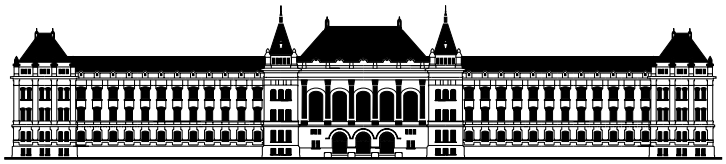
Ph.D. Thesis

János Márkus



Budapest University of Technology and Economics
Department of Measurement and Information Systems

Higher-order Incremental Delta-Sigma Analog-to-Digital Converters

by

János Márkus

M.S. (Budapest University of Technology and Economics) 1999

A thesis
submitted to the Department of Measurement and Information Systems
and the Doctoral Committee of the
Budapest University of Technology and Economics
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy (Ph.D.)
in
Electrical Engineering

Advisor:
István Kollár
Dr. Acad.

2005

Budapest University of Technology and Economics
Department of Measurement and Information Systems
H-1117 Budapest, XI. Magyar Tudósok körútja 2.
Building I., Level E, Room E330.
Tel: +36 1 463 3587, Fax: +36 1 463 4112, Email: markus@mit.bme.hu

Contents

Nyilatkozat (Declaration of Authorship)

Alulírott Márkus János kijelentem, hogy ezt a doktori értekezést magam készítettem és csak a megadott forrásokat használtam fel. Minden olyan részt, amelyet szó szerint, vagy azonos tartalomban, de átfogalmazva más forrásból átvettem, egyértelműen, a forrás megadásával megjelöltem.

A dolgozat bírálatai és a védésről készült jegyzőkönyv a későbbiekben a Budapesti Műszaki és Gazdaságtudományi Egyetem Villamosmérnöki és Informatikai Karának dékáni hivatalában lesz elérhető.

Budapest, 2005. március 9.

Márkus János
jelölt

Abstract

János Márkus
„Higher-order Incremental Delta-Sigma Analog-to-Digital Converters"
PhD thesis

Analog-to-digital conversion, which takes continuous-time, continuous amplitude signals (voltage, temperature, sound, etc.) and converts them into a series of numbers to be used for digital signal processing, is becoming the key element of the scholarly and industrial applications of measurement and data acquisition, and A/D converters are surrounding (though invisible in most cases) our everyday life.

In instrumentation and measurement, there is a growing demand for A/D converters with low or medium bandwidth, but with high absolute accuracy (e.g., sensors, dc-measurement applications). High linearity and small offset are also among the requirements, as well as small power-consumption and low sensitivity to environmental noise (such as the periodic noise coupled from the mains or digital switching noise). One solution to the problem is the incremental (or charge-balancing) $\Delta\Sigma$ converter, which is basically a first-order $\Delta\Sigma$ A/D converter, operated in transient mode. The converter represents a hybrid between the classical dual-slope converter and the $\Delta\Sigma$ one.

This dissertation extends the operation of the incremental converter to higher-order $\Delta\Sigma$ loops. It discusses the basic operation of such a converter, the theoretically achievable resolution, filter design methods for the digital filter following the $\Delta\Sigma$ modulator, and the structure's sensitivity to analog circuit elements imperfections. The introduced general architecture is flexible, thus it is capable to optimize the trade-off between circuit complexity and conversion accuracy.

Design examples and optimization techniques are proposed to help designers selecting the best configuration for a given application. The thesis also compares the results with those found in the literature.

The theoretical results are verified by simulations and also by measurements made on an integrated circuit.

Kivonat (Abstract in Hungarian)

Márkus János
„Többedrendű, számláló típusú Delta-Szigma analóg-digitális átalakítók”
PhD értekezés

Az analóg-digitális (A/D) átalakítás, amelynek során egy analóg jelből (feszültség, hőmérséklet, hang stb.) számítógépes feldolgozásra alkalmas számsorozatot készítünk, egyre inkább kulcsfontosságú szerepet játszik a méréstechnika és adatgyűjtés ipari ill. tudományos alkalmazásaiban, ugyanakkor A/D átalakítók vesznek körül bennünket – bár többnyire észrevétlenül – a mindennapi életben is.

A műszer- és méréstechnikában sokszor van szükség olyan A/D átalakítóra, amelynek sávszélessége közepes vagy kicsi, viszont abszolút pontossága igen nagy (pl. szenzorok, DC-mérő alkalmazások). Sokszor követelmény ilyen alkalmazásoknál a kis linearitási hiba és az elhanyagolható ofszet is, továbbá a kis fogyasztás ill. zajérzéketlenség. Egy megoldás erre a feladatra a számláló típusú (incremental) $\Delta\Sigma$ átalakító, amely az elsőrendű $\Delta\Sigma$ A/D átalakító tranziens működéséből származtatható. Az átalakító egyfajta hibridet képez a klasszikus dual-slope és a $\Delta\Sigma$ átalakító között.

A dolgozat az elsőrendű számláló típusú átalakító működését terjeszti ki magasabb rendű $\Delta\Sigma$ modulátorokra. Tárgyalja a működés alapelveit, az elvileg elérhető felbontást, a modulátort követő digitális szűrő tervezési módszereit, illetve az áramköri elemek pontatlanságára való érzékenységet. Mivel a javasolt általános struktúra többféle modulátortípust magába foglal, így könnyen megtalálható a legjobb kompromisszum az áramköri bonyolultság illetve az átalakító pontossága, gyorsasága között.

Tervezési példák, valamint optimalizációs technikák segítik az ilyen átalakító tervezőjét a legjobb konfiguráció megtalálásában egy adott alkalmazás esetén. A dolgozat összehasonlító elemzést is végez az irodalomban fellelhető módszerekkel.

Az elért elméleti eredményeket szimulációk, valamint egy elkészült integrált áramkör mérési eredményei támasztják alá.

Glossary of Symbols

Symbols

a	Coefficient of a $\Delta\Sigma$ modulator.
b	Scaling coefficient of a $\Delta\Sigma$ modulator at the input.
B	Bandwidth of an analog signal.
c_i	Scaling coefficient of a $\Delta\Sigma$ modulator.
C	Capacitor.
$d_i, D(z)$	Digital output of the modulator.
$D(z)$	Denominator of the NTF of a stabilized $\Delta\Sigma$ modulator.
D_{out}	Digital output of the converter.
$\epsilon[k]$	Discrete-time step-function.
$\varepsilon[k], \varepsilon(z)$	Quantization error of the A/D converter *within* the $\Delta\Sigma$-loop in the sample- or z-domain. See also $q[k], q(z)$.
f_{clk}	Clock rate of an SC circuit in Hertz.
f_N	The minimum sampling rate required for the reversible conversion of an analog signal with bandwidth B, $f_N = 2B$. It is usually referred as Nyquist-rate.
f_s	Sampling frequency.
k	Boltzmann's constant, used in noise analysis. $k = 1.38{\cdot}10^{-23}\text{JK}^{-1}$. k is also used as general variable in other contexts.
l	Number of levels of the internal quantizer and feedback DAC in a $\Delta\Sigma$ modulator.
L_a	Order of the analog $\Delta\Sigma$ modulator.
L_d	Order of the digital filter following the $\Delta\Sigma$ modulator.
m	Mean value of a stochastic signal. Used also as the number of significant samples of an IIR-filter's impulse-response.
N	Number of cycles through an incremental converter operates.
n_{bit}	Number of bits of an incremental converter.
N_i	Decimation ratio of an ith-order sinc-filter following a pure differential $\Delta\Sigma$ modulator.
$N_{i,p}$	Decimation ratio of an ith-order sinc-filter following a $\Delta\Sigma$ modulator with stabilizing poles.
OSR	Oversampling ratio, $OSR = f_s/f_N = f_s/(2B)$.

Φ_i	The ith clock phase of a switched-capacitor circuit.
$q[k], q(z)$	Quantization error of the *whole* incremental $\Delta\Sigma$ A/D converter in the sample- or the z-domain. See also $\varepsilon[k], \varepsilon(z)$.
σ	Standard deviation of a stochastic signal (also the rms-value for signals with zero mean).
σ^2	Variance of a stochastic signal.
S_i	Switch no. i
T	Temperature in Kelvin, used for noise analysis.
T	Time.
T_{clk}	Length of one period of a clock signal in seconds.
u,$u[k]$,$U(z)$	Relative (normalized) input signal of the A/D converter, $u = V_{\text{in}}/V_{\text{ref}}$. u stands for dc-signal, while $u[k]$ and $U(z)$ represents general input signals in the sample- and z-domain, respectively.
U_{max}	Maximum relative (normalized) input signal of the A/D converter, $U_{\text{max}} = V_{\text{max}}/V_{\text{ref}}$.
V_d	Dither signal in volts.
V_{in}	Input signal of the A/D converter in volts.
V_{int}	Output of the integrator of the first-order incremental converter in volts.
V_{lsb}	Equivalent voltage of the LSB, $V_{\text{lsb}} = 2V_{\text{max}}/2^{n_{\text{bit}}}$.
V_{max}	Maximum input signal of the A/D converter in volts.
V_{ref}	Reference signal of the A/D converter in volts.
$w[k]$	Weighting function or impulse response.
$w_d[k]$	Weighting function or impulse response of the IIR-filter $1/D(z)$.

Abbreviations

A/D	Analog-to-Digital
ac	Alternating current (in general: non-constant part of a signal)
ADC	Analog-to-Digital Converter
CIC	Cascaded-Integrators-Comb filter, efficient realization of the digital sinc-filter.
CIFF	Cascaded-Integrators, Feed-forward $\Delta\Sigma$ architecture
CMOS	Complementary Metal-Oxid-Semicondutor, today's most commonly used implementation technology for digital integrated circuits.
CoI	Cascade-of-Integrators digital filter.
D/A	Digital-to-Analog
DAC	Digital-to-Analog Converter
dB	deciBell, logarithmic power-ratio. Used mainly for *SNR* in this thesis.
dc	Direct current (in general: constant part of a signal)

$\Delta\Sigma$	Delta-Sigma modulation. This technique is often referred as $\Sigma\Delta$ (Sigma-Delta) modulation. In this thesis, according to [Norsworthy et al., 1997], the term $\Delta\Sigma$ is used.
ENOB	Effective or equivalent number of bits
FIR	Finite impulse response
HF	High frequency
IC	Integrated circuit
IIR	Infinite impulse response
LF	Low frequency
lhs	Left hand side (of an expression)
LSB	Least significant bit
MASH	Multi-stage noise-shaping or cascaded $\Delta\Sigma$ converter architecture
MSB	Most significant bit
N/A	Used as either Not Available or Not Applicable
NTF	Noise transfer function, transfer function from the internal quantizer to the output of a $\Delta\Sigma$ modulator.
Nyquist-rate	See f_N in symbols
OSR	Oversampling ratio, $OSR = f_s/f_N = f_s/(2B)$.
RC-constant	Time constant of an exponential settling determined by a resistor (R) and capacitor (C). $\tau = RC$.
rhs	Right hand side (of an expression)
rms	Root mean square
SC	Switched-capacitor
$\Sigma\Delta$	See $\Delta\Sigma$
sinc	Sinc-function, $\mathrm{sinc}(x) = \sin(\pi x)/\pi x$.
sinc_d	Discrete-time sinc-function, $\mathrm{sinc}_d(x) = \mathrm{sinc}(Nx)/\mathrm{sinc}(x) = \sin(\pi N x)/(N \sin(\pi x))$, where N is the length of the discrete-time rectangle window.
SNR	Signal-to-noise ratio. In this thesis this term is usually used for Signal-to-quantization-noise ratio.
SoC	System-on-a-Chip, technology to integrate all processing units, memories and I/O circuits onto the same chip
STF	Signal transfer function, transfer function from the input to the output of a $\Delta\Sigma$ modulator.
UGB	Unity-gain bandwidth of an operational amplifier (op-amp).

Preface

This thesis is the collection of the main results achieved in the field of $\Delta\Sigma$ analog-to-digital conversion during my five years long research period conducted at the Budapest University of Technology and Economics (BUTE), Budapest, Hungary and at Oregon State University (OSU), Corvallis, Oregon, USA.

I have started my PhD studies in 1999, right after graduation, at the Department of Measurement and Information Systems (BUTE) under the supervision of Prof. István Kollár. During the first two years I have gained a lot of theoretical and practical knowledge, especially in Quantization Theory, Digital Signal Processing, Matrix Theory and test methods of Analog-to-Digital Converters. From March, 2001 I have spent 14 months as a visiting scholar at Oregon State University, where I was working under the supervision of Prof. Gábor C. Temes, from whom I learnt many aspects of Delta-Sigma modulation and analog circuit design theory and practice.

From Sept. 2002, after I returned to Hungary, I have finalized and prepared for publication the results achieved in the topic of delta-sigma analog-to-digital converters in Oregon. Due to the courses I had to finish as PhD student and various teaching and research activities, I finally started to write this thesis in summer, 2004 and finished it in its first form by the end of the same year.

I would like to thank the colleagues at both the mixed-signal research group at OSU and the Department of Measurement and Information Systems at BUTE for the inspiring environment and the useful discussions. First of all I would like to thank István Kollár, my advisor at BUTE for his help, ideas, feedback and encouragements. I would also like to thank Prof. Gábor C. Temes for my visit to Oregon, his continuous help, critical remarks and his emphasis on writing good technical papers. I thank José Silva, Un-ku Moon and many colleagues at BUTE (especially László Balogh, Balázs Bank, Károly Molnár, József Németh, László Sujbert, Zoltán Szabó, Balázs Vödrös) for their help and the useful discussions during coffee and lunch breaks. Many thanks go to Microchip Technology Inc. and the incremental project team for the circuit-level design and implementation of the prototype chip.

This thesis has been supported by the Faculty of Electrical Engineering and Informatics at BUTE, the NSF US-Hungary grant at OSU, the NSF CDADIC (Center for Design of Analog-Digital Integrated Circuits) project, the Panda Audio Ltd, the László Schnell Instrumentation and Measurement Foundation, the Siemens AG, the Lawrence Livermoore National Laboratory and by the Department of Measurement and Information Systems. I would like to thank for their financial support.

Special thanks to my parents for their constant love and support. I would like to thank my wife, Johanna and child, Barnabás for their patience and understanding and for all the happiness they bring into my life. Finally, I would like to thank God, my heavenly father for the given talents and possibilities in my life, which made it possible to finish this PhD thesis.

Chapter 1

Introduction

In today's engineering technology almost every problem is solved using digital hardware. Digital hardware is more economical, less sensitive or even capable to adapt to the environmental changes and noise, easier to reconfigure or reuse, and in general it is more robust than its analog equivalents. It is true for scholarly and industrial applications (data acquisition, measurement, control loops, etc), as well as in commercial units (e.g., digital thermometer, compact disk, appliances, fuel injection control in vehicles, digital radio receiver, etc.). In order to work properly, most of these applications communicate with the real world through sensors and actuators. As the real world has analog (continuous-time, continuous amplitude) signals and digital hardware can only deal with numbers at a given clock-rate (discrete-time, discrete amplitude signals), every sensor must be accomplished with an analog-to-digital converter (A/D converter or ADC) and every actuator is driven by a digital-to-analog converter (D/A converter or DAC), which performs the required conversion between the analog and digital signals. This thesis focuses on A/D converter design methods.

Integrated A/D converter design started at the same time the first digital processing units become available in the second half of the 20th century (one of the first fully integrated converter was introduced in 1978 by [Hamadé, 1978]). Since then, numerous architecture have been developed, which can be classified many ways. One possible way is based on the ratio of the input signal bandwidth (B) and the converter's conversion rate, usually referred as sampling rate (f_s). It is well known from the Nyquist-theory (see e.g., [Oppenheim and Schafer, 1975]), that an analog signal with bandwidth B can be perfectly reconstructed from its sampled equivalent, if the sampling rate f_s is greater than (or equal to) twice the bandwidth of the signal, i.e., $f_s \geq 2B$. Based on the relationship of f_s and B, converters may be divided into two categories: Nyquist-rate converters ($f_s/(2B) = 1$ or only slightly larger) and oversampling converters ($f_s/(2B) \gg 1$). Typically, Nyquist-rate converters have one-to-one relationship between the instantaneous input signal and a single output value (sample-by-sample conversion). In the other case, oversampling converters operate at much higher rate than twice the signal bandwidth (Nyquist-rate), and the final output sequence is achieved by appropriate digital filtering and decimation. In this case one cannot find sample-by-sample relationships between the analog and digital data, only the waveform and its spectral properties are preserved during conversion.

Table 1.1: A/D converter requirements of different applications

Application	Requirements		
	Resolution	Bandwidth	Power-Consumption
Microcontrollers	Low-Med	Low-Med	Low-Med
LF Measurement	High	Low-Med	Low-Med
Sensor(-arrays)	High	Low-Med	Low
Audio	High	Med	N/A or Low
Control	Med-High	Low-Med	N/A
Video	Med-High	High	N/A
HF, microwave	Med	High	N/A
Telecommunication	Med	High	N/A or Low

Another way of classification is based on the bandwidth, resolution and power-consumption requirements of different applications. The most typical applications with requirements regarding to the A/D converters is listed in Tab. 1.1. These requirements are usually contradicted by each other: high resolution and high bandwidth indicates more complex hardware, which should have low power- and area-consumption (especially for portable, battery-operated equipment), and should have great tolerance on environmental effects (noise, temperature, etc.) at the same time. In addition, today's trend in system design is that the analog and mixed-signal interfaces are integrated into the same integrated circuit (IC) as the digital signal processing units (System-on-a-Chip, SoC design). This gives two serious limitations on high-resolution classical Nyquist-rate A/D converter design: first, in today's widely used low-voltage CMOS digital circuit implementation technology it is not possible to manufacture high-precision analog elements (resistors, capacitors, etc.) on which classical Nyquist-rate converters relies so much. Second, with such an integrated environment, designers have to deal with the switching-noise interference originating from the high-speed clock signal of the digital circuits. In general, as matching of analog elements cannot be made better than 0.1% (which indicates a signal to mismatch error ratio of 1000, equivalent of about 10 bit resolution), classical Nyquist-rate converters with resolution greater than 10 bits can be manufactured either with individual (and thus expensive) laser wafer trimming or has to be designed with sophisticated on-line or off-line self-calibration methods.

To overcome these problems, A/D converters based on Delta-Sigma ($\Delta\Sigma$) or Sigma-Delta ($\Sigma\Delta$) modulation can be a good candidate for high-resolution conversion in an integrated environment, especially if it is realized using switched-capacitor (SC) circuits. Switched-capacitor circuits can be modeled as discrete-time, continuous amplitude systems, and as the information in the circuit is stored in charges delivered in a given time-interval (T_{clk}) rather than voltage or current, it is less sensitive to pulse-like switching noise coupled from the digital part of the circuit. Another useful property of the SC circuits is that they rely only on capacitive matching which can be made as low as 0.1% with careful layout. This matching is about three orders better than the tolerance of the time constant of classical RC-circuits integrated in CMOS environment [Johns and Martin, 1997,

Table 1.2: Classification of different A/D converter architectures. Bold typeset indicates the architecture discussed in this thesis.

Application	Architecture	
	Nyquist-rate	Oversampling
Microcontrollers	Successive apprx., Algorithmic/Cyclic	N/A
dc, LF Measurement, Biomedical app.	Dual-slope, Voltage-to-frequency	**Incremental ΔΣ**
Sensor(-arrays)	Dual-slope	**Incremental ΔΣ**
Audio	Successive apprx.	simple oversampling, ΔΣ
Control	Successive apprx., Algorithmic/Cyclic	N/A
Video	Flash, Pipelined	low-*OSR* ΔΣ
HF, microwave	Flash, Pipelined	Bandpass ΔΣ
Telecommunication	Flash, Successive apprx.	ΔΣ

Chap. 10].

The advantage of using ΔΣ modulation (first introduced by [Inose et al., 1962]) instead of classical Nyquist-rate conversion is that ΔΣ modulator structures do not rely on precise analog elements, but they sample the incoming signal at a much higher rate than the bandwidth of the incoming signal (*oversampling*), and shape the quantization error of the low-resolution (often one-bit) quantizer by means of analog filtering (*noise-shaping*) [Norsworthy et al., 1997, Sec. 1.2], achieving high signal-to-noise ratio (*SNR*) in the signal band. The architecture is also capable to modulate most of the analog imperfection errors out of the band of interest. The oversampled signal is converted back to Nyquist-rate by means of digital low-pass filtering and resampling (*decimation*) [Norsworthy et al., 1997, Sec. 1.3].

This thesis deals with a special ΔΣ modulator topology with optimized circuit complexity and conversion efficiency for dc measuring applications. The proposed structure is called higher-order incremental ΔΣ converters (may also be referred as charge-balancing ΔΣ converter) introduced in the next subsection.

To give an insight into the different applications and different converters used today, and to identify the application area of the proposed architecture, Tab. 1.2 shows different A/D architectures used for different applications (ΔΣ converter structures discussed in this thesis are typeset in boldface), while Fig. 1.1 shows the targeted resolution and bandwidth requirements of the discussed architecture among typical A/D converters (the group of ΔΣ converters discussed in this thesis are in the gray ellipse).

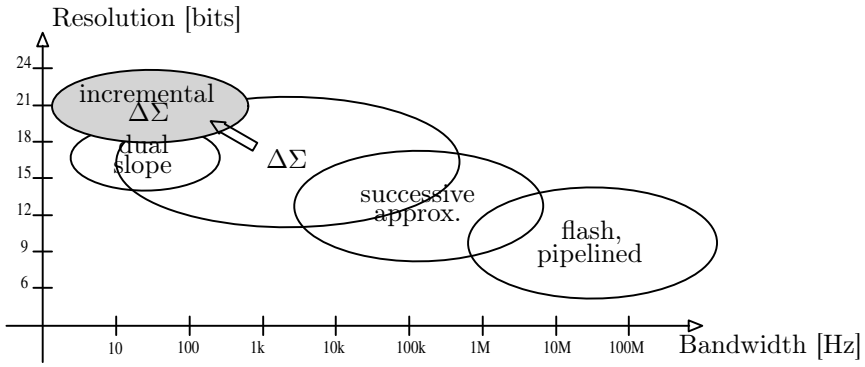


Figure 1.1: Applications of A/D converter structures for typical resolution and bandwidth requirements. Gray area represents the architecture discussed in this thesis.

1.1 Analog-to-digital Conversion for Measurement Applications

In instrumentation, measurement and sensor/transducer applications, often A/D converters with very high dynamic range are required. Such a typical example is a photodiode, which may produce signal currents between the 1pA and 1μA range, spreading about 6 decades in dynamic range. Similar dynamic range may be required in seizmic measurements. In addition to the high resolution demand ($n_{\text{bit}} \geq 20$), these converters also requires high *absolute* accuracy, including high linearity and negligible offset. Moreover, especially in battery-powered sensor and on-the-field measurement applications, power consumption, thus IC area-consumption must also be kept as low as possible. The only property which is easier to handle is bandwidth, since most of these application operates with dc-signals or low-frequency signals (up to a few kHz).

Among Nyquist-rate ADCs, dual-slope and voltage-to-frequency converters have dominated dc measurement applications for many years. However, as the need to integrate analog circuits into SoC (System-on-a-Chip) environment increased, these converters could not be used, as they usually rely on precise, large elements (such as integrating capacitors), and their technology cannot be easily integrated into low-voltage CMOS environment. In addition, these architectures cannot tolerate high-frequency switching noise originating from the digital circuitry. As RC constant mismatch in CMOS technology may have an implementation error of 20%(!), and in capacitor-ratio a maximum of 0.1% mismatch can be achieved with very careful layout [Johns and Martin, 1997, Chap. 10], the realization of high-resolution Nyquist-rate converters becomes very expensive when resolution exceeds 16 bits, either due to the application of individual laser-wafer-trimming, or due to the higher power- and area-consumption, which originates from the required sophisticated off-line or on-line (self-)calibration methods.

For high-resolution, high dynamic range conversion, $\Delta\Sigma$ A/D converter may be a good candidate. However, classical $\Delta\Sigma$ converters, used mainly in telecommunication and consumer electronics applications, are characterized by their signal processing parameters, such as dynamic range and signal-to-noise ratio (*SNR*), as in these applications usually a

running waveform needs to be digitized continuously, and mainly the spectral behavior of the signal is important. Moreover, these converters are mainly dedicated to applications which can tolerate offset and gain errors. On the contrary, in sensor applications the goal is to digitize individual samples or the average value of a noisy dc signal, and must exhibit an excellent sample-by-sample conversion performance (with very low linearity, offset and gain error).

A third candidate for high-precision dc measuring application is the incremental converter and its various extensions. The first first-order CMOS incremental (or charge-balancing) converter has been introduced in [Robert et al., 1987], achieving 16-bit performance in a low-voltage environment. The converter is based on the $\Delta\Sigma$ architecture, however, it operates only up to N clock cycles while it performs one conversion. Its operation represents a hybrid between the classical Nyquist-rate dual-slope converter and an oversampling $\Delta\Sigma$ one. Detailed analysis of its operation, as well as various extensions of the original architecture are discussed in Chapter 2 of the thesis.

This thesis deals with the theoretical and practical aspects of higher-order incremental converters. The operating principles, topologies, specialized digital filter design methods and circuit level issues are all addressed. The theoretical results are verified by showing design examples, simulation results and measurements on implemented circuits. Most of the results discussed in this thesis have been published previously in [Márkus et al., 2004; Temes et al., 2004; Márkus et al., 2003; Márkus, 2003; Márkus et al., 2001].

1.2 Structure of the Thesis

The thesis is divided into six chapters.

Chapter 2 discusses the basic operation of the first-order incremental converter, its similarity and differences to the dual-slope and the first-order $\Delta\Sigma$ converters, and its advantages and disadvantages for dc measurement applications. This chapter also contains detailed analysis of the operation under ideal and non-ideal conditions. It also introduces known extensions to the basic structure.

Chapters 3 and 4 contain the main contributions of this thesis to the topic of incremental converters. Chapter 3 gives the possible extensions of the original architecture to higher-order modulators and discusses their basic properties, while Chapter 4 discusses some more advanced properties of the higher-order converters, addressing practical realization problems and different digital filtering techniques. The theoretical analysis is always verified by simulation results.

Chapter 5 shows design examples and selection guides, with detailed comparison between various architectures. It also contains measurement results on a prototype integrated circuit, which implements a 22-bit incremental A/D converter.

Finally, Chapter 6 gives a short overview of the work and discusses second-order problems to be answered in the future, as well as highlights some novel techniques which may be integrated with the introduced technique to further improve the efficiency of dc measuring A/D converters.

Chapter 2

Incremental $\Delta\Sigma$ A/D Converters

This chapter focuses on the prior art of making incremental $\Delta\Sigma$ converters. It introduces the basic idea, the first-order incremental converter and discusses its operation in details. Some new results regarding to the structure, operation and sensitivity of the first-order incremental converter are also introduced. The different extensions of the introduced architecture found in the literature are also analyzed. The chapter finishes with some concluding remarks about the problems arisen by the basic structure.

2.1 First-order Incremental (Charge-balancing) Converter

The first-order incremental converter was first introduced by van de Plassche (1978). He presented a converter with 5-digit + sign-bit resolution ($\approx$ 17-bit resolution), based on a $\Delta\Sigma$ structure. He implemented his design in bipolar technology, using switched-current sources. Later, Robert et al. (1987) introduced a similar structure with more theoretical details in a low-voltage CMOS environment, achieving 16-bit resolution, naming the converter "incremental $\Delta\Sigma$ converter". As the first-order incremental converter's operation has many similarity to that of the dual-slope Nyquist-rate converter, first this latter's operation is recalled briefly.

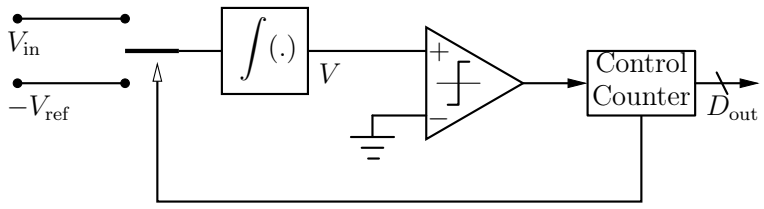


Figure 2.1: Block diagram of the dual-slope converter. V_{in} is the input signal ($V_{\text{in}} \in [0, V_{\text{ref}}]$), V_{ref} is the reference signal, V is the output of the integrator, and D_{out} is the digital output.

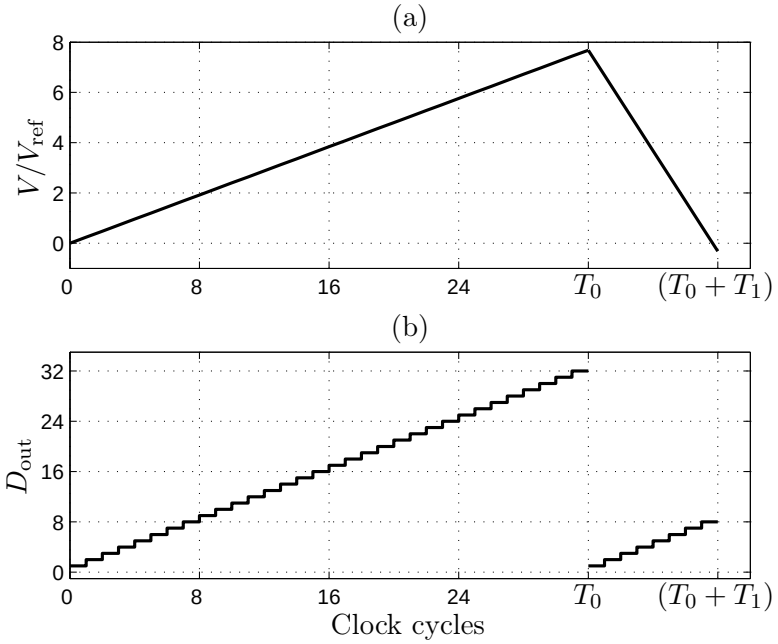


Figure 2.2: Waveforms of a 5-bit dual-slope converter. (a) Normalized output of the integrator (V/V_{ref}), (b) Output of the counter (D_{out}).

2.1.1 Dual-slope Converter

The unipolar dual-slope (or charge-balancing) converter (Fig. 2.1) contains an integrator and a comparator [van de Plassche, 1994, Chap. 7]. It operates in a two-cycle mode (Fig. 2.2). In the first cycle, the unknown input signal ($V_{\mathrm{in}} \in [0, V_{\mathrm{ref}}]$) is entered into the integrator for a given time interval, T_0. Here, T_0 equals to $N = 2^{n_{\mathrm{bit}}}$ periods of the high-frequency clock signal (T_{clk}), where n_{bit} is the required resolution in bits. At the end of the first cycle, a known reference voltage $-V_{\mathrm{ref}}$ is applied to the same integrator until the output of the integrator reaches (to be more exact, crosses) again 0. The length of this cycle is measured (counted) using the same clock. Let the length of the second cycle be T_1, during which the counter counts up to N_{out}. Then, it is readily shown for constant signals that

$$V_{\mathrm{in}}/V_{\mathrm{ref}} = T_1/T_0 + \varepsilon/2^{n_{\mathrm{bit}}} = N_{\mathrm{out}}/2^{n_{\mathrm{bit}}} + \varepsilon/2^{n_{\mathrm{bit}}}, \tag{2.1}$$

where the error (caused by the finite clock frequency) ε satisfies $0 \leq \varepsilon < 1$. Thus, the ratio of the input signal and the reference signal is obtained with n_{bit}-bit resolution [van de Plassche, 1994, Chap. 7].

There are several advantages of the dual-slope converter for measurement applications. First, if the integrator is realized as an analog RC-integrator, it can be easily shown that the output is independent of the RC-constant, as this is cancelled by the double integration (this is the main advantage of the converter compared to the single-slope converter). Thus, the main error source is successfully eliminated. Second, the converter can be implemented with a few elements, thus the converter is power- and area-efficient. Additional advantage of the converter is the capability of periodic noise suppression. The unknown input signal

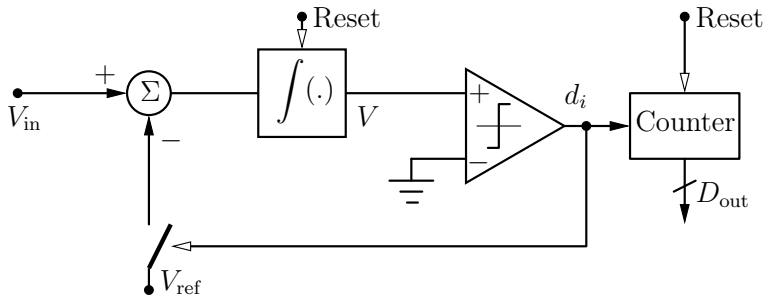


Figure 2.3: Block diagram of the first-order unipolar incremental converter. V_{in} is the input signal ($V_{\text{in}} \in [0, V_{\text{ref}}]$), V_{ref} is the reference signal, V is the output of the integrator, $d_i \in \{0, 1\}$ is the one-bit output sequence of the comparator and D_{out} is the digital output.

is integrated over a fixed time interval ($T_0 = 2^{n_{\text{bit}}} T_{\text{clk}}$). If this time interval is matched to the multiple of the fundamental period of a periodic superimposed signal (such as the 50 or 60 Hz line frequency noise coupled from the mains), the integration cancels this additive noise. For example, in the case of $n_{\text{bit}} = 16$-bit resolution and required suppression of 50 Hz periodic components, $T_{0,\text{min}} = 20$ ms, $T_{\text{clk}} = T_{0,\text{min}}/2^{n_{\text{bit}}} \approx 0.3\mu$s, $f_{\text{clk}} \approx 3.3$ MHz.

In this latter expression, one of the disadvantages can also be observed, i.e., the conversion time of the converter is extremely slow compared to the converter's clock frequency, especially for high ($n_{\text{bit}} \geq 14$ bit) resolution. The worst-case conversion rate (when the input signal is approaching the reference signal) can be calculated as $2^{n_{\text{bit}}+1} T_{\text{clk}}$. Other disadvantages include offset-errors (can be compensated though, at the expense of longer conversion time or even higher clock-rate), large capacitor and resistor values for proper settling of the maximum output of the integrator. These properties make its implementation in low-voltage CMOS technology difficult.

2.1.2 Unipolar First-order Incremental Converter

The unipolar first-order incremental converter (Fig. 2.3) works somewhat similarly to the dual-slope one. The main difference is that the two cycles (integrating the unknown and the reference signal) are interwoven in time. In the following, a discrete-time model of a converter (can be implemented as a SC circuit) is discussed (see Fig. 2.4). The continuous-time version of the converter is usually referred as integrating ΔΣ converter [Robert et al., 1987]. First unipolar operation is assumed, i.e., $V_{\text{in}} \in [0, V_{\text{ref}}]$.

At the beginning of a new conversion, the integrator in the loop and the output counter are both reset. Next, a fixed number ($N = 2^{n_{\text{bit}}}$) of discrete integration steps are performed, where n_{bit} is the required resolution in bits (see the waveforms of Fig. 2.5). Whenever the input to the comparator exceeds zero, its output becomes 1, and $-V_{\text{ref}}$ is added to the input of the analog integrator. After $N = 2^{n_{\text{bit}}}$ steps, the next output of the delaying integrator (which is bounded by $(-V_{\text{ref}}, V_{\text{in}}]$) would become

$$V[N+1] = 2^{n_{\text{bit}}} V_{\text{in}} - N_{\text{out}} V_{\text{ref}}, \tag{2.2}$$

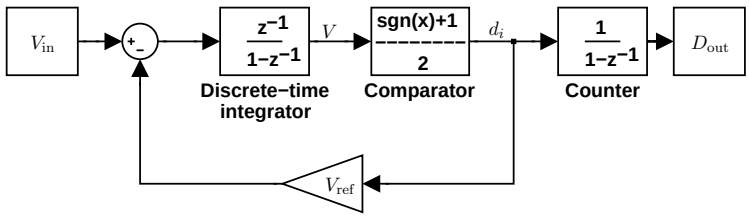


Figure 2.4: Discrete-time model of a first-order unipolar incremental converter. V_{in} is the input signal ($V_{\text{in}} \in [0, V_{\text{ref}}]$), V_{ref} is the reference signal, V is the output of the integrator, $d_i \in \{0, 1\}$ is the one-bit output sequence of the comparator and D_{out} is the digital output.

where N_{out} is the number of clock periods when feedback was applied. Since $V[k]$ must always satisfy $-V_{\text{ref}} < V \leq V_{\text{in}} (\leq V_{\text{ref}})$, it follows that

$$N_{\text{out}} = 2^{n_{\text{bit}}} (V_{\text{in}} / V_{\text{ref}}) + \varepsilon, \tag{2.3}$$

where $\varepsilon \in [-1, 1]$. Generating N_{out} with a simple counter at the output of the modulator, one can easily get the digital representation of the input signal.

In an ideal A/D converter, the analog input signal and the digital output signal can be related by the following equation:

$$D_{\text{out}} V_{\text{lsb}} = V_{\text{in}} + q V_{\text{lsb}}, \text{ or} \tag{2.4}$$

$$V_{\text{in}} = V_{\text{lsb}} (D_{\text{out}} - q), \tag{2.5}$$

where D_{out} is the digital output signal (integer number), V_{in} is the analog input signal, V_{lsb} is the analog equivalent of one bit, and $q \in (-0.5, 0.5]$ or $q \in [-0.5, 0.5)$ is the quantization error. In the case of the unipolar converter discussed above, rearranging Eq. (2.2), one can get

$$V_{\text{in}} = \frac{V_{\text{ref}}}{2^{n_{\text{bit}}}} \left(N_{\text{out}} - (-\frac{V[N+1]}{V_{\text{ref}}}) \right). \tag{2.6}$$

This would imply a $V_{\text{lsb}} = \frac{V_{\text{ref}}}{2^{n_{\text{bit}}}}$, output digital code of $D_{\text{out}} = N_{\text{out}}$ and quantization error of

$$q = -\frac{V[N+1]}{V_{\text{ref}}}. \tag{2.7}$$

However, as V is limited by $-V_{\text{ref}} < V \leq V_{\text{in}} \leq V_{\text{ref}})$, the error q is not limited by $(-0.5, 0.5]$, but by $(-1, 1]$. Simulation results agree well with this statement (Fig. 2.6). One can see (Fig. 2.6(a)) that as the input signal increases, the quantization noise tends to become negative, since its lower limit correlates with $-V_{\text{in}}$ (the integrator's output upper limit is V_{in}). Fig. 2.6(b) shows the inverted output of the integrator, $V[N+1]$, which is exactly the same as q.

There is a easy way to enhance the operation of the converter: simply operate the converter for one more cycle and count one more. In this case the following happens: in this cycle, the output of the integrator ($V[N+1]$) contains the inverse of the quantization error (cf. Eq. (2.7)). If this signal is negative (i.e., the quantization error $q > 0$), the comparator

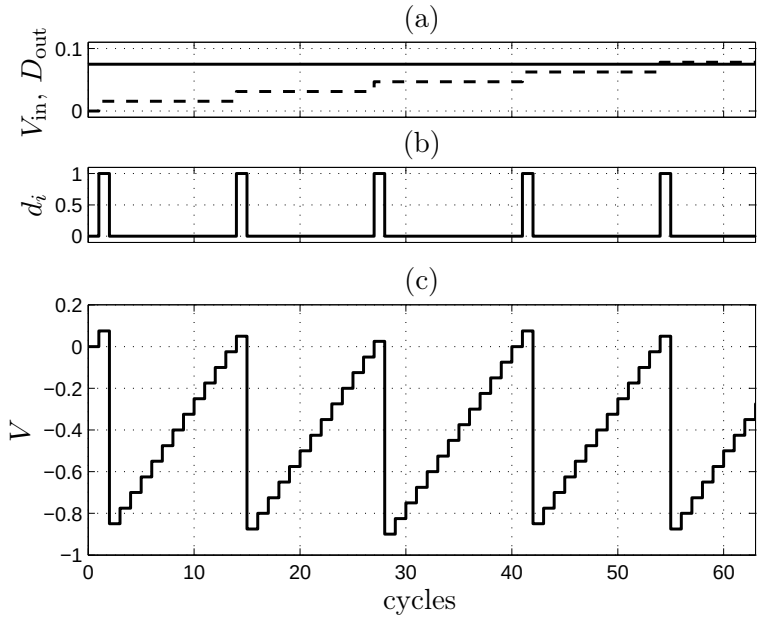


Figure 2.5: Waveforms of the first-order incremental converter. (a) Normalized input signal ($V_{\rm in}/V_{\rm ref} \in [0, 1]$, solid line) and calculated digital output ($D_{\rm out}$, dashed line); (b) output of the comparator (d_i); and (c) output of the integrator ($V/V_{\rm ref}$). $n_{\rm bit} = 6$ bits, $V_{\rm in} = 0.075V_{\rm ref}$, $N_{\rm out} = 5$, indicating a quantized input signal $N_{\rm out}/2^{n_{\rm bit}}V_{\rm ref} = 0.078V_{\rm ref}$. $V_{\rm lsb} = 0.015625V_{\rm ref}$, the quantization error is $q = 0.192V_{\rm lsb}$.

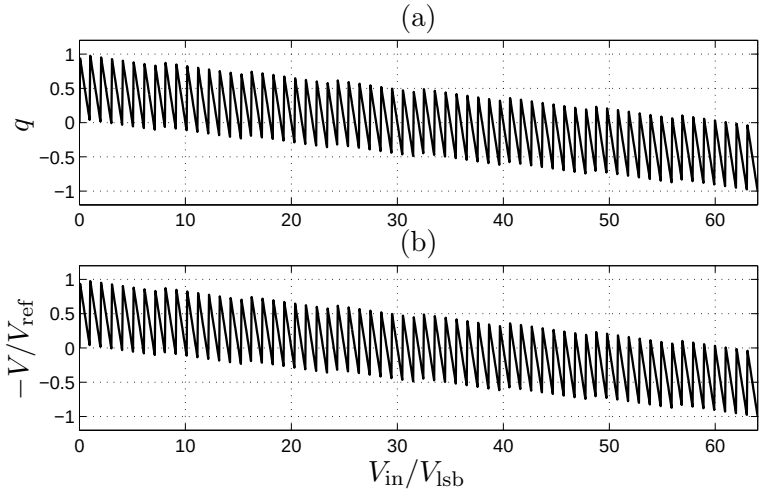


Figure 2.6: (a) Quantization error of a simulated first-order unipolar converter. The error $q \in (-1, 1]$ is not in agreement with its original definition. (b) The inverted output of the integrator ($V[N + 1] = -qV_{\rm ref}$).

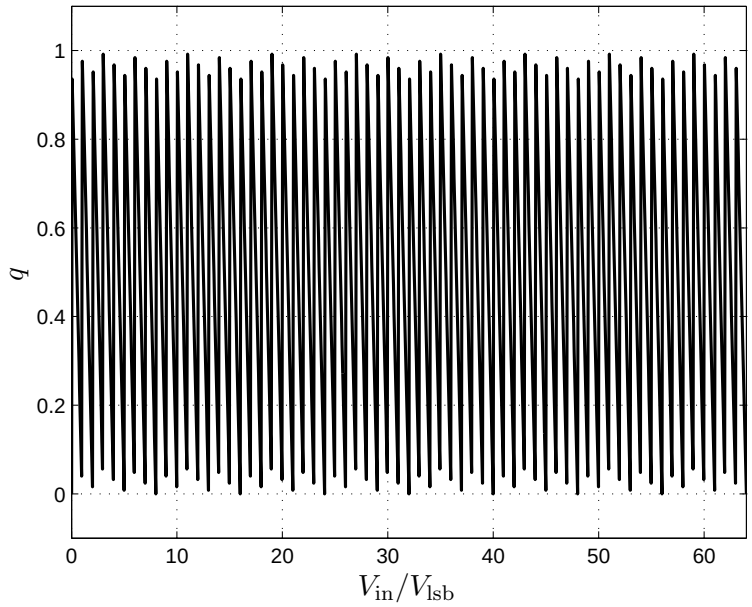


Figure 2.7: Quantization error of a first-order unipolar converter operated through $N+1$ cycles.

is not triggered, thus outputs zero, which does not change the output of the counter. If the output of the integrator is greater than zero (i.e., $q < 0$), then the comparator outputs one, incrementing the counter by one. This means that the quantization error will become $q' = 1 - |q| \in [0, 1)$, a positive number, less than one. With this operation, the quantization error has been successfully mapped into the interval of $[0, 1)$, which causes only a half LSB shift in the output code. Fig. 2.7 shows simulation results agreeing well with the discussion above.

