology is a cost effective alternative for high resolution (greater than 12 bits) converters, which can be ultimately integrated with digital signal processor ICs [5,6,8]. Although the delta-sigma modulator was first introduced in 1962 [9], it did not gain importance until recent developments in digital VLSI technologies.  $\Delta\Sigma$  A/D converters use a low resolution A/D converter (1-bit quantizer), noise shaping and oversampling rate (64 times is a typical value). As the signal is oversampled and the quantization noise is shaped out of band, high resolution is achieved by decimation (sample-rate reduction).

In this chapter first the basic operations required for such conversions are introduced and explained. They include sampling, quantization and encoding. Next, the resulting quantization error and the Signal-to-Noise

Ratio are studied. Then oversampling is discussed and the  $\Delta$  modulator is presented and studied briefly being considered the forerunner of the  $\Delta\Sigma$  Modulator.

## 2.2 The Basic Concept of A/D Conversion

The Analog-to-Digital Conversion process involves two consecutive operations on the original analog signal. The first is uniform sampling whereas the second one is uniform quantization [4,10]. The implementation of these operations is carried out by means of special circuitry, which is essential in building A/D Converters. After that, the encoding follows which gives the final digital signal. All the steps required in a conventional ADC are depicted in Fig. 2.2.

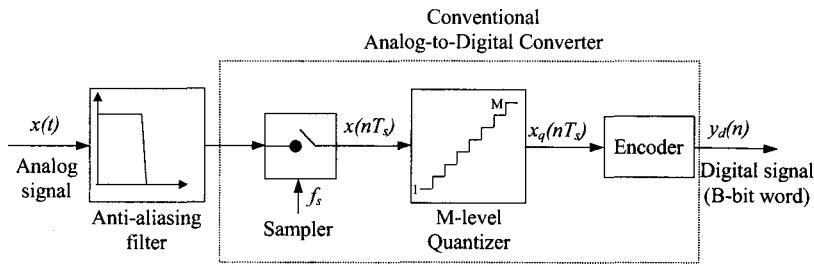


Fig. 2.2 Block diagram for conventional analog to digital conversion

According to Fig. 2.2, the input signal  $x(t)$  is initially sampled at uniform time intervals  $T_s$  by means of the sampler circuit. The resulting sequence  $x(nT_s)$  or simpler  $x(n)$  is called the sampled signal. In order to avoid loss of information the sampling rate  $f_s$  has to be at least twice that of the signal bandwidth. An *antialiasing filter* is used before the ADC to guarantee this condition. Each sample  $x(n)$  will be finally converted to the digital signal  $y_d(n)$  which is a B-bit word. These B bits, which determine  $M=2^B$  distinct equidistant levels in the full dynamic range spanned by the converter, characterize its quality. Initially, the samples  $x(n)$  are quantized, in the M-level quantizer, which means that their values are rounded to reach the closest quantization level. After that each quantized

sample is encoded to the corresponding B-bit binary word. The whole process is discussed in the following sections.

### 2.3 Uniform Sampling

The sampling process is essential for the analog signal to be recorded without loss of information. The condition that has to be fulfilled for a lossless sampling is the well-known *sampling theorem*:

*If the highest spectral component of the signal is  $f_b$ , the signal has to be sampled at least at a sampling rate  $f_s$  twice the size of  $f_b$  i.e.*

$$f_s \geq 2f_b \quad (2.1)$$

If, for example, a band-limited signal with highest frequency equal to 4 kHz has to be converted to digital, a sampling rate of at least 8 kHz is necessary. In case the sampling rate is lower than the one specified by the sampling theorem, the obtained sampled signal contains, at low frequencies, illusory spectral components. This phenomenon is called aliasing. As a result the initial analog signal cannot be reliably reconstructed from its samples.

The analysis of the sampling theorem and a deep insight in the sampling process requires some basic concepts from the Fourier analysis [11]. The ideal sampling is examined by means of the schematic representation shown in Fig. 2.3. The analog signal  $x(t)$  is band-limited with spectral contents in the range  $[-f_b, f_b]$ . This input signal is sampled uniformly using the infinite train of impulses  $p(t)$ :

$$p(t) = \sum_{n=-\infty}^{\infty} \delta(t - nT_s) \quad (2.2)$$

This is achieved by means of the ideal sampler shown in the same Figure. The sampler is controlled by the pulse train  $p(t)$  and is considered ideal in the sense that the pulse train is an ideal signal, not easily realizable. The signal  $x'(t)$  at the output of the sampler consists of equally spaced samples obtained from the multiplication of  $x(t)$  by  $p(t)$ :

$$x'(t) = p(t) \cdot x(t) = \sum_{n=-\infty}^{\infty} x(t) \delta(t - nT_s) = \sum_{n=-\infty}^{\infty} x(nT_s) \delta(t - nT_s) \quad (2.3)$$

On the left side of Fig. 2.3 the signals  $x(t)$ ,  $p(t)$  and  $x'(t)$  are shown and on the right of the same Figure their spectra. The spectrum  $X(f)$  of  $x(t)$  is band-limited while the spectrum of  $p(t)$  is also an impulse train in the frequency domain with distance between the impulses  $f_s = 1/T_s$ . This means that the denser the sampling impulse train is getting the sparser the corresponding spectrum becomes.

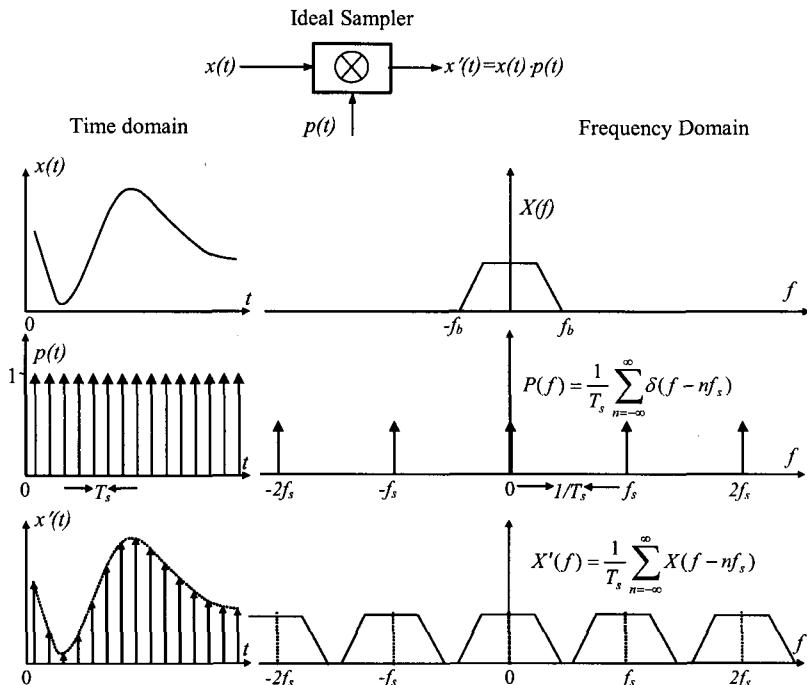


Fig. 2.3 The sampling processes of an analog signal results in a periodic spectral content for the sampled signal.

Finally, the spectrum of the signal  $x'(t)$  is the convolution of the spectra of the signals  $x(t)$  and  $p(t)$ , since multiplication in the time domain implies convolution in the frequency domain. As a result of the convolution process, the spectrum  $X'(f)$  of the signal  $x'(t)$  is a periodic repetition of  $X(f)$  with period  $f_s = 1/T_s$ , which is the period of  $P(f)$ . The spectrum of  $x'(t)$  consists of the frequency band of the initial signal as well as the

repetition of this band, which extends theoretically to infinity. The new bands are called '*images*' of the initial band. The first of them is centered at  $f_s$ .

In case the initial analog signal is to be reconstructed, all the '*images*' have to be eliminated. This can be achieved using some kind of low-pass filtering with cut-off frequency in the middle of the range  $[0, f_s]$ . This presupposes that the '*images*' do not overlap, or, otherwise, the spectral power created from the sampling process will not be spread into the base band. To marginally avoid such an overlapping of the images, as it can be easily concluded from Fig. 2.4, the value of  $f_s$  must be at least double the higher frequency  $f_b$  in the signal. This is actually what the sampling theorem states.

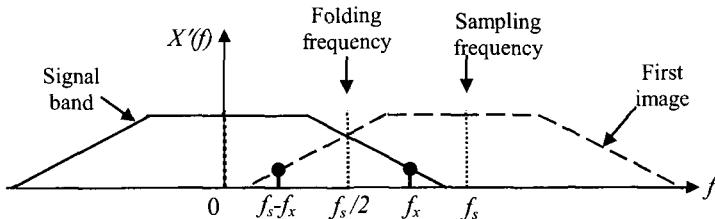


Fig. 2.4 Aliasing. The frequencies of the analog signals, which are above  $f_s/2$ , are found in the reconstructed signal at a lower frequency, symmetric with reference to the so-called 'folding' frequency  $f_s/2$ .

However, if overlapping occurs, then spectral energy of the first image enters the original signal band. In this case every frequency component  $f_x$  in the original analog signal, which is greater than  $f_s/2$ , cannot be resolved in the reconstructed signal in its proper position. It appears folded around  $f_s/2$  (folding or Nyquist frequency) in a new lower frequency position  $f_s-f_x$ . Thus the initial analog signal cannot be properly reconstructed. In the example of Fig. 2.5, since  $f_x=5f_s/6$ , it appears in the reconstructed signal at  $f_s-f_x=f_s/6$  or  $f_x/5$ . The phenomenon is non-linear and is called '*aliasing*' as we saw above. It is depicted graphically and explained theoretically in Figs. 2.4 and 2.5, respectively.

The aliasing can only be prevented by properly low-pass filtering the input signal up to the Nyquist frequency. This low-pass filter, included in Fig. 2.2, is called the *anti-aliasing* filter. Its response must be flat over the frequency band of interest (base band) and attenuate the frequencies

above the Nyquist frequency enough to bring them under the noise floor. Since the analog anti-aliasing filter is a limiting factor in controlling the bandwidth and phase distortion of the input signal, a high performance anti-aliasing filter is required to obtain high resolution and minimum distortion.

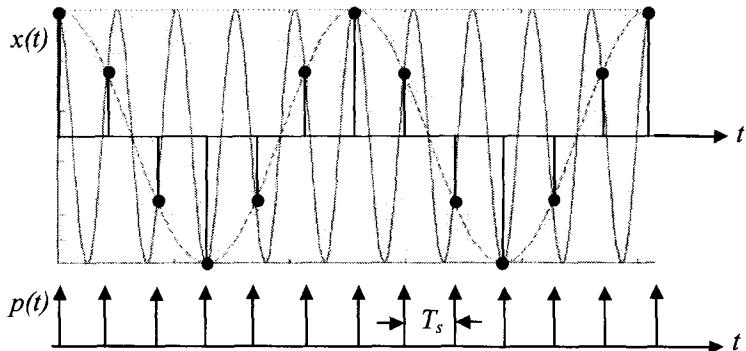


Fig. 2.5 The aliasing phenomenon. A high frequency component  $f_x$  (solid wave) is sampled with sampling frequency  $6f_x/5$ . The frequency of the reconstructed signal (dashed wave) is  $f_x/5$ .

In addition to an anti-aliasing filter, a sample-and-hold circuit is required. Although the analog signal is continuously changing, the output of the sample-and-hold circuitry must be constant between samples so that the signal could be quantized properly. This allows the converter enough time to compare each individual input sample with the internally generated reference levels [7], in order to give a more reliable digital output signal. Furthermore, the impulses  $p(t)$  are not realizable. In practice they have a finite width, which affects the spectrum of the digital signal. Consequently, the sampling process in practice is not the ideal one shown in Fig. 2.3.

The signal is sampled and held constant for all the time period  $T_s$ . The impulse response  $h(t)$  of the sample-and-hold circuit is a pulse of height 1 and width  $T_s$ . Its spectrum  $H(f)$ , shown in Fig. 2.6.a, is  $H(f) = e^{-j\pi f T_s} \sin(\pi f T_s) / (\pi f)$ . Therefore, the spectrum of the sampled signal (Fig. 2.6.c) is that of the ideally sampled signal (Fig. 2.6.b) modu-

lated by the function  $|H(f)|$ . Obviously, this process significantly distorts the base band of the signal. This distortion is equalized during the reconstruction process by means of a specially designed low-pass filter (Fig. 2.6.d).

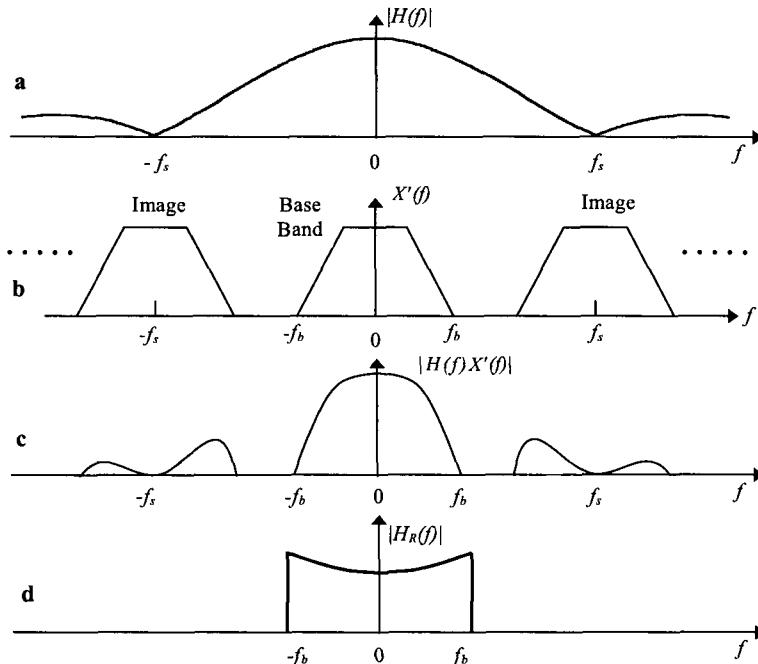


Fig. 2.6 a. Frequency response of the holding circuit. b. Spectrum of the ideally sampled signal c. Spectrum of the sampled signal and d. Low-pass reconstruction filter.

**Example 1.1** An analog signal  $x(t)$  is given in the time domain as follows:

$$x(t)=2 \cos(800 \cdot \pi \cdot t) \text{ Volts}$$

Determine the frequency of the reconstructed analog signal if the sampling rate is 500 Hz.

**Solution** The argument of the cosine term corresponds to  $2\pi f_x t$ . Accordingly the frequency of the signal equals 400 Hz. This signal is not properly sampled by the sampling rate of 500 Hz and using Fig. 2.4 it is concluded that a frequency component is recovered at  $f_s - f_x = 100$  Hz.

## 2.4 Quantization Error and the Linear Model

The sampling process in an ADC is followed by the quantization process. The quantization is the conversion of a continuous valued sample  $x$  (which has infinite resolution by definition), into one of a finite set of discrete values  $q_i$ . The value  $x$  varies in the interval  $(x_{min}, x_{max})$ , while  $q_i$  takes values from the set  $\{q_1, q_2, \dots, q_M\}$ . The number  $M$  of the discrete values  $q_i$  is determined by the type of the quantizer and its transfer function  $q(x)$ . For a uniform quantizer, two types of which are shown in Fig. 2.7a and 2.7b, the sub-intervals  $\Delta = q_{i+1} - q_i$  are equal. This quantizer is more commonly used, although it is not always the most efficient. For an efficient quantizer the sub-intervals  $\Delta$  in the range  $x$  are determined on the basis of the probability density function (PDF) of  $x$ .

The difference  $e(n) = q_i(n) - x(n)$ , which results when  $x(n)$  is approximated by  $q_i(n)$ , is called quantization error and depends on how the signal is being approximated. The quantization error is of the order of  $\Delta$  and can be quite small compared to full-scale (FS) signals, depending on the number of the quantization levels. A full-scale signal has a peak-to-peak variation equal to  $(x_{max} - x_{min})$ . If the input signal exceeds this value, conversion is not properly carried out (overload case), while the quantization error becomes larger than  $\Delta$  (see Fig. 2.7c). On the other hand, as the input signal gets smaller, the quantization error becomes a higher percentage of the total signal.

Since the final digital signal is represented by a binary number of  $B$ -bits, a total of  $M=2^B$  quantization levels are available. Assuming that the sequence  $x(n)$  is scaled such that  $|x(n)| \leq 1$  for fractional number representation, the pertinent dynamic range is 2. For a uniform quantizer, the interval between successive levels,  $\Delta$ , is therefore given by

$$\Delta = \frac{2}{2^B - 1} \quad (2.4)$$

which is called the quantization step size. The sampled input value  $x(n)$  is then rounded to the nearest level. The procedure is demonstrated in Fig. 2.8, for the simple case of representing the signal samples with a 3-bit word. The full signal range contains  $2^3=8$  equidistant levels. The quantization step  $\Delta$  is considered equal to 1, which is  $1/7$  of the total range.

Thus the quantization error, being of the order of  $\Delta$ , corresponds to an error of one least-significant-bit (LSB) of the digital signal. In the quantization process the initial value of each sample is altered so that it can reach the closest level. This alteration gives rise to the quantization error  $e(n)$ , which has maximum absolute value  $0.5\Delta$ . If the converter has a 2V input dynamic range then  $\Delta=0.29V$  and the quantization error is unacceptably large. However, if a 16-bit A/D converter is used, which is the standard for high accuracy A/D converters, a total of  $2^{16}-1 = 65535$  different reference levels are available. For a converter with a 2V input dynamic range, the spacing of these levels is only  $30 \mu V$  apart.

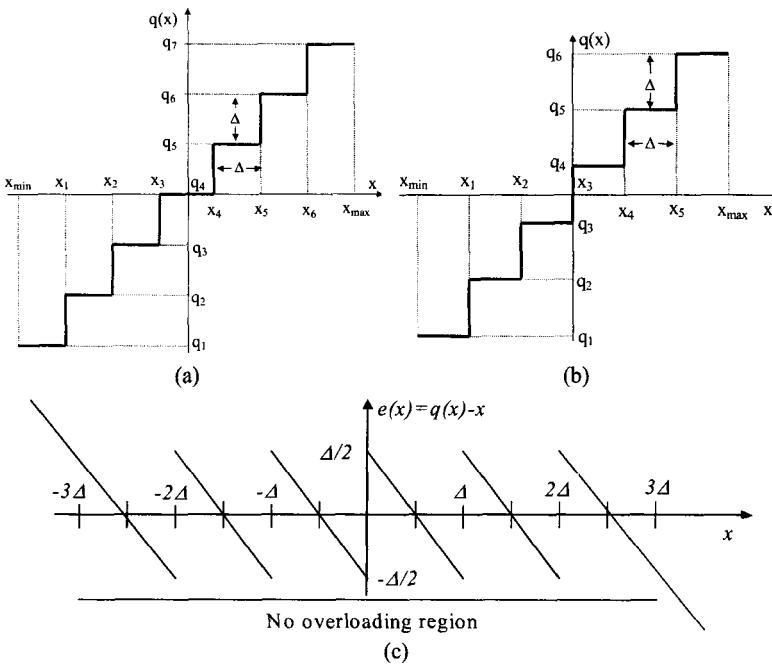


Fig. 2.7 a. Transfer function of a uniform quantizer with 7 levels. One of them is the zero-level (midread). b. Transfer function of a uniform quantizer with 6 levels (midriser). c. Quantization error as a function of the input signal  $x$  (midriser case).

From Fig. 2.8 it follows that the ADC output  $x_q(n)$  is the sum of the sampled signal  $x(n)$  and an error component  $e(n)$ , which is called *quantization noise*. Then

$$x_q(n) = x(n) + e(n) \quad (2.5)$$

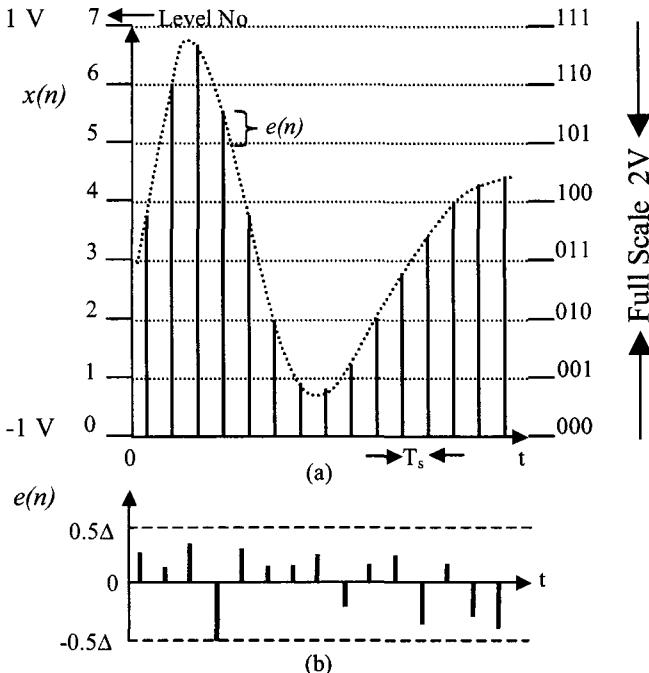


Fig. 2.8 In the quantization process the values of the samples are rounded to the closest available level and then are represented in the respective binary code. The alteration of the initial samples results in the quantization noise  $e(n)$ .

Accordingly, the quantization process can be considered to be a linear operation as shown in Fig. 2.9. In practice of course, this is not true, since the quantizer is a non-linear circuit. However, the linear model for the quantizer can be employed when the number of quantization levels is large and equally probable. In this case the quantization noise  $e(n)$  is almost uncorrelated with the input signal, has a white spectrum and its probability density function (PDF) is uniform in the range  $[-\Delta/2, \Delta/2]$ .

Consequently, the quantization error can be regarded as an independent additive white noise source.

The quantization noise affects the quality of the signal. This is quantitatively expressed using the Signal-to-Quantization-Noise Ratio (SQNR), i.e. the ratio of the signal power over the power of the quantization noise

$$SQNR = \frac{\text{Signal Power}}{\text{Quantization Noise Power}} \quad (2.6)$$

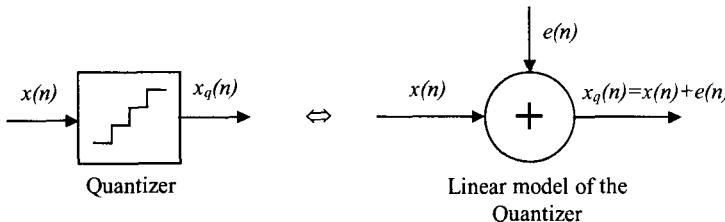


Fig. 2.9 The linear model for the quantizer. The quantization error is considered to be an additive noise source.

For an input signal, which is large compared to the quantization step  $\Delta$ , the error term  $e(n)$  is a random quantity uniformly distributed in the interval  $(-\Delta/2, \Delta/2)$ . In this case its mean value is zero and its power is the variance  $\sigma_e^2$  given by

$$\sigma_e^2 = E\{e^2\} = \int_{-\Delta/2}^{\Delta/2} e^2 f_e(e) de = \frac{1}{\Delta} \int_{-\Delta/2}^{\Delta/2} e^2 de = \frac{\Delta^2}{12} \quad (2.7)$$

where  $E\{\cdot\}$  denotes statistical expectation. If the values of  $e(n)$  are assumed uncorrelated and identically distributed, the quantization noise is white and its power is spread uniformly over the entire frequency range  $[-f_s/2, f_s/2]$  as shown in Fig. 2.10. Thus the power spectral density (PSD) of the noise  $N(f)$  can be expressed as

$$N(f) = \frac{\Delta^2}{12 f_s} \quad (2.8)$$

For a sine wave input signal with full scale amplitude variation  $2A = (2^B - 1)\Delta$ , its power is  $A^2/2$  and the SQNR is expressed as

$$SQNR = 10 \log \left( \frac{A^2 / 2}{\Delta^2 / 12} \right) \cong 10 \log \left( \frac{3 \cdot 2^{2B}}{2} \right) = (6.02 \cdot B + 1.76) \text{ dB} \quad (2.9)$$

Consequently, for a sine wave signal with maximum amplitude, the maximum achieved SQNR and the quantization error possessing the above mentioned properties, depend on the number  $B$  of bits used to represent these samples. Specifically,

*Increasing the number of bits by one, the quality of the digital signal, represented by the SQNR, increases by 6 dBs.*

Furthermore, relationship (2.9) determines the maximum number of bits required for quantizing an analog signal with specific noise floor. So, if this noise floor is determined by a Signal-to-Noise Ratio (SNR) of 58 dBs, 10 bits will be adequate for quantization. More bits in the digitization accuracy would actually contribute to digitizing the noise.

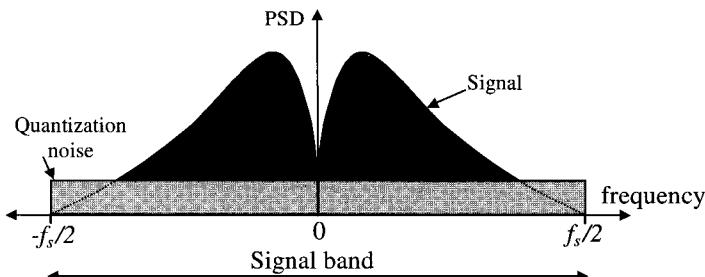


Fig. 2.10 Power spectral density (PSD) at the output of the conventional ADC. The quantization noise spreads all over the signal band.

Another quantity that characterizes the Quantizer is its dynamic range. It is related to its resolving power and depends on the smallest signal it can recognize and quantize. This signal is usually of the order of the maximum value of the quantization error. Accordingly, the dynamic range is expressed in bits as follows:

$$\text{dynamic range} = \log_2 \frac{x_{\max}}{\Delta/2} \quad (2.10)$$

## 2.5 Sampling of Band-Pass Signals

The sampling of band-pass signals is carried out based on the same principles as those applied to band limited signals in Section 2.3. However, it can be shown, that for adequate sampling, it is the bandwidth of the signal that determines the required sampling rate [12]. In Fig. 2.11.a the spectrum of a band-pass signal is demonstrated. The signal consists of two symmetric spectral bands. The positive one (PB) is restricted within the frequency range  $[f_L, f_H]$  resulting in a bandwidth

$$B = f_H - f_L \quad (2.11)$$

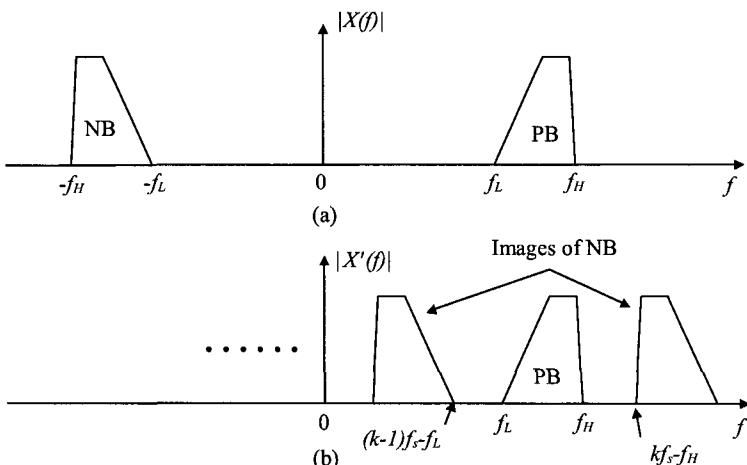


Fig. 2.11 a. Spectrum of the band-pass signal. b. Images resulting from sampling process must not overlap.

The symmetric negative band (NB) displays similar characteristics. The sampling process will create ‘images’ of both bands. When  $f_s < 2f_H$ , some of these ‘images’ will lie in the region  $[-f_L, f_L]$ . However, overlapping of the ‘images’ with the original bands must be avoided. Aliasing occurs when an ‘image’ of the NB overlaps with the PB. This simultaneously means that an ‘image’ of the PB overlaps with the NB. As it is shown in Fig. 2.11.b, in order to avoid overlapping of the PB with the  $k$ -th ‘image’ of the NB the band PB must lie between the  $(k-1)$ -th and the  $k$ -th images of NB i.e.

$$(k-1)f_s - f_L \leq f_L \quad (2.12a)$$

$$kf_s - f_H \geq f_H \quad (2.12b)$$

If  $f_H$  is expressed in terms of  $B$  i.e.  $f_H=rB$ , substituting from (2.11) in (2.12) gives

$$f_s \leq 2B(r-1)/(k-1) \quad (2.13a)$$

$$f_s \geq 2Br/k \quad (2.13b)$$

Equations (2.13) give the functions  $f_1(r, k) = 2B(r-1)/(k-1)$  and  $f_2(r, k) = 2Br/k$ . These functions with variable  $r$  and parameter  $k$  divide the plane  $(f_s, f_H)$  in regions as shown in Fig. 2.12. For each  $k$  a different non-shaded area is obtained, containing the set of points  $(f_s, f_H)$  for which aliasing is avoided, with  $f_H$  and  $f_s$  expressed in units of  $B$ . The lines  $f_1$  and  $f_2$  always intersect, for  $r=k$ , on the line  $f_s=2B$ . Consequently, the sampling frequency cannot be smaller than  $2B$ .

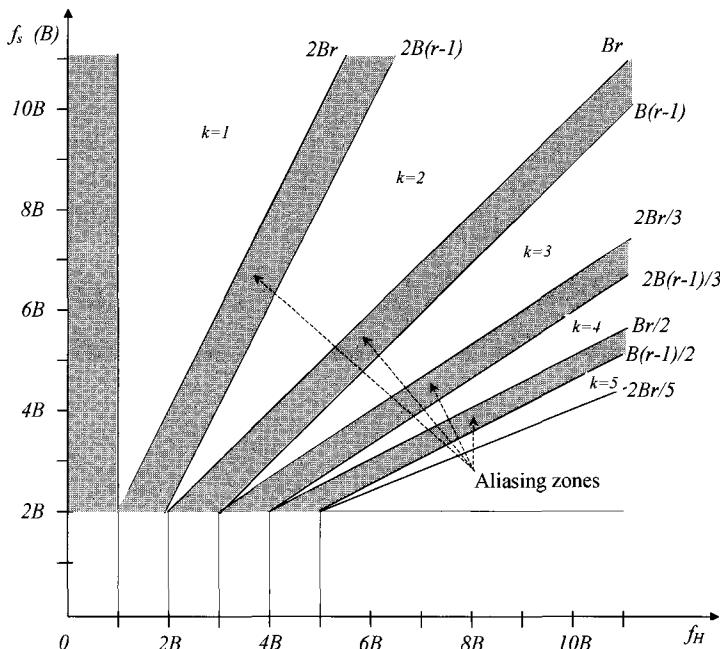


Fig. 2.12 Map for determining the sampling frequency for band-pass signals. White regions contain the appropriate points.

**Example 1.2** Suppose that the spectrum of a band-pass signal lies between  $f_L=7$  and  $f_H=9$  kHz. Which are the appropriate sampling frequencies?

**Solution** The signal band is  $B=2$  kHz. Therefore,  $f_H=4.5B$ . Using Fig. 2.12 we have to draw a vertical line on  $f_H=4.5B$ . The prohibited regions are those parts of this line with ordinates in the aliasing zones.

## 2.6 Oversampling Principles

The high resolution and dynamic range requirements in modern signal processing cannot be fully satisfied by the conventional ADCs, because of limitations in their implementation. Specifically, the limitations of technology refer to the following:

- a. The implementation of uniform quantizers with a high number of quantization levels. For a successive approximation ADC with 16 bits accuracy,  $2^{16}=65536$  equidistant voltage levels must be determined. This can be hardly achieved with current VLSI technology. Although, LASER trimming techniques [13] improve this situation, they result in complicated Nyquist converters.
- b. The implementation of the analog anti-aliasing filter with very strict requirements, such as very narrow transition-band, high attenuation in the stop-band ( $>90$  dBs), very small pass-band ripple (error), phase linearity, low noise etc. Such specifications cannot be achieved in analog integrated circuits.
- c. The presence of the jitter effect, i.e. the uncertainty in timing of the clock pulse edges used in the sampler.

One way of improving the situation is to increase the sampling rate many times higher than that of the conventional ADCs, i.e. above the Nyquist rate  $f_N = 2f_b$ . Of course this requires the various components of the ADC to operate at a much higher frequency. Sampling at a higher rate  $f_s$ , higher than the Nyquist rate  $f_N$ , is called *oversampling*. A measure of this oversampling is the *Oversampling Ratio (OSR)*,  $R$ , defined as follows:

$$R \equiv \frac{f_s}{f_N} \quad (2.14)$$

Usually, the value of OSR is taken to be a power of 2. If the OSR is between 2 and 16 we talk about a mild oversampling, whereas heavy oversampling occurs if the OSR is between 16 and 256.

The result of the oversampling approach is graphically depicted in Fig. 2.13 for  $R=4$ . It is evident that the images of the signal band are not so close to one another and, consequently, the specifications of the antialiasing filter can be relaxed. Furthermore, the quality of the digital signal, as far as the SQNR is concerned, is improved when oversampling is applied. This can be proved by means of Fig. 2.14. When oversampling is used, the quantization noise power is distributed in a larger frequency range. Consequently, the power of the part of the quantization noise lying in the signal band is reduced. The quantization noise lying outside the signal band can be eliminated by means of a high accuracy digital filter. Accordingly, for an oversampling converter with  $OSR=R$  and a full scale input sine wave, the SQNR is evaluated in dBs as follows:

$$SQNR_{oversampling} = 10\log \frac{A^2/2}{(\Delta^2/12)/R} = SQNR_{Nyquist} 10\log(R) \quad (2.15)$$

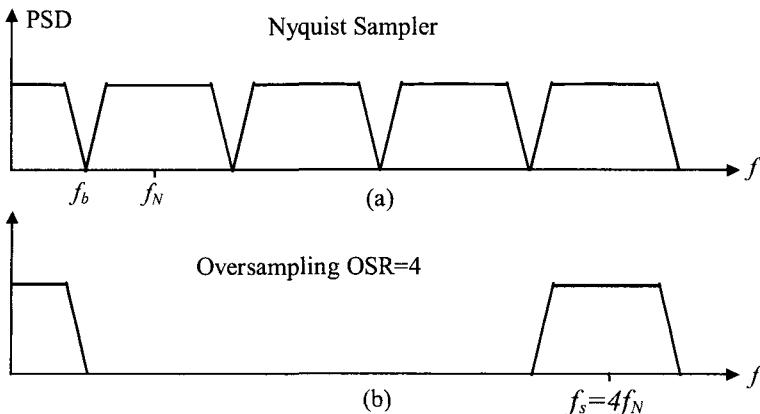


Fig. 2.13 The oversampling process takes apart the images of the signal band.

The last equation shows that the SQNR can be improved by 3 dB if the sampling rate is doubled and the out-of-signal-band noise is eliminated. This improvement is equivalent to using half a bit in the resolution of the quantizer. It is evident that for digitizing audio signals, instead of using a Nyquist ADC with sampling rate of 44.1 kHz and 12 bit accuracy, an oversampling converter can be employed with sampling rate  $4 \times 44.1 = 176.4$  kHz with 11 bit accuracy. This process can be followed to reduce the accuracy of the ADC to one bit, but then the sampling rate should be increased to  $44.1 \text{ kHz} \times 4^{11} = 185 \text{ GHz}$ . In fact this sampling rate is not realizable with present day technology.

However, the 1-bit ADCs that are based on  $\Delta\Sigma$  modulators, achieve a higher improvement in the SQNR for each doubling of the sampling rate, since the quantization noise is almost totally moved out of the signal band by means of noise shaping techniques. Thus a high value for the SQNR can be reached for lower oversampling rates. This technique is discussed thoroughly throughout this book. In this chapter, the basic oversampling ADC of one bit, the Delta Modulator (DM), is introduced.

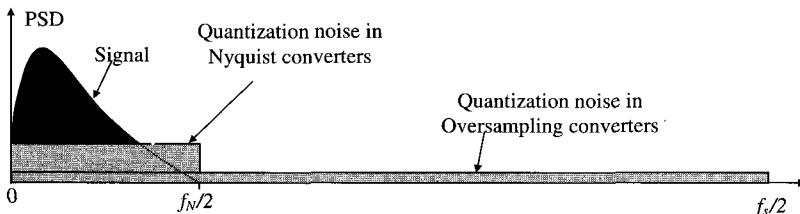


Fig. 2.14 When the sampling rate increases (4 times) the quantization noise spreads over a larger region. The quantization noise power in the signal band is 4 times smaller.

## 2.7 The Delta Modulator

In addition to the substantial improvement in the SQNR, oversampling possesses inherently the motivation for prediction. Specifically, the signal does not change significantly in the interval between successive samples when it is oversampled. This can lead to a reduction in the number of the quantization levels if the difference of two consecutive samples is encoded (Differential Pulse Code Modulation – DPCM). Since the values

of these samples are very close, they are highly correlated and therefore future samples could be predicted from the past ones. A modulator whose operation is based on this prediction principle is called *predictive modulator*. The simplest predictive modulator is the *Linear Delta Modulator* (DM) and its simplified version the *Exponential Delta Modulator* [8,9,13].

The Delta Modulator is a 1-bit ADC that was initially developed for digitizing analog signals in noisy environments. Its basic advantage was its simple circuitry. An implementation architecture of the Linear Delta Modulator is shown in Fig. 2.15a (switched-capacitor [13]). A model for the circuit in Fig. 2.15a is depicted in 2.15b. In 2.15c the sampler is moved to the input allowing the discrete-time representation of the delta modulator. In Fig. 2.15d is schematically demonstrated the reconstruction of a signal after the transmission of the modulated signal through a transmission line.

The DM shown in Fig. 2.15a (and 2.15b) consists of a Differential Loop, where the difference  $x(t) - x_p(t)$  is quantized to give the output sequence  $y(n)$  so that the signal  $x_p(t)$  should track the input signal  $x(t)$  all the time. The operation of DM can be better understood by means of Fig. 2.16. During the initial stages of operation the predicted signal  $x_p(t)$  tries to catch up with the waveform  $x(t)$ . This is the acquisition mode of operation. When this is achieved the predictor locks itself with the waveform  $x(t)$  (tracking mode) and can follow its changes with an error always smaller than  $|e(n)| \leq \Delta/2$ . Nevertheless, if the signal  $x(t)$  changes quickly, faster than  $\Delta/2T_s$ , the predictor starts losing track of the waveform  $x(t)$ . In this case slope overload results. When the signal  $x(t)$  ceases changing rapidly, the system will once more operate in the acquisition mode for a while, until the tracking mode has been achieved again, etc. If the  $x(t)$  waveform is constant or nearly constant the delta sequence  $y_d(n)$  will become periodic with period  $2T_s$ . This is called *Idling Pattern*.

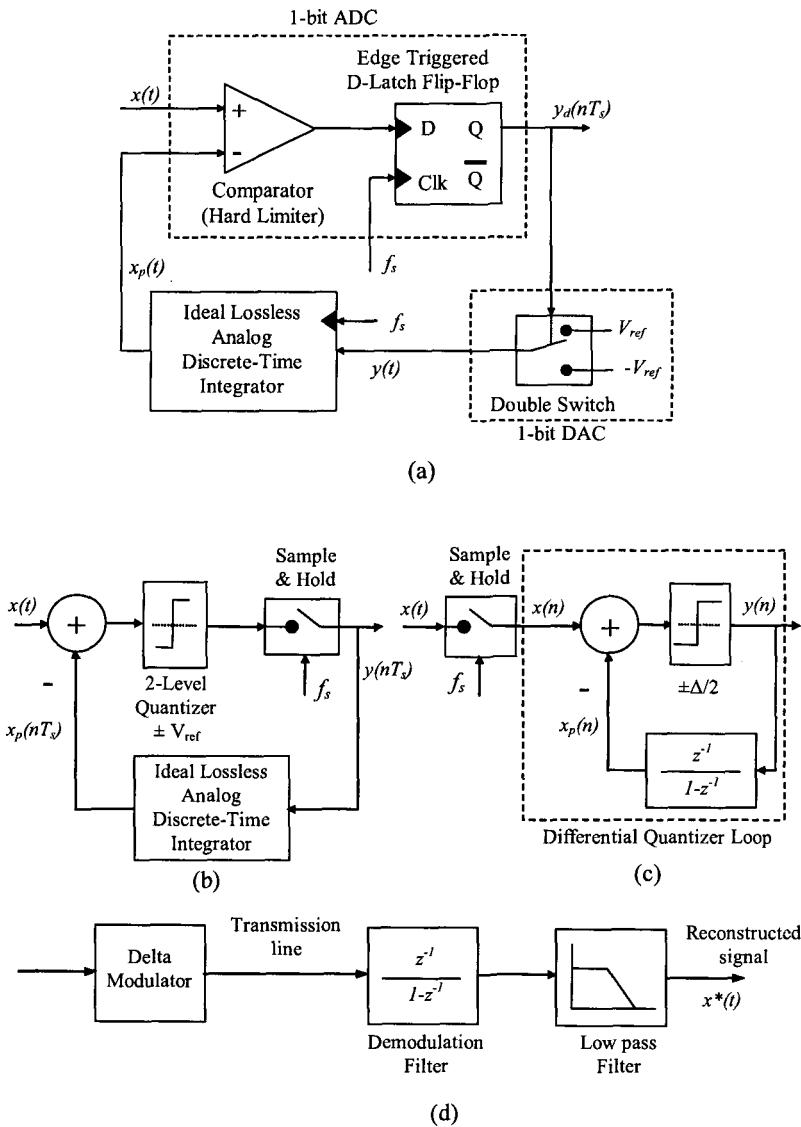


Fig. 2.15 a. Implementation of a Delta Modulator. b. and c. Equivalent block diagrams of the Linear Delta Modulator. d. Transmission and demodulation process.

The condition for no slope overloading is

$$\frac{dx(t)}{dt} \leq \frac{\Delta}{2T_s} \quad (2.16)$$

If  $x(t)$  is a sinusoid, i.e.

$$x(t) = V_o \sin(2\pi f t) \quad (2.17)$$

we get

$$\frac{dx(t)}{dt} = 2\pi f V_o \cos(2\pi f t) \leq 2\pi f V_o \quad (2.18)$$

Combining equations (2.16) and (2.18) gives for  $V_o$

$$V_o \leq \frac{\Delta f_s}{4\pi f} \quad (2.19)$$

Thus the maximum value of input signal amplitude  $V_o$  depends on its frequency and becomes minimum when this frequency is  $f_b$ . For this value we have

$$\frac{V_o}{\Delta/2} \leq \frac{f_s}{2\pi f_b} = \frac{R}{\pi} \quad (2.20)$$

According to Fig. 2.15.b the reconstruction of the signal is easily accomplished by means of a demodulation filter. However, this demodulation filter must be exactly the same as the one in the loop of the modulator (matched units).

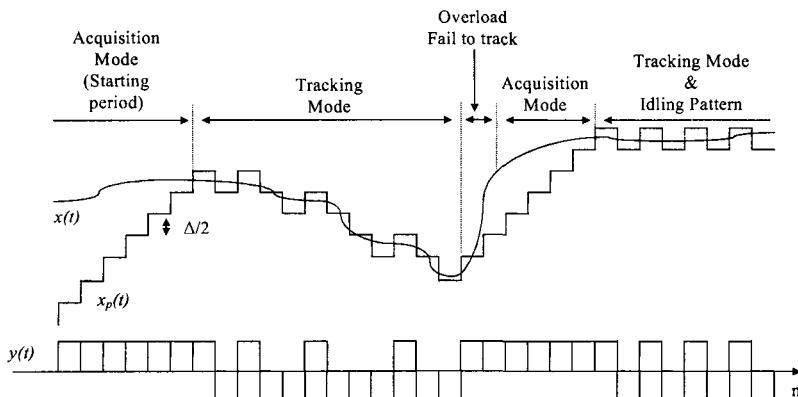


Fig. 2.16 Various signals in the Linear Delta Modulator circuit.

## 2.8 Performance of the Delta Modulator

The performance of the Delta Modulator will be examined in this section employing the methodology used in the analysis of oversampling converters. In this approach the SQNR is used as a quality criterion and is evaluated first. For this purpose the linear model for the quantizer is employed. Thus the delta modulator of Fig. 2.15 is drawn using the linear model, as shown in Fig. 2.17.

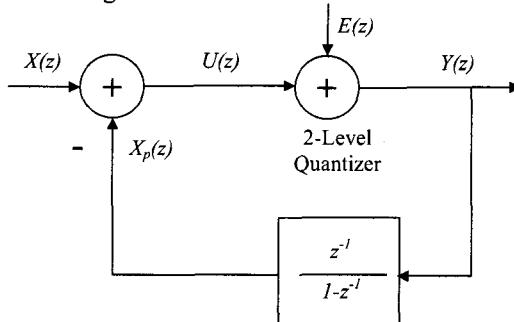


Fig. 2.17 Mathematical model for analyzing the Linear Delta Modulator, employing the linear model for the quantizer.

Analysis of the above figure using the z-transform gives

$$Y(z) = X(z) - \frac{z^{-1}}{1 - z^{-1}} Y(z) + E(z) \quad (2.21)$$

Solving for  $Y(z)$ , the DM can be described by means of two transfer functions, namely the signal transfer function (STF) and the noise transfer function (NTF). Accordingly, the output  $Y(z)$  is expressed as follows:

$$Y(z) = STF(z)X(z) + NTF(z)E(z) \quad (2.22)$$

where

$$\begin{aligned} STF(z) &= 1 - z^{-1} \\ NTF(z) &= 1 - z^{-1} \end{aligned} \quad (2.23)$$

At the receiver the DM sequence is demodulated by the demodulation filter  $1/(1-z^{-1})$  and the original signal  $x(n)$  plus the quantization noise are obtained. The low-pass filter rejects the quantization noise lying out of the signal band improving the SQNR. Thus the SQNR is given by

$$SQNR = \frac{P_s}{P_{e,in}} \quad (2.24)$$

In order to evaluate the quantities  $P_s$  and  $P_{e,in}$ , we consider first that the Delta Modulator is not overloaded. In this case (2.19) is valid and thus

$$P_s \leq \frac{\Delta^2}{32\pi^2} \left( \frac{f_s}{f} \right)^2 \quad (2.25)$$

For the  $P_{e,in}$  we have

$$P_{e,in} = \frac{1}{f_s} \int_{-f_b}^{f_b} \frac{\Delta^2}{12} df = \frac{\Delta^2}{12} \frac{1}{R} \quad (2.26)$$

Consequently, the maximum value for the SQNR without overloading the modulator will be

$$SQNR_{\max} = \frac{3R}{8\pi^2} \left( \frac{f_s}{f} \right)^2 \quad (2.27)$$

which depends on the frequency of the signal and takes its minimum value for  $f=f_b$ .

## 2.9 The Exponential DM

The exponential DM is based on the same principles as its linear counterpart. However, it presents the following advantages:

- a. Its circuitry is simpler than that of the linear DM.
- b. The demodulation filter is not necessary and consequently there is no need for matching the two integrators.
- c. Only the low-pass filter is necessary for the reconstruction of the signal.

Nevertheless, its analysis is more complicated than in the case of Linear DM and the functions  $STF(z)$  and  $NTF(z)$  are different from those given in (2.23).

The implementation of exponential DM is depicted in Fig. 2.18. The only difference from the circuit implementing the Linear DM is the simple RC lossy integrator which replaces the ideal Lossless Discrete Time Analog Integrator. This ideality is very difficult to be implemented

and thus the exponential DM was mostly used instead of its linear counterpart.

The input analog signal  $x(t)$  is compared with the reconstructed waveform  $x_p(t)$ , and the output of the comparator is sampled by a D Flip-Flop. The output sequence  $y_d(n)$  consists of positive and negative pulses such that the waveform  $x_p(t)$  continuously tracks the input  $x(t)$ . The lossy integrator is implemented by a simple RC circuit whose cutoff frequency is

$$f_{cutoff} = 1/2\pi R_1 C_1 \quad (2.28)$$

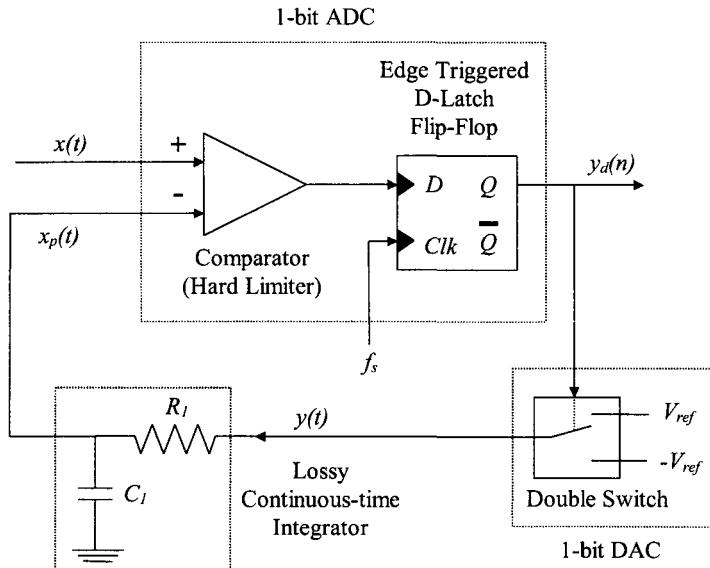


Fig. 2.18. Implementation of the exponential DM.

If this frequency is higher than  $f_b$ , i.e. if

$$\frac{1}{2\pi R_1 C_1} \geq f_b \quad (2.29.a)$$

or equivalently  $R_1 C_1 \leq \frac{1}{2\pi f_b}$  (2.29.b)

we can assume that the transfer function of the integrator will behave nearly as all-pass in the signal frequency band. Furthermore, overloading is to be avoided. For this purpose it has been proved [9] that the value of the time constant  $R_1 C_1$  must be

$$R_1 C_1 \geq \frac{1}{2\pi f_b} \quad (2.30)$$

Since, (2.29.b) and (2.30) should be valid simultaneously the time constant should be

$$R_1 C_1 = \frac{1}{2\pi f_b} \quad (2.31)$$

## 2.10 The Concept of Noise Shaping

The advantages of oversampling were analyzed to some depth in previous sections. At the same time an introduction to the concept of prediction was given by means of an explanation of the operation of Delta Modulator. A further improvement in the SQNR can be achieved by pushing also most of the in-band noise, left after the oversampling, outside the signal frequency band as shown in Fig. 2.19. This is attainable if the  $STF(z)$  is all-pass whereas, and most important, the  $NTF(z)$  is high-pass. This technique is called noise shaping and can be easily and efficiently implemented by modifying the DM system. The idea is to try to encode the integral of the input signal rather than the input signal directly. Clearly, integration (and generally linear filtering) being a linear function does not affect the system function whether it is placed at the end of the system or at the beginning. This means that the demodulation integrator (or filter) can be placed at the input of the DM as well, as it is shown in Fig. 2.20a. Furthermore, for the same reason, the two integrators in this Figure can be replaced by one placed inside the DM loop as shown in Fig. 2.20b. The significant modification of the DM system is that matching the two integrators, the analog and the digital, is not required any more, since the same integrator performs now the function of both. The new DM is called Delta-Sigma ( $\Delta\Sigma$ ) Modulator and achieves

quantization noise shaping i.e. it pushes the quantization noise outside the signal band as it is shown in Fig. 2.19. The letter Sigma ( $\Sigma$ ) is signifying the fact that the input signal is integrated first before entering the DM. If the integrator is substituted by a higher order linear filter, the Delta-Sigma Modulator will be of the same order with the filter. The analysis and the characteristic properties of the 1<sup>st</sup> and higher order  $\Delta\Sigma$  Modulators will be studied in the next chapter.

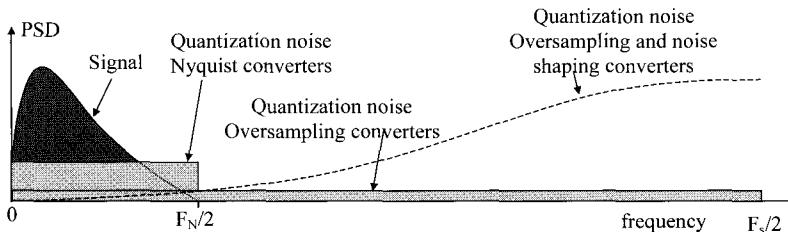


Fig. 2.19 Spectrum at the output of a noise shaping quantizer loop compared to those obtained from Nyquist and Oversampling converters.

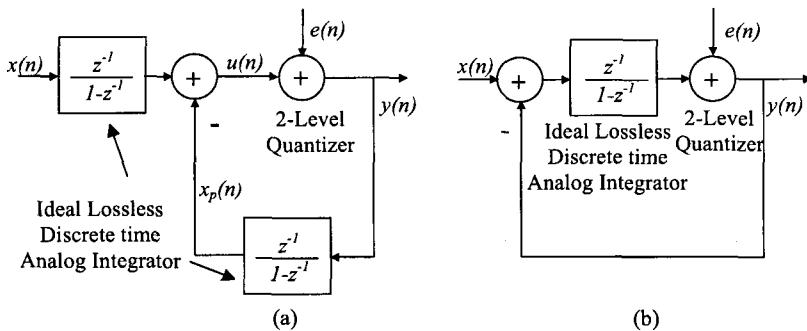


Fig. 2.20 Successive development of  $\Delta\Sigma$  modulator.

## 2.11 Summary

In this introductory chapter the fundamental concepts of Analog-to-Digital conversion were presented. The sampling and quantization stages in an ADC were examined in depth since they significantly affect the quality of the obtained signal. After that, the oversampling principle was

explained while its advantages and disadvantages were presented. **Oversampling** is a technique applied to get rid of a large amount of quantization noise by spreading much of its power out of the signal band.

The delta modulator, which is the simplest Analog-to-Digital conversion circuit employing oversampling, is analyzed. This circuit is the predecessor of  $\Delta\Sigma$  modulators, which in addition employ noise shaping in order to separate almost all of the quantization noise from the signal of interest. In the next chapter a variety of  $\Delta\Sigma$  modulators are presented and their operation is analyzed.

## Problems

2.1 An analog signal  $x(t)$  is given in the time domain as follows:

$$x(t) = 2\cos(800\pi t) \text{ Volts}$$

- a. Determine the lower sampling rate which is required to adequately sample  $x(t)$ .
  - b. If the sampling rate used is 600 Hz, determine the frequency of the reconstructed analog signal.
  - c. Determine the frequency of the reconstructed analog signal if the sampling rate is 300 Hz.
- 2.2 A triangular signal with peak-to-peak voltage variation of 2 Volts, is digitized by an ADC with 2 Volts of maximum input dynamic range. If twelve bits are available for sample representation, evaluate the SQNR achieved with the specific ADC and input signal.
- 2.3 A six-level uniform quantizer has the following quantization levels: {-1, -0.6, -0.2, 0.2, 0.6, 1}. Evaluate the Signal-to-Quantization-Noise Ratio when the following signals are quantized:
- i. dc signal of 0.85 Volts
  - ii. dc signal of 0.195 Volts
  - iii. triangular waveform with amplitude 0.01 Volts and dc offset 0.6 Volts
  - iv. triangular waveform with amplitude 0.8 Volts
- Comment on the way that the SQNR is affected by the signal power. Is it an increasing function of this power?

- 2.4 Suppose that the spectrum of a band-pass signal lies between 6.8 and 8.8 kHz. Which are the appropriate sampling frequencies? (use Fig. 2.12 and  $f_H=4.4B$ ).
- 2.5 A band-pass filter is used for reconstructing a sampled band-pass signal. Using Fig. 2.11b, show that the appropriate sampling frequencies for the band-pass filter requirements to be relaxed are given as follows:
- $$f_s = \frac{2(f_L + f_H)}{2k - 1} \text{ where } k = 1, 2, \dots, k_{\max} \text{ with } k_{\max} = \left\lfloor \frac{f_L + f_H}{2B} + 0.5 \right\rfloor$$
- 2.6 Which is the improvement in the SQNR when a conventional Nyquist converter is used in oversampling mode with  $OSR=10$ ?
- 2.7 Prove eq. (2.30). (Hint: Consider a sinusoidal input equal to  $V_{\text{ref}}$ . The derivative  $dx(t)/dt$  cannot be larger than the slope of  $x_p(t)$  at the output of the RC circuit in Fig. 2.18).
- 2.8 Evaluate the time constant  $RC$  for the integrator in an exponential Delta Modulator when the signal band is [0, 1000 Hz].

## References

- [1] R. Frank, *Understanding Smart Sensors*, Artech House, Boston, 1996.
- [2] R. Pallas-Areny and J.G. Webster, *Sensors and Signal Conditioning*, John Wiley & Sons, New York 1991.
- [3] P.M. Grant, 'Signal Processing Hardware and Software', *IEEE Signal Processing Magazine*, pp. 86-88, January 1996.
- [4] E.C. Ifeachor and B.W. Jervis, *Digital Signal Processing: A practical Approach*, Addison Wesley, Wokingham, England, 1996.
- [5] R.S. Soin, F. Maloberti and I. Franca, *Analogue - Digital ASICs, Circuit Techniques, Design Tools and Applications*, IEE Circuit and Systems series 3, Peter Peregrinus Ltd, London 1991.
- [6] R. Van de Plashe, *Integrated Analog-to-Digital and Digital-to-Analog Converters*, Kluwer Academic Publishers, London 1994.
- [7] *Digital/Analog and Analog/Digital Conversion Handbook*, Motorola Semiconductors, 1980.
- [8] P.M. Aziz, H.V. Sorensen, and Jan Van Der Spiegel, 'An Overview of Sigma-Delta Converters', *IEEE Signal Processing Magazine*, pp 61-83, January 1996.
- [9] R. Steele, *Delta Modulation Systems*, Pentech Press, London, 1975.
- [10] J.G. Proakis and D.G. Manolakis, *Digital Signal Processing: Principles, Algorithms and Applications*, Prentice-Hall, 1996.
- [11] A.D. Poullarikas, *The Transforms and Applications Handbook*, CRC Press, IEEE Press, 1996.
- [12] T. Schilling, *Principles of Communications Principles*, McGraw-Hill, 1990.
- [13] S.R. Norsworthy, R. Schreier and G.C. Temes, *Delta-Sigma Data Converters, Theory, Design and Simulation*, New York, IEEE Press, 1997.

## **Chapter 3**

# **$\Delta\Sigma$ Modulators – Architectures**

### **3.1 Introduction**

Oversampling and noise shaping were introduced and explained in the previous chapter. The former spreads the quantization noise spectrum uniformly over the whole frequency band, thus leaving a small portion of it inside the signal band. The latter, on the other hand, reduces the quantization noise power inside the signal frequency band even further, by pushing most of the in-band noise outside the signal band. Both of these processes are employed in the various types of the  $\Delta\Sigma$  modulator, which are described in this chapter.