Similar result may be achieved if the structure is realized so that the input signal is not delayed in the analog integrator but only the feedback signal (this latter delay is required to avoid delay-free loop). Note that any of these cases the expression between the output of the integrator and the quantization error (Eq. (2.7)) does not hold anymore.

The above analysis about the operation of the unipolar converter was based on the discrete-time model of the first-order converter. This approach models only those delays which are multiple of the sampling time (T_{clk}), i.e., integer powers of z^{-1}. However, in a switched-capacitor implementation, $z^{-1/2}$, $z^{-1/4}$, etc. delays may also be achieved by dividing the clock signal into many non-overlapping phases and switching the different switches at different phases [Johns and Martin, 1997, Chap. 10] [Robert and Deval, 1988]. By using this technique, it is possible to operate the circuit such that the quantization noise is always in $[0, 1)$ [Robert and Valencic, 1985; Robert et al., 1987], but since in this technique information available within one clock period is used to determine the feedback signal, it is not possible to analyze this circuit using z-domain methods. Instead, the operation must be analyzed in the time domain. This method will be illustrated in the next section.

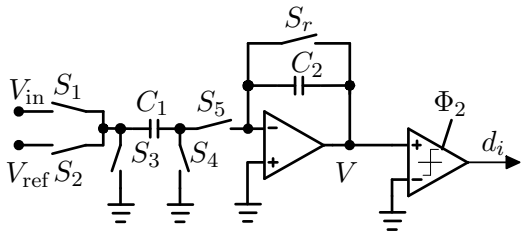


Figure 2.8: Switched-capacitor realization of the first-order incremental converter (without the control logic) operated with four non-overlapping clock-phases.

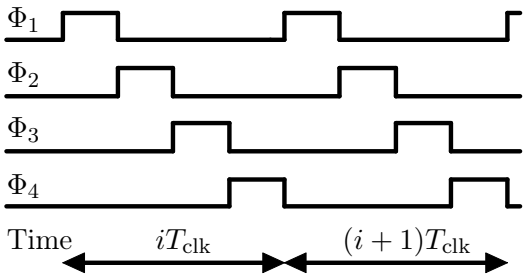


Figure 2.9: Non-overlapping clock-phases for the operation of the first-order incremental converter.

2.1.3 Bipolar Operation

The incremental converter is typically operated with bipolar input signals. This extension is rather straightforward: enabling bipolar input signal ($V_{\text{in}} \in [-V_{\text{ref}}, V_{\text{ref}}]$) and feeding back bipolar reference signal instead of unipolar do the task. However, a detailed analysis in the time-domain is shown here, to introduce the basic operation of switched-capacitor (SC) circuits.

Fig. 2.8 shows a possible SC implementation of the first-order bipolar incremental converter. It consists of a general parasitic-insensitive SC-integrator [Johns and Martin, 1997, Chap. 10] with two input and resetable integrating capacitor, a comparator and control logic (not shown here). Here the switches can be realized as CMOS transmission gates.

In the following, ideal elements are assumed, with $C_1 = C_2$. The circuit needs 4 non-overlapping phase in one clock-period, illustrated in Fig. 2.9.

Before a conversion takes place, S_r is switched on to reset the integrating capacitor C_2. Then, in the ith cycle the switches are operated as follows. In Φ_1, S_1 and S_4 are closed, while all the other switches are open. This causes charging C_1 to V_{in}. In the next phase (Φ_2), S_1 and S_4 are open and S_3 and S_5 are closed. As S_5 is connected to the virtual ground of the op-amp, this phase forces C_1 to discharge. However, this discharging current must flow through C_2, thus, the charge is transferred to C_2, causing V to change to

$$V[i,2] = V[i,1] + \frac{C_1}{C_2} V_{\text{in}}, \tag{2.8}$$

where $[i,k]$ denotes the kth phase of cycle i. During this phase, the comparator is also

enabled. Thus, at the end of this phase it outputs either $d_i = 1$ if $V > 0$, or $d_i = -1$ if $V < 0$. If $d_i = 1$, then $V_{\rm ref}$ is subtracted from V by closing S_3 and S_4 during Φ_3, then closing S_2 and S_5 to transfer $-V_{\rm ref}$ to the output V. Otherwise, when $d_i = -1$, $V_{\rm ref}$ is added to the output similarly to the addition of the input signal, i.e., in Φ_3 S_2 and S_4 are closed to charge C_1, then S_3 and S_5 are closed to transfer this charge to the output. At the end of the cycle,

$$\begin{aligned} V[i,4] &= V[i,1] + \frac{C_1}{C_2}(V_{\rm in} - d_i V_{\rm ref}), \text{ i.e.,} \\ V[i+1,1] &= V[i,1] + \frac{C_1}{C_2}(V_{\rm in} - d_i V_{\rm ref}). \end{aligned} \tag{2.9}$$

One can see from the analysis above, that within one cycle, two integrations take place, and the sign of the second input ($\pm V_{\rm ref}$) depends on the output of the first integration. Thus, analysis of this circuit in the z-domain (assuming that one sample interval is T, consisting of these four cycles) is not straightforward.

A significant difference between the model used for the unipolar operation (cf. Fig. 2.4) and this operation is that it is always assured that during Φ_1 and Φ_4 the output of the integrator is always between $\pm\frac{C_1}{C_2}V_{\rm ref}$, if the input signal $|V_{\rm in}| \le V_{\rm ref}$. This can be proven by induction.

Let $V[i,1] \in [-\frac{C_1}{C_2}V_{\rm ref}, \frac{C_1}{C_2}V_{\rm ref}]$ (which is true for $V[0,1]$ because of the reset signal). If $V_{\rm in} \in [-V_{\rm ref}, V_{\rm ref}]$ is added to this signal, $V[i,2] \in [-\frac{2C_1}{C_2}V_{\rm ref}, \frac{2C_1}{C_2} + 1V_{\rm ref}]$. However, if $V[i,2] < 0$, then $C_1/C_2V_{\rm ref}$ is added to this signal, and if $V[i,2] > 0$ then $C_1/C_2V_{\rm ref}$ is subtracted from this signal during the next two phases. Thus, at the end of Φ_4, $V[i,4] = V[i+1,1] \in [-\frac{C_1}{C_2}V_{\rm ref}, \frac{C_1}{C_2}V_{\rm ref}]$ holds again.

Assuming that the cycle discussed above is repeated up to $N = 2^{n_{\rm bit}}$ cycles, the output of the integrator becomes

$$V = \sum_{i=1}^{N} V_{\rm in} - \sum_{i=1}^{N} d_i V_{\rm ref} = NV_{\rm in} - \sum_{i=1}^{N} d_i V_{\rm ref}. \tag{2.10}$$

For simplicity, let $C_1 = C_2$. Assuming that the input $V_{\rm in}$ is constant and utilizing that V is limited by $\pm V_{\rm ref}$,

$$-\frac{V_{\rm ref}}{N} < V_{\rm in} - \frac{1}{N}\sum_{i=1}^{N} d_i V_{\rm ref} < +\frac{V_{\rm ref}}{N}, \tag{2.11}$$

i.e., the difference of the unknown input signal and the lhs of the expression with known terms (N, d_i, $V_{\rm ref}$) is limited by an interval, which can be made arbitrarily small by increasing N. Thus, an estimate of the input signal is

$$\hat{V}_{\rm in} = \frac{1}{N}\sum_{i=1}^{N} d_i V_{\rm ref}, \tag{2.12}$$

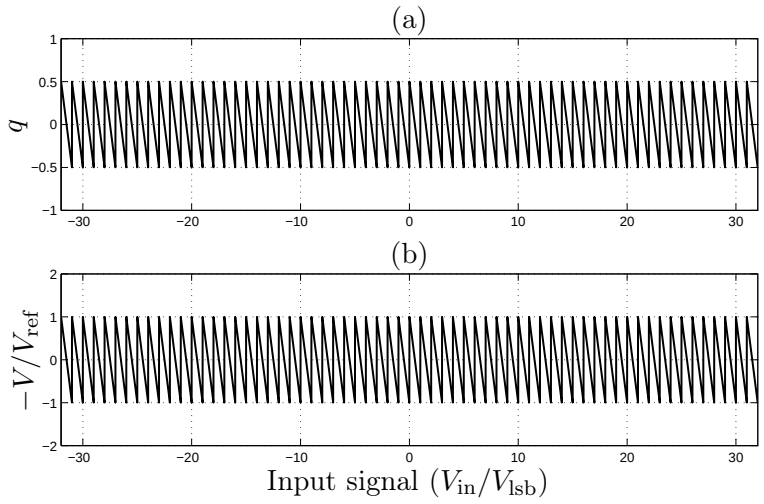


Figure 2.10: (a) Quantization error and (b) inverted output of the analog integrator at the end of the conversion of a 6-bit bipolar first-order incremental converter.

i.e., the digital output can be calculated simply by an up-down counter:

$$D_{\text{out}} = \sum_{i=1}^{N} d_i. \tag{2.13}$$

According to the definition of an ideal A/D converter (Eq. (2.4)), the limits in Eq. (2.11) are equal to $\pm\frac{V_{\text{lsb}}}{2}$, i.e.,

$$V_{\text{lsb}} = \frac{2V_{\text{ref}}}{N} = \frac{2V_{\text{ref}}}{2^{n_{\text{bit}}}}. \tag{2.14}$$

The quantization error q of the conversion can be calculated the following way:

$$q = \frac{\hat{V}_{\text{in}} - V_{\text{in}}}{V_{\text{lsb}}} = \frac{\frac{1}{N}\sum_{i=1}^{N} d_i V_{\text{ref}} - V_{\text{in}}}{\frac{2V_{\text{ref}}}{N}} = \frac{1}{2}\sum_{i=1}^{N} d_i - \frac{1}{2}\frac{NV_{\text{in}}}{V_{\text{ref}}}. \tag{2.15}$$

Comparing this result with Eq. (2.10), it follows that

$$V[N] = -2V_{\text{ref}}q, \tag{2.16}$$

i.e., *the quantization error* of the converter *is available in analog form* at the end of the conversion. This can be used to further refine the digital output: e.g., with minimal effort, the sign of this analog signal can be detected in the next cycle, gaining one more bit precision. More details will be discussed in Sec. 2.2 and in Chaps. 3 and 4.

Fig. 2.10 shows the quantization error (q, Fig. 2.10(a)) and the output of the integrator at the end of the conversion ($V[N]$, Fig. 2.10(b)) for a 6-bit bipolar first-order incremental converter, as functions of the input signal. It can be seen that in this case the quantization error is similar to that of a Nyquist-rate converter, and that Eq. (2.16) is satisfied.

Note that similar operation can be achieved with only two clock-phases. One solution

is (as discussed in the previous subsection) is to operate the circuit for one more cycle to get the correct quantization noise. Although Eq. (2.16) does not hold in this case, it can be proven that

$$V[N+1] = -2V_{\text{ref}}q + V_{\text{in}} \tag{2.17}$$

holds. Thus, switching off the input signal for the last conversion cycle (which does not affect the output signal due to the delay in the loop) causes Eq. (2.16) to be true, thus the analog form of the quantization noise may be reused to refine the output. The advantage of the method is less clock phases, which results either in a lower clock frequency and thus less stringent requirements on the op-amp unity gain bandwidth (UGB), or results in faster conversion time. The disadvantage of the method is that it requires bipolar reference signal to make it possible to subtract or add the reference signal during one clock phase (Φ_2).

Instead of switching off the input signal for the last conversion, another solution is to make the signal path from the input signal to the internal quantizer delay-free, either by using non-delaying integrator (which can be implemented with the same hardware and different switching scheme), or by introducing another input signal path, which feeds forward the input signal directly to the input of the internal quantizer. This latter technique will be used for higher-order converters discussed in the next chapter.

2.1.4 Implementation Details

In the detailed analysis of the previous subsection, ideal elements were assumed. In a real converter several non-idealities may degenerate the performance. These are recalled here briefly.

If the circuit is desired for dc measurement application, offset and charge-injection errors need to be kept very small. Offset error is usually caused by the op-amp, while charge injection is caused by the capacitance of the non-ideal switches used in the circuit. To be able to cancel these errors, first the charge injected into the circuit by the non-ideal switches must be made signal-independent. This can be achieved by delaying the operation of signal-flow switches to those which are connected to fix potential. In particular, in Fig. 2.8, switches S_1–S_3 have to be delayed with respect to those of S_4 and S_5 [Johns and Martin, 1997, Chaps. 7, 10], [Haigh and Singh, 1983]. If these charges are signal-independent, then these introduce an additional input-related offset error, which can be cancelled similarly to the error induced by the op-amp.

Signal-independent constant offset signal can be cancelled many ways. One possible way is to do two conversion, the first one is with zero input signal, and the second one with the unknown signal, and then subtracting the result of the first one from the second. However, this solution doubles the conversion time. Another way is to split the conversion to two parts. During the first part the conversion starts as discussed in the previous subsection, then the output of the integrator ($V[N/2]$) is inverted by the usage of one additional switch, and in the second part of the conversion the input signal is not added, but subtracted from the output of the integrator by using another switching scheme [Robert et al., 1987]. With this inversion, the unipolar offset signal is successfully averaged out from the output. Another method is the usage of auto-zeroing circuit, which cancels the

offset of the op-amp right before the conversion takes place [Robert et al., 1987] [Enz and Temes, 1996].

The analog noise introduced by the switched capacitor circuit is another limiting factor. In switched-capacitor circuits any switched capacitor (C_1 in Fig. 2.8) together with the finite-resistance switches is a noise source with a noise power variance $\sigma^2 = k\mathrm{T}/C$, where k is the Boltzmann-constant, T is the temperature in Kelvin, and C is the value of the capacitor [Johns and Martin, 1997, Chap. 4]. From this, one can calculate the minimum capacitor size for a circuit with 16-bit resolution and 1V reference voltage. In this case, the variance of the quantization noise (assuming white-noise model) is

$$\sigma_q^2 = \frac{V_{\mathrm{lsb}}^2}{12} = \frac{4 \cdot V_{\mathrm{ref}}^2}{2^{2n_{\mathrm{bit}}} 12}, \tag{2.18}$$

i.e., $\sigma_q \approx 8.81$ μV. To make sure that the noise variance from this capacitor is less than this value,

$$\frac{k\mathrm{T}}{C} < \sigma_q^2 \tag{2.19}$$

must hold. In this particular case, $C > 53$ pF is required for this resolution, which is a fairly large capacitor value.

However, in an incremental converter, the input signal is sampled and held by the first capacitor several times and the final result contains the sum of these samples (cf. Eq. (2.10). If the input-referred total noise is Gaussian and has zero mean and σ_g^2 variance, the output of the converter will have a variance σ_g^2/N. Note that this simple averaging is called the best linear unbiased estimator (BLUE) of the input V_{in} with additive zero-mean Gaussian noise. Thus, even higher noise level (even smaller capacitors) may be enabled in the circuit. For example, in [Robert et al., 1987] $C = 10$ pF capacitors were used. Their input-referred rms noise is about 20 μV, but it is divided by $\sqrt{N} = 2^{n_{\mathrm{bit}}/2}$ in the final output, causing an rms error $20/2^{n_{\mathrm{bit}}/2}/V_{\mathrm{lsb}} = 0.002$ LSB in a 16-bit converter with 1 V reference. Note that this drastic reduction in the noise contribution is due to the large number of samples averaged during the conversion.

Another important error source in the circuit is the finite op-amp gain, which causes the leakage of the integrator. Robert et al. (1987) found that the error contribution to the output of the integrator with an op-amp gain $A < \infty$ is

$$E_g = \beta \frac{2^{n_{\mathrm{bit}}}}{A} \tag{2.20}$$

in LSB, where $0.8 < \beta < 1$ depending on A and n_{bit}. This implies a relatively large A, e.g., for 16-bit precision $A > 100$ dB is required, which cannot be easily achieved. Using correlated double sampling [Enz and Temes, 1996] may virtually double the op-amp gain in dB, dropping the op-amp gain requirement to about 60 dB.

Nonlinearities of the capacitors in the circuit were also analyzed and found that capacitors must have a low voltage sensitivity. The circuit is most sensitive to the input sampling capacitor. To achieve the desired linearity, low-voltage sensitivity technology can be used (e.g., metal-metal or poly-poly capacitors with SiO_2 insulator may be used instead

of nitride-oxid) together with design methods (e.g., connecting capacitors with opposite polarity in serial or parallel, to reduce first-order nonlinearity effects).

The introduced converter requires only simple analog and digital circuitry, needs no precision components (as the output is independent of the ratio of the sampling capacitor C_1 and the integrating capacitor C_2). Another advantage is that utilizing the four-phase operation discussed in detail in the previous section, only a single reference is required for bipolar operation. This is essential for high-precision conversion. Note that due to the simple circuitry, the area and power requirements are also very modest [Robert et al., 1987] for moderate resolutions. Over the years several paper has reported successful application of the converter [Yufera and Rueda, 1996; Nakamura et al., 1997; Yufera and Rueda, 1998].

The incremental converter is structurally similar to the conventional first-order delta-sigma ($\Delta\Sigma$) converter, but there are significant differences: (i) the converter does not operate continuously; (ii) both the analog and digital integrators (in general: memory elements) are reset after each conversion; and (iii) the decimating filter following the $\Delta\Sigma$ modulator can be realized with a much simpler structure (in this case, with a simple counter).

2.2 Extensions of the First-order Converter

Over the years, the basic idea of [van de Plassche, 1978] and [Robert et al., 1987] has been modified and improved several way, since the fundamental drawback of the original converter is that it must be operated through $2^{n_{\text{bit}}}$ clock cycles to achieve n_{bit}-bit resolution. Thus, the conversion (output) rate is extremely slow compared to the circuit's clock frequency.

The improvements can be classified into two groups: the first one uses the fact that the quantization error is available in analog form at the end of the conversion. Using this signal it is possible to further refine the resolution. Thus, these methods usually perform a coarse quantization with the first-order incremental converter, then make one or more fine quantization cycles using either different or the same hardware. These modifications are described in the next subsection.

Other methods searched for different (mainly higher-order) structures, which are operated in a similar manner, but due to the higher-order architecture lower conversion time can be achieved, similarly to $\Delta\Sigma$ modulation. These methods are summarized in Sec. 2.2.2 and 2.2.3.

2.2.1 Refining the Quantization Noise

Recalling Eq. (2.16), in a first-order incremental converter the quantization error is available in analog form at the output of the integrator in the Nth or the $N+1$st cycle. The quantization error ($q \in [-0.5V_{\text{lsb}}, 0.5V_{\text{lsb}}]$) is mapped into a signal range $\pm V_{\text{ref}}$ (cf. Fig. 2.10). As this is a large analog signal, it can be easily used as part of a digital correction scheme, to further refine the conversion's resolution. In the related literature, this approach is sometimes called as extended counting conversion.

One of the simplest approach is to use a multi-bit Nyquist-rate converter, which captures this signal and converts it into a fine digital value which may be concatenated to the digital output of the incremental converter. Its one-bit version (i.e., detecting the sign of the residual signal $V[N+1]$ at the end of the conversion) was utilized already in [Robert et al., 1987]. Later, Harjani and Lee (1998) applied a multi-bit Nyquist-rate converter to lower the required number of cycles in the converter and compensate the resolution loss by the multi-bit converter operated at the Nyquist-rate.

More sophisticated ideas use the same hardware to further refine the residue error. Jansson (1995) used successive approximation at the end of the coarse incremental conversion, applying a reduced-by-half feedback signal in every step, and then keeping this signal on or switched off, depending on the output of the comparator. With this extension, the conversion accuracy was greatly improved (16-bit resolution), while conversion time and area-requirements could be well controlled.

Rombouts et al. (2001) introduced an algorithmic (cyclic) converter in which at the end of the incremental conversion the same hardware was used to refine the quantization noise, doubling the residual error by two in every step and use the comparator to detect the next bit.

Mulliken et al. (2002) introduced a two-step algorithmic conversion. In this case, they used the hardware first as a first-order incremental converter, then the residue error at the output of the integrator was resampled and used as an input signal for the next N cycles. Thus, the resulting converter used $2 \cdot 2^{n_{\mathrm{bit}}/2}$ cycles to achieve a resolution of n_{bit} bits. The design resulted in very low power and reduced chip area, applicable to high-density integration such as real-time analog array processing.

2.2.2 Using Different Architecture

Another way to decrease the conversion time is to use different, more complex architecture. In ΔΣ modulators, there are two ways to increase the achievable *SNR* in general: one is using multi-stage noise shaping (MASH or cascaded) architecture, the other is to use higher-order modulators. Applications of these techniques for incremental conversion will be discussed in the following.

Robert and Deval (1988) described the use of two-stage (MASH) incremental converter, consisting of two cascaded first-order modulators. In addition, by detecting the sign of the output of the integrator in the second stage at the end of the conversion, an extra bit of resolution was obtained. Using a 2-stage architecture, the number of clock periods required for 16-bit accuracy was reduced to $N = 362$ (or $N = 257$, if the sign of the last integrator's output was used to pick up an extra bit) from the much larger value 2^{16} needed for the first-order converter. As the circuit cancels the outputs of cascaded stages, it is sensitive to circuit non-idealities such as component mismatches and finite op-amp gains. A similar solution, based on a modular architecture and extended to higher-order MASH structures was proposed in [Nys and Dijkstra, 1993].

2.2.3 Higher-order Modulators

Another way of extending the resolution of incremental converters is to use higher-order single-stage modulators. Even though some commercially available converters [Analog, 2004; Burr-Brown, 2004; Cirrus, 2004; Linear, 2004] may use such structures, their theory and design methodology seems to be unavailable in the open literature. These products are sometimes referred as charge-balancing $\Delta\Sigma$, one-shot or one-cycle $\Delta\Sigma$, or no-latency $\Delta\Sigma$ converters. An example of the few relevant publications is [Johnston, 1991], however it is mainly a data sheet without detailed description of the operation.

A similar approach was introduced in [Lyden, 1993], and [Lyden et al., 1995]. The idea was here to extend the first-order converter to higher-order one by matching the analog processing of the feedback signal in the modulator with the digital filter following the modulator. The idea required precise matching between the analog and digital coefficients or required longer conversion cycle to compensate for the mismatch errors. As this technique is very similar to the one proposed in this thesis, it is analyzed in detail in Sec. 3.2.2.

The next two chapters of the thesis deals with the theoretical operation and properties of higher-order incremental converters. They also address the most relevant practical problems arising from circuit non-idealities, giving several solutions for the different limitations.

Chapter 3

Extensions to Higher-order Architectures

This chapter and the next one focus on the new theoretical results achieved in the field of incremental converters. This chapter consists of two sections. In the next section, the first-order incremental converter is modified with higher-order filtering and dither. Then, three different extensions of the original architecture to higher-order modulators are analyzed in Sections 3.2.1, 3.2.2 and 3.2.3. Many basic properties of the structures are discussed, and at the end of the chapter a comparison of the proposed extensions are given.

3.1 First-order Modulator with Higher-order Filtering

Consider the bipolar first-order incremental converter, consisting of a first-order modulator and a counter. Its bipolar model used for simulations is shown in Fig. 3.1. This circuit models the SC-circuit operated with two clock phases (see the discussion at the end of Sec. 2.1.3), up to $N+1$ cycles. The Enabler block is used to disable the input signal for cycle $N+1$, to ensure the validity of Eq. (2.16).

This structure operates similarly to the one realized as SC-circuit with four clock phases (Sec. 2.1.3). Thus, the output of the converter has a quantization error $q \in [-0.5V_{\text{lsb}}, 0.5V_{\text{lsb}})$, and the output of the integrator after the last cycle, $V[N+2] = -2V_{\text{ref}}q$.

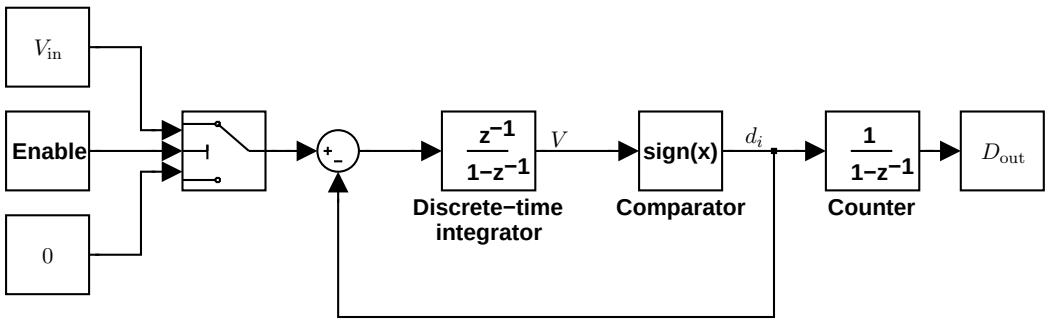


Figure 3.1: Discrete-time model of a first-order bipolar incremental converter.

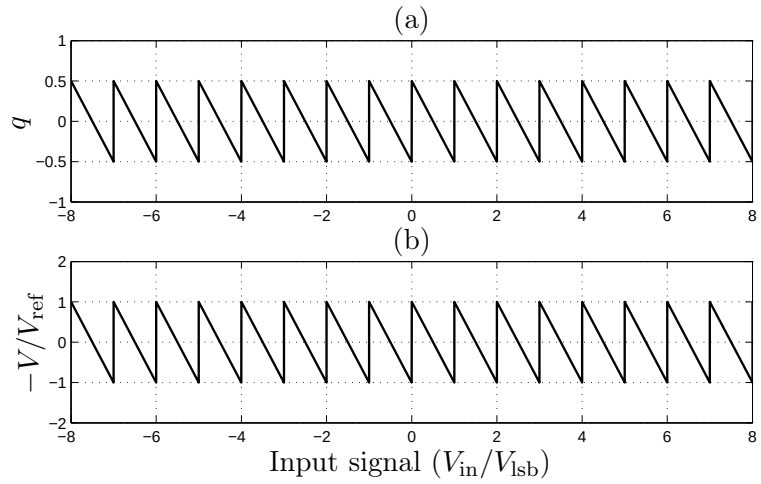


Figure 3.2: (a) Quantization error and (b) inverted output of the analog integrator at the end of the conversion of a 10-bit bipolar first-order incremental converter, around zero input.

This is shown again in Fig. 3.2 for a 10-bit converter, zooming out the converter's error around zero.

As it was discussed in the previous chapter, the converter's biggest drawback is that it requires

$$N = 2^{n_{\mathrm{bit}}} + 1 \tag{3.1}$$

cycles to achieve n_{bit}-bit resolution. As the architecture is similar to that of a $\Delta\Sigma$ modulator, a useful idea is to use different filter at the output, similarly to the decimation of $\Delta\Sigma$ modulators.

Recalling Eq. (2.13), the output of the first-order converter can be calculated simply by using an up-down counter operated through the first $N+1$ cycles. As it can be seen in Fig. 3.1, the counter can be modeled as a discrete-time integrator, operated in transient mode. Thus, it realizes an accumulate-and-dump type decimation filter. The output is calculated as

$$D_{\mathrm{out}}[N+1] = \frac{1}{N}\sum_{i=1}^{N} d_i = \frac{1}{N}\sum_{i'=1}^{N} d_{N-i'}, \tag{3.2}$$

switching to z-domain yields

$$D_{\mathrm{out}}(z) = \frac{1}{N}\sum_{i'=1}^{N} z^{-i'} D(z), \tag{3.3}$$

thus, the transfer function becomes

$$H(z) = \frac{D_{\mathrm{out}}(z)}{D(z)} = \frac{1}{N}\sum_{i'=1}^{N} z^{-i'} = \frac{1}{N}\frac{1-z^{-N}}{1-z^{-1}}, \tag{3.4}$$

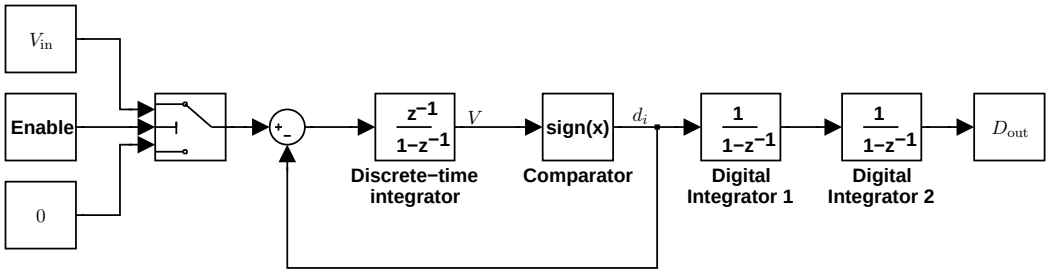


Figure 3.3: Model of a first-order incremental converter with two digital integrators at the output.

and its frequency-response is (using $z = \mathrm{e}^{j\omega T_{\text{clk}}}$)

$$H(f) = \frac{\text{sinc}(fNT_{\text{clk}})}{\text{sinc}(fT_{\text{clk}})} = \frac{\sin(\pi fNT_{\text{clk}})}{N\sin(\pi fT_{\text{clk}})}. \tag{3.5}$$

This filter has zero dB attenuation at dc and at the multiples of $f_{\text{clk}} = 1/T_{\text{clk}}$, and has multiple zeros at the output rate of $1/(NT_{\text{clk}})$ and at its harmonics (except where the harmonics coincide with $1/T_{\text{clk}}$). It is commonly called as first-order digital sinc-filter.

3.1.1 Analysis of Higher-order Filters

Decimation filters following $\Delta\Sigma$ modulators in $\Delta\Sigma$ A/D converters usually consists of a higher-order sinc-filter, which decimates the signal output rate to about four times the Nyquist-rate, and final output rate is achieved by either half-band or other FIR-filters [Norsworthy et al., 1997, Sec. 1.3 and Chap. 13]. First-order digital sinc-filters (like the one discussed above) was in use at a time when it was important to save digital hardware [Candy, 1974]. Later Candy et al. (1976) analyzed the application of second-order sinc-filter at the output of the first-order $\Delta\Sigma$ modulator. Also, a good tutorial about this topic was published in [Candy, 1986]. This tutorial lead to the conclusion that the best trade-off in decimator filter design for $\Delta\Sigma$ applications is to use L_a + 1st order filter for an L_ath-order modulator and decimate the output rate to four times the Nyquist-rate. This rule is still widely used for decimator design, especially since very hardware-efficient implementation techniques (cascaded integrators and comb (CIC) filters) exist [Hogenauer, 1981] [Norsworthy et al., 1997, Sec. 1.3 and Chap. 13].

The idea of using higher-order digital filters for incremental conversion is a natural adoption from the design techniques of $\Delta\Sigma$ converters. However, as it was stated in [Robert and Deval, 1988], by using second-order filter for a first-order incremental converter, the resolution and the average accuracy may be increased, but the quantization error around zero remained the same [Robert and Deval, 1988, Fig. 6]. Their experience has been validated and repeated here: Fig. 3.3 shows a model which has second-order filter at the output. Here only the Cascade-of-Integrators (CoI) filter is considered.

Using the same number of cycles ($N = 2^{10}$) as previously, Fig. 3.4 shows the quantization error at the output of the second integrator as a function of the input dc amplitude. This figure is similar to Fig. 6(b) of [Robert and Deval, 1988]. Theoretically, such a configu-

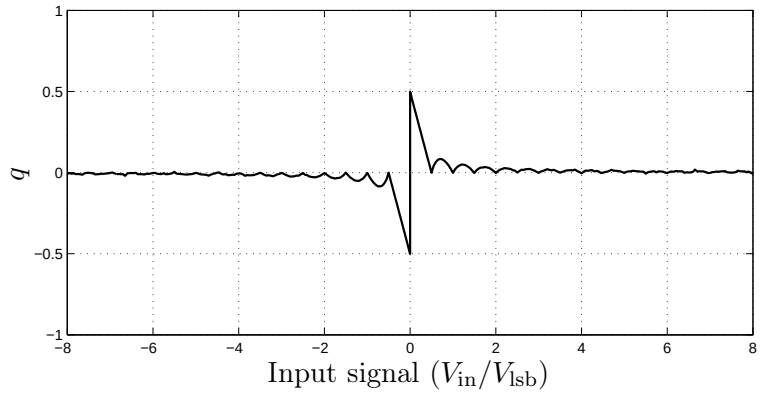


Figure 3.4: Quantization error at the output of the second digital integrator as a function of the input amplitude.

ration could provide $N(N+1)/2$ different output code, which would increase the resolution to $\log_2(N(N+1)/2) \approx 2\log_2(N) - 1$. If the original converter had 10 bit resolution, this improvement would give 19(!) bits of resolution for the same number of cycles.

In reality, comparing Fig. 3.4 and Fig. 3.2, it can be seen that although the average quantization error has been reduced, there is one peak error around zero, which remained the same. Thus, the maximum error signal and the effective/equivalent number of bits (ENOB) around zero remained the same.

The unchanged peak error can be explained by realizing that for dc input signals (especially for very small ones) the linearized models of the internal quantizer and the modulator are no longer valid, i.e., the quantization noise is strongly correlated with the quantizer input. Instead of using linearized (white-noise) modell, the actual operation needs to be analyzed in the time domain. Consider the first-order incremental converter (Fig. 3.1). Comparing Figs. 3.2 and 3.4, one can see that using second-order filter the anomaly arises only when the input signal is within ± 0.5 LSB of the 10-bit resolution converter. Recalling the operation of the first-order (unipolar) incremental converter (Fig. 2.5), it is clear that when the incoming signal is this small, even if it is integrated through N cycles, the output of the integrator does not trigger the comparator during the limited number of cycles (N). Hence, no feedback is applied, and the loop does not become functional. Although in a $\Delta\Sigma$ converter, similar "dead-zones" exist around other input values (typically around low-order fractions of V_{ref}) [Norsworthy et al., 1997, Chap. 1], in the current case the transition of the comparator is triggered at those inputs, thus the effect is most significant around zero.

This "dead-zone" problem can be eliminated, and hence the higher-order filter becomes effective, if the comparator is forced to make decisions and thus the whole loop is forced to operate even for extremely small input signals. This can be achieved by dithering. Injecting a dither signal into the loop right before the quantizer [Norsworthy et al., 1997, Chap. 3] should eliminate the error peak around zero. Fig. 3.5 shows the improved first-order converter with second-order digital filtering and injected dither signal. Note that dithering of $\Delta\Sigma$ converters used for dc measurement applications have been addressed in [Badmirowski and Jackiewicz, 1998; Badmirowski and Jackiewicz, 1999], but there the dither signal was applied at the input and either filtered or subtracted from the digital

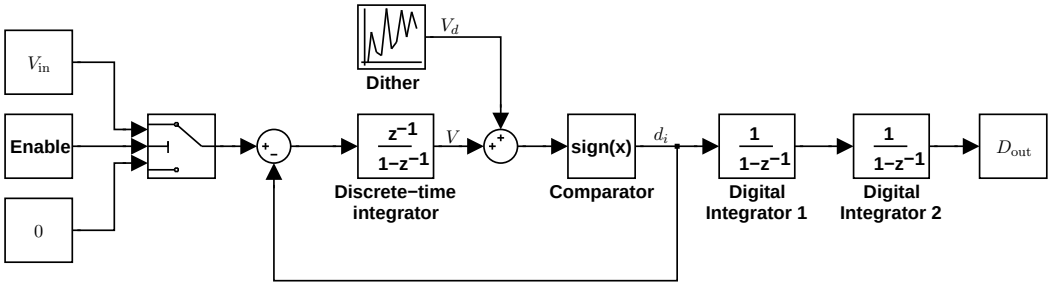


Figure 3.5: Improved first-order converter with second-order digital filter and injected dither signal.

output after conversion, and many conversion was averaged to remove the error peaks and to reduce the noise variance. Here, the structure itself is modified to remove the error peaks during one conversion.

3.1.2 Analysis of the Dither Signal

To analyze the possible error caused by the injected dither in the measurement of the dc component of the input signal, replace the quantizer with an adder, which adds the appropriate quantization error in every clock cycle. Note that this model is equivalent to the original quantizer if no assumptions are taken about the behavior of the quantization error signal. Assuming infinite operation, z-domain analysis of the structure is possible and leads to a result similar to that of a first-order $\Delta\Sigma$ converter:

$$D(z) = z^{-1}V_{\text{in}}(z) + (1 - z^{-1})(V_d(z) + \varepsilon(z)), \tag{3.6}$$

where $D(z)$ is the output of the modulator, $V_{\text{in}}(z)$ is the input signal, $V_d(z)$ is the injected dither signal and $\varepsilon(z)$ is the quantization error of the 1-bit internal quantizer. Switching this equation to the time-domain gives

$$d_k = V_{\text{in}}[k-1] + (V_d[k] - V_d[k-1]) + (\varepsilon[k] - \varepsilon[k-1]). \tag{3.7}$$

Assuming that $V_{\text{in}}[k] = V_d[k] = \varepsilon[k] = 0$ (and also $V[k]$) for $k < 0$, this equation can be used for analysis of the output signal in the time domain. Consider the case when there is no dither signal and only one integrator processes the output signal (Fig. 3.1). With zero initial conditions, it can be readily found that

$$D_{\text{out}} = \frac{1}{N}\sum_{i=0}^{N} d_i = \frac{1}{N}\sum_{i=0}^{N} V_{\text{in}}[i-1] + \frac{1}{N}\sum_{i=0}^{N}(\varepsilon[i] - \varepsilon[i-1]) = V_{\text{in}} + \frac{1}{N}\varepsilon[N]. \tag{3.8}$$

This gives similar result to the one discussed in the previous chapter, however, here not the output of the analog integrator in cycle N is used as a limit for the output quantization error, but the last sample of the quantization error of the 1-bit internal quantizer.

Applying two integrators at the output of the converter leads to the following digital output code:

$$D_{\text{out}} = \frac{2}{N(N+1)} \sum_{j=0}^{N} \sum_{i=0}^{j} d_i =$$

$$= \frac{2}{N(N+1)} \sum_{j=0}^{N} \sum_{i=0}^{j} V_{\text{in}}[j-1] + \frac{2}{N(N+1)} \sum_{j=0}^{N} \sum_{i=0}^{j} (\varepsilon[i] - \varepsilon[i-1]) =$$

$$= V_{\text{in}} + \frac{2}{N+1} \frac{1}{N} \sum_{j=0}^{N} \varepsilon[j]. \quad (3.9)$$

Now the final output contains the input signal and the linear sum of the quantization errors of the internal A/D converter. If the quantization errors of the internal A/D converter during one conversion cycle would consist of independent, zero mean samples, uniformly distributed between $\pm V_{\text{ref}}$, then the output of the second integrator would have a quantization error with a standard deviation

$$\sigma_q = \frac{2}{N+1} \frac{1}{\sqrt{N}} \sigma_\varepsilon, \quad (3.10)$$

where

$$\sigma_\varepsilon = \frac{2V_{\text{ref}}}{\sqrt{12}}. \quad (3.11)$$

For example, operating the converter up to $2^{10} = 1024$ cycles, the original converter (with one integrator) has 10-bit resolution and an output quantization noise variance $1/N\sigma_\varepsilon = 5.6 \cdot 10^{-4}$ (assuming $V_{\text{ref}} = 1$ V). The same converter with two digital integrators and independent quantization noise would have an output quantization error variance $\sigma_q = 3.52 \cdot 10^{-5}$, resulting in a theoretical *SNR* increase of 24 dB.

Unfortunately, $\varepsilon[k]$, the quantization error of the internal A/D converter is not an independent, zero-mean signal for dc inputs. Thus, the model is not valid and one cannot get nearly 15 bit precision out of an incremental converter operated through 1024 cycles using the proposed technique. For example, if the input signal is zero, the output of the modulator $d_k = \pm 1$, the output of the integrator is $V \in \{0, V_{\text{ref}}\}$ or $V \in \{-V_{\text{ref}}, 0\}$, depending on the first feedback signal. The quantization error is then an alternating series of 0 and 1 (or 0 and -1). The final quantization error at the output using two integrators would be either $+1/(N+1)$ or $-1/(N+1)$, resulting in the large peaks shown in Fig. 3.4.

However, adding dither signal improves the performance. From Eqs. (3.7) and (3.9), it can be readily shown that applying dither signal leads to the following output code:

$$D_{\text{out}} = V_{\text{in}} + \frac{2}{N+1} \frac{1}{N} \sum_{j=0}^{N} (\varepsilon[j] + V_d[j]). \quad (3.12)$$

Consider first, when the input signal is zero. In this case, Tab. 3.1 shows the output of the modulator and the quantization noise $\varepsilon[k]$ and $\varepsilon[k] + V_d[k]$ among other samples during the first few cycles. The applied dither signal is uniformly distributed between $\pm\alpha V_{\text{ref}}$,

where $\alpha < 1$.

In this case, in the error sequence $V_d[k] + \varepsilon[k]$, which is averaged by the second digital integrator (cf. Eq. (3.12)), every first term is the sign of a random variable, while every second term is zero. It can be easily shown, that if the dither signal is a zero-mean, uniformly (or more general, symmetrically) distributed signal, then its sign function provides a discrete distribution, with zero mean ($m = -1 \cdot 0.5 + 1 \cdot 0.5 = 0$) and variance of 1 ($\sigma^2 = (-1)^2 \cdot 0.5 + (1)^2 \cdot 0.5 = 1$). As every second sample is zero in the output, averaging N samples yields only in an $N/2$ reduction of the variance. This means that the standard deviation of the error at the output will be

$$\sigma_q = \frac{\sqrt{2}}{(N+1)\sqrt{N}} \sigma_{\varepsilon + V_d}, \tag{3.13}$$

where now

$$\sigma_{\varepsilon + V_d} = 1. \tag{3.14}$$

Thus, operating the converter up to $N = 1024$ cycles, the output variance becomes $\sigma_q = 4.31 \cdot 10^{-5}$, resulting in an *SNR* increase of 22.3 dB.

If there is an input signal greater than zero, the output sequence cannot be predicted as nicely as in Tab. 3.1. However, as the output of the integrator $V[k]$ and also the output of the modulator d_k contains $u[k-1]$, $\varepsilon[k] + V_d[k] = d_k - V[k]$ does not contain the input signal (cf. Eq. (3.7) and Fig. 3.5), thus its nature does not change too much. In a statistical sense, the quantization error of the internal quantizer will not be $\pm V_{\text{ref}}$ in every first and 0 in every second samples, but will be more uniformly distributed between $\pm V_{\text{ref}}$. This means that even better resolution can be achieved, close to the theoretical value of Eq. (3.10). The only problem we can face with is that for large input signals, the quantizer error $\varepsilon[k]$ is not bounded by $\pm V_{\text{ref}}$, thus, for large inputs, the output error will be larger than expected by Eq. (3.13). This can be avoided by limiting the input signal amplitude by proper scaling, e.g., if the dither signal is uniformly distributed between $\pm V_{\text{ref}}/2$, limiting the input signal $|V_{\text{in}}| < 0.5 V_{\text{ref}}$ will avoid overflow and/or saturation error in the loop. Note that introducing this scaling factor means one bit precision loss. At circuit level, using SC-circuits, this scaling may be easily implemented by using two capacitors for the feedback signal and only one for the input signal at the input branch. To remove the error caused by the mismatch of the two capacitors, the two capacitors may be alternated for the input signal. See Sec. 4.3.2 for details of this technique.

Let us assume that the dither signal is uniformly distributed between $\pm 0.5 V_{\text{ref}}$ and the quantization error is uniformly distributed between $\pm V_{\text{ref}}$. In this case, the sum of these two variables has a trapeze-like probability density function, i.e., it is linear between $(-1.5, -0.5) V_{\text{ref}}$ and $(0.5, 1.5) V_{\text{ref}}$ and constant between $(-0.5, 0.5) V_{\text{ref}}$, zero elsewhere. The resulting random variable has a variance

$$\sigma^2_{\varepsilon + V_d} = \left(\frac{1}{12} + \frac{4}{12} \right) V^2_{\text{ref}} = \frac{5}{12} V^2_{\text{ref}} \tag{3.15}$$

Table 3.1: Output of the modulator of a first-order incremental converter with dither signal and zero input.

k	$V_{\text{in}}[k]$	$V[k]$[a]	$V_d[k]$	$d_k = \text{sign}(V_d[k] + V[k])$	$\varepsilon[k] = d_k - V[k] - V_d[k]$	$V_d[k] + \varepsilon[k]$
0	0	0	$V_d[0]$	$\text{sign}(V_d[0])$	$\text{sign}(V_d[0]) - V_d[0]$	$\text{sign}(V_d[0])$
1	0	$-\text{sign}(V_d[0])$	$V_d[1]$	$\text{sign}(V_d[1] - \text{sign}(V_d[0])) = -\text{sign}(V_d[0])$	$-V_d[1]$	0
2	0	0	$V_d[2]$	$\text{sign}(V_d[2])$	$\text{sign}(V_d[2]) - V_d[2]$	$\text{sign}(V_d[2])$
3	0	$-\text{sign}(V_d[2])$	$V_d[3]$	$\text{sign}(V_d[3] - \text{sign}(V_d[2])) = -\text{sign}(V_d[2])$	$-V_d[3]$	0
4	0	0	$V_d[5]$	$\text{sign}(V_d[4])$	$\text{sign}(V_d[4]) - V_d[4]$	$\text{sign}(V_d[4])$

[a] $V[k] = V[k-1] + u[k-1] - d_{k-1}$

If this signal runs through the filter of Eq. (3.12), the output variance is

$$\sigma_q^2 = \frac{4}{(N+1)^2 N^2} N \frac{5}{12} V_{\text{ref}}^2, \tag{3.16}$$

i.e., its standard deviation becomes

$$\sigma_q = \frac{2}{(N+1)\sqrt{N}} \sqrt{\frac{5}{12}} V_{\text{ref}}. \tag{3.17}$$

As this output quantization error is the sum of N (more or less) independent, identically distributed random variable, its distribution is very close to that of a Gaussian signal. Then, one can use the 3-sigma rule to determine a lower bound for the maximum output error. This maximum quantization error can be dedicated as half LSB error:

$$3\sigma_q \leq \frac{1}{2} \frac{2V_{\text{ref}}}{2^{n_{\text{bit}}}}. \tag{3.18}$$

From this equation, either the required number of cycles for a given resolution or the achievable resolution for a given operation time can be calculated, i.e., after some substitutions and rearranging, and also using the approximation of $N + 1 \approx N$,

$$n_{\text{bit}} \leq \log_2 \left(\frac{(N+1)\sqrt{N}}{\sqrt{15}} \right) \approx 1.5 \log_2(N) - 2, \tag{3.19}$$

and for a given resolution

$$N \geq \sqrt[1.5]{\sqrt{15} \cdot 2^{n_{\text{bit}}}} = 2.46 \cdot 2^{\frac{2n_{\text{bit}}}{3}}, \tag{3.20}$$

e.g., if $N = 1024$, $n_{\text{bit}} = 13$, while if $n_{\text{bit}} = 10$, $N = 250$.

Note that this derivation did not include the one bit resolution loss caused by the scaling of the input signal, which prevents the loop from overflow error. Including also this condition results in

$$n'_{\text{bit}} \lessapprox 1.5 \log_2(N) - 3, \tag{3.21}$$

and

$$N \geq 3.9 \cdot 2^{\frac{2n_{\text{bit}}}{3}}. \tag{3.22}$$

Thus, if $N = 1024$, $n_{\text{bit}} = 12$, while if $n_{\text{bit}} = 10$, $N = 396$. Note that the higher the required resolution, the more the saving in the number of cycles. For example, for $n_{\text{bit}} = 14$, the original converter should be operated through $N_1 = 2^{14} = 16384$ cycles, while the modified converter requires only $N_2 = 2516$ cycles.

3.1.3 Simulation Results

Simulation results agree well with the theoretical expectations discussed above. First, Fig. 3.6 shows the quantization error at the output of the second integrator, using the same scale as in Figs. 3.2 and 3.4 for comparison. Here, a dither signal uniformly distributed

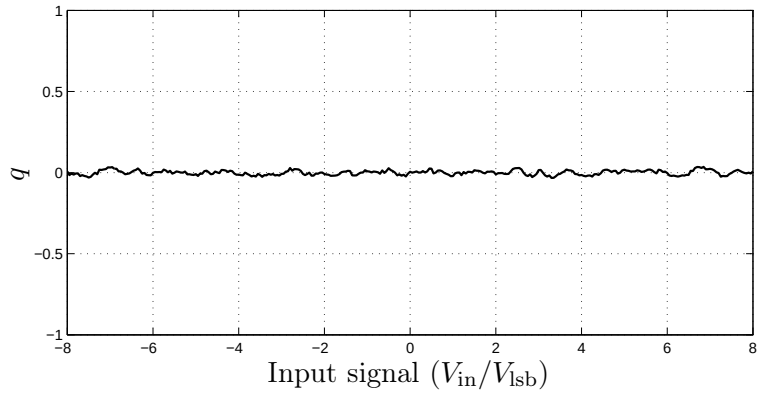


Figure 3.6: Quantization error at the output of the second integrator as a function of the input amplitude, with dither signal injected into the loop.

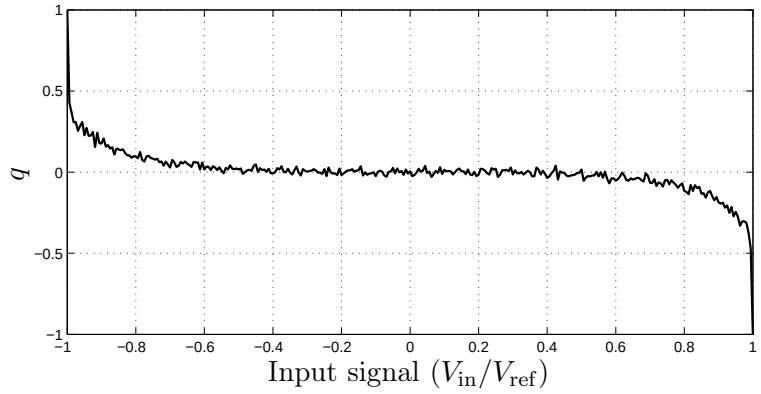


Figure 3.7: Quantization error at the output of the second integrator as a function of the input amplitude range $\pm V_{\text{ref}}$, with dither signal injected into the loop. Note that for large input signals the quantization error is increasing.

between $\pm V_{\text{ref}}/2$ was applied during conversion. It can be seen that the large error peak around zero has been disappeared, and the quantization error (in the range shown) becomes much smaller than 1 LSB of the original 10-bit resolution. $\sigma_q = 2.82 \cdot 10^{-5}$, indicating an *SNR*-increase of 26 dB and ENOB of 13.5 bits, which is slightly better than the theoretical value of Eq. (3.19) and is also better than the conservative value estimated by Eq. (3.13).

To be comparable with Figs. 3.2(a) and 3.4, Fig. 3.6 showed only the output error for small input signals around zero. To verify overload errors, Fig. 3.7 shows a full-scale simulation of the converter with uniform dither between $\pm 0.5 V_{\text{ref}}$. As predicted above, if the input signal is approaching V_{ref}, the quantization error of the internal converter is not limited to $\pm V_{\text{ref}}$, thus, the final quantization error in the output becomes also larger. However, limiting the input signal to $\pm 0.5 V_{\text{ref}}$ eliminates this problem, resulting in one bit resolution loss. Alternatively, dynamic dithering techniques can be used, in which case the dither amplitude is reduced as the input signal amplitude is increasing [Norsworthy et al., 1997, Sec. 3.13]. However, as the introduced converter is dedicated for applications requiring low power- and area-consumption, this method is not considered here.

In the theoretical discussion above, it was proven that for zero input the dither must

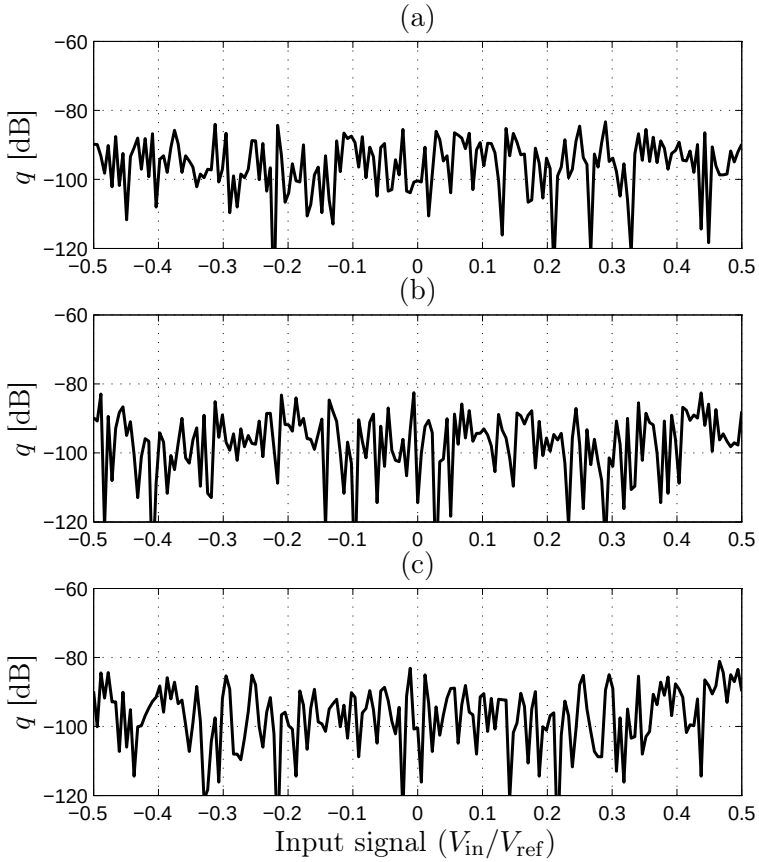


Figure 3.8: Quantization error in dB as function of the input signal amplitude. The probability density function of the injected dither signal is (a) uniform, (b) Gaussian, (c) binary.

have symmetrical distribution, as only its sign is important. Even for larger input, the quantization error in the output does not contain the dither signal itself, only its quantized version. Thus, the converter is expected to be insensitive to the exact distribution of the dither signal. To verify this statement, dither sequences with three different distribution have been applied to the structure. Fig. 3.8 shows the quantization error in dB for uniform dither with limits $\pm 0.4V_{\text{ref}}$ (Fig. 3.8(a)), Gaussian dither with $m = 0$ and $\sigma = 0.2V_{\text{ref}}$ (Fig. 3.8(b)) and binary dither $V_d \in \{-0.3V_{\text{ref}}, +0.3V_{\text{ref}}\}$ (Fig. 3.8(c)). In all cases the peak value of the quantization error remains less than -84dB, indicating an ENOB of about 14 bits. Note that random signals with Gaussian and binary distribution can be easily generated by analog and digital hardware, respectively.

Although the structure is not sensitive to the distribution of the dither signal, it is sensitive to the distribution variance. If the variance is too small, the dither won't remove the error peaks around zero (cf. Fig. 3.2), while if it is too large, the quantization error of the converter may increase even for smaller input signals. The system is most sensitive in the case of binary distribution. This can be improved if the digitally generated dither has a resolution of two or three bits.

As a summary of this section, one can see that using higher-order filtering and ap-

propriate dither signal, the resolution of the incremental converter can be increased by about 2 bits assuming the same number of cycles, without modifying the analog hardware significantly. Thus, while a first-order converter with a first-order digital filter requires $N = 2^{13} = 8192$ cycles to get 13 bits resolution, the same converter using second-order filtering and dither signal can deliver the same performance with approximately $N = 1626$ cycles. More comparisons and design examples will be discussed in Chapter 5.

3.2 Possible Extensions to Higher-order Modulators

If the number of cycles (N) through the converter operates needs to be further reduced, higher-order modulator structures may be used. Similarly to the $\Delta\Sigma$ converters, this leads to less cycles for a given resolution due to the higher loop-gain and the more aggressive noise shaping. Nevertheless, the output quantization error may behave differently to that of a Nyquist-rate converter, due to two effects. First, the final output is calculated by higher-order averaging of many samples, which makes the probability distribution of the output quantization error approximately Gaussian, second, poles have to be introduced in the noise transfer function (NTF) of the converter to stabilize the nonlinear loop. In the following, three possible extensions of the previous results will be discussed. All of the extensions are based on the idea that during one conversion cycle, there are signals in the $\Delta\Sigma$ modulator loop (in particular, either the output of the last integrator or the internal quantization error), which remain bounded during the conversion, independently of N, since existence of the upper and lower bound of these signals is a requirement of stability. As these signals can be used as an upper bound of the higher-order accumulated difference of the unknown input signal and the known output signal, these signals may be used to limit the final quantization error of the converter and to estimate the required number of cycles for a given resolution.

3.2.1 Modulators with Pure Differential Noise Transfer Function

One possible way to extend the operation of the first-order incremental converter to higher-order modulation is based on Eqs. (3.8) and (3.9). In this case, the finite quantization error of the quantizer in the loop is used as a limit for the finale quantization error of the converter.

Theoretical Operation

Consider a higher-order modulator defined by the following equation:

$$Y(z) = z^{-k}U(z) + (1 - z^{-1})^{L_a}E(z), \tag{3.23}$$

where $E(z)$ is the quantization error of the internal quantizer, $U(z)$ is the normalized input signal ($u[k] = V_{\text{in}}[k]/V_{\text{ref}}$, $|u[k]| \leq 1$), L_a is the order of the analog modulator, and $k \leq L_a$ holds. Note that even though converters with $L_a > 2$ and one-bit internal quantizer are not stable, stability of converters based on Eq. (3.23) can always be guaranteed by

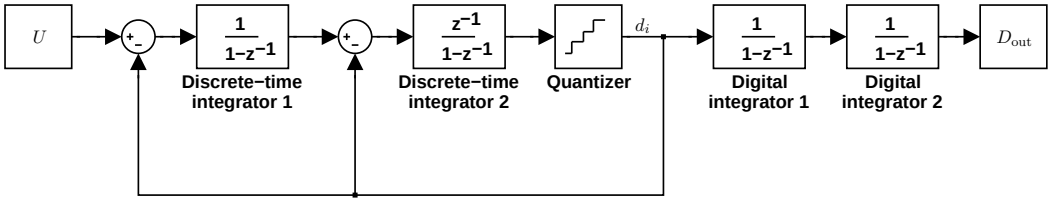


Figure 3.9: Second-order incremental converter with modulator realizing Eq. (3.24).

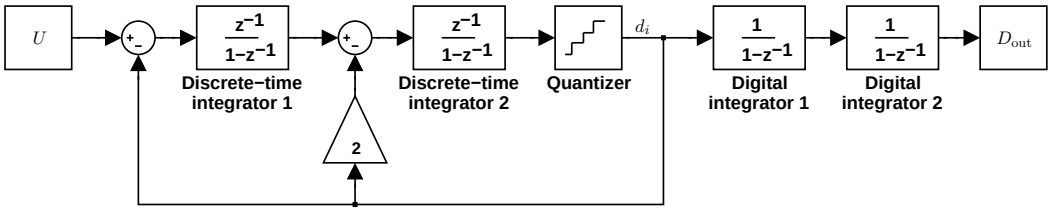


Figure 3.10: Second-order incremental converter with modulator realizing Eq. (3.25).

using multi-bit internal quantizer [Norsworthy et al., 1997, Chap. 4]. As non-ideality of multi-bit feedback DAC may cause severe degradation in the output, its linearity must be improved. For transient operation of $\Delta\Sigma$ modulators, linearization techniques are discussed in Sec. 4.3.7.

Two second-order $\Delta\Sigma$ modulator example realizing

$$Y(z) = z^{-1}U(z) + (1 - z^{-1})^2 E(z) \tag{3.24}$$

and

$$Y(z) = z^{-2}U(z) + (1 - z^{-1})^2 E(z) \tag{3.25}$$

are shown in Figs. 3.9 and 3.10, respectively. As these modulators are second-order ones, they are stable even with one-bit internal quantizer. These figures show also two digital integrators at the output, which are used to calculate the final digital output of the system.

Similarly to the first-order modulator, it can be readily seen that since the output of the modulator contains the L_ath-order derivative of the quantization noise of the internal quantizer, applying $L_d = L_a$th-order integration at the output results in a combined noise transfer function (NTF) of

$$NTF_{\mathrm{total,L_a}} = (1 - z^{-1})^{L_a} \frac{1}{(1 - z^{-1})^{L_a}} = 1, \tag{3.26}$$

while applying one more integrator at the output results in

$$NTF_{\mathrm{total,L_a+1}} = (1 - z^{-1})^{L_a} \frac{1}{(1 - z^{-1})^{L_a+1}} = \frac{1}{1 - z^{-1}}. \tag{3.27}$$

Thus, operating the second-order converter up to N cycles, assuming constant input

and input signal delay $k = 1$, the output of the second digital integrator will contain

$$D_{\text{out},2} = \frac{(N-1)(N)}{2}u + \varepsilon[N], \tag{3.28}$$

while that of the third integrator becomes

$$D_{\text{out},3} = \frac{(N-1)(N)(N+1)}{2 \cdot 3}u + \sum_{i=1}^{N} \varepsilon[i]. \tag{3.29}$$

As a higher-order $\Delta\Sigma$ loop may overload if the input signal is approaching the reference signal, it is usually required to limit the input signal to a fraction of the reference (typically $V_{\text{in}} = 0.8V_{\text{ref}}$, $0.75V_{\text{ref}}$, $0.66V_{\text{ref}}$ or $0.5V_{\text{ref}}$). With these constraints the quantization error of the internal converter may be limited to $\pm V_{\text{ref}}$. If the internal quantizer has l levels, the quantization error is bounded by $\pm V_{\text{ref}}/(l-1)$. Then, the normalized maximum error at the output of the second integrator is

$$\varepsilon_{\text{norm}} = \frac{\max(\varepsilon)}{(N-1)(N)/2} = \frac{2V_{\text{ref}}}{(l-1)(N-1)N}, \tag{3.30}$$

equals to half LSB of the target resolution.

Assuming that the input signal is limited to $V_{\max} < V_{\text{ref}}$, the ratio of the input range and the LSB voltage gives the number of levels in the converter:

$$2^{n_{\text{bit}}} = \frac{2V_{\max}}{2\varepsilon_{\text{norm}}} = \frac{0.5(l-1)(N-1)(N)V_{\max}}{V_{\text{ref}}} = 0.5U_{\max}(l-1)(N-1)N, \tag{3.31}$$

where $U_{\max} = V_{\max}/V_{\text{ref}} \in (0.5, 1)$ is the maximum normalized (dimension-less) input signal. Further refining the equation, the resolution in bits is

$$n_{\text{bit}} = \log_2(U_{\max}0.5N(N-1)(l-1)) \approx 2\log_2(N) + \log_2(l-1) + \log_2(U_{\max}) - 1, \tag{3.32}$$

and the required number of cycles for a given resolution

$$N \approx \frac{\sqrt{2} \cdot 2^{n_{\text{bit}}/2}}{\sqrt{U_{\max}(l-1)}}. \tag{3.33}$$

Thus, for $n_{\text{bit}} = 10$-bit resolution with an $l = 5$-level internal quantizer and a maximum input signal of $0.8V_{\text{ref}}$, $N = 26$(!) cycles are required if 2 integrators are used at the output. This is a great reduction compared to the first-order converter ($N = 1024$ with $l = 2$). For 16-bit resolution, the required number of cycles $N = 203$. With one-bit internal quantizer ($l = 2$) and maximum input signal limited to $U_{\max} = 0.5$, $N = 512$ required. This is also much better than $N = 65536$ with first-order converter.

Further reduction can be achieved if 3 digital integrators are used and the quantization error of the converter is uniformly distributed between $\pm V_{\text{ref}}/(l-1)$. Assuming that these

conditions are valid, the standard deviation of the internal quantizer is

$$\sigma_\varepsilon = \frac{2V_{\text{ref}}}{\sqrt{12}(l-1)}, \tag{3.34}$$

which is reduced by $\sqrt{N}$ in the output, resulting in a finale normalized quantization error of

$$q[N] = \frac{2 \cdot 3}{(N-1)(N)(N+1)} \sum_{i=1}^{N} \varepsilon[i], \tag{3.35}$$

with a standard deviation of

$$\sigma_q = \frac{2 \cdot 3}{(N-1)(N+1)(l-1)} \frac{1}{\sqrt{N}} \frac{2V_{\text{ref}}}{\sqrt{12}}. \tag{3.36}$$

Again, if the input signal is limited to $V_{\max}$, and a lower bound of the maximum output quantization error is estimated as $3\sigma_q$, (since the output error has an approximately Gaussian distribution due to the central limit theorem [Weisstein, 2004]), the number of levels in the converter are

$$2^{n_{\text{bit}}} \leq \frac{2V_{\max}}{2 \cdot 3\sigma_q} = \frac{V_{\max}}{3\sigma_q} = \frac{V_{\max}\sqrt{N}(N-1)(N+1)(l-1)}{6\sqrt{3}}. \tag{3.37}$$

From this equation, the resolution of the converter is

$$n_{\text{bit}} \leq \log_2 \frac{V_{\max}\sqrt{N}(N-1)(N+1)(l-1)}{6\sqrt{3}} \approx \frac{5}{2}\log_2 N + log_2(l-1) + \log_2 U_{\max} - 3.38, \tag{3.38}$$

while the required number of cycles for a given resolution is

$$N \geq \sqrt[5/2]{\frac{6\sqrt{3} \cdot 2^{n_{\text{bit}}}}{U_{\max}(l-1)}} \approx 2.55 \frac{2^{2n_{\text{bit}}/5}}{\sqrt[5/2]{U_{\max}(l-1)}}, \tag{3.39}$$

For 10-bit resolution, with $U_{\max} = 0.8$ and $l = 5$, $N = 26$ cycles are required, which is actually equal to that of the second-order case. However, for 16-bit resolution, N becomes 136, which is considerably less than 203 required with two integrators.

Note that the assumption about the internal A/D quantization error (i.e., uniformly distributed, uncorrelated with the input signal and white) is not valid in the modulator, especially with dc input signals. Thus, dithering is required to make this assumption valid. Applying dither signal, however, causes more severe overload on the internal quantizer, thus, the input signal amplitude range must be even more limited. This can be avoided by applying multi-bit quantizer in the loop. Although errors of the multi-bit feedback DAC may decrease the achievable *SNR* of the converter, Sec. 4.3.7 discusses dynamic methods to improve the linearity of the feedback DAC. The conclusion is that even though an L_ath-order modulator with pure differential *NTF* and $L_d = L_a + 1$st-order digital output filter may work theoretically, most of its benefits are lost due to practical problems.

The above results can be extended to arbitrary order modulators. Consider an L_ath-

order modulator defined by Eq. (3.23), followed by $L_d = L_a$ digital integrators. In this case, the output of the L_dth digital integrator in the converter in the Nth cycle will contain

$$D_{\text{out},l} = \binom{N + L_a - k - 1}{L_a} u + \varepsilon[N], \tag{3.40}$$

where k is the input signal delay, $k \in \{0, 1, \ldots, L_a\}$.

Assuming $k = 0$ (which can be achieved by the architecture discussed in the following section), an input signal limit of $U_{\max}$ and an l-level internal quantizer with uniformly distributed quantization error, the required number of cycles for a given resolution can be calculated from the normalized maximum error.

In the case, when $L_d = L_a$, i.e., the number of digital integrators is equal to that of the analog stages, the normalized maximum error is

$$q_e = \frac{V_{\text{ref}}}{l-1} \frac{1}{\binom{N+L_a-1}{L_a}}, \tag{3.41}$$

which equals to half LSB of the target resolution:

$$q_e = \frac{\text{LSB}}{2} = \frac{V_{\max}}{2^{n_{\text{bit}}}}. \tag{3.42}$$

Substituting Eq. (3.41) into (3.42) and rearranging Eq. (3.42) yields to

$$\prod_{i=0}^{L_a-1} (N + i) = \frac{2^{n_{\text{bit}}} L_a!}{(l-1)U_{\max}}, \tag{3.43}$$

from which the required number of cycles can be calculated. An initial guess may be easily calculated by approximating

$$\prod_{i=0}^{L_a-1} (N + i) \approx (N + L_a/2)^{L_a}, \tag{3.44}$$

from which

$$N_{\text{init}} = \sqrt[L_a]{\frac{2^{n_{\text{bit}}} L_a!}{(l-1)U_{\max}}} - L_a/2. \tag{3.45}$$

Consider now the case, when $L_d = L_a + 1$, i.e., there is one more integrator in the digital filter section, which averages the quantization noise. In this case, the $L_d = L_a + 1$st integrator in the Nth cycle will contain

$$D_{\text{out},l+1} = \binom{N + (L_a + 1) - k - 1}{(L_a + 1)} u + \sum_{i=1}^{N} \varepsilon[i], \tag{3.46}$$

where k is the delay of the input signal. Again, let us assume that $k = 0$.

Since here the output quantization error consists of the average of the internal quantization error, statistical properties of the error must be used to estimate the maximum output

error. If the internal quantization error is uniformly distributed between $\pm V_{\text{ref}}/(l-1)$, its standard deviation is

$$\sigma_\varepsilon = \frac{2V_{\text{ref}}}{(l-1)\sqrt{12}}. \tag{3.47}$$

The normalized output error is the sum of N quantization error samples, thus, its distribution is approximately Gaussian, with a standard deviation of

$$\sigma_q = \frac{1}{\binom{N+(L_a+1)-1}{(L_a+1)}} \frac{2\sqrt{N}V_{\text{ref}}}{(l-1)\sqrt{12}} \tag{3.48}$$

If a lower bound of the maximum output error is estimated by the 3-sigma rule, then $3\sigma_q$ is less than or equal to half LSB of the target resolution:

$$3\sigma_q \leq \frac{V_{\max}}{2^{n_{\text{bit}}}}. \tag{3.49}$$

Substituting and rearranging the above equations, the required number of cycles can be calculated from the following equation:

$$\sqrt{N}\prod_{i=1}^{L_a}(N+i) \geq \frac{\sqrt{3}(L_a+1)!\cdot 2^{n_{\text{bit}}}}{(l-1)U_{\max}} \tag{3.50}$$

By using the approximations

$$\sqrt{N} \approx \sqrt{N+L_a/2} \tag{3.51}$$

and

$$\prod_{i=1}^{L_a}(N+i) \approx (N+L_a/2)^{L_a}, \tag{3.52}$$

an initial value for the required number of cycles can be calculated as

$$N_{\text{init}} = \sqrt[L_a+1/2]{\frac{\sqrt{3}(L_a+1)!\cdot 2^{n_{\text{bit}}}}{(l-1)U_{\max}}} - L_a/2. \tag{3.53}$$

Note that similarly to the second-order modulator case with third-order digital filter, the internal quantization error must be uniformly distributed to make these derivations valid. Dither signal may be added to the internal quantizer to fulfill these requirements.

Architectural Considerations

There are many existing $\Delta\Sigma$ structures which may realize the signal and noise transfer functions of Eq. (3.23) [Norsworthy et al., 1997, Sec. 1.2.3, 5.5], [Schreier, 1993], [Schreier, 2004]. However, there is one structure, which is of particular interest. This is usually referred as Cascaded-Integrators, Feed-forward (CIFF) architecture. Its loop contains cascaded integrators, and their output is fed forward right to the input of the quantizer. The structure becomes even more interesting if the input signal ($u[k]$) is also fed forward to the input of

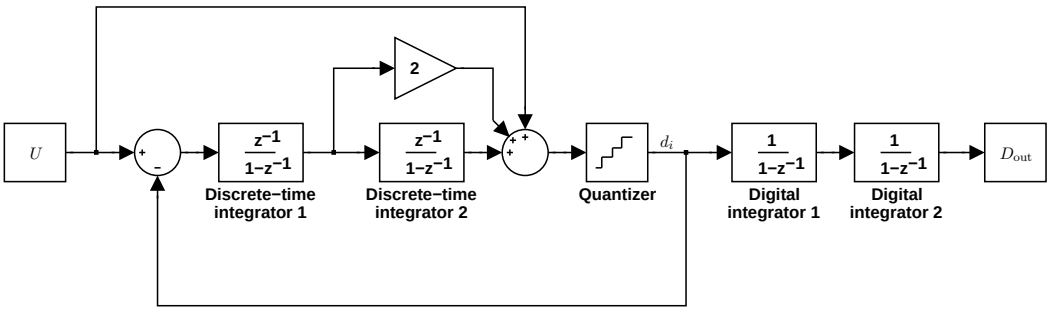


Figure 3.11: Second-order CIFF incremental converter with the input signal fed forward to the input of the quantizer.

the quantizer [Silva et al., 2001; Silva, 2004]. Fig. 3.11 shows a second-order example of the feed-forward structure with the input signal fed forward to the input of the quantizer. Note that this structure can be derived from the distributed feedback structure shown in Fig. 3.10 by using the well-known inversion technique for linear systems.

The main advantage of this structure is that since the input signal is fed forward to the input of the quantizer, the STF of the modulator becomes

$$STF(z) = \frac{H(z)}{1 + H(z)} + \frac{1}{1 + H(z)} = 1, \tag{3.54}$$

where $H(z)$ is the loop filter transfer function. Thus, the output of the converter is

$$Y(z) = U(z) + (1 - z^{-1})^{L_a} E(z). \tag{3.55}$$

Then, the signal at the input of the first integrator,

$$Y(z) - U(z) = (1 - z^{-1})^{l} E(z), \tag{3.56}$$

i.e., only quantization error is processed by the integrators. This has several additional advantages: less sensitivity to the non-linearity of the integrators, less voltage swing at the output of any integrators, only one feedback DAC has to be realized, etc. [Silva et al., 2001; Silva et al., 2004; Silva, 2004]. In addition, as the difference of the input and the feedback signal is processed by L_a integrators, at the output of the last integrator the (delayed) quantization error presents:

$$\left(\frac{z^{-1}}{1 - z^{-1}}\right)^{L_a} (Y(z) - U(z)) = -\left(\frac{z^{-1}}{1 - z^{-1}}\right)^{L_a} (1 - z^{-1})^{L_a} E(z) = -z^{-L_a} E(z). \tag{3.57}$$

This means that *at time step k, the quantization error $\varepsilon[k - L_a]$ is available at the output of the last integrator, if the modulator realizes Eq.* (3.55). Thus, for the incremental converter discussed above, at the end of the conversion (using L_a additional cycles) the quantization error $q[N] = \varepsilon[N]$ is available in analog form if L_a digital integrator are used to calculate the output, similarly to the first-order case. This means that fine quantization is possible

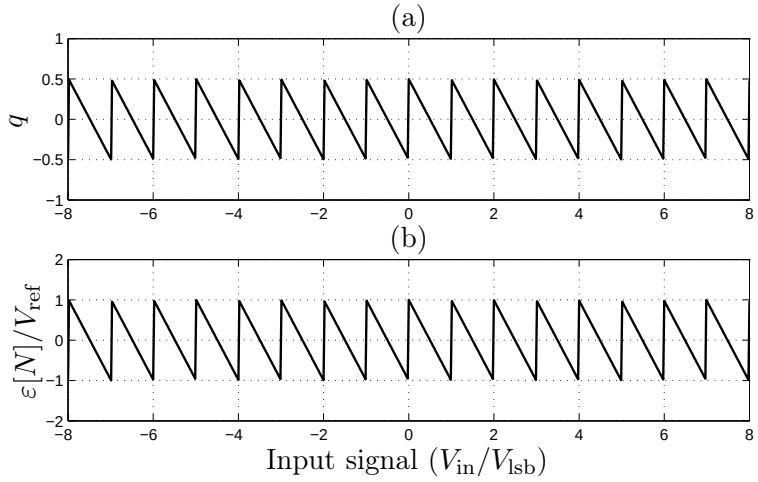


Figure 3.12: (a) Quantization error of the converter (q) and (b) quantization error of the internal quantizer in time step N ($\varepsilon[N]$) around zero in a second-order converter with second-order digital filter.

by using this residual signal.

Simulation Results

Simulations in MATLAB and Simulink verify the theoretical results discussed above. Fig. 3.12(a) shows the quantization error of a 16-bit converter operated through $N = 513$ cycles, while Fig. 3.12(b) shows the internal converter's error $\varepsilon[513]$. The two errors are the same for any input, in agreement with Eq. (3.28).

Fig. 3.13 shows the quantization error of a second-order converter when three digital integrator are used to calculate output ($U_{\max} = 0.8$, $l = 5$, $N = 136$, $n_{\mathrm{bit}} = 16$ bits). Fig. 3.13(a) shows the case when no internal dither signal is used for the conversion. Note that similar peak error exist around zero input to the one discussed in the previous section, except that it is smaller due to the multi-bit internal quantizer. Applying dither signal (with a maximum of half LSB of the internal 5-level ADC), these peaks disappear, and the practical resolution matches well with the theoretical one (Fig. 3.13(b)).

Finally, Fig. 3.14 shows simulation results of the feed-forward converter shown in Fig. 3.11. As predicted by the theory, $2V_{\mathrm{ref}}q = \varepsilon[N] = -V_2[N+2]$ holds for any input signal.

Although this type of extension of the first-order incremental converter has a clear theoretical background, the introduced method has also some limitations. One is that one has to use multi-bit quantizer and thus multi-bit feedback DAC in a converter realizing the transfer functions of Eq. (3.23). Another drawback is that only simulation can show how large are the limits of the quantization error of the internal A/D converter. As the achievable resolution relies on the maximum internal quantizer error, this must always be verified. In addition, the L_ath-order modulator with $L_d = L_a +$ 1st-order digital filter needs dither signal to avoid unwanted peak around zero, while higher-order modulators

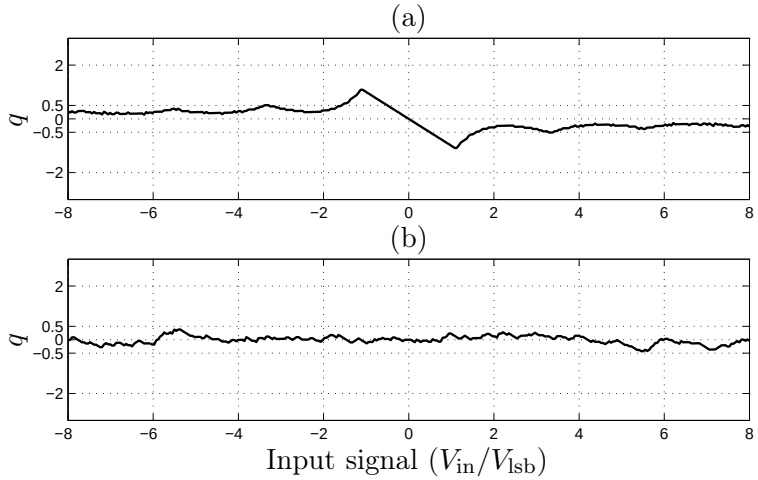


Figure 3.13: (a) Quantization error of the converter (q) without dither signal and (b) quantization error with dither signal in a second-order converter with third-order digital filter, around zero input signal.

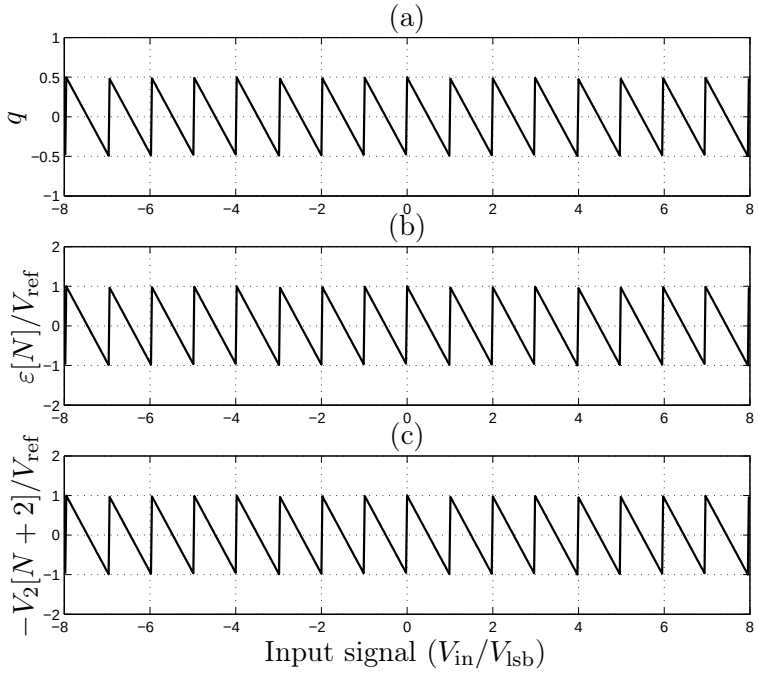


Figure 3.14: (a) Quantization error of the converter (q) (b) quantization error of the internal quantizer in time step N ($\varepsilon[N]$) (c) inverted output of the second integrator at time step $N+2$ ($-V_2[N+2]$) in the converter of Fig. 3.11.

are more sensitive to the magnitude of the dither signal, nevertheless, the maximum input signal must also be reduced to avoid overload errors in the loop.

In the following two subsections, two different extensions are discussed, which are less sensitive to the overload of the internal quantization error and does not rely on pure differential *NTF* i.e., Eq. (3.23) does not need to be satisfied.

3.2.2 Matched Digital Filters

Another possible extension of the first-order incremental converter is to extend its original idea: i.e., using the output of the (last) integrator to determine the required number of cycles for a given resolution. The main advantage of this idea is that in a real circuit the output of any integrator must be limited to a realistic maximum signal (e.g., $\pm V_{\text{ref}}$) to avoid the saturation of the op-amp and the whole integrator. This way one has a cycle- and input-independent value to limit the error of the conversion.

Consider the modulator structure of Fig. 3.10 on p. 33. The output of the first discrete-time integrator during the first N cycles can be readily calculated (assuming that before the conversion, all integrators are reset):

$$\begin{aligned}
V_1[0] &= 0 \\
V_1[1] &= 0 + V_{\text{in}}[0] - d_0 V_{\text{ref}} \\
V_1[2] &= V_1[1] + V_{\text{in}}[1] - d_1 V_{\text{ref}} = (V_{\text{in}}[0] + V_{\text{in}}[1]) - (d_0 V_{\text{ref}} + d_1 V_{\text{ref}}) \\
V_1[3] &= \sum_{i=0}^{2} V_{\text{in}}[i] - \sum_{i=0}^{2} d_i V_{\text{ref}} \\
&\vdots \\
V_1[N-1] &= \sum_{i=0}^{N-2} V_{\text{in}}[i] - \sum_{i=0}^{N-2} d_i V_{\text{ref}} \\
V_1[N] &= \sum_{i=0}^{N-1} V_{\text{in}}[i] - \sum_{i=0}^{N-1} d_i V_{\text{ref}}
\end{aligned} \tag{3.58}$$

The output of the second integrator is

$$\begin{aligned}
V_2[0] &= 0 \\
V_2[1] &= 0 + V_1[0] - 2d_0 V_{\text{ref}} = -2d_0 V_{\text{ref}} \\
V_2[2] &= V_2[1] + V_1[1] - 2d_1 V_{\text{ref}} = V_{\text{in}}[0] - d_0 V_{\text{ref}} - 2d_0 V_{\text{ref}} - 2d_1 V_{\text{ref}} \\
V_2[3] &= V_2[2] + V_1[2] - 2d_2 V_{\text{ref}} = \\
&= V_{\text{in}}[0] - d_0 V_{\text{ref}} - 2(d_0 + d_1) V_{\text{ref}} + \\
&\quad + (V_{\text{in}}[0] + V_{\text{in}}[1]) - (d_0 V_{\text{ref}} + d_1 V_{\text{ref}}) - 2d_2 V_{\text{ref}} = \\
&= 2V_{\text{in}}[0] + V_{\text{in}}[1] - (2d_0 V_{\text{ref}} + d_1 V_{\text{ref}}) - 2(d_0 + d_1 + d_2) V_{\text{ref}} = \\
&= \sum_{i=0}^{1} \sum_{j=0}^{i} V_{\text{in}}[j] - \sum_{i=0}^{1} \sum_{j=0}^{i} d_j V_{\text{ref}} - 2 \sum_{i=0}^{2} d_i V_{\text{ref}}
\end{aligned}$$

$$\vdots$$

$$V_2[N] = \sum_{i=0}^{N-2}\sum_{j=0}^{i} V_{\text{in}}[j] - \sum_{i=0}^{N-2}\sum_{j=0}^{i} d_j V_{\text{ref}} - 2\sum_{j=0}^{N-1} d_j V_{\text{ref}}. \tag{3.59}$$

Note that the same result can be achieved by using z-domain analysis. In this case, the output of the second integrator becomes

$$V_2(z) = \frac{z^{-2}}{(1-z^{-1})^2}(U(z) - Y(z)) - 2\frac{z^{-1}}{1-z^{-1}}Y(z), \tag{3.60}$$

which leads to the same time-domain equations assuming the same initial conditions ($V_{\text{in}}[i] = d_i = V_1[i] = V_2[i] = 0$, if $i < 0$).

if $V_2[N]$ is limited (which must be true, otherwise the converter is not stable and the integrators saturate), then the difference of the double-sum input and output is also limited. Assuming that $|V_2[N]| < 4V_{\text{ref}}$ and assuming constant input, rearranging Eq. (3.59) leads to

$$\left| \frac{(N-2)(N-1)}{2} V_{\text{in}} - V_{\text{ref}} \left(\sum_{i=0}^{N-2}\sum_{j=0}^{i} d_j - 2\sum_{j=0}^{N-1} d_j \right) \right| < 4V_{\text{ref}}, \tag{3.61}$$

i.e.,

$$\left| V_{\text{in}} - V_{\text{ref}} \frac{2}{(N-2)(N-1)} \left(\sum_{i=0}^{N-2}\sum_{j=0}^{i} d_j - 2\sum_{j=0}^{N-1} d_j \right) \right| < \frac{8}{(N-2)(N-1)} V_{\text{ref}}. \tag{3.62}$$

Thus, an estimate of the input signal can be found by realizing

$$D_{\text{out}} = \frac{2}{(N-2)(N-1)} \left(\sum_{i=0}^{N-2}\sum_{j=0}^{i} d_j - 2\sum_{j=0}^{N-1} d_j \right), \tag{3.63}$$

the expression in the right-hand side of the previous equation. Note that this can be realized by using a digital filter, which is the *exact replica* of the analog filter processing the feedback signal. In this case, the digital filter with a transfer function

$$D(z) = \frac{z^{-2}}{(1-z^{-1})^2} - 2\frac{z^{-1}}{1-z^{-1}}, \tag{3.64}$$

starting from zero initial conditions and operated through N cycles can calculate Eq. (3.63), providing an estimate of the input signal.