In Sec. 3.2 of this chapter the first-order  $\Delta\Sigma$  modulator is analyzed and in 3.3 it is compared to the Delta modulators. Then, in Sec. 3.4,  $\Delta\Sigma$  modulation is extended to the second-order and, in Sec. 3.5, to the higher-order single-stage  $\Delta\Sigma$  modulators. The stability problems of the latter are demonstrated in Sec. 3.6. The higher-order multi-stage  $\Delta\Sigma$  modulators are then considered in Sec. 3.7. Next, the use of multi-bit quantizers in  $\Delta\Sigma$  modulators is considered in Sec. 3.8, hybrid modulators in Sec. 3.9 and adaptive ones in Sec. 3.10. Finally, in Sec. 3.11, the concept of  $\Delta\Sigma$  modulation is extended to band-pass signals.

### 3.2 First-Order $\Delta\Sigma$ Modulators

The first-order  $\Delta\Sigma$  Modulator was briefly introduced in Sec. 2.10. This modulator employs oversampling to spread the quantization noise over the  $[0, f_s/2]$  frequency band, as well as noise shaping in order to push most of the in-band noise out of this band to higher frequencies.

The block diagram of  $\Delta\Sigma$  modulator is shown in Fig. 3.1 (a), while for the sake of analysis its linear model is shown in Fig. 3.1 (b). The sampler and the encoder are omitted as they have no impact on the analysis at this level, while the quantizer is replaced by its linear model.

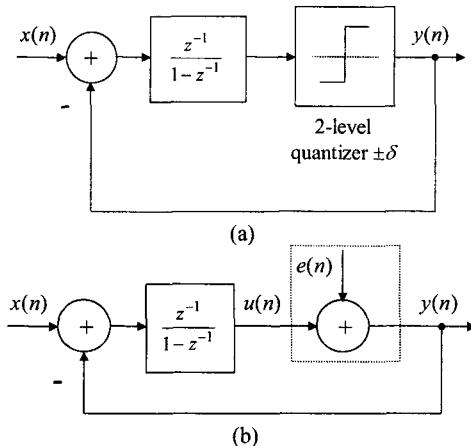


Fig. 3.1 Block diagram of a first-order  $\Delta\Sigma$  modulator (a) and its linear model (b).

By straightforward analysis of the linear system in Fig 3.1 (b), we can easily obtain the following:

$$Y(z) = z^{-1}X(z) + (1 - z^{-1})E(z) \quad (3.1)$$

From (3.1) the STF and the NTF are

$$STF(z) = z^{-1} \quad (3.2)$$

$$NTF(z) = 1 - z^{-1} \quad (3.3)$$

Clearly, the  $STF(z)$  leaves the signal unaltered, just delayed by the period of a single bit, whereas the  $NTF(z)$  high-passes the quantization

error, i.e. it shapes it by suppressing it at low frequencies. Consequently, if oversampling by  $R$  is employed, the quantization noise power inside the signal band  $[0, f_s/2R] \equiv [0, f_b]$  will be lower than its value when no shaping is applied. This is shown below by means of example 3.1.

Under the assumption that the introduced quantization noise is white with power  $P_e$ , the total noise power will be

$$P_{e,tot} = \frac{P_e}{f_s} \int_{-f_s/2}^{f_s/2} |NTF(e^{j2\pi f/f_s})|^2 df = \frac{P_e}{2\pi} \int_{-\pi}^{\pi} |NTF(e^{j\theta})|^2 d\theta \quad (3.4)$$

The out-of band noise can be attenuated using a sharp digital low-pass post filter, while the output samples can be down-sampled to the Nyquist rate. Both of these functions can be performed by the decimator [1-3], which is discussed in Chapter 8. After decimation, the remaining noise that corrupts the signal is the in-band portion of the quantization noise power, which is

$$P_{e,in} = \frac{P_e}{f_s} \int_{-f_b}^{f_b} |NTF(e^{j2\pi f/f_s})|^2 df = \frac{P_e}{2\pi} \int_{-\pi/R}^{\pi/R} |NTF(e^{j\theta})|^2 d\theta \quad (3.5)$$

Assuming sinusoidal input of amplitude  $A$  and frequency  $f$ , the signal-to-noise ratio<sup>1</sup>  $SNR$  is given by

$$SNR = \frac{A^2}{2P_{e,in}} \cdot |STF(f)|^2 \quad (3.6)$$

The SNR, as given by (3.6), increases indefinitely. Of course this is not the case. The linearization of the quantizer is valid as long as its input does not overload it. A quantizer having  $N$  levels around zero and quantization step  $2\delta$  is not overloaded if its input  $u$  is such that the quantization error does not exceed  $\delta$ , i.e.

$$|u| < N\delta \quad (3.7)$$

Unfortunately, due to the feedback in the  $\Delta\Sigma$  modulator, if its order is larger than 1, it is impossible to analytically relate the input of the

<sup>1</sup> Note that the noise mentioned in the signal-to-noise ratios in this chapter is the quantization noise, but for the sake of brevity, the symbol  $SNR$  is used instead of the  $SQNR$  to denote them.

quantizer to the input of the  $\Delta\Sigma$  modulator. Hence (3.7) cannot be used to investigate the stability of  $\Delta\Sigma$  modulators.

**Example 3.1** Assume a first-order  $\Delta\Sigma$  modulator that modulates a signal oversampled  $R$  times. Find the reduction of the in-band quantization noise  $P_{e,in}$  as a function of  $R$ .

**Solution** Using (3.4), the total quantization noise power becomes

$$P_{e,tot} = \frac{P_e}{\pi} \int_0^\pi |NTF(e^{j\theta})|^2 d\theta = \frac{P_e}{\pi} \int_0^\pi |1 - e^{-j\theta}|^2 d\theta$$

It is

$$|1 - e^{-j\theta}|^2 = |1 - (\cos\theta - j\sin\theta)|^2 = 2 - 2\cos\theta$$

Hence

$$P_{e,tot} = 2P_e \quad (3.8)$$

Using (3.5), the in-band fraction of the quantization noise power becomes

$$P_{e,in} = \frac{P_e}{\pi} \int_0^{\pi/R} |NTF(e^{j\theta})|^2 d\theta = \frac{2P_e}{\pi} \left( \frac{\pi}{R} - \sin \frac{\pi}{R} \right)$$

For large values of  $R \gg \pi$ , the sine function can be approximated using a Taylor series expansion, hence keeping up to the third-order term yields

$$P_{e,in} \approx \frac{2P_e}{\pi} \left[ \frac{\pi}{R} - \left( \frac{\pi}{R} - \frac{\pi^3}{6R^3} \right) \right] = P_e \frac{\pi^2}{3R^3} \quad (3.9)$$

Thus the in-band shaped noise is attenuated  $3R^3/\pi^2$  times, compared to the non-shaped noise.

So for the first-order  $\Delta\Sigma$  modulator, the noise is attenuated by the third power of the oversampling ratio. Also note from (3.8) that even though the effect of the NTF on the total noise power is to double it, only a very small fraction of the noise remains in-band.

### 3.2.1 Input and output waveforms

The output waveforms of the first-order  $\Delta\Sigma$  modulator of Fig. 3.1 (a) are presented in Fig. 3.2, for two types of input, namely a DC and a sine wave. It can be seen that the output consists of a sequence of pulses of height  $\pm\delta$ . The relative density of the positive or negative values depends on the amplitude of the input at that region. Hence for small amplitudes (cases (a) and (c) in Fig. 3.2) the positive and negative  $\Delta\Sigma$  sequence values are evenly distributed, while for large amplitudes (cases (b) and (d) in Fig. 3.2) the positive  $\Delta\Sigma$  sequence values are denser than the negative  $\Delta\Sigma$  sequence values at large positive input values and vice versa. Clearly the form of the output varies significantly as the input is changed.

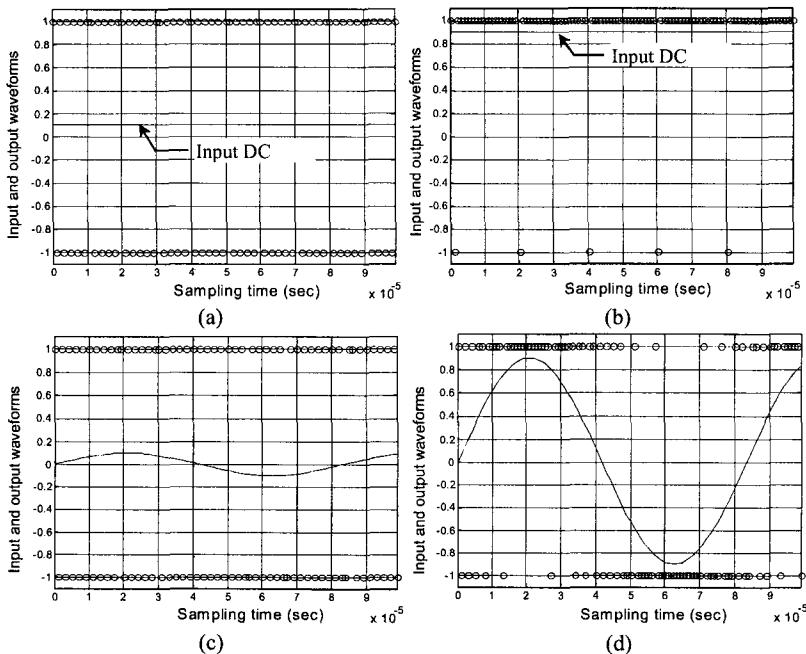


Fig. 3.2 Input waveforms (continuous lines) and output sequences (circles) of the first-order  $\Delta\Sigma$  modulator for DC input of level (a)  $0.1\delta$  and (b)  $0.9\delta$ . Same for sine wave input of frequency 12 kHz and amplitude (c)  $0.1\delta$  and (d)  $0.9\delta$ . The sampling frequency is 1 MHz and  $\delta=1$ .

### 3.2.2 SNR and PSD diagrams

Although the linear analysis performed thus far yields an insight to the performance of the first-order  $\Delta\Sigma$  modulator, in order to correctly evaluate its performance, the first-order  $\Delta\Sigma$  modulator is simulated as a non-linear system. Using a sinusoidal input and sweeping its power, the SNR of the first-order  $\Delta\Sigma$  modulator is obtained. This is depicted in Fig. 3.3, from which some useful conclusions can be drawn. The maximum SNR is not obtained for the maximum input signal, i.e. the full scale of the quantizer, due to its overloading. Also, the dynamic range of the first-order  $\Delta\Sigma$  modulator, defined as the range of input power for which the SNR in dBs remains positive, is limited. This limits the application of the first-order  $\Delta\Sigma$  modulator as an ADC to only low dynamic range systems.

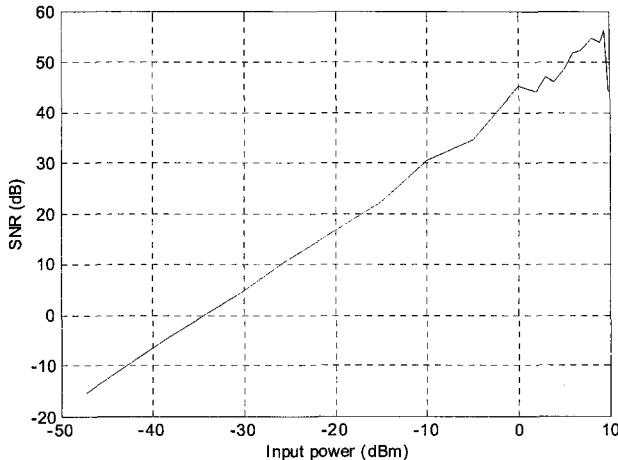


Fig. 3.3 SNR as a function of input power for the first-order  $\Delta\Sigma$  modulator and sinusoidal input. The oversampling ratio is 64. The maximum SNR of 54.7 dB is obtained for sinusoidal input of 8 dBm. The dynamic range of the modulator is 44.2 dB.

The power spectral density of the output signal for small variations of the frequency is depicted in Fig. 3.4. Evidently the spectrum can exhibit discrete tones. These tones can be at multiples of the input frequency (case (a) of Fig. 3.4), very dense at all frequencies (case (c) of Fig. 3.4), or they may not exist at all (case (b) of Fig. 3.4), depending on slight

changes in the frequency of the input signal. This is very undesirable in many applications, like audio systems, since although the total in-band noise is low, the tones can be audible. Due to its limited dynamic range and its noise tones, the applications of the first-order  $\Delta\Sigma$  modulator are limited.

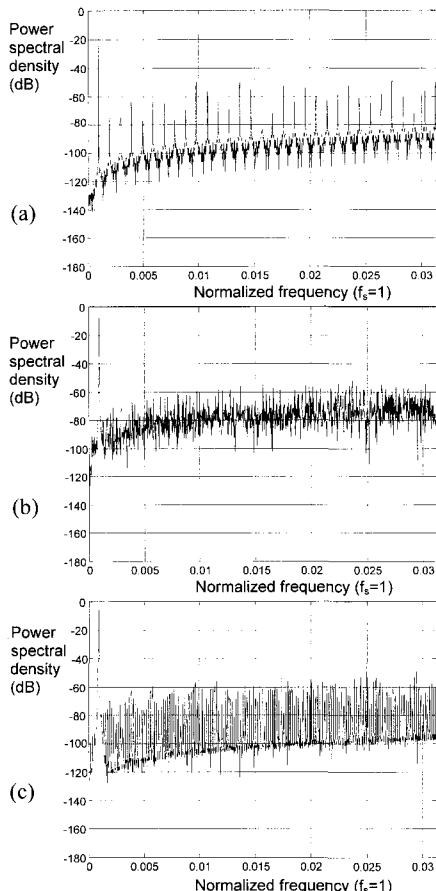


Fig. 3.4 Power spectral density of the output signal of the first-order  $\Delta\Sigma$  modulator for low frequencies and for an 8 dBm sinusoidal input. The input frequency is:  
a.  $9.7656 \cdot 10^{-4}$ , b.  $9.7534 \cdot 10^{-4}$  and c.  $9.6451 \cdot 10^{-4}$  of the sampling frequency.

### 3.3 Comparison of Delta and $\Delta\Sigma$ Modulators

The performance comparison of the DM, with that of the  $\Delta\Sigma$  modulator is done by determining the maximum value of SNR,  $SNR_{\max}$ , for each case. If  $P_s$  is the signal power and  $P_{e,in}$  the in-band portion of the quantization noise power, then

$$SNR \equiv \frac{P_s}{P_{e,in}} \quad (3.10)$$

#### 3.3.1 Linear DM

In order to determine the signal power  $P_s$  we assume that the DM is not slope overloaded. According to (2.27) it is

$$SNR_{\max} = \frac{3R}{8\pi^2} \left( \frac{f_s}{f} \right)^2 \quad (3.11)$$

Clearly  $SNR_{\max}$  is frequency dependent and obtains its minimum value at  $f = f_b$ . Then

$$SNR_{\max} = \frac{3R^3}{2\pi^2} \quad (3.12)$$

#### 3.3.2 Exponential DM

For the exponential DM, assuming no amplitude overload, it is

$$P_s \leq \frac{V_{ref}^2}{2} \quad (3.13)$$

while  $P_{e,in}$  is given by (2.26). Then

$$SNR_{\max} = \frac{\max\{P_s\}}{P_{e,in}} = 6R \frac{V_{ref}^2}{\Delta^2} \quad (3.14)$$

Using (2.20)

$$\Delta = \frac{2\pi}{R} V_{ref}$$

Thus

$$SNR_{\max} = \frac{3R^3}{2\pi^2}$$

### 3.3.3 $\Delta\Sigma$ Modulator

For the first-order  $\Delta\Sigma$  modulator, the in-band portion of the quantization noise power is given by (3.9), which using (3.8) becomes

$$P_{e,in} = P_e \frac{\pi^2}{3R^3} = \frac{\Delta^2 \pi^2}{12 \cdot 3R^3} = \Delta^2 \frac{\pi^2}{36R^3} \quad (3.15)$$

Now there is no slope overload restriction for  $P_s$ , but the input signal should satisfy  $|x(t)| \leq V_{ref}$  or  $|x(t)| \leq \Delta/2$ . Hence

$$SNR_{max} \equiv \frac{P_s}{P_{e,in}} = \frac{(\Delta/2)^2/2}{\Delta^2 \pi^2 / 36R^3} = \frac{9}{2\pi^2} R^3 \quad (3.16)$$

We observe that the SNR of the exponential DM and the first-order  $\Delta\Sigma$  modulator both depend on the third power of  $R$ , i.e. there is a 9 dB improvement per doubling of the  $R$ . Also the SNR of the first-order  $\Delta\Sigma$  modulator is improved by a factor of 3 (4.8 dB) compared to that of the exponential DM. Thus the first-order  $\Delta\Sigma$  Modulator compared to the exponential Delta modulators is superior at least from the maximum SNR point of view. Regarding the linear DM, the SNR depends on the frequency of the input signal, with its minimum value given by (3.12).

## 3.4 Second-Order $\Delta\Sigma$ Modulators

Using a second-order filter in place of the integrator in Fig. 3.1 with transfer function

$$L(z) = \frac{2z^{-1} - z^{-2}}{1 - 2z^{-1} + z^{-2}} \quad (3.17)$$

the second-order  $\Delta\Sigma$  modulator is obtained. Its NTF is a second-order differentiator

$$NTF(z) = (1 - z^{-1})^2 \quad (3.18)$$

while its STF remains just a single sample delay. It can be shown (see exercise 3.1), that although the total quantization noise increases 6 times, the relative attenuation of the in-band fraction as a function of the oversampling ratio is

$$\frac{P_{e,in}}{P_e} \approx \frac{\pi^4}{15R^5} \quad (3.19)$$

i.e. it is attenuated by the fifth power of the oversampling ratio.

The SNR of the second-order  $\Delta\Sigma$  modulator is shown in Fig. 3.5 (a). Both the maximum SNR and the dynamic range are enhanced.

Also, the power spectral density of the output signal for low frequencies is depicted in Fig. 3.5 (b). The noise is now randomized by the higher-order loop and does not have any discrete tones.

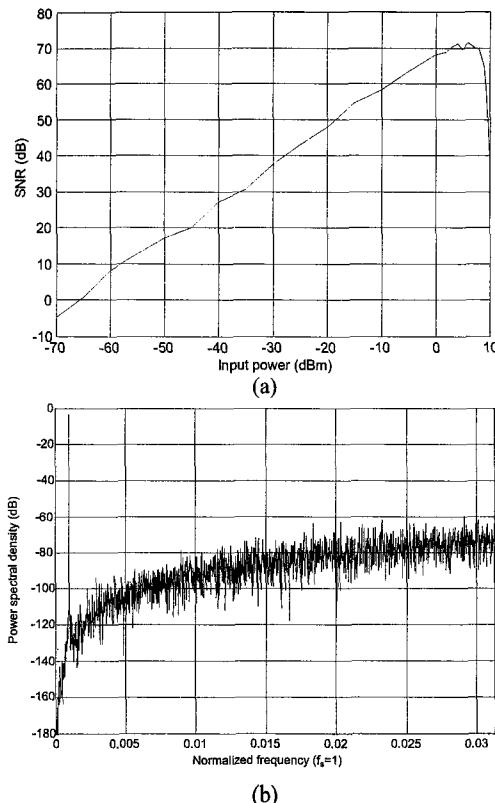


Fig. 3.5 Second-order  $\Delta\Sigma$  modulator with sinusoidal input and OSR 64. a. SNR as a function of input power (dBm). The maximum SNR of 71.5 dB is obtained for 6 dBm input. The dynamic range of the modulator is 76.3 dB. b. Power spectral density of the output signal for low frequencies and for 6 dBm input. The noise no longer forms tones.

### 3.5 Higher-Order ΔΣ Modulators

The advantages of extending the order of the  $\Delta\Sigma$  modulator from first to second have been demonstrated in the previous section. These advantages can be further exploited by increasing the order of the  $\Delta\Sigma$  modulator. Yet, higher-order  $\Delta\Sigma$  modulators are not without problems. Apart from overloading that has been encountered in the lower-order  $\Delta\Sigma$  modulators, their higher-order counterparts suffer also from instability.

There exist a variety of architectures for the implementation of higher-order  $\Delta\Sigma$  modulators [4-9]. The purpose of this and the following sections is to introduce them, point out their relative advantages and disadvantages and explore their suitability as either ADCs or DACs.

The most common approach in the design of high-order  $\Delta\Sigma$  modulators to be used for A/D conversion is the single-stage architectures. The block diagram of a single-stage  $\Delta\Sigma$  modulator to be used in A/D conversion is depicted in Fig. 3.6, where the ADC – DAC pair substitutes the quantizer shown in other block diagrams, e.g. in Fig. 3.1a.

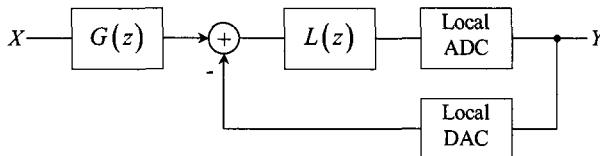


Fig. 3.6 Block diagram of higher-order  $\Delta\Sigma$  modulator.

These modulators comprise a loop filter  $L(z)$ , a local ADC and a local DAC in a single loop. Both of these converters are high-speed to allow for oversampling and low-bit (most usually single-bit). The addition of a pre-filter  $G(z)$  before the summer of the loop in Fig. 3.6 facilitates the implementation of any STF. These filters are usually discrete-time, but they can also be continuous-time. However, when the modulator is used for D/A conversion, then obviously the input is digital. When it is used for A/D conversion, the input can be sampled before the modulator, enabling the use of discrete-time filters. The loop filter is implemented by suitably cascading integrators. There exist various structures for the

loop filter of a single-stage  $\Delta\Sigma$  modulator [4,6,10-14]. There is a trade off between the complexity of each structure and the freedom in choosing the position of the zeros of the NTF and the shape of the STF. The structures are depicted in Fig. 3.7 for third-order  $\Delta\Sigma$  modulators and are classified as follows:

- CIDF The Cascaded Integrators with Distributed Feedback is the simplest structure, shown in Fig. 3.7.a. All the integrators  $I_i$  are delaying integrators. The NTF zeros are fixed to unity, and the shape of the STF cannot be defined independently.
- CIDIDF The Cascaded Integrators with Distributed Input and Distributed Feedback structure is shown in Fig. 3.7.b. The phase of the NTF zeros can be set arbitrarily. The shape of the STF can be defined independently. If all the integrators  $I_i$  are delaying, then the real part of the NTF zeros is fixed to unity. If the even-order integrators  $I_{2i}$  are non-delaying, then the restriction on the real part does not apply.
- CIDIFF The Cascaded Integrators with Distributed Input and summed Feed-Forward structure is shown in Fig. 3.7.c. It employs single output feedback. The phase of the NTF zeros can be set arbitrarily. The shape of the STF can be defined independently. If all the integrators  $I_i$  are delaying, then the real part of the NTF zeros is fixed to unity. If the even-order integrators  $I_{2i}$  are non-delaying, then the restriction on the real part does not apply.

In this classification the loop filters were assumed to be discrete-time filters. However the same classification applies when continuous-time filters are employed. In this case continuous-time integrators are used instead of discrete-time ones. For more details we refer the reader to Chapter 5, where the implementation of  $\Delta\Sigma$  modulators is presented.

Note that the  $c_i$  coefficients (Fig. 3.7) are used for scaling of the internal states, hence are set to unity unless a circuit implementation is considered. Also the  $g_i$  coefficients (Fig. 3.7.b and c) are only used for complex zeros; should the zeros be real, they are set to zero.

If delaying and non-delaying integrators are used, then the first and third integrators are non-delaying, whereas the second is delaying.

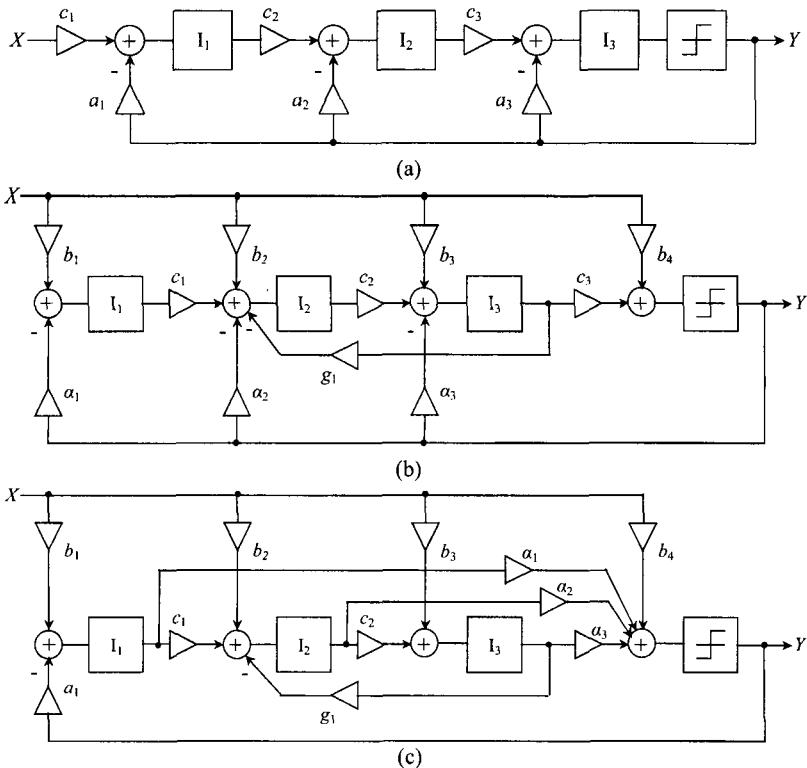
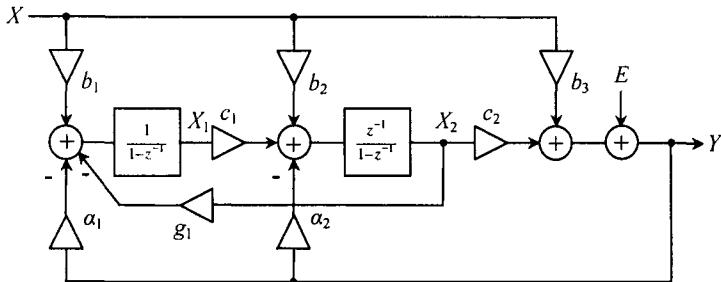


Fig. 3.7 Structures for the third-order single-stage  $\Delta\Sigma$  modulators: a. CIDIF, b. CIDIDF and c. CIDIFF.

**Example 3.2** Using the CIDIDF structure, realize the second-order  $\Delta\Sigma$  modulator with  $NTF = (1 - z^{-1})^2$ . Take advantage of the STF flexibility by setting  $STF(z) = 1$ .

**Solution** Based on Fig. 3.7 (b) and using a non-delaying integrator for  $I_1$ , the block diagram of the second-order CIDIDF  $\Delta\Sigma$  modulator is drawn in Fig. 3.8. Note at this point that since no specific circuit implementation that needs scaling of the internal states is considered,  $c_1 = c_2 = 1$ , and since the NTF zeros are unity,  $g_1 = 0$ . These simplifications are not used in order to derive the NTF and STF in the

Fig. 3.8 Block diagram of the CIDIDF realization of a second-order  $\Delta\Sigma$  modulator.

more flexible form allowed by the CIDIDF structure.

It is

$$X_1 = (b_1 X - a_1 Y - g_1 X_2) \frac{1}{1-z^{-1}} \quad (3.20)$$

$$X_2 = (b_2 X - a_2 Y + c_1 X_1) \frac{z^{-1}}{1-z^{-1}} \quad (3.21)$$

$$Y = c_2 X_2 + b_3 X + E \quad (3.22)$$

Substituting (3.20) and (3.21) in (3.22) yields

$$NTF \equiv \left. \frac{Y}{E} \right|_{X=0} = \frac{1 + (c_1 g_1 - 2) z^{-1} + z^{-2}}{1 + (c_1 g_1 + c_1 c_2 a_1 + c_2 a_2 - 2) z^{-1} + (1 - c_2 a_2) z^{-2}} \quad (3.23)$$

and

$$STF \equiv \left. \frac{Y}{X} \right|_{E=0} = \frac{b_3 + (c_1 c_2 b_1 + c_2 b_2 + c_1 g_1 b_3 - 2b_3) z^{-1} + (b_3 - c_2 b_2) z^{-2}}{1 + (c_1 g_1 + c_1 c_2 a_1 + c_2 a_2 - 2) z^{-1} + (1 - c_2 a_2) z^{-2}} \quad (3.24)$$

As already discussed, the coefficients  $c_1$  and  $c_2$  are for internal state scaling of the modulator, and for the sake of this example are set to unity. Then

$$NTF = (1 - z^{-1})^2 \Leftrightarrow \begin{cases} g_1 = 0 \\ a_1 = 1 \\ a_2 = 1 \end{cases} \quad (3.25)$$

Note that as expected,  $g_1 = 0$ . Also

$$STF = 1 \Leftrightarrow \begin{cases} b_1 = a_1 \\ b_2 = a_2 \\ b_3 = 1 \end{cases} \quad (3.26)$$

Finally, substituting (3.25) in (3.26) yields  $b_1 = b_2 = 1$ .

For the fabrication of the  $\Delta\Sigma$  modulator as an IC, the integrators in the structures of Fig. 3.7 are usually built using SC circuitry for two reasons. Firstly, the accuracy of the coefficients is determined by capacitor ratios that can be very accurate in MOS technology, and secondly because the value of the coefficients does not depend on the sampling rate which can then be easily changed [15]. The quantizer is single-bit to benefit from ease of implementation of the single-bit ADC and DAC and the linearity of the latter [2].

Unfortunately, single-bit single-stage  $\Delta\Sigma$  modulators are potentially unstable systems [2,13,16-18]. This will be discussed further in the next section. Thus the design of their loop filter is a non-trivial task [2,6,10,11,13,17-20,21], and will be the subject of Sec. 4.6. Due to the stability considerations, the SNR improvement as the modulator order increases is diminished for increasing orders [2,21]. The single-stage modulators are quite immune to coefficient errors; such tolerated errors have been reported to be larger than 5% [2,6]. On the other hand, the sensitivity of single-bit single-stage  $\Delta\Sigma$  modulators to analog circuit imperfections is considerably reduced compared to their multi-stage counterparts. Opamp gain as low as the oversampling ratio can easily be tolerated and can be further reduced by the use of gain-compensated SC integrators [2]. Thus single-stage  $\Delta\Sigma$  modulators are relatively insensitive to linear opamp errors. Considerable degradation occurs only if the opamp slew rate is very low due to the limited current supply of the opamp [2]. Limited opamp swing is also a problem, especially for the last integrator, which has very large signal swings. Internal scaling of the modulator limits these swings and alleviates the problem [2]. These circuit realization considerations are investigated in more detail in Chapter 6.

In Fig. 3.9, the SNR versus input power for a sinusoidal input obtained by simulation for certain  $\Delta\Sigma$  modulators of orders 3, 4 and 5 is

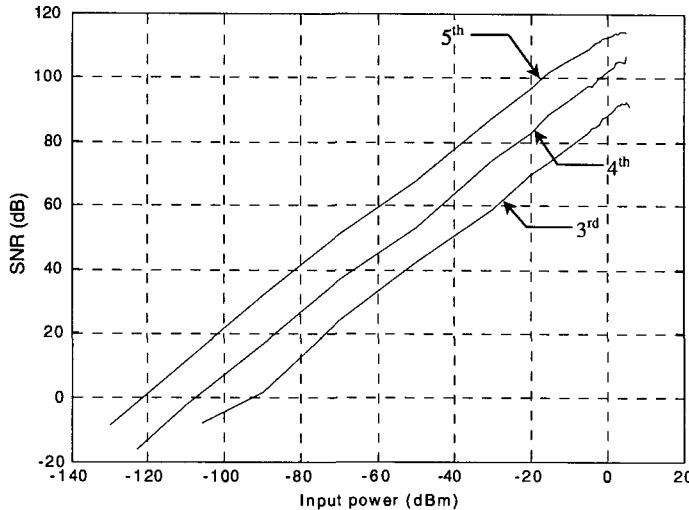


Fig. 3.9 SNR as a function of input power for the higher-order  $\Delta\Sigma$  modulators with sinusoidal input. The oversampling ratio is 64.

shown. These plots and other results given in this section should be considered as indicative of the effect of the NTF order on the SNR.

The corresponding power spectral density plots of their outputs for low frequencies are depicted in Fig. 3.10. Their maximum SNR and dynamic range are given in Table 3.1.

Table 3.1 Maximum SNR, the input power for which it is achieved and dynamic range of the  $\Delta\Sigma$  modulators of orders 1 to 5. The input is sinusoidal and the oversampling ratio is 64. The SNR increase as the order  $N$  of the  $\Delta\Sigma$  modulator is increased to  $N+1$  becomes smaller as  $N$  increases. The increase of the dynamic range also becomes smaller for increasing  $N$ , but is still substantial. The input power for which the maximum SNR is obtained is reduced, indicating faster overload of the quantizer.

Modulator order	Max SNR (dB)	Input power (dBm)	Dynamic range (dB)
1	54.7	8.0	44.2
2	71.5	6.0	76.3
3	92.5	5.0	98.8
4	107	5.0	113
5	114	4.0	126

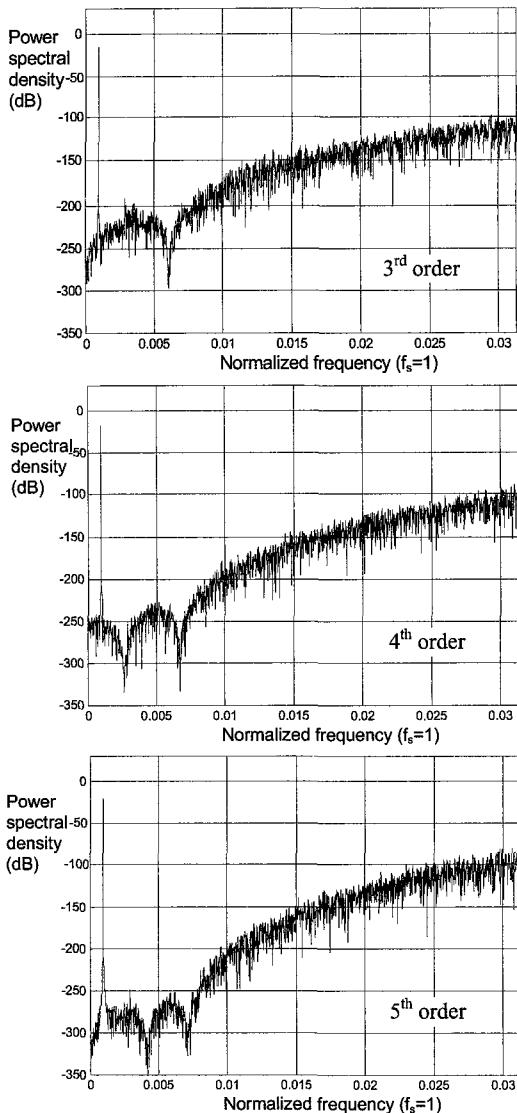


Fig. 3.10 Power spectral density of the output signal of the higher-order  $\Delta\Sigma$  modulators for low frequencies and for sinusoidal input of power that results to the highest SNR for each. The fades of the spectrum are due to the NTF zeros, which are not at DC, but are distributed in the signal frequency band.

The first and second-order modulators are also included in the table for comparison. Increasing the order of the modulators yields wider dynamic range. The improvement of the SNR gets smaller, as less frequency selective NTFs are necessary for the stability of the increasing order modulators.

### 3.6 Stability of Single-Stage $\Delta\Sigma$ Modulators

When the quantizer of a higher-order single-stage  $\Delta\Sigma$  modulator is overloaded, then the system can become unstable. Instability occurs as low-frequency high-amplitude oscillations are excited at the quantizer input. If the quantizer is single-bit, the result is long sequences of consecutive +1 or -1, i.e. the  $\Delta\Sigma$  sequence no longer tracks the input signal. This is demonstrated in Fig. 3.11.

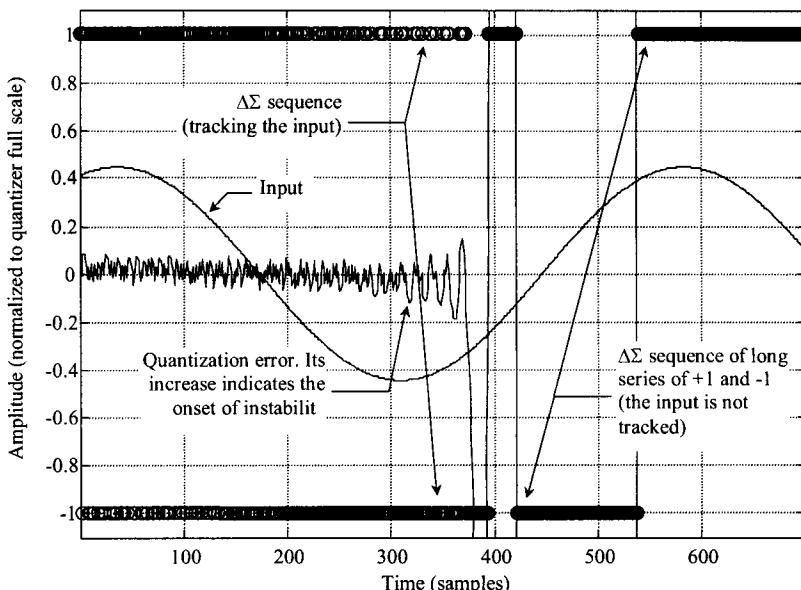


Fig. 3.11 Instability of a higher-order  $\Delta\Sigma$  modulator. Originally the  $\Delta\Sigma$  sequence tracks the input. Then the tracking is lost. The error receives very large values and the  $\Delta\Sigma$  sequence comprises long series of +1 or -1.

Even worse than the loss of tracking of the input, is that these limit cycles, once generated, are very difficult to cease, even if the input signal is removed. If no damping means are used, like clipping the internal states of the modulator or resetting them [13,16], then the limit cycles are impossible to overcome. Thus the modulator becomes unusable.

When these limit cycles are excited, the quantizer is heavily overloaded, thus the quantization noise power obtains extreme values. It is beneficial to determine the maximum input for which stable operation is guaranteed, but it depends on the type of input used and it sometimes takes extremely long input sequences for the limit cycles to be excited. Thus, if simulation is to be used, this must be extensive both in types of input and in length, in order to provide bounds with some confidence.

The stability of single-stage  $\Delta\Sigma$  modulators is further considered in Sec. 4.4, followed by some stability criteria in Sec. 4.5. It is very important to have some means of stabilizing the higher-order single-stage  $\Delta\Sigma$  modulators. Bounding the input is not enough, since some noise can lead to instability. The stabilization techniques are discussed in Sec. 6.10.

### 3.7 Multi-Stage $\Delta\Sigma$ Modulators

The multi-stage  $\Delta\Sigma$  modulators [2,22] consist of first- and/or second-order  $\Delta\Sigma$  modulator stages in cascade. The input to the first stage is the signal to be modulated. The consecutive stages modulate the quantization error of the previous ones. The outputs of the stages are suitably combined so that at the output of the multi-stage  $\Delta\Sigma$  modulator only the input signal and signal-independent noise shaped by  $(1 - z^{-1})^N$  are present. The combination of the stages in a multi-stage  $\Delta\Sigma$  modulator is better described by means of the following example:

**Example 3.3** The Triple First-Order Cascade (TFOC) [22] structure is depicted in Fig. 3.12. By linearizing the quantizers of the TFOC  $\Delta\Sigma$  modulator, verify that it behaves like a third-order  $\Delta\Sigma$  modulator.

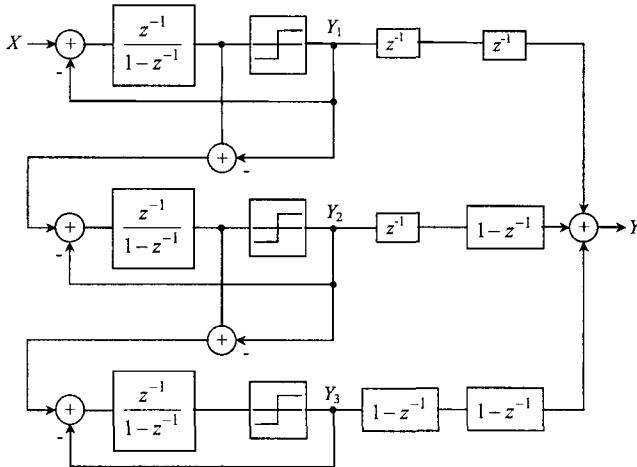


Fig. 3.12 TFOC  $\Delta\Sigma$  modulator. Three first-order  $\Delta\Sigma$  modulators are cascaded, so that the quantization error of the previous stage is the input to the next. The three outputs are differentiated and delayed accordingly, so that the combined output is third-order shaped.

**Solution** Assume that each quantizer in Fig. 3.12 introduces the noise  $E_i$ . Then, the input of the three first-order stages is  $X$ ,  $-E_1$  and  $-E_2$ . Accordingly the outputs  $Y_i$  of the three will be

$$\begin{aligned} Y_1 &= z^{-1} \cdot X + (1 - z^{-1}) \cdot E_1 \\ Y_2 &= z^{-1} \cdot (-E_1) + (1 - z^{-1}) \cdot E_2 \\ Y_3 &= z^{-1} \cdot (-E_2) + (1 - z^{-1}) \cdot E_3 \end{aligned} \quad (3.27)$$

The outputs of the three stages are combined using delays and differentiators as follows:

$$Y = z^{-2} \cdot Y_1 + z^{-1} \cdot (1 - z^{-1}) \cdot Y_2 + (1 - z^{-1})^2 \cdot Y_3 \quad (3.28)$$

Hence substituting Eqs. (3.27) into (3.28) yields for the combined output

$$Y = z^{-3} \cdot X + (1 - z^{-1})^3 \cdot E_3$$

Thus the TFOC  $\Delta\Sigma$  modulator shapes the noise by  $(1 - z^{-1})^3$ , while it

delays the input by three samples, i.e. it is a third-order ΔΣ modulator.

The combination of the output of the stages in an  $N$ th-order multi-stage ΔΣ modulator comprising non-delaying stages is as follows. The output of the first stage is the input signal  $X$  plus the quantization noise  $E_1$  of the quantizer of the first stage shaped by a low-order noise transfer function of the form  $(1 - z^{-1})^{k_1}$ , where  $k_1 = 1$  or 2 is the order of the stage

$$Y_1 = X + (1 - z^{-1})^{k_1} E_1 \quad (3.29)$$

Stage  $n$  has the quantization noise of the previous stage as input. The output is the noise of the  $n-1$  stage plus the noise  $E_n$  from the  $n$ -th quantizer, again shaped by a noise transfer function  $(1 - z^{-1})^{k_n}$ ,  $k_n = 1$  or 2

$$Y_n = E_{n-1} + (1 - z^{-1})^{k_n} E_n \quad (3.30)$$

Differentiating the output of the  $n$ -th stage  $N_n = \sum_{i=1}^{n-1} k_i$  times yields

$$V_n = (1 - z^{-1})^{N_n} E_{n-1} + (1 - z^{-1})^{N_n+k_n} E_n \quad (3.31)$$

As  $N_n = N_{n-1} + k_{n-1}$ , in order to cancel out the shaped quantization noise of all the stages but the last, all the differentiated outputs  $V_i$ ,  $i = 2, \dots, N$ , are added together with the necessary signs. Thus the output signal is the input signal plus the quantization noise of the last stage shaped by  $(1 - z^{-1})^N$ . If delaying first- and second-order stages are to be used, then delays are appropriately used to account for the delay introduced by ΔΣ stages as shown in Fig. 3.12 for the TFOC multi-stage ΔΣ modulator.

The success of multi-stage ΔΣ modulators in randomizing the quantization noise is considerable [23]. This is demonstrated in Fig. 3.13. While the single-stage modulators clearly do not have white quantization noise, the multi-stage third-order ΔΣ modulator has. This is a very interesting result, as the three sections that comprise it have the typical

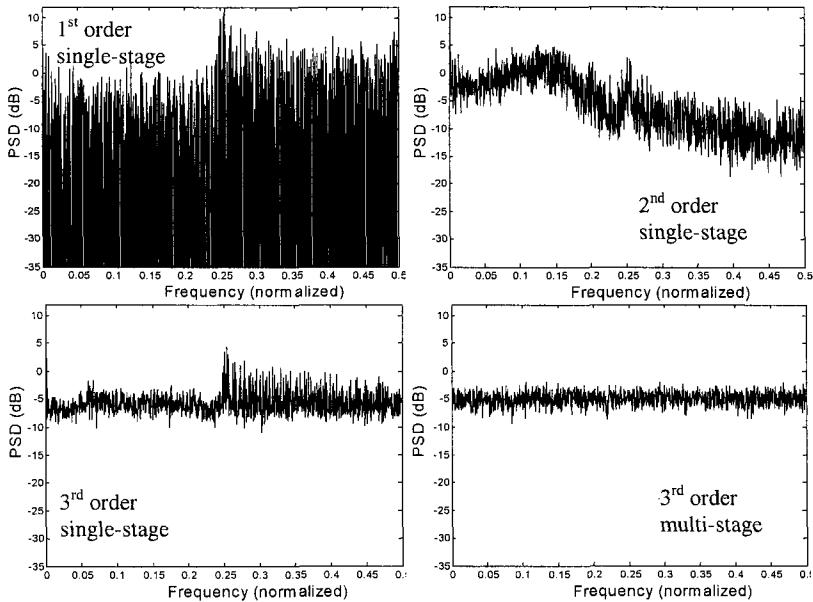


Fig. 3.13 Quantization noise spectra for single-stage  $\Delta\Sigma$  modulators of orders 1 to 3 and the TFOC multi-stage  $\Delta\Sigma$  modulator.

discrete noise spectrum of the first-order  $\Delta\Sigma$  modulator.

Multi-stage  $\Delta\Sigma$  modulators also inherit the excellent stability properties of their low-order stages, thus high SNR is achieved. The SNR of the TFOC  $\Delta\Sigma$  modulator is depicted in Fig. 3.14.

Multi-stage  $\Delta\Sigma$  modulators are commercially used as very successful DACs in CD players. In such a configuration the modulator accepts interpolated PCM signals and converts them to single-bit  $\Delta\Sigma$  signals. The  $\Delta\Sigma$  bit-streams are easily converted to analogue form, and as they are oversampled, the requirements of the analog smoothing filter are relaxed. Also, the quantizers of the modulator stages are realized by just keeping the sign bit of their input. The use of  $\Delta\Sigma$  modulators for D/A conversion is further considered in Sec. 8.4.

Unfortunately, the use of multi-stage  $\Delta\Sigma$  modulators as ADCs is not as straightforward. In such a configuration the modulators comprise an analog part (the low-order modulator stages) and a digital part (the logic

that combines the output of the stages). In order for the digital differentiators to cancel out the noise of all but the last quantizer, the analog modulators must ideally shape the noise of the stages. Finite opamp gain [2,13,17,18,22] makes the ideal shaping impossible, as even SC integrator topologies with low opamp gain sensitivity [2,15,24,25] cause considerable discrepancies from the ideal transfer functions. As a result of the mismatch of the analog and digital parts, lower-order shaped noise leaks into the output of the modulator and hence the achievable SNR is reduced. The effect of mismatch is considered in more detail in Sec. 6.2.1.

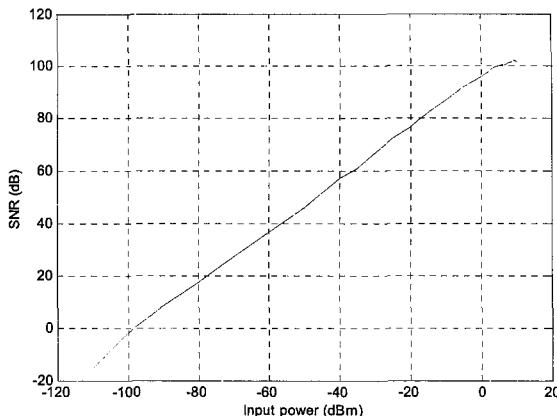


Fig. 3.14 SNR as a function of input power for a given TFOC third-order multi-stage  $\Delta\Sigma$  modulator with sinusoidal input. The oversampling ratio is 64. The maximum SNR is 102.1 dB, achieved at 9.5 dBm of input power. The dynamic range is 108.2 dB. This figure is comparable with the performance of a fourth-order single-stage  $\Delta\Sigma$  modulator.

As there are no stability constraints in the design of multi-stage  $\Delta\Sigma$  modulators, any increase of their order by cascading more stages results to an increase of the SNR. The limit to this increase is the matching of the digital and analog parts. The component of the output noise due to leakage of the lower-order shaped noise becomes relatively larger compared to the increasingly higher-order shaped noise term of the final stage and, finally, renders any improvement of the latter useless.

Finally note that the output of a multi-stage  $\Delta\Sigma$  modulator is not

single-bit, even though the quantizers of the stages comprising it are. The bits are increased by the combining logic.

**Example 3.4** Find the number of bits of the output of the TFOC  $\Delta\Sigma$  modulator.

**Solution** Each of the outputs  $Y_i$  of the three stages is  $\pm 1$ . The two delays of the combining logic for  $Y_1$  do not change the possible values. The differentiator of the combining logic for  $Y_2$  increases the number of possible values to  $-2, 0$  and  $2$ . The two differentiators of the combining logic for  $Y_3$  increase the number of possible values to  $-4, -2, 0, 2$  and  $4$ . Summing the three branches together yields as a possible result the values  $-7, -5, -3, -1, 1, 3, 5$  and  $7$ . These 8 different values of the output of the TFOC  $\Delta\Sigma$  modulator need 3 bits to be represented.

The increased number of bits of the output signal of the multi-stage  $\Delta\Sigma$  modulators increases the complexity of any digital signal processing that follows the modulator.

### 3.8 Multi-Bit $\Delta\Sigma$ Modulators

As it was explained in Chapter 2, increasing the bits  $b$  (or equivalently the levels  $N = 2^b$ ) of any quantizer by one leads to an increased SNR by 6.02 dB. Hence increasing the quantizer bits can increase the SNR of a  $\Delta\Sigma$  modulator. Also, the number of quantizer bits is related to the maximum stable input of a  $\Delta\Sigma$  modulator and its noise and signal transfer functions. To show this, the output  $y(n)$  of a  $\Delta\Sigma$  modulator is written as the sum of the quantizer input  $u(n)$  plus the quantization noise  $e(n)$ . Hence

$$u(n) = y(n) - e(n) \quad (3.32)$$

For the z-transform  $X$  of the input  $x(n)$  to the  $\Delta\Sigma$  modulator,  $Y$  of its output and  $E$  of the additive noise it is

$$Y = STF(z) \cdot X + NTF(z) \cdot E \quad (3.33)$$

Hence in the time domain

$$y(n) = \sum_k stf(k)x(n-k) + \sum_k ntf(k)e(n-k) \quad (3.34)$$

where  $stf(n)$  is the STF impulse response and  $ntf(n)$  is the NTF impulse response. Using (3.34), (3.32) can be written as

$$u(n) = \left( \sum_k stf(k)x(n-k) + \sum_k ntf(k)e(n-k) \right) - e(n) \quad (3.35)$$

Denote the maximum absolute value of the input by  $\|x\|_\infty$  and assume that the quantizer is never overloaded. This guarantees stability of the ΔΣ modulator, although it is a very strict condition; ΔΣ modulators with occasional overloading can be stable. Then, if the full scale of the quantizer is normalized to unity, the quantization noise is bounded by

$$|e(n)| \leq \frac{2}{N-1} \quad (3.36)$$

and the quantizer input is bounded by

$$|u(n)| \leq 1 + \frac{2}{N-1} \quad (3.37)$$

Employing the Schwartz inequality on (3.35) and replacing (3.36) and (3.37) yields

$$N \geq 1 + \frac{2\|ntf\|_1 - 4}{1 - \|x\|_\infty \|stf\|_1} \quad (3.38)$$

where  $\|\cdot\|_1$  denotes the one-norm of a sequence and  $\|x\|_\infty$  is normalized to the full scale of the quantizer. Two facts can be derived from (3.38). First, increasing  $N$  in a given ΔΣ modulator allows for larger maximum stable input. This is shown in Fig. 3.15. Second, increasing  $N$  allows ΔΣ modulators that implement NTFs to offer more in-band noise attenuation without increasing the order, while stability is maintained although the out-of-band noise amplification increases.

**Example 3.5** Using (3.38), find the minimum number of bits required to realize the  $NTF(z) = (1 - z^{-1})^3$ , which is unstable when realized as a single-bit single-stage ΔΣ modulator, and compare it to the actual number of bits obtained by simulation.

**Solution** It is  $\|ntf\| = 8$ , hence (3.38) yields  $N \geq 13$  and  $b \geq 4$ .

Simulation shows that a single-stage  $\Delta\Sigma$  modulator realizing this NTF, becomes stable if the quantizer is 3-bit. The overestimation of the needed quantizer bits by (3.38) is due to the demand for quantizer non-overloading, which is a very strict stability criterion.

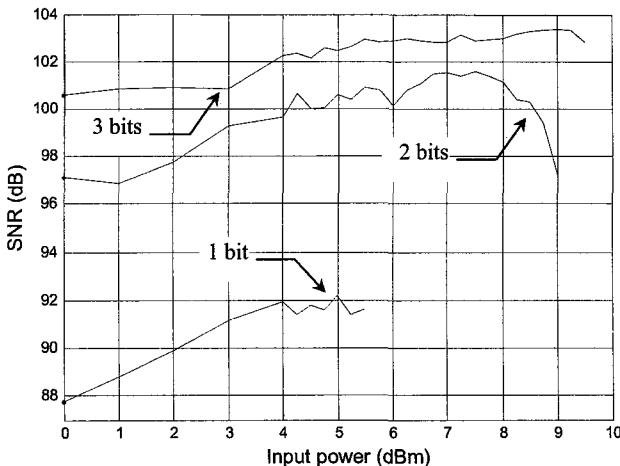


Fig. 3.15 Variation of the SNR of a third-order  $\Delta\Sigma$  modulator as the quantizer bits are increased from 1 to 3. The achievable SNR and maximum input power for which the modulator remains stable increase, although the NTF is constant.

The drawback of multi-bit  $\Delta\Sigma$  modulators is that a multi-bit local ADC must be accompanied by a multi-bit local DAC in the feedback path. This DAC should be fast, but only of a few bits (say  $b$ ) of resolution. Such DACs are usually built using a thermometer-type decoder and  $2^b$  parallel unit elements of the same value. These elements can be resistors, capacitors, or transistors and can be activated by the  $2^b$  decoder outputs. The analog value is the sum of their voltages or currents and their accuracy determines the linearity of the DAC.  $\Delta\Sigma$  modulators are relatively immune to ADC errors due to noise shaping. When the quantizer is single-bit, there exists no DAC error. But in the case of multi-bit  $\Delta\Sigma$  modulators, circuit imperfections cause mismatches in the unit elements of the DAC, thus DAC errors occur, to which the

modulators are very sensitive. Referring to Fig. 3.16, where the quantization error is modeled by an additive white noise source  $E_{ADC}$  and the local DAC error by a similar source  $E_{DAC}$ , the transfer function of the system is

$$Y = \frac{L(z)}{1+L(z)} X - \frac{L(z)}{1+L(z)} E_{DAC} + \frac{1}{1+L(z)} E_{ADC} \quad (3.39)$$

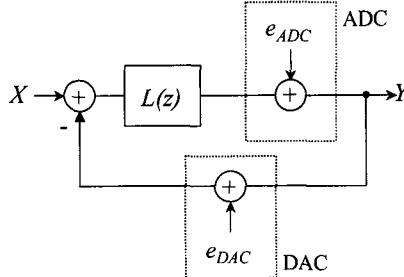


Fig. 3.16 Linearized block diagram of the multi-bit  $\Delta\Sigma$  modulator.

Thus the DAC error is only shaped by the STF, which is approximately all-pass at the signal frequency band, i.e. it is not attenuated. Consequently the DAC linearity should be as high as the desired modulator resolution. This is very difficult to achieve without element trimming, which is impractical for integrated circuit designs.

In order for multi-bit  $\Delta\Sigma$  modulators to be of any practical use, actions have to be taken to increase the DAC linearity, as discussed in Sec. 6.8. Unfortunately, all these actions considerably increase the complexity of the  $\Delta\Sigma$  modulator.

Thus far multi-bit single-stage  $\Delta\Sigma$  modulators have been considered. Multi-stage  $\Delta\Sigma$  modulators can also employ multi-bit quantizers to enhance the SNR [27]. Two important considerations should be taken into account:

- As the quantization noise of the intermediate stages does not appear at the output of the modulator (if properly cancelled out), the use of multi-bit local ADC/DAC pairs at these stages does not improve the SNR. Hence only the final stage employs a multi-bit quantizer.
- The linearity of the local ADC/DAC pair in the last  $\Delta\Sigma$  modulator

should be as high as the resolution of a first or second-order modulator. The local ADC noise is only first- or second-order shaped by the modulator stage; the differentiators of the digital noise cancellation circuitry provide the rest of the noise shaping, and this shaping is applied to both the local ADC and DAC errors. Thus if the modulator is of  $N$ th-order and the last stage is of  $k$ th-order, then the local ADC/DAC linearity error will be shaped by  $(1-z^{-1})^{N-k}$ , whereas the quantization noise will be shaped by  $(1-z^{-1})^N$ , i.e. only one or two orders higher.

Thus the local ADC/DAC block linearity constraints are greatly reduced in a multistage modulator. Should the last stage be a first-order one, the use of a multi-bit quantizer becomes practical. Such modulators can yield high SNR at low oversampling ratios [27], and hence are very useful in high frequency applications.

### 3.9 Hybrid $\Delta\Sigma$ Modulators

Leslie and Singh in [28] have proposed the use of a multi-bit ADC in the feed-forward path and a single-bit DAC in the feedback path of a  $\Delta\Sigma$  modulator. A digital filter cancels out the single-bit quantization noise, so that only the multi-bit quantization noise remains. The block diagram of the resulting modulator is shown in Fig. 3.17.

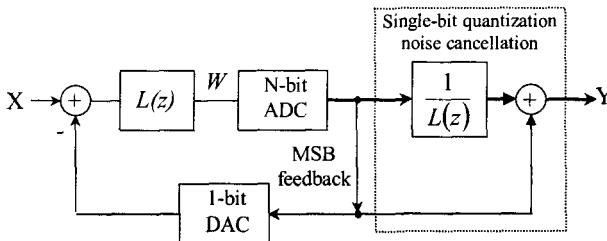


Fig. 3.17 Multi-bit ADC single-bit DAC  $\Delta\Sigma$  modulator.

The  $N$ -bit ADC 1-bit DAC block is equivalent to a single bit quantizer, and introduces error  $e_1$  in the feedback path and the lower digital path.

The  $N$ -bit ADC introduces noise  $e_N$  in the upper digital path. The power of  $e_N$  is  $6.02 \cdot N$  dB lower than that of  $e_1$ . Hence the goal of the system is to cancel out  $e_1$ . The linearized system yields

$$W = [X - (W + E_1)] \cdot L(z) \quad (3.40)$$

and

$$Y = (W + E_N) \cdot \frac{1}{L(z)} + W + E_1 \quad (3.41)$$

Substituting  $W$  from (3.40) into (3.41) yields

$$Y = X + \frac{1}{L(z)} \cdot E_N \quad (3.42)$$

i.e. the 1-bit quantization noise is cancelled out.