This idea (using digital filter exactly matching the analog filter which processes the feedback signal) has been published and also patented by Lyden [Lyden, 1993; Lyden et al., 1995], thus, this model is not analyzed here, the reader is referred to Lyden's works. However, two limiting drawbacks of this method is discussed in the following, and the next subsection proposes an incremental converter which does not suffer from these problems,

even though it is based on the same criteria.

One drawback of the introduced method is that it requires exact matching between the analog and digital filters. Due to the inherent mismatch between the two paths, the processing will not be exactly the same, causing pole and zero error between the two signal path, which results in a gain error at dc. To handle this error, the converter must be operated for more number of cycles (to ensure that the error is below the specification), and the gain error must be compensated by a two-point calibration of the converter.

The second problem of the converter is that the quantization error of the finale output is strongly correlated with the input signal. Consider again the converter of Fig. 3.10, realizing the following transfer functions:

$$Y(z) = z^{-2}U(z) + (1 - z^{-1})^2 E(z). \tag{3.65}$$

In the time-domain, the output signal at time step k can be calculated as

$$y[k] = u[k-2] + \varepsilon[k] - 2\varepsilon[k-1] + \varepsilon[k-2]. \tag{3.66}$$

As in the structure Fig. 3.10 the last integrator's output is followed by only the addition of the quantization noise (by the quantizer), the output of the last integrator is

$$V_2[k] = y[k] - \varepsilon[k] = u[k-2] - 2\varepsilon[k-1] + \varepsilon[k-2]. \tag{3.67}$$

Thus, the output of the last integrator, which is used to limit the quantization error of the conversion, contains the input signal. As a consequence, the final quantization error will also contain u, thus, even if the internal quantization error is a more or less noise-like signal, independent of the input signal (which can be achieved by means of dithering), the final conversion error will be strongly input-dependent. This explains also the large signal swing at the output of the second (last) integrator.

In the following subsection another higher-order structure is described, which does not suffer from these limiting error sources.

3.2.3 Using Cascaded-Integrators, Feed-Forward (CIFF) Structure

Consider a third-order cascaded-integrators, feed-forward (CIFF) structure [Schreier, 1993], [Schreier, 2004], [Norsworthy et al., 1997, Sec. 1.2.3, 5.5]. A third-order example of such a structure is shown in Fig. 3.15. In this example, an input signal path, connecting to the input of the internal quantizer is also shown, for benefits to be discussed later in this subsection. Note that in this structure, the a_i coefficients are used to control the pole-zero map of the NTF, and $b = c_i = 1$ initially. These latter coefficients are used to scale the integrators' maximum output swing (note that a_i must change according to the scaling coefficients, e.g., if $b' \neq 1$, then $a'_i = a_i/b'$ to keep the loop transfer function unchanged.

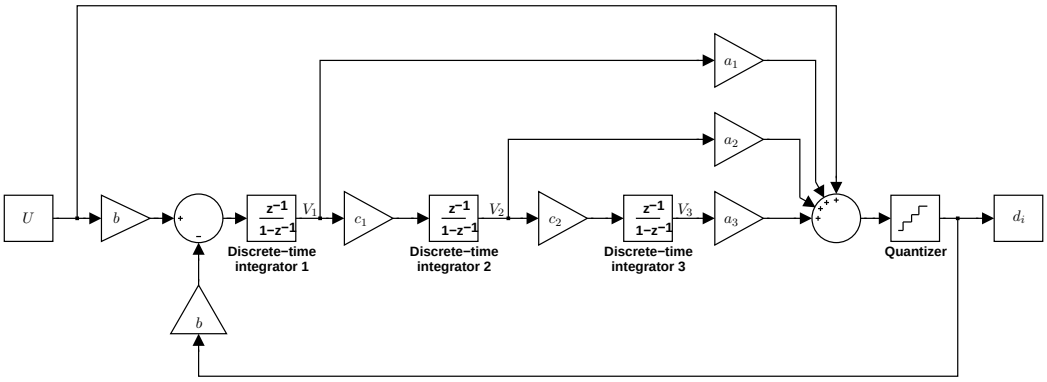


Figure 3.15: A third-order *Cascaded Integrator, Feed-Forward (CIFF)* architecture with the input signal fed forward to the input of the quantizer.

Basic Operation

The operation will be discussed in terms of this third-order structure, and will be generalized later. As in any of the cases discussed above, all memory elements, both analog and digital, must be reset at the beginning of each conversion cycle. Then, V_{in} is applied to the input of the first integrator. Using the notations of Fig. 3.15, the output signals of all integrators can readily be found in the time domain after the first N clock cycles.

The first integrator's output samples are given by

$$\begin{aligned} V_1[0] &= 0 \\ V_1[1] &= b(V_{\text{in}}[0] - d_0 V_{\text{ref}}) \\ V_1[2] &= V_1[1] + b(V_{\text{in}}[0] - d_1 V_{\text{ref}}) = \\ &\quad b(V_{\text{in}}[0] + V_{\text{in}}[1] - d_0 V_{\text{ref}} - d_1 V_{\text{ref}}) \\ &\vdots \\ V_1[N] &= b \sum_{k=0}^{N-1} (V_{\text{in}}[k] - d_k V_{\text{ref}}) , \end{aligned} \tag{3.68}$$

where $d_k = \pm 1$ is the comparator output in the kth cycle.

Similarly, the sequence of the outputs of the second integrator is

$$\begin{aligned} V_2[0] &= 0 \\ V_2[1] &= c_1 V_1[0] + V_2[0] = 0 \\ V_2[2] &= c_1 V_1[1] + V_2[1] = c_1(V_1[1] + V_1[0]) \\ &\vdots \\ V_2[N] &= c_1 \sum_{l=0}^{N-1} V_1[l] = c_1 b \sum_{l=0}^{N-1} \sum_{k=0}^{l-1} (V_{\text{in}}[k] - d_k V_{\text{ref}}) , \end{aligned} \tag{3.69}$$

and that of the third is

$$\begin{aligned} V_3[0] &= 0 \\ V_3[1] &= c_2 V_2[0] + V_3[0] = c_2 V_2[0] = 0 \\ V_3[2] &= c_2 V_2[1] + V_3[1] = c_2 (V_2[1] + V_2[0]) = 0 \\ &\vdots \\ V_3[N] &= c_2 \sum_{m=0}^{N-1} V_2[m] = \\ &= c_2 c_1 b \sum_{m=0}^{N-1} \sum_{l=0}^{m-1} \sum_{k=0}^{l-1} \left(V_{\mathrm{in}}[k] - d_k V_{\mathrm{ref}} \right). \end{aligned} \tag{3.70}$$

Let us assume that the input signal is constant and that the loop is stable for dc input (the case of non-constant input will be discussed in Sec. 4.1, while stable loop design will be addressed in Sec. 3.2.3). In this case,

$$\sum_{m=0}^{N-1} \sum_{l=0}^{m-1} \sum_{k=0}^{l-1} V_{\mathrm{in}}[k] = \frac{N(N-1)(N-2)}{3!} V_{\mathrm{in}}, \tag{3.71}$$

thus,

$$V_3[N] = c_2 c_1 b \left(\frac{N(N-1)(N-2)}{3!} V_{\mathrm{in}} - \sum_{m=0}^{N-1} \sum_{l=0}^{m-1} \sum_{k=0}^{l-1} d_k V_{\mathrm{ref}} \right). \tag{3.72}$$

Rearranging Eq. (3.72),

$$\sum_{m=0}^{N-1} \sum_{l=0}^{m-1} \sum_{k=0}^{l-1} d_k V_{\mathrm{ref}} = \frac{N(N-1)(N-2)}{3 \cdot 2} V_{\mathrm{in}} - \frac{V_3[N]}{c_2 c_1 b} \tag{3.73}$$

The advantage of the feed-forward architecture is that using the scaling coefficients b and c_i, it can always be assured that $|V_3[N]| \leq V_{\mathrm{ref}}$ for any input signal $|V_{\mathrm{in}}| \leq V_{\mathrm{max}}$, where $V_{\mathrm{max}} \in [0.5, 1) V_{\mathrm{ref}}$. The input signal must be limited to a fraction of the feedback reference signal, to ensure the stable operation of the loop. The concrete value of $U_{\mathrm{max}} = V_{\mathrm{max}}/V_{\mathrm{ref}}$ depends on the loop order, for second-order loop $U_{\mathrm{max}} = 0.75$ or 0.667, for third-order loop $U_{\mathrm{max}} = 0.667$ or $U_{\mathrm{max}} = 0.5$ may be preferable. A method to scale down the input signal accurately to the fraction of the feedback signal is discussed in Sec. 4.3.2.

With proper scaling coefficients, $|V_3[N]| \leq V_{\mathrm{ref}}$ is assured. Then, further rearranging Eq. (3.73), and substituting V_{ref}, one can get the following bounds on the difference of the unknown input signal V_{in} and the known terms d_i and N:

$$-\frac{3!}{N(N-1)(N-2)}\frac{1}{c_2c_1b}V_{\text{ref}} \leq V_{\text{in}} - \frac{3!}{N(N-1)(N-2)}V_{\text{ref}}\sum_{m=0}^{N-1}\sum_{l=0}^{m-1}\sum_{k=0}^{l-1}d_k \leq +\frac{3!}{N(N-1)(N-2)}\frac{1}{c_2c_1b}V_{\text{ref}}. \quad (3.74)$$

Thus, after N clock periods, an estimate of $V_{\text{in}}/V_{\text{ref}}$ can be found as

$$D_{\text{out}} = \frac{\hat{V}_{\text{in}}}{V_{\text{ref}}} = \frac{3!}{(N-2)(N-1)N}\sum_{m=0}^{N-1}\sum_{l=0}^{m-1}\sum_{k=0}^{l-1}d_k, \quad (3.75)$$

which means that the input estimate can be calculated from the modulated samples in the digital domain without knowing the exact value of any analog coefficient. This is a great advantage compared to the structure discussed in the previous subsection.

Eq. (3.75) can be realized by three delaying digital integrator. Note that three cycles may be saved during the operation by using non-delaying digital integrators, since

$$\left(\frac{z^{-1}}{1-z^{-1}}\right)^3 = \frac{1}{(1-z^{-1})^3}z^{-3}, \quad (3.76)$$

and the last three delays can be neglected. In the time domain, the output calculation becomes the following:

$$\frac{\hat{V}_{\text{in}}}{V_{\text{ref}}} = \frac{3!}{(N-2)(N-1)N}\sum_{m=0}^{N-3}\sum_{l=0}^{m}\sum_{k=0}^{l}d_k. \quad (3.77)$$

Even though this means that the analog $\Delta\Sigma$ modulator may also be operated through only $N-L_a$ cycles (where $L_a=3$ is the order of modulator), it might be desired to operate it through N cycles due to a property discussed in the following.

Recalling Eq. (3.74), in an A/D converter these lower and upper limits are equal to $\pm V_{\text{lsb}}/2$. From these limits one can find the equivalent value of the LSB voltage as

$$V_{\text{lsb}} = \frac{2\cdot 3!}{(N-2)(N-1)N}\frac{1}{c_2c_1b}V_{\text{ref}}. \quad (3.78)$$

The relative quantization error (in LSBs) can also be found. It is given by

$$q = \frac{\hat{V}_{\text{in}} - V_{\text{in}}}{V_{\text{lsb}}} = \frac{1}{2}c_2c_1b\sum_{m=0}^{N-1}\sum_{l=0}^{m-1}\sum_{k=0}^{l-1}d_k - \frac{1}{2}c_2c_1b\frac{N(N-1)(N-2)}{3!}\frac{V_{\text{in}}}{V_{\text{ref}}}. \quad (3.79)$$

Hence, from Eq. (3.70), it can be readily seen that

$$V_3[N] = -2V_{\text{ref}}q, \quad (3.80)$$

which means that if the output is calculated by Eq. (3.75) or Eq. (3.77), the quantization error can be found in analog form at the output of the last integrator in cycle N, assuming that the digital output is calculated precisely, i.e., it is not requantized to the target resolution of the converter. This property is similar to the first-order case. This signal may be used to further refine the quantization error by using an auxiliary A/D converter. However, this requires the converter to be operated through N cycles, instead of $N - L_a$. In the simplest case, one might determine the sign of the output of the last integrator, to pick up an extra bit precision without any additional hardware, except some logic.

Note that in an ideal A/D converter, the following equation can be used to define the digital output (using integer arithmetic) and the LSB voltage:

$$|D_{\text{out}} V_{\text{lsb}} - V_{\text{in}}| \leq \frac{V_{\text{lsb}}}{2}. \tag{3.81}$$

Applying this equation to the incremental converter would give

$$V'_{\text{lsb}} = V_{\text{lsb}} = \frac{2 \cdot 3!}{(N-2)(N-1)N} \frac{1}{c_2 c_1 b} V_{\text{ref}}, \tag{3.82}$$

while

$$D'_{\text{out}} = \frac{D_{\text{out}}}{V'_{\text{lsb}}} = \frac{c_2 c_1 b}{2} \sum_{m=0}^{N-3} \sum_{l=0}^{m} \sum_{k=0}^{l} d_k, \tag{3.83}$$

which indicate that the knowledge of the exact value of the scaling coefficients is required to calculate the output code. In reality, this is not required, as the input estimate ($V_{\text{lsb}} D_{\text{out}}$) does not contain these scale factors, which are only a gain factor in the output. Note that Eq. (3.80) holds also in this case.

Properties of the Quantization Error

Recalling Eq. (3.80), one can see that the quantization error of the converter is linearly related to $V_3[N]$, the output of the last integrator in cycle N. Thus, to analyze the quantization error, it is enough to examine $V_3[N]$.

Consider first the case, when the input signal is not fed-forward to the input of the quantizer. This solution has several disadvantages, some of them already known from previous sections.

One serious disadvantage is that $V_3[N]$, and so the quantization error, has an input-related term. To verify this, the transfer functions from the input of the modulator to the output of the last integrator has to be evaluated. To find this, first the *NTF* and the *STF* is worth to be calculated. The loop filter of the converter can be easily found (assuming $b = c_i = 1$):

$$H(z) = \sum_{i=1}^{3} \frac{a_i z^{-i}}{(1 - z^{-1})^i} = \frac{\sum_{i=1}^{3} a_i z^{-i} (1 - z^{-1})^{3-i}}{(1 - z^{-1})^3}. \tag{3.84}$$

The NTF of the loop (using the notation $a_0 = 1$) becomes

$$NTF = \frac{1}{1+H(z)} = \frac{(1-z^{-1})^3}{\sum_{i=0}^{3} a_i z^{-i}(1-z^{-1})^{3-i}} = \frac{(1-z^{-1})^3}{D(z^{-1})}, \tag{3.85}$$

while the STF is

$$STF = \frac{H(z)}{1+H(z)} = \frac{\sum_{i=1}^{3} a_i z^{-i}(1-z^{-1})^{3-i}}{\sum_{i=0}^{3} a_i z^{-i}(1-z^{-1})^{3-i}} = \frac{N(z^{-1})}{D(z^{-1})}. \tag{3.86}$$

Usually $D(z^{-1})$ realizes a Butterworth pole configuration to flatten the NTF at high-frequency, so as to ensure the stability of the loop [Schreier, 1993].

From the architecture, it can be readily seen that the transfer function from the input signal to the output of the third integrator is similar to that of the quantization noise, except that this latter one is multiplied by minus one. This latter transfer function can be easily calculated since

$$\frac{V_3(z)}{E(z)} = -\left(\frac{z^{-1}}{1-z^{-1}}\right)^3 NTF, \tag{3.87}$$

thus

$$V_3(z) = \frac{z^{-3}}{D(z^{-1})}(U(z)-E(z)) = \frac{z^{-3}}{\sum_{i=0}^{3} a_i z^{-i}(1-z^{-1})^{3-i}}(U(z)-E(z)) \tag{3.88}$$

If the input signal is dc, it can be readily shown that

$$\left.\frac{V_3(z)}{U(z)}\right|_{z=1} = \frac{1}{a_3}. \tag{3.89}$$

This means that $V_3[N]$ contains the low-pass filtered input signal (including the dc-component) and the low-pass filtered internal quantization error. It follows then that the final quantization error also has strong input-dependence. This also means that the output of the integrators has a large signal swing when the input signal becomes large.

Another problem of the architecture is that the STF of the modulator contains one delay. Thus, the first decision of the comparator has (ideally) a random nature or is based on the offset of the integrators. In addition, in incremental mode, this sample has the highest weight, during calculation of the output (cf. Eq. (3.75)). Even though the loop will later compensate for a possible wrong decision, a better approach is to ensure that even the first sample be based on the sign of the input signal.

As discussed also in Sec. 3.2.1, most of these problems can be avoided by feeding the input signal forward to the input of the quantizer [Silva et al., 2001]. This makes the STF of the converter equal to one, independent of the loop filter. In this case, as $Y(z) = U(z) + NTF(z)E(z)$, the integrators in the loop process $U(z) - Y(z) = -NTF(z)E(z)$,

thus, the output of the third integrator becomes

$$V_l(z) = -\frac{z^{-3}}{D(z^{-1})}E(z), \tag{3.90}$$

a low-pass filtered version of the internal quantization error.

Note that if the NTF of the converter is a pure L_ath-order differentiator, i.e., $NTF = (1 - z^{-1})^{L_a}$, then

$$V_{L_a}(z) = z^{-L_a}E(z), \tag{3.91}$$

i.e.,

$$V_{L_a}[N] = \varepsilon[N - L_a]. \tag{3.92}$$

In this special case (using feed-forward architecture, with feed-forward input signal and pure L_ath-order differential NTF), the two different extensions to higher-order incremental converters discussed in this section and in Sec. 3.2.1, are equivalent.

Note that feeding the input signal forward to the quantizer has many other advantages: even the first feedback signal will contain the input signal; as the integrators are not processing the input signal, the distortion of the integrators are not affecting the input signal; there is less signal swing at the output of the integrators; etc. [Silva et al., 2001; Silva, 2004; Silva et al., 2004].

Scaling of the Coefficients

To validate the derivation discussed above, the scale factors b and c_i must be properly set to ensure that the output of the last integrator does not exceed the reference signal, i.e., $|V_3| < V_{\text{ref}}$ must be held. To set these coefficients, one has to find first the maximum dc input signal for which the modulator remains stable. The stability of the $\Delta\Sigma$ modulator for order greater than two is still an open question, as rigorous analytical results give usually too conservative estimates on the properties of the modulator, if exist at all. An approach based on finding positive invariant set in the state-space for low-order ($L_a \leq 3$) converters was discussed in [Wang, 1992; Schreier et al., 1995; Schreier et al., 1997], and it is also summarized in [Norsworthy et al., 1997, Chap. 4]. Recently, for higher-order modulators with distinct zeros in the NTF, a transformation method was suggested, which transforms the modulator into first- and second-order modulator sections with one common quantizer [Wong and Ng, 2003]. With this transformation, stable dc input bounds can be found for these modulators.

In the case discussed here, the converter is third-order with multiple zeros at dc, which makes its analysis more difficult. Nevertheless, as the converter has intermittent operation and the state variables are reset at the beginning of each conversion, the stabilization of the converter is not too critical. To find the maximum stable dc input bound and the required scaling coefficients, simulation tools can be used [Schreier, 1993; Schreier, 2004] to ensure that the converter is stable.

Consider again the third-order CIFF modulator with Butterworth high-pass NTF filter configuration, with a maximum gain of $H(z)|_{z=-1} = 1.5$. The NTF of such a system

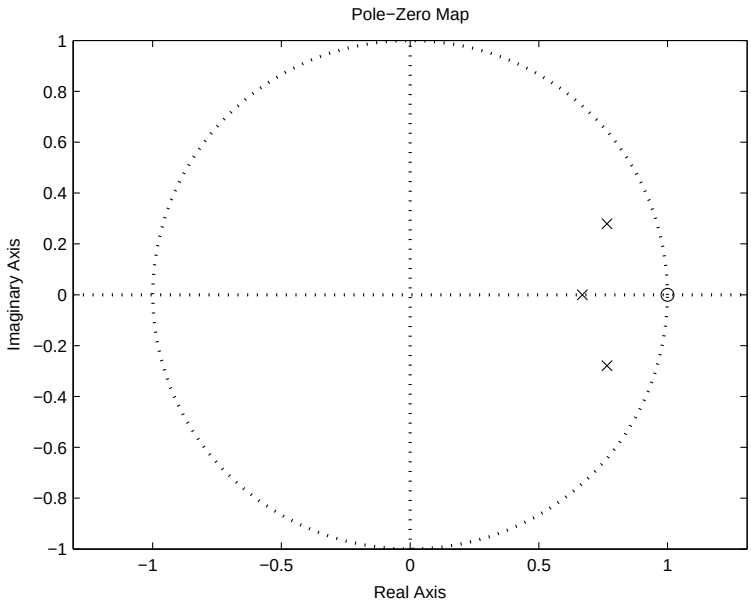


Figure 3.16: Pole-zero map of a third-order $\Delta\Sigma$ modulator with Butterworth low-pass stabilization.

Table 3.2: Unscaled coefficients of the third-order modulator.

a_1	0.7997
a_2	0.2881
a_3	0.0440
b	1
c_1	1
c_2	1

becomes

$$NTF = \frac{(z-1)^3}{(z-0.6694)(z^2-1.531z+0.6639)}, \tag{3.93}$$

resulting in a pole-zero map shown in Fig. 3.16, and a noise transfer function shown in Fig. 3.17. The *NTF* has been designed using the MATLAB toolbox of [Schreier, 2004].

Mapping the coefficients of the modulator (Fig. 3.15) to the coefficients of the *NTF* based on Eqs. (3.85) and (3.93) results in the coefficients listed in Tab. 3.2.

In Sec. 3.2.3, it was shown that the transfer function from the input signal to the output of the third integrator at dc is $1/a_3$, if it is not fed-forward to the input of the quantizer, and it is unconditionally true for that of the internal quantization noise. In this particular case, $1/a_3 = 22.7$, which means that the architecture must be scaled down to prevent the overflow of the integrators. To further justify this, Fig. 3.18(a)–(c) shows the transfer functions from the quantization error to the first, second and third integrator, respectively. It can be seen that large gains exist at different frequencies, so scaling is necessary.

Note that even knowing the transfer functions, the required scaling cannot be calculated, since there is no information about the spectral behavior of the quantization error

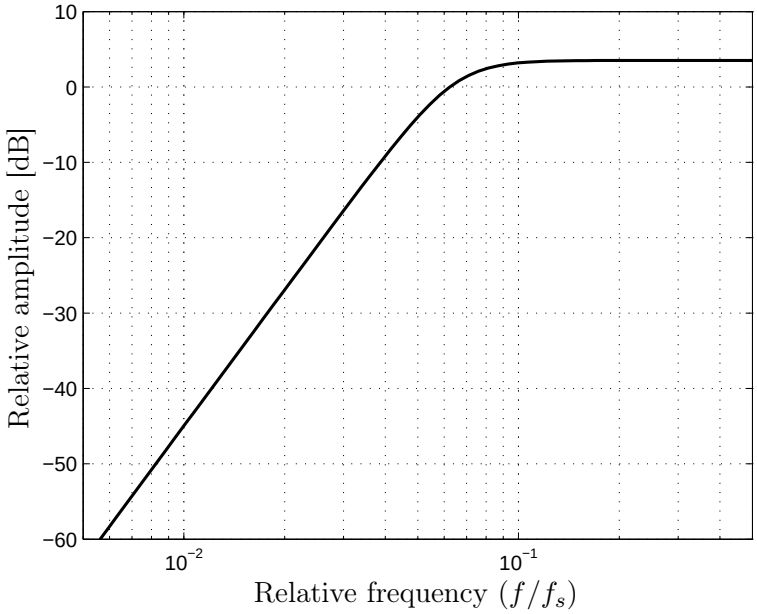


Figure 3.17: Transfer characteristics of a third-order $\Delta\Sigma$ modulator with Butterworth low-pass stabilization.

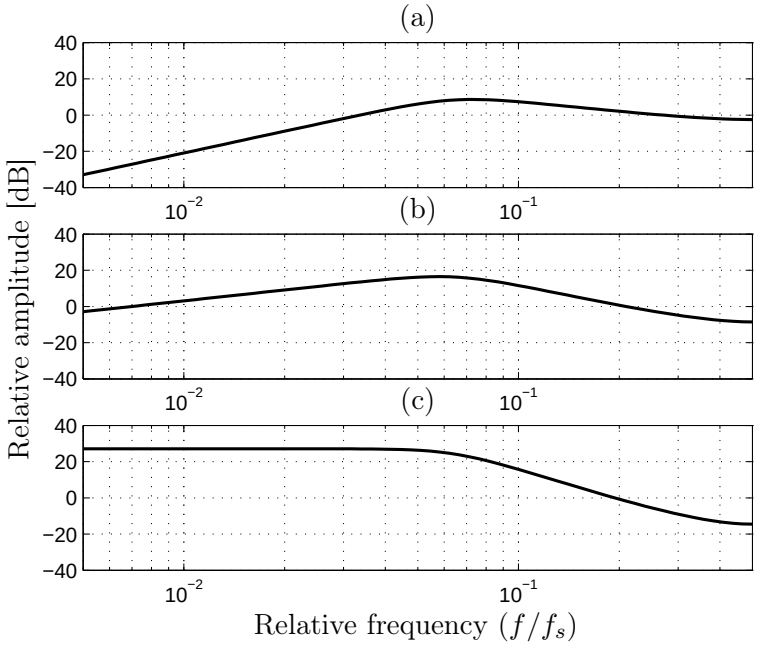


Figure 3.18: Transfer characteristics from the quantization error to the output of the (a) first, (b) second, and (c) third integrator in a third-order $\Delta\Sigma$ modulator.

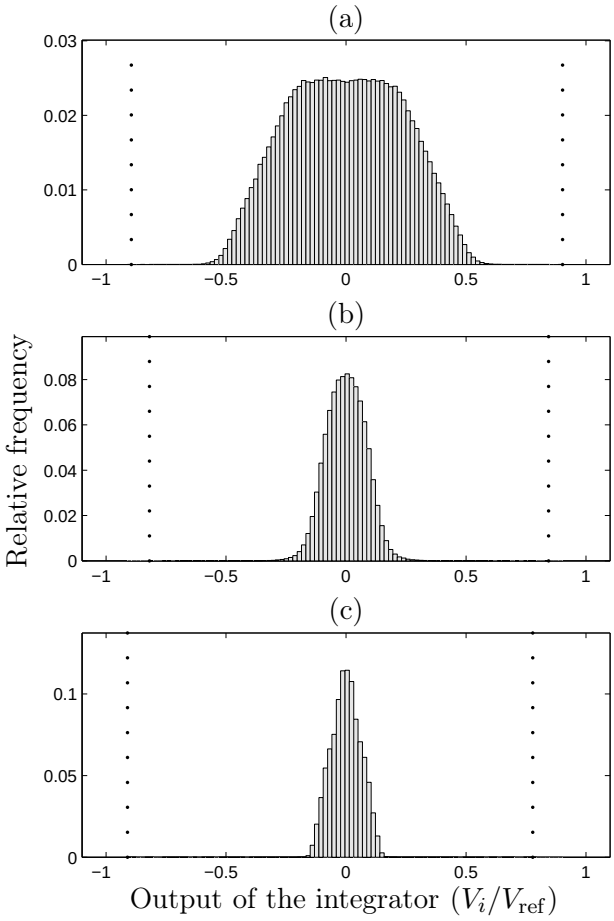


Figure 3.19: Histogram at the output of the (a) first (V_1), (b) second (V_2), and (c) third (V_3) integrator of the scaled architecture (cf. Fig. 3.15 and Tab. 3.3). The input signal is a slowly changing ramp signal ($V_{\text{in}} \in [-0.747V_{\text{ref}}, 0.747V_{\text{ref}}]$). Dotted lines show the upper and lower limits of the histograms.

(the assumption that the error is white-noise is not valid in the case of one-bit $\Delta\Sigma$ converter, especially with dc input). Instead, simulation may be used to find the histogram of the output signal of the integrators to calculate the scaling coefficients. To find these outputs, first the maximum allowable input signal has to be determined. This can be done again by simulation or analytically for some modulator structure [Wong and Ng, 2003]. For a third-order modulator $U_{\max} < 0.75$. One can either use this value to find the required scaling or may use a smaller value (such as $U_{\max} = 0.67$), to gain larger scaling factors for easier implementation. Precise scaling of the input signal will be discussed in Sec. 4.3.2.

According to this discussion, Fig. 3.19 shows the histogram of the state variables (output of the integrators) after scaling. Based on these long-term simulations, the coefficients of the modulator change according to Tab. 3.3.

Note that these results are somewhat conservative, as they are achieved by assuming a maximum input signal $0.7467V_{\text{ref}}$ and long-term operation of the converter ($N = 2^{20}$). In the following section it is shown that the scale factors affect seriously the achievable

Table 3.3: Coefficients of the third-order scaled modulator.

a_1	2.1794
a_2	2.7944
a_3	2.6746
b	0.3670
c_1	0.2810
c_2	0.1595

resolution and/or the number of required cycles for the conversion, thus it is important to find the best trade-off between the scale factors (thus the safe operation) and the operation length.

Resolution of the Converter

Recalling Eq. (3.78), the equivalent number of bits (ENOB) of the converter can be derived as

$$n_{\text{bit}} = \log_2\left(\frac{2\max(V_{\text{in}})}{V_{\text{lsb}}}\right) = \log_2\left(U_{\max} b c_1 c_2 \frac{(N-2)(N-1)N}{3!}\right) \approx$$
$$\approx 3\log_2(N) + \log_2(bc_1c_2) - 2.6, \quad (3.94)$$

where in the last approximation $U_{\max} \approx 1$ and $N \gg 1$ were assumed. This indicates a required number of cycles for n_{bit}-bit resolution

$$N = \text{fix}\sqrt[3]{\frac{3!}{bc_1c_2}\frac{2^{n_{\text{bit}}}}{U_{\max}}} + 2, \quad (3.95)$$

assuming that the final output has been calculated exactly.

In a general L_ath-order modulator the following equation can be used to calculate the required number of cycles

$$\prod_{i=0}^{L_a-1}(N-i) = \frac{2^{n_{\text{bit}}} L_a!}{U_{\max}\left(\prod_{i=1}^{L_a-1} c_i\right) b}. \quad (3.96)$$

In design, one needs to find the lowest value of N consistent with the required resolution. Clearly, the resolution increases rapidly with N, but as $b \leq 1$, $c_1 \leq 1$, and $c_2 \leq 1$ hold, it is reduced by an amount dependent on the product of these scale factors. In practice, however, the scale factors cannot be chosen independently, since they affect the stability of the loop. The most conservative design – discussed in the previous section – is to use long-term simulations and maximum allowable input range to find the scaling coefficients. This however leads to a very small scaling factor product, which may increase the required number of cycles. For example, in an unscaled third-order architecture with $\max(V_{\text{in}}) = V_{\text{ref}}$ (which is a theoretical example, since it cannot be realized due to the overload of the integrators) for $n_{\text{bit}} = 16$-bit resolution, $N = 75$ is required. Based on the conservative

Table 3.4: Coefficients of the third-order scaled modulator with the input signal limited to $\max(V_{\text{in}}) = 0.67V_{\text{ref}}$.

a_1	1.4
a_2	0.99
a_3	0.47
b	0.5674
c_1	0.5126
c_2	0.3171

design discussed previously, $\max(V_{\text{in}}) = 0.7467V_{\text{ref}}$, and $\log_2(bc_1c_2) = -5.93$, resulting in $N = 319$, which is more than 4 times larger than the original value.

One possible way to reduce the required number of cycles is to limit the input signal even more, to, e.g., $0.67V_{\text{ref}}$ or $0.5V_{\text{ref}}$. Even though it increases the required number of cycles due to its direct contribution, however, its secondary effect is the smaller scaling required for the integrators, making the product bc_1c_2 larger. As an example, if the input signal is limited to $0.67V_{\text{ref}}$, the coefficients of the modulator will change according to Tab. 3.4. Thus, $\log_2(bc_1c_2) = -3.43$, resulting in $N = 187$, which is much less than 319, resulting in the same resolution.

Note that large reduction of the input signal may not be advantageous, as this will also limit the dynamic range of the converter and the achievable *SNR* due to the analog noise present in the converter. This means that reducing the input signal to less than $0.4V_{\text{ref}}$ will result in a dynamic range loss of 8 dB, resulting in larger capacitors to reduce the analog noise with the same amount.

There is another way of reducing the required number of cycles. As the converter is operated in an intermittent way and at the start of the conversion the integrators in the loop are reset, its state variables cannot become too large during the finite operation cycles N. Thus, if the input signal is a dc signal and it is guaranteed that this signal does not change during the conversion (which can be assured by using an S/H (sample-and-hold) circuit in front of the converter), the following iterative algorithm may be used to find the lowest required number of cycles:

1. Choose the maximum allowable value of the input signal as a fraction of V_{ref}. In a second-order modulator, the maximum allowable input is around $0.9V_{\text{ref}}$, while in a third-order one it is less than $0.75V_{\text{ref}}$. However, as the integrators' maximum output, the allowable values of the scaling coefficients (b, c_i), and the required number of cycles N all strongly depend on the maximum input signal, it is sometimes advantageous to limit the input signal even more, so as to reduce N. For example, $0.75V_{\text{ref}}$ or $0.67V_{\text{ref}}$ can be chosen for the second-order converter. For a third-order ADC, even $0.5V_{\text{ref}}$ may be advantageous.

2. Find an initial N_{id} by assuming an unscaled architecture, i.e., $b = c_1 = c_2 = 1$, using Eq. (3.94). (For example, 16-bit resolution requires $N_{\text{id}} = 75$ for a third-order loop, while $N_{\text{id}} = 363$ for a second-order one.)

3. Simulate the structure with dc input signals between $(0.7, 1)U_{\max}$ through N_{id} cycles, and get estimates of the scale factors b, c_i from the integrators' maximum output swings.

4. Using the new scale factors, get a new estimate of N using Eq. (3.94).

5. After repeating the previous steps a few ($2 \sim 3$) times, neither the coefficients nor N changes significantly. At this point, the smallest allowable number of cycles N has been obtained.

For example, in the design of a third-order modulator with a maximum noise transfer function (*NTF*) gain of 1.5, using the Delta-Sigma Toolbox in MATLAB [Schreier, 2004], with an input signal reduced to $V_{\text{in}} \leq 0.67V_{\text{ref}}$, the algorithm described above gives $N = 158$ for the optimal number of periods for 16-bit resolution. Note that this is not significantly less than $N = 189$ obtained by more conservative scaling at the expense of a S/H circuit. Especially if higher resolution is required, the required number of cycles found by the algorithm described above is approaching the one based on long-term simulations, as N is increasing when n_{bit} is larger.

All the derivations for the third-order modulator described above can easily be generalized to an arbitrary-order *CIFF* $\Delta\Sigma$ modulator. The general expression to calculate the output is

$$D_{\text{out}} = \frac{\hat{V}_{\text{in}}}{V_{\text{ref}}} = \frac{1}{\binom{N}{L_a}} \underbrace{\sum_{k_{La}=0}^{N-1} \sum_{k_{La-1}=0}^{k_{La}-1} \cdots \sum_{k_1=0}^{k_2-1}}_{L_a} d_k, \tag{3.97}$$

where L_a is the order of the analog loop.

Requantization of the Digital Output

In the theoretical analysis discussed above the digital output of the converter was assumed to be infinitely precise. However, in a real converter the calculated digital output is requantized to the final resolution of the converter. In the following it is shown that rounding gives the best solution for the requantization and even in this case one bit precision loss occurs.

First let us calculate the required digital register width for precise calculation of the output. Filling only ones into the digital filter input, the output of the filter would become

$$D_{\text{out}} = \frac{N(N-1)(N-2)}{3!}. \tag{3.98}$$

This would be the case if the input signal is close to V_{ref}. If the input signal is restricted, then the maximum output of the converter is

$$D_{\text{out}} = U_{\max}\frac{N(N-1)(N-2)}{3!}, \tag{3.99}$$

which requires a register width of

$$n_{\text{bit,reg}} = \text{fix}(\log_2(2D_{\text{out}})) + 1 \approx \text{fix}(3\log_2 N - 1.585 + \log_2 U_{\max}) + 1, \tag{3.100}$$

where fix denotes the truncation or integer-part operation. If $N = 187$ and $U_{\max} = 0.67$, $n_{\text{bit,reg}} = 21$, which is larger than the final resolution ($n_{\text{bit}} = 16$), thus requantization is required.

Unfortunately, this requantization causes one bit resolution loss. Consider a constant input signal $V_{\text{in}} = (k + \varepsilon)V_{\text{lsb}}$, where k is an integer and $\varepsilon \in (0, 0.5)$. According to the previous discussions, with an appropriate N it can be assured that $D_{\text{out}} \in ((k + \varepsilon - 0.5), (k + \varepsilon + 0.5))$. Rounding this signal to the target resolution causes $D'_{\text{out}} \in \{k, k + 1\}$, from which the final quantization error becomes $q \in \{-\varepsilon, 1 - \varepsilon\}$, where this latter one $(1 - \varepsilon) \in [0.5, 1)$. Similar result with opposite polarity can be derived for input signals $V_{\text{in}} = (k - \varepsilon)V_{\text{lsb}}$, where again $\varepsilon \in (0, 0.5)$. This indicates that the final quantization error will be between $\pm V_{\text{lsb}}$, resulting in one bit resolution loss. Note that using different rounding method (floor, ceil, fix, which are various truncation methods), the quantization error may be even larger.

To avoid this resolution loss, two methods can be used. One is to design the converter for $n'_{\text{bit}} = n_{\text{bit}} + 1$ and quantize the final output to n_{bit}. The second alternative is to operate the converter up to N cycles only and increase the resolution with one bit by detecting the sign of the output of the last integrator (cf. Eq. (3.80)).

To verify these results, a third-order converter with a limited input signal $U_{\max} = 0.67$ and coefficients listed in Tab. 3.4 was simulated with $N = 187$, which may give 16-bit resolution according to Eq. (3.94). Indeed, as Fig. 3.20 shows, the output quantization error of the converter calculated with infinite precision is within ± 0.5, as expected. In Fig. 3.21 it is shown that this quantization error is exactly the half of the inverted output of the last integrator in the Nth cycle, according to Eq. (3.80).

However, according to the discussion above, rounding the digital output to 16-bit precision causes the quantization error to be between ± 1 LSB, resulting in one bit resolution loss. This is indicated in Fig. 3.22.

With a little additional logic, the sign of the output of the last integrator can be detected and with this information one bit resolution-increase can be achieved. Fig. 3.23 shows that with infinite precision, the achievable resolution is doubled, i.e., the maximum quantization error is between ± 0.25LSB.

Requantizing this digital output signal to $n_{\text{bit}} = 16$-bit precision, the final quantization error will be bounded by ± 0.5 LSB, fulfilling the resolution requirement (Fig. 3.24).

Note that using the output of the last integrator in the Nth cycle has one more advantage. As the simulated modulator contains the input signal feed-forward path to the input of the quantizer, the output of the last integrator, thus the quantization error does not contain the input signal itself. Still, by looking at Figs. 3.20, 3.21 and 3.22, one can notice that the quantization error's mean value is strongly correlated to the input signal. This shows that the internal quantization error (which is processed by the analog integrators) is not input-independent. However, looking at Figs. 3.23 and 3.24, in which an extra bit of

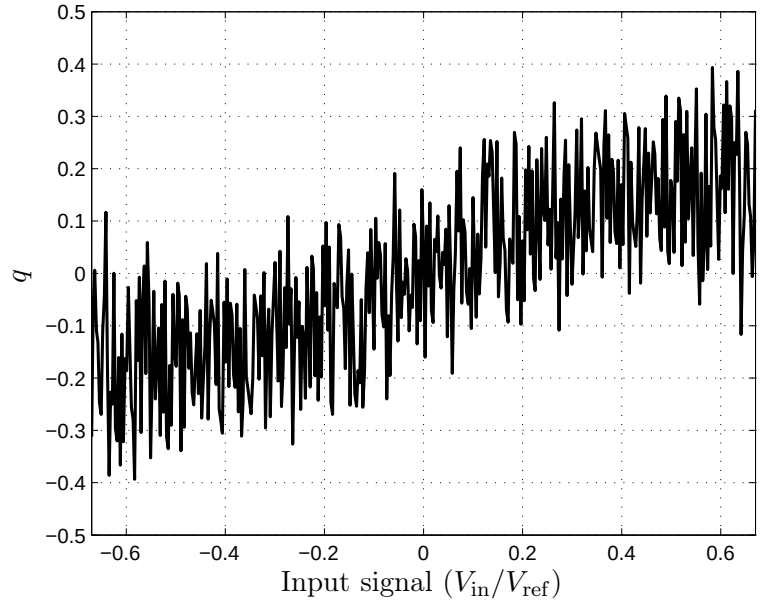


Figure 3.20: Quantization error of a third-order modulator, when the output is calculated with infinite precision.

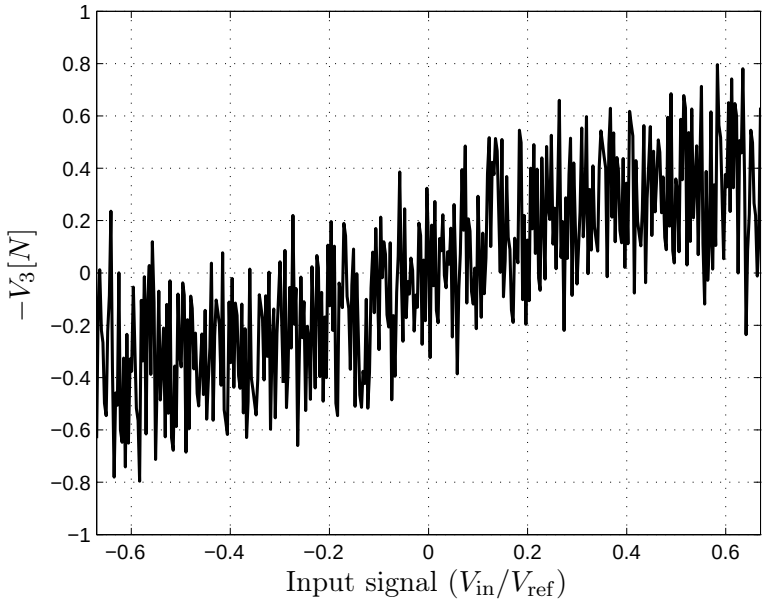


Figure 3.21: Inverted output of the last integrator of the converter in the Nth cycle.

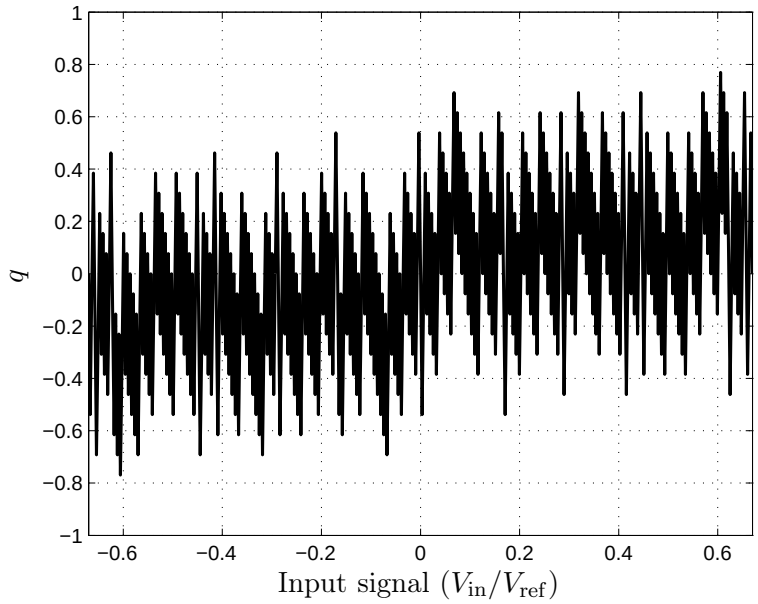


Figure 3.22: Quantization error of a third-order modulator, when the output is calculated with 16-bit finite precision.

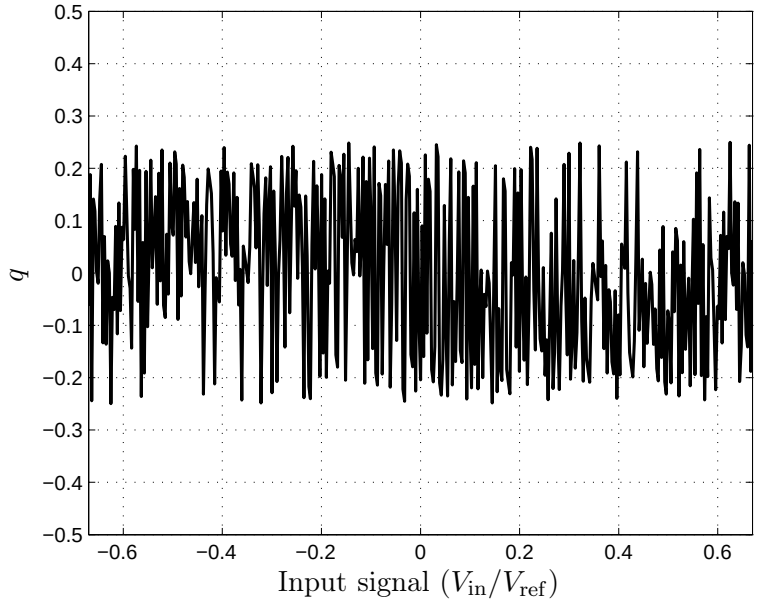


Figure 3.23: Quantization error of a third-order modulator, when the output is calculated with infinite precision and the sign of the output of the last integrator is used to gain one more bit resolution.

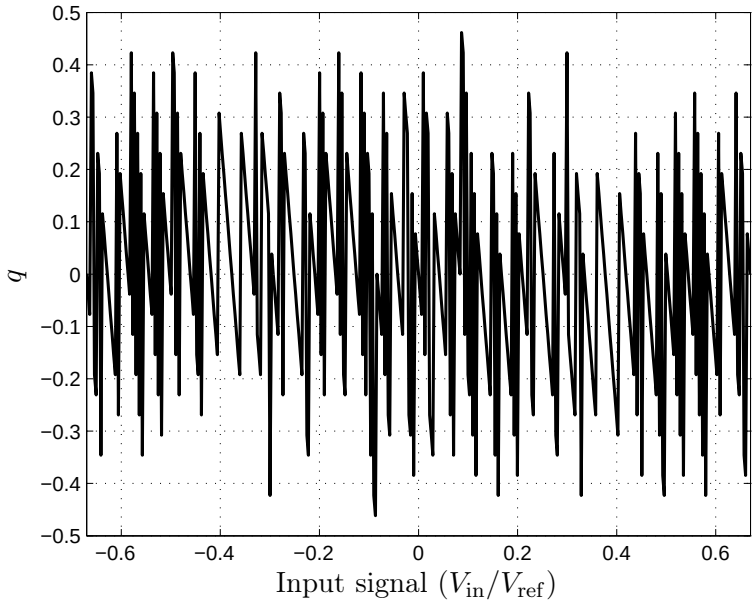


Figure 3.24: Quantization error of a third-order modulator, when the output is calculated with 16-bit finite precision and the sign of the output of the last integrator is used to gain one more bit resolution.

resolution was obtained by detecting the sign of the output of the last integrator, one can notice that this input-related term has been disappeared even from the requantized final quantization error.

3.2.4 Comparison of the Two Extensions

In Secs. 3.2.1 and 3.2.3 two extensions of the first-order incremental $\Delta\Sigma$ converter were proposed.

The first one was based on the observation that using a modulator with a pure L_ath-order differential *NTF*, and applying L_a digital integrators at the output eventually removes the internal quantization errors from the output except the last one ($\varepsilon[N]$), and at the same time the digital integrators accumulate the (constant) input signal. Thus, a great improvement in *SNR* (signal-to-quantization-noise ratio) could be achieved. An approach of using one more integrator was also analyzed.

The second extension used another approach: in a feed-forward architecture the output of the last discrete-time integrator was calculated and a digital filter based on this derivation was used to calculate the final digital output.

It was also shown that in the case of a modulator with pure L_ath-order differential *NTF* and Cascaded-Integrators, Feed-Forward (CIFF) structure with feed-forward of the input signal the two architecture are equivalent, since $-V_3[N + L_a] = \varepsilon[N] = 2q[N]V_{\text{ref}}$ holds in this case.

Chapter 4

Properties of Higher-order Structures

This chapter focuses on general properties of higher-order architectures. It consists of three sections. Sec. 4.1 discusses the effect of input-related noise, while Sec. 4.2 introduces efficient digital filter design techniques, which are capable of periodic noise suppression required for high-precision conversion. Sec. 4.3 addresses practical limitations and proposes different circuit-level solutions to meet the specifications.

4.1 Behavior with Constant Input and Additive Noise

In the previous chapter, the basic architecture-level operation of higher-order incremental converters were discussed. For the derivation of the quantization error, resolution, output calculation, etc., constant input signal was assumed. This can always be achieved by applying a sample-and-hold (S/H) circuit before the modulator, which samples the input signal before the conversion takes place and holds it constant during the conversion. This might be required in several applications, e.g., when the input signal is not always present, or when multiplexed converter is used to collect digital data from several analog sources.

However, there might be some cases when it is not advantageous to use such a S/H circuit. One case is when low power- and area-consumption is critical. Another case is when high precision must be achieved, and the noise, gain, leakage, etc. error of the S/H circuit may not be allowed. A more general case is when the input signal is a dc signal, but noisy. In this case a better result can be achieved by taking more than one samples and averaging out the noise or periodic disturbances, such as the one coming from the power line.

In the following subsections, these cases are addressed in detail.

4.1.1 Constant Input with Additive Gaussian Noise

One important case is when the signal to be converted is a dc signal with additive noise, and the goal is to digitize the mean value of the signal. In a classical Nyquist-rate converter this could be achieved only by taking many individual samples from the signal and average the samples to reduce the noise variance. However, an incremental converter, as it operates in oversampling mode internally, may average the input signal during one conversion, if no

S/H circuit is used in front of the converter. In the following, let us assume that the input signal is the sum of a dc signal and Gaussian noise with zero mean and σ_g^2 variance.

If the modulator contains a feedforward input path, i.e., the input signal is fed to the input of the quantizer (Fig. 3.15), then the output of the modulator is

$$y[k] = u_{\mathrm{dc}} + n_g[k] + \varepsilon[k] * h[k], \tag{4.1}$$

where u_{dc} is the applied dc input signal, $n_g[k]$ is the kth sample of the Gaussian input noise, $h[k]$ is the inverse z-transform of the noise transfer function (NTF), and $\varepsilon[k]$ is the internal quantization error. Thus, in this architecture, the input signal appears in the output without being modified or delayed.

After N cycles, the output samples are weighted by the digital filter and summed. Let us first ignore the quantization error. Theoretically, if the output of the modulator contains only the dc signal and the Gaussian noise, and the weights of the filter are equal ($w_i = 1/N$, i.e., simple averaging takes place), then the variance of the output signal would become $\sigma_{y,\mathrm{id}}^2 = \sigma_g^2/N$. This is the best linear unbiased estimator (BLUE) for the mean value of the input signal based on N samples.

However, using higher-order filtering, the weighting coefficients of the filter are not equal. In this case, the variance of the output signal becomes

$$\sigma_y^2 = \sum_{i=1}^{N} w_i^2 \sigma_g^2 > \frac{\sigma_g^2}{N}. \tag{4.2}$$

In the following the relative increase in the output variance referred to the best linear estimator is calculated for second- and third-order case. For simplicity let us assume that the digital integrators at the output of the converter are operated through N cycles, and none of them are delaying (to achieve exact result, $N' = N - L_a$ should be inserted instead of N into the final equations, where L_a is the order of the analog loop).

In this case, for second-order structure (with second-order digital filter) the output of the filter can be calculated as

$$D_{\mathrm{out}} = \frac{2}{N(N+1)} \sum_{i=1}^{N} \sum_{j=1}^{i} d_j = \frac{2}{N(N+1)} \sum_{i=1}^{N} (N+1-i) d_i. \tag{4.3}$$

Note that the weighting factor $2/(N(N+1))$ is required to ensure that the transfer function of the digital filter is equal to one at dc.

From Eq. (4.3), the weighting coefficients of the filter can be obtained, thus, the ith coefficient of the digital filter's impulse response (weighting function) is

$$w_2[i] = \frac{2}{N(N+1)}(N+1-i),\ i \in (1, N). \tag{4.4}$$

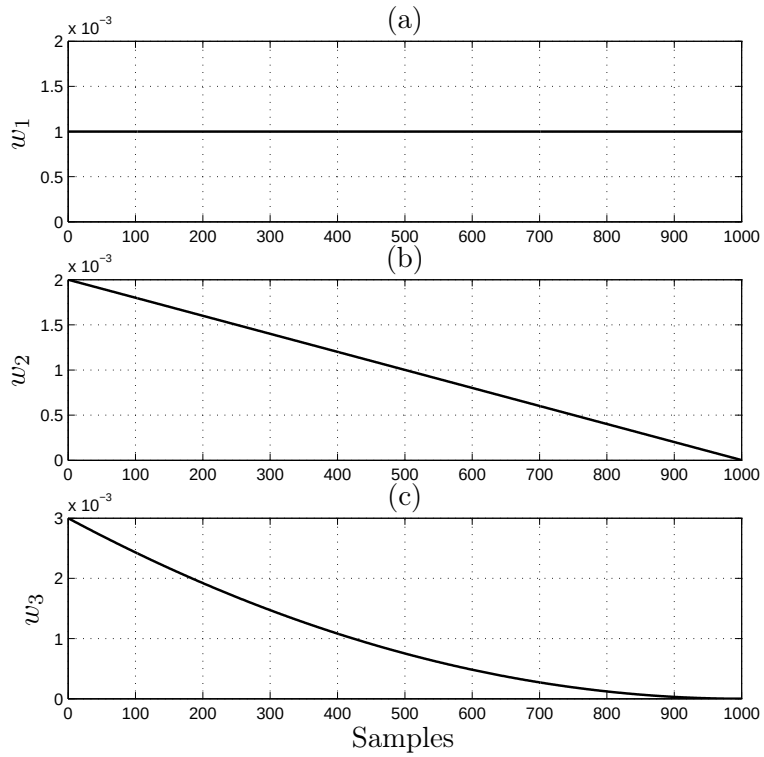


Figure 4.1: Weighting function of the Cascade-of-Integrators filter consists of (a) one, (b) two, and (c) three integrators.

Similarly, for third-order filtering, the output of the digital filter becomes

$$D_{\text{out}} = \frac{2 \cdot 3}{N(N+1)(N+2)} \sum_{i=1}^{N} \sum_{j=1}^{i} \sum_{k=1}^{j} d_k = \\ = \frac{2 \cdot 3}{N(N+1)(N+2)} \sum_{i=1}^{N} \frac{(N+1-i)((N+1-i)+1)}{2} d_i, \quad (4.5)$$

thus, the ith coefficient of the filter's weighting function is

$$w_3[i] = \frac{2 \cdot 3}{N(N+1)(N+2)} \frac{(N+1-i)((N+1-i)+1)}{2}, \; i \in (1, N). \quad (4.6)$$

To illustrate the filter weighting functions, Fig. 4.1 shows the impulse response of three filters, consist of one, two and three cascaded integrators, respectively. Note that a first-order filter puts equal weights on the samples, while higher-order filters behave according to Eqs. (4.4) and (4.6).

The relative increase in the output variance for second-order filter can be calculated by

the following way:

$$\frac{\sigma_{y,2}^2}{\sigma_{y,\mathrm{id}}^2} = N\sum_{i=1}^{N} w_2^2[i] = N\sum_{i=1}^{N}\left(\frac{2}{N(N+1)}\right)^2 (N+1-i)^2 =$$
$$= N\left(\frac{2}{N(N+1)}\right)^2 \sum_{j=1}^{N} j^2 = \frac{4}{N(N+1)^2}\frac{N(N+1)(2N+1)}{2\cdot 3} = \frac{2}{3}\frac{2N+1}{N+1} =$$
$$= \frac{2}{3}\frac{2N+2-1}{N+1} = \frac{4}{3} - \frac{2}{3(N+1)} < \frac{4}{3} \approx 1.33. \quad (4.7)$$

Thus, if the second-order digital filter is operated through N cycles and an input signal consists of a dc signal and additive Gaussian noise with variance σ_g^2 is applied at the input of the converter, the output variance will satisfy

$$\sigma_{y,2}^2 < \frac{4}{3}\frac{\sigma_g^2}{N}. \quad (4.8)$$

Even though this variance greater than $\sigma_{y,\mathrm{id}}^2$, the increase in the variance is negligible, and the input noise has been suppressed by a factor of $4/(3N)$.

For third-order filters similar result can be achieved:

$$\frac{\sigma_{y,3}^2}{\sigma_{y,\mathrm{id}}^2} = N\sum_{i=1}^{N} w_3^2[i] == N\sum_{i=1}^{N}\left(\frac{2\cdot 3}{N(N+1)(N+2)}\right)^2 \times$$
$$\times \left(\frac{(N+1-i)((N+1-i)+1)}{2}\right)^2 = \frac{36}{4}\frac{1}{N(N+1)^2(N+2)^2}\sum_{j=1}^{N}(j(j+1))^2 =$$
$$= 9\frac{1}{N(N+1)^2(N+2)^2}\sum_{j=1}^{N}\left(j^2+j\right)^2 == 9\frac{1}{N(N+1)^2(N+2)^2}\sum_{j=1}^{N} j^4 + 2j^3 + j^2. \quad (4.9)$$

Substituting the known sums [Gradshteyn and Ryzhik, 1994]

$$\sum_{j=1}^{N} j^2 = \frac{N(N+1)(2N+1)}{6}$$
$$\sum_{j=1}^{N} j^3 = \frac{N^2(N+1)^2}{4}$$
$$\sum_{j=1}^{N} j^4 = \frac{N(N+1)(2N+1)(3N(N+1)-1)}{30} \quad (4.10)$$

leads to

$$\frac{\sigma_{y,3}^2}{\sigma_{y,\mathrm{id}}^2} = \frac{9}{N(N+1)^2(N+2)^2}\times$$
$$\times\left(\frac{N(N+1)(2N+1)(3N(N+1)-1)}{30}+\frac{N^2(N+1)^2}{2}+\frac{N(N+1)(2N+1)}{6}\right)=$$
$$=\frac{9}{N(N+1)^2(N+2)^2}\frac{N(N+1)^2(3N(2N+1)+N)}{2\cdot 3\cdot 5}=\frac{9}{5}\frac{N(6N+4)}{6(N+2)^2}=$$
$$=\frac{9}{5}\frac{6N^2+4N}{6(N^2+4N+4)}=\frac{9}{5}\left(1-\frac{20N+24}{6(N^2+4N+4)}\right)<\frac{9}{5}=1.8. \quad (4.11)$$

Again, although the converter does not optimally average the noise, the output variance satisfies

$$\sigma_{y,3}^2 < 1.8\frac{\sigma_g^2}{N}, \quad (4.12)$$

indicating a reasonably good noise suppression.

For a general higher-order modulator, instead of calculating the exact value of $\sum_{i=1}^{N} i^k$ (which exists for any N and k [Gradshteyn and Ryzhik, 1994], but its analytical form becomes very complex), it may be advantageous to use the integral-approximation of

$$\sum_{i=1}^{N} i^k < \int_1^{N+1} i^k di. \quad (4.13)$$

To verify this method, the third-order case is derived also with this approximation:

$$\frac{36}{4}\frac{1}{N(N+1)^2(N+2)^2}\sum_{j=1}^{N}(j(j+1))^2=\frac{9}{N(N+1)^2(N+2)^2}\sum_{j=1}^{N}(j^2+2j^3+j^4)<$$
$$<\frac{9}{N(N+1)^2(N+2)^2}\int_1^{N+1}(j^2+2j^3+j^4)dj=\frac{9}{N(N+1)^2(N+2)^2}\times$$
$$\times\left(\frac{(N+1)^5-1}{5}+2\frac{(N+1)^4-1}{4}+\frac{(N+1)^3-1}{3}\right)=\ldots=$$
$$=\frac{9}{5}\frac{N^4+7.5N^3+21.67N^2+30N+20}{N^4+6N^3+13N^2+12N+4} \quad (4.14)$$

giving comparable, but slightly larger limit than in the exact analysis above. This approximation gives a limit $\sigma_{y,3}^2/\sigma_{y,\mathrm{id}}^2 < 2$ only for $N > 13$ and $\sigma_{y,3}^2/\sigma_{y,\mathrm{id}}^2 < 1.85$ for $N > 53$. The less tight limits on the output variance of this approximation comes from the fact that the lhs of Eq. (4.13) is a very coarse estimation of the integral of the continuous function i^k.

Note that the above discussion did not take into account the quantization noise also present in the output. However, since the output quantization noise is a filtered version of the internal quantization error, it may be modeled as an additive Gaussian noise with a standard deviation $3\sigma_q < V_{\mathrm{lsb}}/2$, i.e., $\sigma_q^2 < V_{\mathrm{lsb}}^2/36$. Thus, the averaged incoming noise

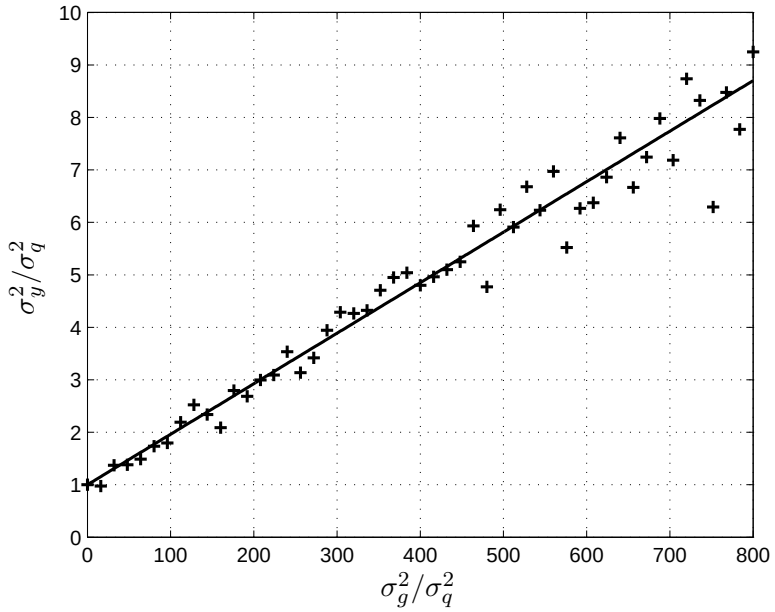


Figure 4.2: Normalized output signal variance as the function of the normalized input noise variance. + signs represent the simulation results, while solid line represents the expected relationship of Eq. (4.16).

must be significantly larger than the quantization error to affect the output. Hence, its variance needs to satisfy

$$k\frac{\sigma_g^2}{N} \gg \frac{V_{\text{LSB}}^2}{36},\ k \in \{4/3, 9/5\} \tag{4.15}$$

i.e. until the input signal noise variance is significantly less than N/k times the quantization noise variance σ_q^2, the final output noise will be suppressed by the quantization noise.

Simulation results agree well with the theoretical expectations discussed above. A third-order modulator ($k = 1.8$) with feed-forward input path was used, with the input signal limited to $0.67V_{\text{ref}}$ and with coefficients listed in Tab. 3.4. 16-bit resolution was assumed, resulting in $N = 187$. The digital output was calculated with infinite precision. Different dc signals plus noise with 50 different variances were applied to the input of the converter and 300 conversions were simulated for each input to get an estimate of the output variance.

Fig. 4.2 shows the normalized output signal variance (σ_y^2/σ_q^2, where σ_q^2 is the variance of the quantization error) as the function of the variance of the normalized analog input noise (σ_g^2/σ_q^2), marked with + signs. The solid line represents the equation

$$\frac{\sigma_y^2}{\sigma_q^2} = 1 + \frac{1.8}{N}\frac{\sigma_g^2}{\sigma_q^2} \tag{4.16}$$

which is the expected theoretical relationship between the input and output noise.

In Eq. (4.16) (see the solid line in Fig. 4.2), the first term (1) in the rhs represents the fact that the variance of the quantization noise always contributes to the final variance, while the last term shows the reduction of the input noise variance. The simulation results

show good agreement with this result, e.g., when the input signal variance $\sigma_g^2 = N\sigma_q^2/1.8 \approx 100\sigma_q^2$, the output variance becomes $\sigma_y^2 = 2\sigma_q^2$, i.e., the output variance doubles compared to the noiseless case. This clearly shows the great reduction capability of the converter for noisy inputs. This also means that if the input noise variance is not significantly greater than that of the quantization error, during the conversion the error will be averaged out and the final quantization and noise error will be within half LSB.

4.1.2 Constant Input with Periodic Noise

One great advantage of the dual-slope converter is that it may suppress periodic disturbances if the integration time of the unknown input signal is matched with the duration time of one or more periods of the disturbing signal. Typical usage of this property in measurement applications is to suppress the periodic noise coupled from the power line.

This property is also inherited in a first-order incremental converter. The only difference is that in this case not the analog input is integrated, but the output of the modulator is summed in the digital domain, i.e., the averaging process is based on samples and not on continuous functions. The noise cancellation is still preserved, since the weights of each samples are equal.

Unfortunately, this property does not hold for higher-order incremental converters, since in these converters the weights of the post-processing digital filter impulse response are not equal and are not even symmetrical. This means that higher-order incremental converters with Cascades-of-Integrators digital filter at the output cannot be used for cancellation of periodic disturbances.

However, first-order cancellation can be obtained by appropriate operation of the converter, as follows. Let us assume that the signal to be measured is a dc signal, the disturbing signal is periodical and symmetrical (e.g., a sine wave), the converter is using a S/H circuit, and two conversions are taking place during one period of the disturbing signal. In this case, the first conversion will convert a dc input $V_{\text{in}} = V_{\text{dc}} + b$, while during the second conversion, the input signal becomes $V_{\text{in}} = V_{\text{dc}} - b$, where V_{dc} is the dc signal to be converted, while $b \in (-A, A)$ is an offset signal, where A is the amplitude of the disturbing signal. Taking the average of these two conversions provides a first-order cancellation of the periodic noise. Note that this operation assumes that the periodical disturbance is a symmetrical function, i.e., only odd harmonics present in it, which may not be the valid model for signals coupled from the power line.

Better periodic noise cancellation, based on symmetrical digital filters will be discussed in Sec. 4.2.

4.1.3 General Case

For arbitrary input signal without S/H circuit in front of the A/D converter, let us assume that the modulator contains a feed-forward path for the input signal, i.e., the input signal samples are not affected by the modulator, they are fed into the digital filter without being delayed or modified. (Note that this assumption is only valid if the quantization error in the loop is uncorrelated with the input signal.) In this case, the spectral behavior of the

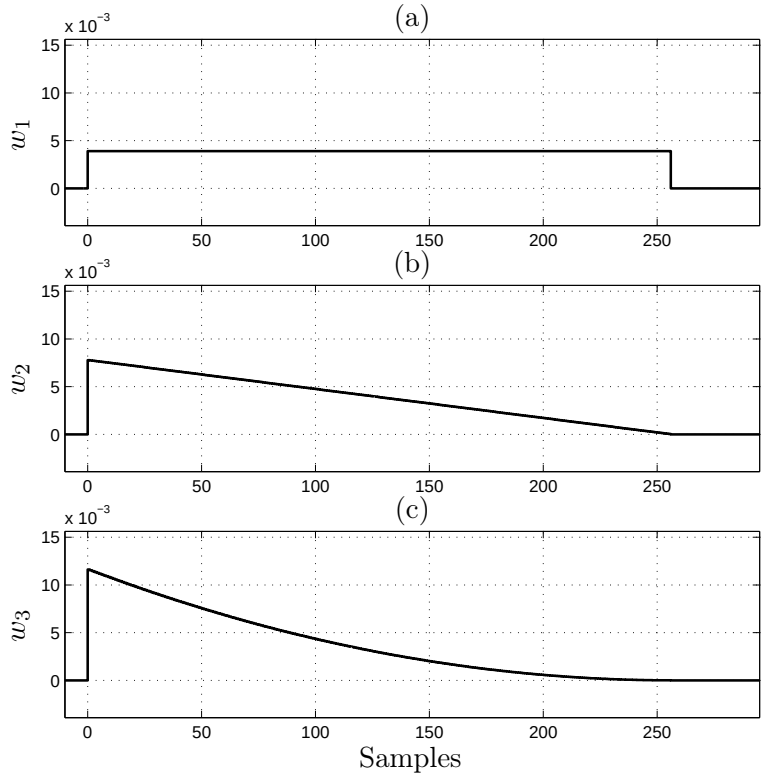


Figure 4.3: Impulse response of the equivalent FIR filter of the (a) first-order (b) second-order (c) third-order transient CoI filter.

input signal is modified only by the CoI (Cascades-of-Integrators) digital filter following the modulator.

As it was discussed earlier, the CoI filter operated in transient mode can be treated as an FIR-filter with the appropriate coefficients (cf. Eqs. (4.3)–(4.6) and Fig. 4.1). To illustrate the FIR-filter impulse response, Fig. 4.1 is repeated here for $N = 256$, moreover, trailing and ending zeros are shown to clearly identify the impulse response of the filter (Fig. 4.3).

The spectral behavior of the CoI filters can be studied on the spectra of these equivalent FIR-filters. The first-order CoI filter's spectrum is the well-known digital sinc-filter, i.e.,

$$S_1(f/f_s) = \frac{\sin \pi N \frac{f}{fs}}{N \sin \pi \frac{f}{fs}}, \tag{4.17}$$

having several zeros at $f_0 = if_s/N$ frequencies, where $i \in \{1..N-1\}$. Even though the higher-order CoI filter's spectrum may be calculated analytically, this has little practical importance. Instead, Fig. 4.4 shows the high-resolution FFT of the three filters. It can be seen that the zeros have been disappeared from the spectrum of higher-order filters, and only the $1/f$ term dominates. This indicates a mild high-frequency attenuation up to $f_s/2$.

As discussed above, this attenuation may not be adequate if line frequency disturbances must be cancelled. In the following section, another digital filter structures are introduced,

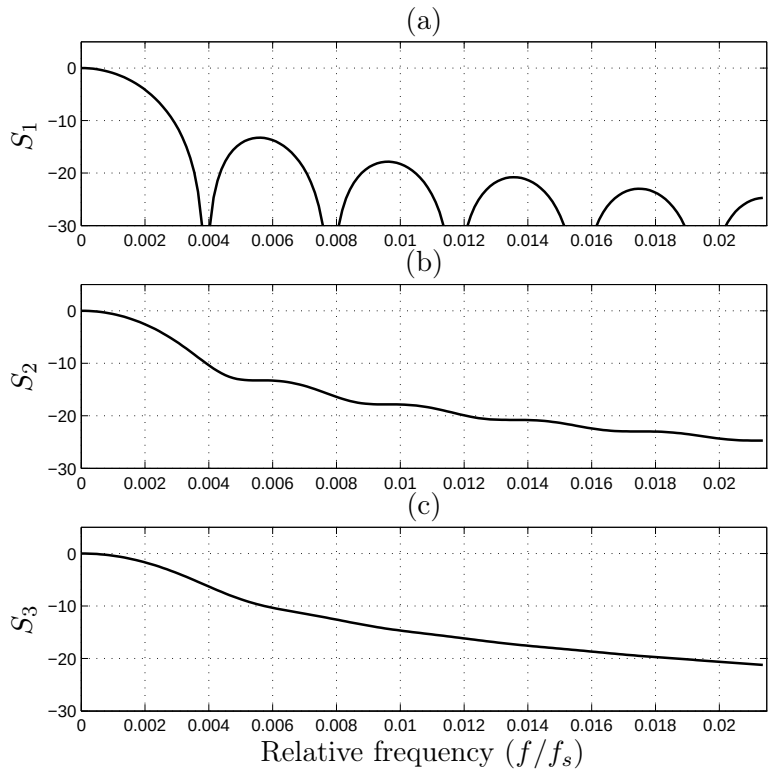


Figure 4.4: Spectrum of the equivalent FIR filter of the (a) first-order (b) second-order (c) third-order transient CoI filter.

which may efficiently eliminate periodic noise from the input signal during conversion.

4.2 Line Frequency Suppression

Up to now it was shown that the dual-slope converter is able to cancel periodic noise disturbances if its integration time matches to the time period of the incoming periodic noise. It was also discussed, that this property is inherited also in a first-order incremental converter, until its output filter is a first-order integrator. However, higher-order converters require higher-order CoI filters, and as it was analyzed throughout the previous sections, these filters does not provide zeros in their transfer characteristic, due to the asymmetry of their impulse response. In the following, lowpass filters with symmetrical impulse response will be discussed, having zeros in their transfer function, making it possible to suppress periodic disturbances.

One of the easiest-to-implement low-pass decimating filter with capability of periodic noise reduction is the averaging filter, which adds the last N samples together and divides the result by N. The filter output is decimated by N, thus the filter operates in an accumulate-and-dump way. This is actually the required filter in the case of the the first-

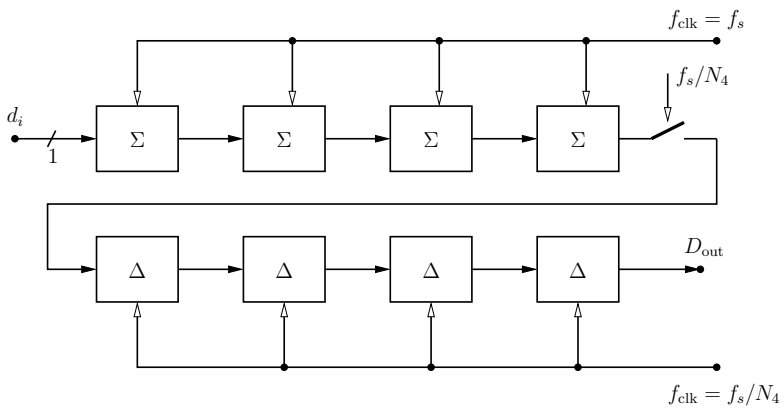


Figure 4.5: An efficient realization of a 4th-order CIC filter

order incremental converter. Its transfer function is the following:

$$H_1(z) = \sum_{i=0}^{N-1} z^{-i} = \frac{1 - z^{-N}}{1 - z^{-1}}, \tag{4.18}$$

and its transfer characteristics is

$$H_1(f) = \frac{\text{sinc}(Nf/f_s)}{\text{sinc}(f/f_s)}, \tag{4.19}$$

where f_s is the sampling rate, f is the frequency and $\text{sinc}(x) = \sin(\pi x)/(\pi x)$.

The rhs of Eq. (4.18) shows an IIR representation of the simple averaging filter. Based on this representation, higher-order "averaging" filters may be defined by the following equation:

$$H_L(z) = \left(\frac{1 - z^{-N}}{1 - z^{-1}}\right)^L. \tag{4.20}$$

These filters are usually referred as Lth-order sinc-filters or Lth-order Cascaded-Integrator-Comb (CIC) filters. This latter name comes from a very efficient realization of the filter, first introduced by Hogenauer (1981). Such a realization is shown in Fig. 4.5.

This filter is among the most popular decimation filters in classical $\Delta\Sigma$ design. It is usually used for the first-stage of the decimator filter and has been analyzed in details by Candy and Benjamin (1981) for $\Delta\Sigma$ decimator design.

In the following subsections these filters will be examined for incremental converter design. Calculation methods are introduced to estimate the required number of cycles for a given precision. It is also shown that the optimal filter is either L_ath-order or $L_a + 1$st-order sinc-filter, where L_a is the order of the $\Delta\Sigma$ modulator used in the converter.

During the derivations it is assumed that either the internal quantization error $\varepsilon[k]$ or the output of the last integrator $V_{L_a}[k]$ in the loop is limited, i.e., $\varepsilon[k] \in (-V_{\text{ref}}/(l-1), V_{\text{ref}}/(l-1))$ or $V_{L_a}[k] \in (-V_{\text{ref}}, V_{\text{ref}})$. In the case when the order of the digital filter $L_d = L_a + 1$, statistical properties of the internal quantization error or the output of the last integrator are also assumed: it is assumed about the internal quantization error that it

is uniformly distributed, uncorrelated with the input signal and the individual samples are uncorrelated with each other (i.e., the noise is white), while in the case when the output of the last integrator is used as a constraint, it is assumed that it has approximately Gaussian distribution.

Even though rigid theoretical analyzes show that most of these assumption are not valid (see, e.g., [Gray, 1989; Gray et al., 1989; Gray, 1990]), in practical circuits, where noise present at the input and also in the circuit itself, these assumptions are at least approximately fulfilled.

Assuming constant input and the above properties of the internal quantization error, the main question to be answered is, how long the converter must be operated to achieve a given resolution, i.e., after how many cycles become the difference between the filtered digital output signal and the original analog signal is less than half LSB of the target resolution.

Unfortunately, to answer these questions most of the derivations in the previous sections are useless, since in those derivations it was assumed that the digital filter following the modulator is the CoI (Cascade-of-Integrators) filter with the same order as that of the $\Delta\Sigma$ modulator. Since the digital filter in this situation is replaced by the higher-order sinc-filters, different methods have to be used to find out the required number of cycles for a given resolution.

4.2.1 Modulators with Pure Differential Noise Transfer Function

First, let us examine those converters which have pure differential noise transfer function (NTF), and the input signal is fed forward to the input of the internal quantizer, i.e., the modulator's output is

$$Y(z) = U(z) + (1 - z^{-1})^{L_a} E(z). \tag{4.21}$$

For simplicity, in most of the following analysis $L_a = 3$ will be used, where L_a is the order of the modulator. Note that $L_a = 2$ and $L_a = 3$ provide the best trade-off between analog circuit complexity and conversion speed for incremental conversion.

Digital Sinc-filter with Order $L_d < L_a$

Filtering the third-order $\Delta\Sigma$ output with a first-order sinc-filter results in the following output:

$$D_{\text{out}} = \frac{1}{N_1}\frac{1 - z^{-N_1}}{1 - z^{-1}} Y(z) = U(z) + \frac{1}{N_1}(1 - z^{-N_1})(1 - z^{-1})^2 E(z), \tag{4.22}$$

where N_1 is the operation length of the first-order sinc-filter, usually referred as decimation ratio (cf. Fig. 4.5). Since the input signal is constant, it is not affected by the averaging filter.

Simplifying Eq. (4.22), the z-transform of the *finite* impulse response of the filter can be easily calculated. It is given as follows:

$$w_1(z) = \frac{1}{N_1}\left(1 - 2z^{-1} + z^{-2} - z^{-N_1}(1 - 2z^{-1} + z^{-2})\right). \tag{4.23}$$

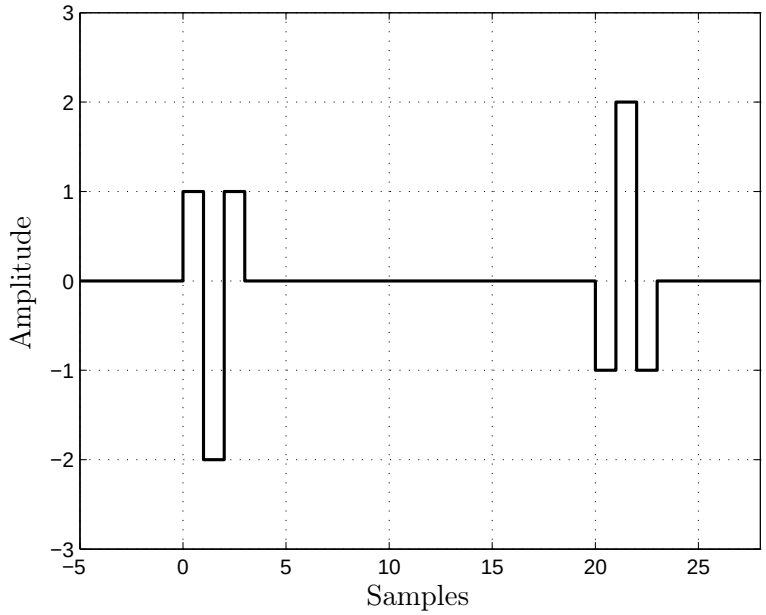


Figure 4.6: Total impulse response of the merged NTF and first-order sinc filter without scaling.

The time-domain impulse response based on this equation is plotted for $N_1 = 20$ in Fig. 4.6 without the scaling factor $1/N_1$. The length of the impulse response is $N_1 + 2$, thus the converter must be operated through $N_1 + 2$ cycles to get a correct output.

Let us first examine the case when the internal quantization error is a random number between $\pm V_{\text{ref}}$. (Note that this is a strict theoretical example, since third-order loop with 1-bit internal quantizer and pure Nth-order differential NTF cannot be realized, since it is not stable.) It is straightforward that the filtered quantization error's maximum swing is

$$\max |q[k]| = \frac{1}{N_1} \sum |w_1[i]| = \frac{8}{N_1}. \tag{4.24}$$

Since the input signal is not affected by the digital filter, this maximum is equal to the half LSB of the final resolution. If the converter has n_{bit}-bit resolution, and the input signal is between $\pm U_{\max}$, then

$$\frac{8}{N_1} = \frac{\text{LSB}}{2} = \frac{U_{\max}}{2^{n_{\text{bit}}}} \tag{4.25}$$

must hold.

This implies

$$N_1 = \frac{8 \cdot 2^{n_{\text{bit}}}}{U_{\max}}. \tag{4.26}$$

This equation suggests, that N_1 must be very large, even larger than the required number of cycles for the first-order incremental converter. This means that using this first-order sinc-filter with higher-order loops is impractical.

Even if the internal quantizer has l multiple levels, the required number of cycles

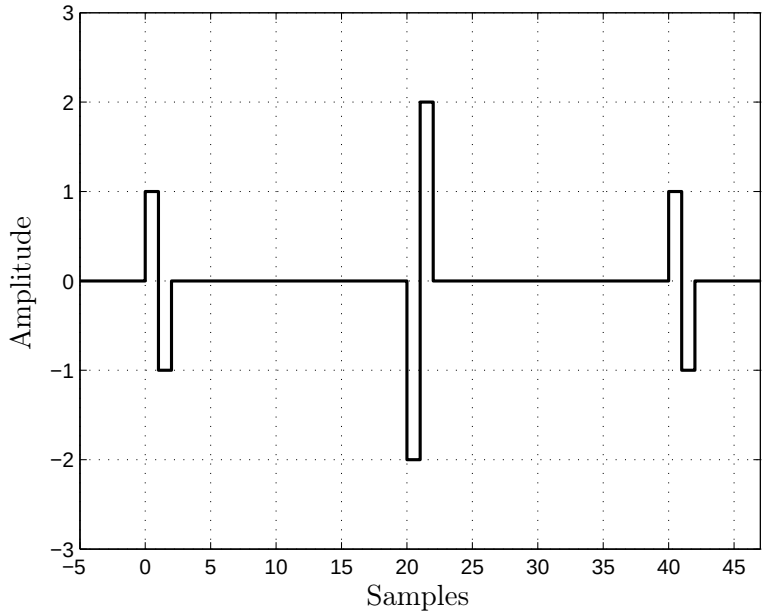


Figure 4.7: Total impulse response of the merged NTF and second-order sinc filter without scaling.

becomes

$$N_1 = \frac{8 \cdot 2^{n_{\text{bit}}}}{U_{\max}(l-1)}. \tag{4.27}$$

For example, if the internal converter has only two levels and 16 bits of precision required (theoretical example), and the maximum input signal is limited to $0.67V_{\text{ref}}$, then $N_1 = 782519$(!), thus the required number of cycles is $N = N_1 + 2 = 782521$, which is impractical for conversion. If the converter has $l = 33$ levels with the same other parameters, then $N_1 = 24454$, thus the required number of cycles $N = N_1 + 2 = 24456$, which still gives very large conversion cycle.

Similar problem exist using second-order filter with third-order modulator. The transfer function from the internal quantization error to the output of the digital decimation filter becomes

$$D_{\text{out}} = \frac{1}{N_2^2}\left(\frac{1-z^{-N_2}}{1-z^{-1}}\right)^2 Y(z) = U(z) + \frac{1}{N_2^2}(1-z^{-N_2})^2(1-z^{-1})E(z), \tag{4.28}$$

where N_2 is the decimation ratio of the second-order sinc-filter.

Again, rearranging the internal quantization noise's filter coefficients results in

$$w_2(z) = \frac{1}{N_2^2}\left((1-z^{-1}) - 2z^{-N_2}(1-z^{-1}) + z^{-2N_2}(1-z^{-1})\right). \tag{4.29}$$

The time-domain impulse response for $N_2 = 20$ can be seen in Fig. 4.7 without the scaling factor $1/N_2^2$. The length of the impulse response is $2N_2 + 1$, thus the converter must be operated through $2N_2 + 1$ cycles to get a correct output.

Using the theoretical example ($\varepsilon \in \pm V_{\text{ref}}$), the filtered quantization error's maximum swing is

$$\max|q[k]| = \frac{1}{N_2^2}\sum|w_2[i]| = \frac{8}{N_2^2}. \tag{4.30}$$

Since the input signal is not affected by the digital filter, this maximum is equal to the half LSB of the final resolution, similarly to the previous case. If the converter has n_{bit}-bit resolution, and the input signal is between $\pm U_{\max}$, then

$$\frac{8}{N_2^2} = \frac{\text{LSB}}{2} = \frac{U_{\max}}{2^{n_{\text{bit}}}}, \tag{4.31}$$

which implies

$$N_2 = \sqrt{\frac{8\cdot 2^{n_{\text{bit}}}}{U_{\max}}} = \frac{2^{\frac{n_{\text{bit}}+3}{2}}}{\sqrt{U_{\max}}}. \tag{4.32}$$

If the internal quantizer has l levels, the required number of cycles becomes

$$N_2 = \frac{2^{\frac{n_{\text{bit}}+3}{2}}}{\sqrt{U_{\max}(l-1)}} \tag{4.33}$$

For example, if the internal converter has only two levels and 16 bits of precision required (theoretical example), and the maximum input signal is limited to $0.67V_{\text{ref}}$, then $N_2 = 885$, thus the required number of cycles is $N = 2N_2 + 1 = 1771$, which is still a very long operation compared to the CoI filter. If the converter has 33 levels with the same other parameters, then $N_2 = 157$, thus the required number of cycles $N = 2N_2 + 1 = 314$.

Digital Sinc-filter with Order $L_d = L_a$

Much better result can be obtained if the order of the digital filter (L_d) is equal to that of the modulator (L_a). In this case, the total noise transfer function becomes

$$w_3(z) = \frac{1}{N_3^3}(1 - z^{-N_3})^3. \tag{4.34}$$

The impulse response for $N_3 = 20$ can be seen in Fig. 4.8 without the scaling factor $1/N_3^3$. Even though the impulse response is even longer, ($N = 3N_3$), due to the very small scaling coefficient ($1/N_3^3$), a small total number of cycles is expected to achieve a given resolution.