This scheme has an analog and a digital part, as the multi-stage ΔΣ modulators. In order for the 1-bit quantization noise to be cancelled out, the two parts must match. Finite opamp gain alters the actual transfer function of the analog loop filter, thus single-bit quantization noise leaks in the output of the modulator.

### 3.10 Adaptive ΔΣ Modulators

The idea of adaptive oversampling converters dates back to the adaptive quantizer step in Delta modulators [29]. That configuration was proposed to reduce the effect of slope overload in Delta modulators. Similar is the goal of the adaptive ΔΣ modulators; if the quantizer is not overloaded, then the modulator is stable. Yu in [30] has considered two types of adaptive ΔΣ modulators. ΔΣ modulators can have adaptive quantizers. The variation of the quantizer step ensures stability. Alternatively, ΔΣ modulators can have adaptive loop filters. The NTF then changes to decrease the noise power amplification. Then, the in-band noise attenuation also decreases, but the modulator remains stable. Yu has shown that the adaptation of even a single coefficient of the loop filter is sufficient. ΔΣ modulators with adaptive loop filters are not suited for ADCs, as the implementation of an adaptive analog filter is very complicated.

### 3.11 Band-Pass $\Delta\Sigma$ Modulators

The advantages of oversampled, noise-shaping methods for A/D and D/A conversion of low frequency signals can also be obtained for narrow-band signals. The zeros of the NTF can be distributed in the narrow signal band, allowing the separation of the signal from the quantization noise.

Band-pass  $\Delta\Sigma$  modulators share the advantages of their low-pass counterparts, and can be derived from them using the following low-pass to band-pass transformation:

$$z \rightarrow -z^2 \quad (3.43)$$

Modulators derived applying (3.43) share the same stability and noise performance as their prototypes [2]. Under this transformation, the signal band is centered at  $f_s/4$ , allowing for simple decimator design [2]. Clearly, to achieve the same noise suppression, band-pass modulators are of twice the order than their low-pass counterparts, although this increase in order does not correspond to a similar increase in circuit complexity, as only the first integrator design is crucial.

**Example 3.6** Transform the NTF of the second-order low-pass  $\Delta\Sigma$  modulator to the corresponding band-pass modulator.

**Solution** The NTF of the second-order low-pass  $\Delta\Sigma$  modulator is

$$NTF_{LP}(z) = (1 - z^{-1})^2$$

Applying (3.43) yields the corresponding band-pass NTF

$$NTF_{BP}(z) = NTF_{LP}(z)|_{z \rightarrow -z^2} = \left(1 - (-z^2)^{-1}\right)^2 = (1 + z^{-2})^2 \quad (3.44)$$

The band-pass NTF is compared to its low-pass prototype in Fig. 3.18.

The  $\Delta\Sigma$  modulator realizing (3.44) has the same SNR and dynamic range as its low-pass prototype and operates on signals around  $f_s/4$ . Its output spectrum is compared to that of the low-pass modulator in Fig. 3.19.

The main application of band-pass  $\Delta\Sigma$  modulators is in RF communication systems; hence over the past few years there is a trend towards higher sampling frequencies and higher SNR at wider signal

bandwidths for band-pass  $\Delta\Sigma$  modulators. The achievements are impressive: In 1993 the state of the art was a fourth-order  $\Delta\Sigma$  band-pass modulator for use in a GSM transceiver; clocked at 26 MHz, the modulator converted signals of 200 kHz bandwidth centered at 6.5 MHz and achieved SNR of 55 dB. The process used was a 1.2  $\mu\text{m}$  / 7 GHz BiCMOS [9]. More recently, higher IF and converted bandwidths with lower power dissipation are addressed. Bazarjani and Snelgrove [31] have proposed a fourth-order SC band-pass  $\Delta\Sigma$  modulator converting a bandwidth of 1.25 MHz centered at 40 MHz with an SNR of 47 dB. The process used was a 0.5  $\mu\text{m}$  CMOS and the chip dissipated 65 mW out of a 3 V supply [31]. Finally, Gao and Snelgrove have announced a modulator that samples signals of 200 kHz bandwidth centered at 950 MHz at a sampling frequency of 3.8 GHz, achieving an SNR of 57 dB. The modulator dissipates 135 mW out of a 5 V supply [32].

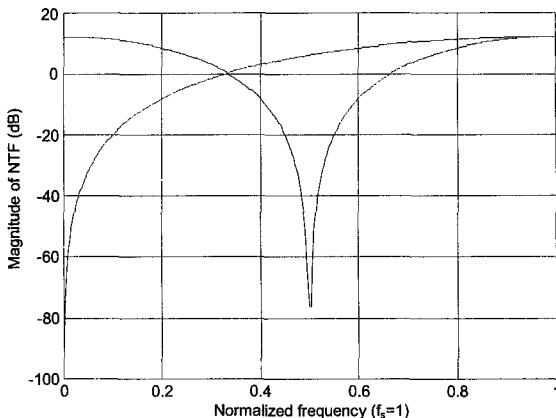


Fig. 3.18 NTF of the low-pass and band-pass second-order  $\Delta\Sigma$  modulator.

The main concern with band-pass modulators is that they have to sample higher-frequency signals, i.e. they need faster samplers. Signals for low-pass modulators have frequencies at least  $2R$  times smaller than the sampling frequency. Band-pass modulators usually have a signal band centered at  $f_s/4$ . Hence, the available time for the settling of the opamps is reduced. A two-path approach in [33] alleviates this problem.

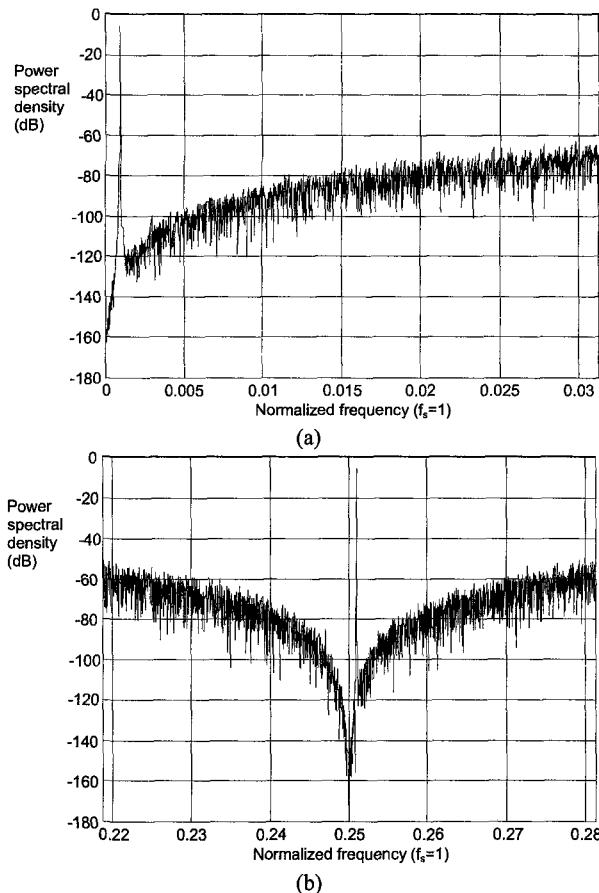


Fig. 3.19 Output spectra of a. the second-order low-pass  $\Delta\Sigma$  modulator and b. its equivalent band-pass.

Should discrete-time circuits be used for band-pass modulators, the sample-and-hold circuit at the input must operate much faster. Should continuous-time circuits be used, the sample-and-hold circuit is in the modulator loop and any errors are noise-shaped. But such designs can be implemented either using LC circuits with high-Q off-chip inductors, or using  $g_mC$  circuits, whose linearity is bounded by the linearity of the transconductance. Lately, IC fabrication process enhancements have

allowed the integration of inductors [34], thus the use of LC circuits seems to be the most promising for high frequency band-pass ΔΣ modulators. Indeed, this is the technology used in [32] for the results at a sampling frequency of 3.8 GHz.

Finally note that not all the architectures of low-pass ΔΣ modulators can be converted into band-pass. For example, the cascaded ΔΣ - pipeline scheme proposed in [5] is restricted in speed by the need of a pipeline ADC, which is slow and cannot operate at the speed necessary for band-pass modulators.

### 3.12 Summary

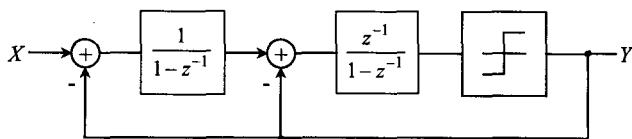
The effect of embedding a filter and a quantizer in a feedback loop is examined using the linear model for the quantizer. The resulting noise-shaping oversampled modulator is investigated both for first- and second-order loop filters, and for higher-order loop filters. The second case gives rise to stability problems. Other architectures for the implementation of higher-order ΔΣ modulators are also considered: multi-stage, multi-bit and hybrid architectures. In addition adaptive ΔΣ modulators are considered. Finally, ΔΣ modulation is applied to band-pass signals.

### Problems

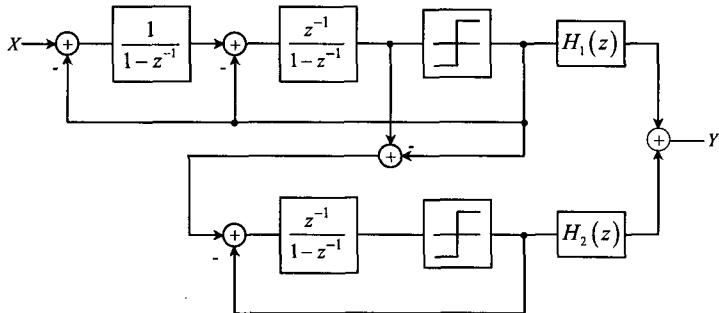
- 3.1 Assume a second-order ΔΣ modulator with  $NTF(z) = (1 - z^{-1})^2$

that modulates a signal oversampled  $R$  times. Find the resulting attenuation of the quantization noise power as a function of  $R$ .

- 3.2 The most usual implementation of a second-order ΔΣ modulator is by cascading a non-delaying and a delaying integrator, as in Fig. 3.20. Demonstrate using linear theory that the system has  $NTF(z) = (1 - z^{-1})^2$ .

Fig. 3.20 Block diagram of second-order  $\Delta\Sigma$  modulator for problem 3.2.

- 3.3 Find the gain in SNR of the second-order  $\Delta\Sigma$  modulator over the first-order using linear theory as a function of the oversampling ratio  $R$ . Compare it to the actual value for sinusoidal input and 64 times oversampling given in Table 3.1.
- 3.4 A cascade of a second-order and a first-order  $\Delta\Sigma$  modulator stages is depicted in Fig. 3.21. Find the necessary digital transfer functions  $H_1(z)$  and  $H_2(z)$  that combine the outputs of the two stages so that the system will be a third-order  $\Delta\Sigma$  modulator with  $NTF(z) = (1 - z^{-1})^3$ . Which is the STF? The resulting two-stage implementation of the third-order  $\Delta\Sigma$  modulator is the Second-Order First-Order Cascade (SOFOC).

Fig. 3.21 Block diagram of two-stage  $\Delta\Sigma$  modulator for problem 3.4.

## References

- [1] J. Candy, ‘Decimation for Sigma-Delta Modulation’, *IEEE Trans. Commun.*, vol. 34, pp. 72-76, 1986.
- [2] S. Norsworthy, R. Schreier and G. Temes, *Delta-Sigma Data Converters, Theory, Design and Simulation*, New York, IEEE Press, 1997.
- [3] S. Park, *Motorola Digital Signal Processors: Principles of Sigma-Delta Modulation for Analog-to-Digital Converters*, Motorola Application Note, 1996.
- [4] P. Aziz, H. Sorensen and J. Spiegel, ‘An Overview of Sigma-Delta Converters’, *IEEE Signal Processing Magazine*, January 1996.
- [5] T. Brooks, T. Robertson, D. Kelly, A. Muro and S. Harston, ‘A Cascaded Sigma-Delta Pipeline A/D Converter with 1.25 MHz Signal Bandwidth and 89 dB SNR’, *IEEE Journal of Solid-State Circuits*, vol. 32, pp. 1896-1906, December 1997.
- [6] K. Chao, S. Nadeem, W. Lee and C. Sodini, ‘A Higher Order Topology for Interpolative Modulators for Oversampling A/D Converters’, *IEEE Trans. on Circuits and Systems*, vol. 37, pp. 309-318, March 1990.
- [7] L. Larson, T. Cataltepe and G. Temes, ‘Multibit Oversampled S-D A/D Convertor with Digital Error Correction’, *Electron. Lett.*, vol. 24, pp. 1051-1052, August 1988.
- [8] W. Press, S. Teukolsky, W. Vetterling and B. Flannery, *Numerical Recipes in C: The Art of Scientific Computing*, Cambridge University Press, 1988.
- [9] G. Troster, P. Sieber, K. Schoppe, A. Wedel, E. Zocher, J. Arndt, H-J. Dressler, H-J. Golberg and W. Schardein, ‘An Interpolative Bandpass Converter on a 1.2mm BiCMOS Analog/Digital Array’, *IEEE J. Solid-State Circuits*, vol. 28, no. 4, pp. 471-477, April 1993.
- [10] R. Adams, P. Ferguson, A. Ganesan, S. Vincelette and R. Libert, ‘Theory and Practical Implementation of a Fifth-Order Sigma-Delta A/D Converter’, *J. Audio Eng. Soc.*, vol. 39, no. 7/8, July/August 1991.
- [11] R. Carley, ‘A Noise Shaping Coder Topology for 15+ Bit Converters’, *IEEE J. Solid-State Circuits*, vol. 24, pp. 267-273, April 1989.
- [12] I. Fujimori, K. Koyama, D. Trager, F. Tam and L. Longo, ‘A 5-V Single-Chip Delta-Sigma Audio A/D Converter with 111 dB Dynamic Range’, *IEEE Journal of Solid-State Circuits*, vol. 32, pp. 329-336, March 1997.
- [13] S. Moussavi and B. Leung, ‘High-Order Single-Stage Single-Bit Oversampling A/D Converter Stabilized with Local Feedback Loops’, *IEEE Trans. on Circuits and Systems-II: Analog and Digital Signal Proc.*, vol. 41, pp. 19-25, January 1994.
- [14] R. Schreier, *Sigma-Delta Toolbox*, Oregon State University, 1997.
- [15] R. Gregorian and G. C. Temes, *Analog MOS Switched-Capacitor and Continuous-Time Integrated Circuits and Systems*, John Wiley & Sons, 1986.
- [16] C. Dunn and M. Sandler, ‘Use of Clipping in Sigma-Delta Modulators’, *IEE Colloquium on Oversampling Techniques and Sigma-Delta Modulation*, London, 1994.
- [17] S. Norsworthy, ‘Optimal Nonrecursive Noise Shaping Filters for Oversampling Data Converters, Part 1: Theory’, *IEEE Proc. ISCAS’93*, vol. 2, pp. 1353-1356, May 1993.

- [18] S. Norsworthy, ‘Optimal Nonrecursive Noise Shaping Filters for Oversampling Data Converters, Part 2: Applications’, *IEEE Proc. ISCAS’93*, vol. 2, pp. 1357-1360, May 1993.
- [19] S. Au and B. Leung, ‘A 1.95-V, 0.34-mW, 12-b Sigma-Delta Modulator Stabilized by Local Feedback Loops’, *IEEE J. Solid-State Circuits*, vol. 32, pp. 321-328, March 1997.
- [20] B. Boser and B. Wooley, ‘The Design of Sigma-Delta Modulation Analog-to-Digital Converters’, *IEEE J. Solid State Circuits*, vol. 23, pp. 1298-1308, December 1988.
- [21] R. Schreier, ‘An Empirical Study of High-Order Single-Bit Delta-Sigma Modulators’, *IEEE Trans. Circuits and Sys.-II: Analog and Digital Signal Proc.*, vol. 40, pp. 461-466, August 1993.
- [22] D. Ribner, A Comparison of Modulator Networks for High-Order Oversampled SD Analog-to-Digital Converters, *IEEE Trans. on Circuits and Sys.*, vol. 38, pp. 145-159, February 1991.
- [23] W. Chou, P. Wong, and R. Gray, ‘Multi-Stage Sigma-Delta Modulation’, *IEEE Trans. Information Theory*, vol. 35, pp. 784-796, July 1989.
- [24] M. Ghausi and K. Laker, *Modern Filter Design*, Prentice-Hall, 1981.
- [25] R. Unbehauen and A. Cichocki, *MOS Integrated Circuits for Signal Processing*, Springer-Verlang, 1989.
- [26] B. Leung and S. Sutarja, ‘Multi-Bit S-D A/D Converter Incorporating a Novel Class of Dynamic Element Matching’, *IEEE Trans. Circuits and Systems-II: Analog and Digital Signal Proc.*, vol. 39, pp. 35-51, January 1992.
- [27] A. Feldman, *High-Speed, Low-Power Sigma-Delta Modulators for RF Baseband Channel Applications*, Ph.D. thesis, University of California, Berkeley, 1997.
- [28] T. Leslie and B. Singh, ‘An Improved Sigma-Delta Modulator Architecture’, *IEEE Proc. ISCAS’89*, vol. 1, pp. 372-375, May 1990.
- [29] R. Steele, *Delta Modulation Systems*, Pentech Press, London, 1975.
- [30] J. Yu, *Design of Fixed and Adaptive Sigma-Delta Modulators*, Ph.D. thesis, King’s College, September 1992.
- [31] S. Bazarjani and W. Snelgrove, ‘A 160-MHz Fourth-Order Double-Sampled SC Bandpass Sigma-Delta Modulator’, *IEEE Trans. Circuits and Sys.-II: Analog and Digital Signal Proc.*, vol. 45, pp. 547-555, May 1998.
- [32] W. Gao and W. Snelgrove, ‘A 950-MHz IF Second-Order Integrated LC Bandpass Delta-Sigma Modulator’, *IEEE J. Solid-State Circuits*, vol. 33, pp. 723-732, May 1998.
- [33] A. Ong and B. Wooley, ‘A Two-Path Bandpass SD Modulator for Digital IF Extraction at 20 MHz’, *IEEE J. Solid State Circuits*, vol. 32, pp. 1920-1934, December 1997.
- [34] M. Steyaert, M. Borremans, J. Craninckx, J. Crols, J. Janssens and P. Kinget, ‘RF CMOS Design, Some Untold Pitfalls’, in *Analog Circuit Design, MOST RF Circuits, Sigma-Delta Converters and Translinear Circuits*, W. Sansen, R. van de Plassche and J. Huijsing, editors, Kluwer Academic Publishers, 1996.

## **Chapter 4**

# **Single-Bit Single-Stage $\Delta\Sigma$ Modulators Modeling and Design**

### **4.1 Introduction**

In the previous chapter various architectures and types of  $\Delta\Sigma$  Modulators were introduced. Some of these are more frequently used than others mostly for reasons of circuit simplicity. In this context, the single-bit, single-stage architecture has gained the largest popularity so far. Consequently, we have chosen this type of  $\Delta\Sigma$  Modulator to study further in this and the following 2 chapters.

One important aspect in studying and subsequently designing  $\Delta\Sigma$  Modulators is the modeling of the quantization noise. An accurate model for this noise is essential for the creation of proper NTFs, which will lead to most effective designs. Thus in this Chapter, modeling of single-bit, single-stage  $\Delta\Sigma$  Modulators is examined first in Sec. 4.2. The quantization noise models used are the linear and the quasi-linear ones, which are reviewed for this purpose. Next the characteristics of a ‘good’ NTF are outlined in Sec. 4.3. Stability of  $\Delta\Sigma$  modulators is explained and criteria for stable circuits are presented in Secs. 4.4 and 4.5, respectively. Then based on all this material, NTFs for optimal performance of the  $\Delta\Sigma$  modulator are derived and discussed in Sec. 4.6. Finally,

the characteristics of these functions are given for comparison in Sec. 4.7, followed by the conclusions in Sec. 4.8.

## 4.2 Modeling of $\Delta\Sigma$ Modulators

A typical structure of a single-bit, single-stage  $\Delta\Sigma$  modulator is shown in Fig. 4.1. For this circuit the output can be expressed in the z-domain as a function of both the signal and the quantization noise as follows:

$$Y(z) = STF(z) \cdot X(z) + NTF(z) \cdot E(z) \quad (4.1)$$

Here  $STF$  is the Signal Transfer Function

$$STF(z) = \frac{L_0(z)}{1 + L_1(z)} = 1 \quad (4.2)$$

and  $NTF$  is the Noise Transfer Function

$$NTF(z) = \frac{1}{1 + L_1(z)} \quad (4.3)$$

The STF should be an all-pass function and this is achieved when  $L_0(z)$  is equal to  $L_1(z)$  and  $|L_1(z)| \gg 1$ . Furthermore, the NTF has to be a high-pass transfer function, since its role is to move the quantization noise power out of the signal band.

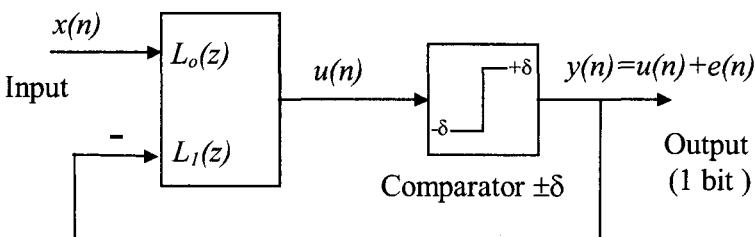


Fig. 4.1 A typical structure of a single-bit, single-stage  $\Delta\Sigma$  modulator.

Knowledge of various parameters of the quantization noise, such as its power spectral density (PSD) and its probability density function (PDF) will help to evaluate the performance of the  $\Delta\Sigma$  modulator. This is carried out by means of the maximum value of the SNR as well as the maximum acceptable value of the signal for

the modulator to remain stable. Modeling of the  $\Delta\Sigma$  modulator with respect to quantization noise is discussed in this section employing the simple linear model [1,2] and the quasi-linear model [3].

#### 4.2.1 Linear model

According to the linear model the quantizer in Fig. 4.1 is represented by the quantization noise source  $e(n)$ . Then the output  $y(n)$  will be as follows:

$$y(n) = u(n) + e(n) \quad (4.4)$$

Ignoring completely the fact that  $e(n) = \delta \cdot \text{sgn}\{u(n)\} - u(n)$ , the model is completed assuming the following:

- a. The noise  $e(n)$  is white in the frequency interval  $[0, f_s]$ .
- b. The PDF  $f_e(e)$  of the quantization error is uniform in the interval  $[-\delta, \delta]$ .

These two assumptions are adequate for evaluating the maximum SNR from (4.3) as follows:

According to the first assumption, the quantization noise power in the signal band is

$$P_{e-in} = \frac{P_e}{\pi} \int_0^{\pi/R} |NTF(e^{j\theta})|^2 d\theta \quad (4.5)$$

where  $R$  is the oversampling ratio and  $P_e$  the total quantization noise power, which according to the second assumption is

$$P_e = \int_{-\infty}^{+\infty} e^2 f_e(e) de = \delta^2 / 3 \quad (4.6)$$

For a sinusoidal input signal with maximum amplitude  $x_{max}$  the  $SNR_{max}$  is

$$SNR_{max} = \frac{x_{max}^2 / 2}{\frac{P_e}{\pi} \int_0^{\pi/R} |NTF(e^{j\theta})|^2 d\theta} \quad (4.7)$$

Since in the first-order  $\Delta\Sigma$  modulator  $x_{max} = \delta$ , it is found that

$$SNR_{\max} = \frac{\delta^2 / 2}{\frac{P_e}{\pi} \int_0^{\pi_R} |NTF(e^{j\theta})|^2 d\theta} \quad (4.8)$$

According to the material exposed so far, it may be thought that the design of  $\Delta\Sigma$  modulators is solved. However, in practice the operation of the modulators is not satisfactory for the following reasons:

- a. Many of them, which are expected to be stable and give high SNR, are in fact unstable. Moreover, most of the  $\Delta\Sigma$  modulators of orders higher than 2, are not stable for the maximum input signal  $x_{max} = \delta$ .
- b. Often, the previously set assumptions are not valid and, consequently, equations (4.5) to (4.8) give a misleading estimation of the  $SNR_{max}$ .

The first of the above remarks is regarded as the most significant. Instability occurs when the transfer functions STF and NTF have at least one of their poles outside the unit circle. The position of the poles can be controlled by an amplification factor  $k$ , which can be introduced as shown in Fig. 4.2. Consequently, the STF and NTF become

$$STF(z) = \frac{k \cdot L_0(z)}{1 + k \cdot L_1(z)} \quad (4.9.a)$$

$$NTF(z) = \frac{1}{1 + k \cdot L_1(z)} \quad (4.9.b)$$

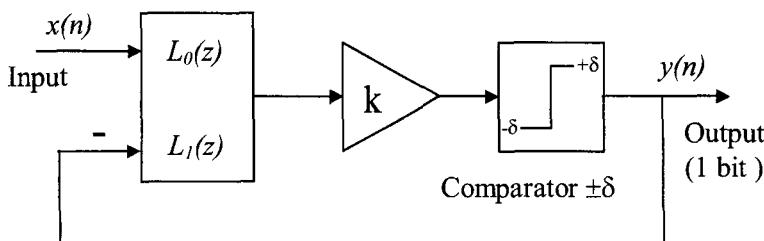


Fig. 4.2 The operation of a  $\Delta\Sigma$  modulator is not affected when a positive gain  $k$  is placed before the quantizer.

In the general case  $k$  is unknown and the transfer functions are not exactly known. This ambiguity is almost resolved in the quasi-linear model, where the values of  $k$  are determined.

#### 4.2.2 Quasi-Linear model

In this model [3], the input to the quantizer  $u(n)$  is divided into  $u_s(n)$  and  $u_e(n)$ . The first of these components is related to the input signal while the second one to the quantization error. These components are multiplied by different amplification factors  $k_s$  and  $k_e$  respectively. Thus the quantizer is replaced by a linear system described by

$$y(n) = y_s(n) + y_e(n) = k_s \cdot u_s(n) + k_e \cdot u_e(n) + e(n) \quad (4.10)$$

which in turn divides the modulator into two linear systems shown in Fig. 4.3.

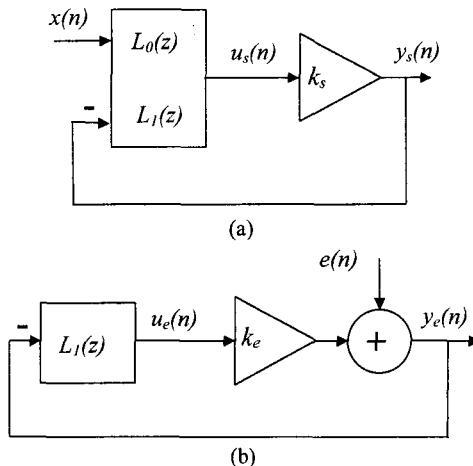


Fig. 4.3 According to the quasi-linear model, the  $\Delta\Sigma$  modulator is split into two linear systems describing its functioning for a. the input signal and b. the quantization error.

Two assumptions are made in this model, too. One refers to the noise  $e(n)$  whilst the other to the component  $u_e(n)$ . These are as follows:

- The quantization error  $e(n)$  is white noise (as in the case of the linear model), uncorrelated with the input signal. No

assumption is made for its probability density function (PDF).

b. The PDF of  $u_e(n)$  is assumed Gaussian with zero mean.

The gains  $k_s$  and  $k_e$  are introduced so that the two hypotheses, namely,

i.  $e(n)$  is uncorrelated to the input signal and

ii.  $e(n)$  is white

would not contradict each other, as it happens with the linear model. For the noise  $e(n)$  to be uncorrelated with the signal  $x(n)$  we should have

$$E\{u_s(n)u_e(n)\} = 0 \quad (4.11.a)$$

and

$$E\{e(n)u_s(n)\} = 0 \quad (4.11.b)$$

Furthermore, if  $e(n)$  is white, then

$$E\{e(n)u_e(n)\} = 0 \quad (4.11.c)$$

Solving equation (4.10) for  $e(n)$  and taking into consideration (4.11), the gains  $k_s$  and  $k_e$  will be obtained from

$$k_s = \frac{E\{y(n) \cdot u_s(n)\}}{E\{u_s(n) \cdot u_s(n)\}} \quad (4.12.a)$$

$$k_e = \frac{E\{y(n) \cdot u_e(n)\}}{E\{u_e(n) \cdot u_e(n)\}} \quad (4.12.b)$$

The same result would have been obtained if the quantization noise power  $P_e$

$$P_e = E\{e(n)^2\} = E\{(y(n) - k_s \cdot u_s(n) - k_e \cdot u_e(n))^2\} \quad (4.13)$$

had been minimized. Applying the linear Mean Square Estimation for the variable  $y(n)$ , by means of the linear estimator  $k_s \cdot u_s(n) + k_e \cdot u_e(n)$ , gives, according to the orthogonality principle, the estimation error  $e(n) = y(n) - k_s \cdot u_s(n) - k_e \cdot u_e(n)$ , to be orthogonal (i.e. uncorrelated) to the signals  $u_e(n)$  and  $u_s(n)$ .

Using the second assumption ( $u_e(n)$  is Gaussian), the quantization noise power  $P_e$  and the gains  $k_s$  and  $k_e$  can be numerically evaluated [3] for specific types of input signals (sine wave or dc). As a result, the signal and noise transfer functions will be written as follows:

$$STF(z) = \frac{k_s \cdot L_0(z)}{1 + k_s \cdot L_1(z)} \quad (4.14.a)$$

$$NTF(z) = \frac{1}{1 + k_e \cdot L_1(z)} \quad (4.14.b)$$

The above quantities, i.e.  $P_e$ ,  $k_s$ ,  $k_e$ , STF and NTF, are not constant but depend on the type and the power of the input signal [3]. This fact lends the name “quasi-linear” to the described model.

Finally, the evaluation of the  $SNR_{max}$  is carried out using (4.7) as in the case of the linear model. In this case, the required in-band quantization error, given by (4.5), depends on the input signal power. Drawing the SNR curve, the estimation for the  $SNR_{max}$  can be made and, consequently, a thorough assessment for the operation of the modulator using the quasi-linear model can be achieved. Nevertheless, the value of the signal amplitude for which instability occurs depends on the type of the signal.

### 4.3 NTF Characteristics

The design of a  $\Delta\Sigma$  modulator requires the selection of an appropriate filter  $L_1(z)$  or, according to (4.3), the corresponding NTF, so that the modulator would be stable and at the same time achieve a high SNR. For this purpose the NTF should possess the following characteristics:

- a. The first term  $h_0$  of the impulse response of the NTF, should always be equal to 1 [1]. This is necessary for the  $\Delta\Sigma$  modulator to be a causal system. The quantization error has to be delayed by at least one sampling period  $T_s$  before returning to the quantizer input.
- b. Suppose that the NTF is of the form

$$NTF(z^{-1}) = H(z^{-1}) = g \cdot \prod_{i=1}^n \frac{(1 - z_i \cdot z^{-1})}{(1 - p_i \cdot z^{-1})} \quad (4.15)$$

where  $g$  is a constant. Using Taylor series expansion around  $z^{-1}=0$ , for the first term of the impulse response we will have

$$h_0 = NTF(0) = g \quad (4.16)$$

Consequently,  $g=1$  and this has to be taken into consideration when an ordinary transfer function is to be selected from tabulated data.

- c. The signal power, which can be properly processed by the  $\Delta\Sigma$  modulator, decreases as the noise power increases. This fact easily proves to be valid, since the output sequence has always constant power  $\delta^2$  which is composed of the signal and the total noise power

$$P_x + P_{e-tot} = \delta^2 \quad (4.17)$$

If the quantization noise spectrum is white, its spectrum at the output of the  $\Delta\Sigma$  modulator will be

$$P_{e-tot} = \frac{P_e}{2\pi} \int_{-\pi}^{\pi} |NTF(e^{j\theta})|^2 d\theta = P_e \cdot A \quad (4.18)$$

where the quantity  $A$  is called the noise amplification factor.

- d. The quantization noise power within the signal band depends on the order of the NTF. It can be shown that, for an ideal high-pass NTF, this noise power is given approximately by [15]

$$P_{e-in} \approx \frac{P_e}{R} \cdot \frac{1}{e} \cdot A^{-(R-1)} \quad (4.19)$$

where  $e$  is the base of natural logarithms. Clearly  $P_{e-in}$  decreases rapidly with increasing  $A$ .

- e. The maximum achievable SNR is usually obtained for moderate signal amplitudes. Since the maximum SNR is given by

$$SNR_{max} = \frac{P_x}{P_{e-in}} = \frac{\delta^2 - P_{e-tot}}{P_{e-in}} \quad (4.20)$$

substituting from (4.19) into (4.20) gives

$$SNR_{\max} = e \cdot \frac{R}{P_e} \cdot A^{(R-1)} \cdot (\delta^2 - P_e \cdot A) \quad (4.21)$$

According to this last equation the maximum SNR is achieved for a high value of  $A$  and consequently for a high order NTF. However, this results in increasing the  $P_{e-tot}$  and accordingly (see eqn. (4.17)) in decreasing the tolerable  $P_x$  in the  $\Delta\Sigma$  sequence. Therefore, for encoding high values of the input signal, a compromise leading to lower SNR must be accepted.

- f. An NTF with minimum phase will lead to higher SNR. A non-minimum phase NTF has the factor  $g$  in (4.15) greater than 1. This results in an increase in the quantization noise power by a factor of  $g^2$ . Using equations (4.17) and (4.20) it becomes evident that the SNR decreases since in (4.20) the numerator decreases whereas the denominator increases. The non-minimum phase NTFs result in the so-called chaotic  $\Delta\Sigma$  encoders [1,13]. Chaotic encoders are more prone to instability and achieve lower maximum SNR. Although the problem of idling tones is not prominent in these circuits, since the nature of quantization noise is quite random, they are not used, due to their disadvantages that were previously mentioned.

#### 4.4 Stability of $\Delta\Sigma$ Modulators

$\Delta\Sigma$  modulators are basically non-linear systems and their stability cannot be analyzed by means of stability criteria employed in linear systems. As far as single-bit, single-stage  $\Delta\Sigma$  modulators are concerned, instability occurs for systems of order higher than one, and is detected from the behavior of internal signals. In this case the amplitude of the internal signal increases rapidly and oscillations take place at low frequencies (limit cycles) [1,4,5]. Using various kinds of simulation, it was concluded that instability oc-

cers when the amplitude or the frequency of the input signal exceeds a specific value. The unstable encoder cannot usually return to its stable operation even if the reasons for instability does not exist any longer.

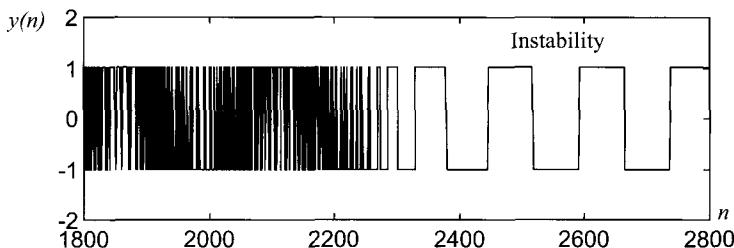


Fig. 4.4 Time evolution of the modulator output  $y(n)$  for sinusoidal input signal in case of unstable modulator.

An example of an unstable modulator is depicted in Fig. 4.4, by means of simulating a third-order  $\Delta\Sigma$  modulator for sinusoidal input signal with 0.6 of the full scale amplitude. The input to the quantizer  $u(n)$  is, till a time instant, varying inside the expected amplitude region. Nevertheless, owing to the nonlinear nature of the modulator as well as to signal dynamics, suddenly,  $u(n)$  increases rapidly even if the input signal is withdrawn. Moreover, as shown in Fig. 4.4, the modulator output does not alternate rapidly between  $-\delta$  and  $+\delta$ , as it does when the modulator is properly functioning.

Various stabilization techniques for restoring the normal operation of the modulator have been proposed [6-8] and are reviewed in Chapter 7. Each time a stabilization technique is applied, the SNR is significantly reduced for a reasonable time period until the modulator totally recovers. This means that a stabilization technique is a necessary but not an optimal solution to avoid instability problems. The most appropriate approach is to design the  $\Delta\Sigma$  modulator so that it will never become unstable.

## 4.5 Stability Criteria

Various stability criteria for selecting the appropriate NTF and thus assessing the functionality of a  $\Delta\Sigma$  modulator can be found in the literature [1]. Three of them are the following:

- a. The sum of the absolute value of the terms in the impulse response of the NTF is bounded [9,10], i.e.

$$S_{|h|} \equiv \sum_{i=0}^{\infty} |h_i| = 3 - x_{\max} \quad (4.22)$$

where  $x_{\max}$  is the maximum signal amplitude for which the modulator remains stable. This criterion is quite strict giving stable circuits, which, however, achieve low  $SNR_{\max}$  though. The criterion is usually applied in the form  $S_{|h|} < c$ , where  $c$  is around 3.5.

- b. The mean-squared value of the magnitude response of the NTF (noise amplification factor)

$$A \equiv \frac{1}{2\pi} \int_{-\pi}^{\pi} |NTF(e^{j\theta})|^2 d\theta < c \quad (4.23)$$

where  $c \approx 2.5$  [9,11]

- c. The maximum value of the frequency response of the NTF has to be smaller than  $c \approx 2$  [9,12], i.e.

$$M \equiv \max \{|NTF(z)|\} < c \quad (4.24)$$

The empirical nature of these criteria is evident, while the value of  $c$  has been estimated after extensive simulations. For a specific  $\Delta\Sigma$  modulator the most suitable NTF is obtained from the family of functions [1,3]

$$NTF_k(z) = \frac{NTF_{initial}(z)}{k + (1-k) \cdot NTF_{initial}(z)} \quad (4.25)$$

This NTF should accomplish the above criteria for certain values of  $k$ . However, the exact amplitude of the input signal beyond which the modulator becomes unstable cannot be determined.

Unfortunately, the estimation of  $x_{\max}$  based on the first criterion is accurate only in a few cases, as for example is the case of

the first-order  $\Delta\Sigma$  modulator. For the rest of cases, the criterion is very strict and the value of  $x_{max}$  is underestimated or is found to be negative, which means that the modulator is unstable. This criterion can be employed for estimating  $x_{max}$  for a known NTF or when designing NTFs posing restrictions on  $S_{|h|}$ . In this last case the criterion is applied using (4.22). In general, when increasing  $S_{|h|}$ ,  $x_{max}$  decreases and the modulator becomes more easily unstable. Consequently,  $S_{|h|}$  must be held small.

As far as the criterion for the value of  $A$  is concerned, for a stable modulator the available power of the signal in the  $\Delta\Sigma$  sequence should not be negligible, since

$$P_x = \delta^2 - P_{e-tot} = \delta^2 - P_e \cdot A > 0$$

requiring

$$A < \frac{\delta^2}{P_e} \quad (4.26)$$

When using the linear model, where  $P_e = \delta^2/3$ ,  $A$  should be smaller than 3, whilst when using the quasi-linear model,  $A$  should be smaller than 2.75, since  $P_e = \delta^2(1 - 2/\pi)$  [1,3] in the latter case. Thus, the employment of the models explains the limitations in the value of the noise amplification factor  $A$ . Using the same justification, the criterion that imposes limitations on the quantity  $M$ , can be explained, since  $M$  is indirectly related to  $A$ . Both quantities  $A$  and  $M$  provide a measure for the available input signal power contained in the  $\Delta\Sigma$  sequence and, consequently, describe the stability of the modulators.

The above discussion about the quantity  $A$ , where

$$A = \frac{1}{2\pi} \int_{-\pi}^{\pi} |NTF(e^{j\theta})|^2 d\theta = \sum_{i=0}^{\infty} h_i^2 = 1 + \sum_{i=1}^{\infty} h_i^2 = 1 + A_1 \quad (4.27)$$

reveals two necessary but contradicting conditions. The quantity  $A$  must be large enough for a high SNR to be achieved. However, the quantity  $S_{|h|}$  has to be small. Consequently, while the increase in  $A$  results in increasing the terms of the impulse response, the decrease in  $S_{|h|}$  demands their decrease. For a given value of  $A$ ,

minimization of  $S_{|h|}$  occurs when only one term of the impulse response is non-zero and all the others are zero [15] (the first term  $h_0$  is always 1). For the first-order  $\Delta\Sigma$  modulator the second term is  $h_1 = -1$ , whereas for higher order modulators the value of  $|h_i|$  can be selected to be much larger than the values of the rest of the terms. In [14,15] it is shown that an alternative stability criterion could be the following:

$$s_{h^2} \equiv \sum_{i=2}^{\infty} h_i^2 < c \quad (4.28)$$

The value of  $c$ , after extensive simulations, is found to be 0.07.

## 4.6 Noise Transfer Function Determination

A desired  $\Delta\Sigma$  modulator should be stable for high values of the input signal and simultaneously achieve maximum SNR. This corresponds to forming the NTF in such a way that the best trade off between maximum achievable input  $x_{max}$  and maximum SNR is reached. Various NTFs can be obtained depending on the requirements and specifications of the modulator. However, they must be high-pass. It is possible to be drawn from tabulated data containing Butterworth, Chebyshev or Inverse Chebyshev functions. The position of poles and zeros of such a function could be as depicted in Fig. 4.5. For high attenuation in the signal band the zeros must be located inside the signal band, while the poles must lie far away from the zeros. For high oversampling ratios the zeros are close to  $z=1$ . In this case the magnitude of the NTF in the signal band is mainly determined by the zeros according to

$$\left|NTF(e^{j\theta})\right|^2 \cong \frac{\left|NUM(e^{j\theta})\right|^2}{\prod_{i=1}^N |1-p_i|^2} \quad (4.29)$$

For small  $\theta$  the denominator can be considered constant and thus the power of the quantization error in the signal band will be proportional to the integral of this quantity. Consequently, the prob-

lem of determining an optimal NTF is related to finding first the best position of zeros and then the position of poles.

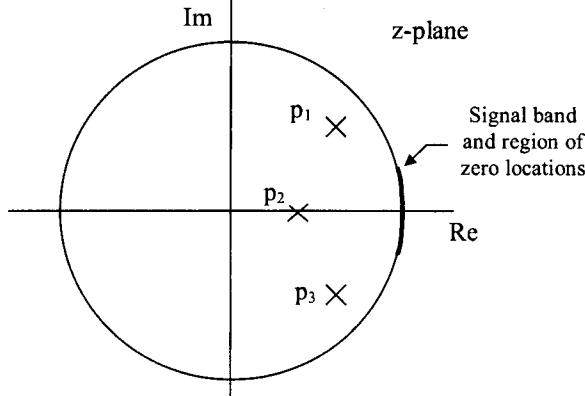


Fig. 4.5 Location of poles and zeros in the unit circle.

#### 4.6.1 Optimizing the position of zeros

The position of zeros should be such that the power of the quantization error in the signal band is minimized [9]. This means that the following integral should be minimized:

$$I_1 = \int_0^{\pi} |NUM(e^{j\theta})|^2 d\theta \quad (4.30)$$

Since the zeros lie on the unit circle, they can be written in the form of  $e^{j\phi_i}$ . Then for  $z = e^{j\theta}$  we will have for even  $N$

$$NUM(z^{-1}) = \prod_{i=1}^{\frac{N}{2}} (1 - e^{-j(\phi_i + \theta)})(1 - e^{j(\phi_i - \theta)}) \approx \prod_{i=1}^{\frac{N}{2}} (\phi_i^2 - \theta^2) = \sum_{i=0}^{\frac{N}{2}} b_i \theta^{2i} \quad (4.31.a)$$

while for odd  $N$

$$NUM(z^{-1}) = (1 - e^{-j\theta}) \prod_{i=1}^{\frac{N-1}{2}} (1 - e^{-j(\phi_i + \theta)})(1 - e^{j(\phi_i - \theta)}) \approx j\theta \sum_{i=0}^{\frac{N-1}{2}} b_i \theta^{2i} \quad (4.31.b)$$

These are valid to a good approximation since  $\theta, \phi_i \leq \pi/R \ll 1$ . It is preferable to evaluate the coefficients  $b_i$  since the frequencies  $\phi_i$  can be determined explicitly only for a NTF of order smaller than 5 [9]. In order to avoid dependence on R, the normalized frequencies will be evaluated (for  $\pi/R = 1$ ).

For  $N$  even we have  $b_{N/2} = (-1)^{N/2}$ . The evaluation of the other coefficients  $b_i$  proceeds as follows: First we insert the expression for  $\text{NUM}(z^{-1})$  from eq. (4.31.a) into eq. (4.30) and evaluate the integral. Then, taking the partial derivatives with respect to  $b_i$ 's and setting them equal to zero leads to the following set of equations:

$$\sum_{i=0}^{\frac{N}{2}-1} \frac{b_i}{2k+2i+1} = -\frac{(-1)^{N/2}}{N+2k+1}, \quad k=0,1,\dots,\frac{N}{2}-1 \quad (4.32.a)$$

A similar procedure followed for odd  $N$  results in the following set:

$$\sum_{i=0}^{\frac{N-1}{2}-1} \frac{b_i}{2k+2i+3} = -\frac{(-1)^{(N-1)/2}}{N+2k+2}, \quad k=0,1,\dots,\frac{N-1}{2}-1 \quad (4.32.b)$$

The desired frequencies  $\phi_i$  are the roots of the polynomials (4.31). The normalized coefficients as well as the normalized roots determined in this way are given in Table 4.1 for orders 2 up to 5. These polynomials will be subsequently referred to as optimal. The corresponding quantities for the equiripple Chebyshev polynomials are presented in Table 4.2. Comparing the two Tables the difference in the positions of the zeros, between optimal and Chebyshev polynomials, can be determined. The normalized frequencies will have to be multiplied by  $\pi/R$ , for denormalization purposes.

In Table 4.3 the performance achieved by the optimal and Chebyshev polynomials when compared to the  $\omega^N$  polynomials, which have all their roots at zero, is presented. This table shows that it is preferable to move the zeros away from dc. Furthermore, the optimal polynomials outperform those of Chebyshev by

approximately 1.2 dB. This fact is expected since the Chebyshev polynomials are equiripple and they do not lead to minimum mean square quantization error in the signal band.

Table 4.1 Coefficients and roots of optimal polynomials for orders 2 to 5.

N	Coefficients				Roots	
	$b_0$	$b_1$	$b_2$	$\varphi_1$	$\varphi_2$	$\varphi_3$
2	-0.3333	1.0		$\pm 0.5774$		
3	-0.6	1.0		0.0	$\pm 0.7746$	
4	0.0857	-0.8571	1	$\pm 0.34$	$\pm 0.8611$	
5	0.2381	-1.1111	1	0.0	$\pm 0.5385$	$\pm 0.9062$

Table 4.2 Coefficients and roots of Chebyshev polynomials for orders 2 to 5.

N	Coefficients				Roots	
	$b_0$	$b_1$	$b_2$	$\varphi_1$	$\varphi_2$	$\varphi_3$
2	-0.5	1.0		$\pm 0.7072$		
3	-0.75	1.0		0.0	$\pm 0.8661$	
4	0.125	-1.0	1.0	$\pm 0.3827$	$\pm 0.9239$	
5	0.3125	-1.25	1.0	0.0	$\pm 0.5878$	$\pm 0.9511$

Table 4.3 Gain in power compared to  $\omega^N$  polynomials.

N	Gain (dB)	
	Optimal	Chebyshev
2	3.5	2.3
3	7.9	6.7
4	12.8	11.6
5	17.9	16.7

#### 4.6.2 Optimizing the position of poles

The quantization noise power in the signal band is reduced as the distance of the poles from the origin of the unit circle increases. However, this results to an increase in the NTF magnitude at high frequencies, and consequently, in an increase in the total quantization noise power  $P_{e-tot}$ . Thus the power of the  $\Delta\Sigma$  sequence available to the signal is reduced and the modulator becomes prone to instability. The selection of an existing function to model the NTF predetermines the position of poles on the complex plane according to the specific function characteristics.

The selection of poles and thus the determination of the NTF using the Inverse Chebyshev functions requires the multiplication of the NTF by  $g$ , so that it takes the form of equation (4.15) with  $g=h_0=1$  (for pass-band gain equal to 1,  $g$  must be less than 1). For Inverse Chebyshev NTFs the position of poles can be uniquely determined by the value of parameter  $A$  or  $M$ , in case the latter should be constrained (Eqs. (4.23) and (4.24)).

The resulting NTFs have their zeros in the signal band as given in Table 4.2. The poles are given in Table 4.4 and depend on the value of  $M$ . The parameter  $BE$  in the same Table expresses the degree of expected stability in increasing order. This means that modulator  $E4$  is more stable than  $E1$ , which implies that it remains stable for higher values of the input signal. For small values of  $M$  the poles are close to the origin of the unit circle.

Another way for selecting the optimum position of poles is to search the whole region inside the unit circle to find the position of poles leading to maximum SNR [14-15]. For this purpose, the maximum value of  $s_{h^2}$  is constrained and the whole region is examined for the poles to minimize the integral in (4.5) and, consequently, maximize the SNR. In Table 4.5 the poles of the NTF are given having been obtained from the above approach. The zeros can be those of Table 4.1 or Table 4.2.

Table 4.4 Poles of the Inverse Chebyshev NTFs for various values of  $M$ .

Order	$BE$	$M$ (dB)	Poles of the NTF
3	E1	5	0.5427, 0.6584 $\pm j0.3720$
	E2	4	0.6191, 0.7314 $\pm j0.3173$
	E3	3	0.7055, 0.8062 $\pm j0.2511$
	E4	2	0.7906, 0.8716 $\pm j0.1830$
4	E1	5	0.7839 $\pm j0.3333$ , 0.6570 $\pm j0.1157$
	E2	4	0.8286 $\pm j0.2834$ , 0.7120 $\pm j0.1010$
	E3	3	0.8769 $\pm j0.2228$ , 0.7777 $\pm j0.0818$
	E4	2	0.9222 $\pm j0.1566$ , 0.8480 $\pm j0.0596$
5	E1	5	0.6997, 0.8486 $\pm j0.2973$ , 0.7353 $\pm j0.1592$
	E2	4	0.7494, 0.8823 $\pm j0.2491$ , 0.7819 $\pm j0.1364$
	E3	3	0.8091, 0.9177 $\pm j0.1914$ , 0.8363 $\pm j0.1078$
	E4	2	0.8657, 0.9468 $\pm j0.1377$ , 0.8866 $\pm j0.0797$

Table 4.5 Poles resulting by means of search over the entire region of the unit circle for maximizing SNR.

Order	BE	Constrain in $s_{h^2}$	Poles of the NTF
3	E1	0.064	-0.0528, 0.7951 $\pm j0.3622$
	E2	0.032	-0.0382, 0.8354 $\pm j0.3024$
	E3	0.016	-0.0309, 0.8695 $\pm j0.2500$
	E4	0.008	-0.0285, 0.9049 $\pm j0.2073$
4	E1	0.064	0.0611, 0.6044, 0.9252 $\pm j0.2800$
	E2	0.032	0.0365, 0.7044, 0.9361 $\pm j0.2354$
	E3	0.016	0.0320, 0.7444, 0.9561 $\pm j0.1900$
	E4	0.008	0.0156, 0.8144, 0.9606 $\pm j0.1600$
5	E1	0.064	-0.0457, 0.9699 $\pm j0.1934$ , 0.8194 $\pm j0.3006$
	E2	0.032	-0.0157, 0.9770 $\pm j0.1634$ , 0.8423 $\pm j0.2420$
	E3	0.016	-0.0157, 0.9770 $\pm j0.1420$ , 0.8851 $\pm j0.1834$
	E4	0.008	-0.0143, 0.9871 $\pm j0.1100$ , 0.9029 $\pm j0.1671$

## 4.7 $\Delta\Sigma$ Modulator Assessment

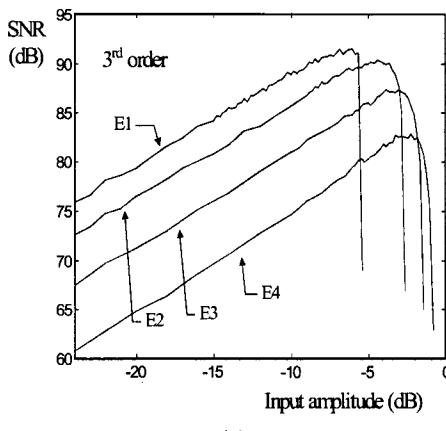
The performance assessment of a  $\Delta\Sigma$  modulator is based on the maximum SNR that can be achieved, as well as on the maximum input signal for which the modulator remains stable. This can be derived from the SNR plot against the input signal amplitude. The spectrum of the  $\Delta\Sigma$  sequence can be used to evaluate both the signal and the quantization noise power, and thus the SNR. In Fig. 4.6, the SNR is given as a function of the input signal applied to third-order  $\Delta\Sigma$  modulators for two different methods for determining the NTF, namely the Inverse Chebyshev and the search for maximum SNR. In Table 4.6 their characteristics are given for various  $\Delta\Sigma$  modulators resulting from simulations. Specifically, the values of the maximum SNR and the maximum signal amplitude for which the modulator remains stable are presented. According to these data, the above-mentioned methods for selecting the NTF are found to be effective. However, the second method based on finding the poles for maximum SNR may lead to  $\Delta\Sigma$  modulators with better performance as can be easily deduced from Table 4.6.

Table 4.6  $\Delta\Sigma$  modulator characteristics for Inverse Chebyshev and maximum SNR type NTFs.

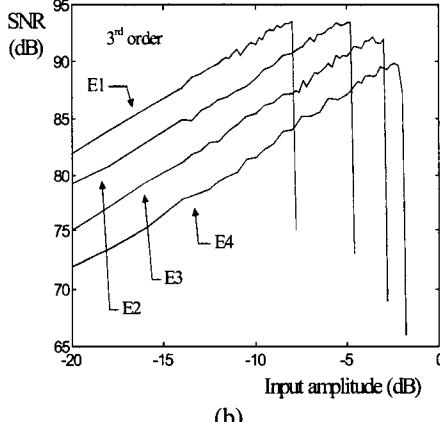
Order	BE	Inverse Chebyshev		Maximum SNR	
		SNR <sub>max</sub> (dB)	x <sub>max</sub> (FS)	SNR <sub>max</sub> (dB)	x <sub>max</sub> (FS)
3	E1	91.5	0.537	93.5	0.407
	E2	90.4	0.741	93.4	0.589
	E3	87.5	0.851	92.2	0.646
	E4	82.9	0.912	89.8	0.813
4	E1	106.3	0.316	107.6	0.295
	E2	105.8	0.575	107.9	0.562
	E3	101.1	0.741	105.1	0.692
	E4	92.9	0.870	101.1	0.776
5	E1	117.8	0.186	122.8	0.355
	E2	118.4	0.501	121.5	0.575
	E3	111.9	0.676	116.9	0.708
	E4	101.8	0.832	111.4	0.794

## 4.8 Summary

In this Chapter the modeling and design of single-bit single-stage  $\Delta\Sigma$  modulators was analyzed. The design of modulators with satisfactory characteristics is based on the selection of NTFs with specific properties. Among them the first term of its impulse response should be significant while the rest negligible.



(a)



(b)

Fig. 4.6 SNR plot as a function of input signal amplitude for Inverse Chebyshev (a) and maximum SNR modulators (b).

The availability of an easily applied stability criterion and a suitable quantization noise model, enables the designer to find an

optimum NTF for the  $\Delta\Sigma$  modulator. The procedure is carried out in two stages. First, the positions of zeros are evaluated analytically so that the minimization of the quantization error power in the signal band is achieved. Then, the optimization of the pole locations is carried out using numerical methods, in order to satisfy a stability criterion and achieve maximum SNR for the specific model simultaneously. Accordingly, the problem of designing single-bit single-stage  $\Delta\Sigma$  modulators is tackled and their behavior and performance is examined. However, other issues concerning the spectral properties of the quantization error for example, should also be considered.

## Problems

- 4.1 Using MATLAB find the poles and zeros of an Inverse Chebyshev high-pass NTF, like one of those given in Table 4.4 (specific order and parameter  $M$ ). Determine the in-band attenuation so that the desired  $M$  is achieved. Draw the plot of  $M$  as a function of the in-band attenuation. Is it verified, this way, that decreasing  $P_{e,in}$  results in increasing  $P_{e,tot}$ ?
- 4.2 Consider as  $NTF_{initial}$  one of the Inverse Chebyshev NTFs given in Table 4.4. Write a program in MATLAB for determining, using eq. (4.25), the corresponding NTFs, for various values of  $k$ . Find the values of  $k$  for which minimization of  $S_{|h|}$ ,  $A$  and  $M$  is achieved. Determine also, the minimum values for  $S_{|h|}$ ,  $A$  and  $M$ . Does  $NTF_{initial}$  obey some of the criteria given in eqs. (4.22), (4.23) and (4.24)?
- 4.3 Using SIMULINK design a  $\Delta\Sigma$  modulator. Make use of Fig. 4.1 letting  $\delta=1$  and in place of  $L_0(z)$  and  $L_1(z)$  the structure in Fig. 3.6. Determine the values of  $G(z)$  and  $L(z)$  so that the STF is an all-pass function with the NTF being one of those in Table 4.4.
  - a. For various values of the amplitude of the input signal (always  $<1$ ) and various frequencies, draw the output pulse sequence as well as the input to the quantizer.

- b. Determine the spectrum of the output using a window of 65536 samples and comment on its noise shaping characteristics.
  - c. Evaluate the SNR for various amplitudes of the input signal and draw the corresponding diagram.
  - d. Using the SNR plot, determine the input signal amplitude for which instability occurs. For this amplitude, simulate the modulator and monitor the input and the output of the quantizer.
  - e. Reduce the accuracy of filter coefficients and examine the change in Noise Shaping.
  - f. Use two separate filters  $L_0(z)$  and  $L_1(z)$  for the structure in Fig. 4.1 (instead of  $G(z)$  and  $L(z)$ ) and simulate the resulting modulator. Monitor the signals at their output. Could this structure be suitable for implementation?
- 4.4 Simulate a  $\Delta\Sigma$  modulator whose NTF is one of those in Table 4.4. Find the output spectrum. Can the observed noise shaping characteristic be approximated by an inverse Chebyshev high-pass transfer function? Justify your answer.
- 4.5 Simulate a  $\Delta\Sigma$  modulator whose NTF is one of the Inverse Chebyshev functions in Table 4.4. Repeat with the same NTF having the same poles and zeros at dc. Verify that the gain in SNR is as given in Table 4.3.