Similarly to the previous discussion, the maximum error at the output becomes

$$\max|q[k]| = \frac{1}{N_3^3}\sum|w_3[i]| = \frac{8}{N_3^3}, \tag{4.35}$$

if $\varepsilon[k] \in (-1, 1)V_{\text{ref}}$. Thus, similarly to the previous derivations, for a given resolution with an l-level internal quantizer,

$$N_3 = \sqrt[3]{\frac{8\cdot 2^{n_{\text{bit}}}}{U_{\max}(l-1)}} = \sqrt[3]{\frac{2^{n_{\text{bit}}+3}}{U_{\max}(l-1)}}. \tag{4.36}$$

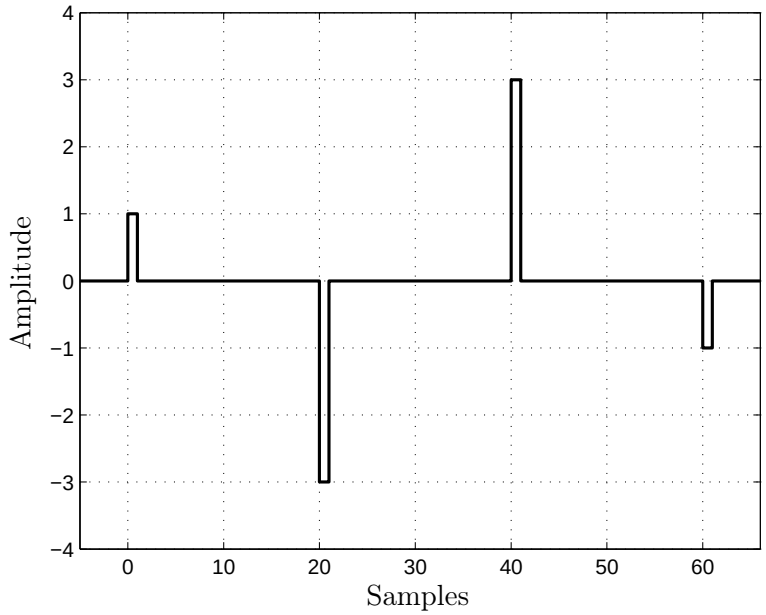


Figure 4.8: Total impulse response of the merged NTF and third-order sinc filter without scaling.

With $l = 2$, $n_{\text{bit}} = 16$ and $U_{\max} = 0.67$ $N_3 = 92$, from which $N = 3N_3 = 277$, giving a very small required number of cycles, in the same order as for the case of CoI filters. If $l = 33$, $N_3 = 30$, $N = 90$.

Note that during the previous three analysis, the distribution of the internal quantization error was not taken into account. The results are valid until the internal error maximum is bounded by $\pm V_{\text{ref}}/(l-1)$, uncorrelated from the input signal and finally, $\varepsilon[k]$ and $\varepsilon[k - N_3]$ are also uncorrelated.

Digital Sinc-filter with Order $L_d > L_a$

In classical $\Delta\Sigma$ design usually the order of the decimation sinc-filter is greater than that of the modulator by one, i.e., $L_d = L_a + 1$. This result was analyzed in detail by Candy and Benjamin (1981), using frequency-domain analysis tools. However, the incremental converter can be analyzed better in the time-domain. In the following the fourth-order sinc-filter with third-order $\Delta\Sigma$ loop will be examined.

First let us examine this case with the method used for lower-order filters (see previous section). The total transfer function from the input and the internal quantization error to the digital output becomes

$$D_{\text{out}} = \frac{1}{N_4^4}\left(\frac{1-z^{-N_4}}{1-z^{-1}}\right)^4 Y(z) = U(z) + \frac{1}{N_4^4}\frac{(1-z^{-N_4})^4}{1-z^{-1}}E(z). \tag{4.37}$$

A major difference between this filter and the previous filters is that the transfer function from the internal quantization error to the output contains one pole. However, it can

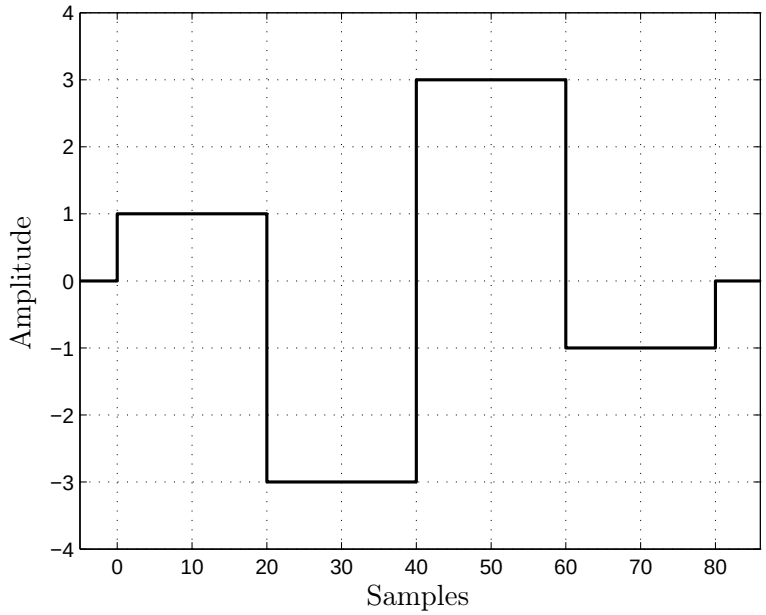


Figure 4.9: Total impulse response of the merged NTF and fourth-order sinc filter without scaling.

easily be shown that the filter has still finite impulse response, since

$$w_4(z) = \frac{1}{N_4^4}\frac{(1-z^{-N_4})^4}{1-z^{-1}} = \frac{1}{N_4^4}\left((1-z^{-N_4})^3\frac{(1-z^{-N_4})}{1-z^{-1}}\right), \tag{4.38}$$

i.e., the impulse response of the filter is the convolution of the third-order filter discussed in the previous section ($w_a(z) = 1 - 3z^{-N_4} + 3z^{-2N_4} - z^{-4N_4}$) and a first-order sinc-filter's impulse response ($w_b[k] = \epsilon[k] - \epsilon[k - N_4]$, where $\epsilon[k]$ is the discrete-time step function). The convolution results in an impulse response shown in Fig. 4.9 for $N_4 = 20$. One can notice that the impulses of the third-order response (cf. Fig. 4.8) are accumulated through N_4 samples. The total transient of the filter, so the required number of cycles for the operation is $N = 4N_4$.

During the previous discussions the maximum output signal was calculated, based on the idea that worst-case output occurs when $\pm\max(\varepsilon[k])$ is weighted by $\mp w_i[k]$. This idea yielded to the product of the maximum input signal and the sum of the absolute value of the filter coefficients. Using this method in the case of fourth-order sinc-filter results in the following equation:

$$\max|q[k]| = \frac{1}{N_4^4}\sum|w_4[i]| = \frac{8N_4}{N_4^4} = \frac{8}{N_4^3}. \tag{4.39}$$

Comparing Eqs. (4.35) and (4.39), it can be seen that this constraint lead to the same equation to calculate the required number of cycles, i.e., there is no benefit using higher-order filter (especially since the total number of cycles is $4N_4$, as opposed to the $3N_3$ in the case of third-order filter). However, this result has been achieved from a worst-case

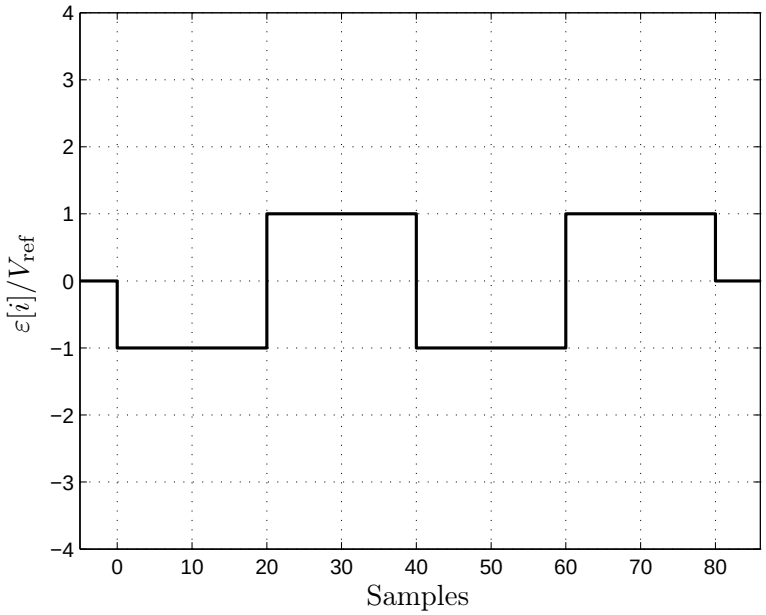


Figure 4.10: Worst-case internal quantization error sequence in the case of fourth-order sinc-filter.

analysis, which assumes that whenever the filter coefficient is negative, the input signal into the filter (i.e., the internal quantization error) $\varepsilon[k] = \max \varepsilon[k]$, and whenever the filter coefficient is positive, $\varepsilon[k] = \min \varepsilon[k]$. Fig. 4.10 shows the worst-case internal quantization error signal sequence from which Eq. (4.39) was derived.

It is clear that the internal quantization error cannot hold its maximum value through N_4 samples, since it would mean that the $\Delta\Sigma$ loop does not operate properly.

Instead of finding the worst-case internal quantization error sequence, one may want to use the statistical property of the internal error to estimate the statistical property of the output quantization error.

Let us assume that the internal quantization error has a uniform distribution between $\pm V_{\text{ref}}$, thus $m_\varepsilon = 0$ and $\sigma_\varepsilon^2 = 4V_{\text{ref}}^2/12$. We would like to find out the output error distribution and its properties.

It is known from the central limit theorem (see, e.g., [Weisstein, 2004]), as used already in some analysis, that the distribution of an y, which is the sum of N i.i.d. (independent, identically distributed) x_i random variable is approximately Gaussian, with $m_y = Nm_x$ and $\sigma_y^2 = N\sigma_x^2$. In our case the first and last N_4 samples are uniformly distributed between ± 1, while the middle $2N_4$ samples are uniformly distributed between ± 3. These samples are summed together, thus, the result is the sum of two Gaussian random variable with $m_1 = m_2 = 0$ and

$$\sigma_1^2 = \frac{1}{N_4^8} 2N_4 \frac{4}{12} \tag{4.40}$$

and

$$\sigma_2^2 = \frac{1}{N_4^8} 2N_4 \frac{36}{12}. \tag{4.41}$$

Since the variance of the first and last N_4 samples is one-ninth of that of the middle $2N_4$ samples, the contribution of these samples to the final output is much less significant. Thus, only the samples in the middle can be used to estimate the output. To estimate a lower bound to the maximum output error, one may use the 3-sigma rule, since it is very unlikely that the output quantization error is greater than $3\sigma_2$. This may be equal to half LSB of the target resolution:

$$3\sigma_2 = \frac{3}{N_4^4}\sqrt{6N_4} < \frac{\text{LSB}}{2} = \frac{U_{\max}}{2^{n_{\text{bit}}}}. \tag{4.42}$$

Rearranging this equation, one can get an estimate of the required number of samples:

$$N_4 > \sqrt[3.5]{\frac{3\sqrt{6}\cdot 2^{n_{\text{bit}}}}{U_{\max}}}, \tag{4.43}$$

while if the internal converter has l levels, then $\sigma_\varepsilon^2 = 4/(12(l-1)^2)$, thus

$$N_4 > \sqrt[3.5]{\frac{3\sqrt{6}\cdot 2^{n_{\text{bit}}}}{U_{\max}(l-1)}}, \tag{4.44}$$

Calculating the required number of samples for 16-bit precision, with $l = 2$ and $U_{\max} = 0.67$, $N_4 = 48$, i.e., N=192, while with a 33-level internal quantizer $N_4 = 18$, N=72. These numbers shows that the required number of samples N dropped to about the half of the case when third-order filter was used.

A more precise derivation takes into account the effect of all filter coefficients. In this case the distribution of the filtered signal still may be modeled with Gaussian distribution, but its variance becomes

$$\sigma'^2 = \sigma_1^2 + \sigma_2^2 = \frac{1}{N_4^8} 2N_4 \frac{40}{12}. \tag{4.45}$$

Using again the 3-sigma rule yields to

$$N_4' > \sqrt[3.5]{3\sqrt{\frac{20}{3}}\frac{2^{n_{\text{bit}}}}{U_{\max}}}, \tag{4.46}$$

which is

$$\frac{N_4'}{N_4} = \sqrt[3.5]{\frac{\sqrt{\frac{20}{3}}}{\sqrt{6}}} \approx 1.01 \tag{4.47}$$

times greater than our previous result. This means that the contribution of the smaller coefficients are negligible according to our previous assumption, causing a maximum increase of 1% compared to the previously calculated N_4 and N.

Note that there is a significant difference in the assumptions used for applying third- or lower order and fourth-order sinc-filter for the calculation of the digital output. In the case of third-order sinc-filter only the following properties of $\varepsilon[k]$, the internal quantization noise were assumed:

- $\varepsilon[k] \in (-V_{\text{ref}}/(l-1), V_{\text{ref}}/(l-1))$;
- $\varepsilon[k]$ and $\varepsilon[k-N_3]$ are uncorrelated;
- $\varepsilon[k]$ and $u[k]$ are uncorrelated.

The first two of these properties are automatically fulfilled under normal operation, while the third one may be easily achieved by injecting a small dither signal into the loop or also automatically fulfilled in practical (noisy) circuits and inputs.

However, for the estimation of N_4, the decimation ratio of the fourth-order sinc-filter, $\varepsilon[k]$ must satisfy much serious conditions:

- $\varepsilon[k]$ is uniformly distributed between $\pm V_{\text{ref}}/(l-1)$;
- Neighboring samples of $\varepsilon[k]$ are uncorrelated;
- $\varepsilon[k]$ and $u[k]$ are uncorrelated.

Here the first two properties are much harder to achieve, e.g., with more intensive dithering.

In addition, in the fourth-order case the output quantization error maximum was set to 3σ, which gives only a probability limit to the output quantization error. It is shown in the following section that this limit is not strict enough and one has to use an upper limit of 5σ for proper quantization error. In the case of using third-order digital filter, the maximum error limit was based on the worst-case internal quantization error sequence, thus it is guaranteed that the final output error will below the half LSB error.

Thus, Eqs. (4.43) and (4.44) can be used only as estimation for the required number of cycles. The final N must be set by using simulations with different input signals.

Using even higher-order filtering ($L_d > L_a + 1$) is not suggested. For example, using fifth-order filter with third-order $\Delta\Sigma$ loop causes double-integration of the individual samples, which gives worse estimate on the average value of the samples than single-integration (i.e., when the individual samples are simply summed together). Thus, $L_d > L_a + 1$ would require more number of cycles than $L_d = L_a + 1$.

Simulation Results

The theoretical results discussed above have been verified by simulations. A third-order $\Delta\Sigma$ modulator with pure L_ath-order differential *NTF* has been designed. To make the modulator stable, an internal quantizer with $l = 33$ levels have been utilized, and to make the input signal transfer function equal to one, a feed-forward input signal path was also used. The model of the modulator is shown in Fig. 4.11. Here a third-order sinc-filter realized by the Hogenauer-structure (Cascaded-Integrator-Comb, i.e., CIC-filter, [Hogenauer, 1981]) is also shown.

Three different cases were simulated for $N_0 = 2^{18}$ different input signal between $\pm U_{\max}$. The output quantization error for second-, third-, and fourth-order sinc-filter are shown in Figs. 4.12, 4.13 and 4.14, respectively.

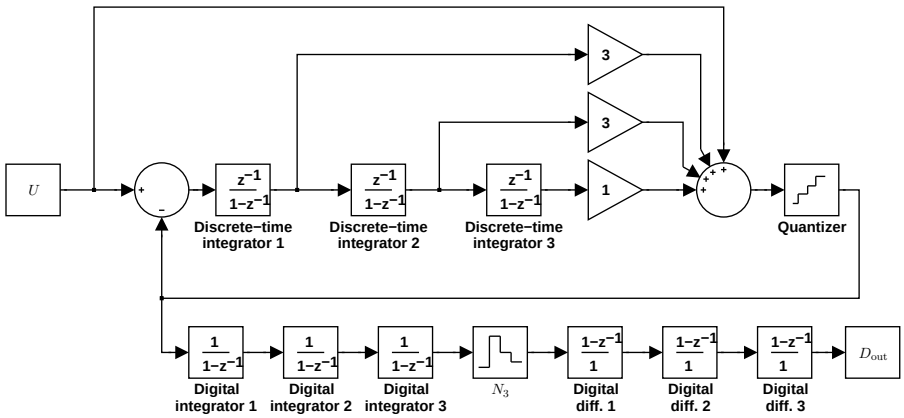


Figure 4.11: Third-order converter with pure third-order differential NTF. Here a third-order sinc-filter following the modulator is also shown.

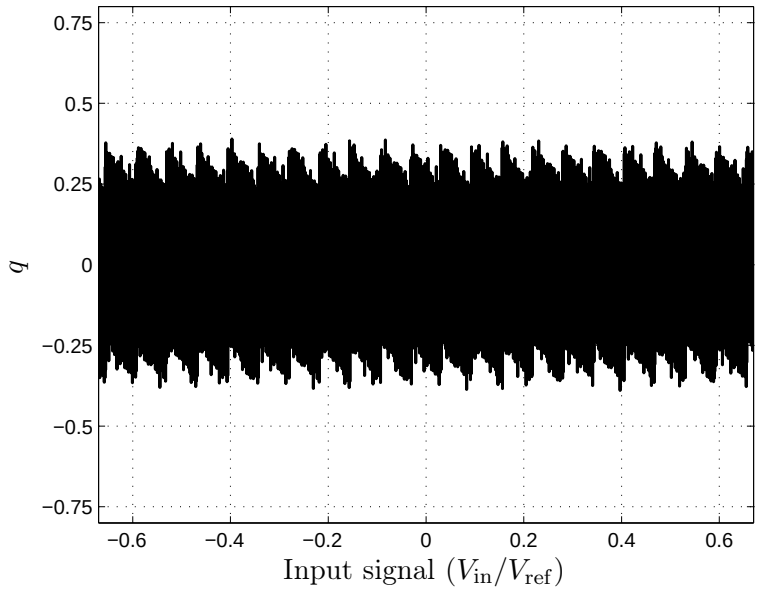


Figure 4.12: Quantization error of a third-order modulator with second-order sinc-filter. $n_{\mathrm{bit}} = 14$, $N_2 = 79$, $N = 159$, $U_{\max} = 0.67$, $l = 33$.

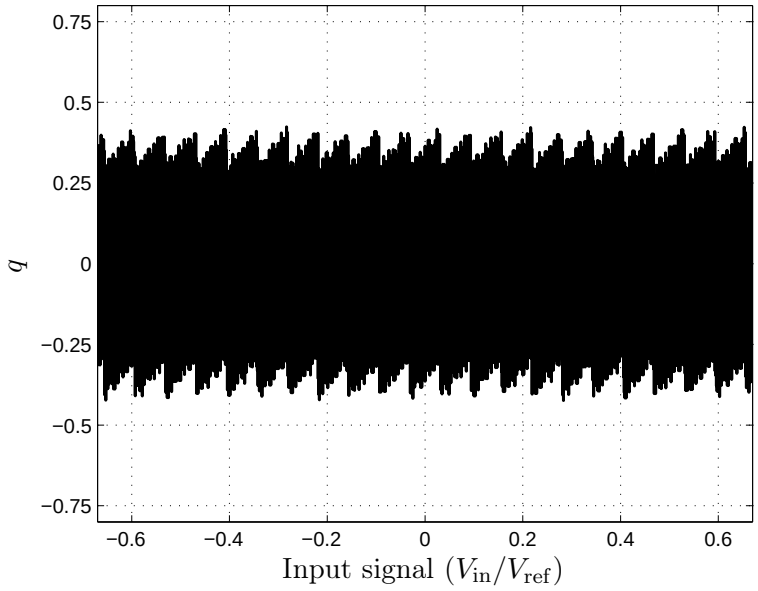


Figure 4.13: Quantization error of a third-order modulator with third-order sinc-filter. $n_{\text{bit}} = 14$, $N_3 = 19$, $N = 58$, $U_{\max} = 0.67$, $l = 33$.

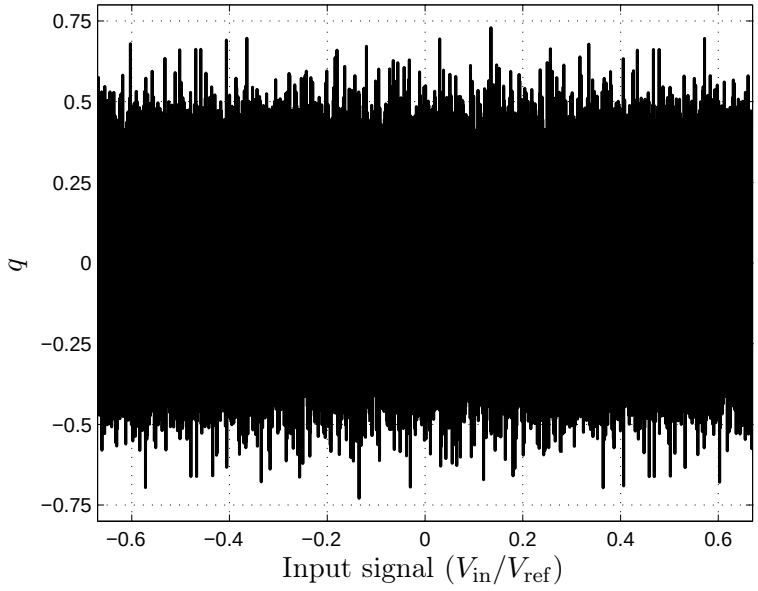


Figure 4.14: Quantization error of a third-order modulator with fourth-order sinc-filter. $n_{\text{bit}} = 14$, $N_4 = 12$, $N = 49$, $U_{\max} = 0.67$, $l = 33$.

The required number of cycles were estimated by the equations derived in the previous section. N was estimated from the worst-case quantization error sequence for the second- and third-order filter, while for the fourth-order filter the estimation of N was based on the approximate output error probability distribution function.

Analyzing Figs. 4.12–4.14, it can be seen that the estimated number of cycles based on the worst-case internal quantization error sequence (Figs. 4.12–4.13) resulted in a somewhat conservative design, since the absolute value of the output quantization error is always less than 0.5LSB, the allowable maximum error. Nevertheless, in the case of fourth-order sinc-filter, the output error is much greater and in many cases it is actually greater than the allowable 0.5LSB. This indicates that in this case the required number of cycles was underestimated by Eq. (4.44).

This estimation error may have two reasons: the first is that the assumption about the property of the internal quantization error are not valid, the second is that the 3-sigma rule is not strict enough, since it allows the quantization error to be greater than 0.5 LSB, only its probability is less than or equal to 0.3%. Calculating the ratio of the overshooting errors and all the errors, it turns out that the probability of the error being greater than the quantization error is $p = 0.2\%$, indicating that the second problem is the dominant, i.e., the 3-sigma rule is not strict enough. One can also see that using a 5-sigma rule would provide an output, whose maximum error is in ± 0.5LSB.

Recalling Eq. (4.44), the derivation of this modified N_4'' is similar, except that the 3 in the expression changes to 5. This yields to the following result

$$N_4'' = \sqrt[3.5]{\frac{5\sqrt{6} \cdot 2^{n_{\text{bit}}}}{U_{\max}(l-1)}}, \tag{4.48}$$

which increases N_4 by $\sqrt[3.5]{5/3} = 1.16$ times of the original N_4, resulting in a 16% increase of the total required number of cycles, too. Taking the example above, N_4'' for 14-bit resolution becomes 14, thus $N = 56$, which is very close to N calculated for third-order sinc-filter. However, for 20-bit precision, $N = 184$ for fourth-order sinc-filter and $N = 222$ for third-order sinc-filter. The difference is much higher using one-bit internal converter, discussed in the next section. A simulation example is shown in Fig. 4.15 using the 5-sigma rule. It can be seen that the final output error is within half LSB of the target resolution of the converter.

The consequences of these simulations are as follows:

- Estimation of the required number of cycles based on Eqs. (4.33) and (4.36) for second- and third-order sinc-filters gives somewhat conservative result.
- (4.44) underestimates the required number of cycles for fourth-order sinc-filter, but using 5-sigma rule gives more adequate result.
- For low precision with multi-bit internal quantizer, it is not advantageous to use $L_d = L_a +$ 1st-order sinc-filter.
- In a real design the calculated number of cycles must always be verified by simulation and/or experimental measurements.

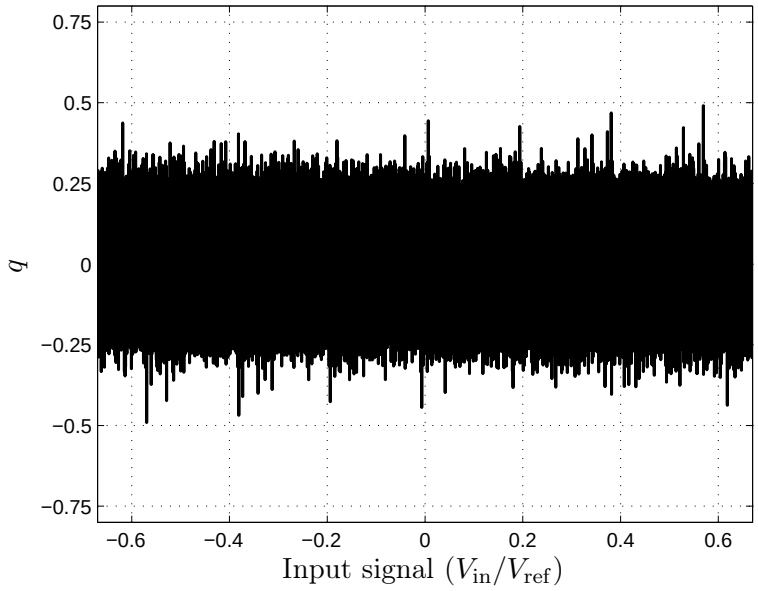


Figure 4.15: Quantization error of a third-order modulator with fourth-order sinc-filter, using 5-sigma rule to estimate N_4. $n_{\text{bit}} = 14$, $N_4 = 14$, $N = 57$, $U_{\max} = 0.67$, $l = 33$.

4.2.2 CIFF Modulators with Stabilized Noise Transfer Function

In today's $\Delta\Sigma$ design, usually the one-bit internal quantizer is preferred to multi-bit one, due to its inherent linearity, easier realization, low power consumption, etc., even though there exist more and more efficient methods to reduce the linearity error of the multi-bit feedback DAC (efficient method for incremental converter will be discussed in Sec. 4.3.7). In this section the required number of cycles for converters with one-bit internal quantizer followed by a digital sinc-filter will be calculated. Here only the same-order sinc-filter ($L_d = L_a$) and the higher-by-one order sinc-filter ($L_d = L_a + 1$) are considered due to the reasons explained in the previous section.

To ensure stability, in higher-order one-bit converters poles are introduced in the NTF, which control the maximum gain in the loop and also in the NTF. Due to these poles, Eq. (4.21) (i.e., the output of the $\Delta\Sigma$ modulator) changes accordingly:

$$Y(z) = U(z) + \frac{(1 - z^{-1})^{L_a}}{D(z)} E(z), \tag{4.49}$$

where $D(z)$ represents the poles of the NTF. Usually these poles are arranged in a Butterworth low-pass configuration.

Due to the presence of the poles, the impulse response from the internal quantizer to the output of the digital filter is not finite (FIR), but infinite (IIR). Thus, models used previously are invalid for this situation, thus, new models have to be developed.

Digital Sinc-filter with Order $L_d = L_a$

Let us consider first a third-order modulator with third-order sinc-filter ($L_d = L_a$). In this case, the normalized transfer function from the internal quantizer to the output of the digital filter becomes

$$w_{3,p}(z) = \frac{1}{N_{3,p}^3} \frac{(1 - z^{-N_{3,p}})^3}{D(z)}, \tag{4.50}$$

where the subscript p denotes that the system has poles. The filter's transfer function is a product of the pure third-order system (Eq. (4.34)) and the filter $1/D(z)$, which is an IIR low-pass filter. In the time domain, the convolution of the two impulse responses determines the filter transient, thus the required operation length.

Even though the filter is an IIR-filter, the required number of cycles is not infinite, since the filter impulse response is fading out exponentially due to the Butterworth low-pass configuration. Thus, one may calculate an amplitude-limit under which the impulse response becomes negligible. Fig. 4.16 shows the impulse response of $1/D(z)$ in linear and dB scale of a third-order modulator (Fig. 4.17) used already in the previous discussions. Here the illustrated transfer function is

$$\frac{1}{D(z)} = \frac{1}{1 - 2.2z^{-1} + 1.689z^{-2} - 0.4444z^{-3}}, \tag{4.51}$$

and its pole-zero equivalent is

$$\frac{1}{D(z)} = \frac{1}{(1 - 0.6694z^{-1})(1 - 1.531z^{-1} + 0.6639z^{-2})}. \tag{4.52}$$

One can see that the impulse response becomes negligible after the first 20-30 samples. Note that this limit depends also on the required resolution.

To find out the required number of samples, the method used in the previous section (estimating the output by worst-case error sequence) cannot be used here, since the filter transient response is the convolution of $D(z)$ and the FIR output (cf. Fig. 4.8), thus many samples are summed together in this configuration. Instead, the method used in the analysis of the one-bit CIFF modulator (cf. Sec. 3.2.3) can also be utilized here. There the output of the last integrator in the loop was used to limit the output quantization error, since

$$|V_3[k]| \leq V_{\text{ref}} \tag{4.53}$$

must hold for normal operation.

It was also shown in Sec. 3.2.3 that in a CIFF modulator with feed-forward input signal path (such as the one shown in Fig. 4.17), the output of the last integrator contains only processed internal quantizer error, i.e.,

$$\frac{V_3(z)}{V_{\text{ref}}} = -\frac{bc_1c_2}{D(z)} E(z), \tag{4.54}$$

where $1/D(z)$ is a Butterworth low-pass filter (cf. Fig. 3.18(c)), and b and c_i are scaling coefficients which ensure the validity of Eq. (4.53) (cf. Sec. 3.2.3).

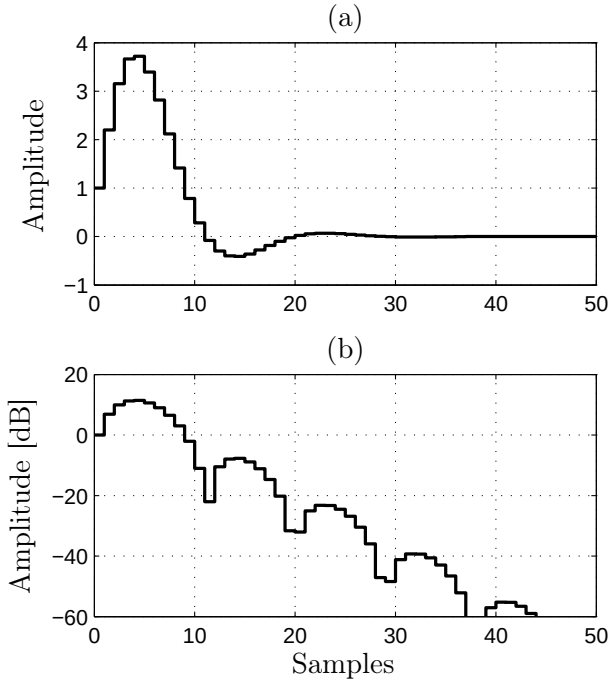


Figure 4.16: Impulse response of a third-order modulator's denominator in (a) linear and (b) logarithmic scale.

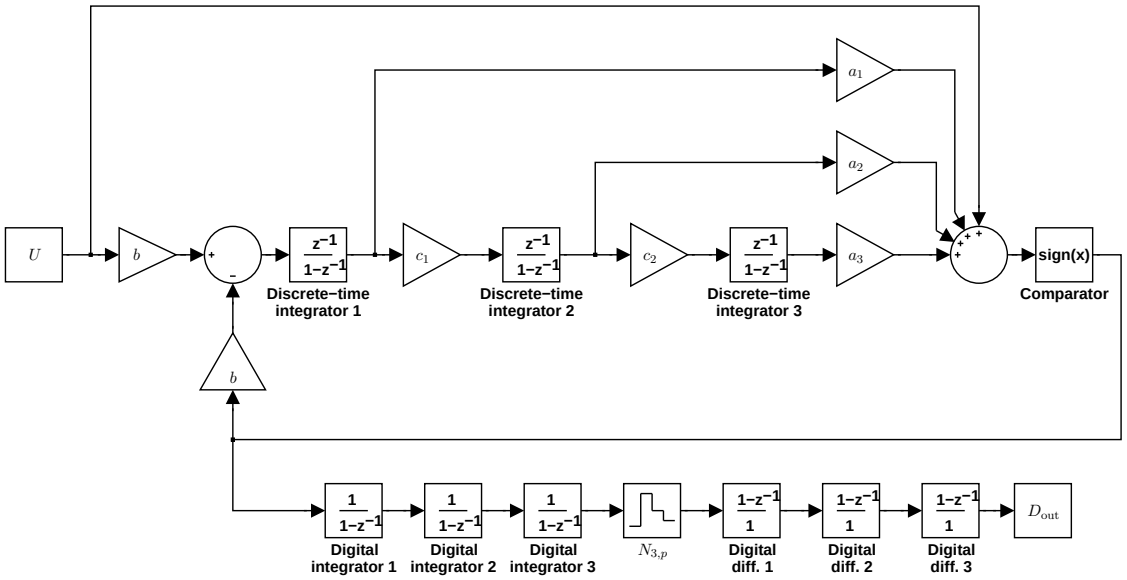


Figure 4.17: Third-order converter with one-bit internal quantizer and stabilized third-order *NTF*. Here a third-order sinc-filter following the modulator is also shown.

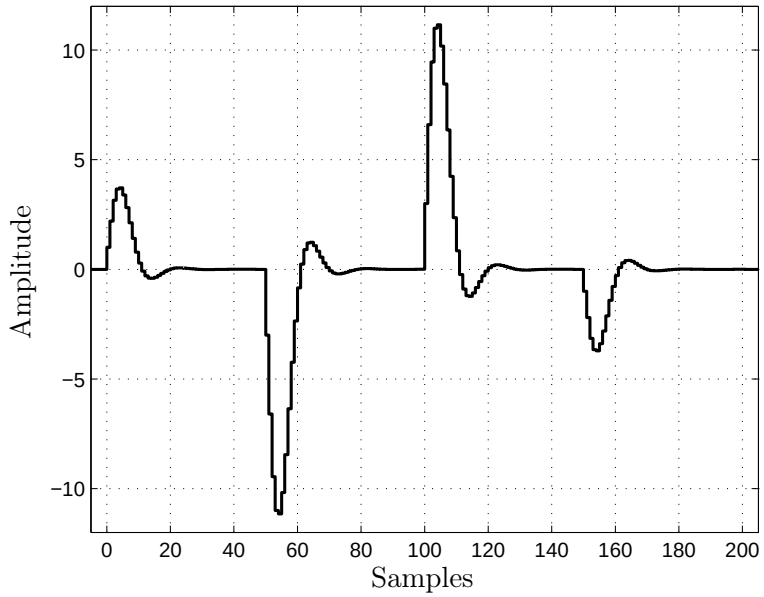


Figure 4.18: Total impulse response of the merged NTF with poles and third-order sinc filter without scaling for $N_{3,p} = 50$.

According to these equations,

$$\left|\frac{1}{D(z)}E(z)\right| < \frac{1}{bc_1c_2}, \tag{4.55}$$

thus, the output quantization error satisfies

$$|Q(z)| = |w_{3,p}(z)E(z)| = \left|\frac{1}{N_{3,p}^3}\frac{(1-z^{-N_{3,p}})^3}{D(z)}E(z)\right| < \left|\frac{1}{N_{3,p}^3}\frac{(1-z^{-N_{3,p}})^3}{bc_1c_2}\right| \tag{4.56}$$

Note that even though this derivation does not contain the number of levels of the internal quantizer directly, the number of levels does have influence on the required number of cycles, since it has a direct relationship with the scaling coefficients b and c_i.

The transfer function $(1-z^{-N_{3,p}})^3$ has already been analyzed in the previous section (cf. Fig. 4.8 and Eq. (4.34)). Here there is one more condition to be satisfied for easy calculation: if the required resolution is high enough, then $N_{3,p}$ is longer than the impulse response of the IIR-filter, thus the convolution of the two filter responses simplifies to the sum of 4 independent IIR-filter impulse-responses (as illustrated in Fig. 4.18). Then, the maximum error can be estimated as

$$|Q(z)| < \left|\frac{1}{N_{3,p}^3}\left(\frac{E(z)}{D(z)} - 3\frac{E(z)}{D(z)} + 3\frac{E(z)}{D(z)} - 1\frac{E(z)}{D(z)}\right)\right| < \frac{1}{N_{3,p}^3}\frac{8}{bc_1c_2}. \tag{4.57}$$

Note that this is a very conservative estimation, since it assumes that $E(z)/D(z)$ takes its maximum and minimum value, when the FIR-filter's coefficient $(1, -3, 3, -1)$ is positive and negative, respectively. Since this is very unlikely, the result is expected to overestimate

the required number of cycles.

This limit equals to the half LSB of the converter:

$$\frac{1}{N_{3,p}^3}\frac{8}{bc_1c_2} = \frac{U_{\max}}{2^{n_{\text{bit}}}} \tag{4.58}$$

from which the required number of samples for a given resolution can be calculated as

$$N_{3,p} = \sqrt[3]{\frac{2^{n_{\text{bit}}+3}}{bc_1c_2U_{\max}}} \tag{4.59}$$

According to Fig. 4.18, if the required resolution is high enough, then $N_{3,p}$ is much greater than the transient of the IIR-filter. In this case, the total number of cycles the converter must be operated is $N = 3N_{3,p} + m$, where m is the length of the transient of $1/D(z)$, the IIR-filter of the delta-sigma loop.

For 14-bit resolution in the case of 2-level internal quantizer, with $U_{\max} = 0.67$ and coefficients listed in Tab. 3.4, $N_{3,p} = 128$ is required, resulting in a total required number of cycles $N = 3N_{3,p} + m = 414$. For 20-bit resolution with the same other parameters, $N_{3,p} = 514$, $N = 1570$.

Digital Sinc-filter with Order $L_d = L_a + 1$

Similarly to the pure differential L_ath-order $\Delta\Sigma$ modulator, the third-order one-bit $\Delta\Sigma$ modulator may also be followed by a fourth-order sinc-filter. In this section the required number of cycles for an incremental $\Delta\Sigma$ converter consists of a third-order one-bit modulator and fourth-order sinc-filter is examined. In this case the digital filter's output becomes

$$D_{\text{out}}(z) = \frac{1}{N_{4,p}^4}\left(\frac{1-z^{-N_{4,p}}}{1-z^{-1}}\right)^4 Y(z) = U(z) + \frac{1}{N_{4,p}^4}\frac{(1-z^{-N_{4,p}})^4}{D(z)(1-z^{-1})}E(z). \tag{4.60}$$

In Sec. 4.2.1, the impulse response of the fourth-order filter was derived from that of the third-order by integrating the pulses through $N_{4,p}$ samples. This case is somewhat different, since the impulse response of the FIR-filter is convoluted by that of the IIR part of the filter. The FIR-filter impulse response is

$$w_{\text{FIR}}[k] = \epsilon[k] - 4\epsilon[k-N_{4,p}] + 6\epsilon[k-2N_{4,p}] - 4\epsilon[k-3N_{4,p}] + \epsilon[k-4N_{4,p}], \tag{4.61}$$

where $\epsilon[k]$ is the step-function. Convoluting this response with that of the IIR part resulting in a total impulse response shown in Fig. 4.19. Note that the transition peaks are changing according to the coefficients in Eq. (4.61), and the final settling in each sections is 1, -3, 3, -1 and 0 times the step response settling of the IIR-filter. The length of the total filter impulse response is $4N_{4,p} + m$.

Similarly to Sec. 4.2.1, in this case it is not advantageous to approximate the internal quantization error with its maximum, since many samples are averaged. Instead, one can

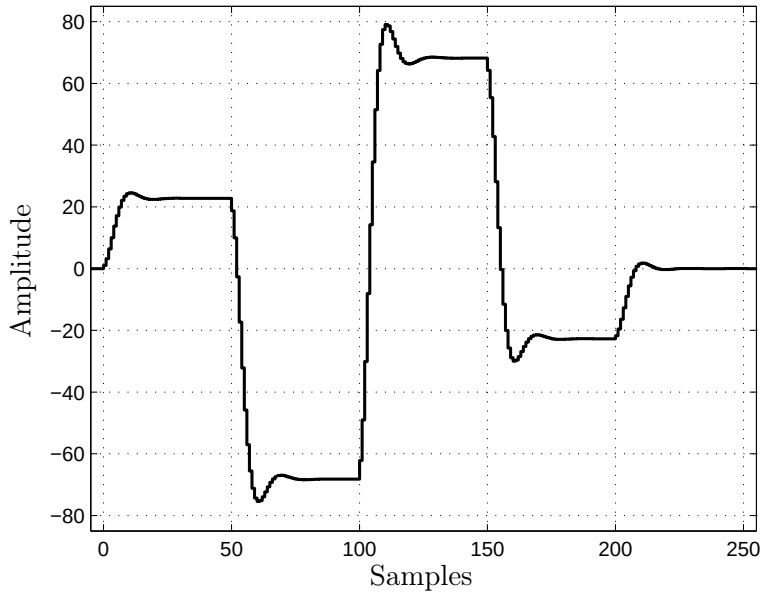


Figure 4.19: Total impulse response of the merged NTF with poles and fourth-order sinc filter without scaling for $N_{4,p} = 50$.

utilize the fact that the output of the last integrator,

$$\frac{V_3(z)}{V_{\text{ref}}} = -\frac{bc_1c_2}{D(z)}E(z), \tag{4.62}$$

is a stochastic variable with approximately Gaussian distribution (cf. Fig. 3.19(c)). Using a k_1-sigma rule, where $k_1 \in [3, 5]$, the standard deviation of this signal may be estimated as

$$\sigma_{V_3} = \frac{V_{\text{ref}}}{k_1}, \tag{4.63}$$

i.e., the standard deviation of the signal $1/D(z)E(z)$ is

$$\sigma_{E(z)/D(z)} = \frac{1}{bc_1c_2k_1}. \tag{4.64}$$

Assuming that the internal quantization error $E(z)$ is uncorrelated, its variance can be calculated as

$$\sigma_\varepsilon^2 = \frac{\sigma_{E(z)/D(z)}^2}{\sum_{i=1}^{m} w_d[i]^2} = \frac{1}{(bc_1c_2)^2k_1^2 \sum_{i=1}^{m} w_d[i]^2}, \tag{4.65}$$

where $w_d[i]$ is the ith element of the impulse response of the filter $1/D(z)$.

This signal, the internal quantization error is filtered by the filter

$$NTF_{\text{total}}(z) = \frac{1}{N_{4,p}^4}\frac{(1 - z^{-N_{4,p}})^4}{D(z)(1 - z^{-1})}, \tag{4.66}$$

to get to the digital output. This filter is the convolution of the IIR-filter $1/D(z)$ and

the FIR $(1-z^{-N_{4,p}})^4/(1-z^{-1})$ (cf. Fig. 4.19). To find the required number of cycles, the convolution of the two numerical impulse response should be calculated. However, an approximate method also exists: if $N_{4,p} >> m$ (which is true for high-resolution converters), then the steady-state part in each section dominates over that of the transient. Then, the filter coefficients in one $N_{4,p}$-long section may be estimated as $w_{\mathrm{FIR}}[k]v_{\mathrm{IIR}}[\infty]$, where $w_{\mathrm{FIR}}[k]$ is 1, -3, 3 and -1 for $k \in [0 : N_{4,p}-1]$, $[N_{4,p} : 2N_{4,p}-1]$, $[2N_{4,p} : 3N_{4,p}-1]$ and $[3N_{4,p} : 4N_{4,p}-1]$, respectively and $v_{\mathrm{IIR}}[\infty]$ is the settling of the step response of the IIR-filter $1/D(z)$. $v_{\mathrm{IIR}}[\infty]$ can be easily calculated, since it is the sum of the samples of the impulse response, which equals to the transfer function at dc ($z=1$).

Thus, the output variance of the filter can be approximated as

$$\sigma_{g,4,p}^2 = \frac{1}{N_{4,p}^8}\sigma_\varepsilon^2 N_{4,p}(1^2+(-3)^2+3^2+(-1)^2)v_d[\infty]^2 = \frac{1}{N_{4,p}^7}\frac{20}{(bc_1c_2)^2k_1^2}\frac{\left(\sum_{i=1}^{m} w_d[i]\right)^2}{\sum_{i=1}^{m} w_d[i]^2}, \quad (4.67)$$

i.e., its standard deviation

$$\sigma_{g,4,p} = \frac{1}{N_{4,p}^{3.5}}\frac{\sqrt{20}}{bc_1c_2k_1}\sqrt{\frac{\left(\sum_{i=1}^{m} w_d[i]\right)^2}{\sum_{i=1}^{m} w_d[i]^2}}. \quad (4.68)$$

Again, as the output error distribution becomes Gaussian, one may use a k_2-sigma rule ($k_2 \in [3,5]$) to determine a lower bound for the maximum possible error and make it equal to half LSB:

$$k_2\sigma_{g,4,p} \le \frac{U_{\max}}{2^{n_{\mathrm{bit}}}}. \quad (4.69)$$

Substituting Eq. (4.68) into Eq. (4.69), and rearranging the given equation, one can get the required number of cycles for a given resolution as follows:

$$N_{4,p} \ge \sqrt[3.5]{\frac{k_2}{k_1}\frac{2^{n_{\mathrm{bit}}}\sqrt{20}}{bc_1c_2U_{\max}}\sqrt{\frac{\left(\sum_{i=1}^{m} w_d[i]\right)^2}{\sum_{i=1}^{m} w_d[i]^2}}} \quad (4.70)$$

Since $k_1 \in [3,5]$ and $k_2 \in [3,5]$ and their value is selectable, selecting $k_1 = k_2$ simplifies the equation somewhat, thus the required number of samples for a given resolution is $N = 4N_{4,p} + m$, where m is the length of the transient of the IIR-filter, while

$$N_{4,p} \ge \sqrt[3.5]{\frac{2^{n_{\mathrm{bit}}}\sqrt{20}}{bc_1c_2U_{\max}}\sqrt{\frac{\left(\sum_{i=1}^{m} w_d[i]\right)^2}{\sum_{i=1}^{m} w_d[i]^2}}}. \quad (4.71)$$

In the case of a third-order CIFF modulator with $U_{\max} = 0.67$, scaling coefficients listed in Tab. 3.4 and Butterworth pole-configuration of Eqs. (4.51) and (4.52), $U_{\max} b c_1 c_2 = 0.0618$ and $\left(\sum_{i=1}^{m} w_d[i]\right)^2 / \sum_{i=1}^{m} w_d[i]^2 = 7.34$, thus

$$N_{4,p} \approx 4.52 \cdot 2^{\frac{n_{\text{bit}}}{3.5}} \tag{4.72}$$

For 14-bit resolution in the case of 2-level internal quantizer, with $U_{\max} = 0.67$, $N_{4,p} = 73$ is required, resulting in a total required number of cycles $N = 4N_{4,p} + m = 322$. For 20-bit resolution with the same other parameters, $N_{4,p} = 238$, $N = 982$.

Note that since this derivation is based on the statistical properties of the internal quantization error, these statistical properties (limited in amplitude and uncorrelated with itself) must be at least approximately satisfied to make the results valid. Fortunately, these properties are more or less fulfilled in a higher-order $\Delta\Sigma$ converter.

Simulation Results

Based on the theoretical derivations, a one-bit third-order converter (Fig. 4.17) was simulated with third- and fourth-order sinc-filter. The required number of cycles was calculated by Eqs. (4.59) and (4.71) for third- and fourth-order sinc-filter, respectively.

One problem in the simulation of these converters is that the minimum required number of cycles is calculated as $N = 3N_{3,p}+m$ and $N = 4N_{4,p}+m$ for third- and fourth-order filter, respectively, where $N_{3,p}$ and $N_{4,p}$ are the decimation ratios of the third- and fourth-order sinc-filter, respectively, and m is the length of the impulse response of the Butterworth lowpass filter in the modulator. However, if the sinc-filter is realized by the Hogenauer-structure, we do not have access to every output sample, since the output is the decimated signal, i.e., every $N_{3,p}$th or $N_{4,p}$th sample are available only. To solve this problem, two method can be used: one is to operate the converter over $N = 4N_{3,p}$ or $N = 5N_{4,p}$ cycles to make sure that the transient is over, however, this results in more cycles than the minimum required. As this method can be easily realized, this was used in the simulations. The alternative way is to delay the operation of the sinc-filter by m.

Figs. 4.20 and 4.21 shows the output quantization error. It can be seen that Eq. (4.59) overestimates the required number of samples (the final quantization error is much smaller than half LSB), since it is based on a worst-case internal quantization error sequence. Nevertheless, Eq. (4.71) gives good estimation for $N_{4,p}$.

4.2.3 Optimized Line Frequency Suppression

Converters designed for high-accuracy dc measurement often require suppression of the line frequency (f_l = 50 or 60 Hz). For dual-slope converters and similarly, first-order incremental converters, this can be achieved by setting the time interval of the incoming signal's integration to be an integer multiple of $1/f_l$.

As discussed in the previous sections, sinc-filters can also be designed for higher-order modulators to provide line-frequency noise suppression. To achieve this, one of the notches

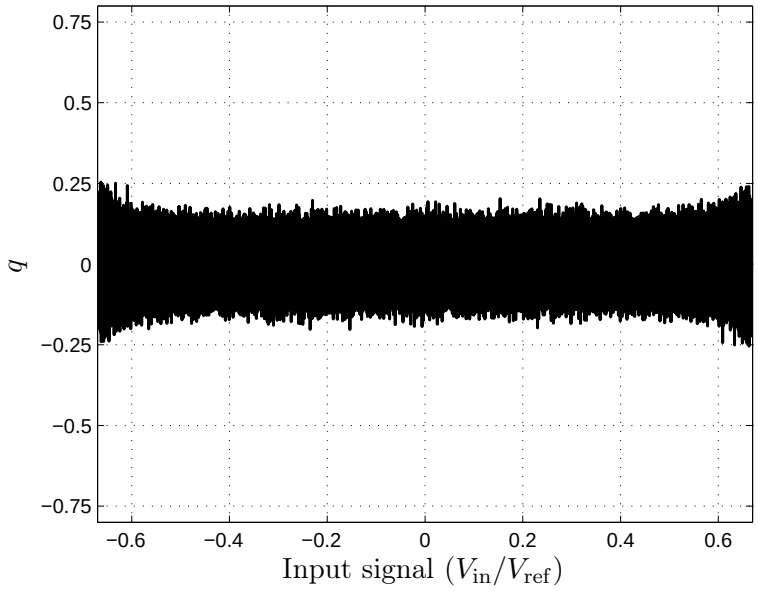


Figure 4.20: Quantization error of a stabilized one-bit third-order modulator with third-order sinc-filter. $n_{\text{bit}} = 14$, $N_{3,p} = 128$, $N = 414$, $U_{\text{max}} = 0.67$.

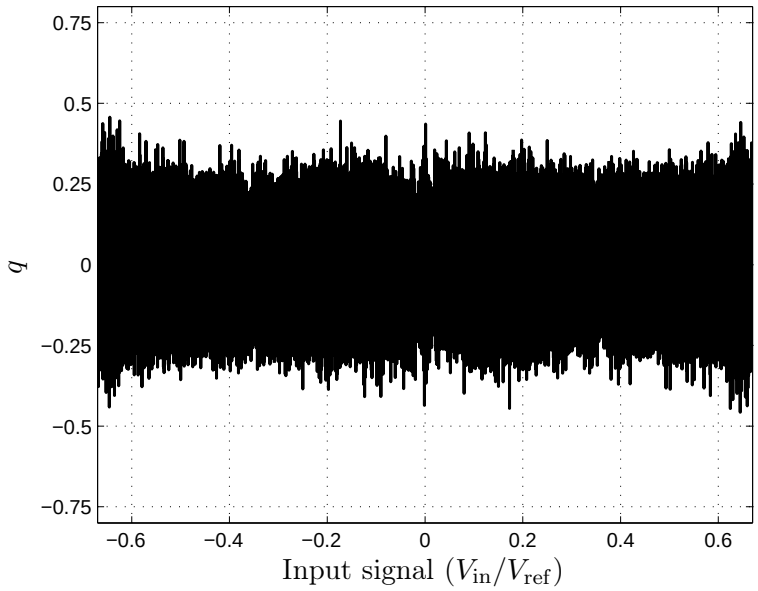


Figure 4.21: Quantization error of a stabilized one-bit third-order modulator with fourth-order sinc-filter. $n_{\text{bit}} = 14$, $N_{4,p} = 55$, $N = 250$, $U_{\text{max}} = 0.67$.

of the filter must coincide with the line frequency f_l. This gives the condition

$$f_s = N_i \frac{f_l}{k} \quad k = 1, 2, \ldots, M-1, \tag{4.73}$$

where N_i is the decimation ratio of the ith-order filter, f_l is the line frequency and f_s is the sampling rate of the modulator.

To make the gain response as flat as possible at low frequencies, and also to obtain reasonably high sampling rate for the analog portion of the circuit to reduce inband thermal noise, normally $k = 1$ is chosen in Eq. (4.73).

In critical applications the suppression available by using straight sinc-filters may not be adequate, especially if the line frequency and/or the on-chip oscillator frequency is inaccurate. In this case the zeros of the sinc-filter can be staggered around f_l, thus widening the frequency range where the rejection is high.

To modify the zeros of the filter, the rotated sinc-filter (RS-filter) introduced by Lo Presti [Presti, 2000] may be used. A second-order factor of its transfer function is of the form

$$H_{\mathrm{dec}}(z) = \frac{1 - 2(\cos N_i\alpha)z^{-N_i} + z^{-2N_i}}{1 - 2(\cos\alpha)z^{-1} + z^{-2}}, \tag{4.74}$$

where $z = e^{j2\pi f/f_s}$, N_i is the decimation ratio of the ith-order sinc-filter and α represents the angle of the modified complex conjugate zeros. If $\alpha = 0$, the expression simplifies to the transfer function of a second-order classical sinc-filter.

If the required suppression is given in a region (say $f_l \pm 5\%$), one can optimize the order of the sinc-filter and the number of RS second-order filter blocks to achieve the given suppression [Presti, 2000].

Since the required frequency range is usually small compared to the line frequency, α is usually also small. Thus, $2(\cos N_i\alpha)$ and $2(\cos\alpha)$ can be implemented as $2 - n_1$ and $2 - d_1$, respectively. Here, the small quantities n_1 and d_1 can be chosen as negative powers of 2. It can be also shown that $n_1 = N_i^2 d_1$. Detailed discussion of this technique can be found in [Presti and Akhdar, 1998].

As $n_1 = N_i^2 d_1$, in cases when N_i is high (say $N_i = 256$), the required register-width may become excessive. In these cases, two-stage decimation may reduce the required precision. The first stage can have a high oversampling ratio (e.g., $N_{i,1} = 32$) and can be implemented with a straight fourth-order structure. The second stage, which implements the staggered zeros, should have a lower oversampling ratio (e.g., $N_{i,2} = 8$), such that $N_{i,1}N_{i,2} = N_i$. With such low $N_{i,2}$, the coefficients n_1 and d_1 are much easier to implement.

Fig. 4.22 compares the achievable rejection around the line-frequency achieved using various filter configurations. If the required attenuation of the line frequency is, say, -110 dB, then the third-order, fourth-order and modified fourth-order filter can obtain this attenuation in the ranges $f_l \pm 1.5\%$, $f_l \pm 4\%$ and $f_l \pm 6.5\%$, respectively.

Similar technique can be used to suppress both $f_l = 50$ Hz and $f_l = 60$ Hz simultaneously using the same clock frequency, which is particularly useful for circuits intended for international use.

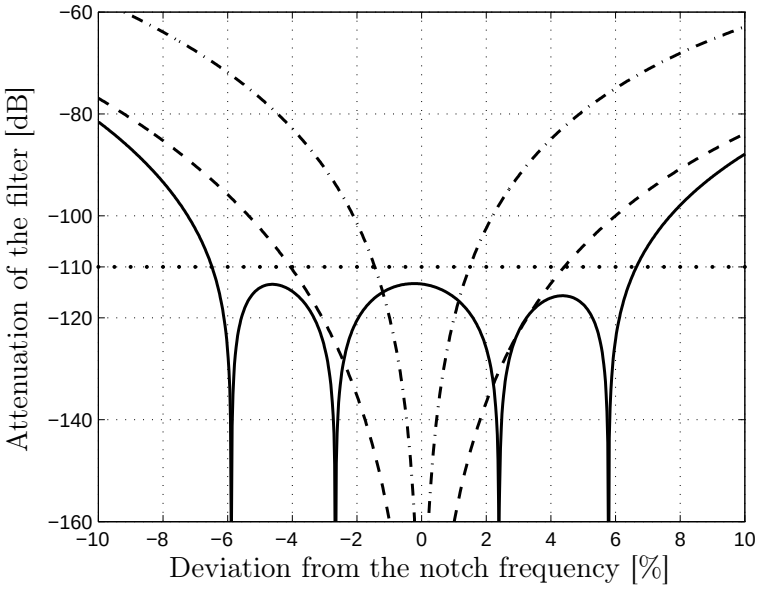


Figure 4.22: Transfer function of different filters around the line frequency: third-order sinc-filter (dash-dot line), fourth-order sinc-filter (dashed line) and fourth-order filter with staggered zeros (solid line).

4.3 Practical Considerations

In the previous sections of this chapter, theoretical results about the operation of the higher-order incremental converters were discussed. Throughout these sections, ideal elements were assumed in the converter, without any noise (except input-related noise), non-linearity, mismatch, etc., which are unavoidable in a real circuit. This section gives an overview about the possible error sources in an implemented converter and gives different circuit-level solution and/or sophisticated algorithms to reduce these errors below the specified level. Here it is assumed that the converter is realized using Switched-Capacitor (SC) technique. A possible realization of a third-order $\Delta\Sigma$ modulator for fully-differential input signal is shown in Fig. 4.23.

4.3.1 Offset and Asymmetry Errors

Since the circuit is intended for dc inputs, offset errors must be kept very small, within an LSB. In addition, charge-injection caused by the non-ideal switches needs to be made signal-independent by properly delaying the operation of floating switches in the modulator [Johns and Martin, 1997, Chap. 10]. Correlated double sampling may be used in the first stage of the analog modulator to reduce its offset [Johns and Martin, 1997, Chap. 10], [Enz and Temes, 1996]. However, asymmetry in the upper and lower halves of the differential circuit can also introduce errors.

Both offset and asymmetry errors can be reduced by using a correction scheme in which the conversion by the $\Delta\Sigma$ loop is performed in two cycles: once with normal inputs, and then with inverted polarity [Robert et al., 1987]. During the second cycle, the output of

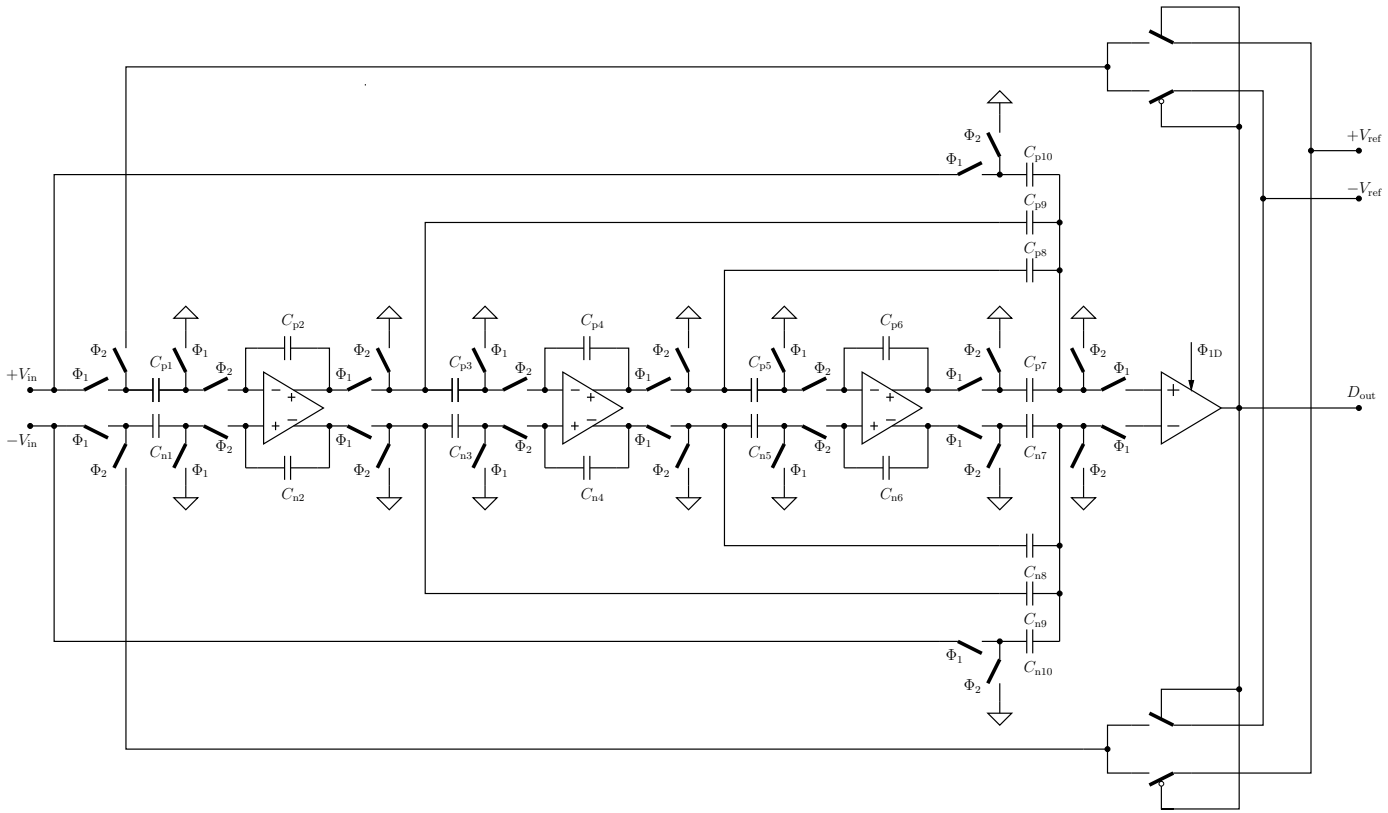


Figure 4.23: A simplified realization of a third-order, one-bit $\Delta\Sigma$ modulator as a fully-differential switched-capacitor (SC) circuit.

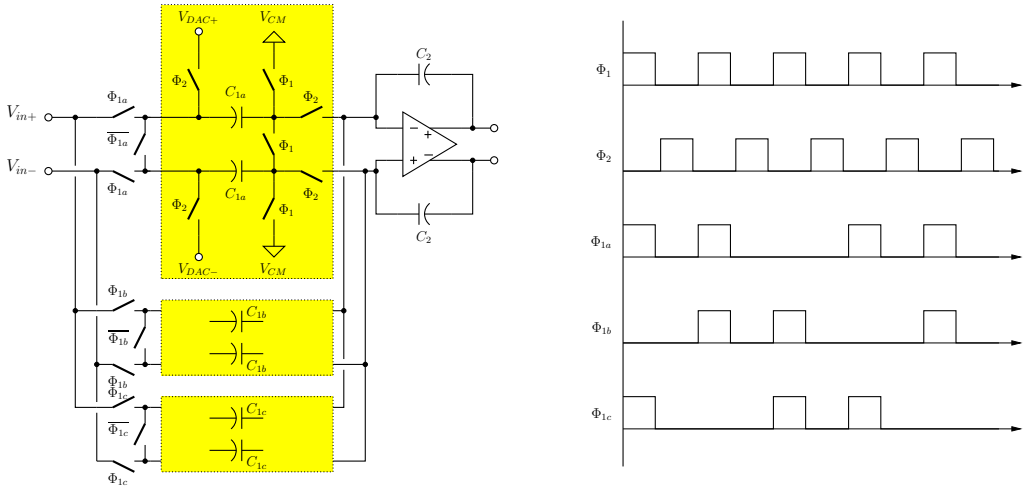


Figure 4.24: The modified (divider) input branch.

the comparator is also complemented. The two output value thus obtained can then be added, and the offset and asymmetry errors will be cancelled. This scheme was used in [Robert et al., 1987] for a first-order loop, and can be adapted for modulators of any order. This method can be easily implemented, as the converter operates in transient mode.

4.3.2 Input Scaling and Gain Error

It is well-known and was also shown earlier that the input signal of higher-order $\Delta\Sigma$ modulators cannot reach the reference signal, because this causes overflow and instability errors. Thus, the input signal must be scaled down. This technique was used throughout the theoretical derivations. Nevertheless, most users of A/D converters expect the input signal to be between $\pm V_{\text{ref}}$. To achieve this, a scaling circuit can be used at the input of the converter to ensure that the input signal is indeed between $\pm V_{\text{ref}}$, but the signal entering into the $\Delta\Sigma$ loop is only a fraction of it (2/3, 3/4, 1/2 or similar). This scaling must be very precise to eliminate possible linearity and gain errors.

In SC implementation this can be achieved by a modified input branch, which delivers a charge of

$$Q_{\text{in}} = C_{\text{in}} \left(\frac{2}{3} V_{\text{in}} - V_{\text{dac}} \right) \tag{4.75}$$

into the input integrator. Fig. 4.24 shows a possible implementation in a differential circuit.

Its operation is as follows. When phase $\Phi_1 = 1$, two of the input capacitors acquire V_{in+} and another two V_{in-}, while the remaining two are connected in parallel and are charged to the difference between the common-mode voltages of V_{in} (and V_{dac}) and the opamp input. When $\Phi_2 \rightarrow 1$, all six capacitors are switched to V_{dac+} or V_{dac-}. The differential input charge will then be given by Eq. (4.75).

Nominally, all six input capacitors are equal to $C_{p1}/3 = C_{n1}/3$. In practice, they will not be perfectly matched, and a dynamic matching scheme can be used to rotate

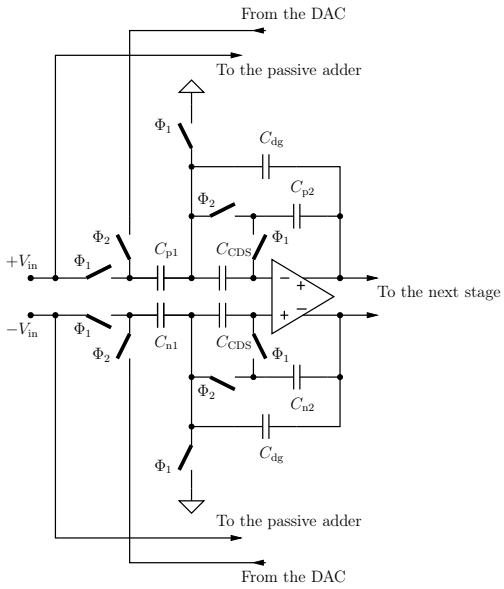


Figure 4.25: Correlated double sampling (CDS) at the input.

their operation such that the average charge delivered exactly satisfies the 2/3 gain factor required for V_{in}. The clock signals needed are shown in Fig. 4.24.

Naturally, it is also possible to introduce any rational scale factor n/m $(m > n)$ by using m capacitors in each input branch in the manner described. The scale factor must be compensated for in the digital domain to restore the validity of the digital output. Also note that the number of total cycles the $\Delta\Sigma$ loop operates must be multiple of $m \cdot n$ to eliminate the mismatch error completely.

This idea has been patented and is discussed in detail in [Temes et al., 2004].

4.3.3 Finite Op-amp Gain and Bandwidth

Finite op-amp gain causes gain and pole error in the integrators. Gain error may be removed by calibration of the converter, while pole error (which causes the integrator to leak) may cause dead-zones in the conversion [Norsworthy et al., 1997].

Since the typical operation frequency of the converter is a few hundred times of the line frequency f_l, the SC circuit switching rate is in the 1–10 kHz range. For proper settling, the op-amp unity-gain bandwidth (UGW) must be about five times this value, which is still less than 100 kHz. For such a low rate it is fairly easy to design a high-gain op-amp (especially in SC circuits, since the op-amp output has only capacitive load). If simple op-amp design techniques cannot make the gain high enough, correlated double-sampling (CDS) op-amp may be used [Enz and Temes, 1996], which is a sophisticated SC-technique, causing the op-amp gain to be squared (thus, it also minimizes offset error). Note that CDS technique is usually used only in the first stage of the $\Delta\Sigma$ loop, since any error introduced in subsequent stages is suppressed by the noise-shaping effect of the first one. One typical example of the realization of a CDS circuit is shown in Fig. 4.25.

4.3.4 $k\mathrm{T}/C$ Noise

In SC circuits, the thermal noise contribution of the finite resistance of the switches depends on the capacitor they switches and its variance is $k\mathrm{T}/C$, where k is the Boltzmann-constant, T is the temperature and C is the value of the capacitor. Only the noise contribution of the first integrator is significant, since the noise in later stages is noise-shaped by the loop.

An advantage of oversampling converters, that due to the decimation, the input-referred noise is averaged during the consecutive cycles. Thus, the noise variance in the output is much smaller than the input-referred noise. This topic was addressed on a theoretical level in Sec. 4.1.1. There it was shown that in the case of third-order converter, the output noise variance

$$\sigma_{y,3}^2 < 1.8\frac{\sigma_g^2}{N}, \tag{4.76}$$

where σ_g^2 is the input-referred noise, N is the number of cycles and $\sigma_{y,3}^2$ is the output variance. One can see that the internal noise is reduced by a factor of N in the final output. If this does not give enough reduction, the converter may be operated for longer cycles than the quantization error would require. In this case, the final signal-to-noise ratio (SNR) will be limited by the analog noise instead of the quantization error.

Another method of reducing the noise contribution is the usage of three-level quantizer. Since in many cases it will provide a feedback of zero, there will be no noise contribution during at least one-third of the feedbacks. This technique is discussed in [Thompson and Bernadas, 1994]. By alternating the feedback capacitors used for positive and negative V_{ref}, one can remove the linearity error caused by capacitor mismatch.

4.3.5 Op-amp Nonlinearity

Op-amp nonlinearity can also degenerate the performance. To reduce this, a low-distortion architecture can be used [Silva et al., 2001; Silva, 2004], which contains a feedforward path for the input signal directly to the quantizer (note that this architecture was used during the derivations, since it has many other advantage). As an additional result, the input signal is not processed by the analog integrators, and hence the effects of op-amp nonlinearities are greatly reduced. To verify the usefulness of this architecture, the reader is referred to [Silva, 2004].

4.3.6 Capacitor Nonlinearity

If high resolution (20 bits or more) is required, it is possible that the linearity of the capacitors available with a given technology cannot satisfy the linearity requirements for the given resolution (note that the circuit is most sensitive to the error of the input sampling capacitor). In this case, simple circuit techniques may be used to alleviate the resulting nonlinear conversion errors. It is possible, for example, to combine two capacitors with opposite polarities in parallel and/or in series to obtain a first-order cancellation of the nonlinear errors.

If such measures are not sufficient to reduce the distortion to acceptable levels, consideration may be given to using multi-bit internal quantization in the $\Delta\Sigma$ loop, as described

in the next section. This will reduce the voltage swing across the input capacitors, and thus reduce the distortion.

4.3.7 Multi-bit Quantization

Using multi-bit quantization (an l-level quantizer and feedback DAC) in the $\Delta\Sigma$ loop has several advantages: it reduces the signal amplitude in the loop, thus reduces capacitor and op-amp nonlinearity error, it reduces the required number of cycles by approximately $\sqrt{l-1}$, and if the quantizer is mid-tread, i.e., in the case of almost zero input it fed back zero, this may reduce also noise (note that this is realization-dependent, e.g., it is true for 3-level quantizers, but not for unit-element feedback DACs). However, the imperfect realization of the A/D and D/A in the loop also introduces errors. The error of the A/D is negligible since its input related error contribution is scaled down by the loop gain (to put it other way, it is noise-shaped by the loop), however, the feedback DAC error has a one-to-one input mapping. Noise is not significant, since it is averaged out by the digital filter, but linearity and gain error may cause errors in the output.

However, the gain error can be eliminated using a two-point calibration of the converter, and the inband mismatch error can be made negligible by using a unit-element DAC incorporating dynamic element matching process such as Data Weighted Averaging (DWA) [Baird and Fiez, 1995]. DWA rotates the usage of the elements, greatly reducing the inband mismatch error.

As the incremental converter works in transient mode, the elimination of inband errors at dc will not be perfect. However, the output error variance can be easily calculated. Consider a feedback capacitor-array DAC with l unit elements (e_1, e_2, $\ldots e_l$), each of them having a relative mismatch error standard deviation of $\sigma_{e_i} = \sigma_e$. With careful layout and design, $\sigma_e = 0.1\%$ may be achieved. In the first cycle, DWA algorithm uses the first k_1 elements (e_1, $\ldots e_{k_1}$), where k_1 depends on the magnitude of the feedback signal. For example, if the feedback signal is $-V_{\mathrm{ref}}$, no capacitor elements are used, if it is 0, $k = l/2$, and if it is full scale (V_{ref}), $k = l$. In the next cycle, the algorithm uses the next k_2 elements, from e_{k_1+1} to $e_{k_1+k_2}$. Naturally, if $k_1 + k_2 > l$, then the usage of the elements starts from the first element again, i.e., the algorithm uses a modulo l arithmetic to switch on the required elements. This means that the elements are used almost equally during a conversion. Whenever all elements are used, the sum of independent error terms causes only a gain error. It can be proven that in a $\Delta\Sigma$ modulator the algorithm transfers the linearity error into first-order shaped noise [Baird and Fiez, 1995].

As the incremental converter operates in transient mode, this first-order shaping will not be perfect. The problem comes from the fact that during the last cycles at the end of the conversion, there will be some elements which are in actual use, while others are not. This can be treated as an input error during the last cycle. The worst-case relative variance of this error is

$$\max \sigma^2_{e,\mathrm{in}} = \frac{l-1}{l}\sigma^2_e, \tag{4.77}$$

i.e., when almost all elements are on. This error term goes into the digital filter in the last sample. However, this value is weighted by a scaling factor, which value depends on

the filter used. In the case of third-order modulator and CoI filters, the scaling factor is $6/(N(N+1)(N+2)$, while in the case of third-order sinc-filter it is approx. $1/(N/3)^3$ and in the case of fourth-order sinc-filter it is approx. $1/(N/4)^4$. This means that the error term is negligible in the output, i.e.,

$$\max \sigma^2_{e,\text{out}} \lesssim \frac{6^2}{N^6} \frac{l-1}{l} \sigma_e^2, \tag{4.78}$$

Using the 3-sigma rule the worst-case output error can be estimated as

$$\max e_{\text{out}} \lesssim \frac{3 \cdot 6}{N^3} \sqrt{\frac{l-1}{l}} \sigma_e \tag{4.79}$$

This means that even if the capacitor mismatch is 5%, $l = 33$ and the number of cycles the converter operates is 300, the output error contribution still allows 24-bit linearity.

Using multi-bit feedback, the required number of cycles can be significantly reduced, thus improving the conversion speed.

Chapter 5

Design Examples

In the previous two chapters, theoretical operation of different incremental $\Delta\Sigma$ converters were analyzed. Modulator structures and digital filter architectures were proposed along with practical considerations. In this chapter, a selection guide is offered for the designers, then the main design equations are repeated along with detailed tables for different applications and architectures. Finally, publicly available data of a 22-bit dc-measuring A/D converter is given, which design was based on the theoretical results discussed in this thesis.

5.1 Selection Guide

Incremental $\Delta\Sigma$ converters main application area is the conversion of a constant signal. Here *constant* has two different meaning: the signal is sampled and held by a S/H circuit or the signal constant part is of interest, while random and/or periodic noise must be cancelled/averaged out.

Depending on the power- and area-consumption and the desired resolution, the family of incremental converters may be divided into two groups:

1. If only moderate resolution (8–12 bits) is needed, and the main requirement is to minimize the chip area and/or the power consumption, then a first-order incremental converter is usually optimal. Depending on other requirements (such as suppression of line-frequency noise, high-speed operation, etc.), different digital filter configurations may be used. Application areas of these converters includes but not limited to: array A/D conversion (e.g., in CMOS sensor arrays, digital cameras, mixed vector-vector multipliers, etc), distributed (battery-operated) sensor network elements, low-power microcontrollers, etc. Modulator and digital filter selection is discussed below in Sec. 5.2.

2. If the main goal is to achieve high resolution (high dynamics) as well as high speed, and circuit complexity (and thus area and power-consumption) is not critical, higher-order incremental converters may be eligible. Application area is high-precision instrumentation and measurement, seizmic application (though here power-consumption may be critical), etc. Again, there are several possibilities, depending on the

other requirements. High-order converter design examples are discussed below in Sec. 5.3.

5.2 First-order Converters

In the case of first-order converters, the $\Delta\Sigma$ modulator structure is given, and consists of a discrete-time integrator and a one-bit quantizer, realized with a comparator. The designer's choice is the digital filter which follows the modulator. According to Sec. 2.1.3 and 3.1, there are two efficient filtering method for first-order converters. The first is a simple digital integrator (realized as a counter), while the second is the use of two cascaded integrators. In this latter case, dither signal must be used to remove error peaks around zero input.

The first method is straightforward to implement, occupies very small chip area, as the integrator is a simple up-down counter, and it is capable of periodic noise suppression. However, for n_{bit}-bit resolution it requires $2^{n_{\text{bit}}}+1$ clock cycles, thus its output rate is very slow compared to its clock frequency.

The second method, using second-order digital filter with dither signal injected in the loop has the advantage of faster operation, but requires more complex digital circuit (two integrators, from which the first one may be realized as counter) and a dither signal generator.

Usually in a design the required resolution is specified. As it was derived in Chap. 3, for a given resolution, a first-order incremental converter requires

$$N_{1,1} = 2^{n_{\text{bit}}} + 1 \tag{5.1}$$

cycles (cf. Eqs. (2.14) and (3.1)) to achieve n_{bit}-bit resolution.

In the improved architecture (cf. Eq. (3.22)),

$$N_{1,2} \geq 3.9 \cdot 2^{\frac{2n_{\text{bit}}}{3}}. \tag{5.2}$$

For typical resolutions ($n_{\text{bit}} = 8$ to 16), Fig. 5.1 shows the required number of cycles. For compact illustration, the number of cycles are shown in a log-scale. Some examples are also tabulated in Tab. 5.1. One can see that the higher the required resolution, the more the benefits the second-order filter has. For example, while the reduction in the required number of cycles in the case of $n_{\text{bit}} = 10$ is about 1/3, it becomes 1/8 for 14- and 1/10 for 16-bit resolution.

In conclusion, if in a given design situation the most important factor is the smallest possible chip area, then a first-order incremental converter with a counter may give the optimum solution. If, on the other hand, chip area, power consumption and speed are all factors to be considered, a first-order converter with a second-order *CoI* filter and injected dither signal can be the solution. Note that different extensions of the first-order converter discussed in Sec. 2.2.1 may also be used to reduce the required number of cycles, but usually at the expense of additional analog hardware, except the method proposed in [Mulliken et al., 2002].

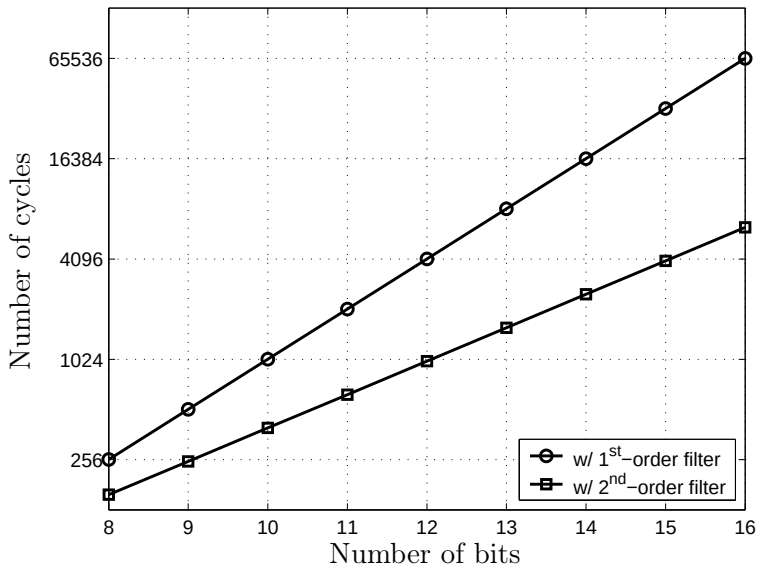


Figure 5.1: Required number of cycles in a first-order incremental converter with first- and second-order filter as a function of the specified resolution.

Table 5.1: Required number of cycles of the first-order incremental converter with first- and second-order filter

n_{bit}	$N_{1,1}$	$N_{1,2}$
8	257	158
10	1025	397
12	4097	999
14	16385	2516
16	65537	6340

Note that improved performance may be achieved by using a three-level quantizer, which has the following benefits:

- It requires almost no additional hardware
- The feedback DAC mismatch error can be eliminated by alternating the feedback capacitors
- It minimizes $k\mathrm{T}/C$ noise due to the feedback of zero
- It reduces the maximum internal quantization error, so less cycles are required for a given resolution
- Due to the smaller internal quantization error, smaller dither signal is required, thus, the input signal magnitude does not have to be limited so drastically.

5.3 Higher-order Converters

Using higher-order converters, it is possible to achieve $n_{\mathrm{bit}} = 16$, 18, 20 or even 24-bit resolution, within reasonable clock rate/sampling rate ratio. As it was discussed in the previous chapter, there are many architectural choices regarding to modulator type, modulator order, internal quantizer resolution, filter type, filter order, etc. These are organized in the next sections to help designers to select between the different trade-offs.

5.3.1 Design Considerations

Again, let us assume that the specification of the converter to be designed contains the required number of cycles and the main goal is to digitize an incoming dc signal (either sampled-and-held by an S/H circuit or continuously averaged by the converter). These specification along gives a wide range of possible solutions using higher-order converters. However, other specification details may narrow the possible choices. These are tabulated in Tab. 5.2.

The suggested $\Delta\Sigma$ modulator architecture is the Cascade-of-Integrators, Feed-forward (CIFF) architecture, with the input signal fed forward right to the input of the internal quantizer. This architecture was used in most of the theoretical discussions, since it has several benefits [Silva, 2004], which have already been discussed in previous chapters. Here a brief summary is given:

- the *STF* of the modulator is 1, i.e., the input signal is neither delayed, nor modified by the loop.
- The input signal is not processed by the integrators in the loop, thus, nonlinearity of the op-amps does not affect the input signal.
- As the input signal is not processed by the integrators, the signal swing in the integrators is much smaller, the required scaling is less severe, thus (i) conversion time

Table 5.2: Specifications and appropriate architectural solutions

Specification	Architecture
Low power- and area-consumption	second- or third-order 1-bit modulator w/ CoI digital filter (Sec. 3.2.3)
Possible lowest delay	CoI digital filter (Sec. 3.2.3)
Lowest number of cycles	CoI digital filter (Sec. 3.2.3)
Suppression of periodic noise	digital sinc-filter (Sec. 4.2)
Wide-range suppression of the line frequency	optimized sinc-filter (Sec. 4.2.3)
Suppression of 60 and 50Hz simultaneously	optimized sinc-filter (Sec. 4.2.3)
1-bit internal quantizer	stabilized *NTF* (Sec. 3.2.3)
Excellent linearity	multi-bit internal quantizer and feedback DAC w/ DWA (Sec. 4.3.7)
Rail-to-rail input signal	Input signal scaling circuit (Sec. 4.3.2)
Uniform output quantization error	Pure Differential *NTF* and same-order CoI filter (Sec. 3.2.1)

is faster, (ii) the capacitor ratio of the largest and smallest capacitor in the circuit is much less, the circuit is less sensitive to parasitic effects and (iii) the output quantization error is independent of the input signal.

- Only one feedback DAC is required.

Additionally, the following rules may help in the architectural decision:

- If the main goal is to provide a valid output signal with the lowest delay and lowest number of cycles, then a high-order (third- or even higher-order) modulator with the same order CoI filter is the right choice.

- If in addition to the fast operation, suppression of the line frequency is also required, then a multi-bit modulator with pure L_ath-order differential *NTF* and same-order sinc-filter may give the optimal solution. Note that the multi-bit feedback DAC must be incorporated with mismatch shaping algorithm.

- If the time-to-market has the highest priority, then traditional, one-bit $\Delta\Sigma$ loops followed by a same or higher-by-one order sinc-filter can be used. In this case, a classical $\Delta\Sigma$ modulator is used in transient mode. Required operation time can be estimated but must also be verified by simulation.

- If wide-range suppression of the line frequency is required, optimized sinc-filter may be used instead of classical sinc-filter.

If the available power budget is limited, it gives another design trade-off: the lower the number of analog stages in the loop, the lower its power-consumption, however, the higher the required number of cycles (i.e., either the clock rate or the conversion time) for a given

resolution. Similarly, the higher the number of levels in the internal quantizer, the lower the required number of cycles, but the higher the power consumption due to the additional analog and digital hardware.

For a quick overview, Tab. 5.3 compares the required number of cycles of several architectures for 16 and 20 bit resolution. In the following subsections some design examples are shown to evaluate the different trade-offs.

5.3.2 Modulators with Pure Differential Noise Transfer Function

Second-order Modulator with Second-order CoI Filter

A possible realization of a second-order modulator with two integrators (second-order CoI filter) is shown in Fig. 3.11 on p. 38. To achieve 16-bit resolution, the required number of cycles can be calculated by Eq. (3.33), which is repeated here for simplicity:

$$N \approx \frac{\sqrt{2} \cdot 2^{n_{\text{bit}}/2}}{\sqrt{U_{\max}(l-1)}}, \tag{5.3}$$

where N is the required number of cycles, n_{bit} is the required resolution in bits, $U_{\max}$ is the relative maximum input signal, and l is the number of levels in the internal quantizer.

Assuming $l = 5$, an input signal limited to $U_{\max} = 0.8$ and $n_{\text{bit}} = 16$-bit resolution, $N = 203$ is required. As discussed in Sec. 3.2.3, the sign of the output of the last integrator in cycle N must also be recognized to pick up an extra bit of resolution, which is lost during the final requantization of the output signal.

Third-order Modulator with Third-order Sinc-filter

$\Delta\Sigma$ modulators with pure differential *NTF* can also be used with digital sinc-filters for incremental conversion. This was analyzed in detail in Sec. 4.2.1. The main advantage of this architecture is that the conversion accuracy is independent of the exact distribution of the internal quantization error, if it is bounded by $\pm V_{\text{ref}}$. Nevertheless, to realize a stable third-order modulator with pure differential *NTF*, multi-bit internal quantizer and feedback DAC is required, and to minimize the DAC linearity error, dynamic element matching technology must be utilized, which increases the required chip-area and power.