## References

- [1] S.R. Norsworthy, R. Schreier and G.C. Temes, *Delta-Sigma Data Converters, Theory, Design and Simulation*, New York, IEEE Press, 1997.
- [2] C. Tonmazou, N. Battersby, S. Porta, *Circuits and Systems Tutorials*, IEEE ISCAS, 1994.
- [3] S.H. Ardalan, J.J. Paulos, 'An Analysis of Nonlinear Behavior in Delta-Sigma Modulators', *IEEE Trans. on Circuits and Systems*, Vol. CAS-34, No. 6, pp. 593-603, June 1987.
- [4] S. Hein, A. Zakhor, 'On the stability of Sigma Delta Modulators', *IEEE Trans. on Signal Processing*, Vol. 41, No. 7, pp. 2322-2348, July 1993.
- [5] R.T. Baird, T.S. Fiez, 'Stability Analysis of High-Order Delta-Sigma Modulation for ADC's', *IEEE Trans. on Circuits and Systems-II: Analog and Digital Signal Processing*, Vol. 41, No. 1, pp. 59-62, January 1994.
- [6] T. Okamoto, Y. Maruyama, A. Yukawa, 'A Stable High-Order Delta-Sigma Modulator with an FIR Spectrum Distributor', *IEEE Journal of Solid-State Circuits*, Vol. 28, No. 7, pp. 730-735, July 1993.
- [7] C. Dunn, M. Sandler, 'Use of Clipping in Sigma-Delta modulators', *The Institution of Electrical Engineers*, 1994.
- [8] S.M. Moussavi, B.H. Leung, 'High-Order Single-Bit Single-Stage Oversampling A/D Converter Stabilized with Local feedback Loops', *IEEE Trans. on Circuits and Systems-II: Analog and Digital Signal Processing*, Vol. 41, No. 1, pp. 19-25, January 1994.
- [9] R. Schreier, 'An Empirical Study of High-Order Single-Bit Delta-Sigma Modulators', *IEEE Trans. on Circuits and Systems-II: Analog and Digital Signal Processing*, Vol. 40, No. 8, pp. 461-466, August 1993.
- [10] D. Anastassiou, 'Error Diffusion Coding for A/D Conversion', *IEEE Trans. on Circuits and Systems*, Vol. 36, No. 9, pp. 1175-1186, September 1989.
- [11] B. P. Agrawal, K. Shenoi, "Design Methodology for  $\Sigma\Delta M$ ", *IEEE Trans. on Communications*, Vol. 31, No. 3, pp. 360-369, March 1983.
- [12] K.C.H. Chao, S. Nadeem, W.L. Lee, C.G. Sodini, 'A Higher Order Topology for Interpolative Modulators for Oversampling A/D Converters', *IEEE Trans. on Circuits and Systems*, Vol. 37, No. 3, pp. 309-318, March 1990.
- [13] L. Risbo, 'On the Design of Tone-Free  $\Sigma\Delta$  Modulators', *IEEE Trans. on Circuits and Systems-II: Analog and Digital Signal Processing*, Vol. 42, No. 1, pp. 52-55, August 1995.
- [14] G.I. Bourdopoulos, ' $\Delta\Sigma$  modulators, Modeling, Design, Implementation, Applications', Ph.D. Thesis (in Greek), Patras, Greece, 1999.
- [15] G.I. Bourdopoulos, A.G. Pneumatakis and T.L. Deliyannis, 'Optimal NTFs for single bit  $\Delta\Sigma$  Modulators', DSP 2002, 14<sup>th</sup> International Conference on Digital Signal Processing, Vol. 2, PP. 877-880, July 1-3, Sandorini, 2002.

## **Chapter 5**

# **Implementation of $\Delta\Sigma$ Modulators**

### **5.1 Introduction**

In chapters 3 and 4 the  $\Delta\Sigma$  Modulator was examined in the signal flow diagram level. The study referred to modeling, design, the characteristics of various structures of  $\Delta\Sigma$  modulators, their efficiency and their stability. In the present chapter we are looking at the implementation of  $\Delta\Sigma$  modulators with emphasis on those used in the design of ADC. We start with the presentation of each of the stages in the block diagram of the  $\Delta\Sigma$  modulator and give various basic circuits useful in the implementation. Next the actual design is presented at the circuit level of first-, second-, third-order low-pass as well as a fourth-order band-pass  $\Delta\Sigma$  single-bit modulators. Also simulation and experimental results with reference to SNR and PSD of these actual circuits are given. The experimental method of obtaining these results is also described. It is hoped that the reader, having gone through this chapter, will be able to build and test his own  $\Delta\Sigma$  modulator at the laboratory.

### **5.2 Basic Blocks of a $\Delta\Sigma$ Modulator for Analog-to-Digital Conversion**

The single-stage  $\Delta\Sigma$  modulator, whether single-bit or multi-bit, is shown in block diagram form in Fig. 5.1. It consists of the following blocks:

- a) The loop filter (LF)

b) The local ADC

c) The local DAC

The loop filter is analog and can be active continuous-time (CT) or discrete-time (DT). The combination of the local ADC and DAC is equivalent to a sampler followed by a quantizer. Sampling is necessary in order to produce the  $\Delta\Sigma$  sequence at the output of the modulator, while with the presence of the DAC the required states of the quantizer are produced in order to be returned to the input of the analog filter.

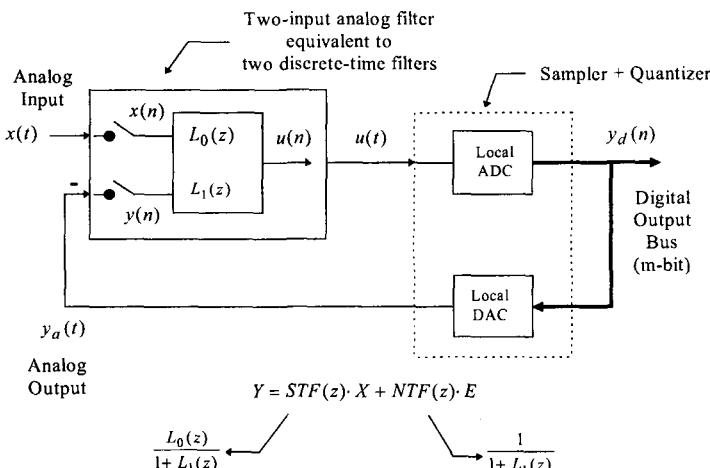


Fig. 5.1 General block diagram of a  $\Delta\Sigma$  modulator for analog-to-digital conversion.

Signals  $x(t)$ ,  $y_a(t)$ ,  $u(t)$  and  $y_d(n)$  are real appearing respectively at the analog input of the modulator, at the output of the local DAC, at the input and output of the local ADC. The first three are continuous time waveforms and can be observed using a CRO, while  $y_d(n)$  is a m-bit digital signal. The signals  $x(n)$ ,  $u(n)$  and  $y(n)$  exist in the discrete-time model of the  $\Delta\Sigma$  modulator, which describes it in the signal flow-level. These may be obtained by sampling ideally the respective continuous-time signals  $x(t)$ ,  $u(t)$  and  $y_a(t)$  at the proper instances assuming ideal components.

When the modulator is single-bit, the local ADC is also one bit, easily implemented by a comparator followed by a D flip-flop, while the

local DAC is also 1-bit. Care should be taken that the loop is not delayless. In the following sections we examine the implementation of each of these blocks in some detail.

### 5.2.1 *The loop filter*

The loop filter is employed in order to obtain the required NTF, which should be high-pass (for low-pass  $\Delta\Sigma$  modulators), as well as the STF, which is desired to be all-pass. Usually  $L_0(z)$  and  $L_1(z)$  are the same function  $L(z)$ .

This filter is analog and can be implemented either in a continuous-time or in a discrete-time active form (switched-capacitor (SC), or switched-current (SI)). After the NTF has been selected, the transfer function of this filter  $L_1(z)$  can be obtained from Eq. (4.3) to be

$$L_1(z) = \frac{1 - NTF(z)}{NTF(z)} \quad (5.1)$$

As has been shown in Chapter 3, in order to realize  $L_1(z)$ , it is customary in  $\Delta\Sigma$  modulation work, to employ methods that are based on the use of integrators in a negative feedback loop. The main reason for this is the low sensitivity of these structures, since multiple negative feedback is applied to a cascade connection of integrators. In Figs. 5.2a and b the inverting and the non-inverting SC integrators are shown respectively [1], together with their respective transfer functions, assuming ideal operational amplifiers. Similarly, the continuous-time active RC integrator [2], shown in Fig. 5.2c, is also used as the basic component in implementing the loop filter. In the latter, of course, sampling is taking place after the signal has passed through it. This is not required in the case of employing SC techniques, because sampling is accomplished by the SC filter itself.

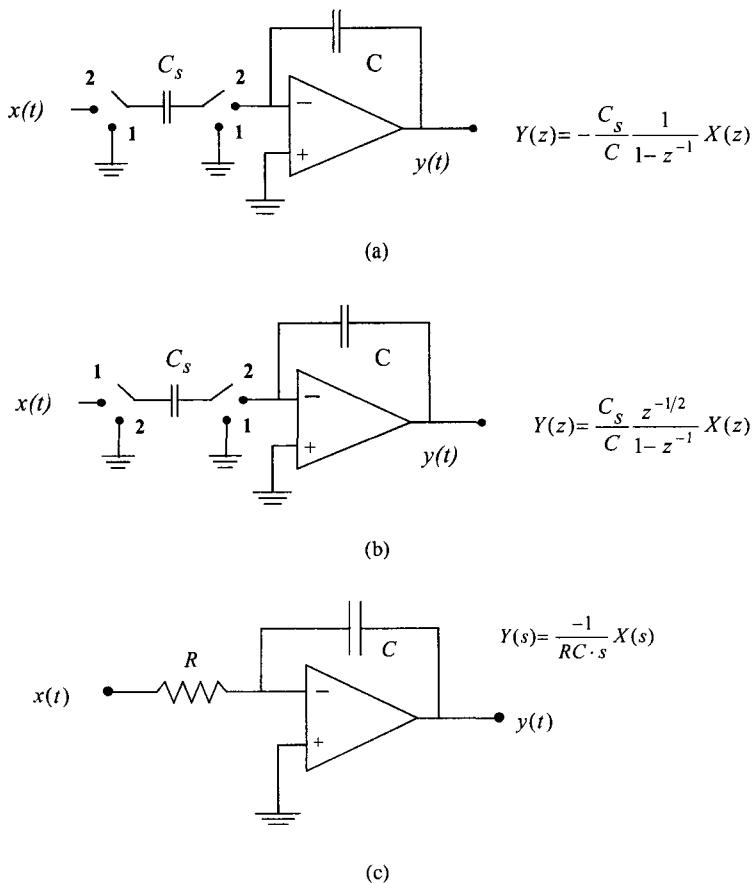


Fig. 5.2 a. Inverting and b. non-inverting single-ended SC integrators and c. the CT integrator.

In Fig. 5.3 fully-differential SC and CT integrators are shown. These are very useful when a high performance  $\Delta\Sigma$  modulator is required at higher frequencies. For operation at even higher sampling rates employment of current-mode continuous-time integrators will be required.

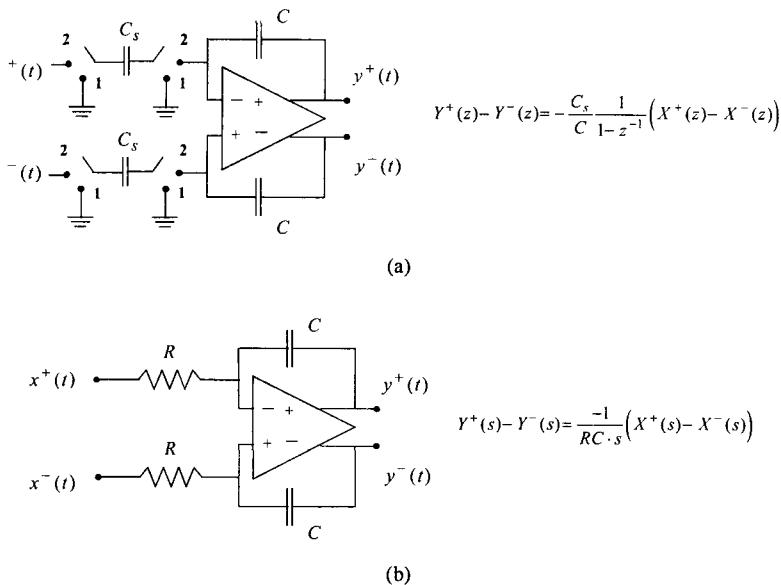
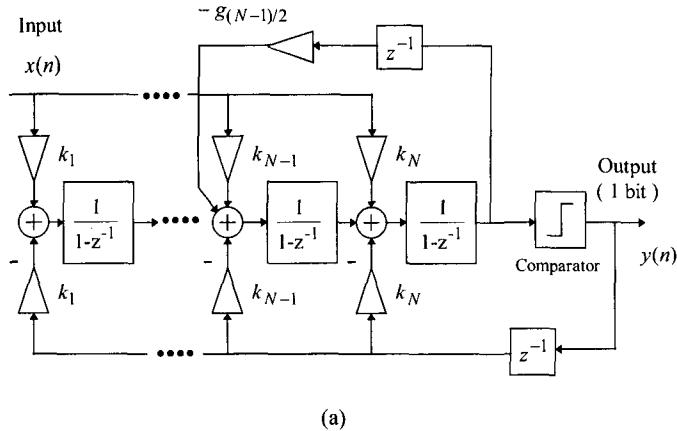


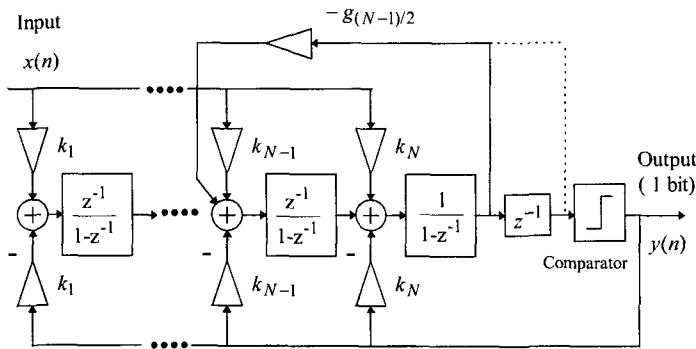
Fig. 5.3 Fully-differential a. SC and b. CT integrators.

As has been mentioned already, the order of the  $\Delta\Sigma$  modulator is identical to that of the loop filter. Thus a first-order  $\Delta\Sigma$  modulator will employ a simple integrator of either type shown in Figs. 5.2 and 5.3, which is a first-order circuit. Higher order modulators will employ loop filters of corresponding orders. These filters will normally consist of the cascade connection of integrators with distributed feedback applied to the cascade. Local negative feedback in two consecutive integrators may also be applied in order to form resonators with high-selectivity. This is clearly shown in Fig. 5.4, which is the general structure and a more detailed repetition of Fig. 3.7b. In case the loop filter is implemented using SC circuits, integrators with or without delay can be used. In the first case the circuit can operate up to higher frequencies as there is no need for multiple opamp settling during the sampling period. In Fig. 5.4a, where the integrators have no delay, the necessary delay in the negative feedback network of the resonator, can be implemented by properly phasing the switches. In Fig. 5.4b, correct phasing of the switches is also

necessary to avoid the double delay. However, in order to simplify the circuit and to avoid the required double opamp settling, the resonator can be implemented as indicated by the broken line [3] (see problem 5.1).



(a)



(b)

Fig. 5.4 Structure of the loop filter as a Chain of Integrators with Distributed Feedback and local resonator feedback (CIDF) a. using delay-less integrators, b. using integrators with delay.

### 5.2.2 Local ADC and DAC

In the single-bit  $\Delta\Sigma$  Modulator, both the local ADC and the DAC are 1-bit. In the multi-bit modulators these should be multibit.

The single-bit local ADC is implemented using a comparator followed by an edge-triggered D Flip-Flop. The frequency of the clock is the oversampling frequency, while the comparator is implemented using an operational amplifier on open-loop having as high a slew rate (SR) as possible. For single-ended structures one of the inputs is grounded, while for fully differential structures, its differential input is fed from the differential output of the last integrator of the loop filter. The D Flip-Flop, operating at the oversampling rate performs the sampling of the comparator output signal. This combination of the comparator and the flip-flop is equivalent to first sampling the signal  $u(t)$  using a sample-and-hold circuit which is then followed by the comparator action. The output  $\Delta\Sigma$  sequence is obtained from the Q output of the D Flip-Flop.

Next the output sequence from the D Flip-Flop enters the local DAC and controls the generation of the analog states at its output. In the single-bit case the simplest circuitry of the DAC consists of a double throw switch connecting its output either to  $+V_{ref}$  or to  $-V_{ref}$  thus closing the loop of the  $\Delta\Sigma$  modulator. In this case the DAC output pulses will be nearly ideal square waves as shown in Fig. 5.5.

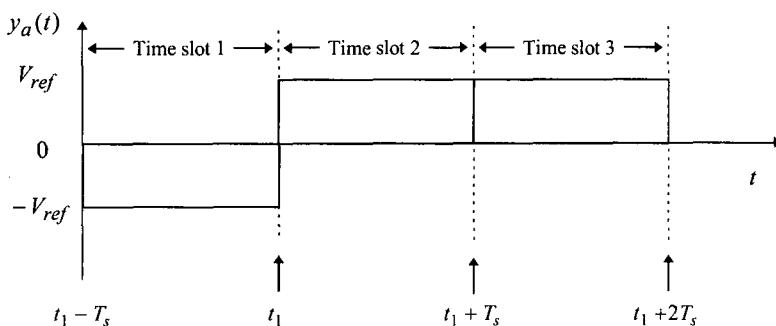


Fig. 5.5 Suitable output pulses from the local DAC in the case of a single-bit  $\Delta\Sigma$  modulator.

In continuous-time  $\Delta\Sigma$  modulators [3,4], to avoid problems created by non-equal rise and fall times of the DAC output pulses (see next chapter), the latter may have the form shown in Fig. 5.6 (Return to zero (RZ) waveform). A circuit generating this type of pulses will be given in the next chapter.

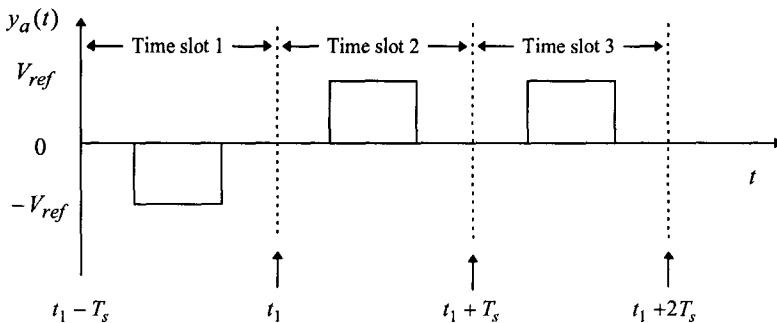


Fig. 5.6 DAC output pulses to reduce the effect of non-equal rise and fall times.

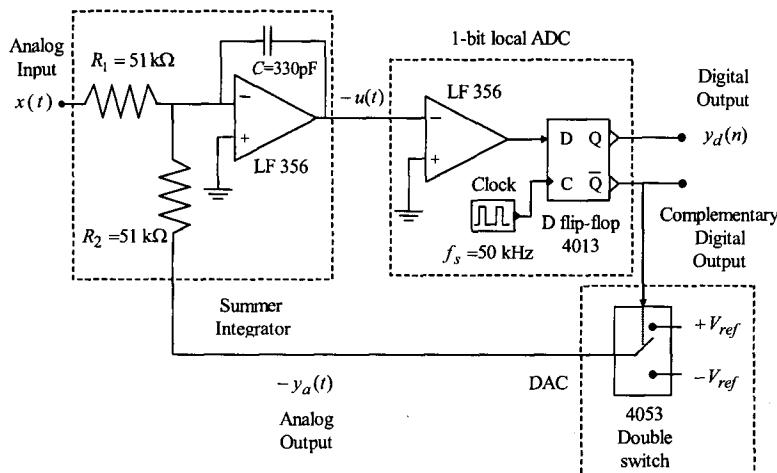


Fig. 5.7 First-order, single-bit  $\Delta\Sigma$  modulator using a CT integrator.

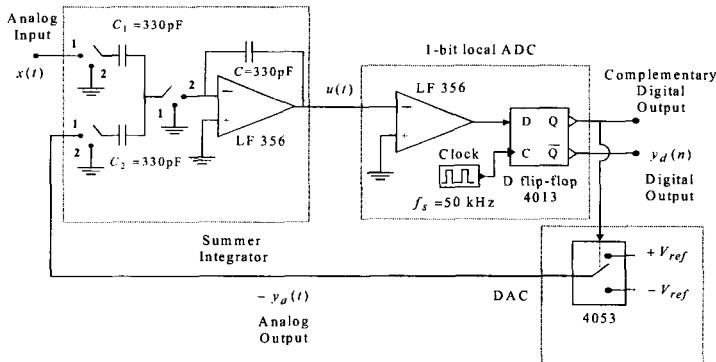


Fig. 5.8 First-order, single-bit  $\Delta\Sigma$  modulator using a SC integrator.

In Figs. 5.7 and 5.8 the implementation of a first-order  $\Delta\Sigma$  modulator is shown. In the first case the loop filter is a continuous-time integrator, while in the second a switched-capacitor integrator. Details of the way the component values have been calculated are given in Sec. 5.3.

In multi-bit  $\Delta\Sigma$  modulators the local ADC is a Flash ADC the output signal  $y_d(n)$  of which consists of m-bit digital words. These words can be in any binary representation i.e. sign-magnitude, one's complement or two's complement, depending on the digital circuitry following the modulator. The local DAC is also m-bit and should be as linear as possible i.e. its output levels should be equidistant.

### 5.3 Continuous-Time Loop Filter Implementation

The purpose of this section is to explain how continuous-time filters can be used in the design of  $\Delta\Sigma$  modulators to implement a discrete-time transfer function [5]. It should be noted that the output of the loop filter is sampled by the local ADC that follows the LF. Also the result at the LF output is dependent on the waveform of the signal at its input, which can be continuous- or discrete-time, as for example the waveform of the DAC output signal (see Figs. 5.5 and 5.6).

To begin with, let the input signal to the continuous-time filter, be an analog discrete-time signal of the form:

$$x_1(t) = \sum_{n=-\infty}^{\infty} x(nT_s) \cdot h(t - nT_s) \quad (5.2)$$

In this equation  $x(nT_s)$  a sampled signal with  $h(t)$  describing the shape of each pulse in the time domain. Thus,  $h(t)$  takes a value within the time limits  $t \in [0, T_s]$  being zero outside these limits. Let also the impulse response of the continuous-time filter be  $f(t)$ .

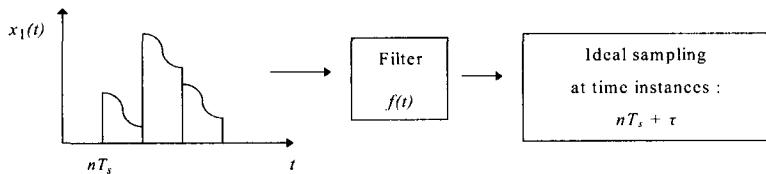


Fig. 5.9 Implementation of a discrete-time filter using a continuous-time filter.

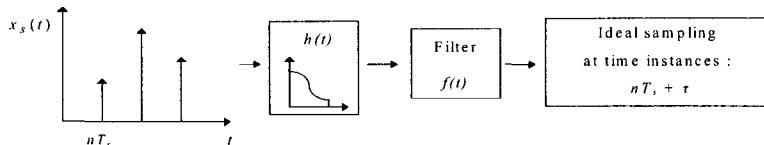


Fig. 5.10 Equivalent procedure to that in Fig. 5.9.

The whole procedure is clearly presented in Fig. 5.9 and equivalently in Fig. 5.10. It can be seen by means of Fig. 5.10, that passing a discrete-time signal  $x_1(t)$  through a continuous-time filter, having a transfer function  $F(s)$ , is equivalent to passing the ideally sampled signal  $x_s(t)$  of  $x(t)$  through a CT filter with transfer  $H(s) \cdot F(s)$ . Taking the Inverse Laplace Transform of the latter the impulse response is

$$hf(t) = L^{-1}\{H(s) \cdot F(s)\} \quad (5.3)$$

Since  $hf(t)$  is sampled at time instances  $nT_s + \tau$ , the impulse response of the discrete-time filter is that obtained by the impulse invariant method [6]. Therefore

$$HF(z) = z^{-\tau/T_s} \cdot \sum_{n=0}^{\infty} hf(nT_s + \tau)z^{-n} \quad (5.4)$$

In the case of  $\Delta\Sigma$  modulators which use a CT filter,  $\tau$  is equal to  $T_s$ . Also when  $y_a(t)$  is of the form shown in Fig. 5.6,  $h(t)$  will be as follows:

$$h(t) = \begin{cases} h(t) = 1 & t \in [\frac{T_s}{4}, \frac{3T_s}{4}] \\ h(t) = 0 & t \notin [\frac{T_s}{4}, \frac{3T_s}{4}] \end{cases} \Rightarrow H(s) = \frac{e^{-s\frac{T_s}{4}} - e^{-s\frac{3T_s}{4}}}{s} \quad (5.5)$$

Using Eqs. (5.3) to (5.5) the following cases are examined:

a) Simple integration:  $F(s) = \frac{1}{T \cdot s}$

$$hf(t) = L^{-1} \left\{ \frac{e^{-s\frac{T_s}{4}} - e^{-s\frac{3T_s}{4}}}{s} \cdot \frac{1}{T \cdot s} \right\} = L^{-1} \left\{ \frac{e^{-s\frac{T_s}{4}}}{T \cdot s^2} - \frac{e^{-s\frac{3T_s}{4}}}{T \cdot s^2} \right\}$$

or  $hf(t) = \frac{(t - T_s/4)u(t - T_s/4) - (t - 3T_s/4)u(t - 3T_s/4)}{T} \quad (5.6)$

where  $u(t)$  is the unit step function. Substituting  $hf(t)$  from Eq. (5.6) into Eq. (5.4) with  $\tau = T_s$  gives

$$HF(z) = \sum_{n=0}^{\infty} \frac{(nT_s + \frac{3}{4}T_s)u(nT_s + \frac{3}{4}T_s) - (nT_s + \frac{T_s}{4})u(nT_s + \frac{T_s}{4})}{T} \cdot z^{-(n+1)}$$

$$\begin{aligned}
 &= \sum_{n=0}^{\infty} \frac{(nT_s + 3T_s/4) - (nT_s + T_s/4)}{T} \cdot z^{-(n+1)} \\
 &= \frac{T_s}{2T} z^{-1} \sum_{n=0}^{\infty} z^{-n}
 \end{aligned} \tag{5.7}$$

Finally Eq. (5.7) can be written (see Eq. (A.1) in Appendix at the end of this chapter) as follows:

$$HF(z) = \frac{T_s}{2T} \cdot \frac{z^{-1}}{1-z^{-1}} \tag{5.8}$$

b) Double integration:  $F(s) = \frac{1}{T^2 s^2}$

$$hf(t) = \frac{(t - T_s/4)^2 u(t - T_s/4) - (t - 3T_s/4)^2 u(t - 3T_s/4)}{2T^2} \tag{5.9}$$

Substituting in Eq. (5.4) and using Eq. (A.2) in Appendix (see problem 5.2) gives

$$HF(z) = \frac{T_s^2}{4T^2} \cdot \frac{z^{-1}(1+z^{-1})}{(1-z^{-1})^2} \tag{5.10}$$

c) Triple integration  $F(s) = \frac{1}{T^3 s^3}$

$$hf(t) = \frac{(t - T_s/4)^3 u(t - T_s/4) - (t - 3T_s/4)^3 u(t - 3T_s/4)}{6T^3} \tag{5.11}$$

Substituting in (5.4) and using Eq. (A.3) in Appendix A (see problem 5.2) gives

$$HF(z) = \frac{T_s^3}{192T^3} \cdot \frac{z^{-1}(1+70z^{-1}+13z^{-2})}{(1-z^{-1})^3} \quad (5.12)$$

It can be easily shown that Eqs. (5.8), (5.10) and (5.12) at low frequencies all reduce to

$$HF(e^{j\omega T_s}) \approx \frac{1}{2} \cdot F(\omega) \quad (5.13)$$

But these discrete-time transfer functions, apart from the one given by Eq. (5.8), do not correspond to pure discrete-time integration of the form  $1/(1-z^{-1})^k$  as one could expect. This forces the designer to make additional calculations in order to implement the  $NTF(z)$  correctly. However in all cases the order is preserved, meaning that the implementation of a third-order  $\Delta\Sigma$  modulator will require three integrators of either type, CT or DT.

In case the waveform  $y_a(t)$  has the shape shown in Fig. 5.5,  $h(t)$  will be as follows:

$$h(t) = \begin{cases} h(t) = 1 & t \in [0, T_s] \\ h(t) = 0 & t \notin [0, T_s] \end{cases} \Rightarrow H(s) = \frac{1 - e^{-sT_s}}{s} \quad (5.14)$$

Then for single, double and triple integration the transfer function  $HF(z)$  will be (see problem 5.3) as follows:

$$HF(z) = \frac{T_s}{T} \cdot \frac{z^{-1}}{1-z^{-1}} \quad (5.15)$$

$$HF(z) = \frac{T_s^2}{2T^2} \cdot \frac{z^{-1}(1+z^{-1})}{(1-z^{-1})^2} \quad (5.16)$$

$$HF(z) = \frac{T_s^3}{6T^3} \cdot \frac{z^{-1}(1+4z^{-1}+z^{-2})}{(1-z^{-1})^3} \quad (5.17)$$

The same method can be applied in order to determine the equivalent discrete-time transfer function when the NTF zeros are not at dc. Such a case is that of band-pass  $\Delta\Sigma$  modulators, which are usually required to operate at high frequencies and thus have to be implemented as CT circuits [7,8].

#### 5.4 Circuit Design of a First-Order $\Delta\Sigma$ Modulator

As a first application example of the analysis in the previous sections, we consider here the circuit design of the first-order  $\Delta\Sigma$  modulators, which are depicted in Figs. 5.7 and 5.8. Let the loop filter  $L_1(z)$  be

$$L_1(z) = \frac{1}{1 - z^{-1}} \quad (5.18)$$

A very important point in the design is the selection of the oversampling ratio (OSR). This is determined by the ratio  $f_s/2f_b$ , where  $f_b$  is the desired signal frequency band and  $f_s$  the sampling frequency. The latter is determined by the frequency limitations of the opamps which are used in the integrators. So, if  $f_s$  can be high, OSR also can be high.

Let us, for example, design a first-order  $\Delta\Sigma$  modulator with the signal frequency band  $f_b=390\text{Hz}$  and oversampling ratio 64. Then the required sampling frequency will be  $f_s=50\text{kHz}$ . It is clear that for SC implementation making use of the integrator in Fig. 5.2b,  $C_1 = C_2 = C$ , say  $330\text{pF}$ .

For CT implementation, using Eq. (5.15), with  $T$  being the time constant  $RC$ , we get  $RC = T = T_s$ . Therefore  $RC = 20\mu\text{s}$ . However, since the amplitude of the signal at the input of the quantizer can be scaled by a factor  $k$  without affecting the result at its output, as was explained in Chapter 4, it is not absolutely necessary for  $RC$  to be exactly  $20\mu\text{s}$ . It should be of that order though, so we select  $R_1 = R_2 = 51\text{k}\Omega$  and  $C = 330\text{pF}$  giving  $RC = 16.8\mu\text{s}$ .

Once the  $RC$  time constant has been determined, the determination of the actual values of  $R$ 's and  $C$ 's will depend on whether the modulator

circuit is going to be built as an integrated circuit or using discrete components. In the first case certain manufacturing limitations like chip size, power dissipation e.t.c. will influence the selection. On the other hand, when building the circuit using discrete components, the values of the resistors should be low and those of the capacitors high. This is so in order to minimize thermal noise and the effects of offset currents, leakage currents and parasitics. Care should be taken though that the impedance level will not be too low, otherwise the resulting high currents will bring the opamps to non-linear operation, which is highly undesirable.

The modulator input signal Full-Scale (FS) range depends on the integrator gain  $G_s$  for the input signal and on that for the feedback signal  $\pm V_{ref}$ , which is the output of the local DAC. If the latter is  $G_f$ , we may write that  $G_s \cdot FS = G_f V_{ref}$ . In Fig. 5.7 for example the gains  $G_s$  and  $G_f$  are given by  $-(R_1 C)^{-1}$  and  $-(R_2 C)^{-1}$  respectively. Thus the choice of the values of  $R_1$  and  $R_2$  as well as that of  $V_{ref}$  fix also the FS of the modulator.

It should be noted that in the above designs we may change the OSR without having to make any change in the circuit. For example these modulators can operate properly with OSR  $64/2=32$  when the signal band becomes  $f_b = 2 \times 390\text{Hz} = 780\text{Hz}$ . This is also true for any NTF with zeros at dc. An additional advantage of the SC implementation is of course the fact that  $f_b$  can be changed simply by changing the clock frequency  $f_s$ .

## 5.5 Circuit Design of a Second-Order $\Delta\Sigma$ Modulator

Consider the CIDF structure of second-order  $\Delta\Sigma$  modulator to be as shown in Fig. 5.11.

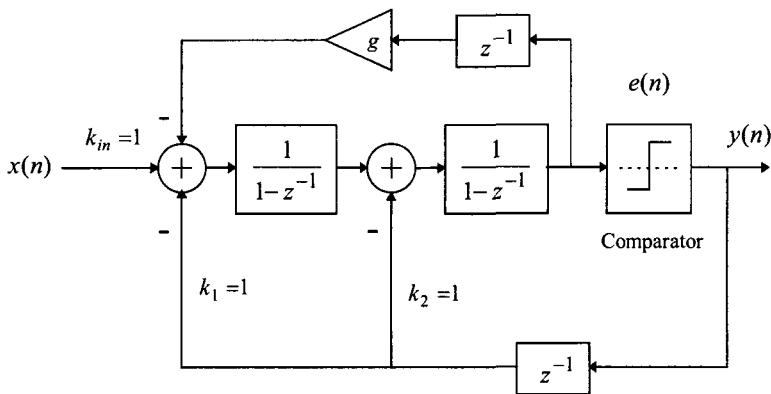


Fig. 5.11 CIDF structure of a second-order  $\Delta\Sigma$  modulator.

Straight forward analysis gives the following  $NTF(z)$ :

$$NTF(z) = \frac{1 - (2 - g)z^{-1} + z^{-2}}{1 + gz^{-1}} \quad (5.19.a)$$

The gain  $g$  is determined by the NTF's zeros. The numerator of the NTF can be written as follows:

$$1 - (2 - g)z^{-1} + z^{-2} = (1 - e^{-j\phi}z^{-1}) \cdot (1 - e^{j\phi}z^{-1}) \quad (5.19.b)$$

Let us select the value of  $\phi$  from Table 4.1 to be equal to  $\phi = 2\pi 0.5774 f_b / f_s$ , where  $f_b$  is the highest frequency in the signal band and  $f_s$  the sampling frequency. Coefficient matching in Eq. (5.19.b) gives (since  $g$  is real)

$$2\cos(\phi) = 2\cos\left(2\pi 0.5774 \frac{f_b}{f_s}\right) = 2 - g \quad (5.20)$$

For OSR=64, i.e.  $f_b/f_s = 1/128$ , the value of  $g$ , obtained from Eq. (5.20), is  $8 \times 10^{-4}$ . This is really very small and makes  $|gz^{-1}| << 1$ . Consequently the  $NTF(z)$  can be simplified, i.e. it is almost equal to the numerator of the right part in Eq. (5.19a). As a result the effect of the gain  $g$  on the NTF is just to move the zeros slightly from dc. This result can be extended for any low-pass  $\Delta\Sigma$  modulators, where for high OSR,  $g$  affects mainly the zeros of the NTF and very little its poles, the position of which are fixed by the coefficients  $k_1$  and  $k_2$ . The proof is left to the reader, who can be helped by the example 3.2.

The block diagram of the second-order  $\Delta\Sigma$  modulator, when this is implemented using CT integrators, is as shown in Fig. 5.12.

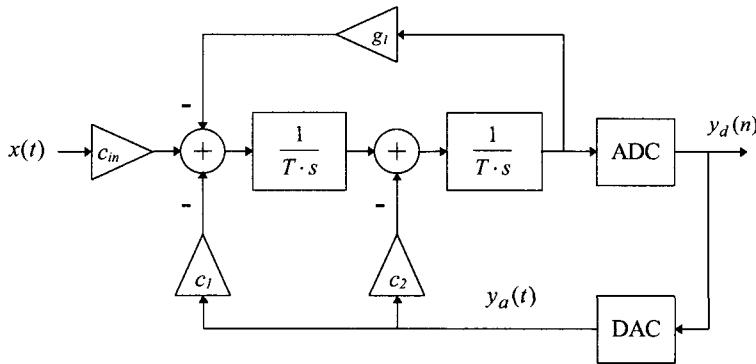


Fig. 5.12 Block diagram of the second-order  $\Delta\Sigma$  modulator implemented using CT integrators.

We want to determine the values of  $c_{in}$ ,  $c_1$ ,  $c_2$  and  $g_1$  in order to implement the NTF given by Eq. (5.19a). The previous result, i.e. that the coefficients  $c_1$  and  $c_2$  fix the position of the poles and  $g_1$  the position of the zeros of the NTF is also valid here. Therefore  $g_1$  is related to  $g$ , while  $c_1$ ,  $c_2$  are related to the coefficients  $k_1$  and  $k_2$ . Coefficient  $c_{in}$  determines the Full-Scale of the input signal. To proceed with the determination of  $c_1$ ,  $c_2$  and  $c_{in}$  we may work as follows :

Ignoring the gain  $g$  we may write  $L_1(z)$  for of the block diagram of Fig. 5.11 as follows:

$$L_1(z) = -z^{-1} \left( \frac{1}{1-z^{-1}} + \frac{1}{(1-z^{-1})^2} \right) \quad (5.21)$$

On the other hand from the block diagram of Fig. 5.12, again ignoring  $g_1$  and using Eqs. (5.8) and (5.10), the equivalent  $L_1(z)$  can be expressed as follows:

$$L_1(z) = -z^{-1} \left( c_2 \cdot \frac{T_s}{2T} \frac{1}{1-z^{-1}} + c_1 \cdot \frac{T_s^2}{4T^2} \frac{(1+z^{-1})}{(1-z^{-1})^2} \right) \quad (5.22)$$

From Eqs. (5.21) and (5.22), equating right hand parts and assuming for simplicity that  $T = T_s$ , we obtain  $c_1 = 2$  and  $c_2 = 3$ .

To obtain the value of  $c_{in}$  we proceed as follows:

From Fig. 5.11 we obtain

$$\begin{aligned} L_o(z) &= \frac{1}{(1-z^{-1})^2} \\ \text{or} \quad L_o(e^{j\omega T_s}) &= \frac{1}{(1-e^{-j\omega T_s})^2} \end{aligned} \quad (5.23)$$

and since the frequency of the input signal is much lower than the sampling frequency we may write approximately

$$L_o(e^{j\omega T_s}) = \frac{1}{(1-e^{-j\omega T_s})^2} \cong \frac{1}{(j\omega T_s)^2} \quad (5.24)$$

Thus the frequency response of the double discrete-time integrator is almost equal to that of a double continuous-time integrator. Therefore  $c_{in} \cong 1$ .

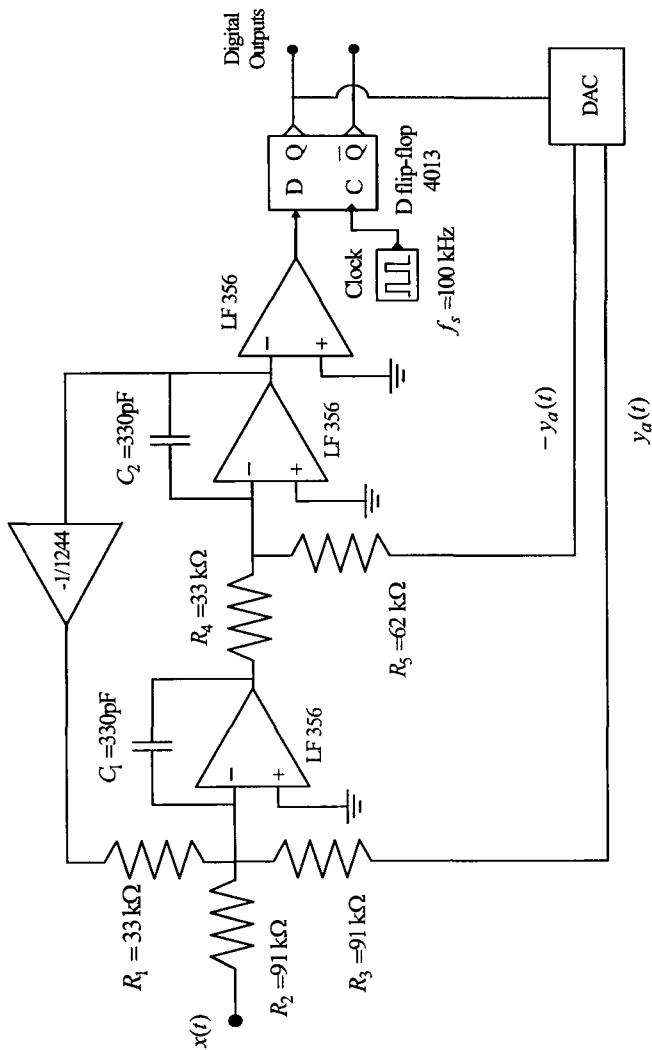


Fig. 5.13 Circuit of a second-order  $\Delta\Sigma$  modulator employing CT integrators (OSR = 64). Nearest standard values of the resistors appear.

For the same reason we will have  $g_1 = g$ . In fact  $g$  affects the frequency response of  $L_o(z)$  and  $L_1(z)$  at low frequencies inside the signal band. But since at these frequencies the two integrator types, discrete-time and continuous-time, have nearly identical frequency responses, for the effect of  $g_1$  to be the same as that of  $g$  requires,  $g_1 = g$ .

The corresponding detailed circuit using CT integrators is shown in Fig. 5.13. Component values have been calculated for  $f_b = 781\text{Hz}$  and  $\text{OSR}=64$ . Thus the sampling frequency  $f_s$  is  $100\text{kHz}$  and  $RC = T_s = 10\mu\text{s}$ . The value of  $C$  was selected to be  $330\text{pF}$ . Thus  $R=30.3\text{k}\Omega$ . Based on this we may calculate the component values to be as follows:

$$C_1 = C_2 = 330\text{pF},$$

$$R_1 = R_4 = 30.3\text{k}\Omega,$$

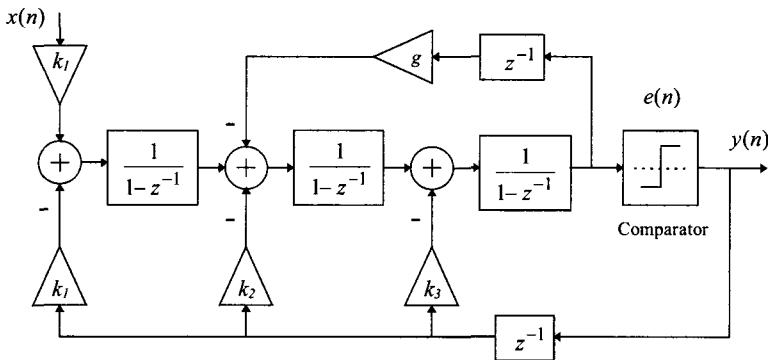
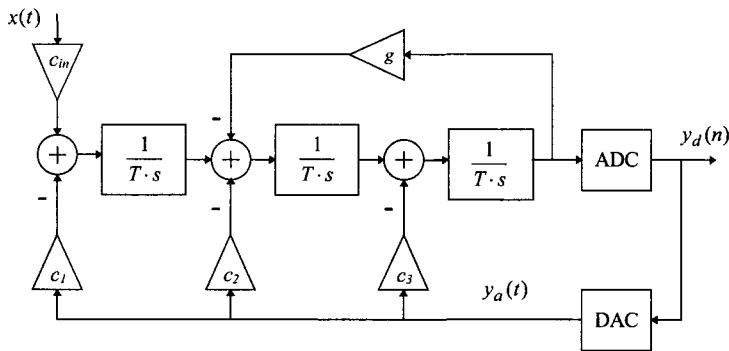
$$R_2 = R/c_{in} = 30.3\text{k}\Omega, R_3 = R/c_1 = 15.15\text{k}\Omega, R_5 = R/c_2 = 10.1\text{k}\Omega.$$

If we want to have the Full-Scale equal to  $2.5\text{V}$  (for  $\pm V_{ref} = \pm 5\text{V}$ ) then  $R_2 = 15.15\text{k}\Omega$ . Finally, if we raise the impedance level of resistors  $R_2$ ,  $R_3$  and  $R_5$  6 times and use standard values, we will obtain the component values appearing in Fig. 5.13 (see also problem 5.6).

## 5.6 Circuit Design of a Third-Order $\Delta\Sigma$ Modulator

The CIDF structure of a third-order  $\Delta\Sigma$  modulator is shown in Fig. 5.14, while the block diagram of its implementation using CT integrators is shown in Fig. 5.15.

The NTF under implementation has the poles and zeros given in Table 5.1. The values of the coefficients  $k_1$ ,  $k_2$  and  $k_3$  of the CIDF structure of the  $\Delta\Sigma$  modulator as well as those of  $c_1$ ,  $c_2$  and  $c_3$  required for its implementation in Figs. 5.14 and 5.15 respectively are also given in the same Table. These values have been determined in a way similar to the way it was followed in the case of the second-order  $\Delta\Sigma$  modulator (see problem 5.4).

Fig. 5.14 CIDF structure of a third-order  $\Delta\Sigma$  modulator.Fig. 5.15 Block diagram of the third-order  $\Delta\Sigma$  modulator implemented using CT integrators.Table 5.1 Characteristics of the selected third-order  $\Delta\Sigma$  modulator

Poles of $NTF(z)$	Frequencies of zeros of $NTF(z)$ in ( $f_b/f_s$ )	Coefficients $k$	Coefficients $c$
-0.0367 $0.8076 \pm j0.2733$	0 $\pm 0.8$	$k_1=0.1158$ $k_2=0.2776$ $k_3=1.0267$	$c_1=0.2316$ $c_2=0.7868$ $c_3=2.4154$

The value of the feedback gain  $g$  in the local resonator is calculated from the equation

$$2\cos\left(2\pi \cdot 0.8 \frac{f_b}{f_s}\right) = 2 - g \quad (5.25)$$

as it was also done in the case of the second-order  $\Delta\Sigma$  modulator. Solving Eq. (5.25) for  $g$ , if the OSR=64, gives  $g = 1/650$ .

The actual circuit for the implementation of this third-order  $\Delta\Sigma$  modulator employing CT integrators is shown in Fig. 5.16.

On the other hand in Fig. 5.17 the actual circuit implementing the same NTF and employing SC integrators is shown. In the calculations an OSR=32 was used to obtain a more practical implementation with discrete components, in case the reader would like to build and test the  $\Delta\Sigma$  modulator circuit in his laboratory. It should be noted that the value of  $g$  in this case is equal to 1/163 being calculated via Eq. (5.25) for  $f_s/f_b = 64$ .

In both cases, the Full-Scale is 2.5V and the sampling frequency 100kHz. For the CT implementation in Fig. 5.16 (OSR=64)  $f_b = 781\text{Hz}$ , while for the SC implementation in Fig. 5.17 (OSR=32)  $f_b = 1562\text{Hz}$ .

To avoid the use of high resistance values in the first integrator in Fig. 5.16 the impedance level has been reduced by a factor of 10 with a corresponding increase in the capacitor value in this integrator.

In the circuit in Fig. 5.17, for reliability reasons, when the circuit is to be built in the laboratory using discrete components, the values of the capacitors have been selected high enough in order to reduce the effects of thermal noise, offset currents, leakage currents and parasitic capacitances. The capacitance level was selected to be 1.5nF, in order to avoid high capacitances that would draw high currents and thus result in non-linear settling of the opamps. Suitable values of the capacitances in the feedback paths whose ratios are nearest to the ratios of the coefficients  $k_1$ ,  $k_2$  and  $k_3$  were 330pF, 92pF and 370pF. The value (15nF) of the storage capacitor of the first integrator has been multiplied by 10 in order the capacitor in the corresponding feedback path not to have a very small value.

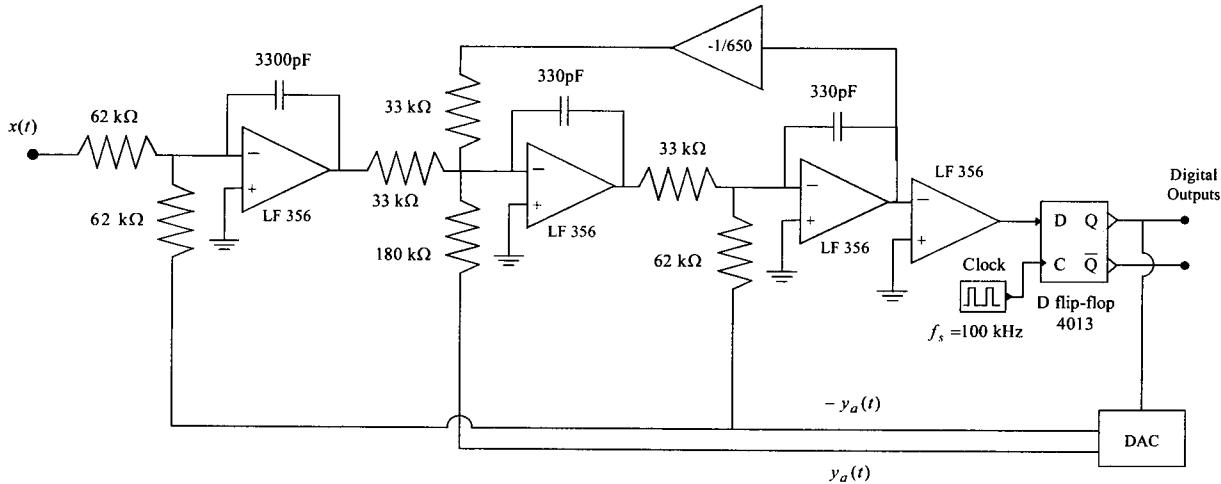


Fig. 5.16 Circuit of a third-order  $\Delta\Sigma$  modulator using CT integrators (OSR = 64).

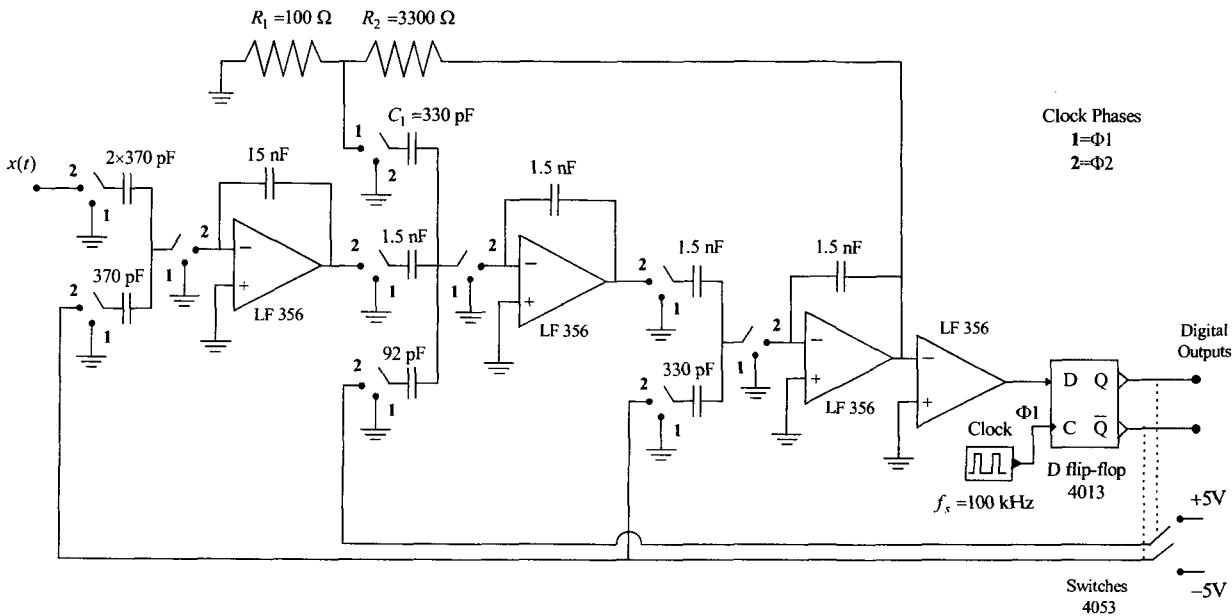


Fig. 5.17 A SC circuit of the third-order  $\Delta\Sigma$  modulator under consideration with capacitance values suitable for a practical construction with discrete components (OSR = 32).

The value of the capacitor at the input is  $2 \times 370\text{pF}$  so that the Full-Scale will be  $2.5\text{V}$ . The value of  $g$  is implemented by means of a resistive voltage divider combined with the capacitance ratio  $330\text{pF}/1.5\text{nF}$  i.e.

$$g = \frac{330\text{pF}}{1.5\text{nF}} \frac{R_1}{R_1 + R_2} = 1/155 \quad (5.26)$$

With these component values the time constant  $(R_1//R_2)C_1$  is about  $33\text{ns}$ , which is much smaller than  $T_s/2 = 5\mu\text{s}$ . Thus the capacitor  $C_1$  has the time to get fully charged.

## 5.7 Circuit Design of a Fourth-Order Band-pass $\Delta\Sigma$ Modulator

It was stated in Chapter 3 that band-pass  $\Delta\Sigma$  modulators can be obtained from the low-pass ones by applying the transformation  $z^{-1} \rightarrow -z^{-2}$  to the NTF of the latter [2]. With this transformation each product term  $(1-z_i z^{-1})$  of the NTF is turned to  $(1+z_i z^{-2})$ , which can be written as follows:

$$\begin{aligned} (1+z_i z^{-2}) &= (1-j^2 z_i z^{-2}) = \\ (1-e^{j(\varphi_i/2+\pi/2)} \sqrt{|z_i|} z^{-1}) &(1+e^{j(\varphi_i/2+\pi/2)} \sqrt{|z_i|} z^{-1}) = \\ (1-e^{j(\varphi_i/2+\pi/2)} \sqrt{|z_i|} z^{-1}) &(1-e^{j(\varphi_i/2-\pi/2)} \sqrt{|z_i|} z^{-1}) \end{aligned} \quad (5.27)$$

Therefore each zero is split in two with arguments  $\varphi_i/2 + \pi/2$  and  $\varphi_i/2 - \pi/2$ , as shown in Fig. 5.18. Arguments  $\pi/2$  and  $-\pi/2$  correspond to frequencies  $f_s/4$  and  $-f_s/4$ , thus the resulting modulators can encode signals with spectra that occupy narrow bands of frequencies around the frequency  $f_s/4$  (see also Sec. 3.11). With  $z = e^{j\theta}$ , the argument  $\theta$  is transformed to  $\pm\pi/2 + \theta/2$ . As a result, the magnitude response of  $NTF_{BP}(z)$  for the range of frequencies  $[-f_s/2, 0]$  and  $[0, f_s/2]$  is exactly the same with the magnitude response of the

$NTF_{LP}(z)$  of the low-pass prototype for the range of frequencies  $[-f_s/2, f_s/2]$ , as shown in Fig. 3.18. The respective frequency band, that is encoded, falls within the range  $[f_s/4 - f_s/4 OSR, f_s/4 + f_s/4 OSR]$ .

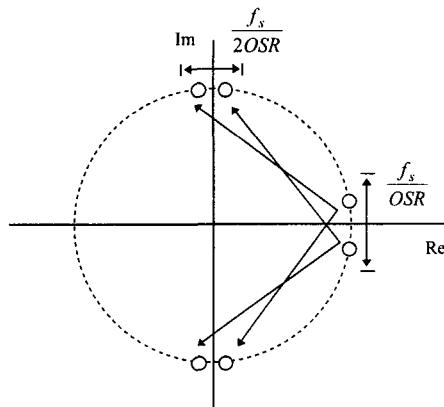


Fig. 5.18 Movement of zeros in the  $z$ -plane when the transformation  $z^{-l} \rightarrow -z^{-2}$  is applied

The structure of a fourth-order band-pass  $\Delta\Sigma$  modulator, which has been obtained by applying the above  $z^{-l} \rightarrow -z^{-2}$  transformation to the second-order  $\Delta\Sigma$  modulator of Sec. 5.5, is shown in Fig. 5.19.

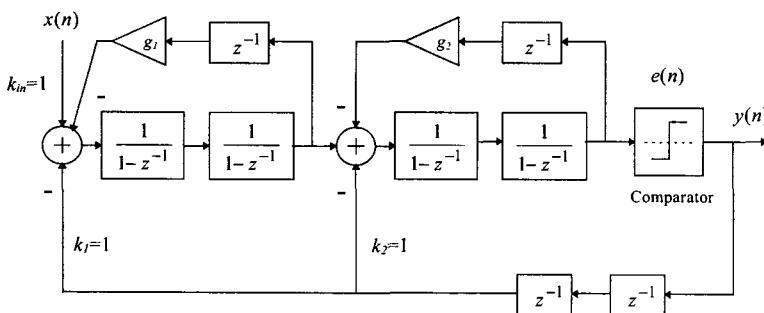


Fig. 5.19 Structure of a fourth-order  $\Delta\Sigma$  modulator.

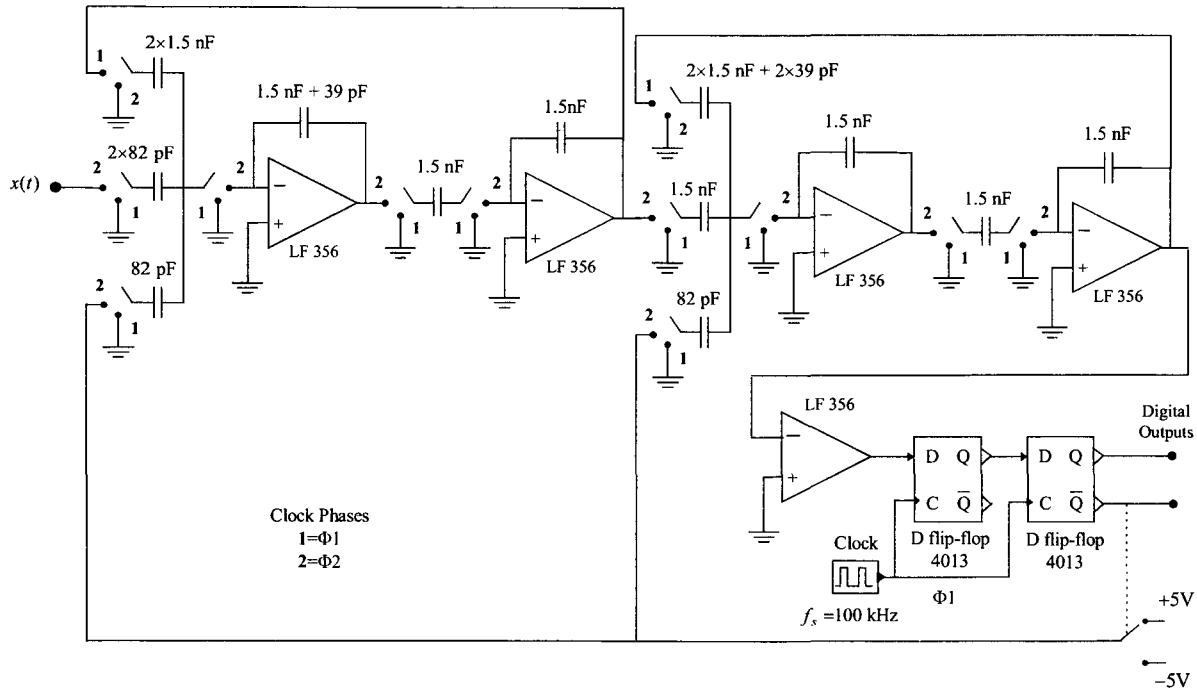


Fig. 5.20 SC implementation of the fourth-order band-pass  $\Delta\Sigma$  modulator.

The values of  $g_1$  and  $g_2$  in the two local resonators, which determine the positions of the  $NTF(z)$  zeros, can be obtained again from Eq. (5.20) by substituting the transformed arguments, i.e.

$$2\cos\left(\frac{\phi_i}{2} + \frac{\pi}{2}\right) = 2\sin\left(\frac{\phi_i}{2}\right) = 2 - g_i \quad (5.28)$$

In this equation  $\phi_i$  are the frequencies of the zeros of the second-order modulator. An OSR=32 results in  $\phi_{1,2} = \pm 2\pi 0.5774/32$ . Then from Eq. (5.28) we get  $g_{1,2} \approx 2(1 \pm 1/36)$ .

It should be mentioned, that in order to avoid overloading the integrators, coefficients  $k_m$ ,  $k_1$  and  $k_2$  may have to be scaled. If  $f_s=100$  kHz and OSR=32 the frequency band of the band-pass modulator under consideration will be [25-0.781, 25+0.781] kHz. The implementation of this modulator as a SC circuit is shown in Fig. 5.20.

For CT implementation of band-pass modulators the determination of gains  $g_i$ , as well as of the feedback coefficients  $c_i$  is more difficult, because the approximations made in Sec. 5.5 cannot be applied. Thus the procedure given in Sec. 5.3 should be followed from start [7,8], (see also problem 5.7).

## 5.8 Testing the Operation of Modulators Experimentally

The experimental test of a  $\Delta\Sigma$  modulator is mainly concerned with the measurement of the SNR and the spectrum of the  $\Delta\Sigma$  sequence.

To take the various measurements, it is useful to possess a workstation in combination with a data acquisition card. The card receives a large number of samples (65536 samples) from the D flip-flop output. The logic states 1 and 0 are taken to correspond to the +1 and -1 states of the  $\Delta\Sigma$  sequence. Applying FFT to these data, using also a window (for example the Blackman window) one obtains the correct spectrum. Using a pure sinewave as the input signal, this appears clearly in the FFT of the pulse sequence. The signal power is considered to be the sum of the squares of all components which are inside the signal frequency lobe. On

the other hand the quantization noise power is the sum of the squares of all components in the spectrum outside the lobe of the main frequency lying of course inside the frequency band. The ratio of these two powers gives the SNR. Applying this procedure for various amplitudes of the input signal the diagrams of the SNR for the  $\Delta\Sigma$  modulators we designed above are obtained. In all cases the input signal is a sine wave.

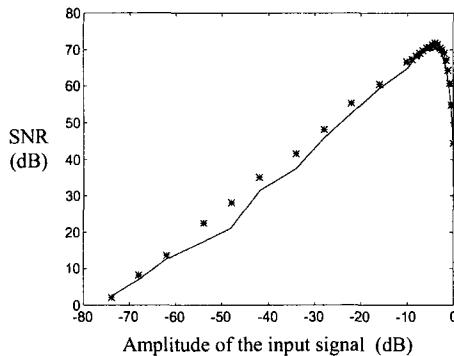


Fig. 5.21 Plot of SNR versus the amplitude of the input signal for the second-order  $\Delta\Sigma$  modulator using CT integrators. Continuous line is obtained by simulation and crosses are the experimental points.

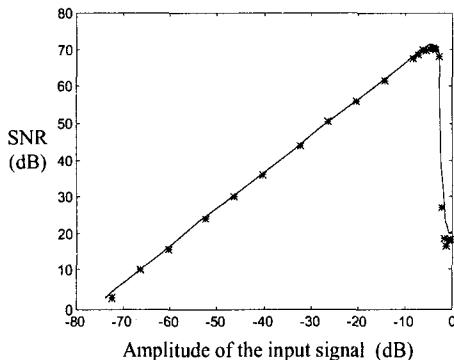


Fig. 5.22 Simulation (continuous line) and experimental (points) plots of SNR in the case of the SC third-order  $\Delta\Sigma$  modulator.

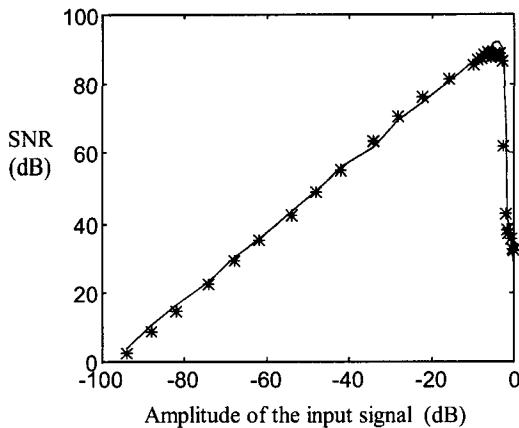


Fig. 5.23 Simulation (continuous line) and experimental (points) plots of the SNR for the third-order  $\Delta\Sigma$  modulator employing CT integrators with OSR= 64.

In Figs. 5.24 and 5.25 are respectively shown separately the spectrum obtained experimentally and that obtained by simulation for the third-order  $\Delta\Sigma$  modulator employing CT integrators and operating at an OSR=64.

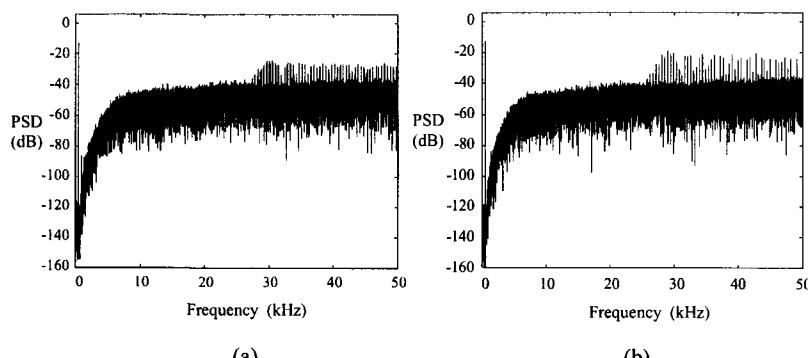


Fig. 5.24 Plot showing the spectrum of the shaped quantization noise at the output of the third-order  $\Delta\Sigma$  modulator a. Experimental, b. by simulation.

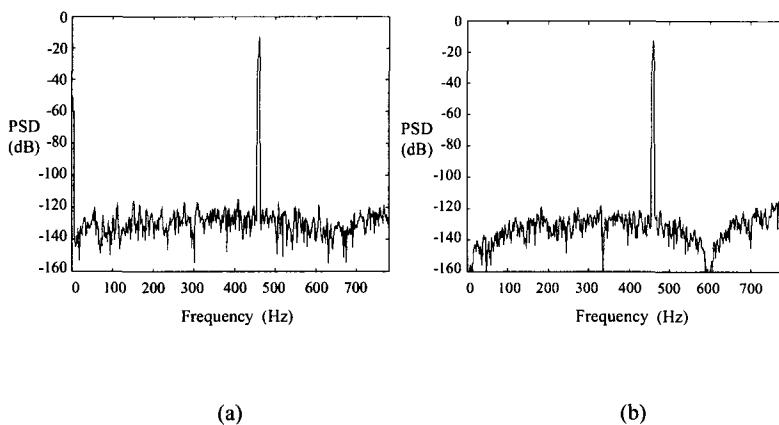


Fig. 5.25 Details of the spectra in Fig. 5.24 inside the signal frequency band a. Experimental, b. by simulation.

In Figs. 5.26 and 5.27 the spectra obtained experimentally and by simulation are shown for the SC fourth-order band-pass  $\Delta\Sigma$  modulator.

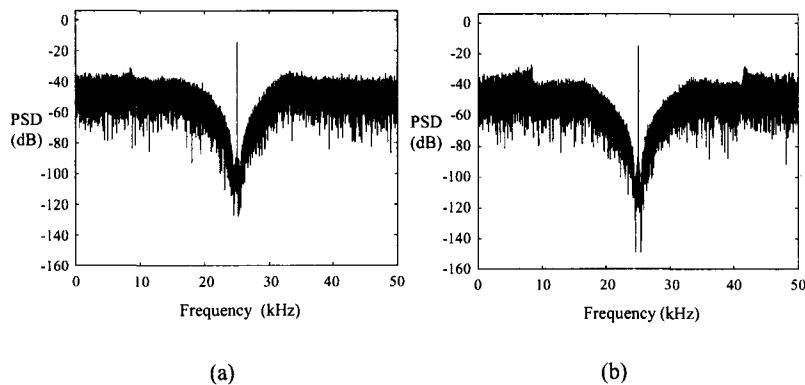


Fig. 5.26 Spectrum of the quantization noise at the output of the SC fourth-order  $\Delta\Sigma$  modulator obtained a. Experimentally, b. by simulation.

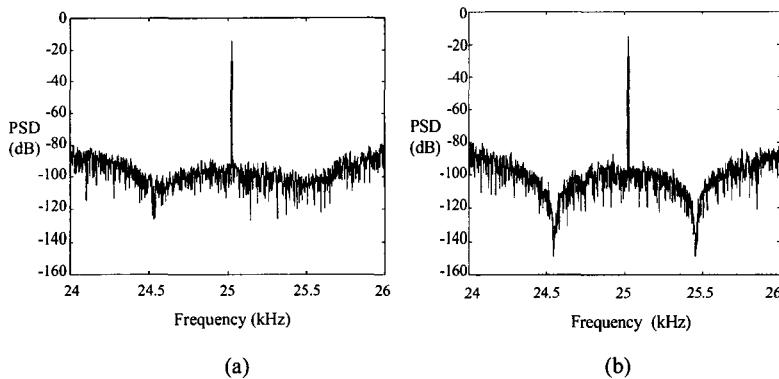


Fig. 5.27 Details of the spectra in Fig. 5.26 within the signal frequency band a. Experimental, b. by simulation.

## 5.9 Low-Power, Low-Voltage $\Delta\Sigma$ Modulators

High technology portable electronic equipment like mobile telephones, lap-top computers etc. operate on batteries that, for obvious reasons, have to be small, light, last long and discharge as slowly as possible. This demands designing the electronic circuitry properly to consume as low-power as possible during operation. On the other hand advanced designs require larger numbers of components on the die, a requirement met by a reduction of the components size. VLSI CMOS technology is most suitable today to be exploited to meet both these demands, i.e. consumption of low-power and the scaling down of the components size.

However scaling while keeping the same supply voltage, say 5V, creates breakdown problems due to the presence of high field strengths in the transistors. Consequently the supply voltage should be reduced too. In digital circuitry, scaling down leads to higher speeds while reducing the supply voltage surely results in low power consumption. However in high resolution analog and mixed-signal circuit designs supply voltage reduction may lead to increased power consumption. Additional important problems created in analog circuits are the following:

- a. reduction in signal swing that can limit the achievable dynamic range and increase harmonic distortion
- b. rapid increase of transistor output conductance not coupled by analogous rise in transistor transconductance. This results in reduced intrinsic gain of the transistor
- c. increased thermal noise
- d. mismatch-induced offsets
- e. passive components, resistors and capacitors may not be available, making
- f. bottom plate parasitics become more important

It must be stressed that the design and implementation of analog circuits with high dynamic range requires the solution of various problems associated with the output swing, common-mode rejection, power supply noise rejection etc.

Therefore in mixed-signal VLSI circuits, the required interfacing circuitry is desirable to be mostly digital. In this case the converters, ADC and DAC, designed to employ oversampling  $\Delta\Sigma$  modulators, prove most suitable, since their analog part is simple, while the most involved processing is performed by the digital part. Thus by scaling down the digital circuitry and reducing the supply voltage the consumed power can be reduced dramatically.

The power in the  $\Delta\Sigma$  loop is mainly dissipated in the first integrator. Clearly the noise performance of the first integrator is more important than the corresponding performance of any other subsequent integrator in the  $\Delta\Sigma$  modulator. This noise is limited by the  $kT/C$  noise, where  $k$  is Boltzmann's constant,  $T$  the absolute temperature and  $C$  the capacitance used in the integrator. The value of this capacitor should be high enough to reduce this noise from the first integrator for the accuracy of the overall modulator to be independent of the oversampling ratio, whereas smaller capacitors can be used in subsequent integrators. But higher capacitance for a given voltage value requires higher power to charge, thus proving the dominating role of the first integrator in the total power dissipation in the  $\Delta\Sigma$  loop.

There are three integrator architectures that may be employed in low-power, low-voltage  $\Delta\Sigma$  modulators. These are the Switched-Capacitor (SC) integrator, the Continuous-Time (CT) integrator and the Switched-

Current integrator. Assuming ideal opamps in the cases of SC and CT integrators, the power dissipation in all three integrators is found to be proportional to the Dynamic Range, and the Nyquist sampling rate corresponding to the signal baseband, while it also depends on the amplifier design.

The integrators, both the SC and the CT, are usually fully-differential, Fig. 5.3. In the case of the CT integrator practical circuit the amplifier settling is not so severely strict as in the case of the SC circuit, while off-chip resistors can be used to increase the input signal swing. There are however some design aspects that make the CT integrator unsuitable for use in digital-audio applications, where the required dynamic range cannot be achieved easily by this integrator. These aspects include the sensitivity to jitter, sensitivity to hysteresis in the feedback reference signal and the need for either off-chip resistors or highly linear on-chip resistors.

Analysis also shows that the switched-current integrator is not suitable for low-voltage operation. Indeed the power dissipation is higher than in the case of the SC integrator, while it is difficult to achieve high linearity when operating at low voltage. The latter is due to

- a. The fact that the input voltage should be a small percentage of the overdrive voltage  $V_{GS}-V_T$  in order to limit the input current swing
  - b. The presence of channel length modulation
  - c. The dependence of the charge injection MOS switches on the signal.
- Following the above argument we are left with the SC integrator to be the most suitable candidate among the three for the design of low-power, low-voltage modulators.

Among the advantages of using a fully-differential SC integrator are the following:

- a. It provides high rejection of common-mode supply and substrate noise
- b. It is relatively immune to switch charge injection errors
- c. It results in a 3 dB increase in the dynamic range, since for a given voltage range, the differential signal has twice the amplitude and four times the power of a single-ended signal, while introducing only

twice the noise power, since the signal flows along two independent paths.

The choice of the opamp architecture is of fundamental importance since this should be the best compromise with regard to various issues such as voltage swing, slewing and settling, low noise ( $1/f$  and thermal), low-power consumption and all these at low voltage. The oversampling ratio has an implicit effect on the amount of power dissipation, which increases dramatically above a certain value depending on the amplifier design.

Low-power, low-voltage single-bit  $\Delta\Sigma$  modulators of higher order ( $>2$ ) are implemented preferably by cascading first-and/or second-order stages. Single-bit is preferable to multibit in order to avoid the high linearity requirement from the DAC in the feedback path of the multibit modulator. Also cascading first-and/or second-order stages is better than using a single-stage, because in the latter the overload level is low, the peak SNR is low and there is presence of strong spectral tones at low input levels.

Finally the comparator and the feedback DAC in each modulating stage should not contribute much to the power dissipation of the modulator. In particular, as we have seen previously (Sec. 5.2.2) the 1 bit DAC is a simple network of switches connected to off-chip reference voltages, Fig. 5.8

## 5.10 Summary

The object of this chapter was to explain how one can implement, build and test in the laboratory single-bit, single-stage  $\Delta\Sigma$  modulators. The aim was to concentrate on the educational aspect of the presentation. This explains the reason why we emphasized on implementations using discrete components, since the corresponding circuits can be easily built and tested in the laboratory.

From the results that were presented above, obtained by simulation and experimentally, one may easily draw some conclusions:

- a. The experimental results obtained without any particular effort nearly coincide with those obtained by simulation

- b. There is a highest value of the amplitude of the input signal for which maximum SNR is achieved. For amplitudes of the input signal above this value the performance of the  $\Delta\Sigma$  modulator is degraded. This is because the modulator is unstable for these amplitudes. However, because the outputs of the integrators are clipped due to limits set by the power supply voltage, the modulator remains stable with decreased SNR.
- c. The performance of the  $\Delta\Sigma$  modulator is not particularly sensitive to the accuracy of the component values.
- d. The performance of a  $\Delta\Sigma$  modulator can be demonstrated in the laboratory more easily and accurately using CT circuits than SC circuits. A lot more about advantages and disadvantages of these types of circuits implementing  $\Delta\Sigma$  modulators are included in the next chapter, which is devoted to the various sources of errors in  $\Delta\Sigma$  modulators in practice.

## Problems

- 5.1 Determine the transfer function of the resonator shown in Fig. 5.4b when there exists a) a single, b) a double delay in the loop. Give your comments. When there exists a double delay in the loop, under what conditions the resonator will be effective?
- 5.2 Derive equations (5.10) and (5.12) following the procedure presented in Sec. 5.3 for the case of Eq. (5.8).
- 5.3 Show the validity of Eqs. (5.15) to (5.17), if the impulse response  $h(t)$  of the circuit that determines the pulse waveform of Fig. 5.5 is that given by Eq. (5.14).
- 5.4 Using the block diagrams in Figs. 5.14 and 5.15 calculate the coefficients  $k_i$ ,  $c_i$ , for  $i=1, 2, 3$  for the  $NTF(z)$  given in Table 5.1. Calculate also the coefficients  $k'_i$  in the case that the integrators have a delay as shown in Fig. 5.4b.
- 5.5 The impulse response of the RC filter of the exponential DM in Fig. 2.18 is  $h(t)=e^{-t/RC}$ . Following the procedure presented in Sec. 5.3 for the case of Eq. (5.8), find the equivalent discrete time

transfer function for the RC filter. Then find the  $STF(z)$  and the  $NTF(z)$  for the exponential DM.

- 5.6 In order to obtain suitable values of the resistors  $R_2$ ,  $R_3$ ,  $R_5$  in the  $\Delta\Sigma$  modulator circuit in Fig. 5.13, the initially determined values have been increased six times. Although this increase does not affect the Full-Scale of the modulator, what is its effect on the amplitude at the output of the integrators? Do we expect, theoretically, any changes in the  $\Delta\Sigma$  sequence?

State all possible ways we can change the Full-Scale of this  $\Delta\Sigma$  modulator.

State all possible ways we may change the Full-Scale of this modulator without affecting the amplitude of the signal at the output of the integrators.

State all possible ways we may scale the output of the two integrators without changing the Full-Scale of the modulator.

- 5.7 Determine the gain  $g_1$  and the coefficients  $c_1$  and  $c_2$  in the block diagram in Fig. 5.12 in order to design a second-order band-pass  $\Delta\Sigma$  modulator, with an equivalent discrete-time transfer function

$$L_1(z) = \frac{z^{-2}}{1+z^{-2}}$$

- a) when the DAC output waveform is that shown in Fig. 5.5 with  $h(t)$  given by Eq. (5.14).
- b) when the DAC output waveform is that shown in Fig. 5.6 (RZ) with  $h(t)$  given by Eq. (5.5).

Hint: Follow step by step the same procedure as in Sec. 5.3 for the derivation of Eq. 5.8. Use also the Z-Transform of Eqs. (A.5) and (A.6).

## References

- [1] R. Gregorian and G. C. Temes, *Analog MOS Integrated Circuits for Signal Processing*, New York, John Wiley & Sons, 1986.
- [2] T. Deliyannis, Yichuang Sun, J. K. Fidler, *Continuous-Time Active Filter Design*, CRC Press, 1999.
- [3] C. Tonmazou, N. Battersby, S. Porta, *Circuits and Systems Tutorials*, IEEE ISCAS' 1994.
- [4] B. P. Del Signore, D. A. Kerth, N. S. Sooch and E. J. Swanson, 'A Monolithic 20-b Delta-Sigma A/D Converter', *IEEE Journal of Solid-State Circuits*, Vol. 25, No. 6, pp. 1311-1316, December 1990.
- [5] S. R. Norsworthy, R. Schreier and G. C. Temes, *Delta-Sigma Data Converters, Theory, Design and Simulation*, New York, IEEE Press, 1997.
- [6] E. C. Ifeachor, B. W. Jervis, *Digital Signal Processing - A practical Approach*, Addison Wesley, 1993.
- [7] O. Shoaei and W. M. Snelgrove, 'Design and Implementation of a Tunable 40MHz - 70MHz GM-C band-pass  $\Delta\Sigma$  Modulator', *IEEE Transactions on Circuits and Systems II, Analog and Digital Signal Processing*, Vol. 44, No. 7, pp. 521-530, July 1997.
- [8] W. Gao, W. M. Snelgrove, 'A 950-MHz IF Second-Order Integrated LC Bandpass Delta-Sigma Modulator', *IEEE Journal of Solid-States Circuits*, Vol. 33, No 5, May 1998, pp. 723-732.
- [9] V. Peluso, M. Steyaert and W. Sansen, *Design of Low-Voltage Low-Power CMOS Delta-Sigma A/D Converters*, Boston, Kluwer Academic Publishers, 1999.
- [10] S. Rabii and B. A. Wooley, *The Design of Low Voltage Low-Power Sigma-Delta Modulators*, Kluwer Academic Publishers, 1999.
- [11] A. D. Poullarikas, *The Transforms and Applications Hand-Book*, CRC Press, IEEE Press, 1996.

## Appendix

The Z-transforms of some common sequences are given below. They can also be found in [6,11].

For  $x(n)=1$

$$X(z) = \sum_{n=0}^{\infty} z^{-n} = \frac{1}{1-z^{-1}} \quad |z| > 1 \quad (\text{A.1})$$

For  $x(n)=n$

$$\begin{aligned} X(z) &= \sum_{n=0}^{\infty} n \cdot z^{-n} = z^{-1} \cdot \frac{\partial}{\partial z^{-1}} \left( \sum_{n=0}^{\infty} z^{-n} \right) = z^{-1} \cdot \frac{\partial}{\partial z^{-1}} \left( \frac{1}{1-z^{-1}} \right) \\ &= \frac{z^{-1}}{(1-z^{-1})^2} \quad |z| > 1 \end{aligned} \quad (\text{A.2})$$

For  $x(n)=n^2$

$$\begin{aligned} X(z) &= \sum_{n=0}^{\infty} n^2 z^{-n} = z^{-1} \frac{\partial}{\partial z^{-1}} \left( \sum_{n=0}^{\infty} nz^{-n} \right) = z^{-1} \frac{\partial}{\partial z^{-1}} \left( \frac{z^{-1}}{(1-z^{-1})^2} \right) \\ &= \frac{z^{-1}(1+z^{-1})}{(1-z^{-1})^3} \quad |z| > 1 \end{aligned} \quad (\text{A.3})$$

For  $x(n)=e^{-pn}$

$$X(z) = \sum_{n=0}^{\infty} e^{-pn} z^{-n} = \frac{1}{1-e^{-p} z^{-1}} \quad |e^p| |z| > 1 \quad (\text{A.4})$$

For  $x(n)=\cos(\alpha n)$

$$X(z) = \sum_{n=0}^{\infty} \cos(\alpha n) z^{-n} = \frac{(1-z^{-1} \cos \alpha)}{1-2z^{-1} \cos \alpha + z^{-2}} \quad |z| > 1 \quad (\text{A.5})$$

For  $x(n)=\sin(\alpha n)$

$$X(z) = \sum_{n=0}^{\infty} \sin(\alpha n) z^{-n} = \frac{z^{-1} \sin \alpha}{1-2z^{-1} \cos \alpha + z^{-2}} \quad |z| > 1 \quad (\text{A.6})$$

## **Chapter 6**

# **Practical Limitations of $\Delta\Sigma$ Modulators**

### **6.1 Introduction**

The implementation of various  $\Delta\Sigma$  modulators, that we presented in the previous chapter, assumed the use of ideal components. However this is not true in practice. For example the opamps do not have infinite gain, infinite bandwidth and infinite slew-rate. There are no ideal DACs, especially multibit DACs. Switches in the SC circuits do not have zero resistance while these circuits suffer from clock feed-through. Also clock jitter problems can appear. All these non-idealities generate additional noise in the practical  $\Delta\Sigma$  modulator, on top of the quantization noise.

In this chapter we examine the effects of these sources of noise on the performance of the  $\Delta\Sigma$  modulator. Then we use these and previous results to compare the  $\Delta\Sigma$  modulators using SC to those using CT loop filters (LF).

### **6.2 Practical Circuit Limitations**

The implemented  $\Delta\Sigma$  modulator is subjected to certain limitations and this applies to both using SC and CT loop filters circuits. These limitations are caused by the non-ideal operation of the various components of the modulator, namely the integrators and the local converters especially the DAC. They may also be due to the jitter of clock pulse edges. More-

over in multistage modulators additional problem exists, namely, the mismatching between the analog and the digital part of the modulator.

The non-ideal operation of the integrators is due to the various limitations of the amplifiers such as finite gain, finite bandwidth, finite slew-rate as well as thermal and flicker noise. Noise can be also produced by the various switches in the SC circuit.

Problems also arise from the operation of the local DAC whose output levels have to be constant and, in the case of the multibit DAC, these levels have to be equidistant, i.e the DAC should be linear.

The result of all these limitations is the appearance of additional noise in the output  $\Delta\Sigma$  sequence, on top of the quantization noise that was examined in preceding chapters. However, although the quantization noise is shaped by the  $\Delta\Sigma$  loop filter, this does not happen in all cases of the noise generated by the above mentioned sources.

In what follows these noise sources and their effects are examined quantitatively and in some cases some remedies are suggested.

### 6.2.1 Noise sources in the $\Delta\Sigma$ loop

For reasons of clarity and better understanding we may redraw the block diagram of the  $\Delta\Sigma$  modulator with the various noise sources indicated on it. This is depicted in Fig. 6.1 in the case of a second-order  $\Delta\Sigma$  modulator.

Clearly, in the case of an ideal modulator the only noise source would be the quantization error  $e_q(n)$ . In a non-ideal  $\Delta\Sigma$  modulator for use in analog-to-digital conversion, there are additional noise sources shown in Fig. 6.1 which may be due to the following:

- External noise accompanying the input signal. This is included in the noise source  $e_1(n)$ .
- Thermal noise of the input stage of the integrators or flicker noise of their opamps, which is included in the noise sources  $e_1(n)$  and  $e_2(n)$ .
- Clock feed-through in the case of SC implementations also included in the noise sources  $e_1(n)$  and  $e_2(n)$ .

- d. Noise due to clock jitter. In the case of SC implementation, when the input signal is sampled, this noise is included in  $e_1(n)$ . In the case of CT implementation this type of noise is included in  $e_3(n)$  and  $e_{DAC}(n)$ .
- e. Noise due to the non-linear operation of the integrators. Due to finite slew-rate or clipping of the output of the opamps, this noise appears at the output terminal of the integrators and thus it is included in the noise sources  $e_2(n)$  and  $e_3(n)$ .
- f. Noise due to the non-linear operation of the local ADC included in the noise source  $e_3(n)$ .
- g. When the local DAC is multibit, as is the case of multibit  $\Delta\Sigma$  modulators, non-linear operation is equivalent to the noise source  $e_{DAC}(n)$ .

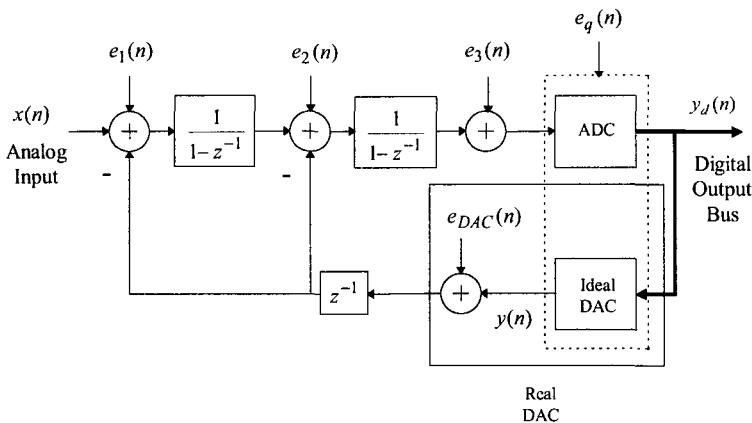


Fig. 6.1 Noise sources in a second-order  $\Delta\Sigma$  modulator for analog-to-digital conversion.

All these noise sources may affect the signal at the output of the system and consequently affect the performance of the  $\Delta\Sigma$  modulator. From Fig. 6.1, using z-transform we may write for the output  $y(n)$  of the modulator (corresponding to the digital words  $y_d(n)$  at the output of the real system) the following:

$$\begin{aligned} Y = X + E_1 - z^{-1}E_{DAC} + (E_2 - z^{-1}E_{DAC})(1-z^{-1}) \\ + (E_q + E_3)(1-z^{-1})^2 \end{aligned} \quad (6.1)$$

Clearly, the most critical noise sources are  $e_{DAC}(n)$  and  $e_1(n)$ , since they appear at the output not shaped. The noise  $e_2(n)$  is subjected to the first-order shaping, while the effect of  $e_3(n)$  is negligible, since it is subjected together with  $e_q(n)$  to second-order shaping and its power spectral density is usually much lower than that of  $e_q(n)$ .

Other type non-idealities like finite bandwidth and finite dc gain of the opamps, mismatching in multistage modulators, e.t.c. affect the NTF. To show their effect consider for example the two-stage  $\Delta\Sigma$  modulator in Fig. 6.2 with  $H_{a_1}$  and  $H_{a_2}$  being the transfer functions of the analog integrators and  $H_d$  the transfer function of the digital differentiator. Assuming ideal local DACs, the output  $y(n)$  of the modulator corresponding to the digital words  $y_d(n)$  is as follows:

$$\begin{aligned} Y = Y_1 - H_d Y_2 \\ = \frac{H_{a_1}}{1+z^{-1}H_{a_1}} X + \left( \frac{1}{1+z^{-1}H_{a_1}} - \frac{H_{a_2}H_d}{1+z^{-1}H_{a_2}} \right) E_{q_1} - \frac{H_d}{1+z^{-1}H_{a_2}} E_{q_2} \end{aligned} \quad (6.2)$$

Note that it is usual  $y_{d_1}(n)$  and  $y_{d_2}(n)$  to be single-bit as shown in Fig. 6.2. However,  $y_d(n)$  is not single-bit any longer because it is the sum of the sequences  $y_{d_1}(n)$  and  $y_{d_2}(n) - y_{d_2}(n-1)$  (see also example 3.4).

For ideal analog integrators,  $H_{a_1}(z)$  and  $H_{a_2}(z)$  will be equal to  $1/(1-z^{-1})$ . Then

$$Y = X - E_{q_2} \left( 1-z^{-1} \right)^2 \quad (6.3)$$

i.e. there is a second-order shaping.

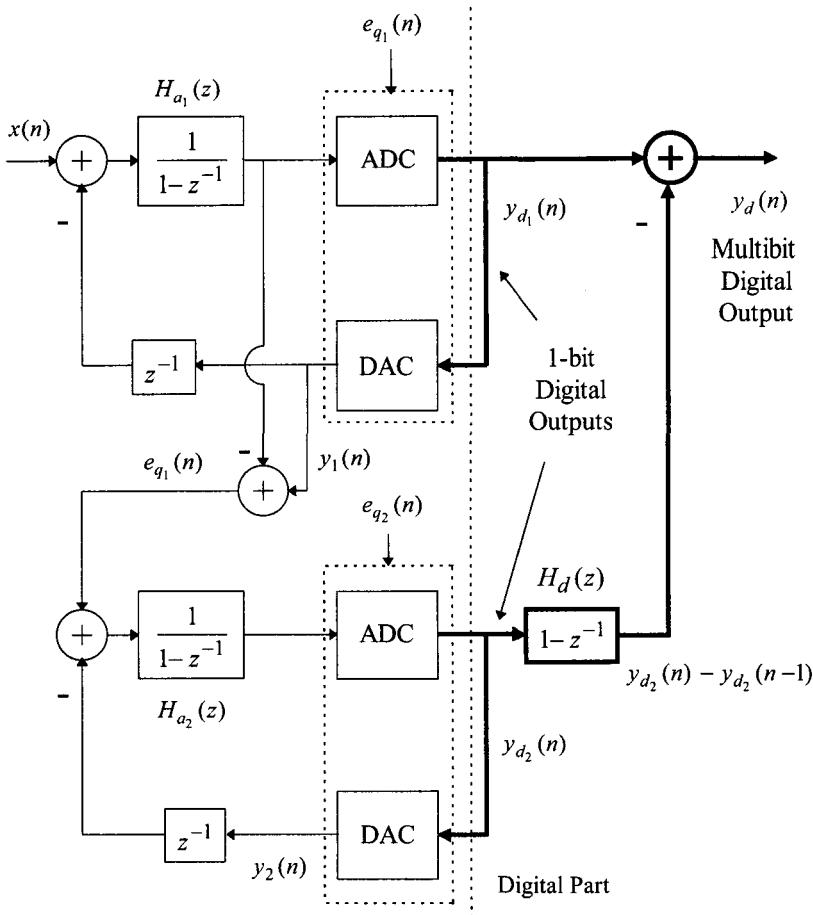


Fig. 6.2 Two-stage  $\Delta\Sigma$  modulator for analog-to-digital conversion consisting of two first-order  $\Delta\Sigma$  modulators.

Consider now the case when the analog integrators  $H_{a_1}(z)$  and  $H_{a_2}(z)$ , due to various imperfections, are not ideal. For simplicity assume that only the first integrator  $H_{a_1}(z) = g/(1-pz^{-1})$  is not ideal, i.e. its gain  $g$  and its pole  $p$  are not equal to unity. Then Eq. (6.3) will become as follows:

$$Y = \frac{gX}{1+(g-p)z^{-1}} + \left( \frac{1-pz^{-1}}{1+(g-p)z^{-1}} - (1-z^{-1}) \right) E_{q_1} - (1-z^{-1})^2 E_{q_2} \quad (6.4.a)$$

Assuming that  $g \approx 1$  and  $p \approx 1$ , to a first-order approximation, we may write Eq. (6.4.a) in the following form:

$$Y \approx X + \left( (1-g)(1-z^{-1})z^{-1} + (1-p)z^{-2} \right) E_{q_1} - (1-z^{-1})^2 E_{q_2} \quad (6.4.b)$$

Clearly the quantization noise  $E_{q_2}$  is subjected to second-order shaping. However, the gain error  $(1-g)$  causes a first-order shaped leakage of  $E_{q_1}$  and the pole error  $(1-p)$  an unshaped leakage of  $E_{q_1}$  to appear at the output of the modulator (see also problem 6.1).

### 6.3 Effect of Thermal Noise

To simplify the study of the effect of thermal noise we split this into two parts:

- We consider that the opamps are noiseless and we calculate the effect of thermal noise generated in the other components and
- We consider the effect of thermal and flicker noise in the opamps the rest of the circuit being assumed noiseless.

The latter is examined separately in Sec. 6.5.3.

#### 6.3.1 SC circuits

Consider the non-inverting SC integrator. In Fig. 6.3 it is assumed that thermal noise is generated in the resistance  $R_{on}$  of each switch (the opamp is assumed noiseless).

At the end of phase 1, i.e. just before switches 1 open, the capacitor  $C_s$  has been charged to the voltage  $x(nT_s) + e_1(nT_s) - e_4(nT_s)$ , where  $e_1$  and  $e_4$  are noise voltages due to noise sources  $n_1(t)$  and  $n_4(t)$ . At the end of phase 2, just before switches 2 open, the capacitor  $C_s$  has been

charged to the voltage  $e_3(nT_s + T_s/2) - e_2(nT_s + T_s/2)$ . Therefore the total charge that will be transferred to capacitor  $C_o$  will be

$$x(nT_s) + e_1(nT_s) - e_4(nT_s) - e_3(nT_s + T_s/2) + e_2(nT_s + T_s/2) \quad (6.5)$$

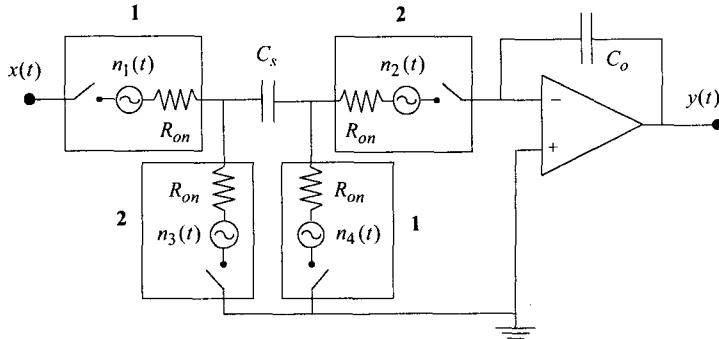


Fig. 6.3 Noise sources in a SC integrator.

Note that  $e_i(t)$  is the voltage across the capacitor  $C_s$  due to noise source  $n_i(t)$ ,  $i=1,2,3,4$  respectively. Each noise source  $n_i(t)$  has a PSD  $P(f)$  given by

$$P(f) = 4kTR_{on} \quad (6.6)$$

which is not bandlimited (goes up to THz). Consequently all  $e_i(t)$  have white spectrum almost up to the frequency  $f_1 = 1/2R_{on}C_s$ , which is much higher than the oversampling frequency  $f_s$ . On the other hand the frequency of  $x(t)$  is much lower than  $f_s$ . As a result  $x(t)$  is sampled without the occurrence of aliasing, while aliasing occurs when each  $e_i(t)$ ,  $i=1,\dots,4$  is sampled. The mean square value of each voltage  $e_i(t)$  across the capacitor  $C_s$  is as follows [1,2]:

$$E\{e_i(t)^2\} = \frac{1}{2\pi} \int_0^\infty 4kTR_{on} \cdot \left| \frac{1}{1+j2\omega R_{on} C_s} \right|^2 d\omega \quad (6.7.a)$$

$$= \frac{4kTR_{on}}{2\pi} \frac{1}{2R_{on}C_s} \tan^{-1}(2\omega R_{on} C_s) \Big|_{\omega=0}^{\omega=\infty} = \frac{kT}{2C_s} \quad (6.7.b)$$

Assuming that the various noise sources  $n_i(t)$  are independent of each other, then the total noise power will be

$$P_{thermal} = \{e_1^2\} + E\{e_2^2\} + E\{e_3^2\} + E\{e_4^2\} = \frac{2kT}{C_s} \quad (6.8)$$

The spectrum of the sampled thermal noise  $e_i(\kappa T_s)$  is white in the frequency band  $[-f_s/2, f_s/2]$  as is the spectrum of  $e_i(t)$ . Aliasing does not affect the shape of the PSD but only its magnitude. Also the power of the sampled thermal noise  $E\{e_i(\kappa T_s)^2\}$  equals the power  $E\{e_i(t)^2\}$ . Therefore the thermal noise power inside the signal frequency band will be

$$P_{th,in} = \frac{2kT}{C_s} \cdot \frac{1}{OSR} \quad (6.9)$$

OSR being the oversampling ratio. This equation explains why the value of the capacitor  $C_s$  of the first integrator should be large when the  $\Delta\Sigma$  modulator is to be produced in integrated form using SC circuits.

### 6.3.2 CT active RC circuits

In the CT integrator the thermal noise source is due to resistor  $R_1$ . The attenuation of this noise  $e(t)$  by this integrator up to the frequency  $f_s/2$  is high enough so as to assume that there will not be any aliasing effects due to its subsequent sampling in the local ADC. Then [3]

$$P_{th,in} = 4kTR_1 \frac{f_s}{2OSR} \quad (6.10)$$

Thus the thermal noise in the CT integrator is proportional to  $R_1 f_s$ , which dictates that the value of  $R_1$  should be small.

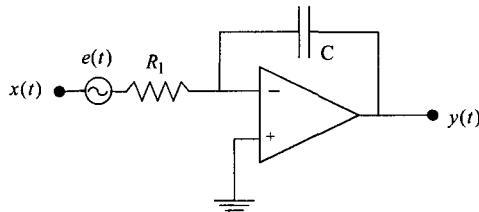


Fig. 6.4 Thermal noise source in a CT integrator.

## 6.4 Effects of the Opamp Non-Idealities

We examine these effects by studying their influence on the operation of the integrators, which are basic building blocks of the  $\Delta\Sigma$  modulators. The non-ideal behaviour of the opamp is mainly due to thermal and flicker noise, dc offset, finite open-loop voltage gain, limited bandwidth and finite slew-rate.

### 6.4.1 Thermal and flicker noise and DC offset of the opamp

In Fig. 6.5a a noise source, including thermal and flicker noise and dc offset of the opamp, is introduced assuming noiseless operation of the switches. The signal at the integrator output, assuming that this is sampled in phase 2, will be as follows:

$$Y(z) = -\frac{C_s}{C_o} \frac{1}{1-z^{-1}} X(z) + \left(1 - \frac{C_s}{C_o} \frac{1}{1-z^{-1}}\right) V_n(z) \quad (6.11)$$

It can be seen from Eq. (6.11) that the noise  $v_n$  and its integral appear at the output.

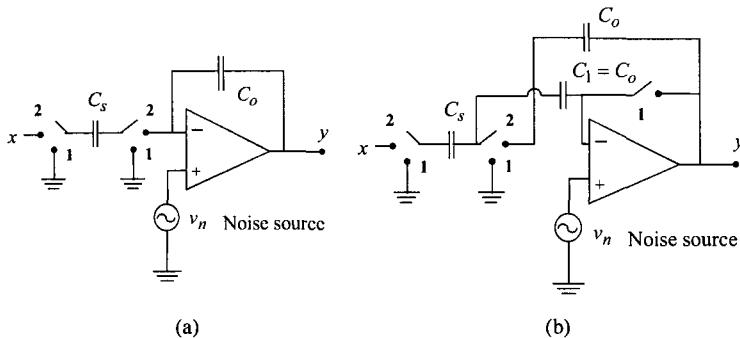


Fig. 6.5 a. Thermal and Flicker noise and dc offset in the integrator, b. Integrator using CDS.

Clearly the presence of flicker or  $1/f$  noise is significant at very low frequencies. Various techniques have been developed for its rejection, as well as the rejection of the dc offset. Most important of them are the following [4]:

- Chopper Stabilization, CHS.
- Autozeroing (AZ) and its special type Correlated Double Sampling, CDS.

The CDS technique, which is popular for SC circuits, can lead to the reduction of the undesirable effects of dc offset, flicker noise and non-infinite gain of the opamp on the operation of the integrator [4]. In Fig. 6.5b a simple example of an integrator using this technique, is shown [5]. Its operation is as follows:

In phase 1, which is the autozero (AZ) phase, the dc offset as well as the noise  $v_n$  are sampled in the capacitor  $C_1$ . In phase 2, when the input is sampled and integrated, the signal at the integrator output will be as follows (see problem 6.2):

$$Y(z) = -\frac{C_s}{C_o} \frac{1}{1-z^{-1}} X(z) + \left(1 - \frac{C_s}{C_o} \frac{1}{1-z^{-1}}\right) \left(1 - z^{-1/2}\right) V_n(z) \quad (6.12)$$

In this case the noise at the output is filtered by the factor  $(1 - z^{-1/2})$  which is high-pass. This is so because the noise is sampled twice in each sampling period. First it is sampled in the AZ phase, phase 1, and then it is sampled again in phase 2, when the input signal is sampled. Thus the noise sample in the AZ phase is subtracted from the noise sample in phase 2. As a result the dc offset can be totally cancelled, while the 1/f noise, the power of which is high mainly at very low frequencies, can be reduced significantly.

#### 6.4.2 Finite opamp gain

We turn now to the effect of the opamp finite open-loop gain on the operation of the integrator.

Due to finite gain of the opamp the transfer function of the SC integrator in Fig. 6.6a is changed to the following (as it was used in Sec. 6.2.1):

$$\frac{Y(z)}{X(z)} = \frac{C_s}{C_o} \cdot \frac{g}{1 - pz^{-1}} \quad (6.13)$$

The pole  $p$  lies inside the unit circle. This is due to the fact that it is not possible for the charge on the capacitor  $C_s$  to be fully transferred to  $C_o$  in each clock period, due to the fact that the latter is not fully discharged during the same period. Applying the charge conservation principle, it can be shown (see problem 6.3) that

$$p = \frac{1}{1 + \frac{C_s}{C_o} \frac{1}{1+A}} \quad (6.14)$$

and

$$g = \frac{1}{1 + \frac{1}{A} \left( \frac{C_s}{C_o} + 1 \right)} \quad (6.15)$$

where  $A$  is the dc open-loop gain of the opamp. Similarly in the case of the CT integrator

$$\frac{Y(s)}{X(s)} = \frac{-1}{R_1 C (1 + 1/A)s + 1/A} \quad (6.16)$$

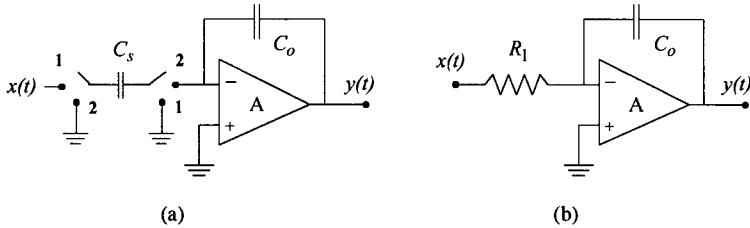


Fig. 6.6 a. Non-inverting SC integrator and b. the CT integrator.

Direct consequence of the modification of the integrator transfer function is that the NTF zeros have been moved inside the unit circle. This leads to a reduction of the attenuation of the quantization noise inside the signal-band and, therefore, to a reduction of the SNR. The effect of the finite dc gain to alter the realized NTF is not so severe in single-stage modulators, as these are quite robust to circuit non-idealities. To demonstrate this, a second-order modulator comprising the cascade of two integrators as shown in Fig. 5.11, each having the transfer function given by Eq. (6.13) and a third-order shown in Fig. 5.14 comprising the cascade of three such integrators, are simulated for various values of the error in the pole position  $p$ . From Eq. (6.14) it can be shown that the pole error is almost proportional to  $1/A$ . The resulting maximum SNR is presented in Fig. 6.7a. Both modulators are quite robust to this error and consequently to the variations of  $A$ , down to very extreme values.

The effect of finite opamp gain becomes more important when higher-order single-stage modulators are realized. Then the variation of the NTF can lead to unstable designs. Finally, as already discussed in Sec. 6.2, the effect of finite opamp gain is very significant in multi-stage modulators [6]. This is demonstrated in Fig. 6.7b.

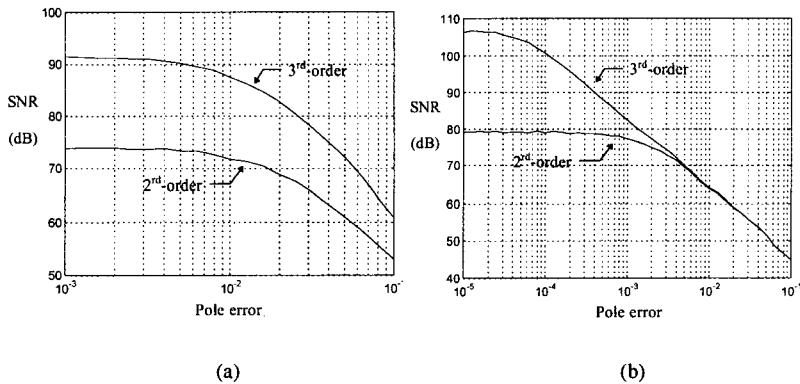


Fig. 6.7 SNR as a function of the error at the pole value. a. In the case of single-stage  $\Delta\Sigma$  modulators, b. in the case of multistage  $\Delta\Sigma$  modulators (cascade of two first-order modulators and cascade of three first-order modulators). The input is sinusoidal and the oversampling ratio is 64.

The technique of Autozeroing has been successfully used in order to improve the performance of the integrators in multi-stage  $\Delta\Sigma$  modulators [7], as well as that of single-stage  $\Delta\Sigma$  modulators [8].

#### 6.4.3 Finite bandwidth and slew-rate of the opamp

We assume that the open-loop gain  $A(s)$  of the opamp follows the single-pole model, i.e.

$$A(s) = \frac{A\omega_p}{s + \omega_p} \quad (6.17)$$

where  $A$  is the dc gain and  $\omega_p$  the pole frequency. It can then be easily shown that the transfer function of the integrator in Fig. 6.6b is as follows

$$\frac{Y(s)}{X(s)} = \frac{-A\omega_p}{R_1Cs^2 + (R_1C(A+1)\omega_p + 1)s + \omega_p} \quad (6.18)$$

Clearly, the operation of the CT integrator is not ideal. For the SC integrators the analysis of the effects of finite bandwidth in combination with the finite gain of the opamp is more complicated and has been studied in detail in [9,10]. Also various types of integrators have been proposed [11-13] in order to achieve a better performance.

In SC integrators the important parameter is the settling time  $\tau$  of the integrator, (assuming an integrator with exponential impulse response), which depends on the opamp bandwidth. In usual SC filters the unity gain bandwidth of the opamp has to be at least 10 times larger than the sampling frequency. However in  $\Delta\Sigma$  modulators no important problems occur when the settling time  $\tau$  and the sampling period  $T_s$  are of the same order [14]. When the settling time  $\tau$  is larger than the sampling period  $T_s$ , the modulator may become unstable as it can be shown by means of simulation [14].

As far as the limited slew-rate of the opamp is concerned, in the case of the CT integrators, this leads to harmonic distortion due to the resulting non-linear operation. In the case of SC integrators if proper settling is achieved within the sampling period, the effect of the slew-rate can be ignored. Otherwise linear operation of the integrator will be necessary in order to avoid harmonic distortion [14] and increase of the inband noise [1].

## 6.5 Effect of Jitter

The term jitter refers to the uncertainty of the time characteristics of a pulse, i.e. the moment of arrival of the pulse edges. The cause of jitter is the intrinsic noise generated by the components of the pulse generating circuit, such as crystals, transistors, resistors etc.

In our study of the effect of jitter in  $\Delta\Sigma$  modulation, we assume that the pulses have zero rise and fall times but there is an uncertainty  $\Delta t$  in the time of the arrival of the edges of the pulse, as shown in Fig. 6.8.

The effect of jitter is equivalent to noise generation at various parts of the circuit, as was explained in Sec. 6.2. However the result is different in the SC from that in CT circuits, where it is much more severe [15], as we explain below.

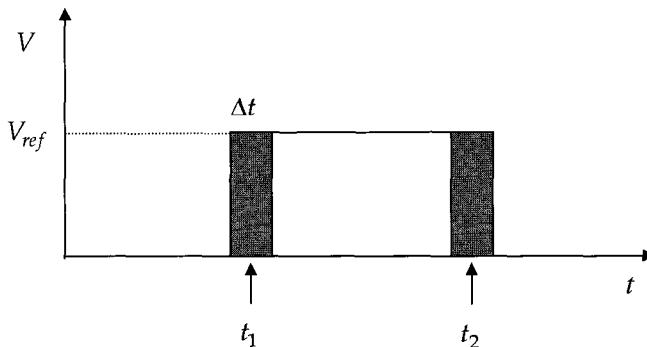


Fig. 6.8 Uncertainty in the timing of the pulse edges.

### 6.5.1 SC circuits

In SC circuits the main problem from jitter arises at the stage of sampling the signal, i.e. at the input of the modulator. Let us suppose that the signal  $x(t)$  is not sampled at the moment  $nT_s$  but at  $nT_s + \Delta t$ . Then the resulting error power will be [14]

$$\begin{aligned} P_{jitter} &= E\left\{\left(x(nT_s + \Delta t) - x(nT_s)\right)^2\right\} \cong E\left\{\Delta t^2 x'(nT)^2\right\} \\ &= E\left\{\Delta t^2\right\} E\left\{x'(nT_s)^2\right\} \end{aligned} \quad (6.19)$$

where  $x'(t)$  is  $dx(t)/dt$ . Clearly, there exists a strong dependence on the power of the derivative of the signal. For a sinusoidal input signal of amplitude  $x_o$  and frequency  $f$  we get

$$P_{jitter} = \frac{x_o^2}{2} (2\pi f)^2 E\left\{\Delta t^2\right\} \quad (6.20)$$

If the uncertainty  $\Delta t$  is uniformly distributed in the range  $[-\tau_{jitter}, \tau_{jitter}]$  Eq. (6.20) gives

$$P_{jit} = \frac{x_o^2}{2} (2\pi f)^2 \frac{\tau_{jit}^2}{3} \quad (6.21)$$

This noise enters the modulator together with the signal. Therefore it is not shaped. In the case of the sinusoidal input of frequency  $f_b = f_s/2R$ , the resulting inband noise power due to jitter will be the following:

$$P_{jit,in} = \frac{x_o^2}{2} \frac{(2\pi f_s/2R)^2}{R} \frac{\tau_{jit}^2}{3} = \frac{x_o^2}{2} \frac{\pi^2}{3R^3} \frac{\tau_{jit}^2}{T_s^2} \quad (6.22)$$

Then the ratio of the noise power due to jitter over the signal power will be

$$\frac{P_{jit,in}}{P_s} = \frac{\pi^2}{3R^3} \frac{\tau_{jit}^2}{T_s^2} \quad (6.23)$$

### 6.5.2 CT active RC circuits

In the CT active RC circuits the noise due to jitter is related to the uncertainty of the characteristics of the pulses of  $y_a(t)$ , i.e. the output pulses from the local DAC in Fig. 5.6 [3]. These pulses subsequently enter the integrator and are integrated with time constant  $R_1C$ . Assuming a pulse similar to that shown in Fig. 6.8, its integration will cause an error  $v_e$  at the output of the integrator given by:

$$v_e = \frac{V_{ref}}{R_1 C} \int_{t_1 + \Delta t_1}^{t_2 + \Delta t_2} dt - \frac{V_{ref}}{R_1 C} \int_{t_1}^{t_2} dt = \frac{V_{ref}}{R_1 C} (\Delta t_2 - \Delta t_1) \quad (6.24)$$

The total error at the output of the integrator, after the integration of the n-th pulse, will be given by the sum of all errors  $v_e(k)$ ,  $k=1,2,\dots,n$ , resulted from the integration of each one of the n pulses. The result is equivalent to that obtained if, instead of having the uncertainty in the pulse timing, we considered that there existed an error  $v_e$  in the ampli-

tude  $V_{ref}$  of each pulse. Thus, the effect of jitter in CT active RC circuits is equivalent to that of a noise source  $v_e(n)$ , which can be included in the noise source  $e_{DAC}(n)$  in Fig. 6.1, and appears at the output of the modulator not shaped.

If  $\Delta t$  is uniformly distributed in the range  $[-\tau_{jit}, \tau_{jit}]$  and assuming that  $\Delta t_1$  and  $\Delta t_2$  are independent variables, the mean value of the power of the noise  $v_e(n)$  will be

$$P_{jit} = E\{v_e(n)^2\} = \frac{V_{ref}^2}{(R_1 C)^2} \left( E\{\Delta t_1^2\} + E\{\Delta t_2^2\} \right) = \frac{2V_{ref}^2 \tau_{jit}^2}{3(R_1 C)^2} \quad (6.25)$$

Usually  $R_1 C$  is of the order of the oversampling period  $T_s$ . If we assume for simplicity that the spectrum of this noise is white, then the part of this noise in the signal band will be

$$P_{jit,in} = \frac{2V_{ref}^2}{3R} \frac{\tau_{jit}^2}{T_s^2} \quad (6.26)$$

Assuming also that the modulator remains stable up to the full-scale  $V_{ref}$ , the highest input signal power will be  $V_{ref}^2/2$ , if the signal is sinusoidal. Therefore the ratio of the power of noise due to jitter over the power of signal will be

$$\frac{P_{jit,in}}{P_s} = \frac{4}{3R} \frac{\tau_{jit}^2}{T_s^2} \quad (6.27)$$

Comparing Eqs. (6.23) and (6.27), it can be seen, that the SC implementation is advantageous as far as the effect of jitter is concerned, since the above power ratio, when the oversampling ratio R increases, is reduced for the SC circuit faster than for the CT circuit. To see what this means in practice, consider the following example. Assuming R=64, for

the  $\Delta\Sigma$  modulator to achieve a maximum SNR of 96dB the requirements will be as follows:

a. SC implementation

$$\frac{\pi^2 \tau_{jit}^2}{3 \cdot 64^3 T_s^2} \leq \frac{1}{2^{32}} \Rightarrow \frac{\tau_{jit}}{T_s} \leq \frac{1}{250} \quad (6.28)$$

b. Active RC implementation

$$\frac{4\tau_{jit}^2}{3 \cdot 64 T_s^2} \leq \frac{1}{2^{32}} \Rightarrow \frac{\tau_{jit}}{T_s} \leq \frac{1}{9450} \quad (6.29)$$

This means that in the case of the CT circuit the uncertainty of the clock pulses should be 40 times lower than in the case of the SC implementation.

## 6.6 Effect of Rise and Fall Times of DAC Pulses in CT Circuits

Non-zero rise and fall times of the DAC pulses create problems in CT  $\Delta\Sigma$  modulators. This is explained here as follows:

Assume that at a certain time period the waveform of the local DAC of a  $\Delta\Sigma$  modulator corresponds to the pulse sequence -1, 1, 1. This waveform will be approximately as shown in Fig. 6.9. Clearly, the transition from -1 to 1 takes some time depending on the rise time of the pulse. However the transition from 1 to 1 does not take any time at all. Thus all positive pulses (and this is true also for all negative pulses) have not the same shape. Consequently, the result of their integration by an analog CT integrator will not always be the same. It will depend on their sequence [15]. For example, the integral of the pulse sequence -1, 1, 1 will give a different result from that of the sequence 1, -1, 1. Such a problem does not exist in a SC integrator, where the result at its output depends only on the final value of each pulse. This is also true in the case of the digital processing of the  $\Delta\Sigma$  sequence.