To realize such a modulator, the model shown in Fig. 4.11 (p. 80) can be used. According to Lee's rule (see, e.g., [Norsworthy et al., 1997, Chap. 4]), and also verified by simulations, the internal quantizer must have at least $2^3 + 1 = 9$ levels to make the modulator stable. The required number of cycles for a given resolution can be calculated by Eq. (4.36), which is repeated here:

$$N = 3N_3, \tag{5.4}$$

where

$$N_3 = \sqrt[3]{\frac{2^{n_{\text{bit}}+3}}{U_{\max}(l-1)}}, \tag{5.5}$$

Table 5.3: Required number of cycles of different higher-order architectures for 16- and 20-bit resolution

Order of modulator	*NTF*	Type of digital filter	# of levels (l)	Input signal ($U_{\max}$)	Resolution (n_{bit})	Decimation ratio (N_i)	# of cycles (N)	Periodic noise suppression
1–1[a]	pure diff.	2nd-order CoI	2	1	16	N/A	362	No
2nd	pure diff.	2nd-order CoI	2	0.5	16	N/A	512	No
2nd	pure diff.	2nd-order CoI	5	0.8	16	N/A	203	No
3rd	pure diff.	3rd-order sinc	33	0.67	20	74	222	Yes
3rd	pure diff.	3rd-order sinc	9	0.5	20	128	384	Yes
3rd	stabilized	3rd-order CoI	2	0.67	20	N/A	468	No
3rd	stabilized	4th-order sinc	2	0.67	20	238	982	Yes

[a] 2nd-order MASH structure of [Robert and Deval, 1988]

where n_{bit} is the required resolution in bits, $U_{\max}$ is the input signal limit and l is the number of levels in the internal quantizer.

To achieve 20-bit resolution with such a configuration, assuming $U_{\max} = 0.67$ and $l = 33$, $N_3 = 74$, $N = 222$. Using a 9-level internal quantizer and $U_{\max} = 0.5$ yields to $N_3 = 128$, $N = 384$, which are very reasonable operation clock rates at the expense of a multi-bit internal quantizer.

5.3.3 One-bit CIFF Modulators with Stabilized Noise Transfer Function

If the designer wants to avoid the trouble with the implementation of a multi-bit internal quantizer and dynamic element matching circuit for the feedback DAC, classical single-bit modulator may also be used for high-precision conversion of dc input signals. The basic operation of this modulator was discussed in Sec. 3.2.3. Cascade-of-Integrators and sinc-filters following the modulator were addressed in Sec. 3.2.3 and Sec. 4.2.2, respectively.

Third-order Modulator with Third-order CoI-filter

One-bit modulator with same-order CoI filter may be used if the main goal is to get the digital output with the lowest possible delay. An example modulator structure is shown in Fig. 3.15 on p. 44. Repeating the results of Sec. 3.2.3, the required number of cycles of a general L_ath-order modulator can be calculated from the following equation (cf. Eq. (3.96)):

$$\prod_{i=0}^{L_a-1}(N-i) = \frac{2^{n_{\mathrm{bit}}} L_a!}{U_{\max}\left(\prod_{i=1}^{L_a-1} c_i\right) b}, \tag{5.6}$$

which simplifies to (cf. Eq. (3.95))

$$N = \mathrm{fix}\sqrt[3]{\frac{3!}{bc_1c_2}\frac{2^{n_{\mathrm{bit}}}}{U_{\max}}} + 2 \tag{5.7}$$

in third-order case, where n_{bit} is the resolution in bits, b and c_i are the scaling coefficients of the analog loop and $U_{\max}$ is the normalized input signal limit.

To achieve 20-bit resolution with a third-order one-bit modulator and same-order CoI digital filter, $N = 468$ is required, assuming $U_{\max} = 0.67$ and the scaling coefficients listed in Tab. 3.4. The required number of cycles (and thus the clock rate of the circuit) is very reasonable and can be easily realized.

Third-order Modulator with Fourth-order Sinc-filter

One-bit modulator with sinc-filter is the proposed solution, if periodic noise cancellation is required during conversion. Such a structure (with third-order sinc-filter) is shown in Fig. 4.17, on p. 85. According to the discussion in Sec. 4.2.2, the number of cycles for a given resolution can be estimated as

$$N = 4N_{4,p} + m, \tag{5.8}$$

where m is the length of the transient of the poles of the stabilized NTF, while

$$N_{4,p} > \sqrt[3.5]{\frac{2^{n_{\text{bit}}}\sqrt{20}}{bc_1c_2U_{\max}}\sqrt{\frac{\left(\sum_{i=1}^{m} w_d[i]\right)^2}{\sum_{i=1}^{m} w_d[i]^2}}}, \tag{5.9}$$

where n_{bit} is the required resolution in bits, b and c_i are scaling coefficients of the loop, $U_{\max}$ is the input signal limit and $w_d[k]$ is the impulse response of the stabilizer low-pass filter of the loop (cf. Eq. (4.71)).

In the case of a third-order CIFF modulator with $U_{\max} = 0.67$, scaling coefficients listed in Tab. 3.4 and Butterworth pole-configuration of Eqs. (4.51) and (4.52), $U_{\max}bc_1c_2 = 0.0618$ and $\left(\sum_{i=1}^{m} w_d[i]\right)^2 / \sum_{i=1}^{m} w_d[i]^2 = 7.34$, thus

$$N_{4,p} \approx 4.52 \cdot 2^{\frac{n_{\text{bit}}}{3.5}}, \tag{5.10}$$

from which $N_{4,p} = 238$, $N = 982$ is required to achieve 20-bit performance. Even though this value is the highest among the examples discussed in this section, it has the advantage of utilizing only one-bit modulator and suppression of periodic noise disturbances by a fourth-order sinc-filter.

5.4 Experimental Results

Based partly on the theoretical and simulation results of this thesis, an integrated circuit has been designed and fabricated by Microchip Technology, Inc., an American analog- and mixed-signal IC manufacturer. The chip is intended to be on the market later in 2005, thus its data sheet is not available yet. However, the following data were taken from the Web page of the Microchip Technology Masters Conference, July 21-24, 2004 [Microchip, 2004]:

The chip target resolution is 22 bits. Measurements on the chip shows that the effective/equivalent number of bits (ENOB) is around 21.6 bits, which indicates very good noise suppression. The chip consists of a one-bit third-order modulator and an optimized fourth-order sinc-filter, providing 120 dB suppression of the line frequency. The sinc-filter's decimation ratio is 512. Conversion rate is 15 Hz, and the zeros of the sinc-filter are at 60 Hz. The chip output contains an rms noise of 0.3 ppm, which is about 1.2 LSB at 22-bit resolution. The current drawn by the chip is typically 250 μA. The $\Delta\Sigma$ modulator uses a switched-capacitor circuit, operated with four non-overlapping clock phases.

The chip main application area is the conversion of wide dynamic range, low frequency signals, including (but not limited to) temperature measurements, weight scales, pressure sensors, and in general, battery-powered portable applications.

Chapter 6

Outlook

This thesis discussed the possible extensions of the first-order incremental converter [Robert et al., 1987], keeping most of its advantages while reducing its disadvantages, especially the high clock-rate (or slow conversion time) associated with the architecture. The main message of my contribution is that it is possible to use higher-order $\Delta\Sigma$ modulators for the conversion of dc signals, if the modulator is used in a repetitive manner. Several architectures have been developed and many theoretical questions have been answered during the research. However, further development based on the result of this thesis may be possible. The remaining tasks can be divided into two groups. The first group (Sec. 6.1) contains those problems, which have arisen during the research, but have not been fully answered due to their secondary importance. The second group of tasks (Sec. 6.2) mainly contains novel techniques and architectures, which may be integrated with the structures I have developed, to achieve even more efficient conversion of high-dynamics, low-frequency signals at the expense of more complex hardware and/or software.

6.1 Further Analysis of the Proposed Structures

One open question regarding to the introduced architectures is the analysis of the effects caused by the non-ideal behavior of the circuit elements of the converter. Even though the basic problems have been addressed and different algorithms or circuit techniques have been proposed to reduce the effect of these errors (cf. Sec. 4.3), exact derivation of the effect of circuit imperfections is missing in some cases. Additional analysis of the following error sources may be required to gain more insight into the operation of the converter:

- Op-amp imperfections (such as finite bandwidth, offset, noise, nonlinearity);
- Hysteresis and offset of the internal quantizer;
- Various errors of the feedback DAC (noise, mismatch error, etc.);
- Sensitivity to the jitter of the clock signal.

Another effect, which has not been analyzed in this work, is the effect of arbitrary input signal. During most of the derivations, it was assumed that the input signal is constant

(i.e., it does not change significantly during the conversion) or contains only such additive Gaussian noise or periodic noise disturbances, whose rms value is much smaller than the input range of the converter. However, analysis of the architecture with arbitrary input signals (large-scale noise, rapid changes in the input signal, etc.) should be carried out to prevent the converter from overflow and saturation errors. Even though this type of input signal is not expected during normal conversion sequence, power-up transients or switching of a multiplexer in front of the converter may generate such unwanted signals which should be detected before or during conversion.

6.2 Possible Future Architectures

One way to extend the results discussed in this thesis is to examine different modulator structures. This thesis focused on modulators with pure differential noise transfer function (cf. Sec. 3.2.1), and modulator realized with Cascade-of-Integrators, Feed-Forward (CIFF) architecture (cf. Sec. 3.2.3). However, there exist many other $\Delta\Sigma$ structures which may be used for conversion of dc signals in incremental mode, even though the developed structures are optimal in several aspects. Among the possible further research areas is the usage of continuous-time $\Delta\Sigma$ modulators in incremental (integrating) mode, since this thesis focused only on converters with discrete-time $\Delta\Sigma$ modulators implemented in switched-capacitor (SC) circuits. The theoretical results and all subsequent analysis was based on discrete-time operation, which may not be the right model for continuous-time circuits. Using continuous-time modulators may be advantageous, since typically they have better power-consumption and put less severe requirements on the op-amp parameters than their SC counterparts, even though they are more sensitive to the clock jitter of the circuit.

It was also shown in the thesis, that using digital sinc-filters at the output, it is possible to suppress periodic noise disturbances (cf. Sec. 4.2). However, in this case the conversion of the input signal is about L_a or $L_a + 1$ times longer than using the originally derived Cascade-of-Integrators filter, where L_a is the order of the analog modulator. This delay in the processing is caused by the transient of the sinc-filter, since its registers must be filled up with valid data to produce a correct output. One possible way to reduce this transient is to operate the $\Delta\Sigma$ converter continuously. This may be advantageous if the circuit is not used in multiplexed mode, but is used for continuous monitoring of a sensor signal (e.g., meteorological temperature or pressure measurement). However, if a $\Delta\Sigma$ converter is operated continuously, limit cycles may limit the achievable performance for input signals around zero and low-order fractions of the reference signal. Methods to eliminate these error sources (dithering, limit cycle observation, etc.) should be developed to achieve comparable performance to the proposed method.

Another possible way to enhance the conversion rate of the converter is to use non-linear decoding of the one-bit stream of the digital output. It was shown earlier that a $\Delta\Sigma$ modulated one-bit stream contains more information about the input signal then the signal reconstructed by linear (low-pass) filtering techniques (see, e.g., [Hein, 1995]). This technique is gaining more attention nowadays [Kim and Brooke, 2005] and may also be useful for incremental conversion. An alternative way for high-precision conversion of dc

signals to be investigated may be the usage of the time encoding machine (TEM), which is capable of error-free decoding of the input signal [Lazar and Toth, 2004].

Building on the results achieved in this thesis and using such enhanced techniques, it might be possible to further reduce the required cycles of operation for a given resolution, at the expense of more complex architecture.

Bibliography

Analog (2004). *AD77xx product family datasheets*, Analog Devices, Inc.
URL: http://www.analog.com.

Badmirowski, K. and Jackiewicz, B. (1998). Effects of noise dither signals on differential linearity and effective resolution of sigma-delta A/D converters in measuring applications, *Proc. of the 15th IEEE Instrumentation and Measurement Technology Conference*, Vol. 2, St. Paul, MN, USA, pp. 1229–1232.

Badmirowski, K. and Jackiewicz, B. (1999). Application of deterministic dither signals in digital voltmeter with sigma-delta oversampled A/D converter, *Proc. of the 16th IEEE Instrumentation and Measurement Technology Conference*, Vol. 3, Venice, Italy, pp. 1659–1662.

Baird, R. T. and Fiez, T. S. (1995). Linearity enhancement of multibit delta-sigma A/D and D/A converters using data weighted averaging, *IEEE Transactions on Circuits and Systems – II. Analog and Digital Signal Processing* **42**(12): 753–762.

Burr-Brown (2004). *ADS124x product family datasheets*, Burr-Brown (Texas Instruments).
URL: http://www.ti.com/.

Candy, J. C. (1974). A use of limit-cycle oscillations to obtain robust analog-to-digital converters, *IEEE Transactions on Communications* **22**(3): 298–305.

Candy, J. C. (1986). Decimation for sigma delta modulation, *IEEE Transactions on Communications* **34**(1): 72–76.

Candy, J. C. and Benjamin, O. J. (1981). The structure of quantization noise from sigma-delta modulation, *IEEE Transactions on Communications* **29**(9): 1316–23.

Candy, J. C., Ching, Y. C. and Alexander, D. S. (1976). Using triangularly weighted interpolation to get 13-bit PCM from a sigma-delta modulator, *IEEE Transactions on Communications* **24**(11): 1268–1275.

Cirrus (2004). *CS55xx product family datasheets*, Cirrus Logic, Inc.
URL: http://www.cirrus.com.

Enz, C. C. and Temes, G. C. (1996). Circuit techniques for reducing the effects of op-amp imperfections: Autozeroing, correlated double sampling, and chopper stabilization, *Proceedings of the IEEE* **84**(11): 1584–1614.

Gradshteyn, I. S. and Ryzhik, I. M. (1994). *Table of Integrals, Series, and Products*, corr. and enl. edn, Academic Press, New York.

Gray, R. M. (1989). Spectral analysis of quantization noise in a single-loop sigma-delta modulator with DC input, *IEEE Transactions on Communications* **37**(6): 588–99.

Gray, R. M. (1990). Quantization noise spectra, *IEEE Transactions on Information Theory* **36**(6): 1220–44.

Gray, R. M., Chou, W. and Wong, P. W. (1989). Quantization noise in single-loop sigma-delta modulation with sinusoidal inputs, *IEEE Transactions on Communications* **37**(9): 956–68.

Haigh, D. G. and Singh, B. (1983). A switching scheme for switched-capacitor filters, which reduces effect of parasitic capacitances associated with control terminals, *Proc. of the IEEE Int. Symp. on Circuits and Systems*, Vol. 2, pp. 586–89.

Hamadé, A. R. (1978). A single-chip all-MOS 8-bit A/D converter, *IEEE Journal of Solid-state Circuits* **13**(12): 785–791.

Harjani, R. and Lee, T. A. (1998). FRC: A method for extending the resolution of Nyquist rate coverters using oversampling, *IEEE Transactions on Circuits and Systems – II. Analog and Digital Signal Processing* **45**(4): 482–494.

Hein, S. (1995). A fast block-based nonlinear decoding algorithm for $\Sigma\Delta$ modulators, *IEEE Transactions on Signal Processing* **43**(6): 1360–1367.

Hogenauer, E. B. (1981). An economical class of digital filters for decimation and interpolation, *IEEE Transactions on Acoustics, Speech and Signal Processing* **29**(2): 155–162.

Inose, H., Yasuda, Y. and Murakami, J. (1962). A telemetering system by code modulation – Δ-Σ modulation, *IRE Trans. on Space Electron. Telemetry* **SET-8**(9): 204–9.

Jansson, C. (1995). A high-resolution, compact, and low-power ADC suitable for array implementation in standard CMOS, *IEEE Transactions on Circuits and Systems – I. Fundamental Theory and Applications* **42**(11): 904–912.

Johns, D. and Martin, K. (1997). *Analog Integrated Circuit Design*, John Wiley & Sons, Inc.

Johnston, J. (1991). New design techniques yield low power, high resolution delta-sigma and SAR ADCs for process control, medical, seismic and battery-powered applications, *Proceedings of the 1st International Conference on Analogue to Digital and Digital to Analogue Conversion*, Swansea, UK, pp. 118–123.

Kim, D. D. and Brooke, M. A. (2005). Iterative bound-based nonlinear decoding of delta-sigma encoded stream and offset calibration, *IEEE Transactions on Circuits and Systems I: Regular Papers*. Submitted for publication.

Lazar, A. A. and Toth, L. T. (2004). Perfect recovery and sensitivity analysis of time encoded bandlimited signals, *IEEE Transactions on Circuits and Systems I: Regular Papers* **51**(10): 2060–2073.

Linear (2004). *LTC24xx product family datasheets*, Linear Technology, Inc.
URL: http://www.linear.com.

Lyden, C. (1993). Single shot sigma delta analog to digital converter, *U.S. Patent 5,189,419*, University College Cork.

Lyden, C., Ugarte, C. A., Kornblum, J. and Yung, F. M. (1995). A single shot sigma delta analog to digital converter for multiplexed applications, *Proceedings of IEEE Custom Integrated Circuits Conference, CICC'95*, Santa Clara, CA, USA, pp. 203–206.

Márkus, J. (2003). Enhancing the resolution of incremental converters with dither, *Proceedings of the 10th PhD Mini-Symposium*, Budapest University of Technology and Economics, Department of Measurement and Information Systems, Budapest, Hungary, pp. 38–39.

Márkus, J., Silva, J. and Temes, G. C. (2001). 20-bit delta-sigma ADC system design progress report, *Technical report*, Oregon State University. 15 p.

Márkus, J., Silva, J. and Temes, G. C. (2003). Design theory of high-order incremental converters, *Proceedings of the IEEE International Symposium on Intelligent Signal Processing (WISP'2003)*, Budapest, Hungary, pp. 3–8.

Márkus, J., Silva, J. and Temes, G. C. (2004). Theory and applications of incremental delta-sigma converters, *IEEE Transactions on Circuits and Systems—I: Regular Papers* **51**(4): 678–690.

Microchip (2004). 840 DSM: Using a 22-bit delta sigma A/D converter, *Microchip Technology Master's Conference*, Scottsdale, Arizona, USA.
URL: http://techtrain.microchip.com/masters2004/downloads/classes/840/840.htm.

Mulliken, G., Adil, F., Cauwenberghs, G. and Genov, R. (2002). Delta-sigma algorithmic analog-to-digital conversion, *Proceedings of the IEEE International Symposium on Circuits and Systems (ISCAS'2002)*, Vol. 4, Scottsdale, Arizona, pp. 687–690.

Nakamura, J., Pain, B., Nomoto, T., Nakamura, T. and Fossum, E. R. (1997). On-focal-plane signal processing for current-mode active pixel sensors, *IEEE Transactions on Electron Devices* **44**(10): 1747–1758.

Norsworthy, S. R., Schreier, R. and Temes, G. C. (eds) (1997). *Delta-Sigma Data Converters – Theory, Design, and Simulation*, IEEE Press, Piscataway, NJ, USA.

Nys, O. J. A. P. and Dijkstra, E. (1993). On configurable oversampled A/D converters, *IEEE Journal of Solid-State Circuits* **28**(7): 736–742.

Oppenheim, A. V. and Schafer, R. W. (1975). *Digital Signal Processing*, Prentice-Hall, Inc., Englewood Cliffs, New Jersey.

Presti, L. L. (2000). Efficient modified-sinc filters for sigma-delta A/D converters, *IEEE Trans. on Circuits and Systems – II. Analog and Digital Signal Processing* **47**(11): 1204–1213.

Presti, L. L. and Akhdar, A. (1998). Efficient antialiasing decimation filter for $\Delta\Sigma$ converters, *Proc. of the 5th IEEE International Conference on Electronics, Circuits, and Systems (ICESC '98)*, Vol. 1, Lisboa, Portugal, pp. 367–370.

Robert, J. and Deval, P. (1988). A second-order high-resolution incremental A/D converter with offset and charge injection compensation, *IEEE Journal of Solid-State Circuits* **23**(3): 736–741.

Robert, J. and Valencic, V. (1985). Offset and charge injection compensation in an incremental analog-to-digital converter, *Proc. of the European Solid-state Circuits Conference*, Toulouse, France, pp. 45–48.

Robert, J., Temes, G. C., Valencic, V., Dessoulavy, R. and Deval, P. (1987). A 16-bit low-voltage A/D converter, *IEEE Journal of Solid-State Circuits* **22**(2): 157–163.

Rombouts, P., de Wukde, W. and Weyten, L. (2001). A 13.5-b 1.2-V micropower extended counting A/D converter, *IEEE Journal of Solid-State Circuits* **36**(2): 176–183.

Schreier, R. (1993). An empirical study of high-order single-bit delta-sigma modulators, *IEEE Trans. on Circuits and Systems – II. Analog and Digital Signal Processing* **40**(8): 461–466.

Schreier, R. (2004). *The Delta-Sigma Toolbox v6.0 (delsig)*. Software Toolbox and User's Manual,
URL: http://www.mathworks.com/matlabcentral/fileexchange/.

Schreier, R., Goodson, M. V. and Zhang, B. (1995). Proving stability of delta-sigma modulators using invariant sets, *Proc. of the IEEE Int. Symposium on Circuits and Systems, ISCAS'95*, Vol. 1, Seattle, WA, USA, pp. 633–36.

Schreier, R., Goodson, M. V. and Zhang, B. (1997). An algorithm for computing convex positively invariant sets for delta-sigma modulators, *IEEE Trans. on Circuits and Systems-II: Analog and Digital Signal Processing* **44**(1): 38–44.

Silva, J. B. (2004). *High-Performance Delta-Sigma Analog-to-Digital Converters*, PhD thesis, Oregon State University, School of Electrical Engineering and Computer Sciences, 97331 Corvallis, OR, USA.

Silva, J., Moon, U.-K. and Temes, G. C. (2004). Low-distortion delta-sigma topologies for MASH architectures, *Proc. of the International Symposium on Circuits and Systems, ISCAS'04*, Vol. 1, Vancouver, Canada, pp. I–1144–I–1147.

Silva, J., Moon, U.-K., Steensgaard, J. and Temes, G. C. (2001). A wideband low-distortion delta-sigma ADC topology, *Electronics Letters* **37**(12): 737–738.

Temes, G. C., Silva, J. and Márkus, J. (2004). *Switched Capacitor Signal Scaling Circuit*, Assignee: Microchip Technology Inc. US patent application, filed on March 23, 2004.

Thompson, C. D. and Bernadas, S. R. (1994). A digitally-corrected 20b delta-sigma modulator, *IEEE International Solid-State Circuits Conference, 1994. ISSCC'94*, San Francisco, CA, USA, pp. 194–195.

van de Plassche, R. J. (1978). A sigma-delta modulator as an A/D converter, *IEEE Transactions on Circuits and Systems* **25**(7): 510–514.

van de Plassche, R. J. (1994). *Integrated Analog-to-Digital and Digital-to-Analog Converters*, Kluwer Academic Publisher, London.

Wang, H. (1992). A geometrical view of $\Sigma\Delta$ modulations, *IEEE Trans. on Circuits and Systems-II: Analog and Digital Signal Processing* **39**(2): 402–5.

Weisstein, E. W. (2004). Central limit theorem, From MathWorld–A Wolfram Web Resource. URL: http://mathworld.wolfram.com/CentralLimitTheorem.html.

Wong, N. and Ng, T.-S. (2003). DC stability analysis of high-order, lowpass $\Sigma\Delta$ modulators with distinct unit circle NTF zeros, *IEEE Trans. on Circuits and Systems-II: Analog and Digital Signal Processing* **50**(1): 12–30.

Yufera, A. and Rueda, A. (1996). SI incremental A/D converter for IC sensor interfaces, *Proc. of the Joint IEEE Instrumentation and Measurement Technology Conference and IMEKO Technical Comittee*, Vol. 2, Brussels, Belgium, pp. 1029–1033.

Yufera, A. and Rueda, A. (1998). (SI)-I-2 first-order incremental A/D converter, *IEE Proceedings – Circuits Devices and Systems* **145**(2): 78–84.

Appendix A

Original Contributions

The new scientific statements concern the design theory of incremental $\Delta\Sigma$ A/D converters. The achieved results are collected into three statements.

Statement 1: I have modified the first-order incremental A/D converter by adding one more digital integrator at the output and injecting dither signal in front of the quantizer. I have analyzed the structure in detail and I have proven that it is more efficient than the original one.

The modified structure is shown in Fig. A.1. The discussion of this statement can be found in Sec. 3.1, on pp. 21–32. Comparative evaluation of the structure can be found in Sec. 5.2.

Statement 1.1: *I have shown that in the new structure much less cycles are required to achieve a given resolution at the expense of the added complexity. The required number of cycles (N) can be calculated as follows:*

$$N \geq 3.9 \cdot 2^{\frac{2n_{\mathrm{bit}}}{3}}, \tag{A.1}$$

where n_{bit} is the required resolution in bits.

Discussion of this statement can be found in Sec. 3.1.2, on pp. 25–29.

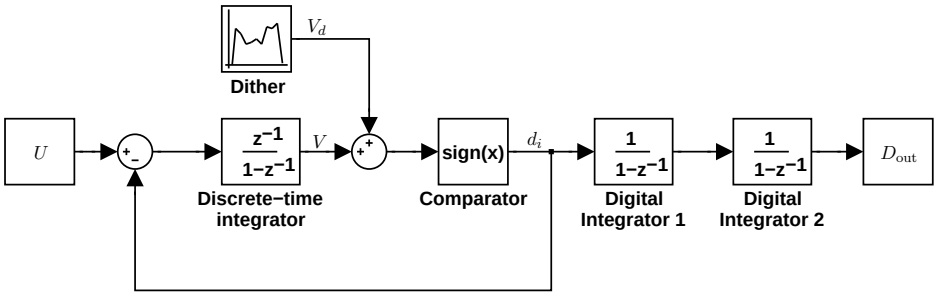


Figure A.1: First-order incremental converter with second-order digital filter and dither signal injected into the loop.

Statement 1.2: *I have derived the quantization error ($q[N]$) of the output signal for zero input signal analytically:*

$$q[N] = \pm\frac{1}{N+1}, \tag{A.2}$$

without dither signal, while

$$q[N] = \frac{2}{N+1}\frac{1}{N}\sum_{i=0}^{N/2}\mathrm{sign}(V_d[2i]), \tag{A.3}$$

with dither signal, where $V_d[2i]$ is the $(2i)$th sample of the injected dither signal. Based on this result, I have shown that the quantization error around zero fulfills the specified accuracy if dither signal is used in the loop.

Discussion of this statement can be found in Sec. 3.1.2, on pp. 25–29.

Since the quantizer in the loop may saturate for large input signals due to the presence of the dither signal, either the input signal or the dither signal must be limited. The next statement is about efficient methods to limit these signals.

Statement 1.3: *To avoid saturation problems, I have suggested using either efficient scaling of the input signal, or three-level quantizer in the loop to decrease the amplitude of the required dither signal.*

Discussion of this statement can be found in Sec. 3.1.2, on pp. 25–29, in Sec. 4.3.2 on pp. 95–96 and in Sec. 5.2, on pp. 102–104.

The following publications contains proofs and discussions of Statement 1: [Márkus et al., 2004; Márkus, 2003].

Statement 2: I have extended the original first-order incremental converter to higher-order $\Delta\Sigma$ modulators and showed that the new architecture requires significantly less cycles to achieve a given resolution.

I proposed two different extensions. The first extension can be used for modulators with pure differential noise transfer function ($NTF = (1-z^{-1})^{L_a}$, where L_a is the order of the modulator) shown in Fig. A.2, while the other extension applies to modulators which have stabilized noise transfer function ($NTF = (1-z^{-1})^{L_a}/D(z)$), and are realized by the Cascaded-Integrators, Feed-Forward (CIFF) architecture, with a feed-forward path from the input signal to the input of the internal quantizer (Fig. A.3).

Detailed discussion about the extensions can be found in Sec. 3.2.1 on pp. 32–41 and in Sec. 3.2.3 on pp. 43–59. Comparative evaluation of the different structures is available in Sec. 5.3.

Statement 2.1: *I have derived that in the case of pure differential NTF, the quantization error of the converter is linearly related to that of the internal quantizer in the Nth cycle, if the digital filter following the modulator is an L_ath-order Cascade-of-Integrators filter, where L_a is the order of the analog modulator (cf. Fig. A.2). The required number of cycles*

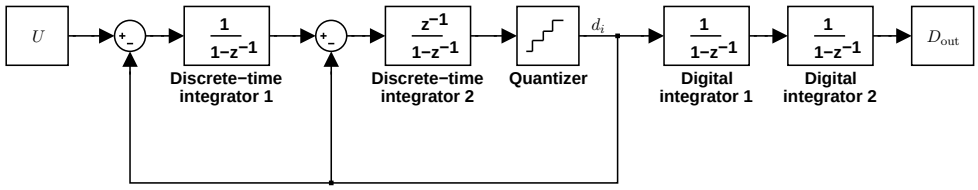


Figure A.2: A possible realization of an incremental converter consists of a second-order modulator with pure differential noise transfer function ($NTF = (1-z^{-1})^2$) and same-order digital Cascade-of-Integrators filter.

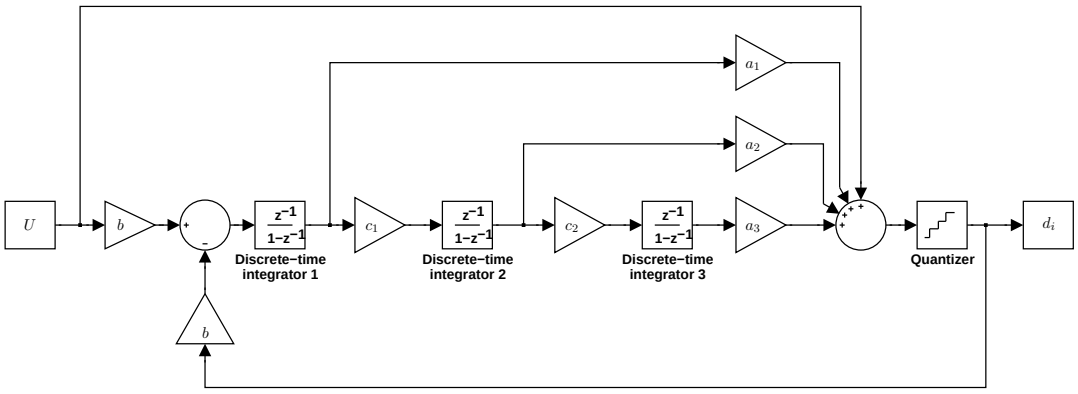


Figure A.3: Third-order Cascaded-Integrators, Feed-Forward (CIFF) $\Delta\Sigma$ modulator architecture, with additional feed-forward path from the input signal to the input of the internal quantizer.

(N) to achieve a given resolution (n_{bit}) can be calculated from the following equation:

$$\prod_{i=0}^{L_a-1}(N+i) = \frac{2^{n_{\mathrm{bit}}} L_a!}{(l-1)U_{\max}}, \tag{A.4}$$

where L_a is the order of the modulator, $U_{\max}$ is the normalized maximum input signal and l is the number of levels in the internal quantizer.

Additional advantage of the proposed structure is that its final quantization error can be modeled as a stochastic signal with uniform distribution, while the distribution of the quantization error of further proposed structures is approximately Gaussian.

I have also examined the case when the output digital filter consists of $L_a + 1$ digital integrators. In this case, however, similarly to the first-order one (cf. statement 1), dither signal is also required, and the structure is efficient for high resolutions only.

Derivation of these statements can be found in Sec. 3.2.1 on pp. 32–37.

Statement 2.2: *I have derived that using CIFF architecture, the digital output may be*

calculated without knowing the value of the coefficients in the analog loop, i.e.,

$$D_{\text{out}} = \frac{1}{\binom{N}{L_a}} \underbrace{\sum_{k_{L_a}=0}^{N-1} \sum_{k_{L_a-1}=0}^{k_{L_a}-1} \cdots \sum_{k_1=0}^{k_2-1}}_{L_a} d_{k_1}, \tag{A.5}$$

where D_{out} is the digital output, N is the number of cycles the converter operates, L_a is the order of the analog modulator and d_{k_1} is the output of the modulator in the k_1th cycle.

Based on this result, I have proven that the quantization error of the converter is linearly related to the output of the last analog integrator in cycle N, if the digital filter following the modulator is an L_ath-order Cascade-of-Integrators filter. The required number of cycles (N) for a given resolution (n_{bit}) can be calculated from the following equation:

$$\prod_{i=0}^{L_a-1} (N-i) = \frac{2^{n_{\text{bit}}} L_a!}{U_{\max} \left(\prod_{i=1}^{L_a-1} c_i \right) b} \tag{A.6}$$

where L_a is the analog modulator order, $U_{\max}$ is the normalized maximum input signal, and c_i and b are scaling coefficients in the loop (cf. Fig. A.3).

Derivation of these statements can be found in Sec. 3.2.3 on pp. 43–47 and pp. 53–55.

Statement 2.3: *I have proven that by realizing a modulator with pure differential NTF using CIFF architecture, the following relationship is true for the quantization error of the converter (q), the output of the last integrator (V_{L_a}) and the error of the internal quantizer (ε):*

$$-V_{L_a}[N + L_a] = \varepsilon[N] = 2V_{\text{ref}} q[N], \tag{A.7}$$

i.e., in this case the two extensions are equivalent.

Detailed analysis of the statement can be found in Sec. 3.2.1, on pp. 37–41, in Sec. 3.2.3, on pp. 47–49, and summarized in Sec. 3.2.4 on pp. 59–59.

Statement 2.4: *I have proven that the introduced converter structure is capable of reducing input noise significantly. If the input noise variance is σ_g^2, then its contribution to the output noise variance (σ_y^2) in second-order case*

$$\sigma_y^2 < \frac{4}{3} \frac{\sigma_g^2}{N}, \tag{A.8}$$

while in third-order case

$$\sigma_y^2 < \frac{9}{5} \frac{\sigma_g^2}{N}, \tag{A.9}$$

where N is the number of cycles the converter operates.

Detailed analysis about the noise suppression can be found in Sec. 4.1.1, on pp. 61–67.

The following publications discuss the results of Statement 2: [Márkus et al., 2004; Márkus et al., 2003; Temes et al., 2004; Márkus et al., 2001].

Statement 3: I have designed optimal higher-order digital sinc-filters for higher-order incremental $\Delta\Sigma$ converters for suppression of periodic noise disturbances.

Detailed discussion of the proposed filter design methods can be found in Sec. 4.2, on pp. 69–90. Comparative evaluation of the different filter structures is available in Sec. 5.3.

Statement 3.1: *I have shown that either same-order ($L_d = L_a$, where L_d is the order of the digital* sinc*-filter, while L_a is the order of the analog modulator) or higher-by-one order ($L_d = L_a + 1$) digital* sinc*-filter gives an optimum between periodic noise suppression and the required number of cycles.*

Details of the statement can be found in Sec. 4.2.1, on pp. 71–79.

Statement 3.2: *I have derived the required number of cycles (N) for converters with pure differential NTF. In the case of third-order modulator and third-order* sinc*-filter,*

$$N = 3N_3, \tag{A.10}$$

where

$$N_3 = \sqrt[3]{\frac{2^{n_{\mathrm{bit}}+3}}{U_{\max}(l-1)}}, \tag{A.11}$$

where n_{bit} is the resolution, $U_{\max}$ is the normalized maximum input signal, while l is the number of levels in the internal quantizer.

In the case of third-order modulator and fourth-order digital filter

$$N = 4N_4, \tag{A.12}$$

where

$$\sqrt[3.5]{\frac{3\sqrt{6}\cdot 2^{n_{\mathrm{bit}}}}{U_{\max}(l-1)}} < N_4 < N_3. \tag{A.13}$$

This statement has been derived in Sec. 4.2.1, on pp. 74–79.

Statement 3.3: *I have derived the required number of cycles for a given resolution for 1-bit, stabilized CIFF modulators. The required number of cycles (N) in the case of third-order modulator and third-order* sinc*-filter (assuming that $N_{3,p} \gg m$, which is fulfilled if the resolution is higher than 12 bit):*

$$N = 3N_{3,p} + m, \tag{A.14}$$

where m is the length of the transient of the stabilizer poles in the system, while

$$N_{3,p} = \sqrt[3]{\frac{2^{n_{\mathrm{bit}}+3}}{bc_1c_2U_{\max}}}, \tag{A.15}$$

where n_{bit} is the resolution, $U_{\max}$ is the maximum input signal, while b and c_i are scaling coefficients in the loop.

In the case of third-order modulator and fourth-order sinc*-filter, applying the same conditions,*

$$N = 4N_{4,p} + m, \tag{A.16}$$

where m is the length of the transient of the stabilizer poles in the system, while

$$N_{4,p} > \sqrt[3.5]{\frac{2^{n_{\mathrm{bit}}}\sqrt{20}}{bc_1c_2U_{\max}}\sqrt{\frac{\left(\sum_{i=1}^{m} w_d[i]\right)^2}{\sum_{i=1}^{m} w_d[i]^2}}}, \tag{A.17}$$

where $w_d[i]$ is the ith sample of the impulse response of the stabilizer poles of the system, while the other parameters are the same as in the previous case.

This statement has been derived in Sec. 4.2.2, on pp. 84–90.

Publications regarding to Statement 3 are: [Márkus et al., 2004; Márkus et al., 2003; Márkus et al., 2001].

Appendix B

List of Publications

Papers in Periodicals in English

[1] J. Márkus and I. Kollár, "Standard environment for the sine wave test of ADC's," *Special Issue on ADC Modeling and Testing of the Measurement Journal*, vol. 31, no. 4, pp. 261–69, June 2002.

[2] T. Z. Bilau, T. Megyeri, A. Sárhegyi, J. Márkus, and I. Kollár, "Four-parameter fitting of sine wave testing results: Iteration and convergence," *Computer Standards and Interfaces*, vol. 26, no. 1, pp. 51–56, Jan. 2004.

published also: in *Proceedings of the 4th International Conference on Advanced A/D and D/A Conversion Techniques and their Applications; 7th European Workshop on ADC Modelling and Testing (ADDA-EWADC 2002)*, Prague, Czech Republic, 26–28 June 2002, pp. 185–190.

[3] J. Márkus, J. Silva, and G. C. Temes, "Theory and applications of incremental delta-sigma converters," *IEEE Transactions on Circuits and Systems—I: Regular Papers*, vol. 51, no. 4, pp. 678–690, Apr. 2004.

[4] J. Márkus and I. Kollár, "On the monotonicity and linearity of ideal radix-based A/D converters," *IEEE Transactions on Instrumentation and Measurement*, 2005, accepted for publication.

Papers in International Conference Proceedings

[5] I. Kollár and J. Márkus, "Sine wave test of ADC's: Means for international comparison," in *Proceedings of the IMEKO TC4 5th European Workshop on ADC Modelling and Testing (EWADC)*, P. Daponte, L. Michaeli, M. N. Durakbasa, and A. Afjehi-Sadat, Eds., Vienna, Austria, 24–26 June 2000, pp. 211–16.

[6] J. Márkus and I. Kollár, "Standard framework for IEEE-STD-1241 in MATLAB," in *Proceedings of the IEEE Instrumentation and Measurement Technology Conference*

(IMTC'2001), vol. 3, Budapest Convention Centre, Budapest, Hungary, 21–23 May 2001, pp. 1847–52.

[7] J. Márkus, J. Silva, and G. C. Temes, "Design theory of high-order incremental converters," in *Proceedings of the IEEE International Symposium on Intelligent Signal Processing (WISP'2003)*, Budapest, Hungary, 4–6 Sept. 2003, pp. 3–8.

[8] J. Márkus and I. Kollár, "On the monotonicity and linearity of ideal radix-based A/D converters," in *Proceedings of the IEEE Instrumentation and Measurement Technology Conference (IMTC'2004)*, vol. 1, Como, Italy, 18–20 May 2004, pp. 696–701.

Foreign Patent

[9] G. C. Temes, J. Silva, and J. Márkus, *Switched Capacitor Signal Scaling Circuit*, Assignee: Microchip Technology Inc., Mar. 2004, US patent application, filed on March 23, 2004.

Presentations at Hungarian Conferences

[10] J. Márkus, "Sine wave test of analog to digital converters," in *Proceedings of the 7th PhD Mini-Symposium.* Budapest, Hungary: Budapest University of Technology and Economics, Department of Measurement and Information Systems, 27–28 Jan. 2000, pp. 24–25.

[11] J. Márkus, "A MATLAB tool to use and test the ADC-standard," in *Proceedings of the 8th PhD Mini-Symposium.* Budapest, Hungary: Budapest University of Technology and Economics, Department of Measurement and Information Systems, 31 Jan. – 1 Feb. 2001, pp. 26–27.

[12] J. Márkus, "Enhancing the resolution of incremental converters with dither," in *Proceedings of the 10th PhD Mini-Symposium.* Budapest, Hungary: Budapest University of Technology and Economics, Department of Measurement and Information Systems, 4–5 Feb. 2003, pp. 38–39.

[13] J. Márkus, "Monotonicity of digitally calibrated cyclic A/D converters," in *Proceedings of the 11th PhD Mini-Symposium.* Budapest, Hungary: Budapest University of Technology and Economics, Department of Measurement and Information Systems, 3–4 Feb. 2004, pp. 8–9, (Best presenter's award in the category of third-year PhD students).

Other Works

Technical Reports

[14] J. Márkus, J. Silva, and G. C. Temes, "20-bit delta-sigma ADC system design progress report," Oregon State University, Tech. Rep., Aug. 2001, 15 p.

[15] J. Márkus, "A comparison of DAC-error calibration algorithms," Oregon State University, Tech. Rep., Sept. 2002, 34 p.

Software with User's Manual

[16] J. Márkus, *ADC Test Data Evaluation Program for Matlab*, Budapest University of Technology and Economics, Department of Measurement and Information Systems, URL: http://www.mit.bme.hu/projects/adctest/, 2002.

Publications Not Directly Related to the PhD Thesis

Papers in Periodicals

[17] János Márkus and Gabor C. Temes, "An efficient delta-sigma ADC architecture for low oversampling ratios," *IEEE Transactions on Circuits and Systems I. – Special Issue on Advances on Analog-to-Digital and Digital-to-Analog Converters*, vol. 51, no. 1, pp. 63–71, Jan. 2004.

[18] Balázs Bank, János Márkus, Attila Nagy, and László Sujbert, "Signal- and physics-based sound synthesis of musical instruments," *Periodica Polytechnica, Ser. Electrical Engineering*, vol. 47, no. 3–4, pp. 269–295, 2004.

Papers in Conference Proceedings

[19] János Márkus and László Sujbert, "Signal model based synthesis of the sound of organ pipes," in *Proceedings of the International Békésy Centenary Conference on Hearing and Related Sciences*, Budapest, Hungary, 24–26 June 1999, pp. 194–199.

[20] János Márkus, "Signal model based synthesis of the sound of organ pipes," in *Proceedings of the 9th Conference and Exhibition on Television and Audio Technologies (TV 2000 Conference)*, Thermal Hotel Helia, Budapest, Hungary, 23–25 May 2000, pp. 151–57, abstract in English and Hungarian.

[21] János Márkus, "An efficient delta-sigma noise shaping architecture," in *Proceedings of the 9th PhD Mini-Symposium*, Budapest, Hungary, 4–5 Feb. 2002, Budapest University of Technology and Economics, Department of Measurement and Information Systems, pp. 52–53.

[22] János Márkus and Gabor C. Temes, "An efficient delta-sigma noise-shaping architecture for wideband applications," in *Proceedings of the 4th International Conference on Advanced A/D and D/A Conversion Techniques and their Applications; 7th European Workshop on ADC Modelling and Testing (ADDA-EWADC 2002)*, Prague, Czech Republic, 26–28 June 2002, pp. 35–38.

[23] János Márkus, "Signal model based synthesis of the sound of organ pipes," in *Végzős konferencia (Conference for graduated students)*, Budapest, Hungary, 28 Apr. 1999, pp. 6–11, in Hungarian.

S.D.G.