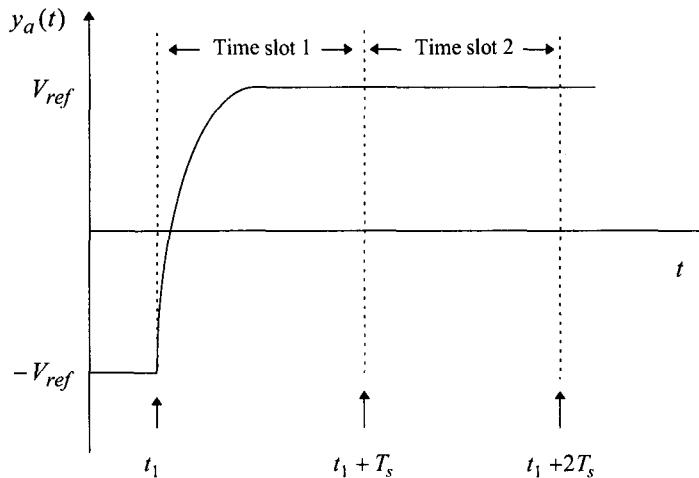
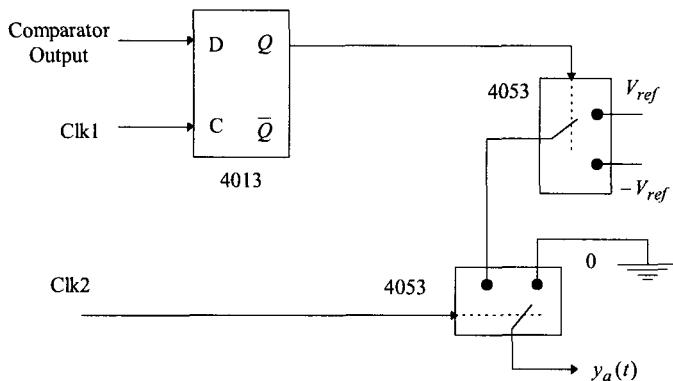
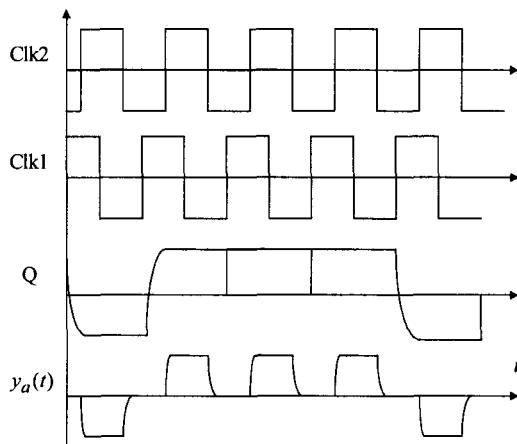


Fig. 6.9 Pulses at the output of the local DAC of a CT  $\Delta\Sigma$  modulator.

The effect of rise and fall times in CT circuits is similar to that of jitter, since the type of rising and falling edges of the pulses appearing randomly is equivalent to noise during the integration of the analog sequence  $y_a(t)$ . The simplest way to avoid this problem is to reduce the duration of the pulses and force them to return to '0' during the rest part of the period  $T_s$ . This technique is called 'return to zero (RZ)' [6,15,16] and is applied to CT  $\Delta\Sigma$  Modulators. The modified pulse waveform is shown in Fig. 5.6, while a simple circuit for achieving this is shown in Fig. 6.10 together with the corresponding waveforms. This was used in the CT  $\Delta\Sigma$  modulators, which were presented in Chapter 5. As it is shown in Fig. 6.10, a second clock Clk2 is used to trigger the D Flip-Flop having a  $T_s/4$  phase difference from Clk1. When the leading edge of Clk2 pulse arrives, the voltages at the outputs  $Q$  and  $\bar{Q}$  of the D Flip-Flop have already been fully settled. Thus all pulses in the analog waveform  $y_a(t)$  will have the same rise and fall times, which depend only on the corresponding times of Clk2 and the speed of the switches.



(a)



(b)

Fig. 6.10 a. A circuit to produce pulses in order to avoid the problem of unequal rise and fall times, b. Timing chart of the various signals.

### 6.7 A Comparison of SC and CT Active RC Circuits

Based on the contents of the previous sections, we summarize here in Table 6.1 some of the important characteristics, advantages and disadvantages of SC and CT active RC circuits.

vantages, of  $\Delta\Sigma$  modulators implemented as SC and as CT circuits. More thorough and detailed discussion on this comparison can be found in Refs. [6,15].

Table 6.1 Comparison of SC and CT  $\Delta\Sigma$  Modulator implementations

Issue	SC	CT
Simulation	Easy	Not easy
Breadboarding	Difficult <sup>1</sup>	Easy
Loop Filter scaling with clock frequency	Yes	No
Jitter	Ininsensitive	Sensitive
Compatibility with VLSI CMOS process	Yes	Not so compatible
High SNR	Reduced by capacitor size	Not limited by capacitor size
Opamp Requirements	Low settling-time	Low noise
Capacitors	Accurate capacitance ratios	Large values
Resistors	Not required	Linear high values <sup>2</sup>
Digital noise pick up	Prone	Not much prone

<sup>1</sup> The required typical values ( $<1\text{pF}$ ) are of the order of the parasitic capacitances.

<sup>2</sup> Laser trimming is required for accurate RC time constants in monolithic designs

## 6.8 DAC Errors

Problems can occur in  $\Delta\Sigma$  modulators because of the non-ideal operation of the local DAC. These depend on whether the  $\Delta\Sigma$  modulator is single-bit (when single-bit DAC is used) or multibit (when the local DAC is multibit). We examine both cases separately here below.

### 6.8.1 Single-bit DAC

We consider the single-bit DAC implemented the way shown in Figs. 5.7, 5.8 and 6.10a. The most probable error in the implementation of this DAC occurs if its two output voltage levels are not exactly  $-V_{ref}$  and  $+V_{ref}$ . Let them be  $-V_{ref} + v_1$  and  $V_{ref} + v_2$ . We may write these values as follows:

$$-V_{ref} - (v_2 - v_1)/2 + (v_2 + v_1)/2 = -V'_{ref} + v \quad (6.30.a)$$

and

$$V_{ref} + (v_2 - v_1)/2 + (v_2 + v_1)/2 = V'_{ref} + v \quad (6.30.b)$$

where

$$V'_{ref} = V_{ref} + (v_2 - v_1)/2 = gV_{ref} \quad (6.31.a)$$

$$v = (v_1 + v_2)/2 \quad (6.31.b)$$

The effect of these errors can be better understood by means of Fig. 6.11a. In Fig. 6.11b the equivalent system is shown of a  $\Delta\Sigma$  modulator using a non-ideal DAC. The input of the quantizer in the two systems of Figs. 6.11a and 6.11b is exactly the same. However, as has been explained in Chapter 4, the output of a single-bit ADC is invariant to any scaling of its input. Therefore the gain  $g$  of the integrator can be ignored in Fig. 6.11b. Consequently the results of these errors are a. the appearance of a dc offset  $-v/g$  at the output of the modulator and b. the attenuation or amplification of the input signal depending on whether the factor  $g = V'_{ref}/V_{ref}$  is larger or smaller than 1 respectively. Both of these results do not create serious problems, since the  $\Delta\Sigma$  modulator is not usually used for encoding dc signals, while the multiplication of the input signal by  $1/g$  does not affect the spectrum of the signal.

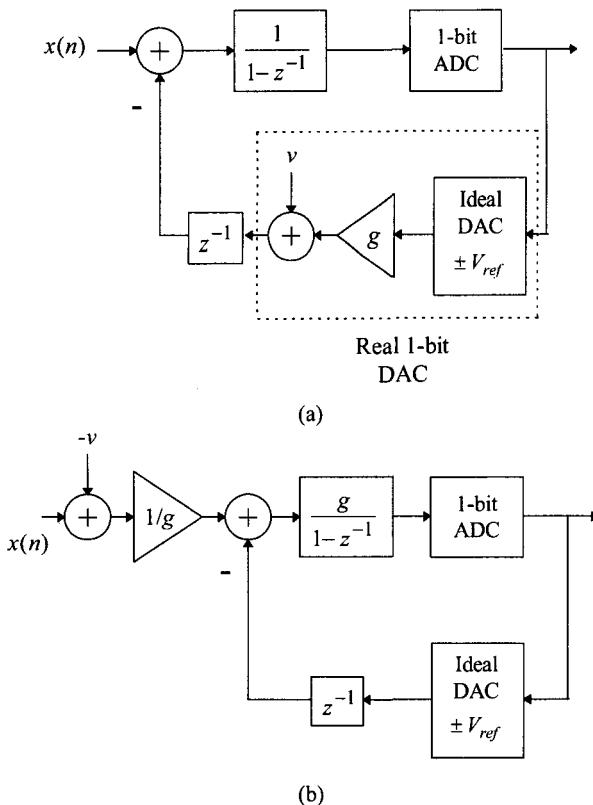


Fig. 6.11 a. Single-bit  $\Delta\Sigma$  Modulator with non-ideal DAC, b. Equivalent system.

### 6.8.2 Multibit DAC

In the case of a multibit  $\Delta\Sigma$  modulator, both the local ADC and DAC are of M-bit and the results are different. The states of an ideal M-bit DAC are  $L_m = (m + 0.5)V_{step}$  with  $m = -2^{M-1}, \dots, 0, 1, \dots, (2^{M-1} - 1)$ , when it is of the even type and  $L_m = mV_{step}$  with  $m = -(2^{M-1} - 1), \dots, 0, 1, \dots, (2^{M-1} - 1)$ , when it is of the odd type.  $V_{step}$  is the voltage step of the DAC output, i.e. the voltage difference between any two successive states. Again in

the implementation, as in the single-bit DAC, an error will appear in the level of the states which will be  $V_m = L_m + v_m$ . Here though  $v_m$  is a random error, which does not allow the assumption that the levels are multiplied by a constant and shifted equally up or down. Simulation shows that when the states are not equally spaced, apart from a dc offset, there appear harmonic distortion, undesirable tones, as well as noise, the spectrum of which overlaps with the shaped quantization noise. All these lead to a degradation of the operation of the  $\Delta\Sigma$  modulator. Why the non-ideality is so serious can be explained by means of the introduction of a noise source  $e_{DAC}$ , as shown in Fig. 6.1.

The most frequently used structure for the implementation of a DAC with increased linearity is based on the use of Unit Elements [6]. These elements may be current sources, resistors, capacitors, etc. An example of such a DAC implementation is shown in Fig. 6.12.

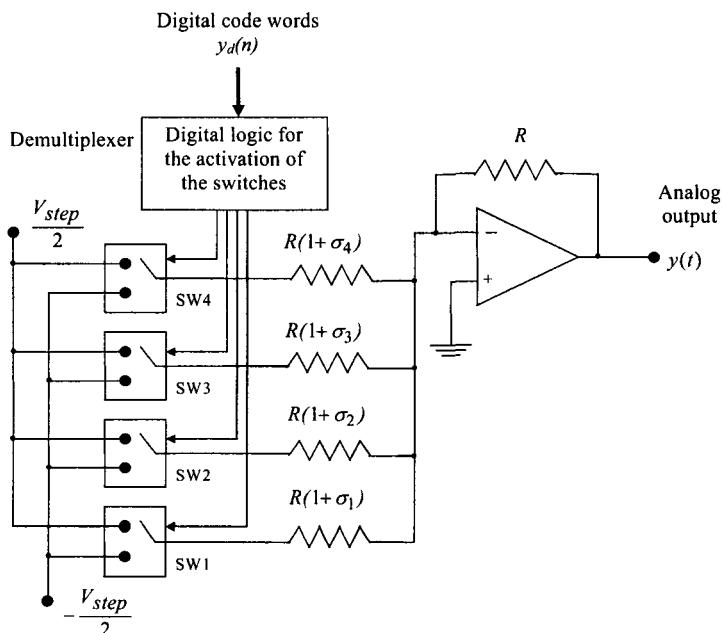


Fig. 6.12 DAC implementation using resistors as unit elements.

The main characteristic of this topology is the relatively lower sensitivity to the errors in the values of the various elements, which leads to increased linearity. In Fig. 6.12 four unit elements (resistors) are used, the values of which should be matched, for a 5 level DAC.

If the values of the unit elements are not matched, the levels of the DAC will not be equally spaced (see problem 6.4) and thus the DAC will not operate linearly. Although this structure of the DAC is less sensitive than others to the errors of the element values, in the case of high accuracy  $\Delta\Sigma$  modulators of the problem still remains.

The problems caused by the DAC non-linearities can be resolved by various methods which fall into the following categories:

- a. Element-Trimming Approaches [6]
- b. Digital Correction techniques [6,17,18]
- c. Self-Calibration techniques [19,20]
- d. Dynamic Element Matching. This method is quite attractive and various techniques have been developed for its applications, namely
  - i. Dynamic Element Randomization [6,21]
  - ii. Dynamic Element Rotation-Barrel Shifter [6]
  - iii. Individual Level Averaging [6,22,23]
  - iv. Noise-Shaped Element Usage [6,24-28]

We do not intend to elaborate on these methods here. The interested reader can consult the relevant references given above.

## 6.9 Summary

In this chapter we reviewed the effects of various circuit component non-idealities on the performance of  $\Delta\Sigma$  modulators. Such non-idealities refer primarily to the opamps and are their finite gain, bandwidth and slew-rate. The finite opamp gain creates a more serious problem in the multistage than in the single-stage modulators. Also problems arise from the various types of noise, namely thermal and flicker noise as well as the clock jitter. The thermal noise constitutes a problem, especially for the SC circuits, while the clock jitter is a more important problem for the CT circuits. Also problems arising from the component inaccuracies, like

the resistor or capacitor values of the local DAC, may degrade the performance of multibit modulators in particular. Multistage modulators are sensitive to the various errors in the component values as well as to mismatching. In general, one may conclude, that single-bit single-stage, modulators are more robust against the various errors in the component values, which justifies their wide use.

### Problems

- 6.1 Determine the effect of mismatching when the second integrator in Fig. 6.2 is not ideal. Repeat when the output of the first DAC is subtracted from the output of the first integrator each of them being multiplied by a non-unity coefficient when feeding the summing node  $e_{q_1}$  in the second stage
- 6.2 Prove Eq. (6.12).
- 6.3 Prove Eqs. (6.14) and (6.15).
- 6.4 Determine the values of the DAC levels in Fig. 6.12. Calculate the mean value of the error and its variance in each state. Generalize in the case of M unit elements.

## References

- [1] V. F. Dias, G. Palmisano, P. O' Leary, F. Maloberti, 'Fundamental Limitations of Switched-Capacitor Sigma-Delta Modulators', *IEE Proceedings-G*, Vol. 139, No 1, February 1992, pp. 27-32.
- [2] S. Rabii, B. A. Wooley, *The Design of Low-Voltage Low-Power Sigma-Delta Modulators*, Boston, Kluwer Academic Publishers, 1999.
- [3] V. F. Dias, G. Palmisano, F. Maloberti, 'Noise in Mixed Continuous-Time Switched-Capacitor Sigma-Delta Modulators', *IEE Proceedings-G*, Vol. 139, No 6, December 1992, pp. 680-684.
- [4] C. C. Enz and G. C. Temes, 'Circuit Techniques for Reducing the Effects of Op-Amp Imperfections: Autozeroing, Correlated Double Sampling and Chopper Stabilization', *Proceedings of the IEEE*, Vol. 84, No 11, November 1996, pp. 1584-1616.
- [5] K. K. Lam and M. A. Copeland, "Noise-Cancelling Switched-Capacitor (SC) Filtering Technique", *Electronics Letters*, Vol. 19, No 20, September 1983, pp. 810-811.
- [6] S. R. Norsworthy, R. Schreier and G. C. Temes, *Delta-Sigma Data Converters, Theory, Design and Simulation*, New York, IEEE Press, 1997.
- [7] M. Rebescini, N. R. Van Bavel, P. Rakers, R. Greene, J. Caldwell and J. R. Haug, 'A A 16-b 160kHz CMOS A/D Converter Using Sigma-Delta Modulation', *IEEE Journal of Solid-States Circuits*, Vol. 25, No 2, April 1990, pp. 431-440.
- [8] P. J. Hurst, R. A. Levinson and D. J. Block, .A Switched-Capacitor Delta-Sigma Modulator with Reduced Sensitivity to Op-Amp Gain', *IEEE Journal of Solid-States Circuits*, Vol. 28, No 6, June 1993, pp. 691-696.
- [9] G. C. Temes, 'Finite Amplifier Gain and Bandwidth Effects in Switched-Capacitor Filters', *IEEE Journal of Solid-States Circuits*, Vol. 15, No 3, June 1980, pp. 358-361.
- [10] K. Martin and A. S. Sedra, 'Effects of the Op Amp Finite Gain and Bandwidth on the Performance of Switched-Capacitor Filters', *IEEE Transactions on Circuits and Systems*, Vol. 28, No 8, August 1981, pp. 822-829.
- [11] K. Nagaraj, K. Singhal, T. R. Viswanathan, and J Vlach, 'Reduction of Finite Gain Effect in Switched-Capacitor Filters', *Electronics Letters*, Vol. 21, No 15, July 1985, pp. 644-645.
- [12] K. Haug, F. Maloberti, G. C. Temes, 'Switched-Capacitor Integrators with Low Finite-Gain Sensitivity', *Electronics Letters*, Vol. 21, No 24, November 1985, pp. 1156-1157.
- [13] K. Nagaraj, T. R. Viswanathan, K. Singhal and J Vlach, 'Switced-Capacitor Circuits with Reduced Sensitivity to Amplifier Gain', *IEEE Transactions on Circuits and Systems*, Vol. 34, No 5, May 1987, pp. 571-574.
- [14] B. E. Boser, Bruce A. Wooley, 'The Design of Sigma-Delta Modulation Analog-to-Digital Converters', *IEEE Journal of Solid-State*, Vol. 23 No. 6, December 1988, pp. 1298-1308.

- [15] R. Adams, ‘Design Aspects of High-Order Delta-Sigma A/D Converters’, Chapter 4.3 in *Circuits and Systems Tutorials*, IEEE ISCAS’ 1994, pp. 235-260, London, UK.
- [16] B. P. Del Signore, D. A. Kerth, N. S. Sooch and E. J. Swanson, ‘A Monolithic 20-b Delta-Sigma A/D Converter’, *IEEE Journal of Solid-State Circuits*, Vol. 25, No. 6, December 1990, pp. 1311-1316.
- [17] M. Sarhang-Nejad, G. C. Temes, ‘A High-Resolution Multibit  $\Sigma\Delta$  ADC with Digital Correction and Relaxed Amplifier Requirements’, *IEEE Journal of Solid-State Circuits*, Vol. 28, No. 6, June 1993, pp. 648-660.
- [18] Tzi-Hsiung Shu, Bang-Sup Song, K. Bacrania, ‘A 13-b 10-Msample/s ADC Digitally Calibrated with Oversampling Delta-Sigma Converter’, *IEEE Journal of Solid-State Circuits*, Vol. 30, No. 4, April 1995, pp. 443-452.
- [19] J. W. Fattaruso, S. Kiriaki, M. de Wit, G. Warwar, ‘Self-Calibration Techniques for a Second-Order Multibit Sigma-Delta Modulator’, *IEEE Journal of Solid-State Circuits*, Vol. 28, No. 12, December 1993, pp. 1216-1223.
- [20] R. T. Baird, T. S. Fiez, ‘A Low oversampling Ratio 14-b 500 kHz  $\Delta\Sigma$  ADC with a Self-Calibrated Multibit DAC’, *IEEE Journal of Solid-State Circuits*, Vol. 31, No. 3, March 1996, pp. 312-319.
- [21] L. R. Carley, ‘A Noise-Shaping Coder Topology for 15+ Bit Converters’, *IEEE Journal of Solid-State Circuits*, Vol. 24, No. 2, April 1989, pp. 267-273.
- [22] B. H. Leung, S. Sutarja, ‘Multibit  $\Sigma\Delta$  A/D Converter Incorporating A Novel Class of Dynamic Element Matching Techniques’, *IEEE Trans. on Circuits and Systems-II: Analog and Digital Signal Processing*, Vol. 39, No. 1, January 1992, pp. 35-51.
- [23] Feng Chen, B. H. Leung, ‘A High Resolution Multibit Sigma-Delta Modulator with Individual Level Averaging’, *IEEE Journal of Solid-State Circuits*, Vol. 30, No. 4, April 1995, pp. 453-460.
- [24] R. Schreier and B. Zhang, ‘Noise Shaped multi-bit D/A Converter Employing Unit Elements’, *Electronics Letters*, Vol. 31, No. 20, Sept. 1995, pp. 1712-1713.
- [25] A. Yasuda and H. Tanimoto, ‘Noise Shaping Dynamic Element Matching Method Using Tree Structure’, *Electronics Letters*, Vol. 33, No. 2, January 1997, pp. 130-131.
- [26] A. Keady and C. Lyden, ‘Tree Structure for Mismatch Noise-Shaping Multibit DAC’, *Electronics Letters*, Vol. 33, No. 17, August. 1997, pp. 1431-1432.
- [27] I. Galton, ‘Spectral Shaping of Circuit Errors in Digital-to-Analog Converters’, *IEEE Trans. on Circuits and Systems-II: Analog and Digital Signal Processing*, Vol. 44, No. 10, October 1997, pp. 808-817.
- [28] G. I. Bourdopoulos, and T. L. Deliyannis, ‘High-Order Vector Sigma-Delta Modulators’, *IEEE Transactions on Circuits and Systems - II Analog and Digital Signal Processing*, vol. 47, No 6, June 2000, pp. 493-503.

## Chapter 7

# Stabilization and Suppression of Tones for Higher-Order Single-Stage $\Delta\Sigma$ Modulators

### 7.1 Introduction

It was demonstrated in Sec. 3.6, that higher-order single-stage  $\Delta\Sigma$  modulators become unstable as their input approaches the full-scale of the quantizer. The input amplitude for the onset of instability is increased if the quantizer is multi-bit, but nevertheless excessive input can lock the modulator in unstable limit cycles. Even if the input is constrained, noise can drive the modulator to instability. Also, some modulators are unstable for a set of initial conditions. In Secs. 7.2 to 7.4, four stabilization methods are presented and compared. Fortunately, one of them, namely clipping, occurs at the supply voltage of the integrators in any  $\Delta\Sigma$  modulator realization, but other methods can have superior performance.

The single-bit  $\Delta\Sigma$  sequence exhibits a repetitive nature if the input amplitude is small and rational. Thus tones exist in the output of a  $\Delta\Sigma$  modulator. Such tones have been encountered in the analysis of the first-order  $\Delta\Sigma$  modulator, in Sec. 3.2. Although the overall SNR can be very high, these tones are troublesome, as they are perceived by the human ear. In Secs. 7.5 to 7.7 the tonal behavior of  $\Delta\Sigma$  modulators is examined and the various strategies to suppress the tones are presented and compared. Fortunately, in ADCs the thermal noise helps to overcome the tonal problem, which needs addressing only in DACs.

## 7.2 Bounds on Quantizer Input Using the Variable Gain Method

The instability of  $\Delta\Sigma$  modulators can be accounted for analytically. The use of the variable gain method provides confident bounds on the input to the quantizer [1]. According to this method, every non-linearity can be substituted by a variable gain  $K$ . The value of the gain changes to the ratio of the output of the non-linearity over its input. Thus the system is transformed into a linear one with a variable parameter  $K$ . A transfer function of the system can be derived in the usual way for linear systems. For sampled-data systems, this function is of the form  $H(z, K)$ . If the poles of  $H(z, K)$  are kept inside the unit circle for all  $K$ , then the system is stable. When a pole is moved outside the unit circle for a particular  $K$ , then the system is no longer stable in the Bounded-Input-Bounded-Output sense (considering the quantizer input as the output of the system, as the  $\Delta\Sigma$  is always bounded). Thus the signal levels in the system tend to increase in an uncontrolled fashion. This results in a change in  $K$ . If this change is such as to move the pole back inside the unit circle, then there exists a stable limit cycle. If on the other hand the pole is kept outside of the unit circle, then the limit cycle is unstable. The system can lock itself into it and become unstable.

To use the variable gain method on  $\Delta\Sigma$  modulators [1,2], the range of the gain  $K$  is calculated. The non-linearity is a 1-bit quantizer, thus the gain is the inverse of the absolute value of the input to the quantizer (if the quantizer output is normalized to  $\pm 1$ ). Thus  $K \in (0, +\infty]$ . The usage of the method involves the calculation of the transfer function of the system and then its poles as a function of  $K$ . Stability is investigated from the pole loci when  $K$  takes all the values in its allowed range.

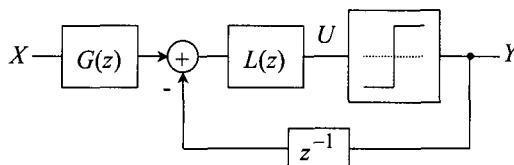


Fig. 7.1 A general  $\Delta\Sigma$  modulator block diagram.

Hence replacing the quantizer in the block diagram of a  $\Delta\Sigma$  modulator shown in Fig. 7.1 by the variable gain  $K$  yields the following transfer function:

$$H(z) = \frac{K \cdot L(z)G(z)}{1 + z^{-1}K \cdot L(z)} \quad (7.1)$$

From (7.1), the pole loci for  $K \in (0, +\infty]$  can be plotted.

Using the variable gain method, the stability of some  $\Delta\Sigma$  modulators is investigated. The results are depicted in Fig. 7.2.

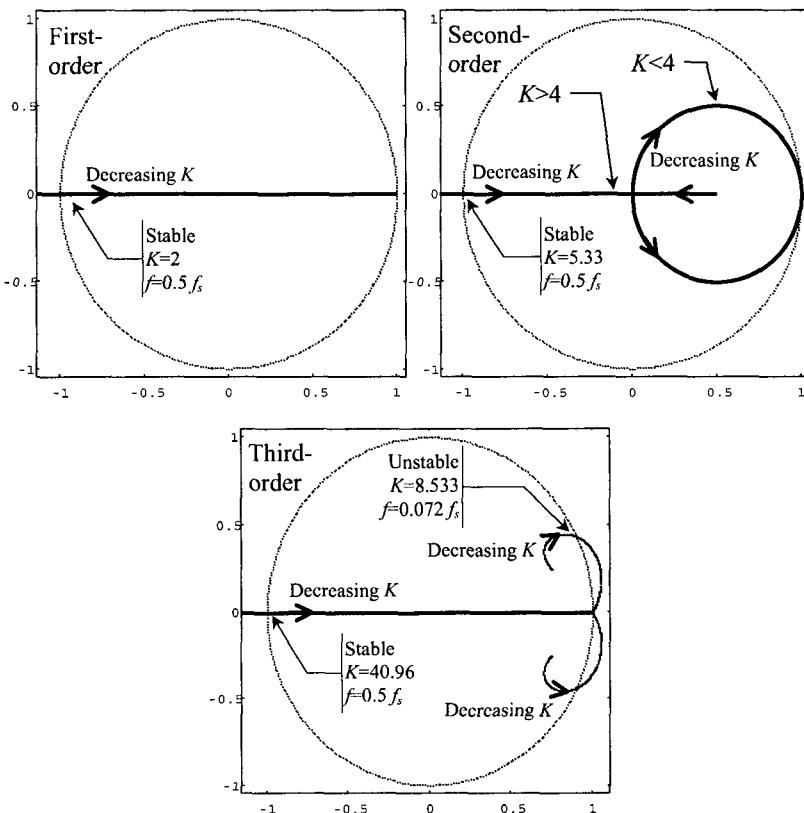


Fig. 7.2 Application of the variable gain method for the quantizer input to  $\Delta\Sigma$  modulators.

The first- and second-order  $\Delta\Sigma$  modulators are shown to be inherently stable, as only a stable limit cycle exists at half the sampling frequency. This limit cycle, termed idling pattern, occurs in modulators of any order when the amplitude of the input to the quantizer is very small. The excitation of the idling pattern is responsible for the idle channel noise of the  $\Delta\Sigma$  modulator [3]. On the other hand, higher-order single-bit single-stage modulators can become unstable.

Using the variable gain method on  $\Delta\Sigma$  modulators, a minimum value of the gain  $K_{\min}$  for stable operation is derived which corresponds to a maximum absolute value at the input to the quantizer,  $u_{\max}$

$$u_{\max} = \frac{1}{K_{\min}} \quad (7.2)$$

This value can be used as a safe bound for the state of the last integrator of the loop filter.

### 7.3 Stabilization Methods

All the methods to stabilize the higher-order single-stage  $\Delta\Sigma$  modulators are based on the detection of states that will possibly lead to instability. These states can be either the excessive values stored in the integrators that comprise the loop filter, or long sequences of identical outputs, the latter being evidence of a large amplitude, low frequency limit cycle at the quantizer input. Hence a detection threshold is set for any of the before mentioned quantities. Upon exceeding it, some action is triggered that stabilizes the modulator. The objectives of any stabilization method are the following:

- a. Recovery, i.e. return to normal operation should the destabilizing input be reduced. The recovery should be fast, in order for the stabilization method to be deactivated and the modulator to be allowed to return to normal operation. This is very important as the modulator always yields higher SNR when it operates normally.
- b. Ease of implementation.
- c. Insensitivity of the method to circuit imperfections, like finite opamp gain.
- d. Limited loss in SNR for the stable input range. As stabilization

actions change the behavior of the modulator, the performance of the modulator might be degraded, should these actions be activated before the onset of instability.

- e. Increase of the SNR of the unstable input range. This is the stabilization gain achieved by the method.
- f. Ease of determination of the detection threshold, which should not vary when different types of input signals are used. It can be analytically calculated using, for example, the variable gain method outlined in the previous section.

### ***7.3.1 Resetting the integrators***

One stabilization method involves the reset of the integrators of the loop filter [4,5]. The instability detection can be based on monitoring the signals in the integrators, or the output states. The latter is found more convenient, thus upon detection of  $n$  same  $\Delta\Sigma$  sequence bits all the integrators are reset and instability is prevented. The method leads to instant recovery from the overload state as, immediately after resetting, the modulator operates normally. Unfortunately, when resetting the integrators, the correlation of the oversampled input samples is lost [4], thus when the method is activated, the SNR is degraded significantly. Variations of the method, like resetting only the last integrator, does not improve matters, as the high signal levels at the rest of the integrators soon overload the highest-order. Any gain due to keeping some of the information in the integrators is lost due to the more frequent need for resetting.

### ***7.3.2 Clipping the integrators***

A second stabilization method involves clipping the integrators to some limiting values [5,6]. It is the easiest method to implement, as integrators are built using opamps, and they clip at their supply voltage. When ideal opamps are used, this method performs very well in terms of SNR. But when a real opamp is saturated, the inverting input is no longer a virtual ground [1]. Time is needed for this voltage to fall back to zero, leading to degradation of the modulator performance. This time reduces the speed of return to normal operation, should the input amplitude be reduced.

Also, as MOS devices are not good limiters [5], the clipping stabilization method is not suited for switched-capacitor  $\Delta\Sigma$  modulators in CMOS.

### 7.3.3 Activation of local feedback loops around the integrators

A third method involves the activation of local feedback loops around the integrators upon the detection of a state that may lead to instability [4,7]. These loops, when activated, decrease the signals stored in the integrators, thus stabilizing the modulator. The block diagram of such a third-order  $\Delta\Sigma$  modulator used in [4,7] is shown in Fig. 7.3.

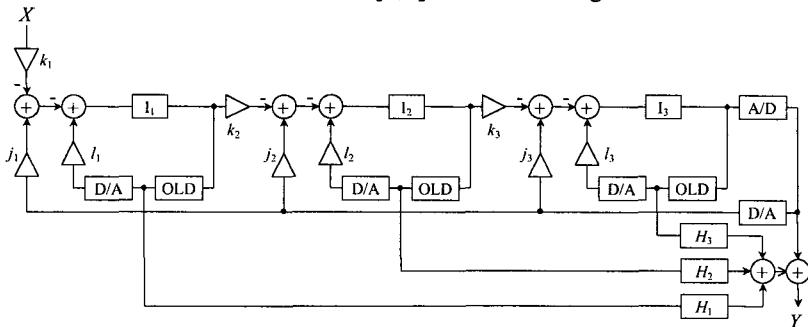


Fig. 7.3 Third-order  $\Delta\Sigma$  modulator stabilized using local feedback loops around the integrators.

The modulator is based on the CIDF structure. The hardware involved is quite complicated as the instability detectors (OLD, Over-Load Detector) in the local feedback loops involve tri-state ADCs and DACs. Also, digital correction filters  $H_i$ ,  $i=1, 2, 3$  are used to counter the effect of the loops at the output. This leads to the problem of matching the digital and analogue parts, already discussed regarding multi-stage  $\Delta\Sigma$  architectures in Secs 3.7 and 6.2.1. Especially the loop around the first integrator leaks non-shaped noise and heavily degrades the performance of the modulator [7]. In order for finite opamp gain not to degrade the SNR as much as in the multi-stage modulators, the loops should rarely be activated. This is achieved in the design proposed in [4,7], but is not the case for any modulator design, like those of Chapter 4. Another disadvantage of this method is that the CIDF loop structure used allows only unity zeros, and so the optimized placement of zeros is impossible. Applying the method

to  $\Delta\Sigma$  architectures that allow non-unity zeros (CIDIDF or CIDIFF) leads to IIR digital correction filters, whose stability is not guaranteed. Finally, the output of the modulator due to the digital correction filters is multi-bit, 9-bit for the modulator proposed in [4], making the implementation of the decimator difficult.

### 7.3.4 Reducing the order of the loop filter

According to the fourth method of reducing the order of the loop filter [8,9], upon detection of a state in the highest-order integrator that can lead to instability, this integrator is cut-off from the loop filter. Thus the order of the modulator is reduced, and the resulting modulator has a wider range of input amplitude for stable operation, i.e. it is more stable. Also, a third-order  $\Delta\Sigma$  modulator is reduced to a second-order one, which is always stable. A third-order  $\Delta\Sigma$  modulator based on the CIDIDF structure and stabilized by reducing the order of the loop filter, is shown in Fig. 7.4. The excess hardware involves the implementation of the  $-\alpha_3$  coefficient, the two switches and their control circuits, the instability detector and the sign comparator.

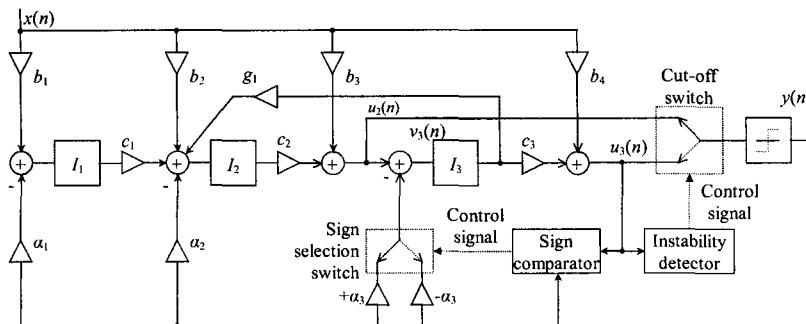


Fig. 7.4 Third-order  $\Delta\Sigma$  modulator stabilized by reducing the order of the loop filter.

For small input signals, the modulator operates in the stable region. Using the variable gain method, the stable region is defined as the range of input amplitudes for which the output of the third integrator is bounded by

$$|u_3(n)| \leq \frac{1}{K_{\min}} \quad (7.3)$$

In the stable region no stabilization actions are taken and the input to the quantizer is the value  $u_3(n)$  of the third integrator. As the input signal increases,  $u_3(n)$  can exceed the limit imposed by (7.3). This range of input amplitudes is defined as the unstable region of operation. For input amplitudes in this region the modulator becomes unstable. To prevent this, the signal  $u_2(n)$  from the second integrator is used as input to the quantizer, effectively cutting-off the third integrator from the loop. The result is a second-order  $\Delta\Sigma$  modulator, which is stable. The third integrator, although its output is not connected to the loop, still integrates its input. If the sign selection switch in Fig. 7.4 is selecting always the  $+a_3$  branch, the input  $v_3(n)$  to the third integrator is given by

$$v_3(n) = u_2(n) - a_3 \cdot y(n) \quad (7.4)$$

where  $y(n)$  is the  $\Delta\Sigma$  sequence, which is fed back. Thus  $u_3(n)$  changes during the cut-off time. The order of the modulator is restored only when the signal at the third integrator satisfies (7.3) once more. A switch selects the output of the third or the second integrator. This is the cut-off switch in Fig. 7.4. The cut-off switch is controlled by an instability detector that performs the comparison in (7.3).

In order to make the cut-off time as short as possible, if the excessive input signal is removed,  $v_3(n)$  should be such that it minimizes  $|u_3(n)|$ . To do so  $v_3(n)$  and  $u_3(n)$  should have different polarity for as many time samples as possible. Note that the  $a_3$  coefficient is positive, it weights the feedback signal and the result is subtracted from  $v_3(n)$ . Thus the feedback signal reduces the value stored in the  $I_3$  integrator, in an attempt to keep  $u_3(n)$  bounded. Should  $I_3$  be cut-off and  $u_2(n)$  and  $u_3(n)$  have different polarity, the feedback no longer reduces the value stored in  $I_3$ . In order to keep reducing the value stored in  $I_3$  and thus facilitate the fast recovery of the modulator from the overload state, it should be

$$v_3(n) = \begin{cases} u_2(n) - a_3 y(n) & \text{if } u_2(n)u_3(n) \geq 0 \\ u_2(n) + a_3 y(n) & \text{if } u_2(n)u_3(n) < 0 \end{cases} \quad (7.5)$$

The sign comparator and the sign selection switch in Fig. 7.4 implement Eq. (7.5) by controlling the sign of  $a_3$ . When (7.3) holds, i.e. the modulator operates normally, the sign comparator effectively compares the sign of the same quantity, namely  $u_3$ . Finding the two signs the same, it selects  $+a_3$ , i.e. the loop remains unchanged. Should  $I_3$  be cut-off, the signs of  $u_2(n)$  and  $u_3(n)$  are compared. If they are the same, then again  $+a_3$  is selected minimizing  $u_3(n)$ . If the signs of  $u_2(n)$  and  $u_3(n)$  are different, then  $-a_3 \cdot y(n)$  has the same sign as  $u_3(n)$ . So the sign selection switch selects the  $-a_3$  coefficient, and thus  $u_3(n)$  is still reduced. Thus the sign comparator and the sign selection switch in Fig. 7.4 are used to minimize the cut-off time of the modulator.

Comparing the stabilized  $\Delta\Sigma$  modulators of Fig. 7.3 and 7.4, the circuitry involved in the stabilization by the method of reducing the order of the loop filter is simpler than that of using local feedback loops. This is evident as stabilization circuitry in the proposed method is used only around the third integrator and not around all of them. Also, the instability detector is a single-threshold device, in contrast to the double-threshold overload detectors of the stabilization using local feedback loop. As for the sign comparator, it can be built using two single-threshold devices and a XOR gate. Stabilization by reducing the order of the loop filter also leads to better performance in most cases, as it will be shown in the next section.

#### 7.4 Comparison of the Stabilization Methods

Two  $\Delta\Sigma$  modulators are used for the comparison of the stabilization methods. Modulator A has NTF poles at  $-0.0367$  and  $0.8076 \pm j0.2733$  and modulator B at  $-0.0554$  and  $0.8070 \pm j0.3510$ . Modulator A has higher maximum stable input (and consequently smaller maximum SNR) than modulator B. As for the zeros, they vary according to the needs of the structure of every stabilization method: For the methods of resetting and

clipping the integrators, as both integrators and resonators can be used, the zeros are also the optimal at 0 and  $e^{\pm j0.038}$  (see chapter 4). For the method of reducing the order of the loop filter, as only integrators are used, the zeros have real parts equal to unity. This does not change them considerably from the optimum placement, thus the SNR gain of the optimum placement is retained. Finally, for the method of activating feedback loops around the integrators, the zeros have to be set to unity, and the resulting modulators are sub-optimal. For every method, many thresholds are used to obtain the optimum. The methods are applied as follows:

- a. For the method of resetting the integrators, all of them are reset upon the detection of a number of  $\Delta\Sigma$  samples of the same value. This number is the threshold  $T$  for the activation of the method.
- b. For the method of clipping the integrators, a modification from [6] is proposed to enhance performance: only the third integrator is clipped at some threshold level  $T$ .
- c. For the method of activating feedback loops around the integrators, activation of the lower-order loops at too low or too high level yielded poor results. Not using the lowest-order loop at all (to avoid the non-shaped noise leakage) resulted to unstable modulators. Thus the two lower-order integrators have their feedback loops activated at twice the theoretical threshold. The loop around the third integrator is activated using a variable threshold  $T$ .
- d. For the method of reducing the order of the loop filter, the method is activated as described in Sec. 7.3.4, only with variable threshold  $T$ .

For each method the SNR as a function of the input amplitude is plotted for the optimum threshold in Fig. 7.5.

The effect of finite opamp gain is to change the loop filter transfer function, hence also the SNR. But as long as the modulator remains in the stable input range, i.e. the stabilization method is not activated, this SNR change is minimal due to the robustness of the  $\Delta\Sigma$  modulator. Of course, the change of the loop filter alters the boundaries of the stable and unstable ranges in an unpredictable way. These changes are responsible for the SNR loss in all stabilization methods. In addition, as the method involving the local feedback loops involves digital correction

filters that, due to finite opamp gain, are no longer matched to the analogue part of the modulator, there is greater SNR degradation for low opamp gains. Opamp gains of 400 and the very low 100 are used to evaluate the robustness of the methods. The resulting SNR curves are depicted in Fig. 7.6.

From the results of the method of resetting the integrators it is evident that the performance in the stable input range is excellent, but in the unstable input range the gain is rather low. Also, the optimum number of consecutive +1 or -1 used to activate the method varies a lot

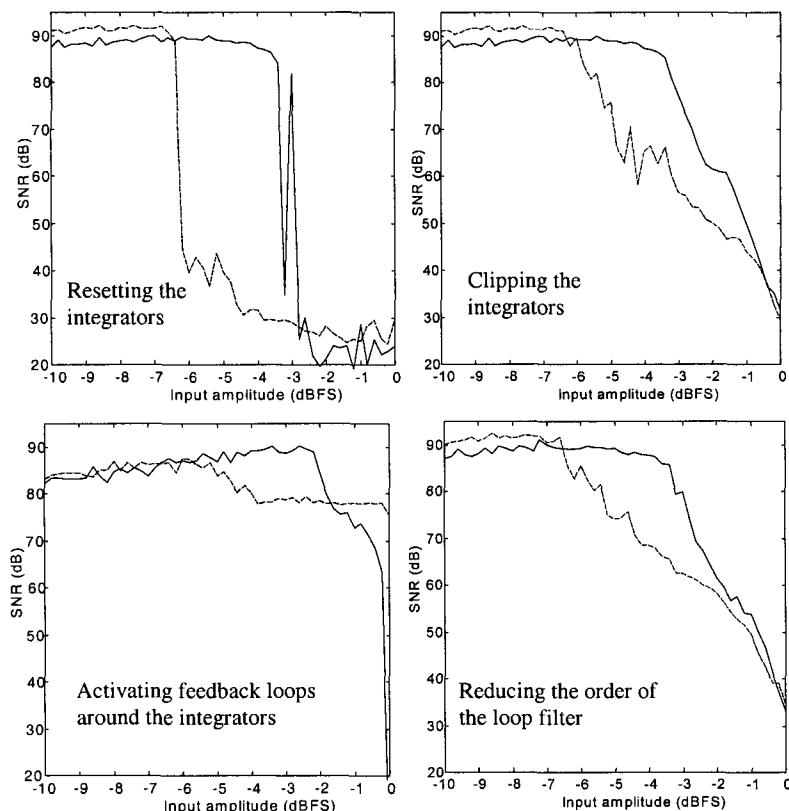


Fig. 7.5 SNR of the two stabilized third-order modulators as a function of the sinusoidal input amplitude. Modulator A is represented by solid lines, while B by the dashed.

as the NTF or the input signal change. On the other hand, modulators thus stabilized are very insensitive to finite opamp gain. Also, the recovery from overload is instant, as the integrators are reset. Finally, the hardware that facilitates resetting is simple, as the integrating capacitors just have to be discharged and the activation of the method is based on observation of the digital output.

The method of clipping the integrators yields far better unstable input range gain than the method of resetting the integrators. This gain can be

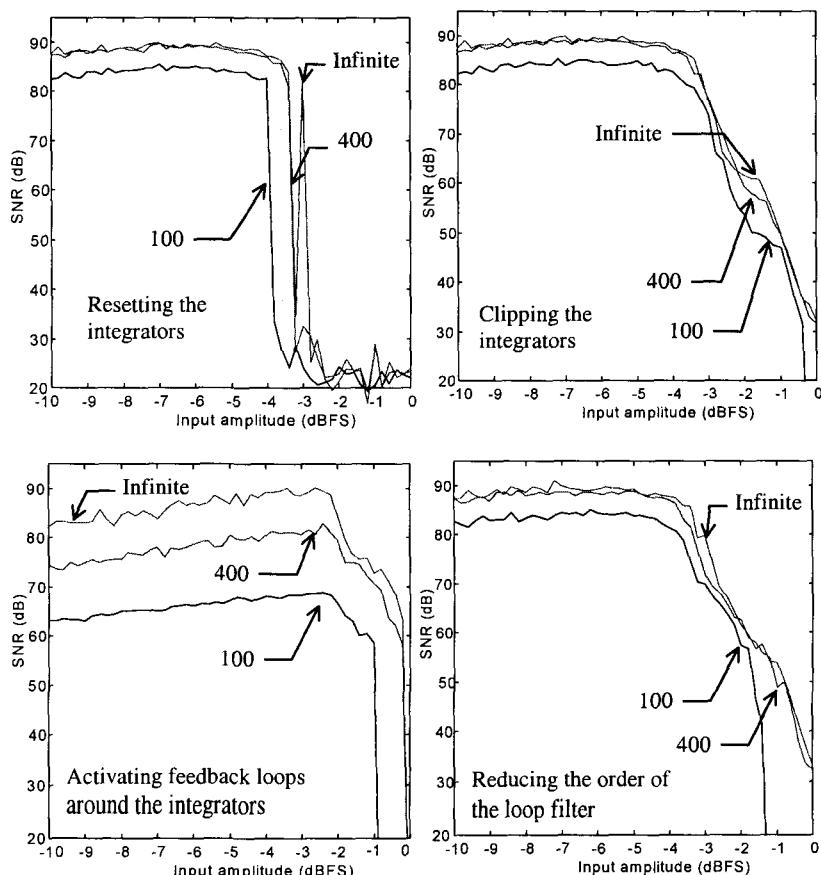


Fig. 7.6 SNR of the stabilized third-order modulator with finite opamp gain, as a function of the sinusoidal input amplitude. The opamp gain is shown on each curve.

further increased should some loss in the stable input range be tolerated. Thresholds around three times the full-scale yield the optimum gains, although even smaller thresholds yield comparable performance. This is also shown in Fig. 7.7, where the range of acceptable thresholds for the various methods and a third-order modulator with sinusoidal input are shown. These thresholds are close to the one predicted by the variable gain method. The activation of the method is rather insensitive to the type of input signal, which is advantageous if the modulator is to be used with real-world signals. Regarding the effects of finite opamp gain, the method proves rather insensitive, unless the opamp gain is too low. Then it exhibits the worst performance from all the methods. In terms of hardware, the only excess circuitry needed for the implementation of the method is that which regulates the supply voltage of the last integrator according to the threshold. On the other hand, opamp saturation reduces the speed of recovery from overload.

The performance of the third-order  $\Delta\Sigma$  modulators stabilized using local feedback loops around the integrators in the stable input range is very good as it exhibits gain. Also the gain in the unstable input range is large, in some cases covering the handicap due to the non-optimal placement of zeros due to the CIDF structure. Unfortunately, as the NTF is different, a direct comparison of these gains to those of other methods is not possible. The gains achieved have a strong dependency on the threshold, as shown in Fig. 7.7. The method is not immune to finite opamp gain, as then the gain in both the stable and the unstable input ranges is greatly reduced. Also, the method is quite complex to implement, as already discussed in the previous section. The optimal thresholds of activation vary between one and two times the full-scale, and are not always constant for different input signal types. Finally the recovery from overload is very fast, even for extreme levels of input.

When the method of reducing the order of the loop filter is used to stabilize the third-order  $\Delta\Sigma$  modulators, very good results are obtained for a wide range of thresholds larger than 1.5 times the full-scale. This is evident from the comparison in Fig. 7.7. Also the optimal thresholds are constant, regardless of the type of input signal. These thresholds are somewhat smaller than those predicted by the variable gain method. There is no loss of SNR in the stable input range and the gain in the

unstable input range is comparable to or even higher than that of clipping the integrators. Also the method is quite insensitive to finite opamp gain, as the gain in the unstable input range remains very high, even for very low opamp gains. The hardware needed for the method of reducing the order of the loop filter is already found less involved than the one of activating the local feedback loops. The recovery from overload is fast, comparable to that of the method of activating feedback loops around the integrators, as long as the input level is kept below full-scale. At full-scale the delay grows to some hundred of samples, which is of no worry as the signals are highly oversampled.

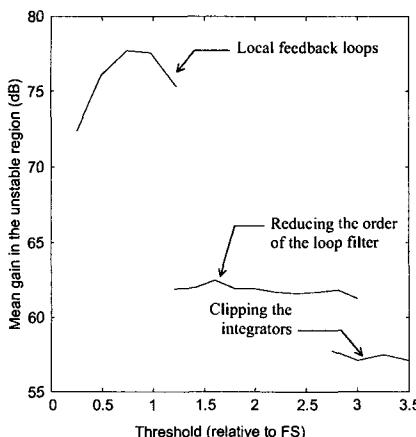


Fig. 7.7 Mean gain in the unstable range for thresholds yielding acceptable mean gain in the stable range. The performance of the method of activating local feedback loops around the integrators depends strongly on the threshold, and the method of reducing the order of the loop filter yields good results for a wide range of thresholds.

Concluding the comparison of the stabilization methods, results are summarized in Table 7.1. It is evident that resetting the integrators is the faster to recover from instability but the worse in terms of SNR gain in the unstable input range. Clipping the integrators has good SNR gain, but is slow in recovering. Activating local feedback loops around the integrators leads to excellent speed of recovery and SNR gain, which unfortunately cannot be directly compared to the rest, as the NTF zeros are forced to unity.

Table 7.1 Comparison of the four stabilization methods.

Method	Excess hardware	Optimum threshold	SNR gain	Speed of recovery	Notes
Resetting the integrators	Overload detector, circuitry for resetting the integrators	Varies for different input signals	Minimal	Instant	Used when speed is important
Clipping the integrators	Theoretical threshold as power supply of the last integrator	Close to theoretical, almost constant for different input signals	Very good	Slow due to saturation of the opamps	Used when gain is important, sensitive to low opamp gain
Activating local feedback loops	Quite complicated	Varies for different input signals	Excellent, but not comparable to the rest	Fast	Only CIDF structure, multi-bit output, very sensitive to low opamp gain
Reducing the order of the loop filter	Overload detector, switch to cut-off third integrator and sign comparator	Wide range below the theoretical, constant	Excellent for all input amplitudes below full-scale	Fast for all input amplitudes below full-scale	

Finally, the method of reducing the order of the loop filter yields better SNR gain than the method of clipping the integrators and is very fast in recovery from overload for all input amplitudes apart from the full-scale. It is also easier to implement and less sensitive to finite opamp gain than the method of activating the local feedback loops.

## 7.5 Tones in $\Delta\Sigma$ Modulators

Tones are reported in the output spectra of  $\Delta\Sigma$  modulators, especially when the input is rational DC or a sinusoid of rational frequency [5,10-13]. It is a common belief that only lower-order  $\Delta\Sigma$  modulators suffer from tones, but this is not true [5] as it is demonstrated in Fig. 7.8, for a third-order  $\Delta\Sigma$  modulator having NTF zeros at 0 and  $e^{\pm j0.038}$  and poles at -0.0367 and  $0.8076 \pm j0.2733$ . Note that this will be the modulator used in the analyses of the next sections, unless otherwise specified.

Tones are mostly evident when the input to the modulator is rational (in level or frequency). When the modulator is used as an ADC, then the probability of such signals is zero and so tones are not very important. But when the modulator is used as a DAC, the signals being digital are rational, and the tones become a serious unwanted artifact.

The existence of high frequency tones can be explained as follows: The only steady-state solution for zero input to the classic first-order modulator is the repeating pattern ..., 1, -1, 1, -1, ... i.e. a limit cycle causing a tone at  $f_s/2$ . If a small DC input  $m_x$  is added, then from time to time two identical codes are generated, as the error between the modulator input and output accumulates in the integrator. The frequency of the generation of the identical codes is  $m_x f_s$ , which explains the dominant out-of-band tone at

$$f_h = (1 - m_x) \frac{f_s}{2} \quad (7.6)$$

This out-of-band tone is very strong, comparable to the full-scale. The in-band tones at multiples of the first in-band harmonic at

$$f_l = m_x f_s \quad (7.7)$$

are due to the fact that the constant output power has to be shared between the strong out-of-band tone and the random components. As a

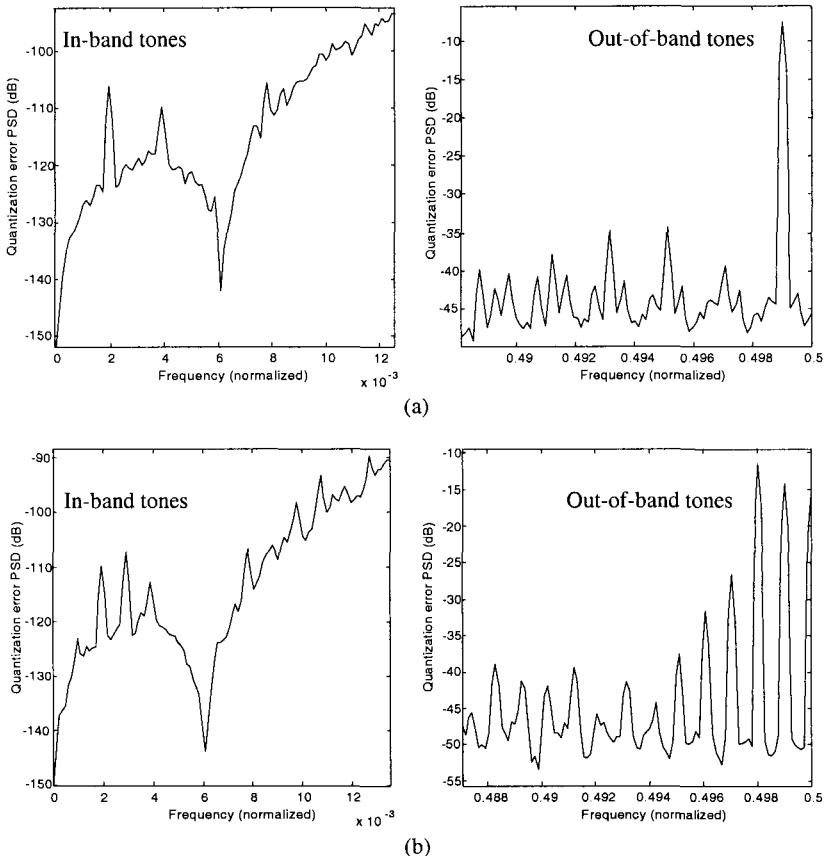


Fig. 7.8 Tones in the quantization error spectra of a third-order  $\Delta\Sigma$  modulator for a. rational DC input level 1/512 of the full-scale, or b. sinusoidal input of rational frequency 1/1024 of the sampling frequency and rational amplitude 1/256.

result the base-band noise power is modulated by  $f_l$  [5]. The in-band tones are of considerably less power than their out-of-band counterparts.

Although the in-band noise power of the modulators is low, the very existence of in-band tones is troublesome, as they can be perceived by the human ear [5]. Out-of-band tones are also troublesome, as circuit non-idealities can cause them to demodulate down to the base-band [5]. As these tones are usually of high power, even a small degree of

demodulation can destroy the SNR. Thus both in- and out-of-band tones should be suppressed.

Methods to suppress the tones are reported in the literature. Traditionally tones in  $\Delta\Sigma$  modulators are suppressed using dithering [5,11-14], or by using chaotic  $\Delta\Sigma$  modulators [5,13]. More recently bit-flipping has been introduced [11,12]. These are the fixed techniques. Also, adaptive techniques like adaptive dithering [5] or adaptive bit-flipping [12] have been proposed.

In the following sections the effectiveness of the above mentioned tone suppression methods is examined, both for the in-band and the out-of-band tones. The input signal is a rational DC level 1/512 of the full-scale. Also, as all these methods degrade the stability of the modulators, this adverse effect is examined by finding the Maximum Stable Amplitude (MSA) using a very slowly increasing ramp input signal. Finally, another adverse effect is the reduction of the SNR of the modulators. This is examined by determining the SNR for a sinusoidal input of amplitude 0.117 relative to the full-scale.

## 7.6 Fixed Techniques

The fixed techniques for suppressing the tones are presented in the following subsections. They are fixed in the sense that the action against the tones is taken, independent of whether the modulator is in a state that generates tones or not.

### 7.6.1 Dithering

The most widely used technique to suppress the tones in any data converter is dithering. Dithering is the addition to the input of the data converter of an uncorrelated signal. As a result the input and the quantization error become uncorrelated. Dithering is investigated theoretically in [14]. It is applied to data converters in [12] and specifically to  $\Delta\Sigma$  modulators in [5,12,13,15].

With the addition of the dither signal the noise in the modulator increases. Thus, care must be taken for the higher-order modulators not to become unstable, i.e. the dither signal should be small. Also, the in-

band noise power increases alarmingly, unless the dither signal is high-pass filtered (or band-reject filtered for band-pass  $\Delta\Sigma$  modulators). This filtering operation can be achieved by simply adding the dither signal directly to the input of the quantizer, thus shaping it by the NTF.

The dither signal is usually a pseudo-random noise. Rectangular and triangular PDF can be used. If the modulator is used as an ADC, then the dither signal must be analog. In order to avoid the use of a DAC, or the implementation of an analog noise source of a given PDF, single-bit dither can be used. This is obtained by single-bit quantizing the output of a digital pseudo-random noise source. The generation of such a signal is done using digital hardware. The single-bit DAC needed is trivial.

The effect of three types of dither signals, namely the single-bit, the rectangular PDF and the triangular PDF on  $\Delta\Sigma$  modulators is investigated, having the power of the dither signal as a parameter. Dithering is certainly successful in the suppression of tones, as is shown in Fig. 7.9. In the same figure, the effect of the three types of dithering on MSA and SNR is demonstrated. The first in-band tone is suppressed using very small dither signals in all three cases. The suppression of the first out-of-band tone is more difficult. Triangular PDF dither is the most efficient. Rectangular PDF and single-bit dither signals should have more than double the power to achieve the same results. The introduction of the dither noise signal makes the modulators less stable. The MSA resulting from the triangular PDF dither signal is less than that from the rectangular PDF and single-bit dither signals of the same power. Finally, although the dithering signal is high-pass shaped, it increases the baseband noise level, hence its introduction reduces the SNR.

### 7.6.2 *Chaotic $\Delta\Sigma$ modulators*

When the NTF of a  $\Delta\Sigma$  modulator is not minimum phase, then the loop filter has at least one pole outside the unit circle [5,13]. The whole feedback system can still be stable, but the dynamics of the loop filter are destabilized, and thus periodic sequences are broken. So chaotic  $\Delta\Sigma$  modulators are not tonal.

Chaotic modulators are obtained from non-chaotic ones by reflecting at least one NTF zero outside of the unit circle. From the linear model point of view, the performance of the modulator does not change if the

reciprocal of some zero is used. But  $\Delta\Sigma$  modulators should have NTF zeros on the unit circle (see Chapter 4), thus this reflection is not possible. Moving the zero away from the unit circle degrades the SNR and thus is not desirable. What is usually done to obtain a chaotic  $\Delta\Sigma$  modulator is the extension of the NTF of a non-chaotic modulator by an all-pass term

$$H_{\text{all-pass}}(z) = \frac{1 - a \cdot z^{-1}}{1 - \frac{z^{-1}}{a}} \quad (7.8)$$

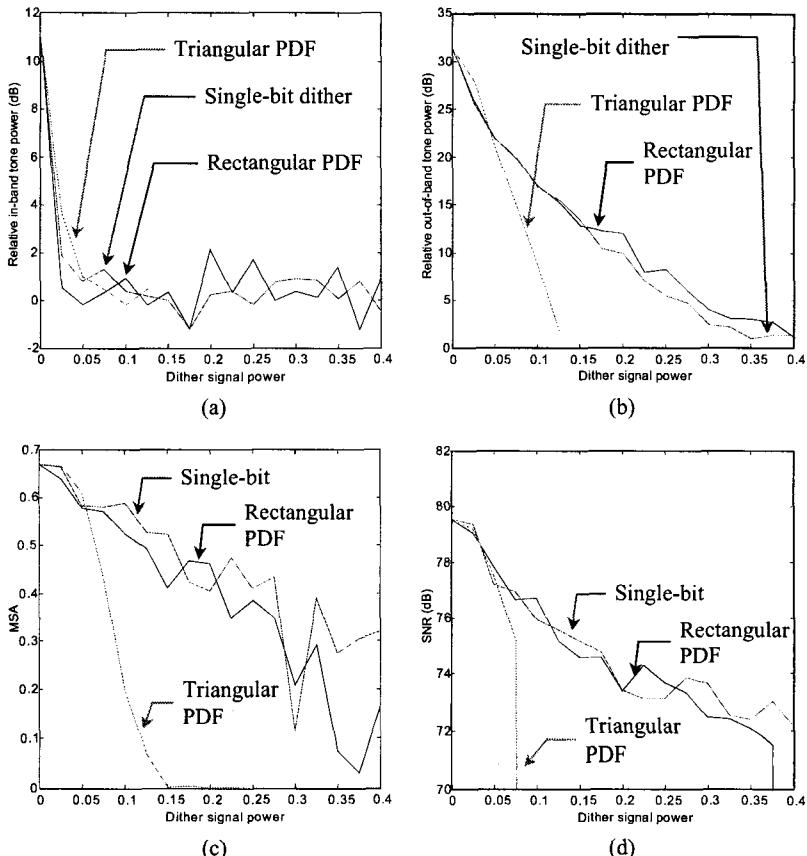


Fig. 7.9 Tone suppression using dither. a. In-band suppression and b. out-of-band suppression. c. Degradation of the MSA and d. Degradation of the SNR.

If  $|a| > 1$  then the resulting modulator is chaotic, and more so as the magnitude of  $\alpha$  increases. The resulting increase of the order of the modulator is not very important from a hardware point of view, as the involved circuitry need not be very expensive. It is reported in [13] that the extension with a negative zero  $a$  is more effective for the out-of-band tone suppression than the use of positive  $a$ .

The disadvantage of the chaotic modulators is that they are more unstable than their non-chaotic counterparts. This can be easily seen as the one-norm of the NTF increases, and does so excessively for negative  $\alpha$ . In some cases it is impossible to obtain stable chaotic modulators from the extension of non-chaotic ones with negative zero.

In the following subsections the effectiveness and penalties of the use of chaotic  $\Delta\Sigma$  modulators to suppress the tones is investigated, both for positive and negative zero extension.

The use of a positive zero in the all-pass term of Eq. (7.8) leads to chaotic modulators with adequate suppression of tones in the base-band. Unfortunately this is not true for the tones lying near half the sampling frequency. The effectiveness of the suppression and its effect on the MSA and the SNR as a function of the positive zero used for the extension is shown in Fig. 7.10. The more chaos is used, the less stable the modulator becomes. Finally, the SNR does not vary considerably until the onset of instability for large  $a$ , as the base-band noise level is fairly constant.

Although the SNR and stability performance of the extended with positive zero modulator is considerable and the base-band tones are adequately suppressed, positive zero extension is inadequate for the out-of-band tones. Thus the performance of the extension with negative zero has to be investigated, even though this can be applied only to sub-optimal modulators. The fifth-order modulator obtained using Chebyshev NTF prototypes is used for this analysis.

The effectiveness in the suppression of tones as a function of the negative zero used for the extension and the degradation of the MSA and the SNR are shown in Fig. 7.11. The suppression of tones in the base-band needs somewhat larger values of  $|a|$  than in the case of positive  $\alpha$ .

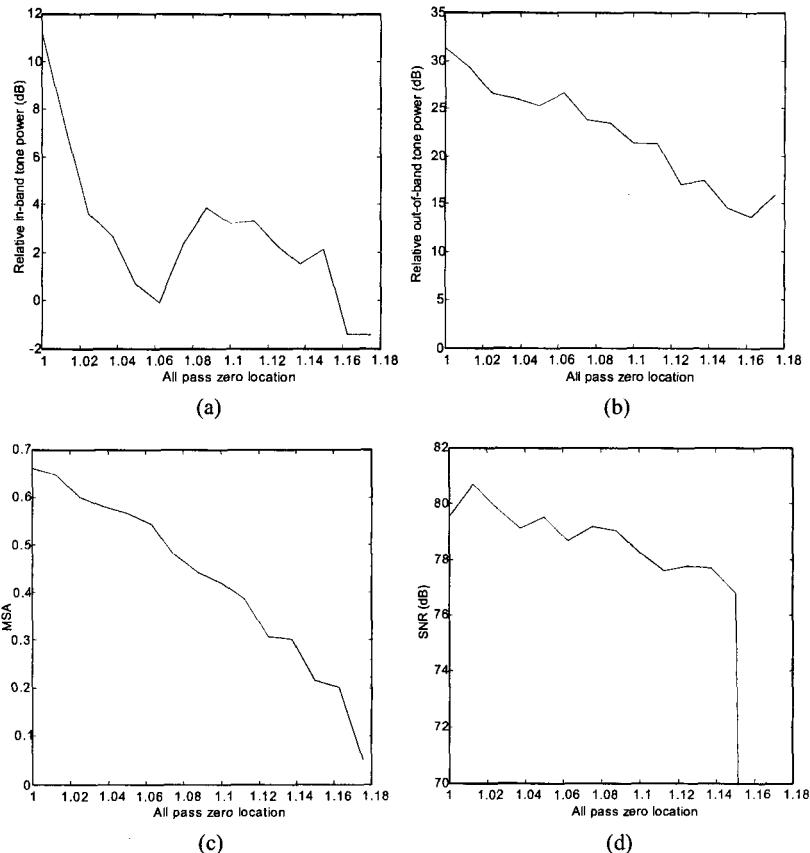


Fig. 7.10 Tone suppression using chaos with positive zero extension. In-band (a) and out-of-band (b) suppression. Also degradation of the MSA (c) and the SNR (d).

Now, also the tones lying near half the sampling frequency are adequately suppressed. Again the more chaos is used, the less stable the modulator becomes. Finally, the SNR does not vary considerably. Even very large values of  $|a|$  still lead to stable modulators.

### 7.6.3 Bit-flipping

The effect of dithering (in single-bit quantizer modulators) is to flip the output bit with probability  $P_f$  given by

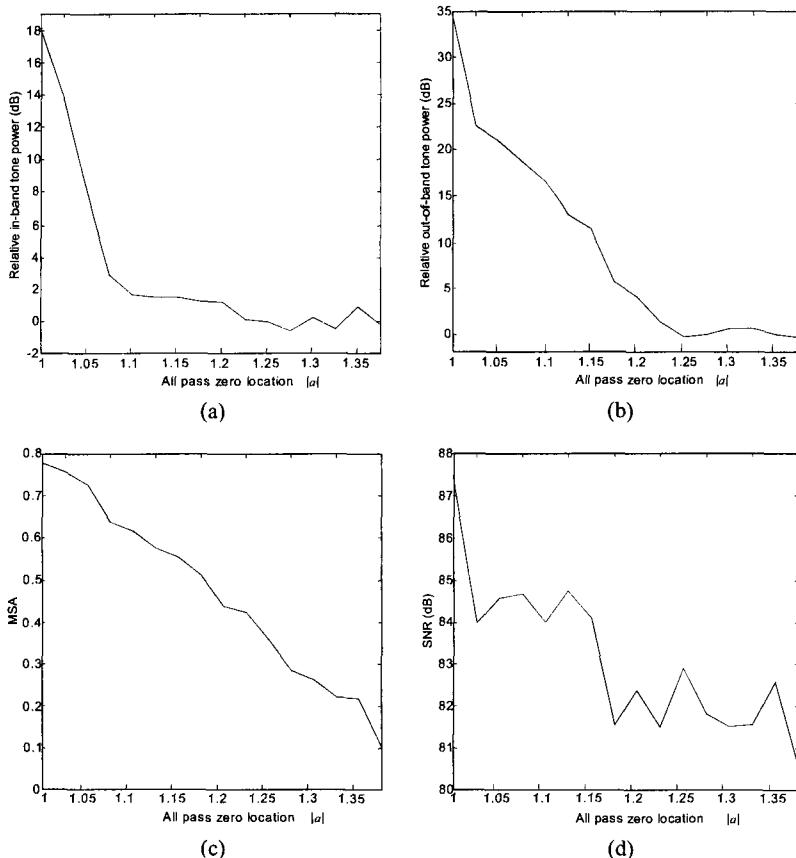


Fig. 7.11 Tone suppression using chaos with negative zero extension. a. In-band and b. out-of-band suppression. c. Degradation of the MSA and d. Degradation of the SNR.

$$P_f = \Pr[d(n) < -u(n)] = \int_{-\infty}^{-u(n)} f_d(x) dx \quad (7.9)$$

where  $d(n)$  is the current sample of the dither signal,  $u(n)$  is the current sample of the output of the loop filter and  $f_d$  is the PDF of the dither signal. So in effect, as  $u(n)$  becomes smaller, the probability of the output of the dithered modulator to be flipped increases. It is this bit-

flipping mechanism that actually breaks up the periodicity of the output bit-stream, effectively suppressing the tones.

Magrath in [11,12] proposes the following scheme. Let the output of the quantizer be flipped with probability 1, if the input to the quantizer is smaller than a threshold value  $D$ . This is equivalent to single-bit dithering with a signal  $\pm D$ , which always has the inverse polarity with respect to that of the input to the quantizer. This scheme is termed bit-flipping.

The effectiveness of this scheme in the suppression of tones and its effect on the MSA and SNR is depicted in Fig. 7.12.

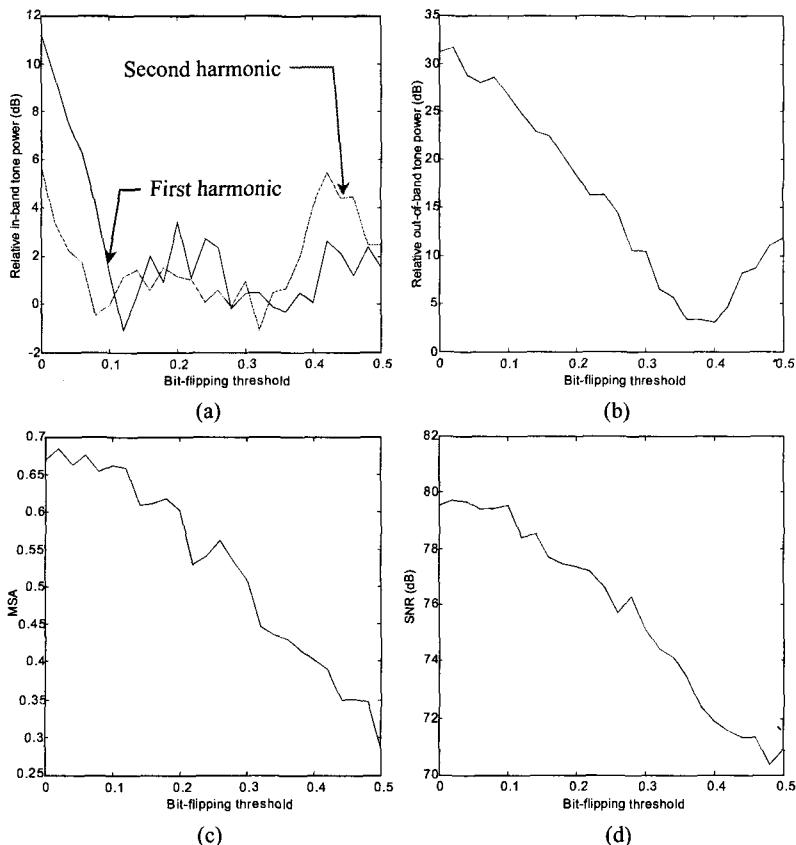


Fig. 7.12 Tone suppression using bit-flipping. a. In-band and b. out-of-band suppression. c. Degradation of the MSA and d. degradation of the SNR.

It is evident that the first in-band harmonic and the tone near half the sampling frequency can be suppressed, but for different threshold values. At the value of  $D$  for which both tones are suppressed adequately, the second in-band harmonic has increased power. Thus simultaneous suppression of tones is very difficult using bit-flipping. The larger the threshold values are used, the less stable the modulator becomes. Finally, the SNR decreases with increased threshold  $D$ .

#### ***7.6.4 Comparison of fixed techniques***

In the preceding sections three fixed techniques for the suppression of tones were presented. Dithering and bit-flipping are closely related and can be applied to any  $\Delta\Sigma$  modulator. Chaos can universally be used only if NTF extension with a positive zero is performed. What is wanted when these techniques are applied is to attenuate all tones at the minimum SNR and MSA costs. The techniques are compared in terms of costs when the dominant in-band tone is attenuated by at least 7dB and the dominant out-of-band tone is attenuated by at least 17dB. The results are summarized in Table 7.2.

Table 7.2 Comparison of the fixed techniques for tone suppression.

Technique	MSA	MSA loss (%)	SNR (dB)	SNR loss (dB)
Uniform dither	0.51	24	75.9	3.44
Triangular dither	0.44	34	75.3	4.04
Single-bit dither	0.53	21	75.8	3.54
Chaos, positive zero	0.26	51	77.2	2.14
Bit-flipping	0.55	18	76.1	3.24

It is evident from Table 7.2 that overall bit-flipping performs best, although care has to be taken about the second in-band harmonic. Chaos preserves the SNR nicely, but it is very inadequate at the out-of-band tone suppression. The resulting MSA is poor. From the fixed dither techniques, the best is single-bit dithering, both in terms of performance and in ease of implementation.

## 7.7 Adaptive Techniques

It is not good practice to take action against the tones, if the modulator is in a state that is likely to generate them, as these tone suppression techniques result in a degradation of the modulator performance. Adaptive techniques on the other hand do monitor the modulator and are equipped with a decision-making mechanism: is the modulator generating tones or not? These techniques are the adaptive dithering [5] and the adaptive bit-flipping [12], and are presented in the next subsections.

### 7.7.1 Adaptive dithering

The generation of tones depends on the input signal. Quiet input signals generate tones, whereas busy input signals do not. Thus it is advantageous to monitor the input signal, and arrange the power of the dither signal injected into the system. If the input level is small, then large dither is needed, but overall stability is not jeopardized. On the other hand, if the input level is large, then no dither is needed. It is also advantageous to monitor the input level in the digital domain. This is readily done by checking the digital output of the decimator, which is a close approximation of the corresponding analog input signal. This technique is the adaptive bit flipping introduced by Norsworthy in [5].

Adaptive single-bit dithering is applied to the third-order  $\Delta\Sigma$  modulator used throughout this chapter. The dither level is adjusted using the level of the output of the first decimator stage as in Table 7.3. Using this scheme, the in-band tones for an input DC level of 1/512 of the full-scale are completely suppressed, while the out-of-band tone at  $f_h$  is suppressed by 25.3dB. This is achieved without any loss in the MSA and only 2.59dB loss in the SNR.

Table 7.3 Adjustment of single-bit dither amplitude using the amplitude of the output of the first stage decimator.

Decimated signal amplitude	Dither amplitude
[0, 0.0625)	0.5
[0.0625, 0.5)	0.1
[0.5, $\infty$ )	no dither

### 7.7.2 *Adaptive bit-flipping*

In adaptive dithering an estimation of the amplitude of the input signal is used to determine whether actions are needed to suppress the tones or not. Limit cycle detection at the output bit stream can also be used. Should limit cycles be detected, then the current output bit is flipped, effectively breaking up the limit cycle. This is adaptive bit-flipping, introduced by Magrath in [12].

The disadvantage of adaptive bit-flipping is that the limit cycle detector is not trivial. The number of possible limit cycles is large, thus a lot of digital logic is needed to detect them. In [12] it is proposed just to detect first-order limit cycles, i.e. ..., 1, -1, 1, -1, ... or second-order limit cycles, i.e. ..., 1, 1, -1, -1, 1, 1, ... longer than a given length. This is tried for a length of 8 bit for both limit cycles and the third-order  $\Delta\Sigma$  modulator used throughout this chapter. The result was the mediocre suppression of the out-of-band tone at  $f_h$  by only 16.7dB. This is just the performance obtained using chaos and positive zero extension. For in-band tones the situation is even worse; they are not suppressed at all. Again there is no loss in the MSA (characteristic of the adaptive techniques), but there is a considerable loss of 8.01 dB in the SNR.

### 7.7.3 *Comparison of the adaptive techniques*

Tone suppression should be carried out using adaptive techniques to minimize performance loss in terms of MSA and SNR. Unfortunately adaptive bit-flipping needs impractical limit cycle detectors. On the other hand adaptive single-bit dithering is very effective with minimal SNR penalty and no MSA penalty. The adaptation hardware is not involved and fully digital. Also no analog noise source is needed. Thus adaptive single-bit dithering is the best candidate for tone suppression.

## 7.8 Summary

Having designed single-stage higher-order  $\Delta\Sigma$  modulators in the previous chapters, here two inherent problems of these modulators, namely stability and tones, are considered, and means to overcome them are presented.

Although careful design can ensure stability up to any input power, noise can drive the modulators into instability. Hence some means of preserving the performance of the modulator and returning it to stable operation are needed. Clipping of the integrators at their supply voltage is inherent in any circuit realization, alleviating the problem, but other methods can be more effective. These methods are presented in the first half of this chapter, and are summarized in Sec. 7.4.

The tones of  $\Delta\Sigma$  modulators are troublesome only in DAC designs, as in ADCs the thermal noise serves as dithering signal, and rational input signals do not exist. In DACs, the tones can be suppressed. In-band tones are easily suppressed, but the out-of-band tones are more persistent. Following the comparison of the tone suppression techniques in Secs. 7.6.4 and 7.7.3, the most successful technique is adaptive single-bit dithering.

## References

- [1] R. Baird and T. Fiez, 'Stability Analysis of High-Order Delta-Sigma Modulation for ADC's', *IEEE Trans. Circuits and Systems*, Vol. 41, pp. 59-62, January 1994.
- [2] T. Okamoto, Y. Maruyama and A. Yukawa, 'A Stable High-Order Delta-Sigma Modulator With an FIR Spectrum Distributor', *IEEE J. Solid-State Circuits*, Vol. 28, No. 7, pp. 730-735, July 1993.
- [3] J. Candy and G. Temes, 'Oversampling Method for A/D and D/A Conversion', in J. Candy and G. Temes, editors, *Oversampling Delta-Sigma Data Converters, Theory, Design and Simulation*, IEEE Press, New York, pp. 1-25, 1991.
- [4] S. Moussavi and B. Leung, 'High-Order Single-Stage Single-Bit Oversampling A/D Converter Stabilized with Local Feedback Loops', *IEEE Trans. on Circuits and Systems-II: Analog and Digital Signal Proc.*, Vol. 41, pp. 19-25, Jan. 1994.
- [5] S. Norsworthy, R. Schreier and G. Temes, *Delta-Sigma Data Converters, Theory, Design and Simulation*, New York, IEEE Press, 1997.
- [6] C. Dunn and M. Sandler, 'Use of Clipping in Sigma-Delta Modulators', *IEE Colloquium on Oversampling Techniques and Sigma-Delta Modulation*, London, 1994.
- [7] S. Au and B. Leung, 'A 1.95-V, 0.34-mW, 12-b Sigma-Delta Modulator Stabilized by Local Feedback Loops', *IEEE J. Solid-State Circuits*, Vol. 32, pp. 321-328, March 1997.
- [8] A. Pneumatiakakis, A. Constantinides, G. Bourdopoulos and T. Deliyannis, 'Stabilization of Third-Order Single-Stage Sigma-Delta Modulators', *Circuits, Systems and Signal Processing*, Vol. 18, No. 2, pp.149-168, 1999.
- [9] A. Pneumatiakakis, *Modelling, Analysis and Design of Higher-Order Single-Stage Sigma-Delta Modulators*, Ph.D. thesis, Imperial College, University of London, May 1999.
- [10] R. Ledzius and J. Irwin, 'The Basis and Architecture for the Reduction of Tones in a Sigma-Delta DAC', *IEEE Trans. Circuits and Sys.-II: Analog and Digital Signal Proc.*, Vol. 40, No. 7, pp. 429-439, July 1993.
- [11] A. Magrath and M. Sandler, 'Digital-Domain Dithering of Sigma-Delta Modulators Using Bit Flipping', *J. Audio Eng. Soc.*, Vol. 45, No. 6, June 1997.
- [12] A. Magrath, *Algorithms and Architectures for High Resolution Sigma-Delta Converters*, Ph.D. thesis, King's College, University of London, October 1996.
- [13] L. Risbo,  *$\Sigma\Delta$  modulators: Stability Analysis and Optimization*, Ph.D. Thesis, Technical University of Denmark, 1994.
- [14] R. Gray, 'Quantization Noise Spectra', *IEEE Trans. on Information Theory*, Vol. 36, pp. 1220-1244, November 1990.
- [15] W. Chou and R. Gray, 'Dithering and its Effects on Sigma-Delta and Multistage Sigma-Delta Modulation', *IEEE Trans. Information Theory*, Vol. 37, No. 3, pp. 500-513, May 1991.

## Chapter 8

# Decimation, Interpolation and Converters

### 8.1 Introduction

In previous chapters we examined the various types of  $\Delta\Sigma$  modulators, with emphasis on those employing single-bit quantizers. In this chapter their particular use in building ADCs and DACs is explained. Because of the use of oversampling in  $\Delta\Sigma$  modulators, the need arises for changing the sampling rate of the signal, decreasing it to the Nyquist rate in the case of an ADC or increasing it in the case of a DAC [1-3]. The former process is called decimation, whereas the latter interpolation. Such rate conversions can be achieved employing high precision FIR filters, usually in multiple stages. The ratio of the initial over the final sampling rate may be fractional as in case of transferring data from a music CD (44.1 kHz) to a Digital Audio Tape (48 kHz).

The basic concepts on rate conversion, namely decimation and interpolation are analyzed in section 8.2. Next, in section 8.3 are given certain filter structures and filtering techniques suitable for rate conversion. In the next section decimation is applied to the output  $\Delta\Sigma$  sequence to facilitate A/D conversion and give the final multibit signal in the Nyquist rate. In the last section interpolation is applied to a Nyquist-rate multi-bit digital signal to increase the sampling rate and facilitate the use of a  $\Delta\Sigma$  Modulator in the D/A conversion process.

## 8.2 Rate Conversion

There are many cases where the sampling rate at which digital signals are processed needs to be changed. Such needs arise when interfacing between two systems with different sampling rates. The process at which the sampling rate is decreased by  $r$  is called decimation, as one every  $r$  samples is kept. The inverse process of increasing the sampling rate by  $r$  is called interpolation, as  $r-1$  appropriate samples are inserted between two original samples. We examine both decimation and interpolation separately in the following subsections.

### 8.2.1 Decimation

The operation of a decimator by  $r$  is straightforward: one every  $r$  samples is kept, the rest are discarded. This way the original sampling frequency  $f_s$  is decreased to  $f_s/r$ . To avoid aliasing of the spectral components above half the final sampling rate, i.e. the spectral components in  $[f_s/2r, f_s/2]$  into the band of the decimated signal, a low-pass decimation filter precedes the down-sampling process (Fig. 8.1).

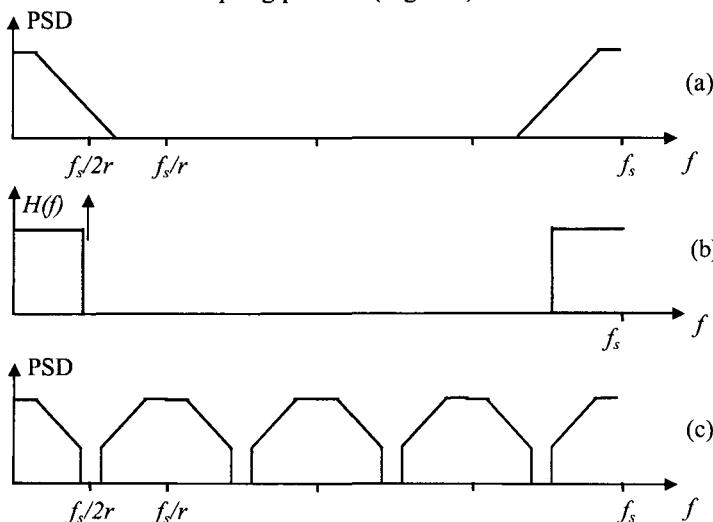


Fig. 8.1 Decimation by  $r=4$ . a. Spectrum of the oversampled signal. b. Low-pass filtering of signal components higher than  $f_s/2r$ . c. Spectrum of the down-sampled signal.

### 8.2.2 Interpolation

The inverse operation of the interpolator by  $r$  involves the insertion of  $r-1$  appropriate samples between two original samples. Three approaches to achieve this are as follows:

A straightforward approach is to insert  $r-1$  zeros. The increase of the sampling rate from  $f_s$  to  $rf_s$  brings the frequency content of the  $r-1$  Nyquist zones  $(m f_s/2, (m+1) f_s/2)$ ,  $m=1, \dots, r-1$ , together with the first one ( $m=0$ ) in the new first Nyquist zone  $(0, rf_s/2)$ , creating  $r-1$  images of the original spectrum. The spectra of the  $r-1$  images equal the wanted spectrum at  $(0, f_s/2)$ , and need to be attenuated below a specified level to avoid excessive signal distortion (Fig.8.2).

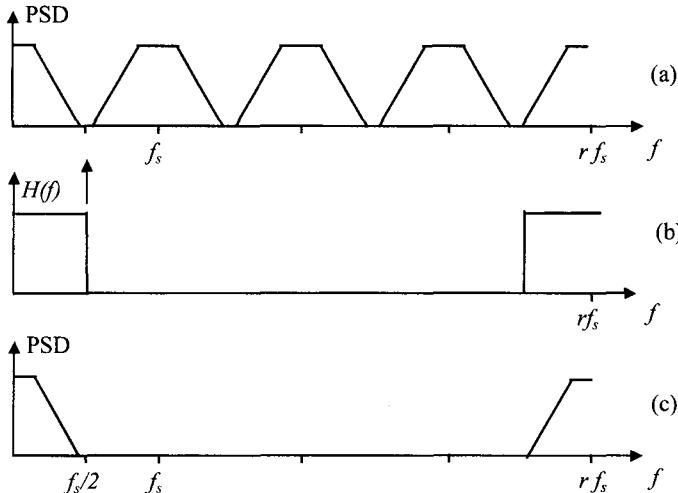


Fig. 8.2 Interpolation by  $r=4$ . a. Spectrum of the original signal. b. Low-pass filtering for canceling  $r-1=3$  images of the original signal band. c. Spectrum of the interpolated signal.

A second approach involves a sample and hold operation, where the sampling performs resampling at the higher rate and the zero-order hold keeps the original sample value for  $k$  samples. The rest  $r-k-1$  samples are zero. In this case the  $r-1$  images also exist, but the signal suffers from  $\sin x/x$  distortion, given by

$$\left| H\left(e^{j2\pi f / f_s}\right) \right| = \frac{\sin\left(\pi f \frac{k+1}{rf_s}\right)}{(k+1)\sin\left(\frac{\pi f}{rf_s}\right)}, \quad k = 0, \dots, r-1 \quad (8.1)$$

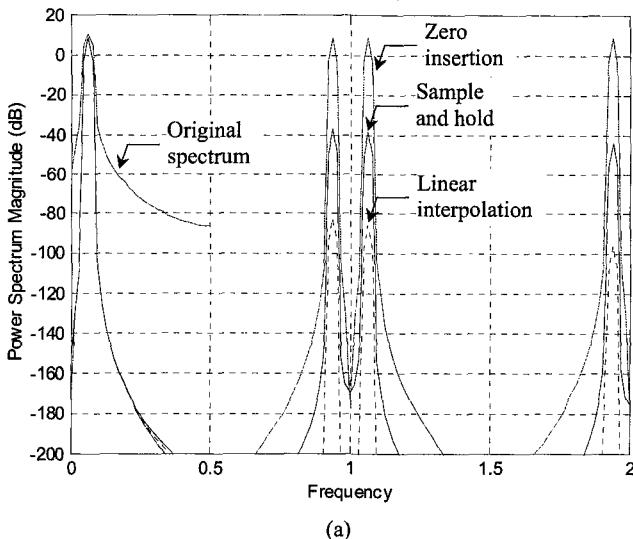
The magnitude of (8.1) is plotted in Fig. 8.3a for  $(r,k)=(16,15)$ , i.e. when all the  $r-1$  inserted samples have the value of the first original sample. In this case the  $r-1$  images are attenuated, but the wanted part of the spectrum in  $(0, f_s/2)$  is degraded.

A third approach is to perform linear interpolation between the two original values. In this case the  $r-1$  images are even more attenuated, but the wanted part of the spectrum in  $(0, f_s/2)$  is also more degraded.

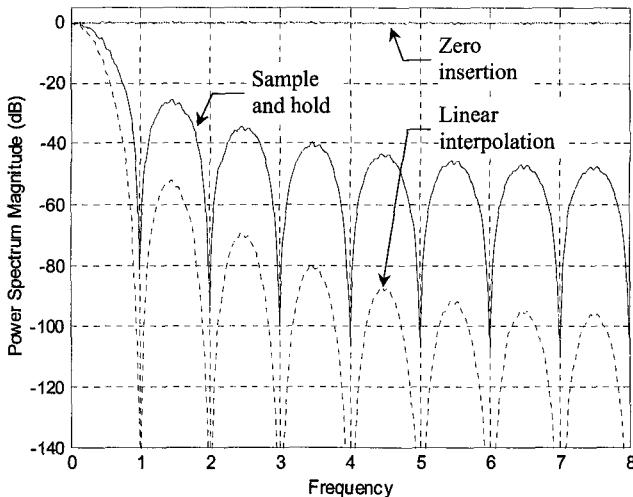
The effect of the three approaches of interpolation on a sinewave and on white noise is depicted in Fig. 8.3 All the interpolation approaches presented result to images of the original spectrum. To avoid distortion they have to be attenuated. The last two approaches are equivalent to some sort of filtering that, to some extent, achieves this suppression. Their drawback is that they also introduce some distortion of the original spectrum. Hence the best approach is to insert  $r-1$  zeros. The attenuation of the images will be performed afterwards, using a filter with smoother pass-band than that of Fig. 8.3b.

**Example 8.1** A HIPERLAN/2 wireless LAN [4] system has two base-band quadrature (I/Q) digital signals sampled at  $f_s = 20$  MHz. Their bandwidth  $B$  is 16.25 MHz. Assume that the low intermediate frequency  $IF$  is also 20 MHz. Describe a system that will interface the base-band transmitter to this  $IF$  using digital modulation.

**Solution** To digitally modulate two quadrature base-band signals to  $IF$  [5], one multiplies the in-phase (I) signal with the cosine of the  $IF$  and the quadrature (Q) signal with the sine of the  $IF$ . Then the two results are added. The resulting digitally modulated signal is centered at  $IF$ . Hence the highest frequency of the modulated signal is  $IF + B/2 = 28.125$  MHz. Evidently the original sampling rate cannot accommodate this frequency. Interpolation of the I/Q signals is needed. The resulting interface to the  $IF$  is shown in Fig. 8.4



(a)



(b)

Fig. 8.3 Effect of the three interpolation approaches on the spectrum of a. a sine wave by 4 and b. white noise by 16. The latter is an indication of the magnitude response of the equivalent filtering process. Inserting zeros results in unfiltered images and no in-band distortion, while sample and hold or interpolation result in attenuated images and in-band distortion. The frequency axes are normalized to the original sampling rate.

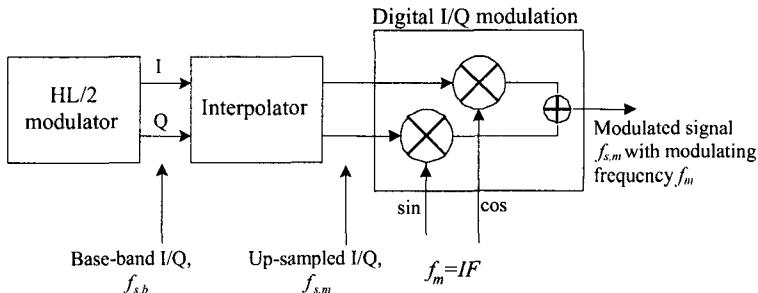


Fig. 8.4 Base-band to IF interface for a HIPERLAN/2 digital modulation system. The base-band I/Q signals are centered at DC and are sampled by  $f_{s,b}=20$  MHz. The interpolated I/Q signals are still centered at DC but are sampled by  $f_{s,m}=80$  MHz. Finally the modulating frequency is  $f_m=IF=20$  MHz and the modulated signal is centered at IF and is still sampled by  $f_{s,m}=80$  MHz.

Selecting to interpolate the I/Q signals by  $r=4$  yields an increased sampling rate of  $rf_s=80$  MHz. Then the IF is one-fourth of the final sampling rate; hence the generation of the sine and cosine of IF and the multiplication with the interpolated I/Q signals are trivial: The cosine samples are 1, 0, -1, 0, ... and the sine samples are 0, 1, 0, -1, ..., so the two multiplications are performed simply by passing through the I/Q samples, reversing their sign or zeroing them.

### 8.3 Decimation and Interpolation Filters

As discussed in the previous section, an interpolating filter follows the increase of the sampling rate from  $f_s$  to  $rf_s$  in an interpolator. Its goal is to attenuate the  $r-1$  images, while keeping the distortion up to  $f_s/2$  below a desired level. If the bandwidth of the signal is  $B$ , then the  $r-1$  images to be suppressed lie in the frequency zones given by

$$\left(2m \cdot \frac{f_s}{2} - \frac{B}{2}, 2m \cdot \frac{f_s}{2} + \frac{B}{2}\right), \quad m = 1, \dots, \begin{cases} r/2 & r \text{ even} \\ (r-1)/2 & r \text{ odd} \end{cases} \quad (8.2)$$

Evidently, the low-pass interpolation filter should have a cut-off frequency of  $B/2$  and should achieve the maximum attenuation at  $f_s - B/2$ . The image suppression filtering is depicted in Fig. 8.5.a.

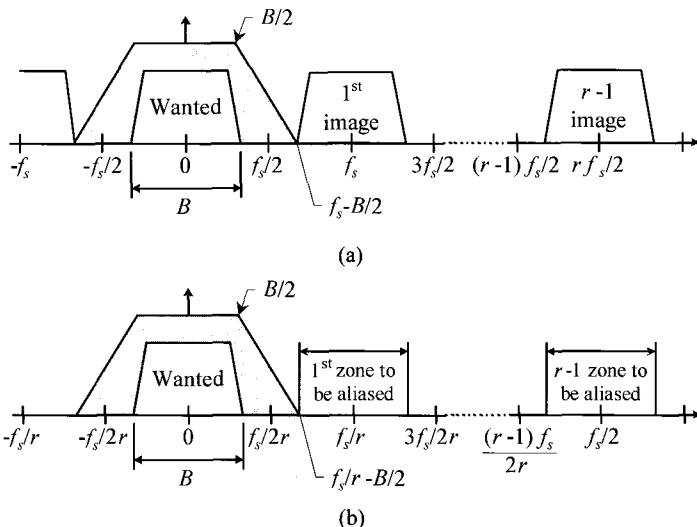


Fig. 8.5 a. Interpolating and b. decimating low-pass filters. The rate conversion is  $r$ , which in the figure is assumed even.

A low pass filter is also used in the decimation process. The rate reduction by  $r$  breaks up the original Nyquist zone ( $0, f_s/2$ ) into  $r$  zones,  $(mf_s/2r, (m+1)f_s/2r)$ ,  $m=0, \dots, r-1$ . The  $r-1$  zones ( $m=0, \dots, r-1$ ) alias into the first one, hence the frequency content in these zones has to be suppressed below a desired level to keep the wanted signal noise-free. As the bandwidth of the signal usually does not span the entire Nyquist zone, the frequency zones that have to be suppressed are

$$\left( m \frac{f_s}{r} - \frac{B}{2}, m \frac{f_s}{r} + \frac{B}{2} \right), \quad m = 1, \dots, \begin{cases} r/2 & r \text{ even} \\ (r-1)/2 & r \text{ odd} \end{cases} \quad (8.3)$$

Evidently, the low-pass decimating filter should have a cut-off frequency of  $B/2$  and should achieve the maximum attenuation at  $f_s/r - B/2$ . The anti-aliasing filtering is depicted in Fig. 8.5.b.

Both the decimating and interpolating filters operate in the higher sampling rate; the decimating filter operates prior to the rate reduction, while the interpolating filter operates after the rate increase. They are

strict filters, as the width of their transition band is a small fraction of the sampling rate on which the filters operate ( $f_{s,high}$ ). This width is given by  $f_{s,high}/r - B$  and decreases as the resampling ratio  $r$  increases.

**Example 8.2** Returning to the HIPERLAN/2 digital I/Q modulator of the example 8.1, design the required interpolation filter. 40 dB of image suppression is required.

**Solution** The sampling rate of the filter is  $4f_s = 80$  MHz. Its cut-off frequency is  $B/2 = 8.125$  MHz. The stop-band, where 40 dB attenuation is required, begins at  $f_s - B/2 = 11.875$  MHz.

Using MATLAB, an equiripple FIR filter fulfilling the above specifications is designed. Its order is 43 (44 symmetrical coefficients). The requirements are fulfilled, as the minimum stop-band attenuation is 41 dB, while the in-band distortion it introduces is  $\pm 0.1$  dB. The magnitude response of the filter is shown in Fig. 8.6.

This number of coefficients is large for the high sampling frequency of  $4f_s = 80$  MHz, as the required rate of multiplications to implement the filter is the non-trivial

$$22 \frac{\text{multiplications}}{\text{sample}} \cdot 80 \frac{\text{samples}}{\mu\text{s}} = 1.76 \frac{\text{multiplications}}{\text{ns}}$$

### 8.3.1 Multi-stage rate conversion

As shown in Example 8.2, it is advantageous to reduce the complexity of the resampling filters. In the last section, the width of the transition band of these filters is increased as the resampling ratio  $r$  is reduced. This gives motivation to perform the resampling in multiple stages, say  $n$  [2-3], where the individual resampling ratios  $r_i$  fulfill

$$r = \prod_{i=1}^n r_i \quad (8.4)$$

In this case each filter is more relaxed and operates at a different sampling rate.

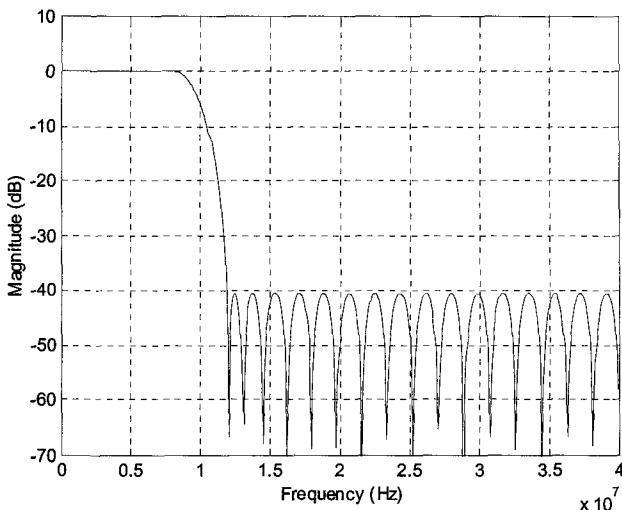


Fig. 8.6 Magnitude response of the interpolation filter of Example 8.2.

**Example 8.3** Perform the interpolation of Example 8.2 in two stages and compare the number of multiplications per ns needed in each case.

**Solution** The interpolation by  $r=4$  can be performed in two steps by  $r_i=2$ ,  $i=1,2$ .

The first interpolating filter operates at a sampling rate of  $r_1 f_s = 40$  MHz. Its cut-off frequency is again  $B/2 = 8.125$  MHz and the stop-band, where 40 dB attenuation is required, begins at  $f_s - B/2 = 11.875$  MHz. Using MATLAB, an equiripple FIR filter of order 25 (26 symmetrical coefficients) is designed, yielding minimum stop-band attenuation of 45.1 dB and  $\pm 0.048$  dB of in-band distortion.

The second interpolating filter operates at a sampling rate of  $r_1 r_2 f_s = r f_s = 80$  MHz. Its cut-off frequency is  $B/2 = 8.125$  MHz and the stop-band, where 40 dB attenuation is required, begins at  $r_1 f_s - B/2 = 31.875$  MHz. Using MATLAB, an equiripple FIR filter of order 8 (8 symmetrical coefficients and a central one) is designed, yielding minimum stop-band attenuation of 47 dB and  $\pm 0.049$  dB of in-

band distortion.

The rate of multiplications now is

$$13r_1f_s + 5r_1r_2f_s = 0.92 \text{ multiplications/ns}$$

i.e. almost half of those needed in the single-stage interpolation of Example 8.2. The overall in-band distortion is  $\pm 0.097$  dB, i.e. approximately the same as the distortion introduced by the single-stage interpolation of Example 8.2. Finally, the overall stop-band attenuation is at least 44.8 dB, almost 4 dB more than that of the single-stage interpolator of Example 8.2. The overall response of the two-stage interpolator is shown in Fig. 8.7.

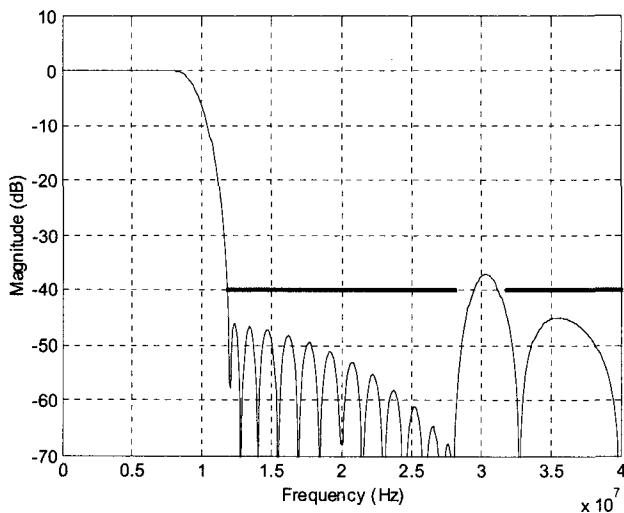


Fig. 8.7 Two-stage FIR interpolation filter for Example 8.3. Note that the response peak above the wanted attenuation of 40 dB does not coincide with any aliased band of the signal.

### 8.3.2 Comb filters

In Examples 8.2 and 8.3, the rate conversion ratio is not large. This is not the typical case for oversampled signals from  $\Delta\Sigma$  modulators, where oversampling rates of 64 are common. In such cases, the rate conversion is performed in multiple stages, with the high-rate stage being

implemented by a comb filter [1,2,6-8].

The comb filters are low-pass FIR filters whose transfer function is

$$H(z) = \left( \frac{1}{r} \sum_{i=0}^{r-1} z^{-i} \right)^N = \left( \frac{1}{r} \cdot \frac{1-z^{-r}}{1-z^{-1}} \right)^N \quad (8.5)$$

Since all their coefficients are unity, they are easily implemented. Also, the second form of (8.5) is very hardware efficient [6,8]. The denominator is implemented as an integrator at the high sampling rate, while the numerator as a differentiator at the low sampling rate. The order of the comb filter is chosen equal to the conversion rate  $r$ , so that the zeros of (8.5) are at the multiples of  $f_s/r$ . The frequency response of a comb filter with  $r = 16$  and  $N = 1$  is depicted in Fig. 8.8.

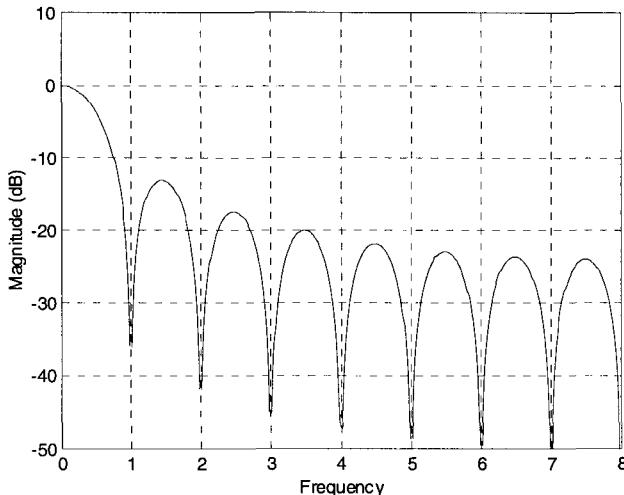


Fig. 8.8 Magnitude response of the 16-tap comb filter, which is used for rate change by 16. The frequency is normalized to the low sampling rate.

Evidently these filters can be used for  $R$ -times rate conversion. Increasing  $N$  results to higher stop-band attenuation.

The drawback of using comb filters for rate conversion is their large pass-band droop, evident in Fig. 8.3.b. For this reason they are only used as a first (high-rate) stage rate conversion filter, where the actual signal

pass-band edge is at least two times less than the pass-band edge ( $f_s / 2r$ ) of the comb filter.

### 8.3.3 Fractional rate conversion

In many cases the resampling factor is not integer, but it is a ratio  $N/M$ , where  $N$  and  $M$  are integers. In this case, the resampling is achieved by interpolating by  $N$  and decimating by  $M$ .

**Example 8.4** The sampling rate used to store music in a CD is approximately 44 kHz, while the rate used in a Digital Audio Tape (DAT) is 48 kHz. How can a CD be recorded onto digital tape?

**Solution** The interface between the two systems comprises a resampler from the CD rate to the digital tape rate. Hence interpolation by  $48/44 = 12/11$  is needed.

As already discussed in the previous sections, the decimation and interpolation are both performed in stages, hence the numerator and denominator of the resampling ratio are written as products of prime numbers

$$\frac{N}{M} = \frac{\prod_{i=1}^K N_i}{\prod_{i=1}^L M_i} \quad (8.6)$$

Next, the order of the  $K$  interpolations by  $N_i$  and the  $L$  decimations by  $M_i$  has to be determined. Two rules should apply to this determination

- The sampling rate at any given step must not be lower than the final sampling rate. Due to aliasing, the bandwidth of the signal is constrained at every decimation step. This low-pass filtering must not constrain the bandwidth more than the final sampling rate dictates.
- The sampling rate at any given step should be kept minimum, for computational efficiency: higher sampling rates require circuits that consume more power, and thus any increase of the sampling rate more than it has to be should be avoided.

**Example 8.5** Assume that a resampling ratio of 14/15 is needed. Demonstrate a system to efficiently achieve this.

**Solution** As it is  $14/15 = (2 \cdot 7)/(3 \cdot 5)$ , four resampling steps are needed. If interpolation by  $N$  is denoted as  $\uparrow N$  and decimation by  $N$  is denoted as  $\downarrow N$ , then the four resampling steps are  $\uparrow 2, \uparrow 7, \downarrow 3, \downarrow 5$ . All 12 possible sequences of these steps are listed and tried:

1.  $\downarrow 3$ , rate less than final
2.  $\downarrow 5$ , rate less than final
3.  $\uparrow 2, \downarrow 5$ , rate less than final
4.  $\uparrow 2, \downarrow 3$ , rate less than final
5.  $\uparrow 2, \uparrow 7, \downarrow 5, \downarrow 3$
6.  $\uparrow 2, \uparrow 7, \downarrow 3, \downarrow 5$
7.  $\uparrow 7, \downarrow 5, \downarrow 3$ , rate less than final
8.  $\uparrow 7, \downarrow 5, \uparrow 2, \downarrow 3$
9.  $\uparrow 7, \downarrow 3, \downarrow 5$ , rate less than final
10.  $\uparrow 7, \downarrow 3, \uparrow 2, \downarrow 5$
11.  $\uparrow 7, \uparrow 2, \downarrow 5, \downarrow 3$
12.  $\uparrow 7, \uparrow 2, \downarrow 3, \downarrow 5$

Sequences 1-4, 7 and 9 result in a temporary rate that is less than the final sampling rate, hence introducing stricter anti-aliasing filtering than needed. Sequences 5, 6, 11 and 12 are the same: first interpolate by 14 and then decimate by 15. The other two possible sequences are 8 and 10:  $\uparrow 7, \downarrow 5, \uparrow 2, \downarrow 3$ , or  $\uparrow 7, \downarrow 3, \uparrow 2, \downarrow 5$ . From these three possibilities, the system that performs  $\uparrow 7, \downarrow 5, \uparrow 2$  and finally  $\downarrow 3$  is selected, as this sequence of resampling steps keeps the intermediate rates at minimum.

## 8.4 $\Delta\Sigma$ Modulators for ADC

$\Delta\Sigma$  modulators are primarily used in the design of data converters [7]. High precision converters can be achieved using noise shaping and oversampling instead of high precision components. The following example demonstrates the needed accuracy of conventional data converters.

**Example 8.6** What is the accuracy of a 16-bit DAC operating with a full scale voltage of 2 V?

**Solution** The required accuracy is  $\frac{1}{2}$  of the LSB, which corresponds to  $2 \cdot 2^{-17} \text{ V} = 15 \mu\text{V}$ . On a 50 Ohm termination, this is comparable to the thermal noise in a 1 GHz bandwidth!

An ADC system employing a  $\Delta\Sigma$  modulator is shown in Fig. 8.9. The  $\Delta\Sigma$  modulator can either be discrete-time or continuous-time. In the former case the sampling is performed before the noise-shaping loop (as depicted in Fig. 8.9), while in the latter, at the quantizer embedded in the loop. In every case the signal is oversampled, allowing for a very relaxed anti-aliasing analog filter. The decimator that follows the modulator reduces the sampling frequency close to the Nyquist rate, while it suppresses the out-of-band quantization noise. The desired resolution of the ADC dictates the oversampling ratio, the order of the  $\Delta\Sigma$  modulator and the level of noise suppression the decimating filters should provide.

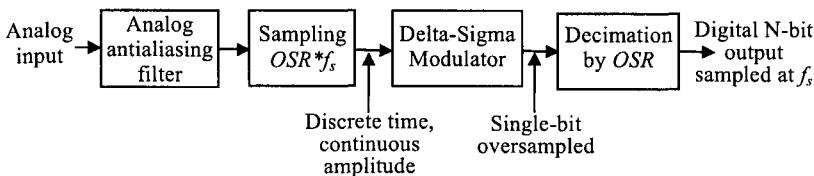


Fig. 8.9 ADC system based on  $\Delta\Sigma$  modulation. The Nyquist rate is  $f_s$ , while the  $\Delta\Sigma$  modulator oversampling ratio is  $OSR$ .

**Example 8.7** Design an 8-bit ADC for audio signals employing  $\Delta\Sigma$  modulation. The desired sampling rate is  $f_s = 48 \text{ kHz}$ .

**Solution** Since the signals to be converted are audio, their bandwidth  $B$  is approximately 22 kHz. The low signal frequency easily allows for large oversampling ratios to be used, hence a second order  $\Delta\Sigma$  modulator can provide the desired SNR of  $8 \text{ bits} \cdot 6 \text{ dB/bit} = 48 \text{ dB}$ .

From the SNR versus input amplitude plot of Fig. 3.5, the peak SNR of a second-order  $\Delta\Sigma$  modulator with oversampling ratio  $r' = 64$  and ideal ‘brick-wall’ decimation is  $SNR' = 72 \text{ dB}$ . Also, from (3.19) the in-band portion of the quantization noise of the second-order  $\Delta\Sigma$  modulator

is inversely proportional to the fifth order of the oversampling ratio. Hence the required  $SNR = 48$  dB is ideally achieved employing oversampling  $r$ , given by

$$SNR/SNR' = (r/r')^5$$

Hence  $r_1 = 21.2$ .

But a practical decimator does not employ ‘brick wall’ filters. Their finite stop-band attenuation and pass-band ripple reduce the SNR. Thus an oversampling ratio  $r_1 = 32$  is chosen, which allows for the reduction of the SNR due to the imperfect decimation filters. Hence the sampling rate for the modulator is  $f_{s,m} = r \cdot f_s = 1.536$  MHz.

Due to the large oversampling ratio, the required analog anti-aliasing filter is quite relaxed. Its cut-off frequency is at  $B$ . Its ripple should be low to avoid distortion, and is set to 0.01 dB. Its stop-band attenuation should be more than the required SNR, and to allow for the various sources of error, it is chosen far below that, at -60 dB. It should be achieved at  $f_{s,m} - B = 1.514$  MHz. This sort of performance is achieved with a third-order Butterworth filter. Note that the anti-aliasing filter needed if a Nyquist rate converter is used is impossible, as its estimated order is 46!

Two-stage decimation is employed. The first filter reduces the rate by  $r_1 = 16$  and the second by  $r_2 = 2$ . The first-stage decimator is the cascade of two 16-tap comb filters ( $N=2$ ). These attenuate the quantization noise that will be folded down into the base-band, as the nulls of the filter’s response are at multiples of  $f_{s,m}/r_1$ . The second-stage decimator is a half-band FIR filter with cut-off frequency at  $B$ , ripple of 0.01 dB and 20 dB of attenuation achieved from  $f_s - B = 26$  MHz onwards. The order of the filter is 32. The performance of the system is depicted in Fig. 8.10. The required 48 dB of SNR are only marginally achieved. Matters are greatly improved if the first-stage decimator is comprised of three comb filters in cascade ( $N=3$ ). Then the degradation due to the decimators from the ideal SNR (using ‘brick-wall’ filters) is very small, and the specifications for the ADC are easily exceeded. Usually, the number  $N$  of comb filters is chosen equal to the order of  $\Delta\Sigma$  modulator plus one ( $N=L+1$ ) [1].

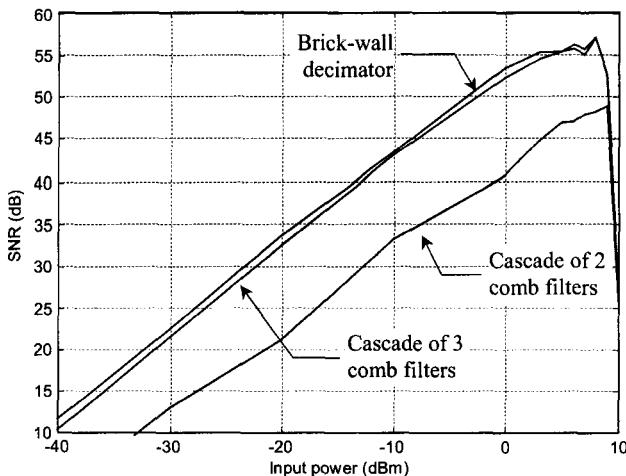


Fig. 8.10 Performance of the ADC system based on  $\Delta\Sigma$  modulation designed in Example 8.7. The SNR resulting from two different first-stage decimators is compared to the ideal, when ‘brick-wall’ decimation filters are used.

## 8.5 $\Delta\Sigma$ Modulators for DAC

High accuracy DACs can be implemented using  $\Delta\Sigma$  modulators in much the same way as  $\Delta\Sigma$  ADCs, by trading the accuracy of the analog components for speed of the digital circuits. A DAC system employing a  $\Delta\Sigma$  modulator is shown in Fig. 8.11.

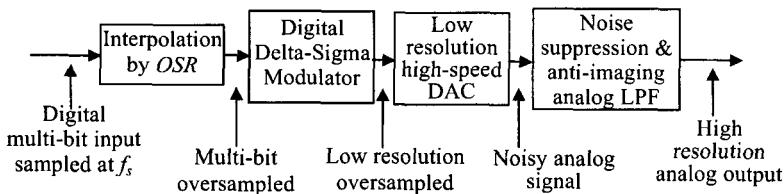


Fig. 8.11 DAC system based on  $\Delta\Sigma$  modulation. The Nyquist rate is  $f_s$ , while the  $\Delta\Sigma$  modulator oversampling ratio is  $OSR$ .

The multi-bit input signal is oversampled and then modulated using a digital  $\Delta\Sigma$  modulator. The modulated signal is a low-resolution (most times single-bit) oversampled digital signal. A low-resolution high-speed

DAC converts this signal to analog. In the case of single-bit  $\Delta\Sigma$  modulator, the DAC is single-bit, which is inherently linear. An analog LPF follows the DAC, which suppresses the images of the DAC and the shaped quantization noise of the  $\Delta\Sigma$  modulator. The loop filter of a digital  $\Delta\Sigma$  modulator is digital. Its quantizer is implemented by means of truncating to the M most significant bits, which constitute the feedback signal [1].

The loop of a  $\Delta\Sigma$  modulator employed in a DAC system is digital, eliminating the need for matching between the digital and analog parts of a multi-stage  $\Delta\Sigma$  modulator. The disadvantage of using such a  $\Delta\Sigma$  modulator is that the output is not single-bit.

## 8.6 Summary

Rate conversion (decimation and interpolation) and the associated digital low-pass filters were considered in the first part of this chapter. Special emphasis was given on multi-stage rate conversion, on the properties of comb filters as well as on the way fractional rate conversion is achieved. Finally, the decimation and interpolation filters were used together with  $\Delta\Sigma$  modulators to build ADC and DAC systems respectively.

## Problems

- 8.1 Design four comb filters with  $R = 16$  and  $N=1,2,3,4$ . Observe the improvement of the stop-band attenuation and the increased pass-band droop.
- 8.2 A low-pass filter is to be used in a decimation process. The original sampling rate is 80 MHz. The final sampling rate will be 10 MHz. The signal band should be restricted to 4 MHz (cut-off frequency) and the stop-band with attenuation 60 dB should start at 5 MHz. Using MATLAB evaluate the order of the equiripple FIR filter to be used as a decimator. Evaluate the order of the equiripple FIR filters employed in case the decimation is to be performed in two successive stages.
- 8.3 Assume that a resampling ratio of 16/5 is required. Demonstrate a system to achieve this efficiently.

## References

- [1] S. Norsworthy, R. Schreier and G. Temes, *Delta-Sigma Data Converters, Theory, Design and Simulation*, New York, IEEE Press, 1997.
- [2] E. Ifeachor and B. Jervis, *Digital Signal Processing. A Practical Approach*, 2<sup>nd</sup> edition, Prentice Hall, New York, 2002.
- [3] R.E. Crochiere and L.R. Rabiner, *Multirate Digital Signal Processing*, Prentice-Hall, Englewood Cliffs, NJ, 1983.
- [4] ETSI TR 101 475: "Broadband Radio Access Networks (BRAN); HIPERLAN Type 2; Physical (PHY) layer", 1999.
- [5] U. Razavi, *RF Microelectronics*, Prentice Hall PTR, Upper Saddle River, 1998.
- [6] A. Kwentus, Z. Jiang and A. Willson, 'Application of filter sharpening to cascaded integrator-comb decimation filters', *IEEE Trans. Signal Processing*, Vol. 45, pp. 457-467, February 1997.
- [7] MOTOROLA, *Principles of Sigma-Delta Modulation for Analog-to-Digital Converters*, APR8/D, Rev. 1.
- [8] S. Chu and C.S. Burrus, 'Multirate Filter Designs Using Comb Filters', *IEEE Trans. on Circuits and Systems*, Vol. CAS-31, No. 11, pp. 913-924, Nov. 1984.

## **Chapter 9**

# **Applications**

### **9.1 Introduction**

As it was stated in the introductory chapter, the applications of  $\Delta\Sigma$  modulation are numerous. One main reason for this is its simplicity and economic implementation. In Chapter 8 we described the use of  $\Delta\Sigma$  modulation in the design of ADC and DAC. Among the many other useful applications, we choose to describe a few that, we consider, to be of interest to most of the readers of this book.

We start in section 9.2 with the application of  $\Delta\Sigma$  modulation in digital radio, because of its high importance in communications today. Next, in sections 9.3 and 9.4 the application of  $\Delta\Sigma$  modulation, in combination with a Phase Locked-Loop, is described in frequency synthesis and in the implementation of digitally programmable clocks with reduced jitter useful in the switched-capacitor filters. Frequency and phase demodulation using  $\Delta\Sigma$  modulation are explained in section 9.5. Finally, the use of  $\Delta\Sigma$  modulation in the implementation of oscillators with high accuracy in the generated frequency, is presented in section 9.6.

### **9.2 $\Delta\Sigma$ Modulation in Digital Radio**

Wireless transmitters up-convert the base-band signal to some Radio Frequency (RF) where wireless transmission is possible. They control the spectrum and power of the transmitted RF signal according to the communication standard they implement. Conversely, wireless receivers

down-convert the received RF signal to base-band. To do so they amplify the received weak signal and filter it to get rid of interferers. The RF signals are analogue, while in modern communication systems, base-band processing is carried out digitally. Hence at some point in the transmitter a DAC is used, whereas at some point in the receiver an ADC is used.

Globalization of modern telecommunications has raised the need for multi-mode mobile handsets, i.e. capable of operation with more than one standard. Thus either multiple branches, or programmable components are employed. Also, the cost, size and power consumption of such handsets should be kept low. These needs force the number of analogue components in the handsets to the bare minimum, placing the burden on the digital part of the radio, where modern DSPs can handle it. These goals form the concept of digital radio [1,2]. Concentrating on the receiver of digital radios, there are two additional design goals, namely the recovery of RF signals having a wide range of power and the suppression of strong nearby interferers.

Selectivity is needed to suppress interferers and accentuate the intended channel. Should this be provided at base-band, the bandwidth and resolution of the ADC have to be large. Also it is not possible to provide the necessary selectivity at RF; existing RF filters select bands of communication standards, not their individual channels. The way around the selectivity problem is the super-heterodyne receiver. The frequency of the received signal can be lowered using a sequence of down-conversion steps, down to Intermediate Frequencies (IF). At each IF stage the pass-band of the filters is located at lower frequency, hence they can be more narrow-band. The various IF filters collectively provide the needed selectivity.

To account for the wide dynamic range of the RF signal, the dynamic range of the ADC should be accordingly large. A way around this is to provide variable amplification at the RF and IF sections, large for small input signals and vice versa. Then the dynamic range of the signal at the input of the ADC is reduced. This is achieved using Variable Gain Amplifiers (VGA) controlled by Automatic Gain Control (AGC) loops.

Alternatively, new architectures that employ direct down-conversion to base-band (zero-IF), or almost to base-band (low-IF) [3] can be em-

ployed. Although the novel receiver architectures are the ultimate goal of digital radio, they impose strict matching requirements to the analogue components [1–3]. Hence such architectures are used for cheap, low-performance equipment like pagers, and are for now unsuitable for implementing the most demanding mobile communication standards. Thus the novelty in the receivers for digital radio lies in the position of the ADC in the chain. Traditionally two ADCs convert the dual base-band signals (I, in-phase and Q, quadrature), but recent technology advances allow IF digitization [2] using conventional converters or  $\Delta\Sigma$  modulators [2,4]. A typical heterodyne receiver with IF digitization is shown in Fig. 9.1.

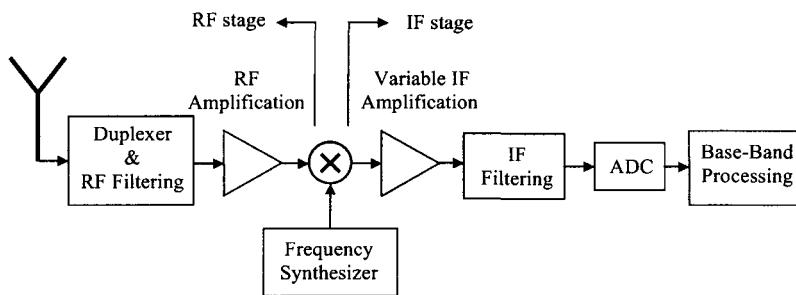


Fig. 9.1. Typical heterodyne receiver with IF digitization.

The IF stage in a heterodyne receiver is necessary to achieve the wanted selectivity of the receiver [3], as narrow-band band-pass filters are not achievable at RF. Hence after the selectivity and amplification of the RF stage, a mixer down-converts the signal to IF, for further band-limiting and amplification. At the RF stage the primary concerns are noise and the selection of the received band of a standard (for example a GSM duplexer and a Low Noise Amplifier). At the IF stage variable gain amplification and (partial) channel selection are sought. To facilitate IF digitization, anti-aliasing filtering is also provided by the IF filter(s).

The digital radio imposes many constraints to the ADC. One constraint is on the power consumption that is very important for the battery life of the handset. Also, to support multi-standard operation, the ADC

has to accommodate a variety of conversion bandwidths, yielding a variety of resolutions. Finally, the ADC has to cope with the large input frequency, as typical IF values for heterodyne receivers are of the order of tens of MHz. As it will be shown in the next two subsections, the last two constraints can be efficiently covered using band-pass  $\Delta\Sigma$  modulators. Also, the decimator that follows the  $\Delta\Sigma$  modulator can serve two roles: suppress the shaped out-of-band noise, and perform the channel selection, with the same circuit.

### 9.2.1 Conversion bandwidth and resolution

The channel bandwidths of modern mobile communications standards ranges from 200 kHz (GSM-like) to 20 MHz (WLAN). Given the channel positioning at IF achieved by the mixer, this is the bandwidth to be converted to digital. On the other hand, the required conversion resolution is adjusted by two factors: the dynamic range ( $DR_{RF}$  in dB) which by definition is the maximum amplitude ( $A_{\max}$ ) minus the minimum ( $A_{\min}$ ) of the RF signals, i.e.:

$$DR_{RF} = A_{\max} - A_{\min} \quad (9.1)$$

and the range of the IF variable gain amplification stage. The latter is important, as large signals can be amplified less by a gain factor  $g_{\max}$ , and small signals can be amplified more by  $g_{\min}$ , resulting to a signal at the input of the converter that has smaller dynamic range. To demonstrate this, assume that the constant gain of the receiver is  $g$  and the gain range of the IF variable gain amplifier is  $DR_{IF}$ , i.e.

$$DR_{IF} = g_{\min} - g_{\max} \quad (9.2)$$

The difference between the maximum and minimum amplitude of the signals at the input of the ADC (in dB) is

$$DR_{ADC} = (A_{\max} + g + g_{\max}) - (A_{\min} + g + g_{\min}) = DR_{RF} - DR_{IF} \quad (9.3)$$

As the RF dynamic range grows or the IF amplifier gain range lessens, the needed resolution of the converter grows. Considering a constant IF amplifier gain range (e.g. to the maximum value allowed by the amplifier), then the needed resolution is proportional to the RF dynamic range.

Usually the mobile telecommunications standards that have large RF dynamic range also have small channel bandwidth. Hence the ADC should be able to adjust a resolution – bandwidth trade off. Such a trade off is easily adjusted using a  $\Delta\Sigma$  modulator as an ADC. Spreading the NTF zeros in a wider band will increase the converted bandwidth, but will also make the noise shaping less favorable, reducing the resolution. This is demonstrated in the next example.

**Example 9.1** Assume a standard with channel bandwidth of 200 kHz. The channel is digitized using a fourth-order band-pass  $\Delta\Sigma$  modulator. What is the effect on the resolution if the channel bandwidth is doubled?

**Solution** The oversampling ratio is

$$R = \frac{f_s}{2B} \quad (9.4)$$

The fourth-order band-pass  $\Delta\Sigma$  modulator is obtained by transforming a second-order low-pass  $\Delta\Sigma$  modulator by  $z \rightarrow -z^2$  as it was explained in Secs. 3.11 and 5.7. The in-band noise power is

$$P_{e,in} = \frac{P_e}{2\pi} \int_{\pi/2-\pi/2R}^{\pi/2+\pi/2R} \left| NTF_{BP}(e^{-j\theta}) \right|^2 d\theta = \frac{P_e}{2\pi} \int_{-\pi/R}^{\pi/R} \left| NTF_{LP}(e^{-j\theta}) \right|^2 d\theta \quad (9.5)$$

Assume optimal placing of the NTF zeros of a second-order low-pass  $\Delta\Sigma$  modulator, from Table 4.1, the frequencies of the roots are  $\phi = \pm 0.5774\pi/R$ . Following similar reasoning as in Sec. 4.6.1, the in-band noise power is proportional to

$$P_{e,in} \propto \int_0^{\pi/R} \left| \phi^2 - \theta^2 \right|^2 d\theta = \frac{4}{45} \left( \frac{\pi}{R} \right)^5 \quad (9.6)$$

As the achieved SNR is inversely proportional to  $P_{e,in}$ , it gives:

$$SNR \propto \frac{1}{P_{e,in}} \propto \frac{45}{4} \left( \frac{R}{\pi} \right)^5 = \frac{45 f_s^5}{4 \cdot 32 \pi^5} \frac{1}{B^5} \quad (9.7)$$

Hence the achievable SNR is inversely proportional to the fifth power of the conversion bandwidth. By doubling the bandwidth, the SNR is reduced by a factor of 32 or 15dB as in usual second-order  $\Delta\Sigma$  modulators.

### 9.2.2 Sampling rate

A band-pass  $\Delta\Sigma$  modulator that results from an equivalent low-pass via the transformation  $z \rightarrow -z^2$  converts a band centered at one fourth of its sampling frequency. In literature various implementations of band-pass  $\Delta\Sigma$  modulators can be found [5–15]. However, an IF of 70 MHz would require a sampling frequency of 280 MHz, which is too high for the modulator<sup>1</sup>, even for the following DSP to handle.

The solution to this problem is to sub-sample the signal, prior to modulating it. In this case the modulator has to have a discrete-time analog part, hence it is implemented using SC circuits. The advantage of this approach is that by band-pass sampling the signal, it is centered around one fourth of the selected sampling frequency. This sampling frequency can be very small, provided that

- a. it remains two times larger than the bandwidth,
- b. the IF filters can provide the necessary anti-aliasing and
- c. the resulting oversampling ratio is sufficient.

---

<sup>1</sup> Note that in [11,14], modulators capable of handling high sampling rates has been presented (3.8 GHz in [11]!), but these results should be considered far from the industrial standard.

Given a signal centered at the IF frequency  $f_{IF}$ , it can be sampled (see problem 2.5) by a set of frequencies given by

$$f_s(k) = \frac{4f_{IF}}{2k+1}, \quad i=0, 1, \dots, k_{\max} \quad (9.8)$$

where the maximum value  $k_{\max}$  is determined by the constraint  $f_s(k_{\max}) \geq 2B$ . Hence

$$k_{\max} = \left\lfloor \frac{f_{IF}}{B} - 0.5 \right\rfloor \quad (9.9)$$

where  $\lfloor x \rfloor$  is the largest integer smaller or equal to  $x$ . Consequently, anti-aliasing filtering must provide the necessary attenuation at the closest image, which lies  $\Delta f(k)$  away, where

$$\Delta f(k) = \frac{1}{2} \left( \frac{4f_{IF}}{2k+1} - B \right), \quad i=0, 1, \dots, k_{\max} \quad (9.10)$$

Finally, the oversampling ratio as a function of  $k$  is

$$R(k) = \frac{2f_{IF}}{(2k+1)B}, \quad i=0, 1, \dots, k_{\max} \quad (9.11)$$

**Example 9.2** The necessary anti-aliasing attenuation, as extracted from the GSM specifications, is 68 dB. The TOKO SA070WA-010 Surface Acoustic Wave filter has a pass-band centered at 70 MHz and provides 68 dB of attenuation at 8 MHz offset from the center of the band. The GSM channel bandwidth is 200 kHz. What is the minimum sampling frequency and the resulting oversampling ratio?

**Solution** From Eq. (9.10), given that the frequency offset  $\Delta f$  is 8 MHz, the maximum  $k$  is

$$k = \left\lfloor \frac{f_{IF}}{\Delta f + B/2} - 0.5 \right\rfloor = 8 \quad (9.12)$$

Then, from Eq. (9.8) the minimum sampling frequency is 16.47 MHz and the associated oversampling ratio is 41.18.

In the above example, the sampling frequency has been set using the constraint imposed by the anti-aliasing filter and employing Eq. (9.10). Similarly the DSP complexity can bound the sampling frequency, in which case Eq. (9.8) is used to find the suitable value of  $k$ . Finally, the modulator may bound the oversampling ratio for a given resolution, whereupon Eq. (9.11) is used.

Note that the value of  $k$  cannot be very high, resulting to many times sub-sampling, as this operation leads to noise aliasing [1]. Although the circuitry has provided enough attenuation up to the sampling frequency to make the noise at this point non-critical, care should be taken to constrain the noise introduced in the receiver.

### 9.3 Frequency Synthesis

An interesting application of  $\Delta\Sigma$  modulation is found in frequency synthesis [16] as well as in Frequency and Phase Demodulation [17]. In these cases it is combined with a Phase-Locked Loop (PLL). Before explaining how these can be achieved we give a short description of the PLL and its operation.

#### 9.3.1 *Principles of the PLL operation*

The Phase-Locked Loop [18] is a very useful circuit in signal processing. It is shown in a simplified block diagram in Fig. 9.2. It consists of a Phase-Detector, a Loop Filter, an Amplifier and a Voltage-Controlled Oscillator (VCO). The output signal from the PLL is proportional to the phase difference of two signals, the input signal and the feedback signal coming out of the VCO. If the VCO signal is in phase with the input signal, there will be no output signal from the phase detector and therefore no output from the PLL. However, when the two signals, input and VCO output, are not in phase there will be an output signal from the phase detector, which is low-pass filtered producing a dc voltage proportional to the phase difference. A difference in frequency can be also described as a

phase difference in which case the signal at the output of the phase detector will be proportional to the difference in frequency. The dc output signal from the low-pass filter will be amplified and then applied to the VCO input in such a way that it will cause the frequency of the VCO output signal to become equal to the frequency of the input signal. When this is achieved we say that the PLL is in lock.

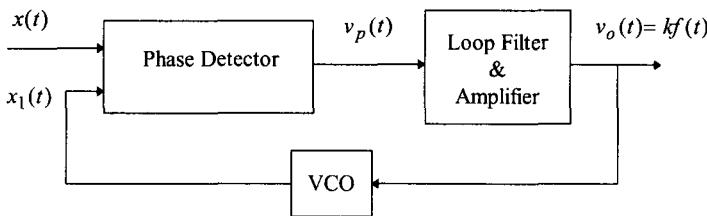


Fig. 9.2 Block diagram of PLL.

Without an input signal the VCO is free running. When the input signal is applied, the PLL operates trying to capture the frequency of the input signal and locks in it. To achieve this the frequency of the input signal has to be within the capture range of the circuit. Also the input signal to the VCO should be of either polarity in order for the VCO to follow slow variations in the frequency of the input signal.

As an example, consider that the input signal is the Frequency Modulated (FM) signal

$$x(t) = A \cos(\theta(t)) \quad (9.13)$$

Clearly, the frequency  $f$  as a function of time will be

$$f(t) = \frac{1}{2\pi} \frac{d\theta}{dt} = f_c + \Delta f \sin(2\pi f_m t) \quad (9.14)$$

where  $f_c$  is the centre frequency,  $\Delta f$  the maximum frequency deviation and  $f_m$  the modulating frequency.

In order to describe mathematically the operation of the PLL we redraw Fig. 9.2 as shown in Fig. 9.3. In Fig. 9.3.a the input signal is the Laplace Transform  $\Theta(s)$  of the phase  $\theta(t)$  measured in radians, while in Fig. 9.3.b the input signal is the Laplace Transform  $\Omega(s)$  of the frequency  $\omega(t)$ . Note that

$$\Omega(s) = s\Theta(s) \quad (9.15)$$

with the output signal given in Volts.

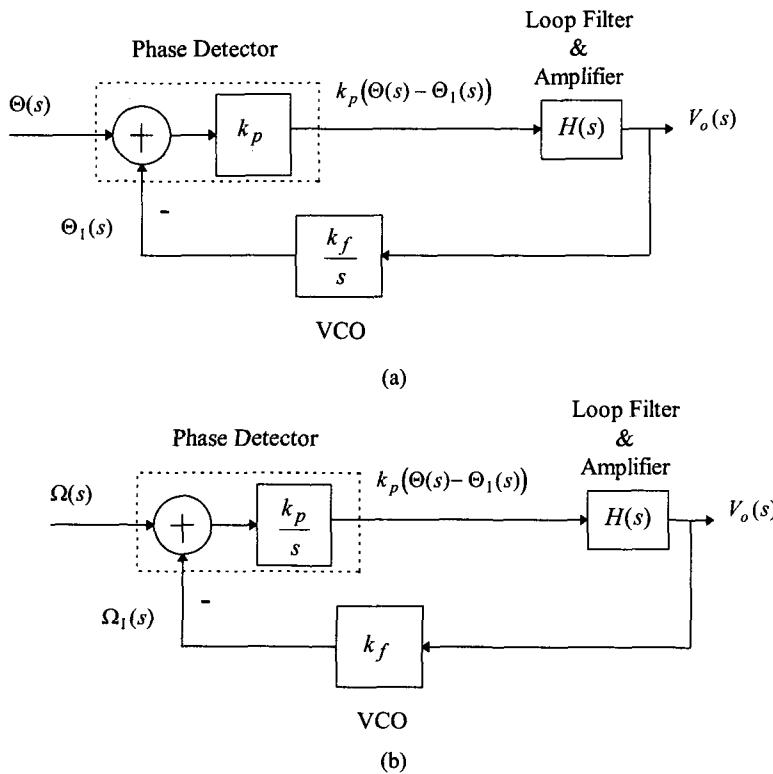


Fig. 9.3 Equivalent PLL block diagrams useful for the analysis of its operation.

In Figs. 9.3.a and 9.3.b, the  $H(s)$  function determines the dynamics of the system, i.e. how the frequency lock-in will be achieved with regard to the initial conditions (phase and frequency) in the VCO (acquisition mode). Also the selection of  $H(s)$  will determine the steady-state error  $\lim_{t \rightarrow \infty} (\theta(t) - \theta_l(t))$ , when the PLL will be operating in the tracking mode. Since this error is desired to be zero, the second-order PLL is mainly used instead of the first-order PLL, the PLL order being the same as that of the loop filter.

### 9.3.2 $\Delta\Sigma$ modulation in frequency synthesis

The purpose of Frequency Synthesis is to generate a frequency  $f_{synth}$  from a reference frequency  $f_{ref}$  with the ratio of the two frequencies being a rational number  $R$ , i.e.

$$f_{synth} = R f_{ref} \quad (9.16)$$

One of the most useful methods in Indirect Frequency Synthesis is based on the use of the PLL [19,16]. Such a system is shown in Fig. 9.4.

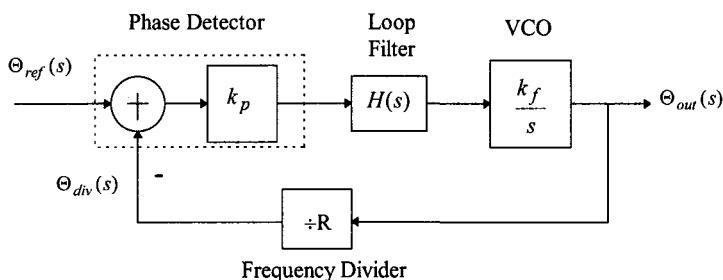


Fig. 9.4 Indirect frequency synthesis based on the use of a PLL.

The basic problem of this method is the implementation of the frequency divider when  $R$  is not an integer. Let us assume that

$$R = r + \frac{k}{m} \quad (9.17)$$

where  $r$  is the integer part of  $R$  and  $k, m$  themselves integers. The frequency division by the rational number  $R$  can be approximately achieved, if on some occasions the division is performed by  $r$  and on the other occasions by  $r+1$ . However this type of operation introduces some kind of noise into the system, which causes the output frequency not to be constant with value  $f_{synth}$ , but to vary in time, assuming the value, say,  $f_{out}(t)$ .

In order to minimize these variations of  $f_{out}(t)$  one may employ  $\Delta\Sigma$  modulation. In such a case the block diagram of the frequency divider becomes as shown in Fig. 9.5. This consists basically of a dual modulus counter, which divides the VCO output frequency some times by  $r$  and some by  $r+1$ , depending on the state of the control input signal.

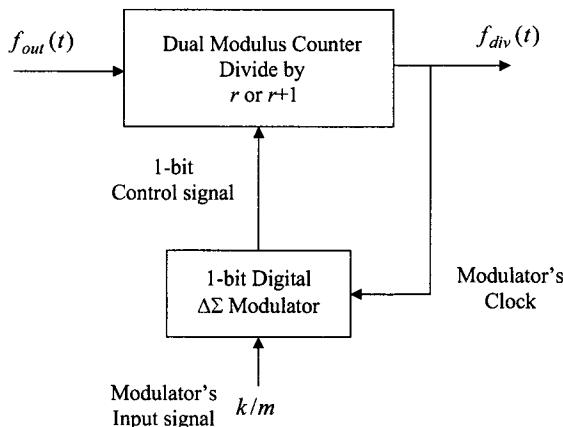


Fig. 9.5 Frequency divider by a rational number ( $\div R$ ) employing  $\Delta\Sigma$  modulation.

The system shown in Fig. 9.5 implements a frequency divider by a rational number. The analysis of its operation will be examined in the following subsection.

### 9.3.3 Analysis of the frequency divider

The division of a constant frequency can be implemented simply by a counter modulo  $r$ . On the other hand the implementation of the division of the varying frequency  $f_{out}(t)$  by an integer can be achieved only approximately. The result is as shown in Fig. 9.6 assuming that the decimation is always by the factor  $r$ .

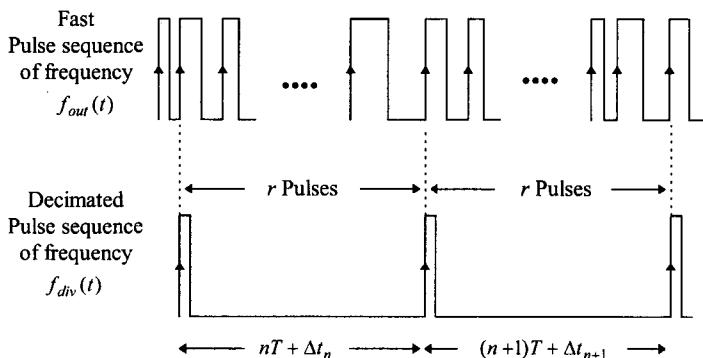


Fig. 9.6 Frequency division by a constant integer.  $T$  is the mean period of the pulses in the decimated sequence

The approximate result is due to the fact that, with this simple frequency division, the period of each created pulse is  $r$  times longer than the mean period of the pulses in the preceding frame of  $r$  pulses. Clearly, for the system to operate as a divisor of the instantaneous frequency, the variation of  $f_{out}(t)$  in the time period of the  $r$  pulses should be small. When the period of the successive pulses changes quickly with time, the division of the instantaneous frequency is meaningless. However, if the variation of their period is relatively small, we may consider that the period is nearly constant and equal to the mean value of the periods, say  $T_{synth}$ , which corresponds to a constant frequency equal to  $f_{synth}$ . The time instances  $t_n$ , when a pulse of the decimated sequence appears, will be

$$t_n = nT + \Delta t_n \quad (9.18)$$

These pulses appear periodically with a mean period  $T$ , which will evidently be equal to  $rT_{synth}$ , while there will be some uncertainty in the time of the arrival of each pulse.

Equation (9.18) is also valid when the decimation is performed some times by  $r$  and some other times by  $r+1$ . In this case the uncertainty  $\Delta t_n$  in the time of arrival of the decimated pulses will be due either to the uncertainty in the period of the fast pulse sequence or to the uncertainty in the variation of the decimation factor ( $r$  or  $r+1$ ). In this case the period  $T$  will be

$$T = \lim_{N \rightarrow \infty} \left( \frac{1}{N} \sum_{n=0}^N r(n) T_{synth} \right) = E\{r(n)\} \cdot T_{synth} \quad (9.19)$$

where  $E\{r(n)\}$  is the mean value of  $r(n)$ . But

$$E\{r(n)\} = p(r+1) + (1-p)r \quad (9.20)$$

where  $p$  is the probability for the division by  $r+1$  and  $1-p$  the probability for the division by  $r$ .

The  $\Delta\Sigma$  modulator used in the frequency divider is shown in Fig. 9.7. This modulator encodes the constant integer  $k$  in a  $\Delta\Sigma$  sequence of digits 1 and 0 which correspond to the integers  $m$  and 0 respectively. The probability that 1 appears at the output of the comparator is equal to

$$p = \frac{k}{m} \quad (9.21)$$

This corresponds to the probability for dividing by  $r+1$ , since the  $\Delta\Sigma$  modulator output controls the Modulus Counter in Fig. 9.5. Substituting from (9.21) into (9.20) and the result in equation (9.19) gives the following:

$$T = \left( r + \frac{k}{m} \right) T_{synth} = R T_{synth} \quad (9.22)$$

Therefore the system in Fig. (9.5) in actual fact operates as a divider of the frequency by  $R$ .

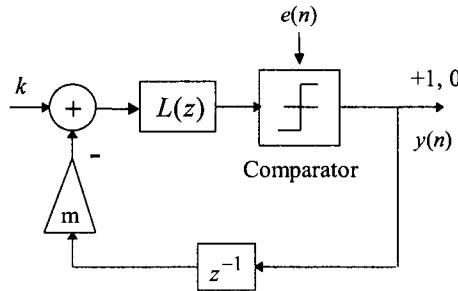


Fig. 9.7 Block diagram of the  $\Delta\Sigma$  modulator used in the frequency divider.

As far as the uncertainty  $\Delta t_n$  in the time of appearance of the pulse is concerned, this corresponds to a sampled noise in the phase

$$e_\theta(nT) = 2\pi \frac{\Delta t_n}{T} = \frac{2\pi}{T} \left( \sum_{i=0}^n r(i)T_{synth} - nRT_{synth} \right) \quad (9.23)$$

But

$$r(i) = r + y(i) = r + \frac{k}{m} + \left( y(i) - \frac{k}{m} \right) = R + q(i) \quad (9.24)$$

where  $q(i)$  are the samples of the quantization noise at the  $\Delta\Sigma$  modulator output. Substituting from equations (9.22) and (9.24) in (9.23) gives

$$e_\theta(nT) = \frac{2\pi}{R} \sum_{i=0}^n q(i) \quad (9.25)$$

Applying z-transform gives

$$E_\theta(z) = \frac{2\pi}{R} \frac{Q(z)}{1-z^{-1}} = \frac{2\pi}{R} \frac{NTF(z)}{1-z^{-1}} E(z) \quad (9.26)$$

Thus, it can be seen that the error in phase is shaped by the NTF of the  $\Delta\Sigma$  modulator and pushed to higher frequencies, which are subsequently dropped by the low-pass filter that is placed in the forward path of the PLL shown in Fig. 9.4. As a result the synthesized frequency  $f_{out}(t)$  is almost constant and equal to the desired  $f_{synth}$ .

#### 9.4 Clock Generators Using $\Delta\Sigma$ Modulation

It is possible to obtain frequency synthesis based on the use of  $\Delta\Sigma$  modulation without the use of a PLL [20] at the same time. In this case the variations of the synthesized frequency are not so small as in the case of the method that was presented in the previous subsection. However the spectrum of these variations is pushed far away from the desired frequency, and the method becomes useful in different applications. This method is as follows:

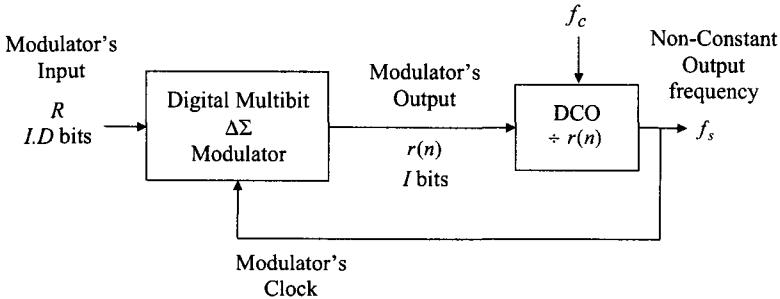


Fig. 9.8 System for the generation of clock pulses with increased but shaped jitter noise.

The system consists of a digital  $\Delta\Sigma$  modulator, which is multi-bit in order to obtain sharper noise shaping, combined with a digital controlled oscillator (DCO) as it is shown in Fig. 9.8. The operation of the DCO is

much the same as the operation of the dual modulus counter as was described in Sec. 9.3.2. The number  $r(n)$  is the instant decimation factor of a high rate pulse train, of constant frequency  $f_c$ . This can be better understood with the help of Fig. 9.9. The desired period is  $RT_c$  with  $R$  being a rational number having an integer part of  $I$  bits and a decimal part of  $D$  bits, while  $T_c$  is the period of the constant clock which synchronizes the system.

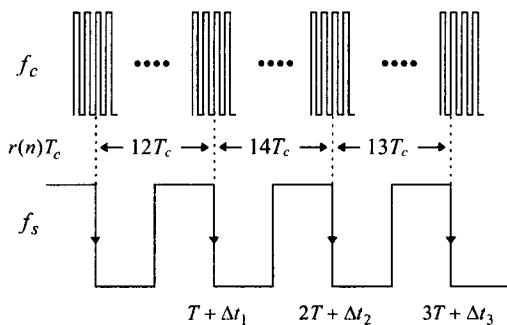


Fig. 9.9 Varying decimation of pulses.

It can be seen that the instantaneous period of the generated pulse train is not constant, but equal to  $r(n)T_c$ . In this case we may assume, as we did in the previous section, that the generated pulse sequence is nearly periodic with period  $T$  and jitter  $\Delta t_n$  in the falling edges. In a similar way as we did before, we may show that the period  $T$  is larger than  $T_c$  by the mean value of the words  $r(n)$ , i.e.  $R$  times. It may be written as follows:

$$T = \lim_{N \rightarrow \infty} \left( \frac{1}{N} \sum_{n=0}^N r(n)T_c \right) = R \cdot T_c \quad (9.27)$$

Also the uncertainty  $\Delta t_n$  will be

$$\Delta t_n = \sum_{i=0}^n r(i)T_c - nRT_c = \sum_{i=0}^n (r + q(i))T_c - nRT_c = T_c \sum_{i=0}^n q(i) \quad (9.28)$$

where  $q(n)$  is the noise at the  $\Delta\Sigma$  modulator output. Therefore the spectral density  $\Delta t(z)$  of the jitter noise will be

$$\Delta t(z) = T_c \frac{Q(z)}{1-z^{-1}} = T_c \frac{NTF(z)}{1-z^{-1}} E(z) \quad (9.29)$$

i.e. it is shaped by the NTF of the  $\Delta\Sigma$  modulator.

The generated pulse sequence is suitable to be used as the clock in discrete-time filters such as the switched-capacitor filters. It has the advantage that through this we may adjust digitally the cut-off frequency of low-pass or the center frequency of a band-pass discrete-time filter, since in these filters the characteristic frequencies are set by the sampling frequency. The problem created by the non uniform sampling of the analog signal, which is caused by the jitter of this clock, is minimized, if the filter is low-pass or band-pass. To show this, assume that the analog signal is  $x(t) = A \cos(\omega t)$ . Then

$$x(n) = A \cos(\omega(nT + \Delta t_n)) \quad (9.30)$$

Expansion in a Taylor series gives

$$x(n) = A \cos(\omega nT) - \omega \Delta t_n A \sin(\omega nT) - \frac{(\omega \Delta t_n)^2}{2} A \cos(\omega nT) + \dots \quad (9.31)$$

As long as  $\omega \Delta t_n$  is much smaller than unity, the noise due to jitter will be determined by the term  $\omega \Delta t_n A \sin(\omega nT)$ . Its power is concentrated mainly at frequencies far from the signal band. Therefore this noise will be attenuated by the filter function  $H(z)$ , provided that this filter is low-pass or band-pass.

### 9.5 $\Delta\Sigma$ Modulation in Analog-Input Digital Phase-Locked Loops for Frequency and Phase Demodulation

$\Delta\Sigma$  modulation can be used in a different way from that explained in Sec. 9.3.2, to obtain an analog-input digital phase-locked loop (ADPLL) which is useful for frequency and phase demodulation. The structure of an ADPLL is shown in block diagram form in Fig. 9.10, where the output  $y(n)$  corresponds to the instantaneous frequency of the input signal  $x(t)$ .

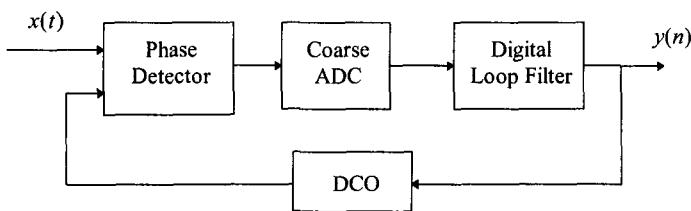


Fig. 9.10 Block diagram of an ADPLL.

In spite of the profound advantage of the ADPLL to achieve digital frequency demodulation, the use of the coarse ADC (low number of bits) introduces quantization noise. In this case it is possible to apply noise-shaping in order to improve the system performance. Since the principle of operation of the resulting system, shown in Fig. 9.11, is similar to that of a conventional  $\Delta\Sigma$  encoder of an analog signal to digital, this type of ADPLL is referred to in the literature as  $\Delta\Sigma$ PLL.

The structure of the  $\Delta\Sigma$ PLL is very similar to  $\Delta\Sigma$  modulator, since at its output we get samples  $f(n)$  of the instantaneous frequency  $f(t)$  of the input signal  $x(t)$ , as well as noise  $q(n)$ , the spectrum of which has been shaped at high frequencies and thus can be attenuated later by a low-pass filter. The operation of the digital controlled oscillator is similar to that described in previous sections. Both the DCO and the ADC are clocked by a constant clock. The signal flow diagram of the  $\Delta\Sigma$ PLL is given [17] in Fig. 9.12. It can be easily shown that

$$Y(z) = cF(z) + \left(1 - z^{-1}\right)^K E(z) \quad (9.32)$$

i.e. the system is a Kth-order  $\Delta\Sigma$  modulator. An IC implementing a  $\Delta\Sigma$ PLL for frequency demodulation of FM signals has been obtained [21].

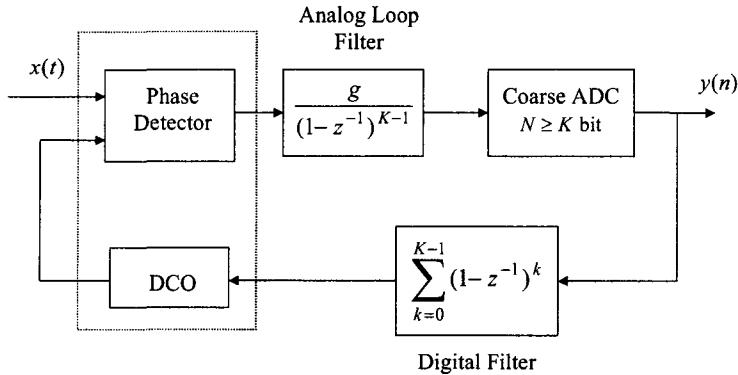


Fig. 9.11 Block diagram of a  $\Delta\Sigma$ PLL.

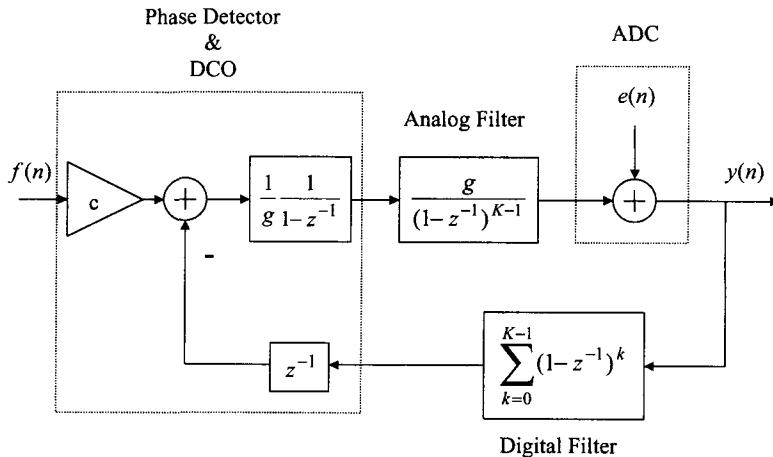


Fig. 9.12 Model for the analysis of the  $\Delta\Sigma$ PLL.

## 9.6 $\Delta\Sigma$ Modulation in Analog Oscillators

$\Delta\Sigma$  modulation can be used in the implementation of digitally programmable analog oscillators thus minimizing the analog as well as the digital circuitry. The simplified structure of such an analog oscillator is shown in Fig. 9.13.

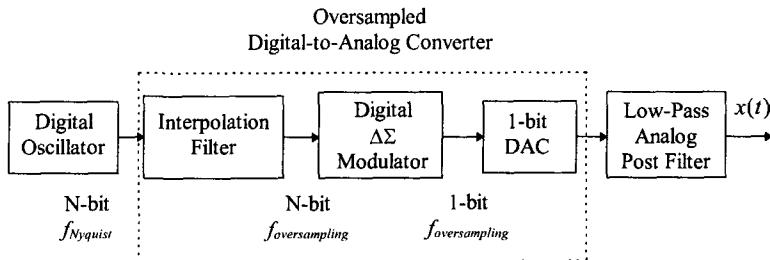


Fig. 9.13 Simple digitally programmable analog oscillator.

The digital oscillator can be based on a Lossless Discrete Integrator (LDI) biquad filter shown in Fig. 9.14

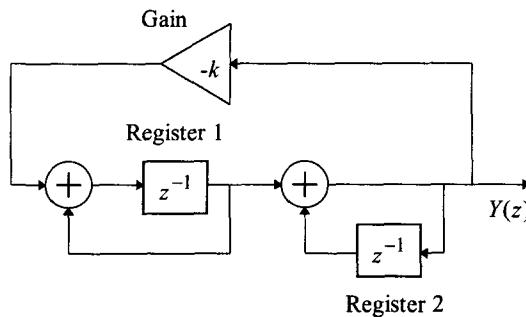


Fig. 9.14 Digital oscillator.

The output signal in Fig. 9.14 is determined by the following equation:

$$z^2 Y(z) + (k - 2)zY(z) + Y(z) = 0 \quad (9.33)$$

As long as  $0 < k < 4$  the output  $y(n)$  will be

$$y(n) = A \sin(2\pi f \cdot nT_s + \phi) \quad (9.34)$$

where the frequency  $f$  can be obtained from the roots of the characteristic equation (9.33)

$$z^2 + (k - 2)z + 1 = 0 \quad (9.35)$$

Thus it can be obtained that

$$2 \cos(2\pi f T_s) = 2 - k$$

or

$$f = \frac{1}{2\pi T_s} \cos^{-1}\left(\frac{2-k}{2}\right) \quad (9.36)$$

The amplitude  $A$  and the initial phase are determined by the initial settings  $x_1$  and  $x_2$  of registers 1 and 2.

It is left to the reader to show that

$$A = \frac{(1-k)x_1 - kx_2}{\sin(2\pi f T_s + \varphi)} \quad (9.37.a)$$

and

$$\varphi = \tan^{-1}\left(\frac{(x_1 + x_2)\sin(2\pi f T_s)}{x_1(1-k - \cos(2\pi f T_s)) - x_2(k + \cos(2\pi f T_s))}\right) \quad (9.37.b)$$

This system can be modified by the introduction of a  $\Delta\Sigma$  modulator in the Digital Oscillator Loop, which makes possible the elimination of the multiplier required for the implementation of the Digital Oscillator in Fig. 9.14, as well as the elimination of the interpolation filter in Fig. 9.13. The modified system is shown in Fig. 9.15 [22,23]. In this figure the multiplier has been replaced by a multiplexer, which chooses either the word  $k$  or the word  $-k$ . The scaling by the factor  $2^{-b}$  (simple shifting and not multiplication) prevents the overloading of the  $\Delta\Sigma$  modulator.

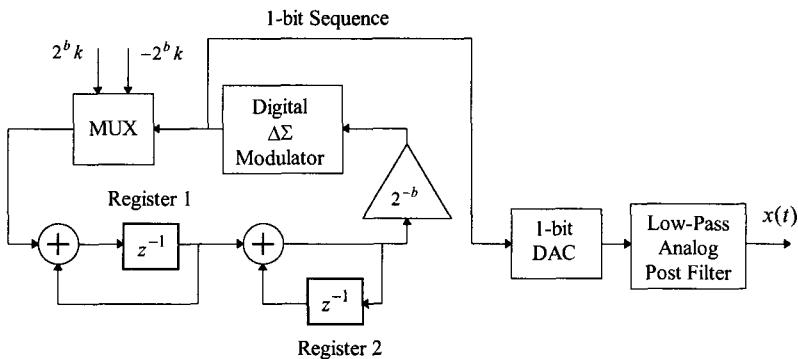


Fig. 9.15 Block diagram of a Digital Oscillator employing  $\Delta\Sigma$  modulation.

Various implementation and simulation results using various types of  $\Delta\Sigma$  modulators (high-order or band-pass) reveal the satisfactory performance of the method with respect to the purity of the produced sine wave [24]. These types of oscillators are useful for analog signal generation with emphasis on simple and compact implementation. They are suitable for on-chip analog circuit testing and built-in self-test of mixed signal integrated circuits [24,25].

## 9.7 Summary

In Chapter 8 we examined the application of  $\Delta\Sigma$  Modulation in Data Converters, ADC and DAC, which is of high importance in digital audio systems. Some additional most important applications of  $\Delta\Sigma$  modulation have been reviewed in this chapter among the many mentioned in Chapter 1. Thus we started with its application in Digital Radio in the form of Band-Pass  $\Delta\Sigma$  Modulators. Next, its application in frequency synthesis as well as in frequency and phase demodulation was reviewed. Also we saw its use in the implementation of digitally programmable clocks with reduced jitter, which are useful in SC circuit operation. Finally, its use in the implementation of analog signal generators with reduced circuit complexity, which are useful for the on-chip analog circuit testing, was examined. However, there are many more uses of  $\Delta\Sigma$  Modulators,

which are not well known, mainly because they may still be in the state of development. In future, no doubt, we should expect more applications of  $\Delta\Sigma$  modulation to appear in the literature and in practice.

## References

- [1] U. Razavi, *RF Microelectronics*, Prentice Hall PTR, Upper Saddle River, 1998.
- [2] S. A. Jantzi, K. W. Martin and A. S. Sedra, ‘Quadrature Bandpass  $\Delta\Sigma$  Modulation for Digital Radio’, *IEEE Journal of Solid-State Circuits*, Vol. 32, No. 12, pp. 1935-1950, December 1997.
- [3] J. Crols and S. Steyaert, ‘Low-IF Topologies for High-Performance Analog Front Ends of Fully Integrated Receivers’, *IEEE Trans. on Circuits and Sys.-II*, Vol. 45, pp. 269-282, March 1998.
- [4] L. Louis, J. Abcarius and G. W. Roberts, ‘An Eighth-Order Bandpass  $\Delta\Sigma$  Modulator for A/D Conversion in Digital Radio’, *IEEE Journal of Solid-State Circuits*, Vol. 34, No. 4, pp. 423-431, April 1999.
- [5] St. A. Jantzi, W. M. Snelgrove, P. F. Ferguson, ‘A Fourth-Order Bandpass Sigma-Delta Modulator’, *IEEE Journal of Solid-State Circuits*, Vol. 28, No. 3, pp. 282-291, March 1993.
- [6] A. M. Thurston, M. J. Hawksford, ‘Phase Correction Techniques for Bandpass Sigma Delta A-D Converters’, *The Institution of Electrical Engineers* 1994.
- [7] D. B. Ribner, ‘Multistage Bandpass Delta Sigma Modulators’, *IEEE Trans. on Circuits and Systems-II: Analog and Digital Signal Processing*, Vol. 41, No. 6, pp. 402-405, June 1994.
- [8] F. W. Singor, W. M. Snelgrove, ‘Switched-Capacitor Bandpass Delta-Sigma A/D Modulation at 10.7 MHz’, *IEEE Journal of Solid-State Circuits*, Vol. 30, No. 3, pp. 184-192, March 1995.
- [9] Bang-Sup Song, ‘A Fourth-Order Bandpass Sigma-Delta Modulator with Reduced Number of Op Amps’, *IEEE Journal of Solid-State Circuits*, Vol. 30, No. 12, pp. 1309-1315, December 1995.
- [10] S. Bazarjani, W. M. Snelgrove, ‘A 160MHz Fourth-Order Double-Sampled SC Bandpass Sigma-Delta Modulator’, *IEEE Transactions on Circuits and Systems II, Analog and Digital Signal Processing*, Vol. 45, No. 5, pp. 547-555 May 1998.
- [11] W. Gao, W. M. Snelgrove, ‘A 950MHz IF Second-Order Integrated LC Sigma-Delta Modulator’, *IEEE Journal of Solid-State Circuits*, Vol. 33, No. 5, pp. 723-732, May 1998.
- [12] S. Chuang, H. Liu, X. Yu, T. L. Sculley and R. H. Bamberger, ‘Design and Implementation of Bandpass Delta-Sigma Modulators Using Half-Delay Integrators’, *IEEE Transactions on Circuits and Systems II, Analog and Digital Signal Processing*, Vol. 45, No. 5, pp. 535-546, May 1998.
- [13] A. K. Ong and B. A. Wooley, ‘A Two-Path Bandpass  $\Sigma\Delta$  Modulator for Digital IF Extraction at 20 MHz’, *IEEE Journal of Solid-State Circuits*, Vol. 32, No. 12, pp. 1920-1934, December 1997.
- [14] O. Shoaei and W. M. Snelgrove, ‘Design and Implementation of a Tunable 40MHz - 70MHz GM-C Bandpass  $\Delta\Sigma$  Modulator’, *IEEE Transactions on Circuits and Systems II, Analog and Digital Signal Processing*, Vol. 44, No. 7, pp. 521-530, July 1997.

- [15] T. Paulus, S. S. Somayajula, T. A. Miller, B. Trotter, K. Choi and D. A. Kerth, 'A CMOS IF Tranceiver with Reduced Analog Complexity', *IEEE Journal of Solid-State Circuits*, Vol. 33, No. 12, pp. 2154-2159, December 1998.
- [16] T. A. D. Riley, M. A. Copeland, and T. A. Kwasniewski, 'Delta-Sigma Modulation in Fractional-n Frequency Synthesis', *IEEE Journal of Solid-State Circuits*, Vol. 28, No. 5, pp. 553-559, May 1993.
- [17] I. Galton, 'Analog-Input Digital Phase-Locked Loops for Precise Frequency and Phase Demodulation', *IEEE Trans. on Circuits and Systems-II: Analog and Digital Signal Processing*, Vol. 42, No. 10, pp. 621-630, October 1995.
- [18] W. F. Efan, *Phase-Lock Basics*, John Wiley & Sons, 1998.
- [19] W. F. Efan, *Frequency Synthesis Phase-Lock*, John Wiley & Sons, 2000.
- [20] P. J. Hurst and B. C. Rothenberg, 'A Programmable Clock Generator that Uses Noise Shaping and Its Application in Switched-Capacitor Filters', *IEEE Journal of Solid-State Circuits*, Vol. 30, No. 4, pp. 403-411, April 1995.
- [21] I. Galton, W. Huff, P. Carbone and E. Siragusa, 'A Delta-Sigma PLL for 14-b, 50kSample/s Frequency-to-Digital Conversion of a 10 MHz FM Signal', *IEEE Journal of Solid-State Circuits*, Vol. 33, No. 12, pp. 2042-2053, December 1998.
- [22] A. K. Lu, G. W. Roberts, D. A. Johns, 'A High-Quality Analog Oscillator Using Oversampling D/A Conversion Techniques', *IEEE Trans. on Circuits and Systems-II: Analog and Digital Signal Processing*, Vol. 41, No. 7, pp. 437-443, July 1994.
- [23] C. M. Zierhofer, 'A Multiplier-Free Digital Sinusoid Generator Based on Sigma-Delta Modulation', *IEEE Trans. on Circuits and Systems-II: Analog and Digital Signal Processing*, Vol. 43, No. 5, pp. 387-396, May 1996.
- [24] X. Haurie and G. W. Roberts, 'Arbitrary-Precision Signal Generation for Mixed-Signal Built-In-Self-Test', *IEEE Trans. on Circuits and Systems-II: Analog and Digital Signal Processing*, Vol. 45, No. 11, pp. 1425-1432, November 1998.
- [25] B. Dufort and G. W. Roberts, 'On-Chip Analog Signal Generation for Mixed-Signal Built-in Self-Test', *IEEE Journal of Solid-State Circuits*, Vol. 34, No. 3, pp. 318-330, March 1999.

This page is intentionally left blank

# Index

- acquisition mode, 30
- adaptive
  - bit-flipping, 193
  - dithering, 192
  - $\Delta\Sigma$  modulators, 69
- ADPLL, 232
- aliasing, 17
- analog-to-digital conversion, 12
- antialiasing filter, 14, 17
- applications, 6, 214
- autozeroing, 148
- band-limited, 16
- band-pass  $\Delta\Sigma$  modulators, 70
- band-pass signals, 25
- basic blocks of a  $\Delta\Sigma$  modulator, 100
- Bazarjani, 71
- bit-flipping, 188
- cascade connection of integrators, 102
- chaotic  $\Delta\Sigma$  modulators, 85, 185
- characteristics of NTFs, 83
- chebyshev polynomials, 91
- chopper stabilization (CHS), 148
- circuit design of  $\Delta\Sigma$  modulators
  - first-order, 113
  - second-order, 114
  - third-order, 119
  - fourth-order, 124
- circuit limitations, 139
- clipping the integrators, 171
- comb filters, 205
- comparison
  - adaptive techniques for the suppression of tones, 193
  - fixed techniques for the suppression of tones, 191
  - SC and CT circuits, 158
  - stabilization methods, 175
- continuous-time loop filter implementation, 108
- conversion
  - bandwidth, 217
  - fractional rate, 207
  - multistage rate, 203
  - rate, 196, 197
- converters
  - analog-to-digital, 12, 208
  - digital-to-analog, 12, 211
- correlated double sampling (CDS), 148
- DAC errors, 159
  - multibit, 161
  - single-bit, 160
- DAC implementations, 162
- DAC output levels, 108, 161
- DAC output pulses, 106
- decimation, 197
- decimation filter, 201

- Delta modulator, 29
  - exponential, 30, 34, 48
  - implementation of, 30
  - linear, 30, 48
  - performance, 33
- $\Delta\Sigma$  modulator, 36
  - adaptive, 69
  - architectures, 41
  - assessment, 95
  - design, 4
  - dynamic range, 46, 50, 56, 63
  - first-order, 42
  - full-scale, 116
  - high-order, 51
  - hybrid, 68
  - implementation, 100
  - low-power, low-voltage, 131
  - modeling, 78
  - multi-bit, 64
  - second-order, 49
  - single-bit, single-stage, 77
  - stability, 58, 85
- $\Delta\Sigma$  modulation for
  - ADC, 208
  - DAC, 211
  - Frequency and Phase demodulation, 232
- $\Delta\Sigma$  modulation in
  - analog oscillators, 234
  - clock generators, 229
  - Digital Radio, 214
  - Frequency Synthesis, 224
- $\Delta\Sigma$ PLL, 232
- demodulation 1
  - filter, 31, 32
  - process, 31
- diagrams of SNR and PSD, 46
- differential integrators, 103
- differential quantizer, 31
- digital-to-analog converter, 12
- dithering, 184
- DPCM, 1, 29
- dynamic range, 24
- dynamic range of  $\Delta\Sigma$  modulators
  - first-order, 46
  - second-order, 50
  - high-order, 56
  - multi-stage, 63
- effects of
  - flicker noise, 147
  - jitter, 152
  - rise and fall times, 156
  - thermal noise, 144, 147
- experimental testing of  $\Delta\Sigma$  modulators, 127
- finite
  - bandwidth, 151
  - opamp gain, 149
  - slew-rate, 151
- flicker noise effects, 147
- folding frequency, 17
- frequency divider, 226
- frequency synthesis, 221
- full-scale, 20
- Gao, 71
- GSM specifications, 226
- heterodyne receiver, 216
- high-order  $\Delta\Sigma$  modulators, 51
- hybrid  $\Delta\Sigma$  modulators, 68
- ideal sampler, 16
- idling pattern, 30
- images, spectrum, 17
- implementation of  $\Delta\Sigma$  modulators, 100
- instability of  $\Delta\Sigma$  modulators, 58, 85
- integrators
  - CT, 102
  - fully differential, 103
  - SC, 102
- interpolation, 198
- interpolation filter, 201
- Inverse Chebychev polynomials, 93
- jitter effect in, 152
  - CT active RC circuits, 154
  - SC circuits, 153

- Leslie, 68
- limitation of practical  $\Delta\Sigma$  modulators, 139
- limit cycles, 85
- linear interpolation, 199
- linear model, 22, 79
- local ADC, 101, 106
- local DAC, 101, 106
- loop-filter, 100, 102
- low-power, low voltage  $\Delta\Sigma$  modulators, 131
- methods for
  - resolving DAC non linearities, 163
  - stabilization, 170
- midread quantizer, 21
- midriser quantizer, 21
- mismatching
  - in DACs, 162
  - in multistage  $\Delta\Sigma$  modulators, 142
- mode
  - acquisition, 30
  - tracking, 30
- model
  - linear, 20, 79
  - quasi-linear, 81
- modeling of  $\Delta\Sigma$  modulators, 78
- modulators
  - adaptive  $\Delta\Sigma$ , 69
  - band-pass  $\Delta\Sigma$ , 70
  - delta, 29
  - delta-sigma, 36
  - hybrid  $\Delta\Sigma$ , 68
  - multi-bit  $\Delta\Sigma$ , 64
  - multi-stage  $\Delta\Sigma$ , 59
  - predictive, 30
    - single-bit single-stage, 77
- modulus counter, 225
- multi-bit  $\Delta\Sigma$  modulators, 64
- multi-stage  $\Delta\Sigma$  modulators, 59
- noise
  - power, 23
  - quantization, 20
- shaping concept, 36
- sources in  $\Delta\Sigma$  modulators, 140
- transfer function (NTF), 78
- NTF characteristics, 83
- NTF determination, 89
- Nyquist
  - Converter, 13, 27
  - frequency, 17
  - rate, 13
  - sampler, 27
- opamp non-idealities, 147
  - dc offset, 147
  - finite bandwidth, 151
  - finite gain, 149
  - finite slew rate, 151
- optimum polynomial
  - coefficients, 91, 92
  - roots, 91, 92
- optimization of position of
  - poles, 93
  - zeros, 90
- overloading, 20
- oversampling
  - principles, 27
  - rate, 13, 27
  - ratio, 27
- PCM, 1
- Phase-Locked loop (PLL), 221
- polynomials
  - Inverse Chebychev, 92
  - Optimal, 92
- Practical circuit limitations, 139
- predictive modulator, 30
- probability density function (PDF), 22, 79
- power spectral density (PSD), 23
- pulse code modulation (PCM), 1
- quantization
  - error, 20
  - levels, 20
  - noise, 22
  - step-size, 20

- quantizer
  - midread, 21
  - midriser, 21
- quasi-linear model, 81
- rate
  - conversion, 196, 197
  - oversampling, 27
  - sampling, 15
- reference voltage, 160
- resetting the integrators, 171
- resonators, 104
- return-to-zero waveform (RZ), 107, 157
- rise and fall times, 156
- sample-and-hold, 18
- sampler circuit, 14
- sampling
  - frequency, 15
  - of band-pass signals, 25
  - rate, 14, 15
  - theorem, 15
  - uniform, 15
- signal-to-noise ratio (SNR), 23
- signal transfer function (STF), 78
- Singh, 68
- Snelgrove, 71
- spectrum images, 17
- spectrum of  $\Delta\Sigma$  sequence, 127
- stability criteria, 87
- stabilization methods, 170
  - activation of local feedback loops, 172
  - clipping the integrators, 171
  - reducing the order of the loop filter, 173
  - resetting the integrators, 171
- suppression of tones, 184
- switched-capacitor integrators
  - inverting, 102
  - non-inverting, 102
- testing  $\Delta\Sigma$  modulators
  - experimentally, 127
  - by simulation, 127
- thermal noise effects, 144
- tracking mode, 30
- triple-first-order cascade (TFOC), 59
- tones in  $\Delta\Sigma$  modulation, 182
- tones suppression, 184
- uniform
  - quantization, 20
  - quantizer, 20
  - sampling, 15
- unit-element DAC, 162
- variable gain method, 168
- voltage-controlled oscillator (VCO), 221